
GAZİ ÜNİVERSİTESİ
Gazi
1926

HİZMET SEKTÖRÜ İÇİN TWİTTER (X) ANALİTİĞİ: EV VE İŞYERİNE YEMEK SİPARİŞİ ÜZERİNE BİR UYGULAMA

Yıldırım GÜNEŞ

DOKTORA TEZİ
TEDARİK VE LOJİSTİK YÖNETİMİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

OCAK 2024

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Yıldırım GÜNEŞ
03/01/2024

HİZMET SEKTÖRÜ İÇİN TWİTTER (X) ANALİTİĞİ: EV VE İŞYERİNE YEMEK SİPARİŞİ ÜZERİNE BİR UYGULAMA

(Doktora Tezi)

Yıldırım GÜNEŞ

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Ocak 2024

ÖZET

Her konuyla ilgili görüşün dolaşımda olduğu Twitter (yeni ismiyle "X"), işletmelerin kendi ürün ve hizmetlerinin kalitesi konusunda da müşteri görüşlerine ulaşabilecekleri bir kaynaktır. Bu kaynaktan faydalanabilmek için, amaca uygun analiz yöntemleri kullanılmalıdır. Twitter (X) sahip olduğu uygulama özellikleri ve üst verisi ile tanımlayıcı, içerik ve ağ analizi olmak üzere üç tür analizin yapılmasına imkan verir. Çalışmada bu üç tür analiz, eve / işyerine çevrimiçi yemek siparişi konusunda bir yıllık (1 Ocak-31 Aralık 2020) döneme ait toplam 35 428 tweet cümlesi üzerinde uygulanmıştır. Pandemi dönemini de kapsayan veri setinin analizleri sonucunda, pandemi dönemi ve sonrasına yönelik analiz ve değerlendirmeler yapılmıştır. Çalışmada veri setinin tanımlayıcı analizleri kapsamında Twitter (X) Veri Seti İçeriğinin Tanımlayıcı Analiz ile Keşfi olarak adlandırılan bir model oluşturulmuştur. Bu safhada özellikle veri setinin analiz sonuçlarını olumsuz etkileyen, reklam gibi içeriklerin ayıklanmasına yönelik yöntemler önerilmiştir. İçerik analizi bölümünde veri setinin, sözlük tabanlı ve makine öğrenmesi yöntemlerinin hibrit kullanımı ile duygu sınıflandırması ve ürün-hizmet sistemleri sınıflandırması yapılmıştır. Söz konusu sınıflandırmalar için, temel ve güçlendirilmiş olarak adlandırılan duygu ve ürün-hizmet sistemleri sözlükleri geliştirilmiştir. Alana özel bu sözlüklerin hazırlanması için, metin sınıflandırmanın önemli problemlerinden yüksek boyutluluğun indirgenmesi ve ayırt ediciliği yüksek özelliklerin seçilmesine yönelik öneriler geliştirilmiştir. Üçüncü analiz türünde, çevrimiçi sosyal ağ yapısı üzerinde etkili kullanıcıların ve ağdaki toplulukların tespiti ve analizi yapılmıştır. Ağ analizinde, ağın yerel ve küresel özelliklerini kullanan merkezilik derecelerinin hibrit kullanımı için yöntem önerilmiştir. Çalışma, çevrimiçi yemek siparişi alanındaki veri setinin aynı anda üç farklı analiz türü ile ele alındığı Türkçe literatürdeki ilk çalışmadır. Çalışmanın, her üç analize yönelik önerileri, alana özel hazırlanan sınıflandırma sözlük ve modelleri ile literatüre katkı sağlayacağı düşünülmektedir.

Bilim Kodu : 92429

Anahtar Kelimeler : Twitter (X) Analizi, Sosyal Ağ Analizi, Sözlük Tabanlı Duygu Analizi, Tanımlayıcı Analiz, Çevrimiçi Yemek Siparişi

Sayfa Adedi : 212

Danışman : Dr. Öğr. Üyesi Murat ARIKAN

TWITTER (X) ANALYTICS FOR THE SERVICE SECTOR: AN APPLICATION ON ORDERING MEAL TO HOME AND OFFICES

(Ph.D. Thesis)

Yıldırım GÜNEŞ

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

January 2024

ABSTRACT

Twitter (now called "X"), where opinions on every subject are in circulation, is a source where businesses can access customer opinions on the quality of their products and services. In order to benefit from this resource, appropriate analysis methods should be used. With its application features and metadata, Twitter (X) allows for three types of analysis: descriptive, content, and network analysis. In this study, these three types of analysis are applied to a total of 35 428 tweet sentences for a one-year period (January 1–December 31, 2020) on the subject of online food ordering from home or the workplace. As a result of the analysis of the data set, including the pandemic period, analyses and evaluations are made for the pandemic period and beyond. Within the scope of the descriptive analysis of the dataset, a model called Exploration of Twitter (X) Dataset Content with Descriptive Analysis is constructed in the study. At this stage, methods have been proposed to remove content, such as advertisements, that negatively affects the analysis results of the data set. In the content analysis section, sentiment classification and product-service systems classification of the dataset are performed with the hybrid use of lexicon-based and machine learning methods. For these classifications, lexicons of sentiment and product-service systems, called basic and boosted, have been developed. In order to prepare these domain-specific lexicons, suggestions for reducing high dimensionality, which is one of the important problems of text classification, and for selecting highly discriminative features were developed. In the third type of analysis, the identification and analysis of users and communities in the network that have an impact on the online social network structure is performed. In network analysis, a method for the hybrid use of degrees of centrality using local and global properties of the network is proposed. This is the first study in the Turkish literature where the online food ordering dataset is analyzed with three different types of analysis at the same time. It is thought that the study will contribute to the literature with its suggestions for all three analyses and the classification dictionaries and models prepared specifically for the field.

Science Code : 92429

Key Words : Twitter (X) Analysis, Social Network Analysis , Lexicon-based Sentiment Analysis, Descriptive Analysis, Online Meal Ordering

Page Number. : 212

Supervisor : Assist. Prof. Dr. Murat ARIKAN

TEŞEKKÜR

Tez konusu seçiminden itibaren her aşamada titiz, detaylı düzeltme ve doğru yönlendirmeleri ile akademik çalışma sürecinin bilimsel ve teknik yönleri ile tüm gerekliliklerini tam manasıyla öğrenmemi sağlayan danışmanım Dr. Öğr. Üyesi Murat ARIKAN'a; tez çalışması başlangıcından itibaren her aşamada desteğini gördüğüm ve tez izleme sürecinde çalışmaya farklı bakış açıları kazandıran tez izleme komite üyesi Prof. Dr. Serpil EROL'a; tez sunumlarımda olumlu eleştirileri ve katkıları ile çalışmayı zenginleştiren ve motivasyonumu artıran tez izleme komite üyesi Prof. Dr. Süleyman ERSÖZ'e; tez sürecinin seminer döneminde ilk makalemin yazım ve yayımlanmasında destek veren Prof. Dr. Saliha ÇETİNYOKUŞ'a, çalışmanın veri setini oluşturan tweet cümlelerinin sınıflandırmasını titizlikle yapan arkadaşlarım Dr. Esat UĞURLU, Erol Aydın AYKAR, Serhan BARDAKÇI ve ailelerine; yaptığımız fikir alışverişlerinde özellikle programlama ve analiz konusunda zihnimde eksik kalan noktaları tamamlamama katkı sağlayan bilişim dünyasından Kaan KARAHAN ve Doğukan DEMİR'e; veri analizleri için kullandığım Python programlamayı öğrendiğim Vahit KESKİN'e; doktora sürecinin yönetiminde karşılaştığım her türlü idari sorunda kolaylık sağlayan, yol gösteren Gazi Üniversitesi ve Fen Bilimleri Enstitüsü Endüstri Mühendisliği idari personeline; uzun soluklu bu çalışma sürecimde her türlü yorgunluğuna rağmen, birlikte olduğumuz tüm zamanlarda olduğu gibi beni destekleme konusunda en ufak bir tereddüt yaşamayan ve doktora sürecinde yaşadığım mutluluk ve heyecanımı ilk paylaştığım, veri seti sınıflandırmasında da görev alan sevgili eşim Ayşe Asena GÜNEŞ'e; zamanlarından çalarak doktora çalışmama ortak ettiğim çocuklarım Bahar, Orhan ve Çınar'ıma, her fırsatta desteklerini ifade eden aile büyüklerime ve dostlarıma teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZET iv

ABSTRACT v

TEŞEKKÜR vi

İÇİNDEKİLER vii

ÇİZELGELERİN LİSTESİ x

ŞEKİLLERİN LİSTESİ xii

SİMGELER VE KISALTMALAR xvi

1. GİRİŞ 1

2. TWITTER (X) PAYLAŞIMLARININ TANIMLAYICI, İÇERİK / METİN VE AĞ ANALİZ YÖNTEMLERİ İLE ANALİZİ VE LİTERATÜR TARAMASI 9

2.1. Twitter (X) mesajları, üst verisi, Twitter (X)'dan veri çekme yöntemleri 11

2.2. Anahtar kelime ve diyez etiketi tanımlama, veri setinin oluşturulması 14

2.3. Tanımlayıcı analiz ve literatür 15

2.4. Metin sınıflandırma, içerik / metin analizi, duygu analizi ve literatür 18

2.4.1. Metin sınıflandırma ve duygu analizi 20

2.4.2. Metin sınıflandırma ve duygu analizi safhaları 26

2.4.3. Modelleme ve sınıflandırma algoritmaları 34

2.4.4. Modellerin performans ölçümü 41

2.4.5. Metin sınıflandırma ve duygu analizi konusunda literatür 44

2.4.6. Metin tabanlı sosyal medya uygulaması Twitter (X) ve yiyecek sektöründeki Twitter (X) metinlerinin sınıflandırmasına yönelik literatür 51

2.4.7. Çevrimiçi alışveriş, yemek siparişi ve sosyal medya 59

2.5. Ağ analizi ve literatür 63

Sayfa

2.5.1. Sosyal medya ağ analizleri (SMAA), Twitter (X) ağ analizi, ağ analizine yönelik literatür 64

2.5.2. Düğüm/aktör seviyesi analiz 71

2.5.3. Ağ seviyesi analiz, topluluk analizi 72

2.5.4. Ağ analizi için kullanılan algoritmalar ve gösterim şekilleri 72

3. ÇEVRİMİÇİ YEMEK SİPARİŞİ KONUSUNDAKİ TWEETLERİN TANIMLAYICI, METİN VE AĞ ANALİZİ YÖNTEMLERİ İLE ANALİZİNDEN ELDE EDİLEN BULGULAR VE YORUMLAR 75

3.1. Anahtar kelime ve etiket tanımlama, koleksiyonun oluşturulması 76

3.1.1. Anahtar kelime / etiket kelimelerini belirleme 76

3.1.2. Koleksiyonun oluşturulması, tanımlayıcı analiz safhasında kullanılacak TVİTA modeli 78

3.2. Tanımlayıcı analiz 85

3.2.1. Tweet istatistikleri 86

3.2.2. Tweetlerdeki kelime ve diyez etiket istatistikleri 90

3.2.3. Tweetlerdeki kullanıcı istatistikleri 94

3.2.4. Tweetlerin URL istatistikleri 99

3.2.5. Tweetlerin zaman grafikleri 100

3.2.6. Tweetlerin coğrafi konumlarına göre dağılımı 108

3.2.7. Tweetlerde kullanılan platform bilgilerinin dağılımı 111

3.2.8. Tanımlayıcı analizle elde edilen bulgular 112

3.3. İçerik/Metin Analizi 116

3.3.1. Koleksiyon içindeki sınıf dağılımları 119

3.3.2. Kelime frekans analizi 120

3.3.3. Sınıflandırmalarda uygulanan yöntem ve önerilen modeller 122

3.3.3.1. Duygu sınıflandırması 126

Sayfa

3.3.3.2. Ürün ve hizmet sistemleri sınıflandırması 127

3.3.4. Önerilen sınıflandırma modellerin ikinci veri seti üzerinde kullanılması 129

3.3.5. İçerik/metin analizi ile elde edilen bulgular 132

3.4. Ağ Analizi 134

3.4.1. Twitter (X) kullanıcılarının merkezilik durumları 139

3.4.2. Ağ içindeki toplulukların tespiti 147

3.4.3. Ağ yapısı için kullanılan diğer algoritmalar 150

3.4.4. Koleksiyonun ağ analizden elde edilen bulgular 154

4. SONUÇ VE ÖNERİLER 159

4.1. Sonuçlar 159

4.2. Müteakip çalışmalar için öneriler 160

KAYNAKLAR 163

EKLER 193

EK-1. Durak kelime listesi 194

EK-2. Tohum kelime listesi 195

EK-3. Alana özel (Çevrimiçi yemek siparişi) sözlüklerin hazırlığında kullanılan yöntemler 196

EK-4. Kelimelerin temsili için gruplandırılan kelimeler ve güçlendirilmiş sözlükler ... 198

EK-5. Reklam içerikli tweetlerin ayıklanması için kullanılan reklam kelimeleri sözlüğü ... 202

EK-6. Twitter (X) kullanıcılarının konum bilgisi yerine yazdıkları ifadeler ve özellikleri 204

EK-7. Koleksiyon sınıflandırması için değerlendiricilere ön bilgilendirme notu 207

EK-8. 66 Twitter (X) kullanıcısı için altı merkezilik ölçütüne göre merkezilik değerleri 211

ÖZGEÇMİŞ 212

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge 2.1. Karışıklık matrisi 42

Çizelge 2.2. Karışıklık matrisinde kullanılan ifadelerin açıklamaları 42

Çizelge 2.3. Yiyecek sektörüne ilişkin literatür özeti-1 53

Çizelge 2.4. Yiyecek sektörüne ilişkin literatür özeti-2 54

Çizelge 2.5. Network X kütüphanesi ağ analiz algoritmaları..... 73

Çizelge 2.6. Ağ yapıları gösterim şekilleri 74

Çizelge 2.7. NetworkX ve Graphviz kütüphanelerinin ağ grafik düzenleri 74

Çizelge 3.1. Veri seti oluşturma ve ayıklama işlemleri 79

Çizelge 3.2. Tüm kullanıcıların tweetlerindeki diyez etiket frekansları..... 91

Çizelge 3.3. Yalnızca potansiyel müşteriler tarafından kullanılan diyez etiket frekansları. 92

Çizelge 3.4. Ürün hizmet sistemleri sınıflandırması için alt konu başlıkları 118

Çizelge 3.5. Alana özel sözlük kelimelerinin koleksiyondaki frekansları..... 120

Çizelge 3.6. Güçlendirilmiş sözlük için kelime temsili örneği..... 124

Çizelge 3.7. Temel ve güçlendirilmiş sözlük kelimelerinin ağırlıklandırılmış değerleri arasındaki farkın bir örnek ile gösterimi..... 125

Çizelge 3.8. Duygu ve ürün-hizmet sistemleri sınıflandırması önerilen modellerin performans değerleri..... 129

Çizelge 3.9. Ağ analizinde kullanılan Twitter (X) kullanıcı numara ve isimleri 136

Çizelge 3.10. Komşuluk matrisinden bir kesit..... 138

Çizelge 3.11. Twitter (X) (66 kullanıcı) kullanıcılarının merkezilik değerler tablosundan bir kesit 145

Çizelge 3.12. Merkezilik değerlerine göre ilk on kullanıcı sıralaması 146

Çizelge 3.13. Kullanıcıların ağırlıklı puan, ortalama ve medyan değerlerine göre sıralaması 146

Çizelge 3.14. 66 Kullanıcıların ortalama ve medyan değerlere göre benzerlik karşılaştırması..... 147

Çizelge **Sayfa**

Çizelge 3.15. 66 Twitter (X) kullanıcısından Girvan-Newman algoritması ile tespit edilen topluluklar 148

Çizelge 3.16. 66 Twitter (X) kullanıcısının etiket yayılım algoritmasına göre yer aldığı topluluklar 149

Çizelge 3.17. Kullanıcıların kümeleme katsayıları 151

Çizelge 3.18. Twitter (X) kullanıcılarının Jaccard katsayıları 152

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 2.1. Büyük veri yönetim aşamaları..... 10

Şekil 2.2. Üç kademeli Twitter (X) analitik çerçevesi 11

Şekil 2.3. Metin sınıflandırma yöntemleri..... 21

Şekil 2.4. Türkçe Twitter (X) mesajlarının duygu analizi için örnek bir sistem modeli. 23

Şekil 2.5. Metin önişlem adımları 26

Şekil 2.6. Dokümanların temsil edilmesi işlemleri ve matris yapının oluşturulması..... 30

Şekil 2.7. Model geçerlilik test yöntemleri 33

Şekil 2.8. Makine öğrenme modelleri 35

Şekil 2.9. Sınıflandırma ve regresyon algoritmaları..... 39

Şekil 2.10. Performans ölçümü için kullanılan ölçütler 44

Şekil 2.11. Duygu analizi ve kategorik sınıflandırma. 50

Şekil 2.12. Twitter (X) analitik çerçevesi.. 56

Şekil 2.13. Müşteri odaklı Twitter (X) veri analitik yaklaşımı. 59

Şekil 2.14. Twitter (X) uygulama özelliklerinin oluşturduğu ağ yapıları 69

Şekil 2.15. Ağ analiz seviyeleri..... 71

Şekil 3.1. Veri havuzundan ayıklanan veri oranları 84

Şekil 3.2. Twitter(X)veriseti içeriğinin tanımlayıcı analizle keşfi için önerilen model . 85

Şekil 3.3. Tweetlerin aldıkları retweet frekans sıralaması..... 86

Şekil 3.4. RT'lenme açısından etkin ilk 25 tweet metni..... 87

Şekil 3.5. Favoriye eklenme/beğeni alma frekans sıralaması..... 89

Şekil 3.6. Favoriye alınma/beğeni açısından etkin ilk 25 tweet metni..... 90

Şekil 3.7. Tüm kullanıcıların kullandıkları diyez etiket frekanslarının kelime bulutu... 92

Şekil **Sayfa**

Şekil 3.8. Yalnızca potansiyel müşterilerin kullandıkları diyez etiket frekanslarının kelime bulutu 93

Şekil 3.9. Özgün içerikli tweet frekans sıralaması 95

Şekil 3.10. En etkili kullanıcılar. 97

Şekil 3.11. Tweetleri en çok RT alarak görünürlük elde etmiş kullanıcılar. 98

Şekil 3.12. Favoriye eklenme/beğeni alma sayıları açısından kullanıcı sıralaması..... 98

Şekil 3.13. URL bulundurma sayılarına göre kullanıcı sıralaması. 99

Şekil 3.14. Tweetlerin aylara göre akışı-zaman grafiği. 101

Şekil 3.15. Tweetlerin mevsimlere göre akışı-zaman grafiği. 102

Şekil 3.16. Haftalık tweet paylaşım frekansları-pandemi öncesi. 103

Şekil 3.17. Günlük tweet paylaşım frekansları-pandemi öncesi. 103

Şekil 3.18. Haftalık tweet paylaşım frekansları-pandemi dönemi..... 104

Şekil 3.19. Günlük tweet paylaşım frekansları-pandemi dönemi..... 104

Şekil 3.20. Saat dilimlerine göre tweet paylaşım frekansları-pandemi öncesi. 105

Şekil 3.21. Saat dilimlerine göre tweet paylaşım frekansları-pandemi dönemi. 105

Şekil 3.22. En çok tweet paylaşımı yapılan tarihler. 106

Şekil 3.23. Resmî tatil günleri tweet frekansları. 107

Şekil 3.24. En çok tweet paylaşımı yapılan günler. 108

Şekil 3.25. Tweetlerin CKB'ne göre dağılımı. 109

Şekil 3.26. Coğrafi konum bilgisi olan tweetlerin illere göre dağılımı 110

Şekil 3.27. Tweet gönderilerinin il nüfuslarına göre oranları..... 111

Şekil 3.28. Tweet gönderilerinin paylaşımında kullanılan platformların dağılımı..... 112

Şekil 3.29. Çevrimiçi yemek siparişine ilişkin tweet metinlerinin sınıflandırılması için önerilen model..... 117

Şekil 3.30. Veri seti duygu sınıfları dağılımı. 119

Şekil **Sayfa**

Şekil 3.31. Veri seti ürün hizmet sistemleri sınıf dağılımı. 120

Şekil 3.32. Alana özel sözlük kelimelerinin koleksiyondaki frekans kelime bulutu..... 121

Şekil 3.33. Alana özel sözlük kelimelerinin koleksiyondaki frekanslarına göre kelime bulutu-ayıklama sonrası..... 122

Şekil 3.34. Temel ve güçlendirilmiş sözlüklerin duygu sınıflandırma performansları. . 127

Şekil 3.35. Temel ve güçlendirilmiş sözlüklerin ürün-hizmet sistemleri sınıflandırma performansları 128

Şekil 3.36. İkinci veri setinin en iyi modellerle sınıflandırması sonrasında ortaya çıkan sınıf oranları 130

Şekil 3.37. İkinci koleksiyonun sınıflandırılmasından sonra hizmet alanında öne çıkan negatif ve pozitif konular..... 130

Şekil 3.38. İkinci koleksiyonun sınıflandırılmasından sonra ürün alanında öne çıkan negatif ve pozitif konular..... 131

Şekil 3.39. Kullanıcılar arasındaki iletişim-bahsetme tweeti. 138

Şekil 3.40. 66 Twitter (X) kullanıcısının derece merkeziliği 140

Şekil 3.41. 66 Twitter (X) kullanıcısının arasındalık merkeziliği. 142

Şekil 3.42. 66 Twitter (X) kullanıcısının yakınlık merkeziliği 142

Şekil 3.43. 66 Twitter (X) kullanıcısının özvektör merkeziliği..... 143

Şekil 3.44. 66 Twitter (X) kullanıcısının sayfa sıralaması. 144

Şekil 3.45. 66 Twitter (X) kullanıcısının Katz merkeziliği. 145

Şekil 3.46. 66 Twitter (X) kullanıcısının Girvan-Newman algoritmasına göre oluşturduğu topluluklar..... 148

Şekil 3.47. 66 Twitter (X) kullanıcısının etiket yayılım algoritmasına göre oluşturduğu topluluklar..... 149

Şekil 3.48. 66 Twitter (X) kullanıcısının bağlı bileşenler algoritmasına göre birbirleri ile kurdukları bağlar..... 152

Şekil 3.49. 66 Twitter (X) kullanıcısından Harel-Koren algoritmasına göre ortaya çıkan topluluklar 153

Şekil **Sayfa**

Şekil 3.50. 66 Twitter (X) kullanıcısından Fruchterman-Reingold algoritmasına göre ortaya çıkan topluluklar 154

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış kısaltmalar, açıklamaları ile aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
ABSA	Özellik/Hedef Tabanlı Duygu Analizi (Aspect-Based Sentiment Analysis)
AdaBoost	Uyarlayıcı Güçlendirme (Adaptive Boosting)
API	Uygulama Programlama Arayüzü (Application Programming Interface)
ARFF	Öznitelik-İlişki Dosya Biçimi (Attribute-Relation File Format)
AU-ROC	Alıcı İşlem Karakteristikleri Eğrisi Altındaki Alan (Area Under Receiver Operating Characteristic)
BV	Büyük Veri
BVA	Büyük Veri Analitiği/Analizi
CART	Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees)
CFS	Korelasyon Tabanlı Özellik Seçimi **(**Correlation-based Feature Selection)
CKB	Coğrafi Konum Bilgisi
CSV	Virgülle Ayrılan Değerler (Comma-Seperated Values)
CV	Sayım Vektörü (Count Vector)
ÇYS	Çevrimiçi Yemek Siparişi
DDİ	Doğal Dil İşleme (Natural Language Process-NLP)
DN	Doğru Negatif (True Negative-TN)
DTM	Doküman Terim Matrisi
DP	Doğru Pozitif (True Positive-TP)

Kısaltmalar	Açıklamalar
EDBC	Etkili Mesafe Tabanlı Merkezilik (Effective Distance-Based Centrality)
e-WOM	Elektronik Ağızdan Ağıza Pazarlama (Electronic Word of Mouth)
GB	Gradyan Güçlendirme (Gradient Boosting)
GBM	Gradyan Güçlendirme/Artırma Makineleri (Gradient Boosting Machine)
GDS	Güçlendirilmiş Duygu Sözlüğü
GÜHSS	Güçlendirilmiş Ürün-Hizmet Sistemleri Sözlüğü
ICC	Sınıflararası Korelasyon Katsayısı (Interclass Correlation Coefficient)
ICF	Ters Koleksiyon/Sınıf Frekansı (Inverse Corpus/Class Frequency)
IDF	Ters Doküman Frekansı **(**Inverse Document Frequency)
ID3	İteratif İkilikçi Ağaç (Iterative Dichotomiser-3)
IGM	Bilgi Kazanım Metodu (Information Gain Method)
İG	İyileştirme / Geliştirme
JSON	Javascript Nesne Gösterimi (Java (JavaScript Object Notation)
KNN	K-En Yakın Komşu (K-Nearest Neighbour-KNN)
LightGBM	Hafif Gradyan Güçlendirme/Artırma Makineleri (LightGBM)
LR	Lojistik Regresyon (Logistic Regression)
MLPC	Çok Katmanlı Algılayıcı Sınıflandırıcılar (Multi Layer Perceptron Classifier)
NB	Naive Bayes Algoritması
NLTK	Doğa Dil Araç Takımı (Natural Language Toolkit)

Kısaltmalar	Açıklamalar
NLP	Doğal Dil İşleme Süreci (Natural Language Process)
NNCI	Ağ Düğüm Merkezilik Indeksi (Network Node Centrality Index)
RANSAC	Rastgele Örneklem Uzlaşması (Random Sample Concensus)
RF	Rastgele / Rassal Orman (Random Forest)
POS Tag	Kelime Görev Etiketi (Part of Speech Tag)
RT	Retweet
SAA	Sosyal Ağ Analizi
SMA	Sosyal Medya Analizi
SMAA	Sosyal Medya Ağ Analizi
SMOTE	Azınlık Sınıf Sentetik Örnekleme Tekniği (Synthetic Minority Over Sampling Technique)
STC	Sözlü Türkçe Derlemi (STC-Spoken Turkish Corpus)
SVM	Destekçi Vektör Makinesi (Support Vector Machine)
TS Corpus	Türkçe Derlem (Turkish Corpus)
TA	Twitter (X) Analitiği/Analizi
TAA	Twitter (X) Ağ Analizi
TAD	Terim Ağırlıklandırma Değeri
TDA	Terim Ayırt Ediciliği (Term Discrimination Ability)
TDS	Temel Duygu Sözlüğü
TF	Terim Frekansı (Term Frequency)
TVİTA	Twitter (X) Verisi İçerik ve Tanımlayıcı Analiz
TXT	Yazı (Text)
TÜHSS	Temel Ürün-Hizmet Sistemleri Sözlüğü
XLS	Microsoft Excel Spreadsheet

Kısaltmalar	Açıklamalar
XGBoost / XGB	Aşırı Gradyan Güçlendirme (eXtreme Gradient Boosting)
UMM	Unigram Karışım Modeli (Unigram Mixture Model)
URL	Tekdüzen Kaynak Bulucu (Uniformed Resource Locator)
VADER	Değerliğe Duyarlı Sözlük ve Duygu Muhakemesi
YN	Yanlış Negatif (False Negative-FN)
YP	Yanlış Positif (False Positive-FP)

1. GİRİŞ

Sosyal medya, tüketicilerin kullandıkları her türlü ürün ve hizmet konusunda serbest bir şekilde fikir beyan edip sohbet edebildikleri ya da şikâyet ve memnuniyetlerini doğrudan üreticilere iletebildikleri web tabanlı araçlar, platformlar sunmaktadır. Son dönemde işletmeler, bu platformlarda paylaşılan ürün ve hizmetlere ilişkin tüketici görüşlerini, analitik yöntemlerden yararlanarak analiz etmekte ve iyileştirme ve geliştirme (İG) için kaynak olarak kullanmaktadırlar. Bir ürün/hizmet üreten işletmeler; İG için telefonla, anketle, çevrimiçi uygulamalarla kullanıcılarından (müşterilerinden) geri bildirim alma çabası içinde iken; kullanıcılar herhangi bir ürün/hizmete ait görüşlerini serbest ve istekli bir şekilde sosyal medya platformları ile paylaşmaktadırlar. Bu paylaşımlar, yeniden paylaşımlarla hızlı bir şekilde çok geniş kullanıcı kitlesine ulaşabilmektedir [1]. Facebook, Youtube, Twitter (X)[1], Instagram gibi müşteri görüşleri ile dolu bu kaynaklardan çekilen ham veriler, analizler sonucunda bilgiye dönüştürülerek daha iyi ürün, daha iyi hizmet olarak günlük hayatımıza girerler. Twitter (X), paylaşılan fikirlere her an yenilerinin eklendiği, yeni eklenen tweetlerle milyonlarcasının da güncelliğini yitirdiği bir platformdur. 2022 yılı itibari ile aylık ortalama olarak 368.4 milyon aktif kullanıcının olduğu dikkate alındığında, Twitter (X) [2]'ın her bir ürün/hizmete ilişkin veriye ulaşma konusunda zengin bir kaynak olabileceği anlaşılmaktadır. Twitter (X)'ın finans, sağlık, enformasyon, gazetecilik, politika, pazarlama, iletişim ve psikoloji gibi birçok alanda etkin olarak kullanıldığı görülmektedir [3]. Bununla birlikte Twitter (X) Analitiği / Analizi (TA)'nin, literatürde özellikle ticari ürün ve hizmete ilişkin konularda kullanımının düşük düzeyde olduğu görülmektedir. Bu nedenle Twitter (X) verilerinin analizlerinde kullanılan araç ve yöntemlerin sektörel olarak geliştirilmesine olan ihtiyaç devam etmektedir.

Metinsel içerikli paylaşım platformu olan Twitter (X) verilerinin, metin sınıflandırma ve duygu analizine tabii tutulması konusunda önemli problemlerin olduğu görülmektedir. Bu problemlerden biri, Twitter (X) gibi sosyal paylaşım platformlarından elde edilen veriler içinde, *gürültü* olarak isimlendirilen reklam ve propaganda içerikli verilerin analiz sonuçlarını etkileme potansiyelidir. Literatürde veri setleri içindeki gürültünün, analiz sonuçlarını olumsuz etkilemeyeceği kabul edilerek çalışmalar yürütülmektedir. Veri seti içindeki gürültünün %100 ayıklanması mümkün olmasa da, son dönemde reklam paylaşımı

[1] Twitter uygulamasının ismi, 24 Temmuz 2023 tarihinden itibaren "X" olarak değiştirilmiştir.

açısından sosyal medya platformlarının kullanımındaki artışın, analiz sonuçlarının yönünü değiştirebileceği düşünülmektedir. Ayrıca Twitter (X)'ın propaganda maksatlı ve yönlendirilmiş paylaşımlarla kamuoyunu etkilediği son dönemlerde sıklıkla gündeme getirilmektedir. Bu nedenle Twitter (X) verilerinden elde edilen analiz sonuçlarının yönlendirilmiş olup olmadığının tespit edilmesi ve bu maksatla kullanılabilecek araç ve yöntemlerin geliştirilmesine ihtiyaç olduğu görülmektedir.

Twitter (X) analizi konusunda dikkat çeken önemli bir başka problem ise, Twitter (X)'ın genellikle siyasi ve güncel konulara ilişkin analizler için kullanılırken, sektörel/alana yönelik çalışmalar konusunda zayıf kalmasıdır. Bu husus özellikle Türkçe literatür açısından daha fazla dikkat çekmektedir. Twitter (X) metinlerinin sınıflandırılması ve duygu analizi konusunda çeviri yöntemi, genel nitelikli hazır sözlükler ya da İngilizce metinsel içeriklerin kullanıldığı görülmektedir. Bu durum, dilin özellikle sektörel, alana yönelik metin sınıflandırma problemlerinin çözümü için kullanılan araç ve yöntemler konusunda zayıf kalmasına neden olmaktadır. Öte yandan metin sınıflandırma ve duygu analizinde, özellikle Twitter (X)'da jargon ya da gündelik dil kullanımı nedeniyle, metinleri temsil kabiliyeti yüksek öznitelik (değişken olarak kullanılacak alana özel kelimelerin) seçimi konusunda güçlükler yaşanmaktadır ve bu konuda yeni yöntemler geliştirilmesine ihtiyaç duyulmaktadır.

Bu çalışma için Twitter (X) analizleri konusunda Türkçe literatürde yer almayan "çevrimiçi yemek siparişi" konusu seçilmiştir. Çalışmanın amacı, Twitter (X) kullanıcıları arasındaki "çevrimiçi yemek siparişi" alanına özel sohbetlerden, İG çalışmalarına kaynaklık edecek yeni fikirlerin *tanımlayıcı analiz* araçları ile nasıl elde edilebileceğini göstermek, bunu diğer alanlara da uygulanabilir bir model ile sunmak, veri setlerinin yönlendirilmiş etkilerden kurtarılarak analiz edilebilmesi için yöntemler geliştirmek, Twitter (X)'dan elde edilen metinsel veri setlerinin sınıflandırılmasında ve duygu analizinde kullanılabilecek yeni yöntem ve *metin sınıflandırma modelleri* geliştirmek, veri setlerinin doğru yorumlanabilmesine *ağ analizi* ile yapılabilecek katkıları ortaya koymak ve Türkçe literatürün bu konularda sektörel/alan esaslı gelişimine katkı sağlamaktır.

Twitter (X) paylaşımlarından oluşan veri setlerinde yapılan analizlere ilişkin literatür incelendiğinde çok farklı konularda çalışmalar yapıldığı görülmektedir. Choudhary vd. [4] Arap Baharı ile 2011 yılında Mısır devrimi sırasında yaşanan olayların Twitter (X)'a

yansımalarını içeren 800 000 tweeti, Çetin ve Amasyalı [5] tweet metinleri üzerindeki duygu analiz çalışmasında 12 000 tweetten oluşan bir veri setini, Li vd. [6] Fukuşima nükleer santral kazasına ait tweetlerden 1520 adetini, Meral ve Diri [7] farklı konulardaki algı analizine yönelik çalışmalarında Haziran-Temmuz 2013 dönemine ait 8 321 adet tweeti, Çevik vd. [8] Toyota'nın gaz pedalındaki bir sorun ile ilgili 2010 yılında paylaşılan tweetlerden 5000 tanesini, Akgül vd. [9] dört aylık bir süreci kapsayan toplam 6 800 tweeti, Koçak vd. [10] hayayolu pazarına yönelik Nisan-Mayıs 2016 dönemine ait 8 762 tweeti, Onan [11] bir aylık bir dönemde topladığı 10 600 tweetten oluşan veri setini, Albayrak vd. [12] sosyal medyaya yönelik veri analiz çalışmasında 10-12 Ekim 2017 dönemine ait toplam 12 739 tweeti, Çoban ve Tümüklü-Özyer [13] Twitter (X) duygu analizi çalışmasında toplam 20 000 tweeti, Taşdelen [14] mültecileri konu alan çalışmasında 1545 tweeti, Akar vd. [15] İstanbul Sözleşmesi'ne yönelik tepkileri ölçmek için Nisan-Kasım 2020 dönemine ait 252 tweeti, Ainapure vd. [16] Mart-Mayıs 2020 dönemine ait ve Covid-19 ve aşılamaya ilişkin 298 791 tweetten oluşan iki veri setini, Qi ve Shabrina [17] sözlük ve makine öğrenmesi tabanlı sınıflandırmaların karşılaştırmalı çalışmasında Ocak-Temmuz 2021 dönemine ait 77 332 tweetten oluşan veri setini incelemiştir.

Literatürde sınıflandırma ve duygu analizini içeren Tweet metinleri üzerinde yiyecek sektörüne yönelik olarak yapılmış çalışmalar şunlardır: Hingle vd. [18] yemek yeme alışkanlıkları, Chae [3] tedarik zinciri içindeki çalışan ve yöneticilerin sohbet içerikleri, Park vd. [19] Çin, Japon, Kore ve Tayland restoran algıları, Mishra ve Singh [1] et tedarik zincirindeki atık kategorileri, Singh vd. [20] sığır et ürünlerindeki memnuniyetsizlikler, El-Khchine, vd. [21] tavuk ürünlerine ilişkin konularda sınıflandırma algoritmaları da kullanılarak çalışmalar yapmış ve model önerileri geliştirmişlerdir.

Bu çalışmada daha önce Türkçe literatürde çalışılmamış "çevrimiçi yemek siparişi (ÇYS)" gibi doğrudan ticari bir ürün/hizmet konusu seçilerek, bu alana yönelik sektörel çalışmalardaki boşlukların doldurulmasına ve alanın gelişimine önemli katkı sağlanacağı öngörülmektedir. Bu maksatla ÇYS konusunda 1 Ocak -31 Aralık 2020 tarih aralığına ait 35 428 adet tweetten oluşan bir veri seti oluşturulmuştur. Söz konusu verinin Twitter (X)'dan çekilmesi için *yemek sipariş*, *yemeksepeti sipariş*, *döner sipariş*, *lahmacun sipariş*, *hamburger sipariş*, *pide sipariş* olarak altı kelime grubu kullanılmıştır. Veri çekme işlemi için açık kaynak kodlu ve ücretli Twitonomy uygulaması, analizlerde MS Excel ve Python programlama dili kullanılmıştır.

Çalışmada, hazır bir uygulama olan ücretli Twitonomy uygulaması ile Twitter (X) Uygulama Programlama Arayüzünün (Application Programming Interface-API) uyguladığı %1'lik, her seferinde son bir aya ait ve 3 200 adet tweet çekebilme sınırları içinde kalınarak ve kamuya açık tweetlerden bir yıllık veri seti oluşturulmuştur [22]. Bu şekilde oluşturulan %1'lik veri setinin, literatürdeki diğer çalışmalarda olduğu gibi tüm veriyi temsil yeteneği olduğu kabul edilerek çalışma yapılmıştır. Çalışma kapsamında yaptığımız ön kabullerden biri de veri setinden %100 olarak ayıklanması mümkün olmayan reklam ya da propaganda maksatlı tweetlerin, analiz sonuçlarını etkilemeyeceği hususudur [19]. Bununla birlikte bu çalışmada literatürden farklı olarak, analizlerin belirli aşamalarında reklam ve propaganda içerikli tweetlerin ayıklanması için model ve yöntemler önerilmiştir.

Çalışmada kapsamında; (i) tweetlerin üst verilerinin (meta veri) ve tanımlayıcı analiz yöntemlerinin kullanıldığı, bu analiz sonuçları üzerinde yapılan sondajlamalarla veri setinin gündem konularının belirlendiği, *Twitter (X) Verisi İçerik ve Tanımlayıcı Analiz (TVİTA)* modeli önerisi getirilmiş, (ii) ileriki çalışmalarda kullanılabilecek, *alana özel reklam kelimeleri sözlüğü* ve alana özel *temel ve güçlendirilmiş sınıflandırma sözlükleri* oluşturulmuş, bu kapsamda dört farklı alan sözlüğü geliştirilmiş, (iii) bu sözlükler kullanılarak ürün/hizmet ve duygu sınıflandırmaları için performansı en yüksek modeller belirlenerek sınıflandırma modelleri oluşturulmuş, (iv) metin analizlerinin önemli problemlerinden *yüksek boyutluluk ve özellik seçimi* konusunda, sözlükteki kelimelerin gruplandırılması (aggregation) ve grupların tek bir kelime ile temsil edilmesi şeklinde uygulanan *bir boyut indirgeme ve dönüştürme yöntemi* geliştirilmiş, gruplandırma işlemi için eş anlamlı olan ya da aynı anlamı ifade etmek için kullanılan kelimeler, aynı anlamdaki jargon kelimelerin bir araya getirilmesi şeklinde uygulanmış, uygulanan özellik azaltma yöntemi ile model performanslarında artış sağlanmış, (v) tüm tweetler için üç terim ağırlıklandırma yöntemine göre sözlükte yer alan kelimelerin ağırlık puanları hesaplanmış, bu puanlar terim doküman matris hücrelerine yerleştirilmiş ve matris yapı kullanılarak altı sınıflandırma algoritması ile tweetlerin sınıfları belirlenmiş, algoritmalarla belirlenen sınıflar değerlendiriciler tarafından belirlenen gerçek sınıflarla karşılaştırılarak performans ölçümü yapılmış, (vi) uygulanan bu yöntemlerle çevrimiçi yemek siparişi konusunda sözlük tabanlı yaklaşım ve sınıflandırma algoritmaları birlikte kullanılarak duygu ve ürün-hizmet sistemleri konusunda *hibrit bir yaklaşımla sınıflandırma modeli* oluşturulmuş, (vii) veri seti analiz sonuçlarının doğru şekilde değerlendirilmesini sağlayan, etkili aktörleri ortaya çıkaran ağ analiz yöntemlerinin, ticari ürün/hizmet içerikli Twitter (X) veri setlerinde nasıl

kullanılabileceği gösterilmiş, (viii) ağ içindeki etkili aktörlerin tespiti için, ağın yerel ve küresel özelliklerini dikkate alan merkezilik derecelerinin ağırlıklandırılmış ortalamasının kullanımı önerilmiştir.

Çalışma kapsamında tanımlayıcı analiz, tweet cümlelerinin içeriğinin keşfedilmesi için içerik analizi (metin analizi) ve Twitter (X) kullanıcıları arasındaki ilişkilerin tespit edilmesi için ağ analizi birlikte kullanılarak, hem bu analizlerin birbirini tamamlayıcı özellikleri incelenmiş hem de her üç analiz türü için yeni yöntemler önerilmiştir. Çalışmanın ileriki bölümleri şu şekilde düzenlenmiştir. İkinci bölümde söz konusu analiz türleri, bu analizlerde kullanılan yöntem ve araçlar ile konuya ilişkin literatür ayrı başlıklar halinde incelenmiştir. Üçüncü bölümde; bu çalışmada geliştirilen yöntem ve araçlarla elde edilen analiz sonuçları, bulgular yorumlanmış ve literatürdeki çalışmalarla karşılaştırılmıştır. Dördüncü bölümde bu çalışmada önerilen yöntemler ve bu yöntemlerin gelecek çalışmalara yönelik kullanım alanlarına ilişkin fikirlere yer verilmiştir.

Tanımlar

Tezin anlaşılırlığını artırmak ve farklı alanlarda kullanımı olup metin madenciliği konusunda özel anlama sahip terimlerin doğru yorumlanmasını sağlamak amacı ile aşağıdaki tanımların yazılmasına ihtiyaç duyulmuştur.

Aykırı Değer Tespiti: Veri seti içindeki gözlemlerden, *normal* kabul edilen bir gözlem profilinin oluşturularak, bu profil özellikleri dışında kalan gözlemlerin aykırı değer olarak ayrıştırılmasını sağlayan yöntemdir [23].

Doküman: Özellikle Python gibi programlama dillerinde, veri tabanları ile yapılan çalışmalar kapsamında, tek bir kayıt, bireysel bir veri/veri birimi doküman olarak nitelendirilir [24]. Bu çalışma kapsamında doküman terimi, veri setindeki üst verisi ile birlikte her bir tweet cümlesini, diğer bir ifade ile her bir tweet gönderisini temsil etmek üzere kullanılmıştır.

Koleksiyon / Korpus: Dokümanların bir araya gelmesi, gruplandırılması ile oluşan, benzer bir amacı paylaşan, aynı kategoriye ait birden fazla dokümanın oluşturduğu topluluk, veri seti ya da söz konusu dokümanların depolama alanıdır [24]. Veri kümeleri halinde

düzenlenmiş özgün metin veya ses derlemi şeklinde de tanımlanabilir ve gazete, roman, yemek tarifi, filmlerden tweetlere kadar her tür veriden oluşabilir [25]. Koleksiyon ile aynı anlamda kullanılan diğer bir ifade ise korpus terimidir ve derlem şeklindeki ifade ile aynı anlamı taşır. Makine öğrenme modellerinin eğitimi için kullanılan kaynak dokümanlardan oluşan koleksiyon olarak tanımlanır [26]. Diğer bir deyişle korpus, analiz edilecek ana veri kütlesidir. Türkçe anlamı *konu külliyatı* şeklinde ifade edilebilecek olan koleksiyon ve korpus terimlerinin her ikisinin de, makine öğrenmesi çalışmalarında kullanıldığı görülmektedir. Çalışma kapsamında, belirlenen anahtar kelimelerle Twitter (X)'dan çekilerek elde edilen tweetlerin oluşturduğu veri seti bir koleksiyon olarak nitelendirilmiştir[2].

Model parametresi: Kullanılan modelin veri seti üzerinden çıkardığı katsayılara denir. İçsel parametre şeklinde de isimlendirilir.

Model hiper-parametresi: Model tarafından ortaya konulmayan, kullanıcı tarafından belirlenen ve algoritmaların çalışmasını kontrol eden parametrelere denir. Dışsal parametre şeklinde de isimlendirilir.

Sosyal Medya Ağ Analizi (SMAA): Topluluk/grup içindeki bireysel aktörler ya da topluluklar arasındaki ağ yapılarını ve ilişkilerini inceleyen analiz yöntemine, sosyal ağ analizi (SAA) denir [27]. Çalışma kapsamında SAA yöntemleri kullanılarak internet tabanlı web uygulamalarından elde edilen veri setleri üzerinde analiz yapılmış ve yapılan bu analizi ifade etmek üzere SMAA, SMAA'nın Twitter (X) verileri üzerindeki uygulamasını belirtmek üzere Twitter (X) Ağ Analizi (TAA) ifadeleri kullanılmıştır.

Uygulama Programlama Arayüzü (Application Programming Interface-API): Bir uygulamaya ait işlevlerin başka bir uygulamada da kullanılabilmesi için oluşturulmuş bir arayüzdür. API'ler genellikle web ve program geliştiricileri (developer) tarafından kullanılmakta ve hizmeti sağlayan servislerin kendileri tarafından, geliştiricilere sunulmaktadır [28]. Örneğin Facebook, Twitter (X) gibi platformlar, belirli sınırlandırmalar dahilinde sahip olduğu yeteneklerinin dışarıdan kullanılabilmesine API'ler aracılığıyla

[2] Çalışmada 4'üncü bölüme kadar literatüre ilişkin hususlarda yer aldığı için hem *veri seti* ve hem *koleksiyon* terimleri aynı anlamı ifade etmek üzere kullanılmış, 4'üncü bölümden çalışmanın sonuna kadar çalışmada analiz edilen veri setini belirtmek üzere yalnızca koleksiyon terimi kullanılmıştır.

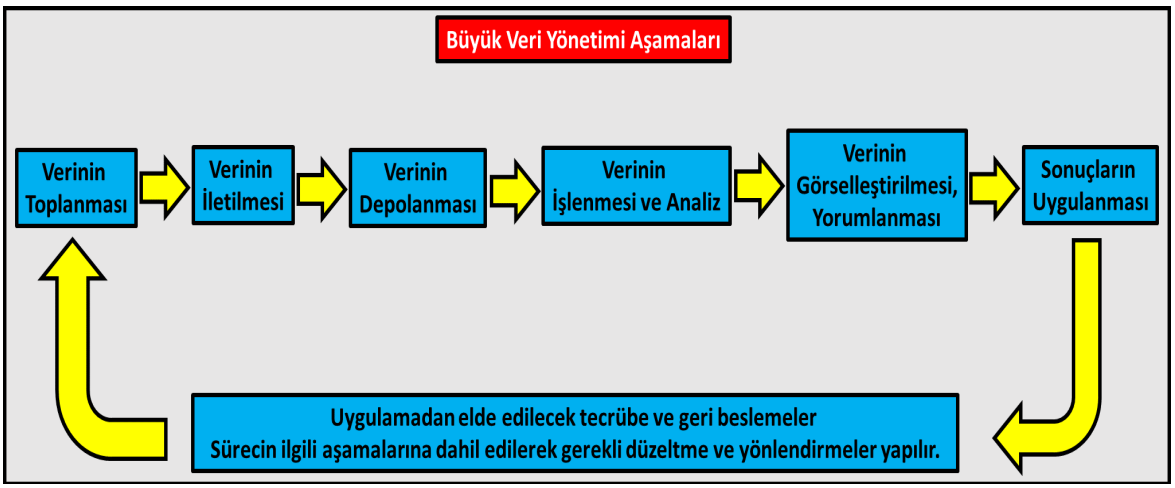
imkan verir. API'ler aracılığı ile uygulama ya da platformların özellikleri, fonksiyonları ya da içerikleri üzerinde istenilen işlemler sınırlamalar ve erişim yetkileri çerçevesinde yapılabilir. Erişim izni yine platformun kendisi tarafından sağlanan erişim/giriş dizgeciği/jetonu (access token) ile verilir [29].

Kütüphaneler (Yazılım Kütüphanesi): Bilgisayar biliminde işlevsellik açısından bir kütüphane görevi gören, yazılım ve program içerikli dokümantasyona sahip ve genellikle yazılım geliştirme maksadı ile kullanılan kaynaklar ve bunların depolandığı alanlardır. Bu kütüphaneler, düzenlenmiş veri, dokümantasyon, yardımcı veri, mesaj taslakları, önceden yazılmış kodları ve bunların değer ve özelliklerini kapsar ve yazılım geliştiricileri, daha önceden yazılmış bu kodlara ulaşabilirler ya da kendi geliştirdikleri kodları buraya depolayabilirler. Bu şekilde yeni programlar daha hızlı bir şekilde geliştirilebilir [30].

Yazılım/Programlama Dili: Bilgisayarlarla iletişim kurmak maksadıyla kullanılan iletişim aracıdır ve bilgisayara yaptırılmak istenen işlem, bu iletişim dili kullanılarak komut haline dönüştürülür. Pyhton, Java, Javascript, C, C++, C#, F#, Go, Scala, Ruby ve PHP [31] hepsi ayrı amaca yönelik kullanılan bir yazılım dilidir.

2. TWITTER (X) PAYLAŞIMLARININ TANIMLAYICI, İÇERİK / METİN VE AĞ ANALİZ YÖNTEMLERİ İLE ANALİZİ VE LİTERATÜR TARAMASI

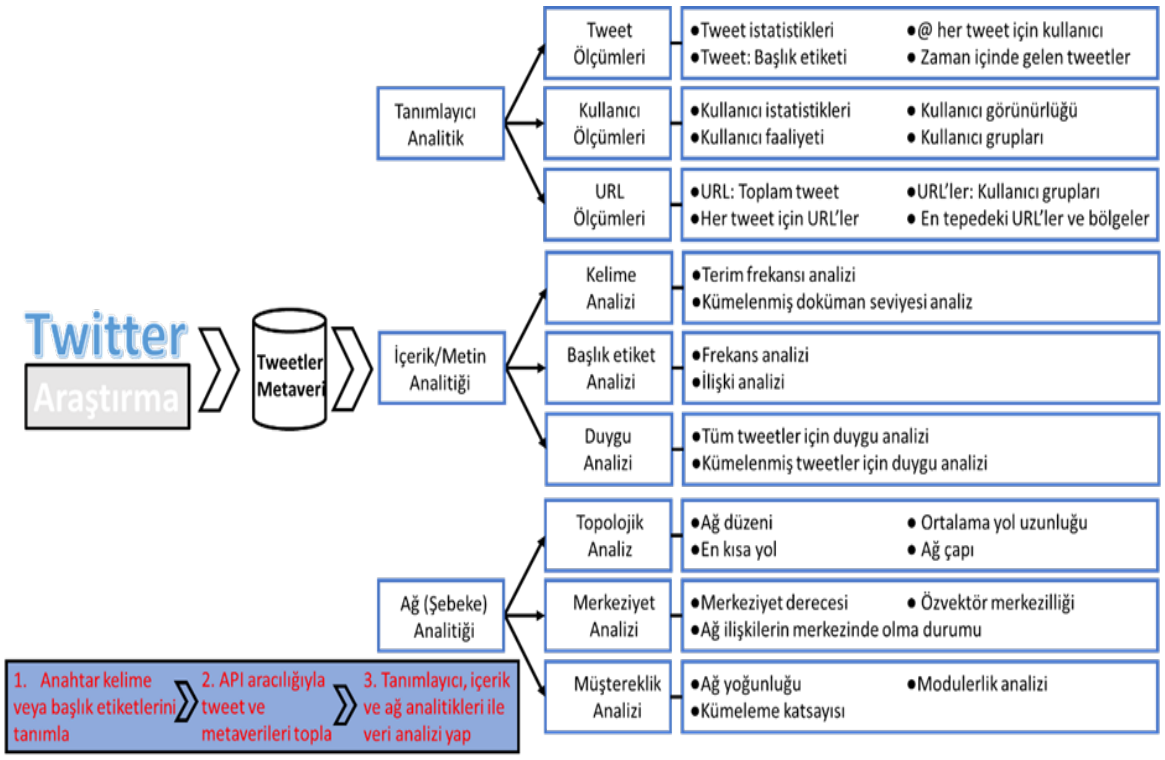
İnternetin, günlük hayatın her alanında kullanılmaya başlaması, büyük miktarda veri üretimine neden olmuştur. Söz konusu verinin yönetimi ve kullanımı için çeşitli yöntemler ve yeni kavramlar ortaya çıkmıştır. Depolanmasından, bilgi üretimine kadar her aşaması kendine özel yönetim süreçlerine ihtiyaç duyan büyük veri (BV) ve yönetim süreçleri için amaçlarına göre değişebilen isimlendirme ve tanımlamalar [32-34] yapılmaktadır. Veriden kullanışlı bilgilerin elde edilmesi için kullanılan yöntemleri içine alan uygulamalar, büyük veri analizi (BVA) olarak isimlendirilmektedir [21, 32, 35, 36]. Karar verme süreçlerini destekleyecek bir disiplin olarak ortaya çıkan BV'de, verinin özellikleri; *kaynak çeşitliliği* (Variety), elde edilme *hızı* (Velocity), *hacim* (Volume), *doğruluğu* (Veracity), *gerçeklik*/*değer* (Value) şeklinde 5V kavramı ile sıralanmaktadır. Bu özellikler, verinin işlenmesi için kullanılan yöntem ve süreçleri belirler. Karar verme süreçlerini etkilemesine, organizasyonel değişikliklere neden olmasına rağmen, büyük veri analizlerinin yöneticiler tarafından yeterli düzeyde etkili bir araç olarak dikkate alınmadığı ve bazı sektörlerde BVA kullanımının %17 gibi düşük seviyelerde gerçekleştiği görülmektedir [37-45]. Öte yandan bulut bilişim sistemi gibi ortaya çıkan yeni sistemler BVA yöntemlerinin kullanımını yaygınlaştırmakta [36], sosyal medya platformları analize konu olan verilerin ulaşılabilirliğini kolaylaştırmaktadır. Süreçlerin ve organizasyonların iyileştirilmesi gibi birçok maksatla kullanılan BVA; finans, ekonomi, sağlık, biyoloji, enformasyon teknolojileri sektörü, lojistik, tedarik zinciri yönetimi ve üretim gibi her alanda uygulanmaktadır [39]. Amaç ve kapsamı açısından farklı isimlendirmeler yapılsa da veriyi kullanılabilir bilgiye dönüştüren veri yönetiminde kullanılan süreçler birbirine benzerdir. İster sosyal medya platformlarından, ister işletmenin iç kaynaklarından elde edilmiş olsun veri setleri genel olarak Şekil 2.1'deki gösterilen safhalar sonunda, işletmeler için ekonomik bir değere sahip bilgiye dönüştürülebilirler [46].



Şekil 2.1. Büyük veri yönetim aşamaları

Veri analizleri için, amaç, kapsam; safhalar; kullanılan teknik, algoritma ve sektörler; veri yapısı; elde edilen sonuçlar dikkate alınarak gruplandırmalar yapıldığı, her analizin kendine özel detaylarla farklılaştığı literatürdeki çalışmalarda görülmektedir [32, 33, 41, 47-54].

Sosyal medya platformlarından elde edilen verilerin analizi için de farklı düzeylerde ve detaylarda uygulamalar yapılmaktadır. Bu çalışmada, farklı düzeyde analitik çalışmaları bir araya getirmesi nedeniyle Chae [3] tarafından Twitter (X) için ortaya konulan analitik çerçeve, bu çalışmanın veri setine özel yeni uygulamalar da dahil edilerek kullanılmıştır. Chae [3] tarafından önerilen analitik çerçeve Şekil 2.2'de gösterilmiştir. Uygulama için, söz konusu analitik çerçeve ile literatürde Twitter (X) analizlerinde kullanılan diğer araç, yöntem ve tekniklerden uygun olanlar seçilmiştir. Söz konusu analitiğin uygulanma detayları literatürdeki kullanım şekilleri ile ileriki maddelerde açıklanmış, analitik çalışmanın uygulaması sonucunda elde edilen bulgu ve yorumlar 3'üncü bölümde verilmiştir.



Şekil 2.2. Üç kademeli Twitter (X) analitik çerçevesi [3]

2.1. Twitter (X) mesajları, üst verisi, Twitter (X)'dan veri çekme yöntemleri

Twitter (X) uygulamasındaki veriler ve kullanımına yönelik özellikler ileriki maddelerde açıklanmıştır.

Twitter (X)'da maksimum 280 karakter olarak yazıma müsaade edilmektedir. Standart bir kullanıcı profili oluşturularak Twitter (X)'ın sosyal ağ ortamına girilmiş olur. Kullanıcı profilinde yer alan bilgiler, kullanıcıya ait bilgileri kapsar ve üst veri olarak ifade edilir. Üst veri araştırmacılar ve işletmecilerin Twitter (X) üzerinde yapacakları analizlerde incelenecek demografik bilgilere kaynaklık eder. Kullanıcılar tarafından, üst verilerin doğru ya da yanlış, eksik bir şekilde girilmiş olması, bu bilgilere dayanarak yapılacak analizlerin doğruluğunu yükseltecek ya da düşürecektir, bu husus yapılacak analizlerde dikkate alınmalıdır. Standart bir Twitter (X) kullanıcı hesabı üzerinden, kullanıcı profili ya da anahtar kelimeler kullanarak sorgulamalar yapılabilir [3].

TA'ne başlamadan önce yapılacak analizlerin doğru yorumlanabilmesi için, Twitter (X)'da kullanılan bazı terim ve uygulamaların anlaşılmasına ihtiyaç vardır. Bu nedenle aşağıda

Twitter (X)'da sık kullanılan ve yapılacak analizlerde de bahsedilecek olan bazı terim ve bu terimlerin kullanımına yönelik hususlar açıklanmıştır. Twitter (X) kullanıcılarının diğer kullanıcılar ile ilişkileri açısından; birincisi diğer kullanıcıları *takip etme*, ikincisi diğer kullanıcılar tarafından *takip edilme* olmak üzere iki durum vardır. Diğer bir ifade ile bir Twitter (X) kullanıcısı hem takip ediyor hem takip ediliyordur. Aradaki bu ilişkiler, çeşitli analiz yöntemleri ile incelenebilen bir yapıya sahiptir. Bir kullanıcıyı takip etme, takip ettiği kullanıcının bir tweet mesajı paylaşması halinde o kullanıcıdan bildirim almak için o kullanıcıya abone olması anlamına gelir. Benzer şekilde kullanıcının kendisinin bir mesaj paylaşması da abonesi olan takipçilerine bildirim gönderdiği anlamına gelir.

- *Açık ve korumalı tweetler:* Twitter (X)'da hesap oluşturulduğunda, varsayılan uygulama olarak tweetler, herkese açıktır, yani herkes hesaba ait tweetleri görüntüleyebilir ve onlarla etkileşim içine girebilir. Eğer tweetlerin korunmaya alınması isteniyorsa, hesap ayarları kullanılarak korumaya alma işlemi yapılabilir. Koruma altına alınmış tweetlere ulaşım ise ancak o hesabı takip etmek isteyen kişinin istek göndermesi ve o isteğin istek gönderilen kullanıcı hesabı tarafından onaylanması ile mümkündür [55]

- *Yanıt (Reply) bahsetme (Mention) tweeti:* Twitter (X)'da kişisel tweetlerle başkalarına cevap vererek ya da başkalarından bahsederek (mention) sohbetlere katılım sağlanabilir. Yanıt, bir başka kişinin tweetine cevap vermektir ve 🗨 simgesi üzerine tıklayarak cevap verilebilir. İki kişi birbirine yanıt verdiği zaman yalnızca ilgili kişiler (yanıt veren kişinin takipçileri ve sohbetteki kişi) bu yanıtı görürler. Bahsetme, tweet metninde başka bir kişinin kullanıcı adının kullanılmasıdır. Belirli bir hesap ya da hesaplara yönelik yazılacak metinlerde kullanıcı adı yazılırken ön tarafına @ işareti (*@kullanıcıadı* şeklinde) konulur. Eğer gönderilen tweet bir yanıtsa, yayınlama simgesi *Yanıtla* olarak görülür ve yayınlandığı zaman, gönderilen tweetteki kullanıcı adı hesabın profil sayfasına giden bir bağlantıya dönüşür [56].

- *Tekrar (Retweet) tweeti:* Bir kullanıcı hesabı tarafından, söz konusu hesabın takipçileri ile herkese açık olarak paylaşılan bir tweet, retweet olarak adlandırılır. Bu genellikle herhangi bir medyanın paylaşılması için kullanılan oldukça pratik yoldur. Retweetlemeden önce, paylaşıma yorumlar eklenebilir. Retweet özelliği kullanıldığında, söz konusu retweet daha önce paylaşılmış olan tweeti referans alır [57]. Retweet, tweet sahibinin orijinal

paylaşımına sadık kalınarak yapıldığı için bu başkasının ifadeleridir anlamında kullanılmış olmaktadır [58].

● *Favoriye alma (Favorites) / Beğenme (Like):*Twitter (X)'da 04 Kasım 2015 tarihine kadar küçük yıldız sembolü ile favorilere ekleme özelliği kullanılmakta iken, bu tarihten sonra beğenme (like) özelliği eklenmiş ve favorilere ekleme özelliği kaldırılmıştır. Favorilere ekleme genel çerçevede beğenmek anlamında kullanılmıştır [58]. Paylaşılan herhangi bir tweetin diğer kullanıcılar tarafından beğenildiğinin ifade edilmesi için beğen düğmesi kullanılır.

● *Diyez etiketi (Hastag):* Diğer uygulamalarda olduğu gibi Twitter (X)'da yaygın olarak kullanılan özelliklerden biri de etiketleme özelliğidir. İngilizce karşılığı hashtag olan ifade için Türkçe karşılık olarak *diyez etiketi, etiket* ya da *başlık etiketi* ifadesi kullanılır, # sembolü ile uygulanır [58]. Twitter (X) diyez etiketi, anahtar kelimeleri veya konuları dizine eklemek için kullanılır. Bu işlev, kişilerin ilgilendikleri konuları kolayca takip edebilmelerini, tweetlerin kategorilere ayrılabilmesini, aramalarda daha kolay ulaşılabilmeyi sağlar [59].

Twitter (X) analitiği için kullanılacak verileri elde etmek için Twitter (X)'dan çeşitli şekillerde veri çekilebilir. Twitter (X) verileri, kullanıcıların herkese açık olarak paylaşmayı tercih ettiği bilgileri yansıtması bakımından diğer birçok sosyal platformun paylaştığı verilerden farklıdır [60]. Twitter (X)'dan veri çekme işlemi Twitter (X) API'leri kullanılarak doğrudan yapılacağı gibi, bu özelliği kullanarak veri sağlayan Twitonomy, Medaiatoolkit gibi ticari uygulamalarla da gerçekleştirilebilir. Twitter (X) tarafından verilen sınırlar içinde kullanılan API'ler aracılığı ile veri çekme genellikle iki şekilde gerçekleştirilir. Birinci durumda, Twitter (X)'den gerekli giriş izni yani kullanıcı anahtarı ve kullanıcı giriş jetonu (Twitter (X) App Consumer Key and Token) izni alındıktan Twitter (X) API'si ile kamuya açık tweetler çekilerek veri seti oluşturulabilir. Bu uygulama ile geçmiş döneme ait belirli tarih aralıkları ile veri çekilmesine olanak sağlanmaktadır. Veri çekme işlemi için ikinci olarak Twitonomy, Exporttweet, Keyhole, Rite Tag, Mediatoolkit, Viralheat gibi ticari uygulamalar kullanılır. Söz konusu uygulamalar ile tweetlerin toplanması ve toplanan tweetlerin excel ve pdf ortamına aktarılması mümkündür. Bu uygulamalarda, son bir haftalık veriler çekilebilmektedir. Söz konusu ticari uygulamalar incelendiğinde, her birinin Twitter (X)'ın farklı analitik değerlerine yöneldiği görülmektedir. Her iki durumda da anahtar

kelime ve başlık etiketi ile yapılan veri çekme işleminde 3200 adet tweet metni çekilebilmektedir [60-62].

2.2. Anahtar kelime ve diyez etiketi tanımlama, veri setinin oluşturulması

TA'nde önemli adımlardan ilki araştırma yapılacak konuya ve maksada yönelik doğru verilere ulaşabilmek için doğru anahtar kelime ve etiketinin belirlenmesidir. Veri kümelerinin oluşturulabilmesi, konuya ilişkin bilginin Twitter (X)'dan çekilmesi için anahtar kelimenin tespit edilmesi gerekir. Anahtar kelime tespiti için kullanılan yöntemlerden biri, müşteri şikayetlerinin incelenmesi olabilir [3]. Müşteri şikayetlerine, çevrimiçi olarak internet üzerinden ulaşılabileceği gibi, işletmelerin veri tabanlarındaki müşteri geri bildirimlerinden, konuya ilişkin anket ve çalışmalardan ulaşılabilir ve elde edilen müşteri görüş ve bildirimlerinin hangi konular üzerinde yoğunlaştığı incelenerek anahtar kelime ve diyez etiketi tespit edilebilir. Anahtar kelime ve diyez etiketlerinin belirlenmesinden sonra analiz için veri çekme işlemleri yapılır. Literatürde yer alan çalışmalarda, analiz edilen veriler içinde reklam içerikli tweetlerin olabileceği dikkate alınmalı ve veri kümesinin oluşturulmasında ayıklama işlemi yapılmalıdır.

Sosyal medya analizlerinde anahtar kelime ve diyez etiketi belirlenmesi veri havuzu oluşturmanın ilk aşaması sayılabilir, tespit edilen kelimelerle sorgulama ve veri çekme işlemi gerçekleştirilir ve sorgulama sonucunda veri havuzu elde edilmiş olur. Bu çalışma için Twitter (X)'dan çekilen veri setinde çok sayıda reklam içerikli tweet olduğu görülmüştür. Literatürdeki çalışmalarda reklam içerikli tweetlerin veri setlerinden ayıklanmadığı ve bu tweetlerin analiz sonuçlarını etkilemeyeceğinin kabul edildiği görülmektedir.

Standart veri çekme işlemlerinde, kullanıcılar tarafından korumalı/özel olarak işaretlenmiş tweetler, API ya da uygulamalar aracılığı ile veri toplama esnasında veri havuzuna çekilemezler [22]; bunun sonucu olarak veri setleri, yalnızca kamuya açık olan tweetlerden meydana gelir. Bu durum, verilerin temsil yeteneği açısından bir kısıtlılık ortaya çıkarabilir. Ayrıca reklam ya da propaganda maksatlı yazılmış, tweet ve retweetlenmiş tweetlerin ayıklanması mümkün olamayacağından, analiz kalitesinde düşüş olabilir [19].

Veri setinin oluşturulmasında doğru verilere ulaşılabilmesi için veri çekmede kullanılacak anahtar kelimelerin amaca uygun olarak belirlenmesi gerekir. Belirlenen anahtar kelimeler ile veri havuzu oluşturulurken ilgisiz verilerin de veri havuzuna girmesi mümkündür [63].

2.3. Tanımlayıcı analiz ve literatür

Tanımlayıcı analitik, tweet sayısı, farklı tweet tiplerinin dağılımı, diyez etiketi sayıları gibi tanımlayıcı istatistiklere odaklanır. Twitter (X) verisinin zenginleştirilmiş doğal yapısı, tweetler, kullanıcılar, etiketler, URL'lere ilişkin geniş çaptaki ölçüm setlerini kullanarak istihbarat elde edilmesine imkân verir [3]. Kullanıcı, tweet ve URL istatistiklerinin tümünün zaman dilimleri ile ilişkilendirilerek grafiğinin çıkarılması mümkündür. Bu maksatla çalışmanın amacına uygun olarak saat, gün, hafta, ay, yıl gibi farklı zaman dilimleri kullanılabilir. Bruns ve Stieglitz [64] çalışmalarında; (i) kullanıcı ve kullanıcı gruplarının katkıları, (ii) aktivitelerin zamanla ilişkilendirilmiş tüm resmi, (iii) ilk iki ölçüm aracının birleştirilerek belirli kullanıcı ve kullanıcı gruplarının zamanla bağlantılı aktivitelerinin ortaya çıkarılması şeklinde üç tür kullanıcı/kullanıcı grubu aktivitesinin önemine vurgu yapmışlardır. Kullanıcılara ait zamansal ölçümler, analiz edilen konuya ilişkin tweetlerin sayılarındaki kayda değer artış ve azalışların olduğu zaman dilimlerinin tespit edilmesini, kullanıcıların zamana bağlı üretkenliklerinin ölçülmesini sağlarken, diğer yandan işletmeler açısından mutlaka etkilenmesi gereken müşterilerin tespit edebilmesine de [64] imkan verir.

Veri seti içindeki tweetlerin kaçının orijinal (sohbet başlatan) tweet, kaç tanesinin retweet (RT) olduğu veri seti içindeki gündem konularına ilişkin bilgilerin elde edilmesini sağlarken, en çok hangi tweetlerin favoriye eklendiğinin/beğeni aldığının tespit edilmesi ise analiz konusu ürün/hizmetin olumlu olarak hangi yönlerinin ortaya çıktığını gösterir. Tanımlayıcı analitik kapsamında tweet miktarı, kelime sayısı, tweet başına düşen kullanıcı sayısı, diyez etiket miktarı gibi ölçüm setleri kullanılır ve bunlar üzerinde yapılacak incelemeler başka ölçümlerin yapılmasına imkân verir. Kim tweetledi, kim cevapladı ve kim retweetledi sorularının cevapları, iş alanına yönelik değer katmak isteyen uygulamacılar ve bir fenomeni araştıran araştırmacılar açısından önemli iken bu çalışmada amaca uygun olarak tweet sayısı, kelime sayısı, her bir tweet için kullanıcı sayısı, diyez etiket sayısı [3] gibi istatistiki değerler üzerinden inceleme yapılmıştır. Tweet (sohbet başlatan) ve RT sayıları sohbet ortamındaki konu çeşitliliği hakkında fikir verir. Bir kullanıcı doğrudan bir tweet gönderisi paylaşımı ile bir sohbet başlatmış olur, başlatılan bu sohbet eğer RT edilmiş ise orijinal tweetin ortamda

yayılması sağlanmış olur. Veri seti içinde orijinal tweet sayısı ne kadar çoksa çalışma yapılan konuya ilişkin gündem çeşitliliğinin yüksek olduğu söylenebilir. Orijinal (sohbet başlatan) bir tweetin kaç tane RT aldığı o konunun kullanıcılar arasında ne kadar çok yayıldığını gösterir. Çok sayıda RT almış bir tweet ise, işletmelere bu tweeti yazan kullanıcıların olumlu etkilenmesi gereken müşteri hedef kitlesi olduğu yönünde bir işaret verebilir. Eğer en çok RT alan bir tweet ürün/hizmet hakkında olumsuz görüşler içeriyorsa, işletmeler bu konuyu öncelikle ilgilenilmesi/iyileştirilmesi gereken bir konu olarak ele almak durumunda olacaktır. Tweet istatistiklerinden bir tweetin favoriye eklenme/beğeni alma sayısı, analiz edilen ürün/hizmetin müşteriler tarafından en çok beğenilen yanını gösterir. Ürün/hizmetin bu beğenilen yanı, işletmeler için kalite seviyesinin korunması gereken yönünü gösterir. Yine bu istatistikler dakika, gün ya da aylık dilimler şeklinde düzenlenerek kullanıcıların üretkenlikleri ölçülebilir [64]. Zamansal ölçümler, aktiviteler ve görünürlülükler, zamana bağlı değişimlerin ve lider konumundaki kullanıcılardaki değişikliklerin görülmesini sağlar.

Tanımlayıcı analizin ikinci aşamasında kullanıcıya ait ölçümler yapılır; en aktif kullanıcılar, kullanıcı gruplarının faaliyetleri ve kullanıcılara ilişkin diğer bilgiler incelenir. Kullanıcı istatistikleri ile kullanıcı başına düşen ortalama tweet sayıları, en aktif, en etkili, en görünür (görünürlüğü en yüksek), en beğenilen kullanıcıların kimler olduğu belirlenir, işletmelerin hedef kitleleri arasında önceliklendirme yapabilmesine imkan verir. Ayrıca bir konuyla ilişkili fenomen araştırması yapan işletmeler de kullanıcı istatistiklerine yönelirler. Kullanıcı istatistikleri ile iletişim ağında (veri seti içindeki iletişim içinde) merkezi konumda olma durumuna ilişkin emareler elde edilir [3].

Kullanışlı bilgiye ulaşmak için elde edilen analitik değerlerden biri, aktif kullanıcının tespit edilmesidir. Veri seti içinde en aktif kullanıcı, yazıp paylaştığı orijinal (sohbet paylaşımı başlatan) tweet sayılarına göre belirlenir, bu ölçüm kullanıcının ortamdaki aktivitelerinin sayılması işlemidir. Kullanıcının etkinlik ölçümü ise takipçi sayıları ile yapılır, bununla birlikte bir kullanıcının yaptığı paylaşımının ne kadar beğeni (like) aldığı da o paylaşımın ortamdaki etkisini gösteren bir değerdir [3, 64]. Kullanıcıya ait ölçüm araçlarından bir diğeri de ortamdaki görünürlüğün tespitidir. Görünürlük, alınan cevap tweeti (Reply) ile alınan RT'lerin toplamı (ya da bu her iki değerin ayrı ayrı frekansları) şeklinde ifade edilebilir. Twitter (X) kullanıcısının ortamda görünür olup olmadığının ölçütü olarak, veri kümesindeki tweetlerden bahsetme tweetlerinin çıkarılmasıdır, diğer bir ifade ile bu bir kullanıcı tarafından alınan bahsetme tweetlerinin sayısıdır [3]. Görünürlük ifadesinin

kullanımındaki bir başka karşılığı ise, diyez etiketinin kullanımı ile ilişkilidir. İnternet ortamında kullanıcıların yaptıkları sorgulamalarda ulaşılabilirliği artırmak için diyez etiketi kullanılır. Yazılan tweetin konusu ile ilgili kelimelerin tweete, diyez (#) sembolü eklenerek ayrıca yazılması, o kelime ile başka kullanıcıların yaptıkları aramalarda görünür olmayı sağlar. Bu şekildeki görünürlük uygulamasını özellikle reklam içerikli tweet paylaşımlarında görmek mümkündür, bu maksatla reklam tweetlerinde çok sayıda diyez etiketi kullanılır. Bu durum bu çalışmada kullanılan veri setindeki diyez etiketi frekans değerleri ile de tespit edilmiştir. Kullanıcılara ait istatistikler, işletmelere, sosyal paylaşım ortamında amaçlarına uygun olarak hangi kullanıcıları hedef kitle olarak seçmeleri gerektiği konusunda bilgi verebilir. Ayrıca Twitter (X) kullanıcılarının değerlendirmesi için kullanıcı aktiviteleri, etkinlikleri ve görünürlülük değerlendirmeleri veri setinin toplam değerleri açısından yüzdelik dilimlere ayrılarak, önceliklendirmeler yapılabilir. Bir başka ifade ile veri seti içindeki en aktif, en etkili, en görünür kullanıcılardan hangilerinin veri seti toplamında ilk %1'lik, orta bölümdeki %90'lık ve son bölümdeki %9'luk dilimler içinde kaldığı tespit edilebilir. Bu şekilde dilimlere ayırma, işletmenin amacı ve imkanlarına uygun olarak hedef kitle politikaları belirlemesine yardımcı olabilir. Bruns ve Stieglitz [64] çalışmasında verilen bu yüzdelik dilimler; en üst seviye, orta seviye ve en alt seviyede kalan kullanıcıları gruplandırma yöntemi olarak verilmiştir. Dilimlere ayırma, işletmenin amaç ve imkanları ve veri seti toplam sayıları dikkate alınarak, ilk 25 ya da 50 kullanıcı, son 25 ya da 50 kullanıcı ve bu ikisi arasında kalan orta dilim şeklinde de yapılabilir. Kullanıcının aktiviteleri ile görünürlülüğü arasında anlamlı bir karşılaştırma yapılabilir, bu karşılaştırma herhangi bir kullanıcının mesajlarının sohbet ortamındaki etkisine vurgu yapmaktadır. Eğer kullanıcı çok sık tweet gönderiyor ama az sayıda geri dönüş alıyorsa, bu kullanıcı mesajlarının etkisinin düşük olduğu söylenebilir. Tersi durumda ise yani kullanıcı az sayıda tweet gönderiyor ve çok sayıda geri dönüş alıyorsa değer birden büyük çıkar ve bu durum kullanıcı mesajlarının oldukça etkili olduğunu gösterir [64].

Kullanıcılar, tweetlerini sohbet konusuna ilişkin görsel, metin, belge, rapor, makale ile genişletmek, detaylandırmak için URL kullanırlar. Tweetlerdeki URL analizleri de kullanıcılar arasındaki esas gündem konularının tespit edilmesine yardımcı olur, en çok ilgi çeken konuların tespit edilmesini [3] ve içeriğine göre konu ve kullanıcı hakkında detay bilgiler elde edilmesini de sağlayabilir. URL analizlerinde; (i) tweetlerin kaç tanesi bir veya daha fazla URL içeriyor, yüzdesi nedir, (ii) URL içeren tweetlerde kaç farklı URL mevcuttur, (iii) aktif ve görünür kullanıcıların tweetleri URL içeriyor mu, kaç tanesi içeriyor,

yüzdesi nedir, (iv) URL içeren tweetlerin tepesinde yer alan tweetlerin ne tür özellikleri vardır (işletmelerin internet sayfası, çevrimiçi haber sitesi, çevrimiçi haber sayfası, konuyla ilgili makale veya yazı, iş ilanı, linkedin sayfası). Bu soruların cevapları ile veri setine ilişkin önemli bilgiler elde edilebilir.

Tweetlerin üst veri olarak bulundurdukları coğrafi konum bilgisi (CKB), bölgesel çıkarımlar yapılmasını sağlayabilir. CKB üst verisi ile veri setine ilişkin yeterli seviyede anlamlı bilgi elde edilebilmesi, kullanıcıların profil bilgilerinin tam ve doğru olarak girilmiş olmasına bağlıdır. İşletmeler genellikle coğrafi konuma ilişkin politikaların geliştirilmesi için öncelikle kurumsal verilerini ve diğer bölgesel istatistiki verileri kullanabilirler. Daha geniş saha araştırmaları için Twitter (X) üst verisindeki CKB kullanılabilir. Twitter (X) aracılığı ile elde edilen bilgilerin, diğer kaynaklardan ulaşılan sonuçlarla karşılaştırması yapılabilir ve alınacak kararları destekleyici bilgi olarak kullanılabilir.

Tweet üst verisinde tweetin hangi platformdan paylaşıldığını gösteren bilgiler yer alır. Bu kapsamda tweetlerin üst bilgisinde paylaşımların hangi platformdan yapıldığını gösteren "iPhone for Twitter (X), "Android for Twitter (X)", "Web for Twitter (X)" şeklinde bilgiler bulunur. Bunların dışında tweetin üçüncü taraf bir uygulamadan gönderildiğini gösteren istemci adı da görülebilir. Üçüncü taraf istemciler, yazarların tweetlerini ve pazarlama kampanyalarını yönetmek, reklamcılık performansını ölçmek, müşteri desteği sunmak ve reklam göstermek üzere belirli insan gruplarını hedeflemek için kullandıkları uygulamalarıdır ve geliştiriciler tarafından kullanılan yazılım araçlarıdır. Tweetler ve kampanyalar, doğrudan insanlar tarafından oluşturulabilir veya bazı durumlarda bir uygulama tarafından otomatikleştirilebilir. Tweet verisinin ne şekilde üretildiği, hangi kaynaktan üretildiğinin tespit edilmesi özellikle reklam maksadı ile üretilmiş tweetlerin veri setinden ayıklanması konusunda yardımcı olabilir. Platform bilgisi ayrıca işletmeler tarafından çevrimiçi sistem ve uygulamalarını, web sayfalarını geliştirmelerinde hangi platformlara öncelik vermeleri gerektiği yönünde bir fikir verebilir.

2.4. Metin sınıflandırma, içerik / metin analizi, duygu analizi ve literatür

Twitter (X) gibi sosyal medya platformlarından elde edilen metin içerikli veri setlerinin analizi için, genel olarak kelime ve diyez etiket frekanslarının tespit edilmesi, metinlerin sınıflandırması ve duygu analizi yapılır. Kelime sayıları, cümle uzunlukları, frekans tespiti

ya da frekanslar arasında yapılan oranlar; incelenen metnin konusu, kapsamı, metni yazanların eğilimleri gibi bilgilerin elde edilmesini sağlar. Tweetlerin duygu sınıflandırması, kelime frekans tespiti ve kelime analizi, kelimelerin sınıf frekansları içerik analizi altında incelenir. Diyez etiket frekansları da koleksiyon hakkında değerli bilgilerin ortaya çıkarılmasına yardımcı olur.

Kelime analizi, içerik analizinin başlangıç noktasıdır. Kelime analizi, terim frekans analizi ve doküman seviyesi analizi olmak üzere iki seviyede gerçekleştirilir. TF, metinlerde en çok kullanılan terimlerin ortaya çıkarılarak metinlerin ana gündem maddesinin ve konuşmaların ağırlık noktalarının ne olduğunun belirlenmesini sağlar. Tespit edilen kelime frekansları görselleştirilerek koleksiyon içeriği hakkında bilgilerin yorumlanması mümkündür. Bu maksatla kullanılan kelime bulutu, kelimelerin gösterimdeki boyutlarını incelenen koleksiyondaki frekansa göre orantılayan, doküman içeriği konusunda ilk bakışta fikir edinilmesini sağlayan bir görselleştirme aracıdır. Bu yöntemle bir dokümanda yer alan ana konseptin ve dokümandaki ana düşünce ve yardımcı düşüncelerin ortaya çıkarılması ve tespit edilmesi, kullanılan algoritmalarla tespit edilen konu başlıkları altındaki düşüncelerin kümelenmesi ve böylece her ana başlık altındaki kavramların anlaşılması sağlanabilir. Tüm koleksiyon için uygulanan kelime analizi, ayrıca sınıflandırılmış koleksiyon seviyesinde de kullanılarak o sınıf koleksiyonuna yönelik çıkarımların yapılmasına, koleksiyonların bulundukları kategoriler açısından detaylı analizine yardım eder [3].

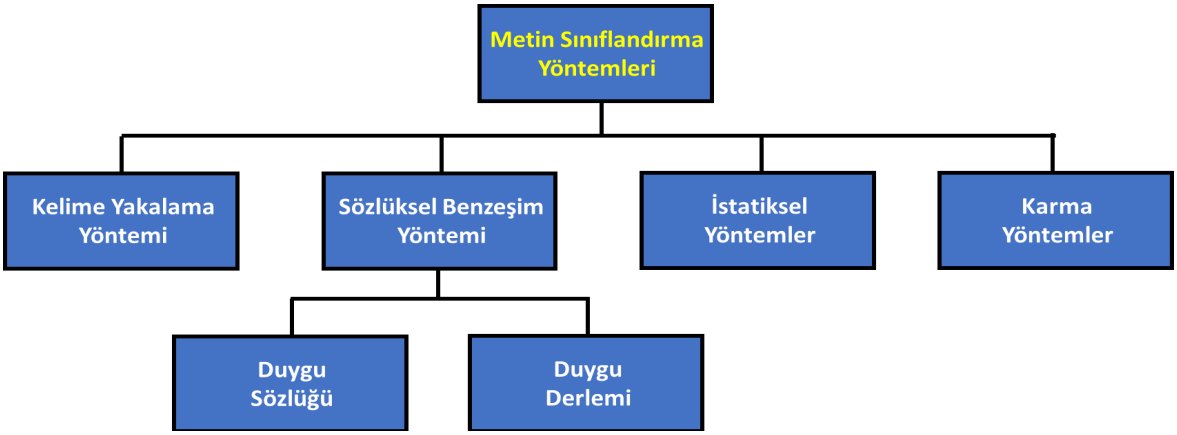
Dünya üzerinde diyez etiket (#hashtag) kullanımı ilk kez Twitter (X) üzerinde başlatılmış ve sonrasında tüm sosyal ağlar tarafından kullanılır hale gelmiştir. Diyez etiketi, kullanıcıların belli konudaki görüşlerini kategorize etmelerini sağlamaktadır. Kullanıcılar ilgi alanlarına göre etiketler üzerinden paylaşımlar yapmakta ve bu sayede aynı ya da benzer ilgi alanına sahip topluluklar bir araya gelebilmekte, markalar potansiyel müşterilerine ulaşabilmekte, sivil toplum kuruluşları da bu sayede hedef kitleleri ile etkileşime geçebilmektedir. Diyez etiketleri, akademik çalışma/araştırmalar ve blog yazılarının kategorize edilmesini sağlayan ana ilgi alanlarını gösterdikleri için tweetlerin önemli bir bileşenidir. Etiketlerin frekans analizi, diyez etiketlerinin popülaritesini gösterir ve koleksiyonların ana gündem maddelerinin belirlenmesini sağlar [3, 65]. Kelime ve etiket frekans analizleri, veri seti içeriğine ilişkin ön bilgiler verir, bununla birlikte esas bilgiler tweetlerin metin sınıflandırma ve duygu analizine tabii tutulması ile elde edilebilir.

2.4.1. Metin sınıflandırma ve duygu analizi

Twitter (X), metin içerikli bir sosyal paylaşım platformu olması ve çok sayıda kullanıcı görüşlerine ulaşılması nedeniyle metin sınıflandırma çalışmalarına konu olmaktadır. Metin sınıflandırma için kullanılan yöntem ve araçlar, bazı farklılıklarla Twitter (X) kullanıcıları tarafından paylaşılan tweetlerin sınıflandırılması ve duygu analizi için de kullanılır. Duygu analizi için literatürde; duygu madenciliği, düşünce madenciliği, tarama madenciliği, düşünce çıkarımı, sezgi analizi, algı analizi [66] gibi farklı isimlendirmeler de yapıldığı görülmektedir.

Metin sınıflandırma, metinlerden bilgi çıkarılmasına odaklanır, dilin özelliklerini kullanarak metinleri anlamlı bilgiye dönüştürür, bu dönüştürme süreci dilin işlenmesi ile ilgili güçlükler içerir. Metin sınıflandırma, örüntülerin doğal dil metinlerinden çıkarılması nedeniyle diğer veri madenciliği uygulamalarından farklılık gösterir ve yapay zekâ, doğal dil işleme (DDİ-Natural Language Processing/NLP), istatistik, bilgiye erişim ve bilgi yönetimi gibi bazı teknikleri kullanır [67-69]. Bu alanda kullanılan teknikleri, Bach vd. [70] dilbilimsel, istatistiksel ve makine öğrenimi; Zohar [71] bilgiye erişim, bilgi çıkarımı, web madenciliği, kümeleme şeklinde gruplandırılmaktadır.

Metin sınıflandırma ve duygu analizi için amaç, kullanılan araç ve yöntemler dikkate alınarak çeşitli sınıflandırmalar yapılmaktadır. Ravi ve Ravi [72] tarafından duygu analizi için; öznellik sınıflandırma, duygu sınıflandırması, yorumun kullanışlılığı ölçümü, duygu sözlüğü oluşturma, görüş kelimelerini/ürün durumunu çıkarsama ve diğer duygu analiz uygulamaları şeklinde altı başlık altında gruplandırma yapılmıştır. Metin sınıflandırma makina öğrenme tekniklerinin kullanıldığı önemli alanlardan biridir. Makina öğrenmesi, sözlük tabanlı yaklaşım, gözetimli, gözetimsiz öğrenme, hibrit yaklaşımlar şeklinde gruplandırılır [73]. Gözetimli öğrenme esas olarak sınıflandırma işlemi olmakla birlikte, söz konusu makina öğrenmesi teknikleri çeşitli şekillerde metin sınıflandırma işlemi için kullanılmaktadır. Metin sınıflandırma için yapılan bir başka gruplandırma Şekil 2.3'te gösterilmiştir [74].



Şekil 2.3. Metin sınıflandırma yöntemleri

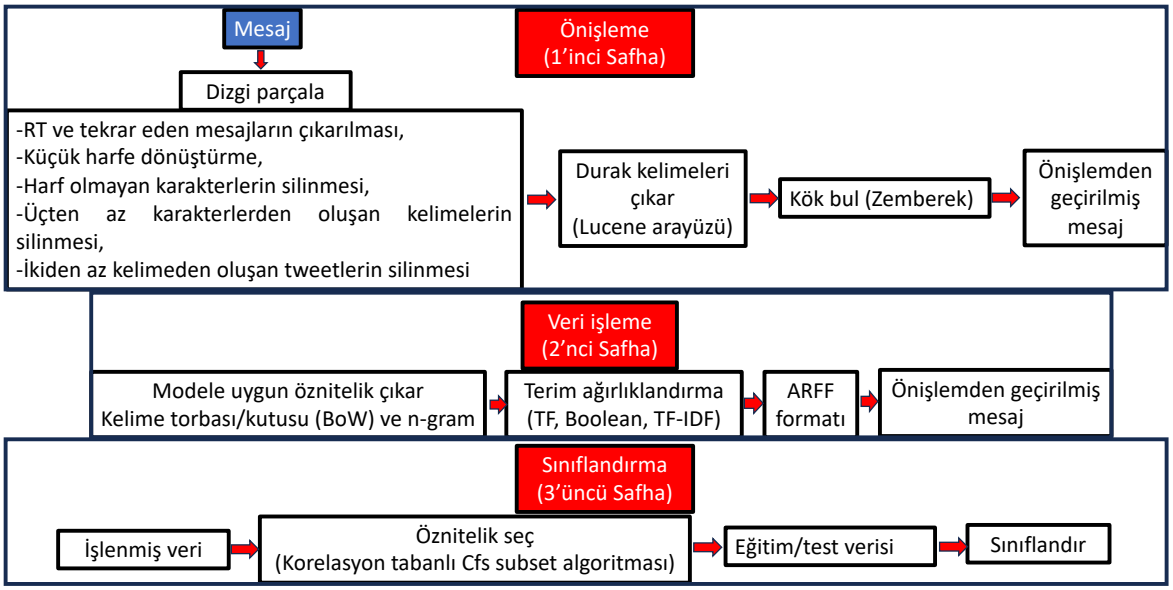
Duygu analizi, tweet metni içindeki sübjektif bilgiyi ortaya çıkarmakla ilgilenir. Diğer bir ifade ile duygu analizi, metni yazanın duygu ve düşüncesini hedefler [75, 76]. Duygu analizi birçok araştırmacı tarafından, bir sınıflandırma problemi olarak tarif edilmektedir. Duygu analizi, metin içindeki duygunun pozitif, negatif gibi hangi yönde olduğunun tespit edilmesi şeklinde bir sınıflandırmadır. Bu sınıflandırmalar pozitif, negatif gibi ikili; pozitif, negatif, nötr olarak üçlü [72, 76-81]; az olumlu, çok olumlu, az olumsuz, çok olumsuz gibi dörtlü [82]; pozitif, negatif, nötr, karışık şeklinde dörtlü [83] ile kullanıcıların yıldız değerlendirme sistematiğinden esinlenilen çok negatif, negatif, nötr, pozitif, çok pozitif şeklinde beşli [84] olarak amaca göre düzenlenebilir. Bu çalışmada ikili ve üçlü sınıflandırmalar yapılmıştır. Bu kategoriler dışında, mutluluk, sevinç, kızgınlık, şaşkınlık gibi duyguları da kapsayan sözlüklerle sınıflandırmalar yapılabilir [74]. Metin içindeki duyguların kategorilere ayrılmasında, hangi yöntem ve algoritmanın kullanılacağı, metin sınıflandırma işlemlerinin önemli bir problemidir ve kullanılacak sınıflandırıcının iyi eğitilip eğitilememesi analiz sonuçlarının başarısına olumlu ya da olumsuz etki eder. Duygu analiz çalışmalarında sınıf ayrımı ile pozitiflik/negatiflik derecesi de tespit edilebilir, bu yönüyle duygu analizi bir regresyon problemi olarak kabul edilir [85]. Diğer bir ifade ile metinde geçen özel anlama sahip kelimelere dayalı çeşitli skorların ve çeşitli duygu tonlarının hesaplaması yapılabilir [63]. Duygu analizi kapsamında bazı çalışmalarda metnin analizi ile metne bağlı diğer öğeler de analiz konusu yapılabilir. Bu kapsamda özellikle sosyal medya platformlarından elde edilmiş olan koleksiyonların birçoğu video, görsel, ses gibi eklentilere sahiptir ve bunların analizi metnin duygusunun ortaya çıkarılmasında etkilidir.

Araştırmacılar tarafından geliştirilen duygu analiz metodolojileri; ürün yorumları, film yorumları, otel yorumları, forumlar, bloglar, Facebook gibi sosyal paylaşım siteleri,

haberler, haberlere yapılan yorumlar ve Twitter (X) gibi mikroblog veri setleri üzerinde uygulanmaktadır [86].

Metin sınıflandırma işlemi, metinlerin hangi özellik/nitelik/özniteliklerle temsil edileceği ve hangi algoritma ile sınıflandırma yapılacağı şeklinde iki temel alt probleme indirgenebilir [5]. Metin madenciliği safhaları ile benzer süreçler içeren sınıflandırma işlemleri; önişlemler, verinin temsil edilmesi, öznitelik seçimi, sınıflandırma ve performans değerlendirmesi şeklinde safhalandırılır [87]. Bu safhalardan metinlerin, analize hazır bir yapıya dönüştürülmesini kapsayan önişlemler, veri analizi için kullanılan toplam süre ve enerjinin % 50 - % 85'ini alabilmektedir [88]. Bununla birlikte metinleri ön işlemden geçirmek metnin sınıfının bulunması sırasındaki hesaplama yükünü azaltmakta ve sınıflandırma başarısını artırmaktadır [89].

Duygu analizinin, sosyal medya platformlarından metin içerikli Twitter (X)'a uyarlanmış bir örneği Şekil 2-4'te verilmiştir [90], gösterilen içerik ve detaylar çalışmaya göre farklılık gösterebilir. Günlük hayatın bir parçası haline gelen internet ve sosyal medya, metin sınıflandırma ve duygu analizinin yaygın olarak kullanıldığı alanlardan biridir. İnsanlar aldıkları ürün ya da hizmet konusunda edindikleri izlenimleri, olumlu ya da olumsuz düşüncelerini, herhangi bir konuyla ilgili bakış açılarını internet tabanlı uygulamalar üzerinden rahatlıkla ifade edebilir ve bunu diğer kullanıcılarla paylaşabilirler. Müşteri geri bildirimi niteliğindeki bu düşünce ve görüşlerin, toplanması ve işletmelerin kullanımına sunulması maksadıyla çeşitli platform, uygulama, puanlama, geri bildirim, yıldız verme gibi değerlendirme sistemleri geliştirilmiştir. Bu geri bildirim alternatiflerinin dışında, kullanıcıların herhangi bir ürün ya da hizmet hakkındaki samimi duygu ve düşüncelerini sohbet ortamı içinde evde, işyerinde dinlenirken ya da çalışırken, yoğun trafikte beklerken diğer kullanıcılarla paylaşma imkânı da bulunmaktadır ve bu türden samimi paylaşımlar diğer kullanıcılar üzerinde daha etkili olabilir. Facebook, Instagram, Twitter (X) gibi uygulamalar, kullanıcılara bu imkânı sağlayan platformlardandır. Twitter (X) ise diğer platformlardan farklı olarak metinsel paylaşımları esas alan bir uygulamadır. Bu ortamlarda paylaşılan duygu ve düşüncelerin analiz edilmesi, incelenmesi ve işletmeler için kullanılabilir hale getirilmesi diğer metin dokümanlarından farklı özellikler içerir, özel dil ve metin işleme tekniklerine ihtiyaç duyar. Ayrıca sosyal medya uygulamaları kendi kullanım özelliklerine bağlı olarak, farklı analiz safhaları, bu uygulamalarda günlük dil kullanımının yaygın olması nedeniyle farklı analiz zorlukları içerebilir.



Şekil 2.4. Türkçe Twitter (X) mesajlarının duygu analizi için örnek bir sistem modeli[90].

Metinsel verilerin analizindeki zorluklarının aşılması için, doğal dil işleme (DDİ) araçları kullanılır. DDİ, metinden bazı anlamsal bilgilerin elde edilmesine yönelik olarak, cümlenin parçalara ayrılması, metin içindeki parçalarının etiketlenmesi, özellik çıkarımı, yazım denetimi gibi dilbilimsel işlemler için kullanılır [91, 92]. Bu işlemler, her dilin kendi özellik ve yapısına göre gelişme gösterir ve bir dil için geliştirilen yöntem ya da araçlar diğer diller için uygun olmayabilir. Örneğin yazım düzeltmesi amacıyla diğer diller için geliştirilen ve sonlu bir sözcük dağarcığı kabulüne dayanan tekniklerin; bitişken bir dil yapısına sahip olan ve yapım ekleri, bileşik kelimeler gibi birçok nedenle analizi diğer dillerden farklılık gösteren Türkçe için geçerli olmadığı belirtilmektedir [93, 94].

DDİ süreci çalışmanın içeriğine bağlı olarak; kelimenin kök ve gövde gibi yapısı ile ilgilenen *biçimsel,* kelimenin kendisi ile ilgilenen *sözcüksel, gramer yapısal,* isim ve sıfat olması gibi kelimenin türü ile ilgilenen *sözcük türü (Position of Subject-POS),* kelime anlamı ile ilgilenen *anlamsal*, *içerik* ve *fonetik* gibi farklı analiz seviyelerinde yapılabilir [91, 95, 96].

Metin sınıflandırma için DDİ süreçleri kullanılarak, dilin özelliklerine uygun modeller geliştirilir. Her dilin kendine has özellikleri bu süreci değişik şekillerde zorlaştırabilir. Türkçe için DDİ sürecini etkileyen, zorlaştıran özellikler arasında genel olarak şu özellikler belirtilir: Kelimelerin sondan eklemeli olması; cümle içinde kelimelerin serbest dizilimi; *kutupsal dönüştürücü ya da tutum değiştirici etken* olarak isimlendirilen "değil" gibi kelimelerin kullanımı; kelimenin anlam yönünü değiştirebilen -ma, -me, -suz ve -lu, -li gibi

eklerin, kelime kökleri kullanılarak yapılan sınıflandırmalarda göz ardı edilmesi; cümledeki iki olumsuz kullanımın olumlu anlama dönüşmesi; *kuvvetlendirici ve zayıflatıcı (çok, az, fazla gibi kelimeler)* ifadelerin polarite değerinin belirlenmesi; koşul cümlelerinin nasıl çözümleneceği; soru gibi gözüken cümlelerin puanlandırılması; alaycı anlam kullanımları; deyimlerin içerdiği kelime anlamlarından farklı anlamlar taşıması; kelimelerin metnin ilgili olduğu alana özel olarak anlamının değişebilmesi; aynı cümle içinde olumlu ve olumsuz anlamların birlikte kullanımı gibi konular DDİ süreçlerinin çözüm aradığı başlıca zorluklardır [86, 97-99]. Diğer dillerde de ortaya çıkabilen bu tür etkenler, her dilin yapısına bağlı olarak farklı çözümlere ihtiyaç duyabilir. Duygu sınıflandırması için kullanılan bu çözümlerden biri *sözlük tabanlı sınıflandırma yaklaşım*dır. Sözlük tabanlı yaklaşımda, belirlenen her kategori için o kategoriyi temsil eden sözlükler hazırlanır, kategori sözlüğü ile sınıf etiketlemesi yapılacak doküman arasında karşılaştırma yapılır ve karşılaştırmadan elde edilen sonuca göre dokümana sınıf etiketi verilir. Duygu analizinde pozitif, negatif, nötr, öfke, tatmin, güvensizlik gibi sınıflandırmalar yapılabilir ve sınıflara göre sözlükler hazırlanır. Bu sözlükler duygu sözlüğü (sentiment lexicon) olarak isimlendirilir. Metin sınıflandırma için, analiz edilecek veri seti derleminden elde edilen sözlüklerden ya da genel sözlüklerinden istifade edilir [63, 100].

Kapsam ve hazırlanma amacına göre duygu analizi ve sınıflandırmalar için kullanılan sözlükler; genel ve özel (alana özel) amaçlı sözlükler şeklinde iki gruba ayrılır. Yapılacak çalışmaya, veri setine, kullanılacak algoritma ve araçlara bağlı olarak hangi sözlüğün tercih edilebileceği konusunda karar verilebilir. *Genel amaçlı sözlük* kullanımında cümle içindeki terimler, bağlamına bakılmaksızın değerlendirilmektedir. Bu sözlükler ilave bir çabaya gerek kalmadan her alan için kullanılabilirken, bazı özel alanlarda kullanımında performans kaybına neden olabilmektedir. Alana özel terimleri içermemesi ve terimler için bağlamına bakılmadan tek sabit bir değerin kullanılması özel amaçlı sözlük kullanımının tercih edilmesini sağlamaktadır. *Özel amaçlı sınıflandırma sözlükleri,* terimin kullanıldığı metin içindeki özel anlamının dikkate alınmasını sağlar, ayrıca terimlerin bu sözlüklerdeki polarite değerleri genel amaçlı sözlüklerden farklılık gösterebilmektedir [74, 86]. Sınıflandırma için eğer hazır bir sözlük kullanılmayacaksa, çalışmaya uygun bir sözlüğün hazırlanması gerekecektir. Sözlük hazırlama için kullanılan iki yöntem vardır. Biri başka bir dilden çeviri, diğeri tohum kelime listesi yöntemidir. Çeviri yönteminde otomatik, manuel çeviri ve hibrit yöntem olmak üzere üç farklı yol kullanılabilir. Otomatik çeviri ile hazırlanan sözlükler hızlı

bir sonuç verirken, duygu sözlüğünün bulunmadığı durumlarda ihtiyacı karşılayacak kadar dar kapsamlı bir duygu sözlüğü hazırlanabilir [86].

Sözlük oluşturmanın en maliyetli yolu, sınıflandırma ya da duygu sözlüğünü *manuel hazırlama yaklaşımıdır*. Bu tür yaklaşımda, sözlüğün kapsamı nispeten dar olmakla beraber, doğruluk derecesi diğer yaklaşımlardan daha yüksektir. Literatürde bu yaklaşım, duygu sözlüğünün bulunmadığı veya edinilemediği durumlarda, ihtiyacı karşılayacak kadar dar kapsamlı bir duygu sözlüğünün geliştirilmesi şeklinde kullanılır [86]. Sözlük geliştirme için manuel hazırlanan bir tohum kelime listesi eş-zıt anlamlı kelimelerle genişletilerek kullanılabilir, bu genişletme için kullanılabilecek kaynaklardan biri WordNet [101] sözlüğüdür. WordNet İngilizce'deki duygu analiz çalışmaları içinde temel teşkil etmektedir ve İngilizce için oluşturulmuş sözlüksel bir veri tabanıdır. Sözlükteki 117 000 eş anlamlı kelime, kavramsal olarak diğer anlamdaş kelime kümeleriyle ilişkilendirilmiştir [74]. En yaygın kullanıma sahip duygu sözlüğü SentiWordNet ise, İngilizce genel amaçlı bir duygu sözlüğüdür ve birçok çalışmada farklı dillere çevrilerek kullanılabilmektedir. Çeviri ile yapılan sözlük çalışmaları için, Google Translate API, Tureng, Zargan ve tr.bab.la gibi sözlük uygulamalarından istifade edilebilmektedir. Metin sınıflandırma ve duygu sınıflandırması için istifade edilen diğer sözlükler arasında Bing, Afinn, Laughran, NRC, Textblob, Bing Liu Görüş sözlüğü yer almaktadır.

Tohum kelime ile sözlük oluşturmada, az sayıdaki kelime listesi, eş-zıt anlamlar gibi yöntemlerle ihtiyaca göre genişletilmiş bir sözlüğe dönüştürülebilir. Bu kapsamda yaptığı çalışmada Hu ve Liu [102] duygu polaritesi bilinen 30 kelimelik bir tohum kelime listesini, müşteri yorumlarının duygu analizinde kullanmak maksadı ile WordNet kütüphanesindeki eş ve zıt anlamları kullanarak sözlük yapısına kavuşturmuştur. Söz konusu çalışmada, içerisinde 4.783 negatif ve 2.006 pozitif kelime, kelimelerin yanlış yazılmış halleri, kelimelerin morfolojik çeşitleri ve argo kelimeleri barındıran bir sözlük oluşturulmuştur [74]. Alqasemi vd. [103], Arapça dilinde otomatik duygu sözlüğü oluşturmak için tohum terimleri ve derlem terimleri arasındaki kosinüs uzaklığı kullanarak terim ayrım vektörü olarak isimlendirilen yeni bir yaklaşım önermişlerdir. Blair-Goldensohn vd. [104], Remus vd. [105] tohum kelime listesi yöntemi ile duygu sözlüğü geliştirmişlerdir.

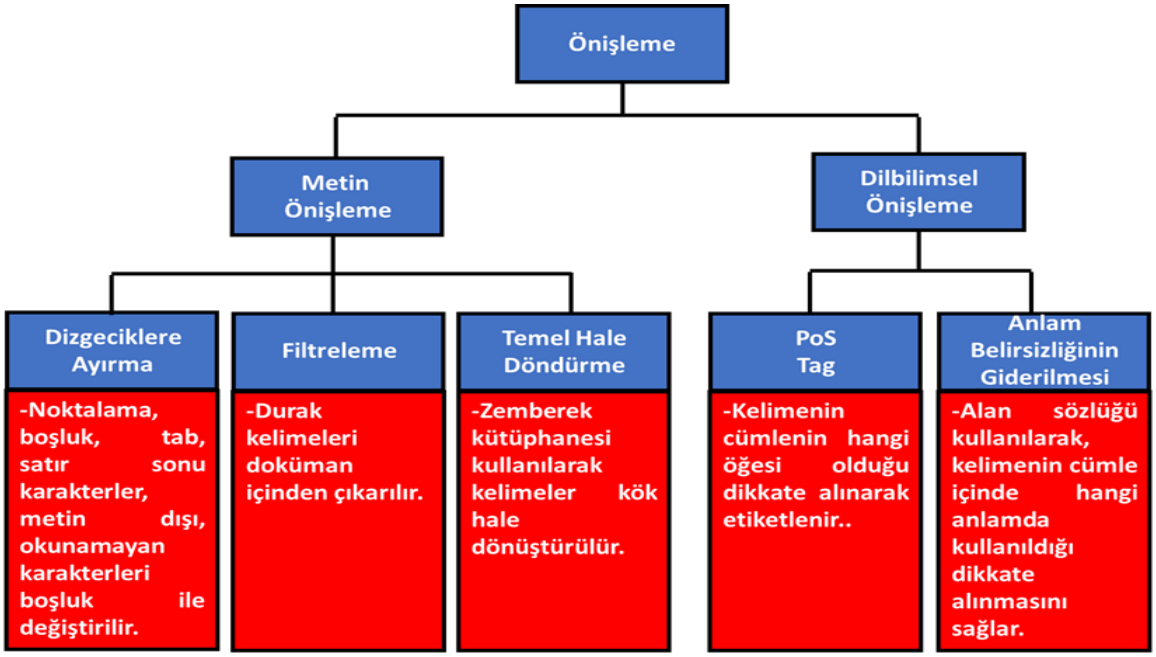
Literatürde tohum kelime listesinden üretilmiş alana özel sözlükler, otomatik ya da manuel çeviri yöntemi ile geliştirilmiş sözlükler, deyim ve atasözlerinin de dahil edildiği özel

sözlüklerin kullanıldığı çok sayıda metin sınıflandırma çalışmaları mevcuttur [9, 16, 97, 101, 104, 106-116]. Türkçe metinlerin sınıflandırması için kullanılabilecek, genel amaçlı bir Türkçe Derlemi (Turkish Corpus-TS Corpus), Sözlü Türkçe Derlemi (STC-Spoken Turkish Corpus), SentiTürkNet [117] gibi çalışmalar mevcuttur.

2.4.2. Metin sınıflandırma ve duygu analizi safhaları

Metin analizleri genel olarak önişlemler, verinin temsil edilmesi, verinin sınıflandırma algoritmaları ile işlenmesi, sonuçların görselleştirilmesi ve değerlendirmesi aşamalarından meydana gelir. Bu aşamaların detayları ileriki maddelerde sıralanmıştır.

Metinsel veri setleri için uygulanan ilk basamak, metin önişleme ve temizlik aşamasıdır. Bu aşama, analiz sürecinin en fazla zaman alan aşamasıdır. Yapılacak analiz çalışmasına göre içeriği değişmekle birlikte metin önişlem aşamasında yapılan işlemler özet olarak Şekil 2.5'te gösterilmiştir.



Şekil 2.5. Metin önişlem adımları

Veri seti içinden, gereksiz, tutarsız, boş verilerin ya da algoritmaların hatalı işlem yapmasına neden olabilecek verilerin çıkarılmasına veri temizleme denir [118]. Eğer veri seti derlemi oluşturularak analiz yapılacaksa metin içindeki konuyla ilgisi olmayan tüm kelimelerin metinden temizlenmesi gerekir. Burada amaç dokümanda, sadece sınıflandırma için

kullanılacak etkili ve daha fazla bilgi barındıran kelimelerin kalmasını sağlamak, ilgisiz kelimeleri ayıklamaktır. Öte yandan konuyla ilgisiz kelimelerin tespit edilmesi de kolay bir işlem değildir [49]. Eğer analiz için sözlüksel yaklaşım metodu ile bir sınıflandırma yapılacaksa bu durumda metin önişleme ve temizliğin kapsamı değişecektir. Her iki durumda da analiz öncesi metin önişlem ve temizlik işlemlerini, analiz için metinde nelerin kullanılacağı ya da kullanılmayacağı belirleyecek, yapılacak çalışmanın amaç ve kapsamına göre bu işlemlerin önceden tespit edilmesi gerekecektir.

Tweetler gibi sosyal medya platformlarından elde edilen metinsel içeriklerin ayıklanması diğer metinlerden farklı işlemler gerektirir. Örneğin tweetlerden oluşan bir veri setinde, retweetlerin veri setinden çıkarılıp çıkarılmayacağına karar verilmesi gerekir. Bazen retweetler, veri kümesinde çoğunluğu oluşturarak alınacak sonuçları değiştirebileceği, analiz sonuçlarında tutarsızlıklar oluşturabileceği için veri setinden ayıklanabilir [64]. Bazen de kullanıcıların paylaştıkları ve kabul edilen görüşleri yansıttığı değerlendirilerek veri setinden ayıklanmasına ihtiyaç duyulmayabilir. Eğer kullanıcı tarafından retweet yapılırken yorum eklenmiş ise, bunun ayrıca dikkate alınarak veri setine ve analize dahil edilmesi gerekir. Çünkü RT'ye eklenen yorumlar yeni görüşler içerebilir ve bu açıdan diğer tweetlerden bir farkı yoktur. RT'nin başkasına ait bir görüşün paylaşılması ve eğer olumsuz bir görüş ilave edilmemişse paylaşılan görüşün onaylanması anlamına geldiği ve analize dahil edilmesi gereken yeni bir metin/doküman ortaya çıktığı kabul edilebilir.

Metin veri setleri için uygulanabilecek önişlem ve temizlik işlemleri ileriki maddelerde özet olarak açıklanmıştır.

- *Dönüştürme, dizgeciklere ayırma:* Veri setlerinde öncelikle, xml ve html dosyaları şeklindeki metinlerin düz metne dönüştürülmesi; Türkçe/İngilizce karakterlerin dönüşümü, büyük/küçük harfe dönüştürme, yazım hatalarının düzeltilmesi (Zemberek kütüphanesinin önerme ve *deasciification*[3] fonksiyonu ve Apache Lucene'in uzaklık algoritması gibi araçlar kullanılabilir), kelime içinde tekrar eden ve kelimenin orijinal halinde olmayan karakterlerin temizlenmesi gibi dönüştürme ve standartlaştırma işlemleri yapılabilir [7, 51, 119, 120].

[3] Türkçe karakter yazım hatalarını ASCII kod tablosunu kullanarak yapılan düzeltme işlemine verilen isimdir, bu işlem ile Türkçe metin içindeki İngilizce karakterlerin dönüşümü yapılır. Örneğin "caldim" ifadesi deasciification fonksiyonu kullanılarak "çaldım" şekline dönüştürülür.

Metinlerin standart hale dönüştürülmesi için uygulanan bir başka işlem de dizgeciklere ayırmadır. Metinlerden tüm noktalama işaretleri, kelimeler arasındaki birden fazla boşluklar, metin dışı işaret, şekil ve semboller (simge, emoji, rakam, resim gibi) temizlenir ve boşlukla değiştirilir [51]. Bu işlemlerin yapılacak analize göre değişebileceği göz ardı edilmemelidir. Örneğin metin içindeki rakamlar, analiz için kullanılacaksa veri setinden ayıklanmayabilir. Metin temizliğinde kelimeler arası boşlukların tek boşluğa dönüştürülmesi işlemi, diğer tüm temizlik işlemlerinden sonra yapılmalıdır. Çünkü Excel ile yapılacak temizlik işlemleri sonrasında yeni boşluklar ortaya çıkabilmektedir.

- *Filtreleme:* Tek başına bir anlamı olmayan edat, bağlaç ve zamir gibi durak kelimeler, 3 karakterden daha kısa kelimeler ve özel isimlerin metinlerden çıkarılması yaygın olarak kullanılan filtreleme yöntemlerindendir [82, 119, 120]. Bu çalışma kapsamında hazırlanan durak kelime listesi EK-1'de verilmiştir.

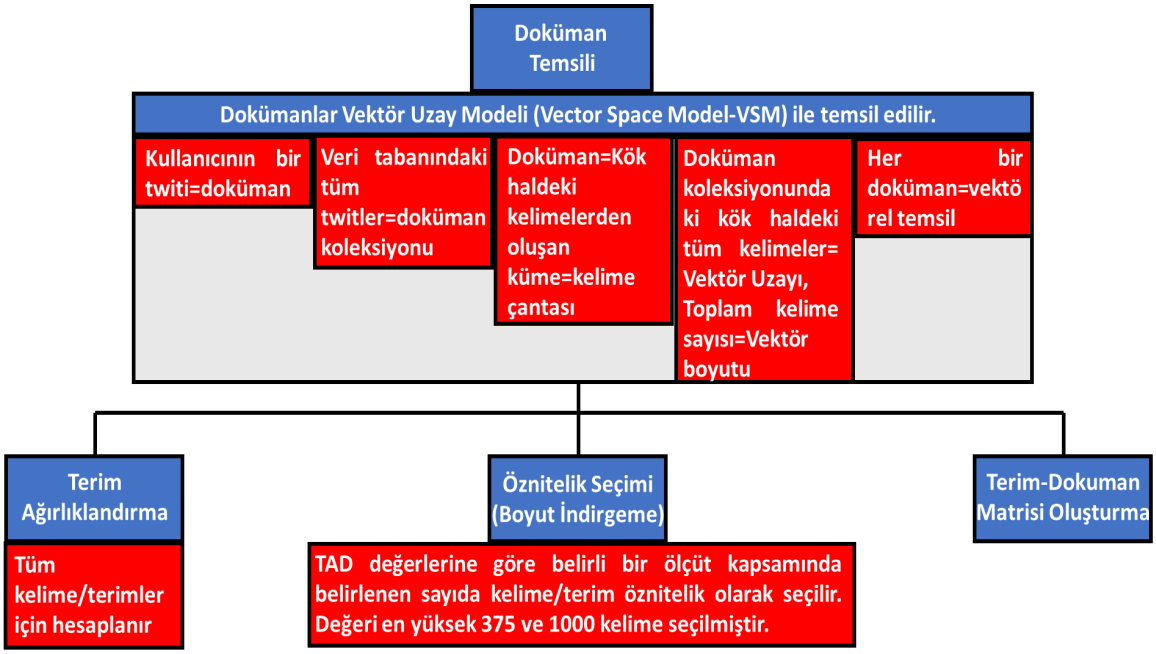
Sözlüksel yaklaşım metodu için, metin temizliği öncesinde sınıflandırma için kullanılacak sözlüklerde yer alan kelimeler, metin içinden temizlenmesi planlanan kelime/kelime grupları ile karşılaştırılmalıdır. Her iki listede aynı kelimeler yer almamalıdır. Burada maksat sınıflandırma sözlüğünde yer alan bir kelime/kelime grubunun dokümanlar içinden temizlenmesini engellemektir. Örneğin, üç harfli olduğu için dokümanlardan temizlenen bir kelime, bir başka kelime ile kelime grubu oluşturarak sınıflandırma sözlüğünde yer almış olabilir, ancak eğer kelime koleksiyondan silindiyse kelime sayımında frekans değeri sıfır olarak ortaya çıkacak ve sınıflandırmaya etki edemeyecektir.

- *Kök hale dönüştürme (stemming), temel hale dönüştürme (lemmatization):* Metin içindeki aynı kelimelerin algoritmalar tarafından yakalanabilmesi için, kelimelerin eklerinden kurtarılması gerekir. Bu işlem kelimelerin öncelikle kök hale dönüştürülmesi ile yapılır. Çoğul ekleri çıkarılarak kelimeler tekil forma, fiiller kök hale getirildikten sonra mastar hale dönüştürülür [51]. Türkçe'nin sondan eklemeli bir dil olması, terim uzayının büyüklüğü nedeniyle zaman alıcı bu işlemlerden genellikle köke indirgeme tercih edilen bir yöntemdir [121]. Uçan [74] çalışmasında kelimelerin hem ekli hem de kök halleriyle yapılan deneylerde kök halinin %2 daha yüksek başarı ürettiğini tespit etmiştir. Yapılan bazı çalışmalarda ise Türkçe metinlerin sınıflandırılmasında kök bulmanın etkisinin çok az olduğu ifade edilmektedir [122]. Kelimelerin gövde köklerinin bulunması maksadıyla kullanılabilen uygulamalar arasında; PCKimmo [123], Türkiye Türkçesi için yaklaşık

30.000 kök içeren NLP kütüphanelerinden Zemberek [119, 124], NLTK (Natural Language Toolkit-Doğal Dil Araç Takımı) kütüphanesi ile entegre olan Turkish Stemmer bulunmaktadır.

- *Cümle öğesinin etiketlenmesi (POS Tag), anlam belirsizliğini giderme (word sense disambiguation):* Cümle öğesinin etiketlenmesi, sözcüğün cümlenin hangi öğesi olduğunun tespit edilmesi ve etiketlenmesidir. Aldıkları bu etiketler, analize göre kelime puanlarının ağırlıklandırılması için kullanılabilir. Anlam belirsizliğini giderme, sözcüğün cümle içindeki anlamının belirlenerek dikkate alınmasını sağlayan yöntemlerdir. WordNet sözlüksel veritabanı gibi sözlük kullanımına yönelik yaklaşımlar, bu belirsizliklerin giderilmesi için kullanılabilir [51, 125].

Dokumanların temsil edilmesine yönelik işlem safhaları Şekil 2.6'da gösterilmiştir. Öncelikle kelimelerin, doküman ve koleksiyon açısından ayırt edicilik değeri / ağırlığı *terim ağırlıklandırma* yöntemleri ile belirlenir. Ağırlıklandırılmış terimlerden seçilenler (öznitelik/nitelik/değişken seçimi), sınıflandırılacak dokümanlarla matris yapıda ile bir araya getirilir. Öznitelik seçimi, işlem hacminin azaltılması, işlem süresinin kısaltılması için gereklidir. Aksi takdirde küçük miktardaki dokümanlardan bile çok yüksek boyutlu matrisler ortaya çıkabilir. Dokümanların temsil edilmesinde sıklıkla kullanılan yöntemlerden biri diğeri de bir dizilimdeki tekrar (tekrar eden harf/kelime) oranını bulmaya yarayan n-gram'dır. Arama algoritmalarında da kullanılan n-gram yöntemi, tekrar eden kelimeleri ortaya çıkarır ve bu tekrarlar için bir oran hesaplar [126].



Şekil 2.6. Dokümanların temsil edilmesi işlemleri ve matris yapının oluşturulması

Analiz edilen bir doküman içinde geçen terimlerin, doküman açısından nasıl bir anlam (terimin doküman ile ne derece ilgili olduğunun tespiti için) ifade ettiğinin, ölçülebilir bir değere dönüştürülmesi, analiz açısından gerekli bir işlemdir [127]. Terim ağırlıklandırma, veri kümesindeki kelimelere bir ağırlık değeri ataması yaparak; önemli kelimeleri ortaya çıkarır, dokümana bir katkısı olmayan kelimelerin tespit edilmesini sağlar. Bunun sonucunda dokümanın ait olduğu sınıf tespit edilebilir. Terim ağırlıklandırma yöntemlerinden hesaplamalarda; sınıf bilgisini kullananlar *gözetimli terim ağırlıklandırma*, sınıf bilgisini kullanmayanlar *gözetimsiz terim ağırlıklandırma* yöntemi olarak isimlendirilir [128].

Terim ağırlıklandırma için kullanılan çok sayıda klasik yöntemler olduğu gibi geliştirilen yeni yöntemler de mevcuttur. Kelime torbası gibi sayım tabanlı klasik yöntemlerden, cümle içindeki kelime frekansını esas alan sayım vektörü (Count Vector-CV); terim frekansı (Term Frequency-TF), ters doküman frekansı (Inverse Document Frequency-IDF), varyans değeri, TF-IDF metodu, çoklu sınıflandırmaya uygun TF-bilgi kazanım metodu (Information Gain Method-IGM), terimlerin sınıf frekanslarını dikkate alan eden TF-IDF-ters koleksiyon frekansı (Inverse Corpus Frequency-ICF) ve TF-IDF-ters küme boyutu dokuman frekansı (ICSDF) gibi istatiksel yöntemlerle başarılı şekilde özellik seçimleri yapılabilmektedir. Ayrıca n-gram istatistikleri gibi temel teknikler de yaygın olarak kullanılmaktadır. Klasik yöntemlerin dışında PATricia (PAT) ağacı, SWN kelime grupları, grafik-tabanlı yöntem, popüler derin öğrenme tekniği olan Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri

(Bidirectional Encoder Representations from Transformers (BERT) mimarisi ve bu mimari üzerine kurulan BERTweet gibi yöntemler de özellik seçimi için kullanılan yeni yöntemlerdendir [49, 129-136].

Boyut indirgeme işlemi, analiz sürecinin ilk adımı olan önişlemler ile başlar ve yukarıda sıralanan terim ağırlıklandırma yöntemleri kullanılarak yapılan puanlandırma sonrasındaki seçimle tamamlanır. Metin önişleme ve temizlik işlemleri bir yandan koleksiyondaki dokümanların işlenmesi için standart bir format oluşturmayı hedeflerken, durak kelimesinin ayıklanması gibi işlemler koleksiyon boyutunun küçültülmesini sağlar.

Öznitelik seçimi, metin madenciliğinin önemli sorunlarından *çokboyutluluk (curse of dimensionality)* ve *öznitelik seyrekliğinin (feature sparsity)* işlem hacminde neden olduğu büyümeyi önlemek ve algoritmaların hızlandırılması için gereklidir. Tweet metinleri gibi gündelik ifadeleri (sokak ağzı, jargon, argo kelimelerin kullanılması) içeren doğal dil dokümanlarındaki özniteliklerin sayısının çokluğu ve seyrekliği, diğer metin türlerine göre daha çok karşılaşılan bir problemdir [[49, 68, 137].

Boyut indirgeme için bütün dokümanlarda çok sık ya da çok az sayıda geçtiği için istatistiksel olarak belirgin bir yararı olmayan sözcükler koleksiyondan çıkarılabilir. Örneğin dokümanların %95'inden daha fazlasında bulunan ya da en fazla sayıda kullanılan ilk beş kelimenin, en az sayıda kullanılan son beş kelimenin çıkarılması gibi yöntemler boyut indirgeme için kullanılabilir. Durak kelimelerine benzer şekilde edat, bağlaç, zamir gibi kelime türlerinin doküman sözlüğünden çıkarılması işlemi de bir başka boyut indirgeme yöntemi olarak kullanılabilir ya da sözlüklerde fiil, isim, sıfat, ünlem, zarf kelime türlerinden seçilecek olanların kullanılması boyut indirgeye hizmet eder [51, 82, 138]. Terimlerin ağırlık puanları hesaplandıktan sonra, belirlenecek bazı kurallar çerçevesinde terimlerin bir kısmı öznitelik olarak seçilirken bir kısmı ayıklanarak boyut indirgeme sağlanabilir. Burada amaç hangi miktarda özniteliğin en iyi performans gösterdiğinin tespit edilmesi, çok boyutluluk ve seyreklik probleminin etkisinin azaltılmasıdır.

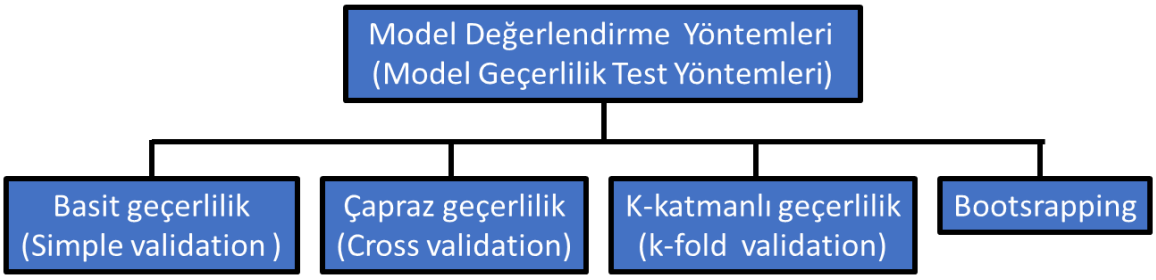
Öznitelik seçiminde kullanılan bu yöntemler iki grup altında toplanabilir. Eşik değer, artırımlı, azaltımlı seçim yöntemleri gibi *filtre tabanlı/ istatiksel tabanlı yöntemler* [127, 139, 140] ve *sarmalama tabanlı/sınıflandırmaya dayalı yöntemler* olmak üzere iki temel yöntem kullanılmaktadır. Örneğin Akba [137]'nın çalışmasında artırımlı bir yöntemle, aynı veri

setindeki kelime sayısını 250, 375, 500, 750, 875, 1 000, 1 125 ve 1 250 şeklinde uygulayarak 8 farklı matris oluşturmuş ve 375'lik kelime grubunun en iyi performansı verdiğini tespit etmiştir. Artırımlı yöntemin tersten uygulaması da azaltımlı yöntem olarak isimlendirilir. Sarmalama tabanlı öznitelik seçim yöntemlerinde ise belirli bir sınıflandırma algoritmasının başarımına dayalı olarak öznitelikler seçilir [139]. Hangi öznitelikler daha iyi sınıflandırma başarısı veriyorsa onları seçip diğerlerini eleme şeklinde uygulanır. Örneğin başlangıçta tespit edilen doküman sözlüğündeki tüm terimler öznitelik olarak seçilir ve müteakiben öznitelikler teker teker ya da belirli miktarlarda azaltılarak uygulama yapılır ve en iyi performansı veren öznitelik miktarı seçilir [140].

Sıralanan bu öznitelik seçim yöntemleri, aynı zamanda algoritma ve programlar tarafından otomatik olarak kullanılmaktadır. Örneğin metin analiz programı WEKA'nın CfsSubsetEval fonksiyonu, filtre tabanlı öznitelik seçim yöntemini kullanır. Söz konusu fonksiyon, korelasyon tabanlı özellik seçim (Correlation-based Feature Selection-CFS) algoritmasını kullanır. XGB ve Hafif Gradyan Güçlendirme Makineleri (Light Gradient Boosting Machines-LightGBM) gibi algoritmaların kullandığı sıralı özellik seçim (Sequential Feature Selection-SFS) ve tekrarlamalı özellik eleme (Recursive Feature Elimination-RFE) yöntemleri ise sarmalama tabanlı öznitelik seçim yöntemleridir.

Koleksiyonun sınıflandırma için algoritma tarafından işlenebilmesi için doküman, değişken görevindeki öznitelikler ve bu özniteliklerin doküman bazında değerini gösteren ağırlıklandırılmış değerlerinin matris bir yapıya dönüştürülmesi gerekir. Bu matris yapıda ayrıca dokümanların gerçek sınıflarını gösteren hedef değişkenler de yer almalıdır. Bu yapı doküman terim matrisi (DTM) olarak adlandırılır. DTM'de satır başlarına, koleksiyonu oluşturan dokümanlar, sütun başlarına öznitelikler/değişkenler yazılır. Doküman ile değişkenlerin kesiştikleri hücreye değişkenin o doküman için hesaplanmış terim ağırlık değeri yazılır. DTM'nin son sütununda ise dokümanların gerçek sınıf değerleri yazılır. Sınıf değerleri, iki sınıflı bir koleksiyon için 0 ve 1 değerleri şeklindedir [63].

Veri setleri makine öğrenmesi yöntemleri için önce eğitilir, sonra test edilir. Bu maksatla veri setinin bir kısmı eğitim veri seti, bir kısmı test seti olarak ayrılır. Bu ayırma işlemi için kullanılan yöntemler Şekil 2.7'de gösterilmiştir.



Şekil 2.7. Model geçerlilik test yöntemleri

Basit geçerlilik yönteminde veri seti ikiye bölünür, literatürde genelde veri setlerinin %70-80'i eğitim, geri kalanı test seti olarak kullanılır. Ancak veri setinin dengesizliği dikkate alınarak bu oranda değişiklik yapılabilir. Akpınar [88] çalışmasında veri setinin %5-%33 arasındaki bölümünün test veri seti olarak ayrılabileceğini belirtmiştir.

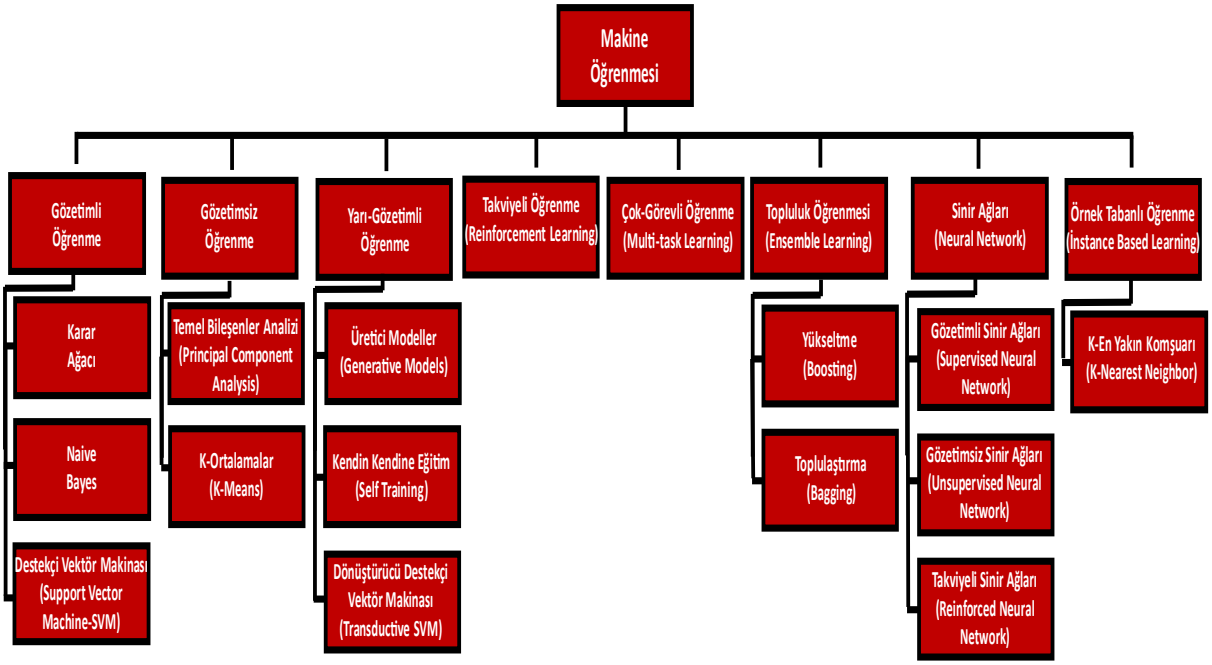
Sınırlı sayıda dokümanın olduğu veri setlerinde *çapraz geçerlilik yöntemi* kullanılabilir. Bu yöntemde veri kümesi tesadüfi olarak iki eşit parçaya ayrılır, her iki parça dönüşümlü şekilde eğitim ve test seti olarak kullanılır ve elde edilen ortalama değerler kullanılır. Eğitim ve test veri seti olarak ayırma için kullanılan en yaygın yöntemlerden biri *k katmanlı çapraz doğrulama* yöntemidir. Bu yöntem, gerek eğitim ve gerekse test veri setlerinin yeterli temsil yeteneğine sahip olması ve sınıflandırma başarımını artırmak maksadı ile kullanılan bir yöntemdir. Bu yöntem, çapraz doğrulama yönteminin farklı bir uygulamasıdır. Bu yöntemde veri seti ikiden daha fazla (k sayısı kadar) sayıda eşit parçaya bölünür ve tüm parçalar dönüşümlü olarak eğitim ve test seti olarak kullanılır. Bu yöntem birkaç bin veya daha az satırdan meydana gelen küçük veri tabanlarında tercih edilir ve her seferde elde edilen hata oranlarının ortalaması, kurulan modelin tahmini hata oranı olarak kabul edilir. Genelde çalışmalarda k sayısı 5 ya da 10 olarak belirlenmektedir. Küçük veri kümeleri için modelin hata düzeyinin tahmininde kullanılan bir başka yöntem *bootstrapping* yöntemidir. Çapraz geçerlilikte olduğu gibi model bütün veri kümesi üzerine kurulur. Daha sonra en az 200, bazen binin üzerinde olmak üzere çok fazla sayıda öğrenim kümesi tekrarlı örneklemelerle veri kümesinden oluşturularak hata oranı hesaplanır. Bu yöntemler karma şekilde de kullanılabilmektedir [88, 121].

2.4.3. Modelleme ve sınıflandırma algoritmaları

Sınıflandırma modelleri için kullanılan algoritmalar, veri seti özellikleri ve model parametrelerine ilişkin hususlar ileriki maddelerde incelenmiştir. Veri yığınlarının büyüklüğü, veri havuzundan değerli bilginin ortaya çıkarılması, verinin analiz edilmesi işlemlerinin otomatik sistem ve modellerle gerçekleştirilmesini zorunlu kılmaktadır. Otomatik işlemler için, bilgisayar sistemlerinin aynen insan eğitimi gibi eğitilmesi ya da bir sistemin öğretilmesi gerekir. Öğrenme, gözlem ve gerçek dünya ile karşılaştırmalardan elde edilen veri ve bilgilerle destekleyerek zaman içinde geliştirmeye olanak tanıyan bir bilim [141] olarak tanımlanmaktadır.

Makine öğrenmesi, temel olarak gözetimli (denetimli-supervised) ve gözetimsiz (denetimsiz-unsupervised) olmak üzere iki kategoride ele alınmaktadır. Sınıflandırma ve regresyon problemleri de gözetimli öğrenme modelidir. *Gözetimli öğrenmede*, işlem eğitilmiş veriler üzerinden gerçekleştirilir, sistemin girdi ve çıktısı önceden belirlenmekte ve girdi ve çıktı verileri arasındaki ilişki ile makinenin işleyişi öğrenmesi sağlanmaktadır. Öğrenme işlemi gerçekleştirildikten sonra, makineye yeni girdiler verilerek çıktıyı tahmin etmesi, sınıflandırması beklenmektedir [142]. *Gözetimsiz öğrenmede* daha önceden tanımlanmış etiketleme veya eğitim veri seti yoktur. Bunun yerine algoritma bir model kullanarak metindeki örüntüleri tespit eder. Burada araştırmacının etki edebileceği tek şey, metnin kaç sınıfa ayrıştırılacağı gibi bazı parametreleri belirlemesidir [114]. Gözetimsiz öğrenmede, sınıfları bilinmeyen girdi kümesi parçalara ayrılarak, verilerin gruplandırılması/kümelemesi yapılır. Kümeleme, en temel gözetimsiz öğrenme yöntemidir [142]. Kümeleme algoritmalarının seçimi, mevcut veri türüne, çalışmanın amaç ve uygulamalara bağlıdır [143]. Gözetimli ve gözetimsiz öğrenme, makine öğrenme modellerinin temelini oluşturmakla birlikte, öğrenme yöntemlerine göre makine öğrenme modelleri için yapılan bir sınıflandırma Şekil 2.8'de verilmiştir [144]. Algoritmaların öğrenme yöntemi olarak kullandıkları yaygın yaklaşımlardan biri, topluluk öğrenme yöntemleridir. Makine öğrenme algoritmalarının istikrarını artırmak, doğruluğunu geliştirmek, zayıf yönlerini güçlendirmek maksadıyla kullanılan bu yöntemler; *torbalama (bagging), yükseltme / güçlendirme (boosting), istifleme (stacking)* şeklinde üç grup olarak kategorize edilebilir. Bu yöntemler, algoritmalar için farklı yaklaşımlar kullanırlar. Örneğin, önyükleme (bootstrap) ve toplama (aggregation) olmak üzere iki kademeli işlemin

birleşimini ifade eden *toplulaştırma / torbalama (bagging) yöntemi*, birçok zayıf öğrencinin birleştirilerek tek bir güçlü öğreniciden daha güçlü bir öğrenici elde edilmesine imkan verir. Torbalama, algoritmaların aşırı öğrenme probleminden kaçınmalarına yardım eder. En yaygın torbalama algoritmalarından biri rastgele / rassal ormandır (Random Forest-RF) [145-147]. *Güçlendirme (boosting) yöntemi*, zayıf öğrenicilerin, güçlü öğrenicilere dönüştürülmesi metodudur. Zayıf sınıflandırıcıların oluşturduğu serilerle modeller kurulur, her bir model bir önceki modelin tahmin artıkları / hatalarını kullanır. Torbalama yönteminden farklı olarak bu yöntemde ağaçlar birbirine bağımlıdır. Uyarlayıcı Güçlendirme (Adaptive Boosting- AdaBoost), GBM, XGBoost algoritmaları özelliklerini bu yaklaşımdan almaktadır [148, 149]. *Yığın genelleme / yığınlama (stacked generalization / stacking),* yukarıda sıralanan iki yöntemden farklı olarak, veri setinin tamamını kullanır. Veri seti üzerinde kurulan farklı modellerden en iyi performanslar birleştirilerek bir meta öğrenici kullanılır [150].



Şekil 2.8. Makine öğrenme modelleri

Makine öğrenmesinde kullanılacak algoritmanın seçimi, çözülecek problem türü, problemdeki değişken sayısı ve probleme uygun model türü gibi etkenlere bağlıdır. Algoritmaların temelindeki öğrenme yöntemlerini gösteren Şekil 2.9'daki modellerden bazıları aynı isimli algoritma olarak kullanılmaktadır, literatürde model ve algoritmaların tasnifi konusunda farklı gruplandırmalar bulunmaktadır.

Veri sınıflandırma problemlerinde analiz sonuçlarının doğru değerlendirilebilmesi için, veri setlerindeki sınıf dağılımlarının dengeli olup olmadığı kontrol edilmeli, analiz sonuçları değerlendirilirken sınıf dağılımları dikkate alınmalıdır, dengesiz veri setlerinde mükemmele yakın elde edilecek doğruluğun çok başarılı olduğu düşünülmemelidir. Çarpık sınıf oranlarına sahip bir sınıflandırma veri kümesi dengesiz veri seti olarak, veri kümesinin büyük kısmı büyük sınıf, diğer kısmı azınlık sınıfı olarak isimlendirilir [151, 152]. Veri setlerindeki hangi dağılım oranının dengesizlik olduğu konusunda, Google Analitik tarafından şu tanımlamalar yapılmaktadır: (i) Ilımlı dengesizlik düzeyi veri kümesinin %20-40'ı, (ii) orta düzey dengesizlik veri kümesinin %1-%20'si, (iii) olağanüstü dengesizlik veri kümesinin %1'i olarak kabul edilir [152]. Dengesiz veri seti problemi, bu tür veri setlerinde daha iyi sonuçlar üreten algoritmaların tercih edilmesi ile aşılabilmekle birlikte bu konuda uygulanan genel yöntemler şu şekilde sıralanabilir: (i) Azınlık sınıftaki gözlem sayısı, bu sınıftaki gözlemlerden rastgele yöntemle kopyalama ile çoğaltılarak büyük sınıfın veri sayısı ile eşitlenir. (ii) Büyük sınıftaki gözlemlerden rastgele seçim ile ayıklama yapılarak bu sınıftaki gözlem sayısı, azınlık sınıftaki gözlem sayısı ile eşitlenir. (iii) Azınlık sınıftan yeniden örnekleme ile yeni örneklerin üretilmesini sağlayan Azınlık Sınıf Sentetik Örnekleme Tekniği (Synthetic Minority Over Sampling Technique-SMOTE) [153] algoritması ile veri setindeki azınlık sınıf gözlem sayısı çoğaltılır. (iv) Analiz sonuçlarının başarısının ölçümü için doğru ölçüm aracının kullanılması, dengesiz dağılımın analiz sonucuna etkisini ortadan kaldıracaktır. Bu maksatla ikili sınıflandırmalar için Alıcı İşlem Karakteristikleri Eğrisi Altındaki Alan (Area Under Receiver Operating Characteristic-AU-ROC) tekniğinin kullanılması doğru bir seçim olacaktır. Çünkü AU-ROC tekniği, veri setindeki sınıf dağılımından bağımsız çalışır, AU-ROC, sınıflandırıcının iki sınıf arasında ayrım yapma yeteneğini ölçer, bu durum modelin performansının dengeli bir şekilde ölçülerek değerlendirilmesini sağlar. (v) Yetersiz oranda temsil edilen azınlık sınıftaki gözlemlerin sınıflandırma hataları için, ceza puanı kullanarak sınıflar arasında dengeli bir model oluşturulmasını sağlayan cezalandırıcı özellikli Destekçi Vektör Makinesi (Support Vector Machine-SVM) algoritması kullanılır. Bu maksatla modellerdeki *sınıf ağırlığı* parametresi ile modelin azınlık ve çoğunluk sınıflardan eşit şekilde öğrenmesi sağlanır. (vi) İkili sınıflandırmalarda, her iki sınıftan gözlemleri dikkate alarak modelin eğitilmesini sağlayan RF ve GB gibi ağaç yapılı algoritmalar kullanılır. (vii) Azınlıkta kalan sınıflar birleştirilerek bir sınıf altında temsil edilir, böylece azınlık sınıf ortadan kaldırılmış olur. (viii) Azınlık sınıfı için, Rastgele Örneklem Uzlaşması (Random Sample Concensus-RANSAC), yerel aykırı faktör, izolasyon ormanı gibi aykırı değer tespiti yöntemleri ve

gözlemler arasında dengeli bir model oluşturulmasını sağlayacak matematiksel parametrelere sahip algoritmalar kullanılır [154]. (ix) Veri setine sınıflar arasında dengeli dağılımı sağlayacak sayıda yeni gerçek veri eklenir ve dengesizlik ortadan kaldırılır. (x) Sınıflandırma için birden fazla algoritma kullanarak modeller oluşturulur, oluşturulan tüm modellerin performansları aynı yöntemle ölçülerek karşılaştırması yapılır.

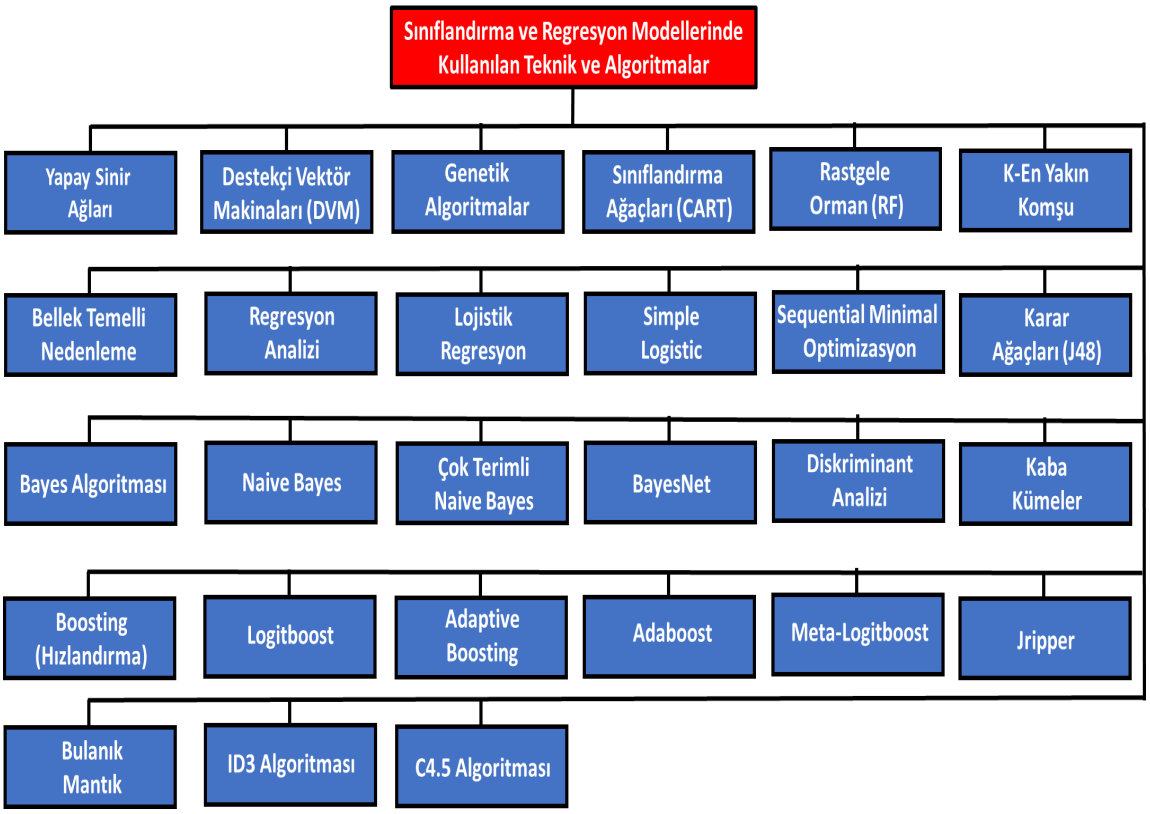
Sınıflandırma modellerinde karşılaşılan bir başka problem, makine öğrenmesi algoritmalarının aşırı ya da yetersiz öğrenmesi durumudur. Algoritmanın genel bir doğruya ulaşmak yerine veriyi ezberlemesi ve sadece o veri setine özel yüksek başarı elde edilmesi durumu aşırı öğrenme durumudur. Modellerde istenen sonuç, veriyi ezberlemesi değil, genel bir örüntünün elde edilmesi ve bunun diğer veri setlerinde uygulanabilmesidir. Algoritmanın veriler arasındaki örüntüyü ortaya koyamaması durumu ise yetersiz öğrenme/zayıf öğrenme/az öğrenme olarak nitelendirilir. Zayıf öğrenmede model, verilerin arasındaki ilişkinin mantığını yakalayamadığı doğru sonuçlar üretemeyecektir [155-157]. Bu tür öğrenme problemlerinin çözümü için yukarıda açıklanan k-katmanlı çapraz doğrulama yöntemi kullanılabilir [156]. Bunun dışında algoritmalar tarafından iterasyon sayısının sınırlandırılması, tahminlemeye katkısı düşük olan düğümlerin budanması, ceza puanı ile düzenlileştirme, sınıflandırıcı özelliği düşük değişkenlerin çıkarılması, öğrenmenin gürültüsü düşük daha fazla sayıdaki veri seti ile yapılması, torbalama ve artırımlı yöntemler gibi topluluk öğrenme yöntemlerinin kullanımı şeklinde yöntemler uygulanır [156].

Hiper-parametreler, modellerde kullanılan algoritmaların çalışmasını kontrol altında tutan, modelin eğitim sürecini denetlemeye olanak sağlayan ayarlanabilir parametrelerdir. Farklı değerler alabilen bu parametre değerlerinde değişiklikler yapılarak modelin en iyi performans sergilediği değerler ve buna bağlı optimize model tespit edilmiş olur. Örneğin, sinir ağlarını kullanan algoritmalarda gizli katmanların sayısına ve her katmandaki düğüm sayısına kullanıcı karar verir [158]. Benzer şekilde K-En Yakın Komşu (K-Nearest Neighbour-KNN) algoritmasında k değeri, SVM'de hangi kernel fonksiyonunun kullanılacağı, katman sayısı, maksimum derinlik gibi modele göre değişen hiper parametreler bulunmaktadır. Hiper parametre seçimi çeşitli değer aralıklarının model üzerinde denenmesi ile bulunabilir, öte yandan kullanıcının bu konudaki tecrübesi ve veri setinin özellikleri de bu seçim işleminde etkilidir. Bu işlem, hiper parametre seçimi, hiper parametre ayarı, hiper parametre iyileştirmesi (hyper parameter fitting / hyper parameter tuning) ya da hiper parametre uyumlaştırması olarak isimlendirilir [159]. Yüksek hesaplama

süreleri nedeniyle hiper parametre iyileştirmesi, model eğitim süresini uzatmaktadır. Zaman maliyetine rağmen bu işlem, her zaman model performansının artırılmasını garanti edemez. Tran vd. [160], hiper parametrelerin iyileştirildikten sonraki değerinin ya da algoritmanın varsayılan değerinin kullanımı arasında bir tercih yapabilmek için yöntem geliştirmişlerdir. Öte yandan hiper parametre iyileştirmesi için kullanılan teknikler, farklı veri setleri ve algoritmalarda model performanslarını artırma konusunda farklı sonuçlar üretebilir [161]. Bu çalışmada kurulan modellerde, hiper parametre iyileştirme işlemleri ile modellerin performanslarında genel olarak pozitif yönde artış olduğu gözlenmiş ve tüm modeller için bu işlem gerçekleştirilmiştir.

İşletmelerin işlemek zorunda oldukları içsel verilerin yanı sıra, dışsal veri kaynağı olarak sosyal medya ortamından yığınla veri akışı gerçekleşmektedir. Bu verinin yönetilmesi, faydalı bilgiye dönüştürülmesindeki güçlükler, makine öğrenme modellerinde kullanılan algoritmalara olan ihtiyacı artırmakta, yeni ortaya çıkan problemler algoritmaların geliştirilmesini sağlamaktadır. Algoritmaların performansları, işlenen veri setlerinin içerik ve özelliklerine bağlı olarak değişiklik gösterebilir ya da bir algoritma, başka bir algoritmanın zayıf bir yönünü geliştirdiği için daha iyi performans sergileyebilir.

Literatürde sınıflandırma ve regresyon modellerinde Naive Bayes (NB), Karar Ağaçları, Yapay Sinir Ağları, SVM algoritmalar ve Örnek Tabanlı ve İstatistiksel Dil Modeli Tabanlı Sınıflandırıcılar yaygın olarak kullanılmakla birlikte, bu alandaki diğer algoritmalar Şekil 2.9'da gösterilmiş, ileriki maddelerde çalışma kapsamında da kullanılan en yaygın algoritmalar özetle açıklanmıştır [88, 162-165]. Söz konusu algoritmalar, Şekil 2.9'da gösterilenlerle sınırlı değildir ve bu algoritmalara geliştirilen yenileri eklenmektedir. *Lojistik regresyon (LR)*, ikili sınıflandırmalar için uygun doğrusal bir modeldir [166]. LR, bağımlı değişkenin sınıflara ait olma olasılığı ile ilgilenir, bağımlı değişkenin alacağı değer ile ilgilenilmez. Lojistik modelde, bağımsız değişkenler ile iki ya da çok sınıflı kategorik bağımlı değişken arasındaki ilişki matematiksel doğrusal bir model ile tanımlanır [167]. LR modelinde bağımsız değişken katsayısı, bağımsız değişkende meydana gelen bir birim değişikliğin bağımlı değişkende meydana getireceği değişim miktarını ve yönünü gösterir.



Şekil 2.9. Sınıflandırma ve regresyon algoritmaları

Karar ağaçları, eğitim ve testinin hızlı olması, sonuçlarının daha kolay yorumlanabilmesi sebebiyle çok sık kullanılan sınıflandırma yöntemlerinden biridir. Karar ağaçlarında temel amaç, bir veri kümesini niteliklerine göre sorgulamalar yaptırarak en hızlı şekilde sonuca ulaşabilmektir [142]. Karar ağaçlarında sınıf etiketleri bilinen veri örneklerinden yola çıkılarak kural tabanlı bir ağaç yapısı oluşturulmaya çalışılır. Karar Ağaçları, Torbalama, RF, Aşırı Gradyan Güçlendirme (eXtreme Gradient Boosting-XGBoost) algoritmalarında karar ağaç yapıları kullanılmaktadır [168]. *Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees-CART)*, problemlerin çözümünde karar ağaçlarını kullanana istatiksel bir metottur. Ele alınan bağımlı değişken kategorik ise yöntem sınıflandırma ağaçları, sürekli ise regresyon ağaçları olarak adlandırılır [169]. CART, her aşamada ilgili kümeyi kendinden daha homojen olan iki alt kümeye ayırarak ikili karar ağaçları oluşturan bir yapıya sahiptir [170]. CART, bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi incelerken aynı zamanda bağımsız değişkenlerin birbirleri etkileşimlerini de ortaya koymaya çalışır [171]. CART'ın sahip olduğu algoritma, benzerlik gösteren değişkenlerin aynı ağaç düğümünde toplanmasına dayalıdır [172]. *Bayes ağı algoritmasında;* her bir değişken düğümlerle gösterilir, değişkenler arasındaki etkileşimi, şartlı olasılıklarla ve grafik olarak temsil ederler. Bu grafik, kullanıcı ara yüzü olarak

kullanılır ve modellenen problemin değişkenleri arasındaki etkileşimi görsel olarak yansıtır. Eldeki problemin doğası gereği ya da verilerin eksik olması nedeniyle net bir sonucun elde edilmediği durumlarda başarılı bir performans göstermektedir [142]. *NB algoritması ise*, Bayes teoremini temel alan, olasılık temelli bir algoritmadır [118]. Amaç belirli bir gözlemin her bir sınıfa ait olma olasılığını, koşullu olasılık temelli olarak hesaplar. Hesaplama ve sınıflandırma doğruluğu nedeniyle yüksek boyutlu metin verilerinin analizinde tercih edilen bir algoritmadır [173]. *KNN algoritması,* sınıflandırılması istenen bir veriyi, en yakın k komşuluktaki verileri temel alarak sınıflandıran bir yöntem olup, benzerlik fonksiyonunu kullanır ve birbirine yakın olan gözlemlerin birbirine benzeyen gözlemler olduğu esasına dayanır [142, 174]. Yöntemde, sınıfı belirlenmek istenen veri örneğinin, sınıfı belli olan diğer tüm veri örneklerine olan uzaklıkları Öklit formülü, Manhattan ve Minkowski yöntemlerinden birine göre hesaplanır. Uzaklıklar hesaplandıktan sonra en yakın k tanesi (k değeri genellikle 1, 3, 5 gibi tek sayılardan seçilir) seçilerek bu örneklerden çoğunluğu hangi sınıfa ait ise sınıfı belirlenmek istenen veri örneğinin de o sınıfa ait olduğu kabul edilir [118, 142, 174, 175]. *SVM*, sınıflandırma algoritmaları arasında popülerliği yüksek, istatistiksel öğrenmeye dayalı bir sınıflandırma algoritmasıdır, birçok alanda kullanılmaktadır [142]. SVM'ler, büyük verilerle çalışabilmeleri, az sayıda hiper parametreye sahip olmaları, uygulamadaki iyi sonuçları, yüksek güvenirliği, ezberleyerek öğrenmeye olan dayanıklılığı ve doğrusal olmayan sınıflandırmadaki başarı düzeyleri nedeniyle yaygın olarak kabul görmektedir [176]. Yüksek boyutlu özellik uzayını genelleştirme yeteneği ile SVM, özellik seçimi ihtiyacını da ortadan kaldırır [177]. *Yapay Sinir Ağlarının (Artificial Neural Networks-ANN)* en temel görevi, veri yapısından elde ettiği sonuçlarla genelleştirmeler yapmasıdır. ANN'ler genel olarak doğrusal ve doğrusal olmayan problemler için tek ve çok katmanlı algılayıcılar olarak kullanılmaktadır [178]. Biyolojik sinir ağlarına benzetilen ANN'ler veriden öğrenebilme, genelleme yapabilme, sınırsız sayıda değişkenle çalışabilme gibi özelliklere sahiptir [142, 179, 180]. *Çok Katmanlı Algılayıcı Sınıflandırıcı* (MLPC-Multilayer Perceptron Classifier), probleme ilişkin bilgilerin alındığı giriş katmanı, sınıflandırıcı ara katmanlar ve işlenen bilginin dışarıya iletildiği çıktı katmanından oluşan bir yapı üzerine kuruludur [180]. *RF* algoritması, birden çok karar ağacının ürettiği tahminlerin bir araya getirilerek değerlendirilmesine dayanır ve temel konsepti, ortak akıl şeklinde nitelendirilir. Karar ağaçlarının önemli problemlerinden biri, aşırı öğrenme durumu, diğeri algoritmada oluşan ağaç fazlalığı nedeniyle ortaya çıkan yavaşlamadır [181-183]. Karar ağacı tabanlı bir algoritma olan *GBM,* zayıf öğrenicilerin bir araya getirilip güçlü bir öğrenici ortaya çıkarmayı hedefler ve bu güçlendirme (boosting) yöntemi olarak

bilinmektedir. GBM, Adaboost algoritmasının sınıflandırma ve regresyon problemlerine uyarlanmış ve genelleştirilmiş halidir. İteratif bir yöntemle sınıflandırma yapan AdaBoost, her sınıflandırma sonunda hatalı sınıflandırılan gözlemler üzerindeki hataları gidererek çalışır. GBM; CART, RF, AdaBoost torbalama yöntemlerinden bazı özellikler bir araya getiren ve ağaca dayalı yöntemlerin çerçevesi olarak kabul edilen bir algoritmadır [184].

Karar ağacı tabanlı *XGB,* gradyan güçlendirme/artırma (Gradient Boosting-GB) algoritmasının çeşitli düzenlemeler ile optimize edilmiş yüksek performanslı halidir. Paralel işleme, ağaç budama, eksik değerleri işleme ve aşırı öğrenmeyi önleme tekniklerini kullanarak optimize edilmiştir, büyük veri setleri için tasarlanmıştır Hız ve performansı yüksek olan XGBoost, temel GBM çerçevesini kullanır [168, 185-188]. Bu sıralananlar dışında analizlerde, entropi kurallarını temel alan *ID3* [189, 190], ID3 algoritmasının yetersizliklerinin giderilmesi için geliştirilen *C4.5* [191, 192] ve C4.5'in geliştirilmiş bir şekli *C5.0* [193] gibi algoritmalar kullanılmaktadır.

Metin sınıflandırma ve regresyon problemlerinde; çok sayıda programlama dili, kütüphanesi ve bunlarla geliştirilmiş uygulama, analiz aracı olarak kullanılabilir. Söz konusu araçlar, *IBM Watson Doğal Dil Anlama* platformu [194]; *Google Bulut Doğal Dil Uygulama Arayüzü (Google Cloud Natural Language API)* [195]; *Microsoft Azure Metin Analitikleri (Microsoft Azure Text Analytics)* [196]; *Python NLTK (Natural Language Tool Kit)* [65, 197]; *Değerliğe duyarlı ve duygu muhakemesi (Valence Aware Dictionary and sEntiment, Reasoner-VADER) sözlüğü, Stanford Temel DDİ Süreci (Core Natural, Language Process-CoreNLP) kütüphanesi*, *TextBlob*, Gizli Dirihlet Ayrımı (Latent Dirichlet Allocation-LDA) kütüphanesi, *FastText*, *Keras ve PyTorch*, *Scikit-learn* NB, NumPy, Pandas, Matplotlib, Seaborn, Statsmodel, TensorFlow MaxEntClassifier, Transformers, spaCy gibi Python kütüphaneleri [198]; *RapidMiner* platformudur [199]; *Lexalytics* [200]; *Ağa Genel Bakış Keşif ve Araştırma Excel Eklentisi (Network Overview Discovery and Exploration add-in for Excel-NodeXL)* ve *Waikato Bilgi Analizi Ortamı (Waikato Environment for Knowledge Analysis-WEKA)* uygulamaları [201] şeklinde sıralanabilir ve bunlarla sınırlı değildir.

2.4.4. Modellerin performans ölçümü

Makine öğrenmesi algoritmaları, bir uygulama üzerinde sınandığı zaman hangi oranda başarı elde edildiğinin ölçülmesi gerekir. Bu maksatla yapılacak performans testleri için

duyarlık (precision), anma (recall), doğruluk (accuracy), hata oranı ve F1-skor gibi değerler ölçüm kriteri olarak kullanılır ve hesaplamalarda *karışıklık matrisi (hata matrisi)* esas alınır [202]. Bu çalışmada yapılan sınıflandırmalarda, algoritmaların performans ölçümleri için karışıklık matrisi[4] esas alınmıştır.

İkili bir sınıflandırma modelinin model sınıflandırması ile gerçek sınıf değerleri karşılaştırılması sonucunda ortaya çıkan dört farklı durum için karışıklık matrisi Çizelge 2.1 ve bu çizelgede kullanılan kısaltmaların (Türkçe ve İngilizce kısaltmalar) açıklamaları Çizelge 2.2'de verilmiştir.

Çizelge 2.1. Karışıklık matrisi

	Algoritma Tarafından Tahmin Edilen Sınıf		
Gerçek Sınıf		Sınıf 1 (Pozitif)	Sınıf 2 (Negatif)
	Sınıf 1 (Pozitif)	Doğru Pozitif (DP)	Yanlış Negatif (YN)
	Sınıf 2 (Negatif)	Yanlış Pozitif (YP)	Doğru Negatif (DN)

Çizelge 2.2. Karışıklık matrisinde kullanılan ifadelerin açıklamaları

Kısaltma	Anlamı	Açıklaması	Model Tarafından Yapılan Sınıflandırma İşleminin Doğruluğu
DP (TP)	Doğru Pozitif (True Positive)	Gerçekte *pozitif* olarak etiketlenmiş olan verinin, algoritma tarafından da *pozitif* etiketlenmesi durumu	Doğru
YN (FN)	Yanlış Negatif (False Negative)	Gerçekte *negatif* olarak etiketlenmiş olan verinin, algoritma tarafından *pozitif* olarak etiketlenmesi durumu.	Yanlış
YP (FP)	Yanlış Pozitif (False Positive)	Gerçekte *pozitif* olarak etiketlenmiş olan verinin, algoritma tarafından *negatif* olarak etiketlenmesi durumu.	Yanlış
DN (TN)	Doğru Negatif (True Negative)	Gerçekte *negatif* olarak etiketlenmiş olan bir verinin, algoritma tarafından da *negatif* olarak etiketlenmesi durumu.	Doğru

[4] Algoritmanın sınıflandırmasının gerçek sınıflandırmaya göre doğruluğunun gösterildiği bu çizelge için *karışıklık matrisi* ifadesi yerine *karmaşıklık matrisi* ifadesinin de kullanıldığı görülmektedir ve İngilizce karşılığı *comlexity matrix* şeklindedir.

Makine öğrenme algoritmalarının performans değerlendirmesi için Şekil 10'da verilen ölçütler kullanılır. Sınıflandırma başarısı genelde duyarlık ve anma değeri ile ölçülür ve her bir kategori için hesaplanır [74].

Duyarlılık değeri; algoritma ile yapılmış doğru sınıflandırmaların oranları şeklinde hesaplanır. Gerçekte ulaşılması gereken bilgilere ne oranda ulaşılmış olduğunu gösteren bir değerdir. *Anma değeri*, kategoriye ait elemanların ne kadarının tahmin edildiğini ölçümlemek için hesaplanmaktadır. *Doğruluk*, algoritma tarafından doğru sınıflandırılmış örnek sayısının veri setindeki toplam örnek sayısına oranıdır ve analizin gerçek değere ne kadar yakın olduğunu gösterir. *Hata oranı,* algoritma tarafından yanlış sınıflandırılmış örnek sayısının veri setindeki toplam örnek sayısına olan oranıdır. Hata oranı, doğruluk değerinin 1 değerinden çıkarılmasıyla da elde edilebilir. *F1 değeri*, duyarlık ve anma değerlerinin harmonik ortalamasının alınması ile elde edilen bir değerdir 0 ile 1 arasındaki bu değer, yapılmış olan testin doğruluğunu ifade eder. Veri madenciliği ve makine öğrenimi alanlarında yapılmış çalışmalar için en yaygın kullanılan başarı ölçütüdür [74, 87, 118, 137].

Belirtilen ölçütlerden doğruluk değeri, genellikle sınıf dağılımı dengeli olan veri setlerinin performansının ölçülmesi için kullanışlı bir ölçüttür. Dengesiz veri setlerinde daha sağlam bir ölçüm aracı olan F1 değeri ise duyarlılık ve anma değerleri arasında bir denge oluşturarak daha sağlam sonuç verebilir. Ancak F1 hesaplamasında da sınıflardaki gözlemlerin (örneklerin) oranı dikkate alınmamaktadır. Bu nedenle, F1 değerinin sınıfların dağılım oranlarını dikkate alan ağırlıklandırılmış ortalama değeri, dengesizlik bulunan veri setlerinde tercih edilen bir yöntemdir.

Elde Edilen Sonuçların Performans Ölçüm Araçları ve Formülleri	
Ölçüm Aracı	**Formül**
Duyarlık (Precision)	$Duyarlık\ (Pozitif\ için)\ Değeri = \frac{\mathrm{DP}}{\mathrm{DP}+YP}$ $Duyarlık\ (Negatif\ için)\ Değeri = \frac{\mathrm{DN}}{\mathrm{DN}+YN}$
Anma Değeri (Recall)	$Anma\ Değeri = \frac{\mathrm{DP}}{\mathrm{DP}+\mathrm{YN}}$
Doğruluk (Accuracy)	$Doğruluk = \frac{\mathrm{DP}+\mathrm{DN}}{\mathrm{DP}+\mathrm{YP}+\mathrm{DN}+\mathrm{YN}}$
Hata Oranı	$Hata\ Oranı = \frac{\mathrm{YP}+\mathrm{YN}}{\mathrm{DP}+\mathrm{YP}+\mathrm{DN}+\mathrm{YN}} = 1\text{-Doğruluk}$
F1 Ölçümü	$F1-Değeri = 2 * \frac{\mathrm{Duyarlık}*\mathrm{Anma}}{\mathrm{Duyarlık}+\mathrm{Anma}} = \frac{2\mathrm{DP}}{2\mathrm{DP}+\mathrm{YP}+\mathrm{YN}}$
Kısaltmaların Açıklamaları	
Kısaltma (İngilizce)	Anlamı (İngilizce)
DP (TP)	Doğru-Pozitif (True-Positive)
DN (TN)	Doğru-Negatif (True-Negative)
YP (FP)	Yanlış-Pozitif (False-Positive)
YN (FN)	Yanlış-Negatif (False-Negative)

Şekil 2.10. Performans ölçümü için kullanılan ölçütler

Çalışmada performans karşılaştırmaları için; veri setinde pozitif ve ürün sınıflarının hafif düzeyli dengesizliği (20% ve altında) de dikkate alınarak, sınıf oranlarını hesaplamalara dahil eden ağırlıklandırılmış F1 değeri kullanılmıştır.

$$\text{Ağırlıklıklandırılmış F1 değeri} = \frac{Sınıf\ içindeki\ örnek\ sayısı}{Veri\ setindeki\ toplam\ örnek\ sayısı} * \frac{2 * Duyarlılık * Anma}{Duyarlılık + Anma}$$

2.4.5. Metin sınıflandırma ve duygu analizi konusunda literatür

Twitter (X) kullanıcılarının paylaşımlarındaki metin içeriklerinin formatsız, düzensiz, anlamı belirsiz, imalı, jargon ve argo içerikli gibi farklı şekillerde olduğu görülmektedir. Bu tür formatsız metinlerde alana özel duygu terimlerinin kullanımı, kişilerin sahip oldukları duyguları ifade etme ve düşünme yöntemlerindeki farklılıklar, yazım hataları, örtülü anlam ve belirsizlikler, kısa uzunluktaki metinler, duygu analizi ve metin sınıflandırmayı zorlaştırmaktadır [203, 204]. Metin sınıflandırma ve duygu analizi için kullanılan; makine öğrenmesi, sözlük-tabanlı ve derlem-tabanlı, hibrit, derin öğrenmeye dayanan, dönüştürücü

yaklaşımların her biri, veri seti özelliklerine bağlı olarak farklı avantaj ve dezavantajlara sahiptir [205, 206]. Bu yaklaşımlar, birbirlerinin eksikliklerini tamamlayacak şekilde hibrit olarak kullanılabilir. Son dönem çalışmalarında hibrit yaklaşımların sınıflandırma model performanslarını artırdığına vurgu yapılmaktadır.

Metin ve duygu sınıflandırması esas olarak kelime odaklı bir çalışmadır ve kelimeye ilişkin özellikler incelenir. Kelimelerden elde edilen değerlerle bir metnin; doküman, cümle ve özellik-tabanlı (Aspect-Based Sentiment Analysis-ABSA) olarak üç seviyede sınıflandırması yapılabilir. Bu sınıflandırmalarda sıklıkla kullanılan bir yöntem duygu puanlarının toplulaştırılması yöntemidir. Bu yöntem kelimenin ağırlık puanlarını birleştirme [207], farklı sınıflandırıcıların sınıflandırma sonuçlarını bir araya getirme [208] gibi değişik şekillerde uygulanabilir. Mirtalaie vd. [209], belirlenen hedef bir özelliğin duygu değerini belirlerken farklı özelliklerin istenen özellikle ilişkisini dikkate alarak duygu polaritelerini toplulaştırır. Toplulaştırma yöntemlerini incelediği çalışmasında Basiri vd. [210] kelimelerin değerlerini tespit ettikten sonra bu değerlerin cümle ve genel seviyede toplulaştırmasını yapmıştır.

Kelimenin doğal kullanıma en yakın duygu değerinin tespit edilmesi konusunda; kelimenin cümle içindeki yeri, özel anlamı, örtük anlamlar, olumsuz cümlelerin çözümlenmesi gibi güçlüklere rağmen [77, 211, 212], metin sınıflandırmalarda farklı sektörlere yönelik yüksek performanslı modeller üretilebilmektedir [213-215]. Bunlar arasında, makine öğrenmesi ve sözlüğe dayalı sınıflandırmanın hibrit kullanımını esas alan çok sayıda çalışma yer almaktadır. Makine öğrenmesinin, hazır ya da veri setinin kaynak olarak kullanıldığı derlem tabanlı bir sözlükle birlikte kullanılması metin sınıflandırma performansını artırabilir [216, 217]. Öte yandan sözlük kullanımında şu hususların dikkate alınması gerekecektir: (i) Hazır bir sözlük ile sınıflandırma yapıldığında, sözlükte olmayan bir terim dikkate alınmayacaktır. (ii) Derlem tabanlı sözlük yaklaşımı ile sınıflandırma yapıldığında, koleksiyonda ayıklanamayan ilgisiz kelimeler nedeniyle çok boyutluluk problemi ortaya çıkabilecektir. (iii) Kelimelerin anlam ve kullanımında zaman içinde değişiklikler ve yeni kelimeler ortaya çıkabilmektedir, bu değişikliklerin sınıflandırma için kullanılacak sözlüklere yansıtılması gerekir [129, 217-219].

Metin sınıflandırma için gerekli sözlük ve eğitim verileri açısından yeterli kaynağa sahip olmayan dillerde doğru model kurulması, model performanslarının yükseltilmesi önemli bir sorun oluşturabilmektedir. Metin sınıflandırma ve duygu analizi konusundaki çalışmaların önemli bir bölümü başlangıçta İngilizce'de yapılmış, söz konusu çalışmalar müteakiben diğer Avrupa dillerine de uyarlanmıştır. Türkçe'nin sondan eklemeli bir dil olması ve cümle içindeki dizilişin Avrupa dillerinden farklı olması nedeni ile İngilizce'den uyarlanarak yapılan çalışmalarda düşük performanslar elde edildiği belirtilmektedir [137]. Duygu analizinde kullanılan yöntemlerin, Türkçe dilinin özelliklerini de dikkate alan, alana özel çalışmalarla güçlendirilmesinin, modellerin Türkçe veri setleri üzerindeki performanslarını artırabileceği düşünülmektedir.

Her dilin kendi özellikleri nedeniyle ortaya çıkan sorunların çözümü için, alana özel çalışmalar, diller arası gömme yöntemi gibi transfer öğrenme modelleri ile katkı sağlanabilmektedir [220]. Kılıçer vd. [221] yaptığı çalışmada Türkçe için yapılan duygu analizlerinde, Türkçe sınıflandırma sözlüklerinin çeviri sözlüklerden, hibrit yaklaşımların diğer yaklaşımlardan daha yüksek performans gösterdiğini; öte yandan hibrit çalışma sayısının düşük düzeyde kaldığını belirtmiştir.

Makine öğrenmesi için gözetimli, gözetimsiz, yarı-gözetimli, takviyeli (reinforcement), çok-görevli (multi-task), topluluk (ensemble), örnek-tabanlı (instance-based) öğrenme, sinir ağları gibi çok sayıda yöntem kullanılabilir [144, 222]. Makine öğrenmesinde algoritmanın seçimi, çözülecek problem türü, problemdeki değişken sayısı ve probleme uygun model türü gibi etkenlere bağlıdır. Kullanılan algoritmaların birbirlerine göre üstünlükleri değişkenlik göstermekte, yeni geliştirilen yöntem ve algoritmalarla önceki yöntemlerin zayıf yönleri güçlendirilmekte ve algoritmaların yeni sürümleri kullanıma girmektedir [88, 162-164, 223]. Literatürde önerilen model ve karşılaştırmalı analizlerle derin öğrenme dahil makine öğrenmesinin üstün performans sergilediği çok sayıda çalışma yer almaktadır [129, 206, 214, 224-230]. Ayrıca birden fazla sınıflandırıcının kombinasyonundan elde edilen performansın genellikle tek bir sınıflandırıcının kullanımından daha iyi deneysel sonuçlar verdiği belirtilmektedir [231, 232].

Hibrit modellerin uygun mimari ve hassas hiper parametre seçimi ile tüm modellerden daha iyi performans gösterebileceği belirtilmektedir [110, 218, 231, 130, 233-239]. Dey ve Das [240] modifiye edilmiş TF-IDF terim ağırlıklandırma yöntemini önerdikleri bir yaklaşıma

dayanan duygu analiz çalışmasında, farklı veri setleri üstünde 62.1%-89.2% aralığında performans elde etmişlerdir. Erşahin vd. [241], otel ve sinema konusundaki tweetlerden oluşan üç farklı veri seti üzerinde ve üç farklı sınıflandırma algoritması (SVM, NB ve J48) ile yaptıkları çalışmada önerdikleri modellerde 73%, 86.32% ve 91.96% oranında performanslar elde etmişlerdir. 68 analizin derlendiği bir çalışmada performansların en yüksek; sözlük tabanlı çalışmalarda 88.85%, makine öğrenmesi çalışmalarda 98.29% ve hibrit çalışmalarda 91.96% şeklinde gerçekleştiği tespit edilmiştir [221]. Bir başka çalışmada ise derin öğrenme, makine öğrenmesi ve hibrit yaklaşımların son dönem çalışmalarda elde edilen performanslarının 74%-91% aralığında değiştiği belirtilmektedir [242].

Tweetlerin kısa cümlelerden oluşması, bu cümlelerde jargon kelimelerin bulunması, tweetlerdeki az sayıdaki özellik içeren kelimelerin tespit edilebilmesini kritik bir işlem haline dönüştürür. Ayırt edici özelliği olan kelime sayısının seyrekliği, bu kelimelerin sınıflandırma için kullanılan skor değerlerinin düşük kalmasına neden olmaktadır. Müşteri yorumlarının kelime sayısının fazla olması, yorumun ürüne yönelik daha fazla bilgi içerdiği kabulü ile yorumun güvenirliliğini artıran bir unsur olarak kullanılmaktadır [243]. Bu açıdan Twitter (X) doğrudan bir ürün ya da hizmetle ilgili değerlendirme içeren yorumlardan farklılaşır. Tweet cümlelerinin söz konusu özellikleri, sınıflandırmalar açısından terim ağırlıklandırma ve öznitelik seçimini daha önemli hale getirmektedir. Alexandrovna vd. [244] çalışmasında özellik ağırlıklandırma yöntemlerine dikkat çekerek, verimli özellik ağırlıklandırma yapılarak performansın artırılabileceğini ifade etmiş ve özellik boyutunu azaltan bir metodoloji ile doğruluk değerinde 4%-5% oranında bir iyileşme elde etmiştir. Metni en iyi temsil edecek terimlerin tespiti için, Bandhakavi [245] alana özel sözlük bilgisi ile unigram karışım modelini (Unigram Mixture Model-UMM); Sarayna [246] tweetlerden oluşan veri setinde TF-IDF yöntemini, Kaur [247] ise TF-IDF yöntemini n-gram ile birlikte kullanmıştır. Alshehri and Algarni [248], TF ve seçilen özellikleri ayırt edicilik derecelerine göre gruplandıran ve her grubun katkısına bağlı olarak ağırlıklandıran terim ayırt ediciliğini (Term Discrimination Ability-TDA); Sharma ve Kumar [249] terimin istatistiksel, anlamsal ve bilimsel adlandırılmış varlık özelliklerinin kullanıldığı çoklu özellik tabanlı kavram sıralama algoritmasını kullanmıştır.

Metin sınıflandırma ve duygu analizi konusunda Türkçe literatürde yer alan çalışmalar ileriki maddelerde sıralanmıştır.

Eroğul [250] çalışmasında film yorumlarını n-gram model, SVM algoritması ile olumlu ve olumsuz kategorilerde ikili sınıflandırmaya tabi tutmuştur. Makine öğrenmesini kullanarak Kaya vd. [251] Türkçe gazete köşe yazıları üzerinde duygu analizi gerçekleştirmiş, Boynukalın [252] yaptığı sınıflandırma çalışmasında %80 civarı başarı elde etmiş, Akbaş [253] Twitter (X) üzerinden topladığı verileri üzerinde duygu analizi yapmış, duygusal kelime listesi kullanmış ve duygu sınıflandırmasında %85 civarı başarı elde etmiştir. Tokcaer [254] çalışmasında, Türkçe duygu analizi üzerine yapılan güncel çalışmaları ve bu çalışmalarda kullanılan yöntemleri, duygu analizi için kullanılan uygulama açık kaynaklı kütüphaneler ve veri tabanlarını, incelediği 55 Türkçe kaynak üzerinden derlemiştir.

Aytekin [255] çalışmasında, Türkçe blog sayfalarındaki müşteri yorumlarını pozitif ve negatif olarak sınıflandıran yarı denetimli bir duygu analizi modeli önermiş ve çalışmada çeviri ile elde edilen duygu sözlüğünü kullanmıştır. Vural vd. [115]'nin çalışmasında SentiStrength [256] altyapısı ile duygu analizi yapılmış, çeviri ile elde edilmiş sözlük kullanılarak sınıflandırma yapılmıştır.

Özsert ve Özgür [257] yaptıkları çalışmada duygu analizinde kullanılan kelime polaritelerini belirlemek için yarı otomatik yöntem geliştirmiş, WordNet kütüphanesinden faydalanarak duygu sözlüğü oluşturmuşlardır. Türkmenoğlu ve Tantuğ [116] çalışmalarında, sözlük temelli yaklaşım kullanmışlar, SentiStrength [256] sözlüğünden istifade ederek ve çeviri yöntemi kullanarak duygu sözlüğü elde etmişlerdir. Uçan [74], otomatik çeviri yöntemini kullanmış, kaynak olarak SentiWordNet'ten istifade etmiş, sözlüğün performansını otel ve film yorumları üzerinde ölçmüş ve sözlüğün performansını %80 olarak tespit etmiştir.

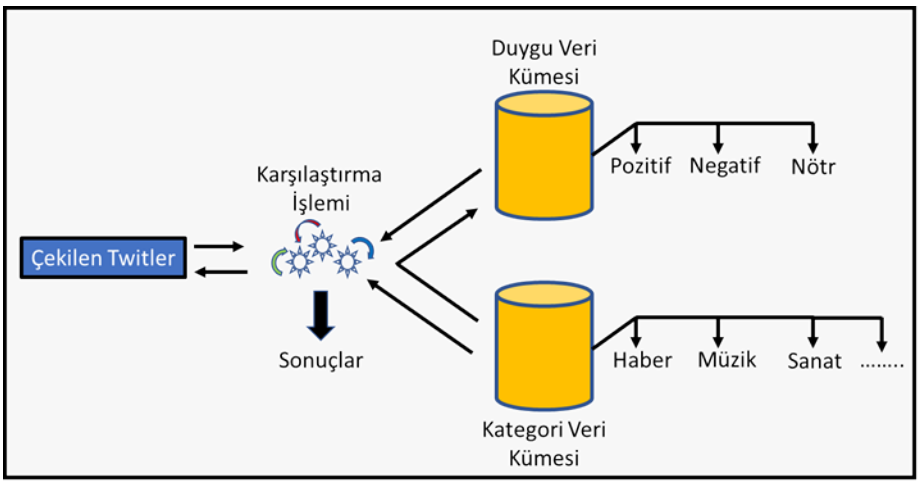
Meral ve Diri [7]'nin çalışmalarında, toplanan tweetlerin manuel olarak olumlu, olumsuz ve nötr şeklinde etiketlemesi yapılmış, etiketleme işlemi birden fazla kişi tarafından yapılarak objektifliği sağlanmıştır. Çalışmada DDİ modülü olarak, n-gram ve kelime tabanlı iki farklı yöntem kullanılmış, jargon sözlük oluşturulmuş ve sözlükteki her bir ifade sınıflandırıcı için bir özellik olarak kabul edilmiştir. Bu çalışmada 8.500 civarı tweet elle etiketlenmiş, kelime yakalama bakış açısı ve makine öğrenmesi yöntemlerinden Rastgele Orman, NB ve SVM sınıflandırıcıları kullanılmıştır. Çalışmada, Twitter (X) verileri kelime tabanlı bir DDİ sürecinden geçirilerek sınıflandırıcılar eğitilmiş, korelasyon tabanlı öznitelik seçim yöntemi kullanılmış ve sınıflandırma performansında %90'a yakın başarı elde edilmiştir.

Tutar vd. [66], sezgi analiz çerçevesi oluşturdukları çalışmalarında, Sentiwordnet kütüphanesinden istifade ederek Türkçe sezgi sözlüğü hazırlamışlar, sözlükteki kelimelerin anlamlarını üçüncü parti bir İngilizce-Türkçe sözlükten, sezgi (duygu) değerlerini ise SentiWordNet'ten almışlardır. Dehkharghani [258] tarafından Türkçe genel amaçlı duygu sözlüğü oluşturulması için SentiTurkNet çalışması yapılmıştır.

Akgül vd. [9] tarafından yapılan Twitter (X) üzerindeki duygu analizi çalışmasında, iki kişi (objektifliği sağlamak maksadıyla) tarafından tweetlerin manuel olarak pozitif, negatif ve nötr şeklinde etiketlemesi yapılmış ve etiketleme sonuçları otomatik etiketleme sonuçları ile karşılaştırılmıştır. Çalışmada, sözlük ve n-gram olmak üzere iki farklı model kullanılmıştır. Karakter tabanlı n-gram (2, 3 ve 4-gram uygulamaları) modeli uygulamalarından en başarılı uygulama sonucuna 3-gram model ile ulaşılmış olmakla birlikte, çalışmada kullanılan sözlük tabanlı yöntemlerin daha başarılı sonuçlar verdiği belirtilmiştir. Çalışmada ulaşılan bir başka sonuç, sistemin eğitilmesinde kullanılan veri setindeki veri sayısı arttıkça duygu analizindeki tweetleri doğru etiketleme konusundaki başarı yüzdesinin arttığıdır.

Baykara ve Göktürk [259] tarafından yapılan çalışmada tweet analizi için iki veri kümesi ile işlem yapılmıştır. Birincisi, tweetlerin duygu analizinin yapılabilmesi için oluşturulan veri kümesi, ikincisi tweetlerin kategorize edilmesinde kullanılmak üzere oluşturulan veri kümesidir. Birinci gruptaki veri kümesi, tweetlerin duygu açısından pozitif, negatif, nötr olarak ayrıma tabii tutulması için kullanılmış, ikinci gruptaki veri kümesi ise tweetlerin kategorize (haber, spor, magazin, sanat gibi) edilmesi için kullanılmıştır. Bu çalışmada, duygu analizi ve kategorik sınıflandırma için oluşturulan model Şekil 2.11'de gösterilmiştir.

Çetin ve Eryiğit [97] çalışmalarında, "pozitif", "negatif" ve "nötr" olmak üzere üçlü sınıflandırma ile duygu analizi yapmışlardır. Çalışmada, duygu analizi için Lexicon-NRC Emotion duygu sözlüğünün Türkçe bölümünü kullanmışlardır. Bu çalışmada lineer sınıflandırma tekniği, sınıflandırıcı olarak Liblinear kütüphanesinden Logistic Regression algoritması kullanılmıştır. Çalışmadaki sınıflandırıcı için metinsel özellikler, duygu tanımlama grubundaki hedef kategori ve hedef terimler belirlenmiştir.



Şekil 2.11. Duygu analizi ve kategorik sınıflandırma [259]

Bilgiç ve Koçak [114] tarafından yapılan çalışmada, Atatürk ve Sabiha Gökçen havalimanları hakkında İngilizce olarak yapılan müşteri yorumları üzerinde görüş ve duygu madenciliği analizi yapılmış ve müşterilerin şikayetçi oldukları konular tespit edilerek işletmeciler için yol gösterici, bu alanda verilen hizmetlerin iyileştirilmesine katkı sağlayacak çıkarımlar yapılmıştır. Çalışmada hazır sözlükler kullanılarak duygu analizi yapılmıştır.

Sağlam [86] tarafından yapılan çalışmada dokümanların yalnızca tek bir duygu ile etiketlenmemesi gerektiği ve dokuman içinde farklı duygu dalgalanmalarının olabileceği eleştirisi ile *Duygu Haritası Modeli* geliştirilmiştir. Bu kapsamda cümle, paragraf ve metinlerde yapılacak duygu analizlerinde, sadece bir yönde (pozitif, negatif ve nötr şeklinde) yapılacak bir etiketlemenin gerçekçi olmayabileceği belirtilmiştir. Twitter (X) analizlerinde veri seti büyüklükleri ve kapsanan zaman dilimleri çalışma alanlarına göre farklılık gösterebilmektedir. Örneğin kamuoyunu kısa süreli olarak etkileyen konularda, bu etkinin ortaya çıktığı birkaç gün, bir hafta gibi kısa bir zaman dilimi için veri toplanırken; bir ürün ve hizmet konusunda müşterinin görüşlerinin elde edilmesine yönelik çalışmalarda birkaç haftadan başlayıp daha uzun döneme uzanan zaman dilimi için veri toplama işlemi gerçekleştirilebilmektedir. Bu konuda çalışmanın konusu, amacı, ürün, hizmet çeşidine göre karar verilebilir.

2.4.6. Metin tabanlı sosyal medya uygulaması Twitter (X) ve yiyecek sektöründeki Twitter (X) metinlerinin sınıflandırmasına yönelik literatür

Kullanıcıların, sosyal medya platformlarda her konuda bilgi paylaşmaları, bu platformları tüketici görüşlerine ulaşılabilen kaynaklar haline getirmektedir. Bu platformlar sayesinde edilgen durumdan etken hale geçen tüketiciler, işletmelerin pazarlama stratejilerinde daha merkezi bir konuma geçmektedirler. Söz konusu platformları ürün ve hizmet pazarlama açısından etkili bir araç olarak kullanabilen işletmeler ise rekabet üstünlüğü elde edebilmektedir. Internet üzerinde görüş ve değerlendirmelerin paylaşıldığı Facebook, Twitter (X), Instagram gibi sosyal paylaşım uygulamalarından elde edilen ham veriler analizlerle, faydalı bilgilere dönüştürülebilmektedir. Genel olarak sosyal medya uygulamalarında analitik yöntemlerin kullanılması, literatürde *sosyal medya analizi (SMA), SAA*, ya da analitik uygulama hangi platform için yapıldı ise o platform ismi ile (*Facebook analizi, TA* gibi) isimlendirilmektedir. Analiz yöntemleri, söz konusu sosyal medya platformlarının uygulama özellikleri de dikkate alınarak, platformlara özel modeller geliştirilir.

SMA, "iletişim ve etkileşimi kolaylaştıran sosyal medya verisini; toplama, izleme, analiz etme, özetleme, görselleştirmeyi gerçekleştiren ve kullanışlı örüntüleri ve istihbaratı ortaya çıkarmak maksadıyla kullanılan bilişim araç ve çerçevelerini geliştirme ve değerlendirme" şeklinde tanımlanmaktadır [260, 261]. Karabulut ve Küçüksille [65]'nin çalışmasında ise SMA için "sosyal ağlarda paylaşılan bir içeriğin eriştiği kişi sayısı, aldığı etkileşim sayısı, yorum sayısı gibi veriler ile yapılan çalışmalar" şeklinde tanımlama yapılmaktadır. Twitter (X) karşılıklı etkileşime imkan vermesi, takipleşme ve paylaşılan mesajların bulaşıcı bir etki ile dağılımını sağlayan retweet özelliği ile öne çıkmaktadır [262]. Mesajların bu şekilde bulaşıcı bir etki ile dağılımı *söylenti teorisi* olarak isimlendirilmektedir [263]. Yapılan bir çalışmada ise Twitter (X)'da mesajların kullanıcılar arasında yayılma özelliğinin ağızdan ağıza pazarlamaya (Word of Mouth-WOM) etkileri incelenmiştir [264].

Twitter (X), diğer kullanıcı verilerine ulaşabilmeye imkan sağlaması nedeniyle, işletmeciler ve araştırmacılar tarafından kullanılabilmektedir [3]. Twitter (X) API'si kullanılarak Twitter (X)'dan veri çekilebilmektedir[5]. Bu arayüz yazılımları açık veri olarak adlandırılan ve

[5] Twitter (X) geliştiricilere tweet çekme işlemi için Twitter (X) API ile sağlanan ücretsiz giriş imkanı, Twitter (X) uygulama politikalarında yapılan değişiklikle 9 Şubat 2023 tarihinden itibaren kaldırılmış, Twitter (X) ile geliştiricilerin birlikte çalışmalarının devamı için ücretli API uygulamasına geçilmiştir. 9 Şubat 2023

üçüncü taraflarla paylaşılan xls (Microsoft Excel spreadsheet), csv (comma-seperated values) gibi veri dosyalarını kullanarak işlem yaparlar; txt (yazı-text) metin şeklinde kaydedilmiş veri yapısını, csv virgülle ayrılmış veri yapısını gösterir [265]. API aracılığı ile veri toplama işlemi, temel olarak anahtar kelime veya etiket (hashtag) kullanılarak yapılır. API arayüzünün; (i) araştırma API'si (search API) ve akış API'si (stream API) olmak üzere iki bileşeni vardır. Araştırma API'si ile seçilen kritere göre geçmişe ait veri toplanabilirken, akış API'si ile spesifik bir yer veya kullanıcılarla ilgili ya da özel anahtar kelimelerle ilişkili sürekli tweet akışı sağlanabilir [1]. API metodu, kamuya açık olan Twitter (X) verisinin ancak %1'inin kullanılmasını sağlar. Ayrıca GNIP, DataShift gibi Twitter (X) firehose olarak adlandırılan veri sağlayıcılar aracılığıyla, kritere dayalı Twitter (X) verisinin %100'ü de temin edilebilir. İdeal olmakla birlikte bu yöntem oldukça pahalı bir seçenektir [1]. Bu konuya ilişkin Morstatter vd. [266] yaptıkları çalışmada Twitter (X) Akış API'sinden elde edilen veri ile Firehose'dan elde edilen verilerin örneklem temsil yeteneği yönüyle uyumlu olduğunu göstermişlerdir.

Küresel bir sosyal ağ olan Twitter (X), başlangıçta kullanıcıları için mesajların uzunluğunu 140 karakter ile sınırlandırmışken bu sınırlama sonradan 280 karaktere çıkarılmıştır. Twitter (X)'de tweetler, cevaplar ve retweetler olmak üzere üç tip tweet bulunmaktadır [267]. İstatistiksel verilere göre Twitter (X) aylık 326 milyon aktif kullanıcı sayısı ile günde 500 milyon adet tweet kaydetmektedir, mobil Twitter (X) kullanıcıları toplam kullanıcıların % 80'ini meydana getirir, günlük aktif kullanıcı sayısı 139 milyondur, Twitter (X) kullanıcılarının % 34'ü kadındır. Twitter (X), en çok ziyaret edilen sosyal ağlar arasında ikinci ve kullanıcı sayısı açısından sosyal platformlar arasında 12'nci[6] sıradadır. Twitter (X), bu özellikleri ile birçok alanda toplum üzerinde birçok araştırmanın yapılmasını kolaylaştırmakta, tüketici davranışlarına yönelik gerçek örüntülerin, tüketiciler arasındaki ağların ortaya çıkarılmasını sağlamakta ve sosyolojik olaylar ve toplumsal duyguların izlenmesine imkan sağlamaktadır [268-272]. Bu çalışmada incelenen yiyecek hizmet sektöründe yapılan Twitter (X) analitiği ile ilgili literatürde yer alan çalışmalar, Çizelge 2.3 ve Çizelge 2.4'te özet olarak verilmiş ve ileriki maddelerde detaylandırılmıştır. Yiyecek sektöründeki duygu analiz çalışmaları az sayıda olmakla birlikte performansı tatmin edici

tarihinden itibaren Twitonomy uygulaması ile veri çekme işlemi sonlandırılmıştır. Twitonomy yalnızca kullanıcının kendi profiline ilişkin istatistiki verileri paylaşmaya devam etmektedir.

[6] Sıralama; Facebook, Youtube, What's up, MS Messenger, WBXN/WECHAT, Instagram-1 milyar, QQ, Qzone, Douyn/Tiktok, Sina Weibo, Reddit, Twitter şeklindedir.

düzeyde model ve yöntemler geliştirildiği görülmektedir. Ayrıca duygu analizi konusunda özellikle alana özel çalışmalara ihtiyacın devam ettiği belirtilmektedir [213-215].

Çizelge 2.3. Yiyecek sektörüne ilişkin literatür özeti-1

Yiyecek Sektörüne İlişkin Literatür Özeti-1					
Yazar	Amaç	Veri	Araç / Yöntem	Sonuç / Katkı	Müteakip Çalışma
Hingle vd. [18]	Yemek yeme alışkanlıklarını tespit etmek (zaman, yer, kiminle)	50 kişilik katılımcı grup, -Tarih/Zaman: Anlık veri akışı ile iki gün hafta içi ve bir gün hafta sonu 773 tweet, -Günü dört dilime ayırmış.	-VIBE (Anlık veri yakalama) -Standart haritalama (Gmap) -Graphviz	-Yiyecek tüketimi ve davranışsal faktörlere ilişkin veri toplamaya ve aralarındaki ilişkiyi tanımlamaya uygun metot geliştirmiştir	Fenotip, genotip kalıtım ve Sosyo-demografik veriler eklenerek ilave çalışma yapılabilir.
Chae [3]	Twitter (X)'ın TZY'nde kullanım durumunu tespit etmek	-22 399 tweet, -Tarih: 5 Şubat-10 Nisan 2013, -Anahtar kelime: SC, SCM	-Tanımlayıcı, Metin, Ağ analizi, -Script, -Wordnet sözlük, -Twitonomy, -n-gram, -Makine öğrenmesi	Twitter(X)'ın TZ'de kullanımı, TZ pratikleri için Twitter(X)'ın potansiyeli ve rolü ortaya konulmuş, Twitter (X) analizi için çerçeve önerilmiştir.	-Endüstriyel uygulamaları tasarlayabilmeleri için Twitter (X) analizine yömnelik uygulama rehberi geliştirme, Sosyal medyaya yapılan yatırımların, TZY'ne etkileri
Mishra ve Singh [1]	Tedarik zinciri süreçlerinde sığır eti atıklarının minimize edilmesini sağlamak ve önleyici tedbirler geliştirmek	-Veri seti 5 kategoriye ayrılmıştır. -Tarih:Ocak 2015-Ocak 2016,	-Tanımlayıcı analiz, -Duygu, içerik, atık kategori analizi, -Streaming API, Wordnet sözlük	-Sığır etinde ortaya çıkan azaltılması gerektiği ve en fazla atığın tüketici seviyesinde ortaya çıktığı, müşteri memnuniyetsizliğinin perakendecinin saygınlığını düşürdüğü tespit edilmiştir.	-Uygulama daha uzun zaman dilimini kapsayacak şekilde genişletilebilir. -Koyun eti ve diğer sektörlerdeki tedarik zincirlerinde uygulanabilir.
Singh vd. [20]	Sığır et ürünlerine yönelik memnuniyetsizlikleri ve çözüm yollarını tespit etmek, müşteri odaklı tedarik zinciri geliştirmek	-26 269 -Tarih:23.03-13.04.2014	-İçerik analizi, hiyerarşik kümeleme, birliktelik kuralları, kümeleme, -Streaming API, -Duygu analizi, NLP, -SVM,	Sığır et ürünlerindeki memnuniyetsizliklerin, tedarik zincirinde hangi süreçleri ilgilendirdiği tespit edilmiştir.	Önerilen analiz yönteminin gıda sektöründe diğer ürünlerin, tedarik zinciri süreçlerine yönelik uygulamalar yapılabilir.

Çizelge 2.4. Yiyecek sektörüne ilişkin literatür özeti-2

Yiyecek Sektörüne İlişkin Literatür Özeti-2					
Yazar	Amaç	Veri	Araç / Yöntem	Sonuç / Katkı	Müteakip Çalışma
El-Khchine [21]	Müşteri odaklı tedarik zinciri gelişimine katkı sağlamak, tavuk ürünlerindeki ana ilgi alanlarını tespit etmek	-13 639 tweet eğitim seti ve 11 032 tweet test seti (%80 benzer tweetler, veri setinden ayıklanmıştır) -10 günlük zaman aralığı,	-Search API, Python script, -KNN, LR, SVM, -Duygu (sözlük) analizi	Tavuk ürünlerine ilişkin ana ilgi/endişe alanları ve kalite konusundaki problem sahaları tanımlanarak, tüketici merkezli tedarik zincirinin geliştirilmesine yönelik öneriler geliştirilmiştir.	Anahtar kelime miktarı, veri çekme tarih aralığı ve tweet sınıflandırması üzerine zaman serisi analizi uygulanarak ileri seviye analizler yapılabilir, algoritmalar değiştirilerek uygulanabilir.
He vd. [261]	6 dizüstü bilgisayar markası hakkında en çok konuşulan konuları, olumlu, ve olumsuz konulşulan konuları ve bilgisayarların en çok konuşulan özelliklerini tespit etmek	-385 614 tweet, -Tarih:01.11-31.12.2014, -Anahtar kelime: Dell, Lenova, Toshiba, Acer, Asus, HP	-İstatistiki, kullanıcı ve duygu analizi, -VOZIP, -VOZIQ: Apache Sqlr, Hadoop, MySQL, JavaServer Pages, Amazon AWS, -SVM, NB, -Duygu analizi (Lexalytics, Wordnet, SentiStrengt, Social mention, Trackur, Sysomos,Viralheat)	Markaların zayıf ve güçlü tarafları, geliştirilmesi gereken özellikler, kullanıcı ihtiyaç ve beklentileri ortaya çıkarılmıştır.	Aynı çalışmanın, İngilizce dışındaki dil ve Twitter (X) dışındaki platformlarda yapılarak karşılaştırmalar yapılabilir.

Adak vd. [273] hazır yiyecek sektörüne yönelik duygu analiz çalışmalarını incelediği literatür çalışmasında, bu alanda çoğunlukla sözlük tabanlı ve makine öğrenmesi metotlarının kullanıldığını, derin öğrenme tekniklerini uygulayan sınırlı sayıda çalışma olduğunu ve derin öğrenme tekniklerinin diğer sektörlerde daha iyi performans gösterdiğini, bu sektördeki modellerin 77%'sinin doğası gereği yorumlanmasında güçlüklerle karşılaşıldığını belirtmiştir. Ahmed vd. [274] sınıflandırma başarısını artırmak için örtük anlamları da dikkate alan bir model geliştirmişlerdir. Alan özel anlamı ve hedef özellikleri birlikte kullanan söz konusu modelin restoran ile ilgili veri seti üzerinde kullanılmasından 89% başarı elde edilmiştir. Aktaş vd. [275] yiyecek ve gıda sektörüne yönelik makine öğrenmesi ile yaptığı çalışmada 86% performans elde etmiştir.

Hingle vd. [18], genç yetişkinlerin beslenme alışkanlıkları, yemek yeme nedenleri ve Twitter (X)'ın bu alanda kullanımına yönelik bir çalışma yürütmüşlerdir. Çalışma, 50 kişilik grup tarafından kullanılan Twitter (X) uygulaması üzerinden yeme alışkanlıklarına yönelik hususların raporlanması ve bunların analiz edilmesi şeklinde yürütülmüş ve Gmaps haritalama yazılımı ile görselleştirme için kullanılmıştır. Görselleştirmede bir gün dört zaman dilimine ayrılarak tweet yoğunlukları açısından karşılaştırmalar yapılmış, yiyecek çeşidi, zaman dilimi, konum ile bağlantılı davranışlar arasında ilişkiler harita üzerine aktarılmıştır. Çalışmada yapılan analiz ile Twitter (X)'ın beslenme alışkanlıklarının tespit edilebilmesi, bu alışkanlıkların yer ve zaman dilimi bilgisi kullanılarak çeşitli karşılaştırmalar yapılabilmesi için uygun bir araç olduğu tespit edilmiştir.

Koçak vd. [10]; API sunucularından java tabanlı program ile veri çekilmesi, imla ve yazım kurallarına göre metnin düzenlenmesi, argo ifadelerden temizlenmesi, gönderilerdeki olumlu, olumsuz ve nötr duyguların 0,1,2 olarak ARFF (Attribute-Relation File Format / Öznitelik-İlişki Dosya Biçimi) dosyasına işlenmesi, işlenen metin ve duygu polaritelerinin WEKA programında sınıflandırma algoritmaları (SMO) ile analiz edilmesi şeklinde dört aşamalı bir model oluşturmuşlardır. Girgin ve Şahin [276], önişleme, polarite hesaplama modülü, olumsuzluk işleme modülü, özellik üretme modülü ve sınıflandırma modülünden oluşan beş aşamalı bir sınıflandırma modeli geliştirilmiştir.

Tweetlerin analiz edilerek maksada uygun bilgilerin elde edilmesine yönelik Chae [3] tarafından TA modeli geliştirilmiştir. Chae [3]'nin TA'sı; tanımlayıcı analitik, içerik analitiği, ağ analitiği olmak üzere üç farklı analitik araştırma tekniğini birleştirir. Twitter (X) verisi, tweetler ve üst veri (meta data) ile çok miktarda bilgiyi üstünde taşır. *Tanımlayıcı analitik*; bu bilgilerdeki tweet sayısı, farklı tweet tiplerinin dağılımı, etiket sayıları gibi tanımlayıcı istatistiklere odaklanır. Etiket analizi ve anahtar kelime analizi, Twitter (X)'dan çalışma yapılan konuyla ilgili tweetlerin çıkarılmasını sağlar. Tanımlayıcı analitik, tedarik zinciri yönetimi araştırmalarında yaygın bir şekilde kullanılmaktadır. *İçerik analizi,* Twitler'deki yapılandırılmamış metinlerin temizliği ve işlenmesi süreçlerinden geçirilmesinden sonra, bilgi çıkarılması maksadıyla, kelime/terimlerin frekanslarının çıkarılması, bunların analizi ve kümelenme işlemlerinin yapılmasını kapsar. Chae [3], TA çerçevesinde yaptığı içerik analizinde; kelime analizi (terim frekans analizi, kümelenmiş doküman seviyesi analiz), etiket analizi (başlık etiketi frekans analizi, farklı başlık etiketlerinin birliktelik/ilişkisellik analizi) ve duygu analizi yapılmıştır. Duygu analizinde;

tüm tweetlerin sübjektifliği ve kümelenmiş/kategorilendirilmiş[7] tweetlerin objektifliği şeklinde iki seviyeli olarak incelenmiştir. *Ağ analizi* kapsamında toplojik analiz, merkezilik analizi (düğüm seviyesi ölçütü), topluluk analizi (ağ seviyesi ölçütü) ile veri setine ilişkin ağ özellikleri ortaya çıkarılabilir [3]. Birçok akademik çalışmada kullanılan ağ analizi, ağ içinde hangi aktörün diğerleri ile ne seviyede ilişkisi olduğunu ölçerek ağ yapısı ve aktörler hakkında bilgi verir, bu bilgiler genellikle aktörlerin ağ içindeki merkeziliklerinin ölçülmesi ile elde edilir [277, 278].

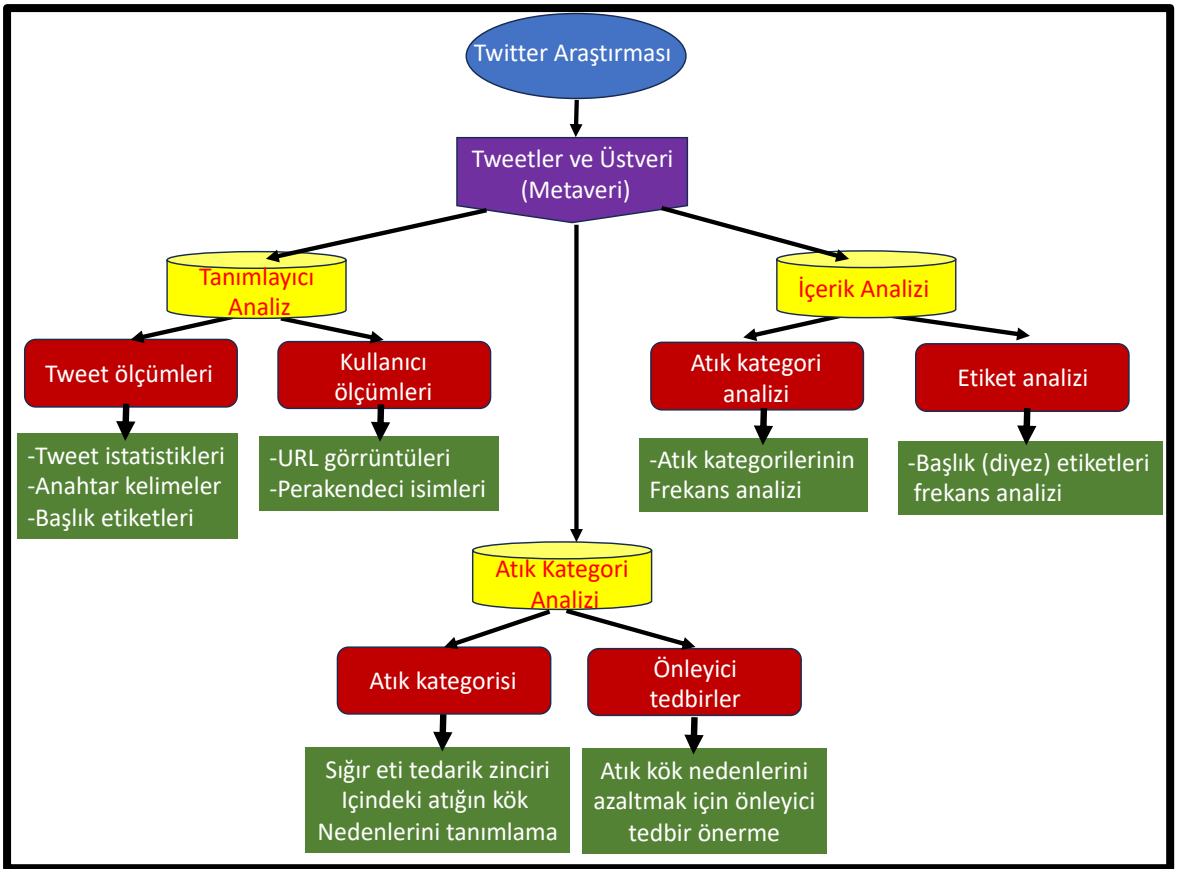
Bruns ve Stieglitz [64]; Twitter (X) sohbetlerini konu alan diğer çalışmalarda da kullanılabilecek tweetlerin ölçümlerine yönelik standart bir katalog hazırlamışlardır. Bu kapsamda bireysel katılımcıların görünürlülükleri ve aktiviteleri, kullanıcı grupları, geçici sohbet akışları, daha fazla ya da daha az aktif olanlar, sohbetlerin zamanla bağlantıları gibi durumların ortaya konulmasını sağlayan ölçüm standartlarını belirlemişlerdir. Kullanıcı ölçümü için esas olarak *kullanıcı aktiviteleri* ve bu *kullanıcıların görünürlülükleri* olmak üzere iki ölçüm aracı kullanılır. Aktivite ölçümü basit olarak herbir kullanıcı tarafından gönderilen tweet sayısıdır. Genelde pek az sayıda kullanıcı yüksek seviyede aktivite gösterirken, diğerleri sadece retweetleyerek sohbet içinde yer alabilir. Kullanıcı aktivitelerinin ölçümü için; her bir kullanıcının gönderdiği tweet sayısı, gönderilen orijinal (diğer bir kullanıcıdan bahsetmeyen) tweetler, başka bir kullanıcıdan bahseden yani @ sembolü kullanılan tweetler incelenir. Kullanıcı ölçümleri, aktif kullanıcıları, faaliyetler açısından kullanıcı gruplarını ve kullanıcıya bağlı diğer bilgileri ortaya çıkarır [64].

Diğer bireylerin görüşlerini araştırma ve bu görüşlere güvenme eğiliminin yüksek olduğu [279] ağırlama endüstrisinde, lokanta algılarının ölçülmesi maksadıyla Park vd. [19] TA'ni kullanarak Çin, Japon, Kore ve Tayland restorantlarına ilişkin lokanta algılarını ortaya çıkarmayı amaçlamışlardır. Çalışmada konuya ilişkin anlamlı örüntüleri, popüler kelimeleri ve duygusal cümleleri bulmak için metin madenciliği, duygu analizi ve en etkin Twitter (X) kullanıcılarını tespit etmek için takipçi ilişkilerini kullanarak ağ analizi yapılmıştır. Sonuçta Çin restorantları hakkında ortalama duygu miktarının diğer Asya mutfaklarına göre daha düşük olduğu, olumlu tweetlerin yiyecek kalitesi hakkında, olumsuz tweetlerin hizmet kalitesi ve yiyecek kültürü hakkında olduğu tespit edilmiştir. Park vd. [19]'nın çalışmasında, veri toplama ve arşivleme için, ScraperWiki çevrimiçi web sayfası [280], metin madenciliği

[7] Çalışmada CSR, risk, lojistik, üretim, IT olarak beş kategori kullanılmıştır.

ve kelime frekans analizi için RapidMiner 5.3, SentiStrength 2.2 ve veri tarama, filtreleme ve sınıflandırma gibi diğer analitik süreçler Microsoft Excel 2013'ü kullanmıştır.

Mishra ve Singh [1] Twitter (X)'dan elde edilen görüşleri kullanarak tanımlayıcı analiz, içerik analizi ve atık kategori analizi yapmış, sığır eti atığının tedarik zinciri süreçlerindeki atıkların minimize edilmesine yönelik bir analiz modeli (Twitter (X) analitik çerçevesi) ile bu konuda önleyici tedbirleri geliştirmiştir. Önerilen model, Şekil 2.12'de gösterilmiştir. Mishra ve Singh [1]'in çalışması sonucunda, sığır eti ürünlerinin atık nedenleri olarak; etin raf ömrünün dolmasından önce et rengindeki değişim, etin tüketimi esnasında hissedilen etin sertliği, et ürününde fazla yağ ve kıkırdak bulunması, kötü lezzet, koku ve çürüme ve ürüne yabancı madde karışımı ve hasarlı paketleme tespit edilmiştir. Ayrıca yapılan analiz ile atığın oluşmasındaki temel nedenlerin tedarik sürecinin hangi aşamalarında (çiftlik, mezbaha, lojistik faaliyetler, perakendeci) ortaya çıktığı tespit edilmiştir.



Şekil 2.12. Twitter (X) analitik çerçevesi [1]

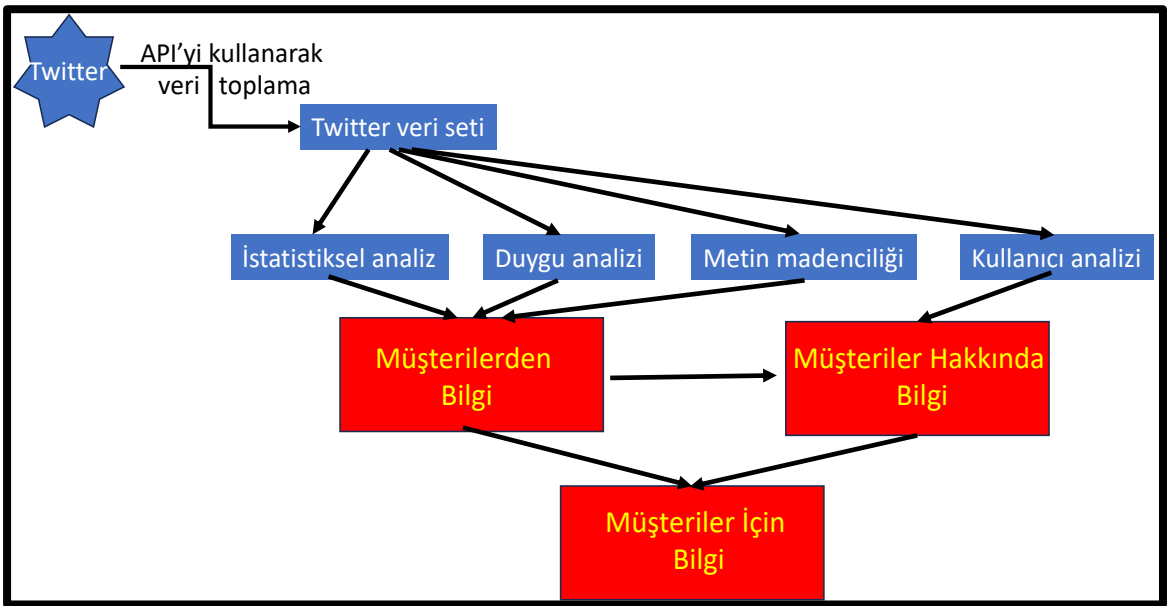
Singh vd. [20], Twitter (X)'dan elde edilen verilerin incelenmesi için SVM'yi kullanan metin analizi ve hiyerarşik kümelemenin uygulandığı bir Twitter (X) analitik modeli önermişlerdir. Önerilen model ile sığır et ürünleri konusundaki memnuniyetsizlikler ve bunların giderilmesi için tedarik süreçlerinde kimler tarafından ne tür tedbirler alınması gerektiğine yönelik hususlar tespit edilmiştir. Bu yaklaşım ile tedarik zinciri karar vericileri için; müşteri geri bildirimi, yiyecek ürün kalitesi ve akışı konularında olmak üzere iki konuda bilgiler ortaya çıkarılmıştır. Bu sonuçla müşteri odaklı tedarik zinciri oluşturulması hedeflenmiştir.

Singh vd. [20] tarafından önerilen model çerçevesinde yapılan vaka analizinde; coğrafi bölge sınırı konulmadan tweetlerin dünyanın hangi bölgesinden yazıldıkları harita üzerine çıkarılmış, sığır eti ve biftek kelimeleri ile üç haftalık bir zaman dilimi için arama yapılmış (beef ve steak kelimeleri ile), anahtar kelime tespiti için İngiltere'de perakendeci ziyaretleri gerçekleştirilmiş, işyerleri veri tabanından müşteri şikayetlerine ilişkin bilgiler incelenmiş, çalışanlar ve tüketicilerle ayrı ayrı mülakatlar yapılmış, oluşturulan anahtar kelime listesi ile tweet filtrelemesi yapılmış, her ülkeye karşılık gelen market isimleri kullanılarak, tweetlerin ülkelere göre sınıflandırması yapılmış, duygu analizinde SVM kullanılmış, hiyerarşik kümeleme (yeniden örneklemeyle çoklu önyükleme) işlemi gerçekleştirilmiş, toplanan tweetlerde 1 000'den fazla kullanılan etiket tespit edilmiş, en etkin Twitter (X) kullanıcıları frekans ve yüzdelik olarak tespit edilmiştir.

El- Khchine, vd. [21]'nin çalışmasında, tavuk tedarik zinciri yönetiminde Twitter (X) verisinin ve KNN, LR ve SVM algoritmalarının kullanımına dayanan bir analitik yaklaşım önerilmiştir. Çalışma, tavuk ürünlerine ilişkin ana ilgi alanlarını tanımlayarak tüketici merkezli tedarik zincirinin gelişmesine öneriler getirmiştir.

He vd. [261] tarafından yapılan çalışmada, Twitter (X) araştırma API'sinden bazı laptop markalarına ait tweetler çekilerek, istatistiki analiz, metin madenciliği, duygu analizi teknikleri uygulanmış ve elde edilen sonuçlar görselleştirilerek çıkarımlarda bulunulmuş, şirketlere müşteriden elde edilen müşteri hakkındaki bilgilerin yine müşteri için nasıl kullanılabileceğine yönelik hususlar ortaya konulmuştur. Müşteriyi merkeze alan çalışmasında He vd. [261] müşteri bilgisini; organizasyonlar tarafından kendi ürünlerini daha iyi anlamalarını sağlamak için organizasyonlardan müşterilere gönderilen bilgi (*müşteriler için bilgi*), organizasyonun müşteri ihtiyaçlarını daha iyi anlayabilmesi için

ihtiyaç duyduğu bilgi (*müşteriler hakkında bilgi*), bir hizmet veya ürünü kullanan ve bu nedenle ürün ya da hizmet hakkında elde edilmiş olan ve müşteride bulunan tecrübe ve bilgi (*müşterilerden elde edilen bilgi*) şeklinde üç grupta kategorize etmişlerdir. Müşteri ilişkileri yönetiminde en değerli bilgi kaynağına [281] yönelen bu çalışmada analizlere yönelik; metin madenciliğinin özel bir uygulaması olan otomatikleştirilmiş duygu analizi [282]; metin araştırması için VOZIQ, Apache Solr; büyük veri analizi için Hadoop; işlenmiş veriyi depolamak için MySQL ve web tabanlı görselleştirme için JavaServer Pages [283] uygulamaları kullanılmıştır. Çalışmada önerilen Twitter (X) veri analitik yaklaşımı aşağıdaki Şekil 2.13'te gösterilmiştir.



Şekil 2.13. Müşteri odaklı Twitter (X) veri analitik yaklaşımı [261]

He vd. [261] analizler neticesinde en çok konuşulan laptop markaları, markalar hakkında ne oranlarda hangi konularda olumlu ve olumsuz konular konuşulduğu, laptopların en çok hangi özelliklerinden bahsedildiği gibi çeşitli bilgilere ulaşmıştır.

2.4.7. Çevrimiçi alışveriş, yemek siparişi ve sosyal medya

Tüm dünyada olduğu gibi Türkiye'de de her geçen gün internet kullanıcı sayısı artmaktadır. Türkiye 50 milyon üzerinde nüfusa sahip ülkeler arasında yapılan internet kullanıcı sayısındaki net artış sıralamasına göre, bir önceki yıla göre 5 000 000 kullanıcı sayısındaki

(% 9.3) artış ile 14'üncü sırada yer almaktadır [268]. Internet kullanımı arttıkça doğal olarak internet aracılığıyla platformlar üzerinden yapılan çevrimiçi alışverişlerin de artması beklenmelidir. Tüketiciler ürün ve hizmet alımı konusunda çeşitli nedenlerle, çevrimiçi alışveriş yapmayı tercih etmektedirler. Her geçen gün kullanıcı sayısı artan çevrimiçi alışveriş ortamına, yeni alışveriş platformları ve siteleri eklenmektedir. Tüketiciler, alacakları ürün ve hizmetlerde maksimum faydaya ulaşmaya çalışırken, üretici/işletmecilerde bir yandan mevcut müşterisini muhafaza etmeye, diğer yandan yeni müşteri ve pazarlara ulaşmaya çalışmaktadırlar. Bu süreç içinde internet ortamının sağladığı avantajları kullanabilen üretici ya da tüketiciler, önemli avantajlar elde ederek, sağladıkları faydayı artırmaktadırlar. Tüketiciler, ürün ve hizmet alımı için zaman ve para ayırırlar ve kullanacakları ürün ya da hizmet alımı esnasında olumsuz anlamda herhangi bir sürprizle karşılaşmadan deneyim yaşamak istemektedirler ve bu nedenle söz konusu ürün ve hizmetler hakkında çeşitli bilgi kaynaklarından bilgi toplamaktadırlar [284]. Bu bilgi kaynaklarının en önemlilerinden biri de diğer tüketiciler tarafından yapılan değerlendirmelerdir ve bu değerlendirmelere herkes için elektronik ortamda hızlı ve doğru bir şekilde ulaşma imkânı bulunmaktadır. Üreticilerin de bu değerlendirme ortamını kendi lehine olacak şekilde kullanabilmesi, rekabet avantajı elde etmelerini sağlayacaktır. Kullanılan ürün ve hizmetlerin tüketicilerin en doğal şekilde yorumladığı, olumlu ya da olumsuz görüşlerini diğer tüketicilerle paylaştıkları elektronik ortamların başında sosyal medya platformları gelmektedir. Sanal ortamların kullanımından önce kişiler, ürün ve hizmetler konusunda doğrudan iletişim ile tanıdıkları kişi ve çevrelerinden tavsiye almakta iken sanal ortam kullanımının yaygınlaşması ile bu tavsiyeleri sanal ortamlardaki değerlendirmelerden almaya başlamışlardır.

Qualman'ın [285] çalışmasında yer alan WOM tekniği için yapılan tanımlama, artık dijital ortamlara geçiş yapmış ve *elektronik ağızdan ağıza pazarlama* (Electronic-Word of Mouth, e-WOM) adını almıştır. Bu şekilde sosyal ağ ortamlarında müşteriler ürün ve hizmetler konusundaki tecrübelerini, pozitif ve negatif görüşlerini ya da söz konusu ürün ve hizmetlerin güçlü ve zayıf yönlerini, geliştirilmesine ihtiyaç duyulan özelliklerine yönelik hususları serbest bir metin ile ya da şekil ve sembol ya da puanlamalarla dile getirebilmektedirler. Ağızdan ağıza iletişimin yükselişi, tüketicilerin güveninin şirketlerden tüketicilere kaymasına neden olmaktadır. Nielsen Global Şirketinin bir araştırmasına göre; şirketlerin yaptığı reklama az sayıda tüketicinin güvendiği, araştırmaya katılanların % 90'ının tanıdıkları insanların önerilerine güvendiklerini, tüketicilerin % 70'inin internette

yayınlanan tüketici görüşlerine inandıklarını ifade ettikleri belirtilmiştir [284]. ABD'de yapılan bir araştırmada; tüketiciler, tavsiyeye başvurma ve tavsiye sağlama konusunda en çok ağızdan ağıza iletişime başvurulan ürün hizmet gurubunun, yiyecek ve içecek işletmeleri olduğu belirtilmiştir [286]. Twitter (X), serbest ve kendine özgün samimi bir sohbet ortamı olan platformdur. Bu özelliği dikkate alındığında Twitter (X)'ın, hizmet ve ürünler için e-WOM tekniğinin kullanılabildiği bir platform olduğu söylenebilir. Eğer marka ve firmalar kendileri için, kullanıcıların herhangi bir reklam mesajından daha çok güvendiği sosyal medya platformundaki değerlendirmeler arasında yer bulamazlarsa, önemli bir pazarı kaybetmiş olacaklardır.

Pazarlamaya yönelik stratejilerin belirlenmesi açısından insanların neden sosyal ağlarda olumlu ya da olumsuz görüş paylaştıklarının bilinmesi, marka ve firmaların kendi pazarlama stratejilerini geliştirmeleri açısından önemlidir. Özellikle kendi ürün/hizmetleri hakkında sosyal ağlarda paylaşılmış olumsuz görüşlerin yönünü olumluya çevirmeye çalışan ticari kuruluşlar açısından tüketicilerin görüş paylaşma konusundaki motivasyonlarının bilinmesi gerekir. Literatürde tüketicilerin ağızdan ağıza negatif görüş bildirmeye neden olan faktörlerini inceleyen Bovee vd. [287]'nin çalışmasında bu konuya etki eden faktörler; (i) tüketicinin memnuniyetsizliği, tatminsizliği, (ii) işletmenin şikâyetlere nasıl karşılık verdiğine dair tüketici algılamaları, (iii) sorunların çözülmemesi karşısında tüketicinin duyduğu rahatsızlık, (iv) sorunların işletmenin kontrolü altında olup olmadığına ilişkin tüketici algılamaları, (v) başarısız hizmet ya da ürün sonucunda müşterinin olası davranışsal tepkileri olarak sıralanmıştır. Sundaram vd. [288] tüketicileri negatif ağızdan ağıza iletişime iten motivasyon faktörlerini; endişeyi azaltma, intikam alma, tavsiye arama, başkalarını düşünme şeklinde sıralamış ve tüketicilerin bu konudaki motivasyonlarını, ürün performansı, çalışan davranışları, işletme tarafından problemlere karşı verilen cevaplar, fiyat-ürün değeri şeklinde dört esas başlık altında toplamıştır.

Armağan ve Eskici [289] çevrimiçi yemek siparişlerine ilişkin çalışmasında, tüketicilerin söz konusu hizmetten istifade etme motivasyonunu incelemişlerdir. Bu kapsamda motivasyon unsuru olarak; haz ve zevk alma, yeni tatlar deneme arzusu, önceden çevrimiçi satın alma ile yaşanan deneyimin sonraki alışverişlerde kolaylık ve rahatlığı artırması, tüketicinin zaman tasarrufu sağlaması, daha ucuz ürün ve hizmete ulaşma imkanı, kullanım sonrası ürün hakkında değerlendirme yapma imkanı, istenilen ürün çeşidine hızlı bir şekilde zaman ve mekan sınırlaması olmadan ulaşılabilmeyi tespit etmişlerdir.

Özaslan ve Uygur [284] tarafından yiyecek ve içecek sektörüne ilişkin yapılan çalışmada; restoran, fastfood, kafeterya, pastane ve diğer sınıf işletmeler şeklinde sınıflandırma altında inceleme yapılmıştır. İncelenen bu sektörlerde müşterilerin en çok karşılaştıkları problemler; servis yetersizliği ve yavaşlığı, servis ve yemeğe göre fiyatın yüksekliği, lavabo ve tuvaletlerin temiz olmaması, çalışanların nezaketsizliği, tanıtımlarda sunulan ile sunulanın aynı olmaması, lezzet, ısı, koku ve sunumun iyi olmaması, ortamın genel olarak temiz olmaması, porsiyonların küçük olması, ortamın kalabalık olması ve diğer hususlar şeklinde sıralanmıştır.

Namkung ve Jang [290] yiyecek ve restoran sektörüne yönelik çalışmalarında, menü çeşitliliği, sağlıklı alternatifler, tat, sunum, tazelik, yemeğin olması gereken sıcaklıkta servis edilmesi şeklinde kategorilendirme yapmıştır. Aynı konuda Liu ve Jang [291] sektörü değerlendirmek için tat, menü çeşitliliği, gıda/yiyecek güvenliği konularını içine alan *yiyecek kalitesi*; tutarlılık, çalışanların ilgi ve alakası/yardımseverliği başlıklarını kapsayan *hizmet kalitesi*; temizlik ve tasarım, tasarım içinde müzik, renk, ışık, koku ve dekorasyona ait unsurları içeren atmosfer olmak üzere üç kategori başlığı kullanmışlardır.

Zahoor vd. [292] Karaçi'deki restoranlar hakkındaki yorumların duygu analizi ve tat, ambiyans, hizmet ve değer başlıkları altında kategorizasyonuna ilişkin çalışmalar; Alamoudi ve Alghamdi [293] yiyecek, hizmet, ambiyans, fiyat hedef özellikleri çerçevesinde duygu sınıflandırmaları yapmışlardır. Liapakis vd. [294] yiyecek ve içecek sektörü için, 2018 yılı bir aylık bir dönemine ait müşteri yorumlarını analiz ettikleri çalışmada, analiz için yiyecek kalitesi, müşteri hizmetleri, şirket imajı, fiyat, ürün miktarı olmak üzere beş özellik belirlemişlerdir. Çetin ve Eryiğit [97] hedef tabanlı olarak restoran yorumlarını içeren veri seti üzerinde yaptıkları çalışmada hedef varlıklar olarak restoran, yiyecek, içecek, ambiyans/ortam, servis/hizmet, lokasyon/yer/konum; hedef özellikler olarak, genel, fiyat, kalite, tarz/stil ve opsiyonlar/seçenekler/alternatifler/çeşitlilik ve diğer başlıkları şeklinde analiz yapmışlardır.

Yukarıda sıralanan literatür çalışmaları ve Twitter (X) metinlerinin özellikleri dikkate alındığında, çeşitli nedenlerle ÇYS veren müşterilerin söz konusu hizmete ilişkin memnuniyet ve memnuniyetsizliklerinin, *hizmet kalitesi*, *ürün kalitesi* ve *diğer konular* şeklinde üç ana başlık altında toplanabileceği öngörülmektedir.

2.5. Ağ analizi

Ağ yapıları, yapıyı oluşturan elemanlar arasındaki iletişim, etkileşim, bağ ya da ilişkileri gösteren, düğüm ve bu düğümleri birbirine irtibatlayan kenarlardan oluşan yapılardır. Ağ modelleri ve analizi mühendislik, bilgisayar, elektronik, pazarlama gibi birbirinden farklı birçok alanda farklı sistemlere uyarlanarak kullanılabilir.

İçinde insan ilişkilerinin yer aldığı ağlar, genel olarak sosyal ağlar olarak ifade edilir, ağ yapılarının analizi, söz konusu ağdaki ilişkilerin yorumlanarak çıkarımlar yapılmasını ve ağdaki bilginin ortaya çıkarılmasını sağlar, bu analiz ile ağdaki insan ilişkilerinin yapısal özellikleri, yapı içindeki kişilerin birbirlerine göre konumlanmaları, ilişkilerin gücü ve ağın topolojik yapısı tespit edilir [295]. Ağ analizi tanımlamalar, görselleştirme ve istatistiksel modellemelerle bahsedilen ağ yapısının anlaşılmasını hedefler [296]. Ağ içindeki aktörler araştırma konusuna göre birey, örgüt veya hayvan olabilmektedir. Bağlar ise arkadaşlık, işbirliği, danışma, güven veya nefret gibi ilişkiyi niteleyebilmektedir [295]. Ağdaki ilişkilerin yapısal özelliklerinin analizi için sosyal ağ yapılarının farklı karakteristik özelliklerinin bilinmesi gereklidir. Sosyal ağdaki her aktör, farklı ağ türlerinde farklı etkilere sahiptir. Bir aktör ilgili ağ türünde merkezi konumundan fayda sağlarken, başka bir aktör farklı bir ağda diğer bireylere olan yakınlığından fayda sağlayabilir [297].

Ağ yapısı içinde ilişki ve iletişimin tek taraflı ya da çift taraflı olması, aradaki bağın gücü, ağdaki ilişkilerin yapısının anlaşılmasını sağlar. Analiz edilen ilişkinin doğası, ağın yapısını belirlemektedir. Örneğin, akrabalık ilişkilerinin incelendiği bir ağda, ilişkilerin karşılıklı ve okların iki taraflı olması ilişkinin niteliğinden kaynaklanmaktadır. Tam tersine tavsiye ağlarında (ilişkilerin yönlendirme ve tavsiyelerle kurulduğu ağlar) okun yönünün tek taraflı olması beklenmektedir [298]. Aktörler arasındaki ikili ilişki, yönlendirilmiş ağ ve yönlendirilmemiş ağ ile açıklanır.

İnsan ilişkilerinin internet ortamına taşınması ile yeni sosyal ilişkiler ortaya çıkmış, farklı kişi, grup, toplulukların birbirleri ile iletişime girebilmeleri hız kazanmıştır. Bu da, bilgi akışı için yeni kanallar oluşturmuş, ortaya çıkan yeni bilgi akış kanallarının ağ analizi ile incelenmesi de faydalı bilgilere ulaşılmasını sağlamıştır. Bu bilgiler, sosyal medya ile ortaya çıkan yeni iletişim kanallarını ve bunların oluşturduğu ağları kontrol edebilen, etkileyebilen, yönlendirebilen merkezi konumdaki aktör ve ağ içindeki topluluk bilgilerini kapsar ve ticari

işletmeler açısından önemlidir. Çünkü bu aktörlerin herhangi bir ürün ve hizmet hakkında kişi ve toplulukları yönlendirebilme gücünü kullanarak, kitlelere ulaşma imkânı ortaya çıkmıştır. Ağ içindeki bilgi akışı ve ağdaki aktörlerin analizi sayesinde firmalar ürün ve hizmetlerine yönelik müşteri kümelerini tanımlayabilir, onlara ulaşabilir ve bu müşteri kümelerine göre hedef pazarlarını yönetebilirler [299].

2.5.1. Sosyal medya ağ analizleri (SMAA), Twitter (X) ağ analizi, ağ analizine yönelik yönelik literatür

SAA kavramından, bireyin diğer bireylerle etkileşime girdiği sosyal çevre ve ilişkilerden doğan sosyal yapıların incelenmesi anlaşılmaktadır [295]. Sosyal medya kullanımının etkinlik kazanması ile internet ortamında da yeni bir sosyal çevrenin oluştuğu, bu çevre ile farklı düzeylerde etkileşime girildiği, bu etkileşimlerin işletmeler tarafından kullanılabilecek yeni ortamlar oluşturduğu görülmektedir. Sosyal medyadaki bu yeni ortama ait veri setleri üzerinde yapılacak analizlerin doğru yorumlanabilmesi için, ortamdaki etkenlerin iyi anlaşılması gerekir. Veri setlerinin bir etki ve yönlendirme ile mi ortaya çıktığının tespit edilmesi, böyle bir etki varsa bu etkinin hangi kişi ya da topluluklar tarafından oluşturulduğunun ve bu etkinin nasıl yönetilebileceğinin bilinmesine ihtiyaç vardır. Bu ihtiyaç, sosyal ağdaki etkili aktörlerin, ağ yapılarında ortaya çıkan toplulukların ortaya çıkarılması ile mümkündür [300, 301].

Sosyal ağlarda, kendi görüşlerini doğrudan paylaşan bireyler, müşterileri satın almaya yönlendiren ya da gündemi istenen yönde etkilemeye çalışan çevrimiçi etkileyiciler, yönlendirici kişi ya da gruplar, sektör temsilcisi büyük organizasyonlar, ulusal, uluslararası resmi kurum ve kuruluşlar gibi irili ufaklı birçok aktör, rekabet içinde tek başına ya da değişik şekillerde gruplar oluşturarak etkinlik göstermekte ve ilgili olduğu alanı şekillendirmeye çalışmaktadır. Böyle bir ortamdan elde edilecek bilgilerin yönlendirilmiş etkilerden arındırılarak doğru yöntem ve araçlarla okunması gerekir.

Sosyal ağ ortamının şekillendirilmesi çabası, bu kaynaktan istifade eden kullanıcılar açısından bazı güçlükler ortaya çıkarmaktadır. Bu güçlüklerden biri; sosyal ağlardan elde edilen verilere dayalı kamuoyu gündemine yönelik araştırmalarda ya da herhangi bir ürün ve hizmete ilişkin analizde, yönlendirilmemiş gerçekçi ve doğru bilgiye ulaşmaktır [302]. Bir diğer güçlük ise ilgilenilen alanda istenen bir bilgiyi istenen kullanıcılara

ulaştırabilmektir. Birinci güçlüğün aşılması için ilgilenilen konuda sosyal platformlarda yönlendirici etkisi olan kişi ve grupların tespit edilerek araştırma ve analiz sonuçlarının bu etkiden ayıklanmasından sonra yorumlanması, ikinci güçlüğün aşılması için bilgiyi istenen kullanıcılara ulaştıracak etkili kişi ve gruplarla iş birliği yapılması gerekecektir.

SAA, ağ ortamının özelliklerinin, yapısının ortaya çıkarılmasını, ağdaki güçlü aktörler ve aralarındaki bilgi akışı ve iletişimin anlaşılmasını sağlar [303]. Ağ analizi ile önceden bilinmeyen, diğerlerine göre daha kuvvetli bağlanmış düğümlerin ve bunların oluşturdukları toplulukların tespit edilebilmesi açısından önemlidir [304].

SAA, aradaki iletişim kanallarının yapı ve özelliklerinin araştırılmasını gerektirir [296]. Sosyal ağ ortamında bilgi akışının gerçekleştiği kanalları kontrol edebilen, etkileyebilen, yönlendirebilen aktörler, ticari işletmeler açısından önem kazanmıştır. Çünkü bu aktörlerin herhangi bir ürün ve hizmet hakkında olumlu ya da olumsuz görüş bildirmeleri kendi kitleleri üzerinde etki yaratır hale gelmiştir. Bu nedenle müşteri hedef kitlesini ve gruplarını tanımlamaya, yönetmeye, yönlendirmeye çalışan, işletmeler, söz konusu aktörleri göz ardı edemezler [299].

Twitter (X) gibi sosyal ağlarda takipleşme her zaman bağlantı için yeterli bir gösterge olmayabilir [305]. Arkadaşlık ilişkisi ile oluşturulmuş bağlantılar olmadan da kullanıcılar arasında ilişki kurulabilir. Bir kullanıcı Twitter (X)'ın özelliklerini kullanarak, bir ağ yapısı içinde yer alır. Bu yapı içindeki kullanıcıların ilişkileri, etkinlikleri, konumlarına ilişkin bilgiler graf teorisindeki teknik ve ölçümlerle tespit edilebilir. Twitter (X) kullanıcılarının ağ teorisindeki teknik ve ölçümleri kullanarak Twitter (X) verisinden ağ bilgisini ortaya çıkarmak mümkündür [277]. Twitter (X)'a yönelik ağ analizi; (i) Twitter (X) kullanıcıları, (ii) Twitter (X)'daki içerik paylaşım (retweet) özelliği, (iii) sohbet etme ya da başlatma özelliği (tweet yazma ve cevap verme) ve (iv) bir kullanıcının takip ettiği kullanıcılar ve bu kullanıcıyı takip eden diğer kullanıcı bilgilerini kapsayan takipçi ilişkilerine dayandırılarak yapılabilir ve bu ilişkiler üzerinden incelenen ağa ilişkin bilgiler ortaya çıkarılabilir.

Çevrim içi ağlarda etkili kullanıcı ve toplulukların tespit edilmesi için; takipçi (statik/yapısal ilişkiler) ve etkileşim (dinamik/işlevsel ilişkiler) ilişkileri ve bunların karması kullanılır [306]. Söz konusu ilişkilerin yapısı ağın yerel, genel (küresel) merkezilik dereceleri ve bu

ikisinin hibrit kullanımı ile ortaya çıkarılır. Bu konuya ilişkin literatür ileriki maddelerde sıralanmıştır.

Ağ özellikleri, ağırlık merkezleri, düğüm ve bağların gücü ve etkinliklerinin tespiti için genellikle derece, yakınlık, arasındalık, özvektör, katz merkezilikleri ve sayfa sıralaması algoritması kullanılır. Bu merkezilik değerleri, farklı yöntemlerle ağırlıklandırılarak ağ yapılarındaki etkili kullanıcılar tespit edilir ve kullanıcılar arasında sıralamalar yapılır. Bu maksatla temel merkezilik ölçütlerinin minimum, maksimum değerlerini Fei vd. [307] en yüksek medyan puanıyla; Isfaq vd. [306] Katz merkeziliği ile birleştirerek kullanmıştır. Zareie vd. [308], komşuların bağlantı yapılarının benzerliğini dikkate alan bir küme sıralaması yaklaşımı geliştirmiştir. Yang vd. [309], ağ yapısını ve yerel merkezilik değerlerini birlikte kullanan, komşuluk ilişkilerinde öklit uzaklık hesabını kullanan ağ düğüm merkezilik indeksi (Network Node Centrality Index-NNCI) ismini verdiği bir yöntem önermiştir. Ullah vd. [310] etkili mesafe tabanlı merkezilik (Effective Distance-Based Centrality-EDBC) olarak isimlendirdiği önerisinde; düğüm derecesi, etkili mesafe, komşu etkisi veya komşuluk potansiyeli gibi faktörleri kullanmıştır. Önerisini, özvektör, arasındalık, yakınlık merkezilikleri, hiperlink kaynaklı konu arama, sayfa sıralaması gibi mevcut tekniklerle karşılaştırmıştır. Wan vd. [311], aktör etkinliği tespitinde ağ yapısının yayılımına odaklanmış, k-kabuk ve derece merkeziliklerini kullanarak ağırlıklandıran "bağlantı önemi" isimli bir yöntem önermiştir. Zhong vd. [312], düğümlerin merkezilik derecelerini ve bunların bir eşik değerle farklarını kullanarak etkilerini hesaplamış ve "kapsamlı etki" adıyla bir yöntem geliştirmiştir.

Rehman vd. [313] çalışmasında, Twitter (X)'de sohbet başlatan ve kanaat önderi olarak isimlendirilen etkili kullanıcıların durumlarını, iç-dış merkezilik dereceleri ve arasındalık merkezilikleri ile ele alarak incelemiş, kanaat önderlerini bu merkezilik derecelerine göre sıralamıştır. Kullanıcıları, etkileşim şekline göre sınıflandıran Rehman vd. [313] sohbeti başlatan kullanıcının ağı kontrol üzerinde etkisinin düşük olduğunu ancak bilginin yayılmasını sağladığını belirtmiştir.

Zhao vd. [314], temel merkezilik ölçütlerinin ağın farklı yönlerini öne çıkarırken, diğerlerini ihmal ettiğini belirterek, analiz için ağın yerel ve küresel bilgilerini birlikte kullanan yarı-yerel bir yöntem sunmuştur. Söz konusu yöntemle elde edilen etkili kullanıcı sıralamaları

derece merkeziliği, bağlı bileşenler, arasındalık merkeziliği, sayfa sıralaması ile elde edilen sonuçlarla karşılaştırılmıştır.

İspir ve Deniz [315], köşe yazarlarının Kasım 2015 seçimlerine ilişkin Twitter (X) gündeminin en önemli 11 konu başlığını belirlemişler, bu konuların önem derecelerini merkezilik dereceleri ile tespit etmişler; çalışmada UciNet, NetDraw yazılım araçlarını kullanmışlardır.

Arslan vd. [316], ekolojik aktivist örgütlenmelerin sosyal medya kullanımlarına ilişkin durumu ortaya çıkarmak maksadıyla, ekoloji ağında aktif olan aktörlerin ilişkilerini SAA tekniğiyle incelenmişlerdir. Demir ve Ayhan [317], tweetler üzerinde sosyal ağ analiz yöntemini, NodeXL yazılımını, merkezilik dereceleri ölçümlerini kullanarak, politika ve kamu gündeminin belirlenmesinde aktörlerin etkinlik ve önem derecelerinin tespit edilmesini göstermiştir.

Bakan [318], 40 sanat okulunun, kurumsal Twitter (X) kullanım düzeylerindeki farklılıkları ortaya çıkarmak amacıyla Twitter (X) veri seti üzerinde yaptığı analizde, bireyler arası etkileşimleri sosyal ağ analiz yöntemi ve merkezilik dereceleri ölçümlerini kullanarak incelemiştir. Kobak [319], "Tiktok kapansın" gündemine ilişkin topladığı veriler üzerinde Twitter (X) kullanıcıları arasındaki iletişim ve kümeleşmeyi tespit etmek için, NodeXL uygulaması kullanarak SAA yapmış, sonuçta düğümlerin genel yapısının bağımsız olduğunu, düğümler arası birebir etkileşimin düşük olduğunu tespit etmiştir.

Yee vd. [320], turizm sektörünün akışını (turist hareketleri, konaklama yerleri, turist davranış kalıpları) SAA ile ele alan çalışmaları inceleyen bir literatür çalışması yapmış, metodolojik açıdan sosyal ağ analiz çerçevesini bibliyometrik olarak incelemiştir. Eshaghi ve Karamidehkordi [321], bir göl restorasyonu kapsamında birlikte çalışan iştirakçilerin, paydaşların aralarındaki iletişim ve ilişkilerini, projenin uygulanması konusunda etkileşim yoğunluğunu, paydaşlar arasındaki grupları ve etkileşimdeki zayıflığın projeyi nasıl etkilediğini sosyal ağ analiz teorisi ile ortaya koymuştur.

Rodda ve Bhavani [322], bir romandaki karakterlerin, topluluk oluşturma ve ayrışmalarını sosyal ağ analiz teorisi ile incelemiştir. Costa ve Ralha [323], sürekli gelişen dinamik sosyal ağlarda topluluk ve kritik aktörlerin tespiti için, modüler yoğunluk fonksiyonunun lokal

optimizasyonuna dayalı derin güçlendirilmiş öğrenme stratejisi ile bir model önerisi geliştirmiştir.

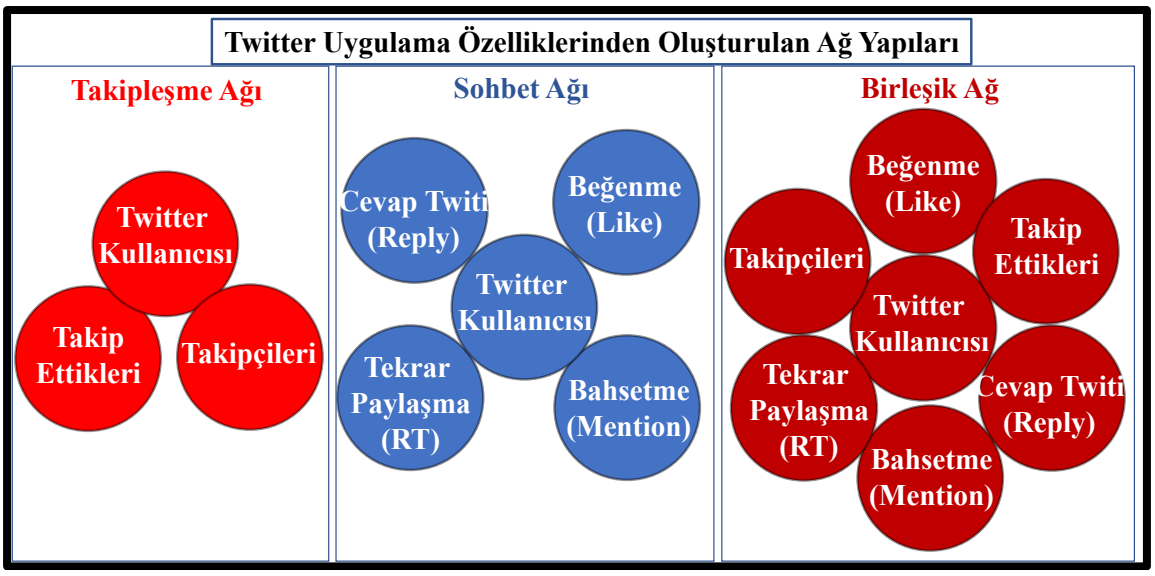
Hastuti vd. [324], seçimlere ilişkin tweetlerden oluşan veri seti üzerindeki görüş analizi ile ilgili çalışmalarında, derece merkeziliği hesaplamalarının, önemli bireylerin tespitinde kullanılabileceğini, ancak bu şekilde tespit edilen aktörlerin popülaritelerinin (takipçi sayılarının) de değerlendirmelere dahil edilmesi gerektiğini ifade etmişler, analiz için Gephi yazılımını kullanmışlardır.

Aragon [325], deniz arkeolojisi alanında uyguladığı SAA'nde, ülkeler arasındaki kültürel temaslarda mekânsal ve mekânsal olmayan unsurları entegre ederek arkeolojik kalıntılara yeni yorumlar getirmiştir. Bu çalışmada sosyal ağ analiz çerçevesi kullanılarak, gemi enkazları ile ülkelerin malzeme sevkiyatı, güç kontrolü gibi merkezilik durumları analiz edilmiş, analizde görselleştirme için Gephi yazılımı kullanılmıştır.

Pang [326], Aşk ve Entrika isimli drama eserindeki kahramanları SAA yöntemi ile inceleyerek başkahramanın tespit edilmesi için merkezilik derecelerini kullanmıştır.

Purbasari vd. [327], dijital inovasyon ekosisteminde en önemli rolleri oynayan aktörleri bulmak ve ağ yapısını elde etmek için Gephi uygulamasını kullanmış, uygulamada SAAyaklaşımı ile aktörlerin ağ yapısı içinde merkezilik derecelerini tespit etmiştir.

Ağ analizlerinde safhalandırma, genel olarak Twitter (X) gibi metinsel veri setlerinin analizinde de kullanılan safhalarla benzer şekilde kullanılabilir. Twitter (X) uygulama özelliklerinin ağ yapıları Şekil 2.14'te gösterilmiştir. Twitter (X)'da oluşturulan ağ türleri iki başlık altında toplanmıştır; (i) takip eden, takip edilen bilgisine dayanan arkadaşlık ağı, (ii) Twitter (X) kullanıcıları arasında kişilerarası ilişki başlatan sohbet (reply ve bahsetme, retweet) ağı [3]. Bu türlere birleşik analiz şeklinde sohbet ve takipçi ağının karma şekilde kullanımı ile üçüncü bir tür eklenebilir [299]. Çalışmaya göre farklılık göstermekle birlikte, ağ analiz seviyeleri için genelde birbirine benzer isimlendirmeler yapılır. Kim vd. [328] düğüm ve ağ seviyesi analizler; Chae vd. [3], Twitter (X)'daki takipçi ve sohbet ağları için topolojik, merkezilik ve topluluk analizleri, Ağcasulu [295] tüm ağ düzeyi, ilişki düzeyi ve aktör düzeyi başlıkları altında analiz çalışmaları yapmışlardır.



Şekil 2.14. Twitter (X) uygulama özelliklerinin oluşturduğu ağ yapıları

Twitter (X)'daki sohbetler açısından bağların yönü ile ilgili üç durum söz konusu olabilir; (i) eğer bir kullanıcı ilk defa bir tweet metni yazarak sohbet başlatıyor, bir başka kullanıcının tweetini RT yapıyor ya da bir kullanıcının adından bahsederek tweet yazıyorsa ok yönü dışa doğru, (ii) bir başka kullanıcıdan yazılan tweet için beğenme/beğenmeme yapılıyorsa ok yönsüz, (iii) bir başka kullanıcının tweeti sadece okunuyorsa/tıklanıyorsa ok yönü içe doğru olur.

Ağ analizi için kullanılan merkezilik ölçütlerin her birinin, en etkili düğümü bulmaya yönelik varsayımı birbirinden farklıdır [316], bu ölçütler bireysel düğümlere odaklanır ve en fazla bağa sahip kullanıcının ve ağdaki etkili aktörlerin ortaya çıkarılmasını sağlar [3].

Genellikle güçlü bağlar, birbiri ile ilişkili kişiler (aile, iş, arkadaş grupları gibi) arasında olur ve bu kişiler birbirine benzer özelliklere sahiptirler yani ortak noktaları ve özellikleri vardır, Granovetter [329], zayıf bağların farklı özellikteki grup ve kişileri bağlaması nedeniyle kritik özellikte olduğu ve kişiye heterojen yapıdaki daha geniş gruplara ulaşma imkanı sağladığı belirtilmektedir. Benzer şekilde Shi vd. [262]'nin çalışmalarında da, Twitter (X) kullanıcılarına uyarlanarak Twitter (X)'daki çift yönlü bağların güçlü, tek yönlü bağların ise zayıf bağ olduğu ifade edilmiştir.

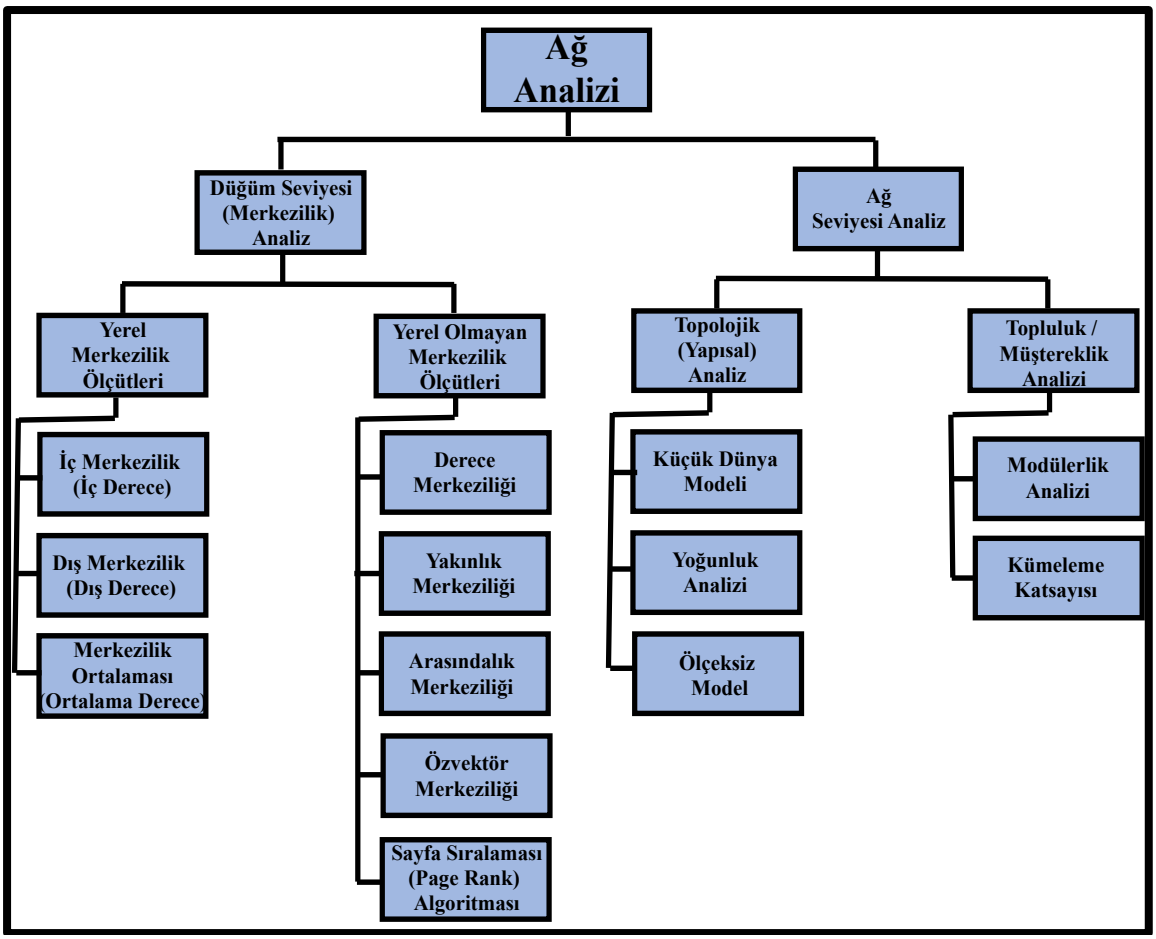
Ağ çalışmalarında önemli konulardan biri ağ büyüklüğünün belirlenmesi, ağ için dış sınırların belirlenmesidir [295]. Bunun, Twitter (X)'da yapılan ağ analizi içindeki karşılığı

ise sohbet şebekesinin ya da takipçi ağının sınırlarının nerede durdurulacağının belirlenmesi hususudur. Hangi aktörlerin ağ yapısına dahil edileceği konusunda; aktörlerin aralarındaki ilişkiler üzerinden hareket eden kartopu örneklemi ve başlangıçta hazırlanan bir liste üzerinden aktör belirleyen sayım örneklemi yöntemlerinden biri kullanılabilir [330].

SAA'leri, analizin amacına göre farklı aktörlerin ağ içindeki konumlarının tespit edilmesine yöneltilebilir. Etkili aktör ve topluluk tespiti; viral pazarlama, virüs engelleme, sosyal ağlarda istenmeyen bilgi yayılımını engelleme, veri setini doğru temsil eden örnek kümenin belirlenmesi gibi çeşitli amaçlarla kullanılabilir [309, 331]. Örneğin sosyal medya fenomenini araştırmaya yönelik analizler, ağdaki aktörleri merkeze alarak inceleme yaparken; ticari bir ürün ya da hizmete ilişkin analizler ağ içindeki müşteri gruplarının tespitini analizin odağı olarak belirleyebilir. Buna başka bir örnek, politik konulara ilişkin bir çalışmada, toplum algısını etkilemek için hangi aktörlerin kullanılabileceğinin tespiti ya da kamuoyunda ortaya çıkarılmış olan bir algının hangi aktörler tarafından sosyal medyada yaratıldığının tespit edilmesi ağ analizine konu olabilir. Diğer bir örnekte ise sosyal medyadaki çevrimiçi etkileyicilerin takipçileri ile birlikte oluşturduğu ağlardan hangilerinin daha güçlü ve etkili bir yapıya sahip olduklarının tespit edilmesi ve bu tespit sonucuna göre ticari bir ürün ya da hizmet için reklam verilebilecek çevrimiçi etkileyicilerin tespit edilmesi amaçlanabilir. Bu amaçlara uygun olarak ağ analizleri, aktörleri ya da kullanıcı topluluklarını incelemenin merkezine alabilir.

Ağ analizinde yapılacak analiz seviyesi dikkate alınarak, uygun ölçütlerin kullanılması gerekir. Literatürde yer alan ölçüt ve analiz seviyeleri birbirleri ile ilişkileri dikkate alınarak düzenlenmiş ve Şekil 2.15'te gösterilmiştir[8]. Yapılacak çalışmalarda, incelenecek konuya, amaca ve cevaplandırılacak sorulara göre uygun analiz ve algoritmaların seçilmesi ve yönetilebilir bir analiz ve ölçüm listesi hazırlanması gerekir [3].

[8] Düğüm ve ağ seviyesinde kullanılan analiz yöntemleri Şekil 3-3te listelenen yöntemlerle sınırlı olmayıp, yaygın kullanıma sahip yöntemler şekle dahil edilmiştir.



Şekil 2.15. Ağ analiz seviyeleri

2.5.2. Düğüm/aktör seviyesi analiz

Bir sosyal ağ içerisindeki aktörlerin, katkılarını ifade edecek şekilde önem derecelerine göre belirlenmesi, SAA'nin konusu olan merkezilik ölçütleri ile mümkün olabilir. Birey ya da aktörün hangi derecede merkezi bir konumda oldukları ağ merkeziyeti tanımlaması ile ifade edilmektedir. Ağ analizinde mikro düzey olarak kabul edilen aktörün ağ içinde ne kadar merkezde olduğu onun ağ ilişkilerine katılma derecesi ile belirlenir [295]. Merkezilik aktörün doğrudan ve dolaylı olarak sahip olduğu bağların sayısıdır ve bu aktörün merkeziliği etki alanında yer alan bütün aktörlerin ortalama uzaklığını ifade eder [315]. Ağdaki diğer aktörlerle en fazla bağı olan aktörleri ortaya çıkaran bu bağlantılar açısından aktörün ağdaki durumunu belirleyen merkezilik analizi, odak aktörlere en yakın aktörlere odaklanır [3]. Ağ içindeki merkezilik arttıkça, kişilerin/düğümlerin/aktörlerin sahip olacağı sosyal güç, bilgi ve yaratıcılık gibi özellikler de artmaktadır [332]. Ticari işletmeler için merkezi durumdaki aktörlerin tespit edilmesi, hedef müşteri kitlesinin bu aktörler üzerinden etkilenmesi

açısından önemlidir. Merkezi durumdaki aktörlerin istenen yönde etkilenmesi, sosyal medyayı kullanan topluluklar açısından zincirleme şekilde etki yaratarak yeni aktör ve topluluklara ulaşılmasını sağlayabilirler. Çünkü merkezi konumdaki aktörler, farklı toplulukların arasında köprü oluşturur ya da topluluklara liderlik ederek ağda önemli rol üstlenirler. Böyle bir etki sadece işletmeler açısından değil, kamuoyunu etkilemeye çalışan her organizasyon açısından önemlidir.

İşletmeler ağ yapısı içinde yer alan bir aktör/düğümün sahip olduğu bağ sayısı, diğer düğümlere ne kadar yakın olduğu, hangi düğümler arasında kalarak güç kazandığı, ağ içinde akışı engelleyebilme gücü, hangi düğümlerin bağlanmasında en kısa yol üzerinde bulunduğu, sahip olduğu bağlardan ne kadarının sıkı ve gevşek bağlardan oluştuğu gibi özelliklerini tespit ederek, işletme amacı doğrultusunda kullanabilirler [333-336].

2.5.3. Ağ seviyesi analiz, topluluk analizi

Tüm ağ seviyesindeki analiz, ağ yapısına kuşbakışı bakarak tüm ağın yapısını ortaya çıkarmaya çalışır, ağ içinde topluluklar, kümeler oluşup oluşmadığı, aktörlerin/düğümlerin ağ içinde ne tür yapılar oluşturdukları tespit edilmeye çalışılır. Topluluk analizi, tüm ağ içindeki toplulukların sayıları, en büyük topluluk, en küçük topluluk, bu toplulukların üstlendiği konuların neler olduğu, toplulukların yüzdeleri gibi ölçüm araçları kullanılarak ağ yapısı hakkında bilgileri ortaya çıkarmayı hedefler. Bireysel düğüm noktalarının grup oluşturma eğilimleri üzerine ölçümler yapılır, ağ içindeki alt gruplar belirlenir. Ağ içindeki topluluğun topluluk analizi ile tespiti, bir yandan grup içi bağ yoğunluğunu maksimize ederken diğer yandan gruplar arası bağları minimize ederek bulunur [295, 299].

2.5.4. Ağ analizi için kullanılan algoritmalar ve gösterim şekilleri

Ağ yapılarının ortaya çıkarılması için kullanılan ve düğümlerin ağ oluşturma eğilimlerini inceleyen üç tür model yaygın olarak kullanılmaktadır. Bu modeller; (i) *klasik ağ modellerinde* düğümlerin bağ kurma olasılıklarının eşit değerler etrafında oluşmasını ve ağda homojen bir dağılım olduğunu esas alan rastgele grafik model (random graph model) [337], (ii) *basit ağ modellerinin* rasgele düğümler arasında kısa yol uzunlukları olabileceğine dayanan küçük dünya modeli (small-world model) [338] ve (iii) *karmaşık yapılara sahip* ve yeni düğümlerle sürekli büyüyen büyük ağ modellerinin, tercihli olarak daha güçlü

bağlantılara doğru yöneldiği fikrine dayanan ölçeksiz modeldir (scale-free model) [339]. Ağ yapılarının analizi ve görselleştirilerek yorum yapılmasını sağlayan Gephi, Cytoscape, NetworkX gibi çeşitli yazılım araç ve kütüphaneleri mevcuttur. Python ile kullanılan NetworkX ve Graphviz kütüphaneleri, ağ analizine yönelik algoritmaları kullanarak, ağdaki düğümler hakkında bilgi edinilmesini ve ağ yapısının ortaya çıkarılmasını sağlar [340]. Bu algoritmalar özellikleri ile Çizelge 2.5'te verilmiştir. Çalışmada Python programlama dili ile NetworkX kütüphanesi kullanılarak ağ yapısına yönelik analiz ve değerlendirmeler yapılmıştır.

Çizelge 2.5. Network X kütüphanesi ağ analiz algoritmaları

Algoritma	**Algoritmanın kullanım amacı / özelliği**
Derece merkeziliği [341]	Düğümlerin ağ içindeki merkeziliklerinin, önem derecelerinin tespiti
Arasındalık merkeziliği [341]	
Yakınlık merkeziliği [341]	
Özvektör merkeziliği [341]	
Katz merkeziliği [341]	
Louvain metodu [342]	Topluluk tespiti
Girvan-Newman [343]	
Etiket yayılım (label propagation) [344]	
Erdős-Rényi [345]	Ağın yüksek ya da düşük kümeleme katsayılarını kullanarak ağ yapısının özelliklerinin tespiti
Watts-Strogatz [346]	
Mesafe algoritmaları [347]	En kısa yol, ortalama en kısa mesafe, ağ çap ve yarıçapı gibi istatistiki özelliklerin tespiti
Bağlantı algoritmaları [348]	Ağ içindeki topluluk oluşturan bileşen ve izole kalmış düğümlerin tespiti
Kümeleme algoritmaları [349]	Topluluk oluşturma eğilimlerinin tespiti
Adamic-Adar indeksi [349]	Düğümler arası bağlantı ihtimalini öngörme
Jaccard katsayısı [351]	
Tercihli bağlantı (preferential attachment) [352]	

Şebekelerin görselleştirilmesi için en çok kullanılan algoritmalardan biri diğeri Fruchterman – Reingold algoritmasıdır. Bu algoritma kuvvet yönlendirmeli algoritmalardandır. Bu algoritmada düğümleri birleştiren bağlara birer kuvvet ataması yapılır. Kuvvet yönlendirmeli algoritmalar, makul süre içinde yüksek kalitede çözümler üretme özelliğiyle ön plana çıkmaktadırlar [353]. Özellikle yönsüz sosyal şebekelerde kullanılan bir başka etkili görselleştirme algoritması ise Harel – Koren Hızlı Çok Ölçekli (Harel-Koren Fast Multiscale Layout) algoritmasıdır. Söz konusu algoritma ile çok büyük boyutlu grafikler, düz çizgili olarak estetik bir şekilde çizilebilmektedir. Bu algoritma, kuvvet yönlendirmeli algoritmanın hızını önemli ölçüde iyileştirebilmektedir [354]. Ağ analizi için amaca uygun düğümün gücü ya da topluluğun yapısına yönelik algoritmalar kullanılır. Söz konusu

algoritmaların, analize ilişkin bilgilerin anlaşılması için kullandıkları farklı gösterim şekilleri gruplandırılmış olarak Çizelge 2.6'da gösterilmiştir [355].

Çizelge 2.6. Ağ yapıları gösterim şekilleri

Ağ Yapıları Gösterim Şekilleri
Kuvvet yönlendirmeli gösterim
Açı yönelimli düzen
Sıkıştırılmış gösterim
Grid gösterim
Matris yapı şeklinde gösterim
Doğrusal yönlü gösterim
Ana halka düzeni
Parçalı örtüşen kenar düzeni
Gen ve ağaç düzenleri (hiyerarşik düzen, ana hat ağaç düzeni, radyal ağaç düzeni, akıllı ağaç düzeni, dönen ağaç düzeni)
İlişkisel ve metabolik düzen
Göreceli ana hat düzeni ve uzamsal (mekânsal) yayılım düzeni
Diagram kenarları düzeni
Üç boyutlu gösterim

Çalışma kapsamında da bir kısmı kullanılan NetworkX ve Graphviz kütüphanelerinin grafik düzen gösterimleri Çizelge 2.7'de verilmiştir [356, 357].

Çizelge 2.7. NetworkX ve Graphviz kütüphanelerinin ağ grafik düzenleri

NetworkX ve Graphviz Kütüphanelerinde Kullanılan Ağ Grafik Düzenleri
İki parçalı düzen
Dairesel düzen (Circo)
Kamada-Kawai düzeni
Yeniden ölçeklendirilmiş düzen
Kabuk düzen
Yay düzeni
Spektral düzen
Sarmal düzen
Çok parçalı düzen
Rastgele düzen
Düzlemsel düzen
Hiyerarşik ya da katmanlı nokta düzeni (dot)
Yay modeli (Neato) düzeni
Kuvvet yönlendirmeli yerleşim düzeni (Force-Directed Placement-FDP)
Ölçeklenebilir kuvvet yönlendirmeli yerleşim düzeni (Sparse Force-Directed Placement-SFDP)
Dairesel düzen
Radyal düzen (Radial Layout-Twopi)
Sıralı spektral birleşim (Ordered Spectral Agglomeration-Osage)
Yamalı çalışma düzeni (Patchwork)

3. ÇEVRİMİÇİ YEMEK SİPARİŞİ KONUSUNDAKİ TWEETLERİN TANIMLAYICI, METİN VE AĞ ANALİZİ YÖNTEMLERİ İLE ANALİZİNDEN ELDE EDİLEN BULGULAR VE YORUMLAR

Bu bölümde çevrimiçi yemek siparişleri konusunda Twitter (X)'den toplanarak oluşturulan veri setinin üç farklı analiz yöntemi kullanılarak elde edilen sonuçlar verilmiştir.

Birinci aşamada toplanan veri setinin tanımlayıcı analiz araçları ile analizi yapılmış; buradan elde edilen sonuçlarla, işletmelerin İG maksadıyla kullanabilecekleri müşteri görüşlerinin veri seti içinden nasıl çıkarılabileceğini gösteren ve Twitter (X) Veri Seti İçeriğinin Tanımlayıcı Analiz ile Keşfi olarak adlandırılan bir model sunulmuştur. Modelde analiz sonuçlarının reklam içerikli metinlerin etkisinden kurtarılması için analiz öncesi ve sonrasında uygulanabilecek yöntemler gösterilmiştir. Ayrıca bu analiz yöntemi ile çalışma sonunda, ÇYS'ne ilişkin Twitter (X)'da yapılan paylaşımların ana gündem konuları tespit edilmiştir.

İkinci aşamada, veri setindeki tweetlerin iki kademeli sınıflandırması için model önerileri geliştirilmiş ve performans ölçümü yapılmıştır. Bu sınıflandırma modelleri biri duygu sınıflandırması, diğeri ürün-hizmet sistemleri sınıflandırmasıdır. Bu iki metin sınıflandırma modeli ile tweetler müşterinin ürün ve hizmet konularından memnuniyetinin/memnuniyetsizliğinin tespit edilmesini sağlayarak işletmelerin ürün ve hizmetlerinin iyileştirilmesi ve geliştirilmesi konusunda yön verebilecektir. Bu aşamada ayrıca metinlerin sınıflandırılmasında karşılaşılan problemlerden boyut indirgeme ve öznitelik seçimi için önerilen yöntem sunulmuş, sınıflandırmalar için temel ve güçlendirilmiş olarak adlandırılan sınıflandırma sözlükleri geliştirilmiştir. Sınıflandırma modellerinin geliştirilmesi için 10 makine öğrenme algoritması, üç terim ağırlıklandırma yöntemi, temel ve güçlendirilmiş sözlükler kullanılarak duygu sınıflandırması için 72 ve ürün-hizmet sistemleri sınıflandırması için 72 olmak üzere 144 model karşılaştırması yapılmıştır.

Üçüncü aşamada, Twitter (X) gibi sosyal medya platformlarından çekilen veri setlerinin analiz sonuçlarının, yönlendirici etkilerden arındırılarak doğru şekilde yorumlanabilmesi, sosyal medya platformlarında marka görünürlüğünün artırılabilmesi maksadıyla, platformlardaki etkili aktör ve topluluklar ağ analiz yöntem ve teknikleri ile tespit edilmiştir.

Çalışmayla, anahtar kelimelerle oluşturulmuş bir veri setinin, ilişkisel bir yöntem olan ağ analiz yöntemi ile nasıl analiz edilebileceği gösterilmiştir.

3.1. Anahtar kelime ve etiket tanımlama, koleksiyonun oluşturulması

3.1.1. Anahtar kelime / etiket kelimelerini belirleme

Veri havuzunun istenen şekilde oluşturulabilmesi için sorgulamalarda kullanılacak ifadelerin doğru belirlenmesi, tüm analiz sonuçlarını doğrudan etkileyen önemli bir aşamadır. Veri havuzunun oluşturulması için, Twitter (X)'dan anahtar kelime ya da etiket (hastag) kelimeleri kullanılarak arama ve veri çekme işlemi yapılabilir.

Çalışmada anahtar kelime kullanılarak veri havuzu oluşturulmuştur. Anahtar kelime tespiti için; çalışma kapsamında sırasıyla şu işlemler yapılmıştır: (i) İnternet üzerinden Türkiye'de eve/işyerine hazır yemek servis hizmeti veren şirketlerin kullandığı platformlarda yer alan yorum, şikayet, öneri, konuya ilişkin tartışma/karşılaştırma yazıları [358] içeren metinler manuel olarak incelenmiştir. (ii) Facebook/Twitter (X) kişisel kullanıcı hesabı üzerinden konuya ilişkin yorum, şikayet ve öneriler incelenmiştir. "Yemek", "sipariş", "pizza" kelimeleri için inceleme yapılan 143 tweet yorumunun tarih aralığı, 26 Kasım 2009-28 Mart 2018'dir. "Yemek", "sipariş", "döner" kelimeleri için inceleme yapılan 78 tweet yorumunun tarih aralığı, 26 Ocak 2010-8 Ağustos 2019'dur. (iii) Konuya ilişkin yorum, şikâyet ve öneriler ayrıca Google arama motoru üzerinden "yemek, sipariş, şikâyet, memnuniyet, pizza, döner, lahmacun, hamburger, pide" kelimeleri kullanılarak sorgulanmış, rastgele seçilen internet adreslerindeki [359-366] yorumlar incelenmiştir.

Bu sorgulama sonuçlarından EK-2'deki tohum kelime listesi çıkarılmıştır. Tohum kelime listesi; eş ve zıt anlamlı sözlükler, literatürdeki duygu sözlükleri ve yiyecek alanındaki çalışmalar kullanılarak genişletilmiş ve bu işlem sonucunda; pozitif, negatif sınıflardan oluşan *temel duygu sözlüğü*, ürün ve hizmet sınıflarından oluşan *temel ürün ve hizmet sistemleri sözlüğü* olmak üzere iki sözlük hazırlanmıştır. Bu sözlükler, 4'üncü bölümde detayları açıklandığı şekilde *güçlendirilmiş sözlükler* olarak adlandırılan bir yapıya dönüştürülmüştür. Alana özel bu sözlüklerin hazırlanmasında kullanılan yöntemler EK-3'de, temel sözlüklerdeki kelimelerden birleştirilen ve güçlendirilmiş sözlük yapısına dahil edilen kelimeler EK-4'te verilmiştir.

Anahtar kelimelerle oluşturulan ver havuzundaki yemek siparişine ilişkin metinler üzerinde yapılan incelemelerde aşağıda belirtilen hususlar tespit edilmiştir.

- Twitter (X)'da yer alan yemek siparişine ilişkin görüşlerin genelde sipariş öncesine ait duygu ve düşünceleri içerdiği görülmüştür.

- Bu konudaki tweetlerin genel içerikli olduğu; pizza, döner, lahmacun, hamburger, pide gibi herhangi bir hazır yemek ürünü ile ilgili ismin az sayıda zikredildiği, örneğin son bir hafta içinde ürün isminin de yer aldığı tweet metni sayısının ancak 6-7 tweet civarında olduğu tespit edilmiştir. Bu nedenle ürün isminin de yer aldığı tweetlerle oluşturulacak koleksiyonun tweet sayısının düşük kalacağı anlaşılmış, analizin ürün bazında değil sektörel düzeyde yapılmasına karar verilmiştir. Öte yandan koleksiyona dahil olacak tweet sayısını ve müşteri görüşlerindeki çeşitliliği artırabilmek için veri havuzu oluşturulurken *pide, lahmacun, döner, hamburger* gibi ürün ismi içeren kelime grupları da anahtar kelime olarak kullanılmıştır. *Pizza* kelimesi de sorgulamalarda kullanılmış ancak çok fazla sayıda yabancı dilde yazılmış tweet metninin veri havuzuna girmesine neden olduğu görüldüğünden bu kelime anahtar kelime listesinden çıkarılmıştır.

- Yemek siparişi konusunda hizmet veren işletme isimleri kullanılarak yapılan Twitter (X) sorgulamalarının, işletmeler arasında karşılaştırmalı bir çalışma yapılmasına ve dengeli düzeyde bir koleksiyon oluşturulmasına imkan verecek sayıda olmadığı görülmüştür.

- Anahtar kelime olarak; "*yemek sipariş, online yemek siparişi, yemek, sipariş*" kelimeleri üzerinde yapılan sorgulamalarda; (i) *online yemek siparişi* kelimesi ile yapılan aramalarda üç kelimeden oluşan bir kelime grubu ile sorgulama yapıldığı için çok az sayıda tweet elde edilebildiği, (ii) *yemek* ve *sipariş* kelimeleri ile müstakil olarak arama yapıldığında konu kapsamı dışında çok fazla sayıda reklam, yemek tarifleri ve lokantalara ilişkin tweetin veri havuzuna dahil olabildiği, (iii) bu kelimeler içinde yalnızca *yemek sipariş* kelime ikilisinin anahtar kelime olarak kullanılmasının uygun olacağı (*yemek sipariş* kelime çifti ile yapılan sorgulamada *yemek siparişi* kelime çifti ile yapılan sorgulama sonucundan daha fazla sayıda veri elde edilebildiği için "i" harfinin çıkarılması tercih edilmiştir) tespit edilmiştir.

- Yemek siparişi konusunda ayrıca "yemeksepeti sipariş" ifadesinin de yoğun olarak yorum metinlerinde yer aldığı tespit edilmiş, veri havuzundaki yorum çeşitliliğinin artırılması için bu kelime ikilisinin de anahtar kelime olarak kullanılabileceği tespit edilmiştir.

Sonuçta anahtar kelime olarak; *yemek sipariş*, *yemeksepeti sipariş*, *döner sipariş*, *lahmacun sipariş*, *hamburger sipariş*, *pide sipariş* kelime grupları belirlenmiş ve veri havuzu bu kelimelerle yapılan sorgulama ile oluşturulmuştur.

Anahtar kelimelerle sorgulamalar, Twitonomy uygulaması kullanılarak yapılmıştır. Uygulamadaki arama motoru, arama bölümüne yazılan anahtar kelime grubunun tamamını aynı şekli ile kapsayan tweet metnini sorgulama sonucu olarak listelemektedir. Örneğin, *yemek sipariş* kelime grubunun aranması sonucunda, hem *yemek* hem *sipariş* kelimesini aynı anda ve ardışık yazılmış şekli ile içeren tweet metinleri sorgu sonucu olarak gösterilmektedir. Twitter (X) aynı konuda tek seferde çekilebilen tweet sayısını 3200 tweet ile sınırlandırmaktadır ve o konudaki mevcut tweetlerin %1'ini sorgu sonucu olarak vermektedir. Bu kısıtlamalarla Twitonomy uygulaması ve yukarıda belirtilen anahtar kelimeler kullanılarak, yalnızca kamuya açık tweetler çekilmiş ve koleksiyona dahil edilmiştir. Koleksiyondaki tweetlerin paylaşım tarih aralığı, Aralık 2019-Aralık 2020'dır ve 42 255 adet tweet cümlesinden meydana gelmiştir. Ancak bu sayı, analiz kalitesini yükseltmek için çalışma kapsamında hazırlanan *alana özel reklam sözlüğü* ile yapılan ayıklama sonrasında düşmüştür.

3.1.2. Koleksiyonun oluşturulması, tanımlayıcı analiz safhasında kullanılacak TVİTA modeli

Tanımlayıcı analizler çok detaylı önişlemler gerektirmese de veri setinin içeriği, kapsamı, elde ediliş yöntemi, amacı gibi hususlar yapılacak önişlemlerin detaylarını belirleyebilir. Bu çalışmada önişlem olarak yapılan ayıklama işlemleri Çizelge 3.1'de gösterilmiştir. Çizelge 3.1'de gösterilen ayıklama işlemleri her çalışmaya özel olarak değişebilir. Örneğin bu çalışmada veri içinde yer alan haber içerikli tweetlerin müşteri algısına olumlu ya da olumsuz yapacağı tesirin de dikkate alınması gerektiği düşünüldüğünden veri setinden çıkarılmamıştır.

Çizelge 3.1. Veri seti oluşturma ve ayıklama işlemleri

<table>
<tr><td colspan="3">Veri Seti Oluşturma, Veri Ayıklama</td></tr>
<tr><td>Esas İşlem Konusu</td><td>Yapılan İşlemler</td><td>Kullanılan Araçlar / Programlar</td></tr>
<tr><td>Anahtar Kelime Tespiti</td><td>Konuya ilişkin en çok kullanılan anahtar kelimelerin tespiti</td><td>Internet üzerinden müşteri şikâyet ve yorumlarının manuel incelenmesi</td></tr>
<tr><td>Veri Çekme</td><td>Anahtar kelimeler (yemek sipariş, yemeksepeti sipariş, pide sipariş, lahmacun sipariş, döner sipariş, hamburger sipariş) ile sorgulama ve Twitter (X)’dan veri çekme</td><td>Twitonomy, ücretli uygulama</td></tr>
<tr><td rowspan="6">Veri Ayıklama (Veri havuzundan tamamen çıkarılan tweetler)</td><td>İki kelime ve daha az sayıda kelime içeren tweetler</td><td rowspan="6">Excel, Pyhton, Manuel</td></tr>
<tr><td>Türkçe dışındaki dillerde yazılmış tweetler</td></tr>
<tr><td>Boş tweetler</td></tr>
<tr><td>Yalnızca görselden oluşan tweetler</td></tr>
<tr><td>Veri setine birden fazla girmiş aynı tweetler</td></tr>
<tr><td>Reklam içerikli tweetler</td></tr>
</table>

Belirlenen anahtar kelimeler ile koleksiyon oluşturulduktan sonra, aynı tweetlerin ve reklam tweetlerinin ayıklanması şeklinde iki ön eleme işlemi yapılmış, bu işlemlere ilişkin detaylar ileriki maddelerde izah edilmiştir.

- Tüm satır bilgisi aynı olan tweetlerden sadece biri veri havuzunda muhafaza edilmiş, diğerleri koleksiyondan çıkarılmıştır. Koleksiyonun içine aynı tweetin tekrarla girmesinin nedeni, veri çekme işlemi için konuya ilişkin birden fazla kelime/kelime grubunun anahtar kelime olarak belirlenmesi ve veri çekilmesidir. Bu çalışmada koleksiyonun oluşturulmasında, altı kelime çifti kullanılmıştır. Bu kelime çiftlerinin içindeki ortak kelimeler nedeniyle, aynı tweetin birden çok defa (aratma işleminde ifadenin tamamı aynı şekli ile aratılmış olmasına rağmen) koleksiyona girdiği görülmüştür ve bu tweetlerin biri dışındakilerin koleksiyondan çıkarılması gerekir. Burada tekraren koleksiyona girmiş olan tweetlerin koleksiyondan ayıklanması ile kastedilenin, retweetlerin ayıklanması olmadığı hususuna dikkat edilmelidir ki bu çalışmada retweetler, ayrıca analiz edilmiştir. Burada tekraren aynı tweetin koleksiyona girmesi, anahtar kelime olarak birden fazla kelime çifti kullanılarak veri çekildiği için ortaya çıkmış bir durumdur. Anahtar kelime olarak tek bir kelime/kelime grubu ile veri çekme işlemlerinde böyle bir ihtiyaç ortaya çıkmayacaktır. Buradaki ayıklama işlemi, MS Excel’in veri analiz özellikleri kullanılarak yapılmıştır.

Ayıklama işlemi gerçekleştirilirken, tüm satır bilgisinin (Date, Handle, Name, Text, URL, Platforms, Foloowers, Retweets, Favorites) karşılaştırılması neticesinde birebir aynı olan tweetlerin elenmesi şeklinde yapılmıştır, örneğin farklı tarih bilgisine sahip aynı metin içerikli tweetler ayıklanmamıştır. Örneğin bir kişinin tweeti, birçok kişi tarafından yorumsuz olarak RT yapılabilmektedir. Bu şekilde metin bölümü aynı olan çok sayıda tweet ortaya çıkmaktadır. Bu RT'lerin bir kısmı reklam maksadı ile yapılırken, diğer bir kısmı ise müstakil kullanıcılar tarafından yapılan RT'ler şeklindedir. Reklam maksatlı tweetler bir sonraki adımda elenmiştir.

- İkinci adımda, koleksiyon içinde çok sayıda olduğu tespit edilen reklam tweetlerinin çıkarılması sağlanmıştır. Bu işlem MS Excel'in veri analiz özelliği kullanılarak gerçekleştirilmiştir. Reklam tweetlerinin ayıklanması için basit bir eleme usulü seçilmiştir. Veri havuzundan rastgele 325 tane reklam içerikli tweet seçilmiştir. Bu reklam tweetleri incelenerek, metin içinde çoğunlukla kullanılan kelimelerden *alana özel reklam sözlüğü* hazırlanmıştır. Sözlükte yer alan kelimelerin, reklama has nitelikleri ön plana çıkaran kelimeler olmasına dikkat edilmiş, bu kelimelerin öznitelik olarak seçilmesinde sezgisel bir yaklaşım kullanılmıştır. Öznitelik olarak seçilen bu kelime listesi, alana özel reklam sözlüğü olarak kabul edilebilir. EK-5'teki reklam sözlüğü kullanılarak sınıflandırma yapmak maksadıyla; "Reklam sözlüğündeki kelimelerden bir tanesinin tweet cümlesi içinde yer alması durumunda, bu tweet cümlesi reklam tweeti olarak kabul edilecektir" şeklinde kural tanımlaması yapılmış ve reklam içerikli tweetler bu kural çerçevesinde belirlenmiştir. Bu şekildeki ayıklama işleminde, veri havuzundan çıkarılan tüm tweetlerin reklam tweeti olduğu elle yapılan kontrolde teyit edilmiştir. Burada yapılmak istenen, yüksek miktardaki reklam tweetlerinin basit bir ön eleme işlemi ile koleksiyondaki oranının düşürülerek, analiz sonuçlarının yönünü değiştirebilecek etkisini azaltmaktır. Bu şekilde literatüre kelimelerin yanlış, hatalı yazımlarının da dahil edildiği alana özel bir reklam sözlüğü kazandırılmıştır. EK 6'da verilen sözlük, çevrimiçi yemek siparişlerine yönelik müşteri görüşlerinin yer aldığı koleksiyon içindeki reklam içerikli metinlerin tespiti ve bunların ayıklanması maksadıyla yine koleksiyondaki metinler içinden seçilen kelimeler kullanılarak, alana özel reklam jargonu/gündelik kullanımı da dikkate alınarak hazırlanmış bir sözlüktür.

Literatürde reklam içerikli tweetlerin, analiz sonuçlarını etkilemeyeceği kabulü ile çalışmalar yürütülmektedir. Ancak bu çalışma için oluşturulan koleksiyonda reklam içerikli tweetlerin oranının yüksek olduğu ve koleksiyondan ayıklanması gerektiği öngörülmektedir.

Ayrıca yapılacak çalışmalarda, reklam içerikli tweetlerin farklı zaman dilimlerine göre artış/azalış gösterebileceği de dikkate alınarak, analizin etkilenip etkilenmeyeceği hususunda karar verilebilir. Örneğin bu çalışma kapsamında, yiyecek tüketimlerinin değişikliğe uğradığı Ramazan ayında reklam içerikli tweetlerin sayısında artış olduğu tespit edilmiştir. Analizlerde buna benzer durumlar için ayrı değerlendirme yapılması gerektiği açıktır. Öte yandan sosyal ağların kullanımının yaygınlaşması, bu ağlarda reklam yapma maliyetinin düşük olması, bireysel reklamlardan büyük işletmelere kadar birçok reklam içeriğinin sosyal ağlarda dolaşıma girmesine neden olmaktadır. Doğrudan yapılan reklamların yanı sıra, işletmeler çevrimiçi etkileyicileri[9] de kullanarak reklamlar yapmaktadırlar ki bu tür reklamların tespit edilmesi diğerlerine göre daha güç olmaktadır. Bu nedenle reklam içerikli tweetlerin ayıklanmasının sadece başlangıç aşamasında yapılmasının yeterli olmayabileceği, ayrıca veri setlerinden reklam içerikli tweetlerin % 100 olarak ayıklanmasının mümkün olmayacağı öngörülmektedir. Koleksiyonun analizi öncesinde reklam içerikli tweetlerin koleksiyondan ayıklanmaması durumunda, reklamların etkisinin ne olabileceğinin analiz sonuçlarının değerlendirilmesinde dikkate alınması gerekir.

EK 6'daki sözlük kelimelerini ayıklama konusunda performans ölçümü yapılmadığı için, bu çalışmadaki koleksiyona özel nitelik taşıdığı kabul edilebilir. Ancak EK 6'da yer alan kelimeler, başka çalışma alanlarına uyarlanmak üzere tohum kelime listesi olarak kullanılarak sözlük kapasitesi genişletilebilir ve restoran, mutfak gibi yemek sektörünün farklı alanlarına yönelik reklam ayıklama maksatlı sınıflandırma çalışmalarında kullanılabilir.

Koleksiyon üzerinde elle yapılan incelemelerde, reklam tweetlerinin koleksiyondan ayıklanmasında ileriki maddelerde sıralanan beş yöntemin müstakilen ya da birlikte kullanılabileceği ve bunun farklı veri setlerine uygulanabileceği öngörülmektedir. Bu çalışma kapsamında ilk ve son sıradaki yöntemler birlikte kullanılarak ayıklama işlemi yapılmıştır.

[9] Influencer, sosyal medyada aktif, hedef kitlesi geniş ve paylaşımları ile dijital tüketicileri etkileme gücü olan, paylaşım içeriklerinde örtük biçimde markaların reklamını yaparak tüketicilerin tercihini etkileyebilen, paylaşımlara yorum yaparak hedef kitlelerinin dikkatini ürün ya da işletmelere çekebilen sosyal medyadaki kanaat önderleridir. Çok sayıda takipçisi olan makro ve tematik paylaşımları ile daha sınırlı takipçi sayısına sahip mikro influencer olarak iki grup olarak isimlendirilirler.

- Bu çalışmada olduğu gibi alana özel reklam sözlüğü kullanılarak ve sözlükteki kelimelerin bulunduğu metinler reklam kabul edilerek koleksiyondan ayıklanabilir.

- Sosyal ağ hesabının kullanıcısı (Handle) ya da kullanıcı adı (Name) kontrol edilerek, eğer o alanda faaliyet gösteren bir işletme ismi tespit edilirse, bu kullanıcı tarafından paylaşılmış tweetler reklam içerikli kabul edilerek koleksiyondan ayıklanabilir. Ayrıca sektörde faaliyet gösteren işletmelere ait tweetlerin, işletme ile müşterisi arasındaki paylaşımlar olabileceği de dikkate alınarak, bu tweetlerin koleksiyondan ayıklanıp ayıklanmaması konusunda karar verilmesi gerekebilir.

- Sektördeki bir işletmenin paylaşımını, yorumsuz şekilde RT'leyen (metin kısmı "RT@alandakiişletmeninkullanıcıismi" şeklinde başlayan) tweetler, reklam tweeti kabul edilerek koleksiyondan ayıklanabilir.

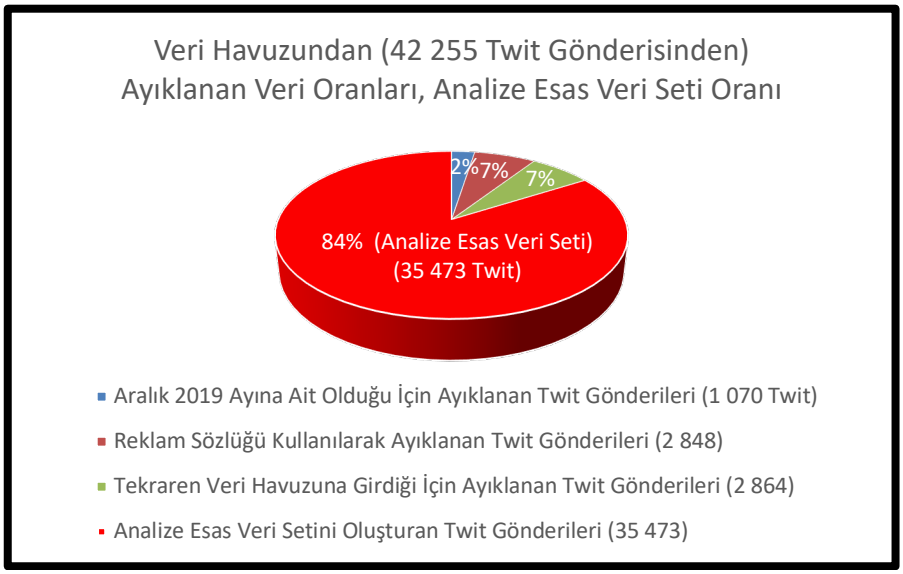
- Tweetin paylaşım yapılan platform bilgisi de reklam ayıklama işleminde bir araç olarak kullanılabilir. Platform bilgisini kullanarak yapılan ayıklama işleminde; otomatik tweet gönderisi yapmayı sağlayan, bireysel ya da kurumsal sosyal ağ hesaplarının yönetilmesini kolaylaştıran yönetsel uygulama ve platformların kullanılıp kullanılmadığı kontrol edilir, bu tür yönetsel araçları kullanan hesapların reklam içerikli tweet olduğu kabul edilerek, koleksiyondan ayıklanabilir. Kullanıcıların günlük tweet paylaşımları arasında reklam tweetleri yayımlamak, bu reklamların zaman içinde kaybolmadan belirli aralıklarla yenilenmesine devam etmek için sosyal ağ kontrolü, düzenlenmesi, takip edilmesi, bu işlemlere hız kazandırılması maksadıyla çeşitli yönetsel araçlar, platformlar, web uygulamaları, ssd'ler gibi farklı araçların kullanımına ihtiyaç duyulmaktadır ve bu araçlar reklam maksadı ile tweet yayımlama işini kolaylaştırmakta, hızlandırmakta ya da düzenli aralıklarla otomatik reklam üretebilmektedir ve bunlar tweet metaverisi olarak platform bilgisinde görünmektedir. Bu çalışmada kullanılan koleksiyondaki tweetlerin platform bilgileri incelendiğinde; "Twitter (X) for Android", "Twitter (X) for iPhone", "Mobile Web (M2)", "Twitter (X) Web App", "IFTTT", "Clarabridge Engage" ve "Instagram" şeklinde bilgilerin yer aldığı görülmüştür. İlk iki sırada yer alan platform bilgisine sahip tweetlerin reklam içerikli olmadıkları, bunun dışındakilerin ise reklam içerikli oldukları tespit edilmiştir. Sonuç olarak otomatik üretilmiş tweet metinlerinin koleksiyondan ayıklanması için platform bilgisi tweet metaverisinden alınabilecek ve reklam içerikli tweetlerin ayıklanmasında kullanılabilecektir.

- Analiz sonuçlarında reklam ayıklama, analiz sonuçlarında yapılan sıralamalarda hangi tweetin kime ait olduğunun tek tek incelenmesi ve tespit edildiğinde reklam içerikli tweetin ayıklanması işlemidir. Bu ayıklama işlemi doğrudan bir ayıklama işlemi olmamakla birlikte, reklam tweetleri analiz başlangıcında koleksiyon içinden ayıklansın ya da ayıklanmasın analiz sonuçları üzerinde dikkat çekici unsurlar sırasıyla ve derinlemesine incelenerek reklam etkisi ile ortaya çıktığı tespit edilenler göz ardı edilerek değerlendirme yapılmalıdır.

Yukarıda sıralanan bu beş yöntem, reklam twetlerinin veri setlerinden ayıklanması maksadıyla bu çalışma kapsamında önerilen yöntemlerdir. Bu yöntemlerin veri setleri için genel geçerliliğinin tespit edilebilmesi için, performans ölçümlerinin de yapılması gerekir. Bu yöntemler, müstakilen ya da birlikte kullanılarak veri setlerinden reklam ayıklama işlemleri yapılabilir.

Veri setleri içinde bu çalışmada da olduğu gibi bazı haber tweetlerinin yer aldığı görülmektedir. Söz konusu haberlerin yemek siparişleri ile ilgili haberler olduğu ve kişilerin görüşlerini olumlu ya da olumsuz olarak etkileme potansiyeline sahip olduğu, koleksiyon içinde söz konusu haberlere ilişkin yorum tweetlerinin yer aldığı ve RT yapılarak sosyal ağ içinde etki yarattığı görülmektedir. Bu nedenle doğrudan müşteriye ait görüşler olmasa da bu tweetlerin ana gündem maddeleri içinde yer alabileceği, işletmelerin bu konuları dikkate almaları gerektiği kabul edilmiş ve koleksiyondan ayıklanmamıştır. Bu tür tweetler, yapılacak çalışma alanına ve çalışmadan elde edilmek istenen sonuca bağlı olarak koleksiyon içinde tutulabilir ya da koleksiyondan çıkarılabilir.

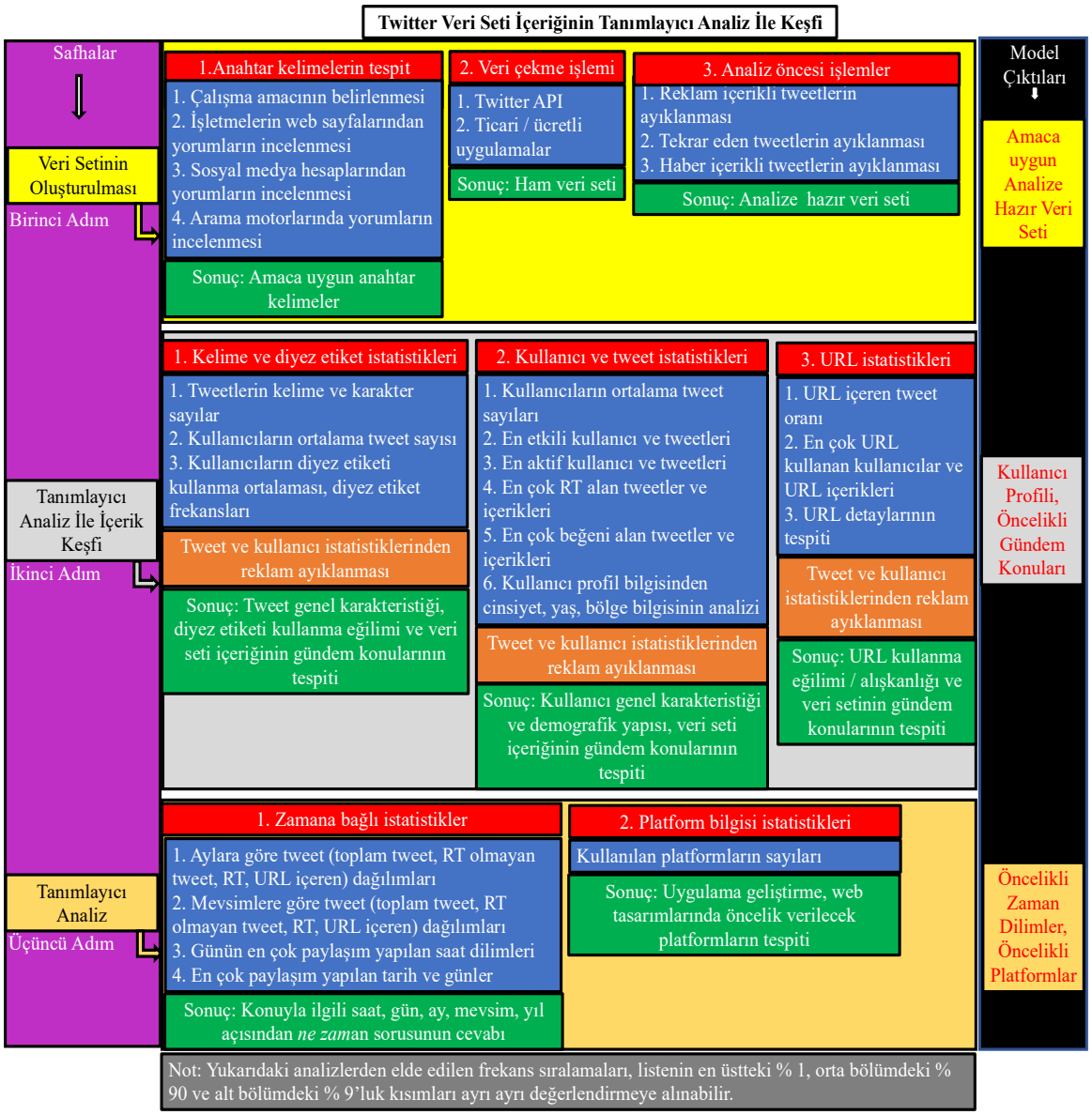
Veri havuzunun oluşturulması için, veri toplama işlemi Aralık 2019-Aralık 2020 tarih aralığında gerçekleştirilmiştir. Bu çalışmada yapılan iki ayıklama yöntemi ile veri havuzundan ayıklanan tweet gönderisi sayıları Şekil 3.1'de gösterilmiştir.



Şekil 3.1. Veri havuzundan ayıklanan veri oranları

Sonuç olarak tanımlayıcı analiz kapsamında analizler, CKB ile ilgili analizler hariç bir takvim yılı içindeki yemek siparişine ilişkin 35 428 tweet üzerinden yapılmıştır. Tweetlerin CKB'ne ilişkin bilgiler uygulama üzerinden sadece harita ve sadece CKB şeklinde liste elde edilebildiğinden bu bilgilerin müstakilen hangi tweetlere ait olduğu tespit edilememiştir. Bu nedenle CKB'deki topluca elde edilen verilerden herhangi bir ayıklama yapılmadan, 42 255 tweet gönderisi ile coğrafi konum analizi yapılmıştır.

Söz konusu koleksiyon, çalışma kapsamında hazırlanan Twitter (X) Verisi İçerik ve Tanımlayıcı Analiz (TVİTA) Modeli aşama ve işlemlerinden geçirilerek elde edilen sonuçlar bölümün devamında sunulmuştur. Hazırlanan model, Şekil 3.2'de gösterilmiştir.



Şekil 3.2. Twitter (X) veri seti içeriğinin tanımlayıcı analiz ile keşfi için önerilen model

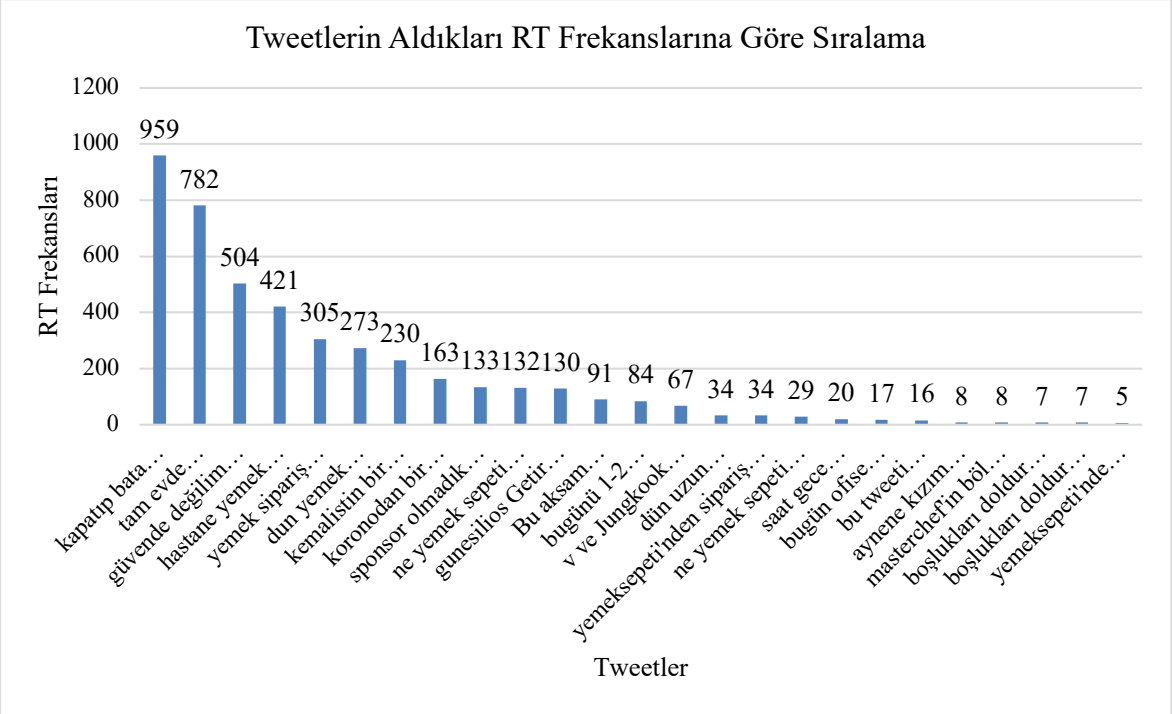
3.2. Tanımlayıcı Analiz

Koleksiyonda toplam tweet sayısı 35 428'dir, bunların 23 507'si orijinal tweet olup koleksiyonun %66'sını oluşturur. Koleksiyondaki toplam müstakil kullanıcı sayısı 2 884'dür. ÇYS konusunda sohbet eden kullanıcıların bir yıl içinde ortalama 1,23 tweet paylaştıkları görülmektedir. Koleksiyondan RT'ler çıkarıldıktan sonra geriye 23 507 orijinal tweet kalmaktadır ki bu sayı kullanıcı sayısına bölündüğünde bir kullanıcının özgün içerikli paylaşım sayısının bir yıl içinde 0,82 tweet olduğu görülmektedir. Buna göre kullanıcıların ÇYS konusunda özgün tweet paylaşımlarının yoğun olmadığı söylenebilir.

3.2.1. Tweet istatistikleri

Koleksiyodaki tweetlerin 2/3'ünün ortamdaki gündemi belirleyen ve yönlendiren konuşmaları içeren tweetler (diğer bir ifade ile sohbet başlatan yeni tweet) olduğu görülmektedir. Bu oran koleksiyonun konu çeşitliliğini ya da konunun farklı yönlerine ilişkin sohbet oranını da göstermektedir. Bu koleksiyonda 1/3 oranında RT olduğu dikkate alınırsa, ortamda etkili olan tweet metinlerinin (etkili kullanıcıların) olduğu söylenebilir. Bir başka ifadeyle koleksiyonda, sohbet başlatan tweetlerden 1/3'ünün RT yapılarak ortamdaki görünürlüğünün diğerlerine göre arttırılmış olduğu ve bunların diğer sohbet başlatan tweet metinlerine göre daha etkili olduğu söylenebilir. Koleksiyonda en çok RT'lenen tweetlerin frekans grafikleri ve metinlerinden (diğer tweetlere göre daha etkili olan, gündeme etki eden konuları içeren) ilk 25 tweeti Şekil 3.3 ve Şekil 3.4'te verilmiştir.

Koleksiyonun %66'sını tweetler, %34'ünü RT'ler meydana getirmektedir. Tweet sayısı koleksiyonun konu çeşitliliğinin yüksek; RT sayısı ise koleksiyonda tweetlerin yayılmasını sağlayan etkili kullanıcılar ve daha çok kullanıcıya ulaşarak görünürlüğü yükseltilmiş tweetler olduğunu gösterir. Tweetlerin RT frekansları, koleksiyon içindeki etkili ve gündem belirleyen mesajların görülmesini sağlar.



Şekil 3.3. Tweetlerin aldıkları retweet frekans sıralaması

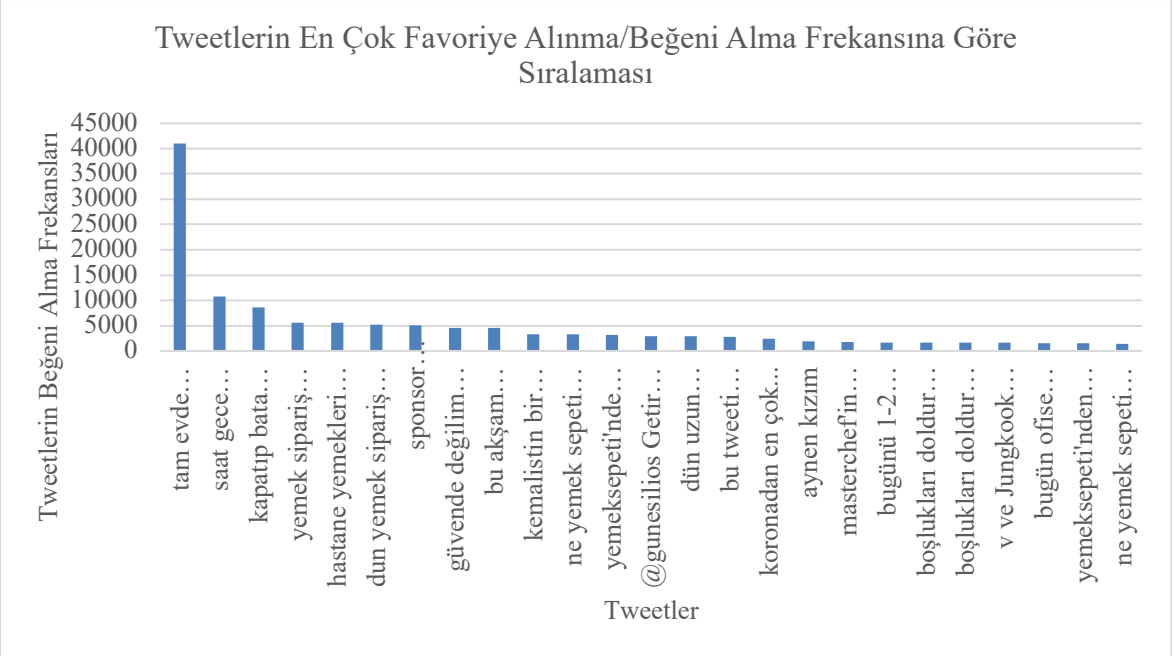
Sıra Nu.	Twitin RT'lenme Sayısı	**RT'lenme Açısından İlk 25 Twitin Metni**
1.	959	kapatıp batacağımıza açıp batalım dedik😊#gelal #eveservis devam..@yemeksepeti ile çalışıyoruz,mahalleye komşulara servisi 3 şefimiz/ben bisikletle yapıyoruz..#bizdeböyle sipariş için 02163566010@delykalamis#trattoria #pizzeria #segafredo #kalamış https://t.co/prdx2bI1Hg
2.	782	tam evde yemek yok ne sipariş versem diye kara kara düşünürken üst komşu ölmüş sağolsun https://t.co/UtNMrAqH5O
3.	504	Güvende değilim. Evimde güvende değilim, işte değilim, okulda değilim, sokakta değilim, yurtta değilim. Her an öldürebilirler beni, ve kendi kendimi öldürdüğümü iddia edebilirler. Birinin bana musallat olması için yemek sipariş etmem bile yeterli olabilir. Güvende değilim.
4.	421	Hastane yemekleri yenilmiyor her nöbet dışarıdan sipariş vermek zorunda kalıyoruz. Eğer bu twite 1000 fav gelirse cumartesi nöbete yemek,pasta börek gelecek. Ben diyeceğimi dedim gerisi siz de zalımlar:) https://t.co/xo9aD76aBn
5.	305	- Yemek siparis ediyorumdur -Me: Siz bi sey istiyo musunuz ?Arkadaslarım: Yok biz yemeyiz ya- Yemek gelince arkadaslarım - https://t.co/SiGbp4u80p
6.	273	Dun yemek siparis eden arkadasimi telefondan taciz edip ciplak fotograf yollamis bi kurye ve daha once baskasina yapmadiysa onun oldugunu anlayacagi icin ve evinin adresini bildigi icin sikayet veya ifsa etmeye korkuyoruz, evet turkiyede kadin olarak bu kadar guvende hissediyoruz
7.	230	Kemalistin biriyle DM'den atıştık. Beni çok sinirlendirdi. Ver lan adresini dedim. Hemen verdi. 1938 tane lahmacun sipariş ettim adresine. Biraz da o sinirlensin şimdi :)
8.	163	Koronadan en çok etkilenen sektörlerden restoranlara, TOBB ve @yemeksepeti işbirliğiyle orijinal bir kampanya...Restoranlar iki ay sonra vereceğiniz sipariş için indirimli çek satarak nakit akışı problemini çözüyor. Restoran kapanırsa da zararı biz üstleniyoruz. Emeksepeti...
9.	133	Sponsor olmadıkları için Burger King yerine Terra Pizza'dan yemek sipariş ettim. https://t.co/jkFKVlzgxd
10.	132	ne yemek sepeti ne de sipariş veren insanlar seni hak etmiyor kral https://t.co/OBXlbR44XP
11.	130	@gunesilios "Getir" i bilmiyorum fakat yemek sepeti'nde sipariş notu kısmı var. Orqya "kurye arkadaş sakın acele etmesin. Önce can sağlığı "yazıyorum. Size de tavsiye ederim. Gelen gençler minnetle teşekkür ediyor. Bir çeşit destektir onlara bu.
12.	91	Bu akşam @DominosTR'den yemek sipariş ettim ardından siparişi getiren kişi arayarak adres teyit etti, sipariş geldiğinde yenecek şekilde değil dağılmıştı, yemek sepetini arayarak iade oluşturdum. Tekrardan siparişi getiren kişi gelip aldı.
13.	84	bugünü 1-2 twitle anlatmak imkansız,@yemeksepeti nden sipariş verip göndermeyin ihtiyacı olana verin diyenler,İstanbulun her köşesinden gelenler(#gelal)başka şehirlerden 'iban' isteyip 'askıda pizza' önerenler. ve şunu çok net anladım medyaya hemen dönmem lazım! teşekkürler🙏❤️ twitter.com/bilgingokberk/…
14.	67	[200505] V ve Jungkook VLive yayınından JK: Karanfil buketlerine geçmeden önce, son zamanlarda ne yapıyordun anlatsana? V: Çilek toplayıp şarkı yazıyordum. JK: Doğru şu sıralar müzik yapmayla epey meşgulsün. V: Yemek falan yiyorum. JK: Sen mi pişiriyorsun? V: Hayır, sipariş. https://t.co/FeGPPdgOQk
15.	34	Dün uzun zamandır ilk kez dışarıdan yemek siparişi verdik, sipariş verdik diye kızıyolar mıdır yaa falan diye düşünürken içinden bu not çıktı💛 https://t.co/VSFDd56UgI
16.	34	Yemeksepeti'nden sipariş verirken manitayla böyle oluyoruz joker moker çıkar belki diye ben de bakiyorum twitter.com/MovieGrafMG/st…
17.	29	ne yemek sepeti ne de sipariş veren insanlar seni hak etmiyor https://t.co/FNftZLyNaB
18.	20	Saat gece 01:03 ve 3 tane acili lahmacun siparis verdim pisman degilim
19.	17	Bugün ofise yemek sipariş ederken "talep ediyorum" demişim. TALEP EDİYORUM.
20.	16	bu tweeti begenen 4 kisiye yemeksepetinden hamburger menü sipariş ediyorum aksam 21.00'da sonucu acıklarım.. herkese bol şans.. https://t.co/2cGaiCSrbp
21.	8	aynen kızım yemek sepeti siparis veriyoz falan vayt caklıt moka iciyoz aynen https://t.co/hLAlMUAncz
22.	8	masterchef'in bölümlerini takip edemeyenlere yemeksepeti sipariş geçmişimi gönderebilirim. ne yapıyolarsa tak diye sipariş ediyorum. aha bugün köftedeyiz https://t.co/JigSIbHGyh
23.	7	Boşlukları dolduralım:🌯İlk sipariş ettiğin tatlı/ yemek:🍔En çok sipariş ettiğin tatlı/ yemek:🍩Son sipariş ettiğin tatlı/ yemek:
24.	7	Boşlukları dolduralım:🌯İlk sipariş ettiğin tatlı/ yemek:🍔En çok sipariş ettiğin tatlı/ yemek:🍩Son sipariş ettiğin tatlı/ yemek:
25.	5	yemeksepeti'nde semtimin muhtarı olarak (şaka değil muhtarım....) karantinada bile restoranlardan 24 saat sipariş verebilme olayı aşırı iyi. dün gece tulumba tatlısı söyledik bu gece için planımız ekler. vanilyalı. kalmamışsa çikolatalı da olur. neli varsa olur

Şekil 3.4. RT'lenme açısından etkin ilk 25 tweet metni

Şekil 3.4'teki toplamda 4459 RT alan 25 tweet incelendiğinde şu bilgiler tespit edilebilir: (i) Birinci ve sekizinci sıradaki tweetler ticari kuruluşların reklam içerikli tweetleri oldukları için işletme tarafından dikkate alınmayacaktır. (ii) 14'üncü sıradaki tweetin bir müzik grubuna ait canlı yayından alıntılar içerdiği, doğrudan yemek siparişi ile ilgili bir görüş içermediği görülmektedir. Bu tweetler ancak söz konusu müzik grubunun müşteri kitlesinin ilgini çekmesi açısından reklam maksatlı kullanılmak üzere değerlendirilebilir. (iii) 23'üncü ve 24'üncü sıradaki tweetlerin müşteri görüşlerini toplamaya ya da müşteri ilgisini çekmeye yönelik işletmeler tarafından hazırlanmış bir anket olduğu görülmektedir. Anket sorusu içeren bu tweet metni de esas müşteri gündemini yansıtmadığı için bir değeri yoktur. Bu şekilde listede yapılan ayıklamalardan sonra, bir yıllık dönem içinde işletmeler açısından bu alanda dikkate alınacak, işletme politikalarının geliştirilmesine yön verecek esas konuların; 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 25 numaralı tweetlerde yer aldığı görülmektedir. Ayrıca 14 numaralı tweetin de müşteri ilgisini çekebilme potansiyeline sahip olduğu ve işletme politikaları ve mali kapasitesi ile uyuştuğu takdirde reklam aracı olarak kullanılabileceği değerlendirilebilir.

Bu değerlendirme sonucunda koleksiyonun önemli gündem konularının; yemek siparişi verme konusunda yaşanan kararsızlık, yemek siparişi ile ilişkili olarak ürün ve hizmete duyulan güven, yemek siparişi veren kadınların, işletme çalışanlarına karşı duyduğu güvensizlik nedeniyle ortaya çıkan müşterilerin kişisel güvenlik endişesi, hastanelerdeki hastane yemeklerinin beğenilmemesi, yemek siparişi sonrasında yemeklerin arkadaşlarla paylaşılması, kuryelere emniyetli ve güvenli çalışma koşulları sağlanması, ürün sunumunun beklenen şekilde gerçekleşmemesi, yemek siparişi üzerinden yardımlaşma, çevrimiçi yemek işletmelerinin indirimli/joker uygulamaları, gece saatlerinde verilen yemek siparişi ve pişmanlık duygusu, yemek siparişi üzerindeki bir televizyon yemek programının etkisi ve yemek siparişi ile karantina (COVID-19 nedeniyle) ilişkisi olduğu söylenebilir.

Koleksiyondaki tweetlerin favoriye eklenme/beğeni (*like*) alma sayıları tespit edilerek yapılan tweetler ve frekans sıralaması Şekil 3.5 ve Şekil 3.6'da verilmiştir.



Şekil 3.5. Favoriye eklenme/beğeni alma frekans sıralaması

Şekil 3.6 incelendiğinde; 3'üncü, 16'ncı, 20'nci, 21'inci ve 22'nci tweetlerin müşteri görüşünü yansıtmaması, reklam içerikli olması ve anket uygulaması olması nedeniyle değerlendirme dışında tutulacaktır. Geri kalan tweetler incelendiğinde koleksiyonda müşteriler arasındaki esas gündemi oluşturan konuların, yemek siparişi konusundaki kararsızlık, gece yarısı yemek sipariş verme konusunda pişmanlık, sipariş edilen yemeğin arkadaşlarla paylaşılması, hastanelerdeki yemeklerin beğenilmemesi, yemek siparişi ve güven konusu arasındaki ilişki, yemek konusundaki bir televizyon programının etkisi ile sipariş verme, yemek siparişi ile yardımlaşma ve ÇYS ile ilgili sitelerin indirim/joker uygulamaları gibi konuların olduğu görülmektedir. Şekil 3.3 ve Şekil 3.5 karşılaştırıldığında sıralamaları farklı olmakla birlikte elde edilen gündem konularının benzer olduğu görülmektedir. Bu konular sektörde yer alan işletmelerin, ürün ve hizmetin iyileştirilmesine yönelik hangi konulara yönelmeleri gerektiğini göstermektedir.

Sıra Nu.	Favoriye Eklenme/Beğeni Alma Sayıları	Favoriye Alınma Sayısı Açısından İlk 25 Twitin Metni
1.	40976	tam evde yemek yok ne sipariş versem diye kara kara düşünürken üst komşu ölmüş sağolsun https://t.co/UtNMr
2.	10772	Saat gece 01:03 ve 3 tane acili lahmacun siparis verdim pisman degilim
3.	8637	kapatıp batacağımıza açıp batalım dedik😊#gelal #eveservis devam.. @yemeksepeti ile çalışıyoruz,mahalleye komşulara servisi 3 şefimiz/ben bisikletle yapıyoruz..#bizdeböyle sipariş için 02163566010@delykalamis#trattoria #pizzeria #segafredo #kalamış https://t.co/prdx2bI1Hg
4.	5613	Yemek siparis ediyorumdur -Me: Siz bi sey istiyo musunuz?Arkadaslarım: Yok biz yemeyiz ya -Yemek gelince arkadaşlarım - https://t.co/SiGbp4u80p
5.	5612	Hastane yemekleri yenilmiyor her nöbet dışarıdan sipariş vermek zorunda kalıyoruz. Eğer bu twite 1000 fav gelirse cumartesi nöbete yemek,pasta börek gelecek. Ben diyeceğimi dedim gerisi siz de zalımlar:) https://t.co/xo9aD76aBn
6.	5155	Dun yemek siparis eden arkadasimi telefondan taciz edip ciplak fotograf yollamis bi kurye ve daha once baskasina yapmadiysa onun oldugunu anlayacagi icin ve evinin adresini bildigi icin sikayet veya ifsa etmeye korkuyoruz, evet turkiyede kadin olarak bu kadar guvende hissediyoruz
7.	5128	Sponsor olmadıkları için Burger King yerine Terra Pizza'dan yemek sipariş ettim. https://t.co/jkFKVIzgxd
8.	4583	Güvende değilim. Evimde güvende değilim, işte değilim, okulda değilim, sokakta değilim, yurtta değilim. Her an öldürebilirler beni, ve kendi kendimi öldürdüğümü iddia edebilirler. Birinin bana musallat olması için yemek sipariş etmem bile yeterli olabilir. Güvende değilim.
9.	4575	Bu akşam @DominosTR'den yemek sipariş ettim ardından siparişi getiren kişi arayarak adres teyit etti, sipariş geldiğinde yenecek şekilde değil dağılmıştı, yemek sepetini arayarak iade oluşturdum. Tekrardan siparişi getiren kişi gelip aldı.
10.	3330	Kemalistin biriyle DM'den atıştık. Beni çok sinirlendirdi. Ver lan adresini dedim. Hemen verdi. 1938 tane lahmacun sipariş ettim adresine. Biraz da o sinirlensin şimdi :)
11.	3296	ne yemek sepeti ne de sipariş veren insanlar seni hak etmiyor kral https://t.co/OBXlbR44XP
12.	3162	yemeksepeti'nde semtimin muhtarı olarak (şaka değil muhtarım....) karantinada bile restoranlardan 24 saat sipariş verebilme olayı aşırı iyi. dün gece tulumba tatlısı söyledik bu gece için planımız ekler. vanilyalı. kalmamışsa çikolatalı da olur. neli varsa olur
13.	2966	@gunesilios "Getir" i bilmiyorum fakat yemek sepeti'nde sipariş notu kısmı var. Orqya "kurye arkadaş sakın acele etmesin. Önce can sağlığı "yazıyorum. Size de tavsiye ederim. Gelen gençler minnetle teşekkür ediyor. Bir çeşit destektir onlara bu.
14.	2886	Dün uzun zamandır ilk kez dışarıdan yemek siparişi verdik, sipariş verdik diye kızıyolar mıdır yaa falan diye düşünürken içinden bu not çıktı💛 https://t.co/VSFDd56UgI
15.	2728	bu tweeti begenen 4 kisiye yemeksepetinden hamburger menü sipariş ediyorum aksam 21.00'da sonucu acıklarım.. herkese bol şans.. https://t.co/2cGaiCSrbp
16.	2372	Koronadan en çok etkilenen sektörlerden restoranlara, TOBB ve @yemeksepeti işbirliğiyle orijinal bir kampanya...Restoranlar iki ay sonra vereceğiniz sipariş için indirimli çek satarak nakit akışı problemini çözüyor. Restoran kapanırsa da zararı biz üstleniyoruz. Emeksepeti...
17.	1956	aynen kızım yemek sepeti siparis veriyoz falan vayt caklıt moka iciyoz aynen https://t.co/hLAlMUAncz
18.	1734	masterchef'in bölümlerini takip edemeyenlere yemeksepeti sipariş geçmişimi gönderebilirim. ne yapıyolarsa tak diye sipariş ediyorum. aha bugün köftedeyiz https://t.co/JigSIbHGyh
19.	1656	bugünü 1-2 twitle anlatmak imkansız,@yemeksepeti nden sipariş verip göndermeyin ihtiyacı olana verin diyenler,İstanbulun her köşesinden gelenler(#gelal)başka şehirlerden 'iban' isteyip 'askıda pizza' önerenler.. ve şunu çok net anladım medyaya hemen dönmem lazım! teşekkürler🙏❤️ twitter.com/bilgingokberk/…
20.	1629	Boşlukları dolduralım: 🌯İlk sipariş ettiğin tatlı/ yemek: 🍔En çok sipariş ettiğin tatlı/ yemek:🍩Son sipariş ettiğin tatlı/ yemek:
21.	1628	Boşlukları dolduralım: 🌯İlk sipariş ettiğin tatlı/ yemek: 🍔En çok sipariş ettiğin tatlı/ yemek:🍩Son sipariş ettiğin tatlı/ yemek:
22.	1626	[200505] V ve Jungkook VLive yayınından JK: Karanfil buketlerine geçmeden önce, son zamanlarda ne yapıyordun anlatsana? V: Çilek toplayıp şarkı yazıyordum. JK: Doğru şu sıralar müzik yapmayla epey meşgulsün. V: Yemek falan yiyorum. JK: Sen mi pişiriyorsun? V: Hayır, sipariş. https://t.co/FeGPPdgOQk
23.	1525	Bugün ofise yemek sipariş ederken "talep ediyorum" demişim. TALEP EDİYORUM.
24.	1508	Yemeksepeti'nden sipariş verirken manitayla böyle oluyoruz joker moker çıkar belki diye ben de bakiyorum twitter.com/MovieGrafMG/st…
25.	1455	ne yemek sepeti ne de sipariş veren insanlar seni hak etmiyor https://t.co/FNftZLyNaB

Şekil 3.6. Favoriye alınma/beğeni açısından etkin ilk 25 tweet metni

3.2.2. Tweetlerdeki kelime ve diyez etiket istatistikleri

Koleksiyondaki 35 428 tweet, toplam 28 824 kullanıcı tarafından paylaşılmıştır. ÇYS konusunda sohbet eden kullanıcılar, bir yıl içinde ortalama 1,23 tweet gönderisi paylaşmıştır. Toplam tweetlerin 32 128 adetinde diyez etiketi kullanılmadığı tespit edilmiştir. Buna göre koleksiyonun %91'i diyez etiketi içermemekte, %9'unda diyez etiketi kullandığı

görülmektedir. Koleksiyondaki diyez etiketleri dahil toplam kelime sayısı, 698 051 kelimedir. Bunlardan toplam 10 690 kelime diyez etiketi olarak kullanılmıştır. Bu durumda toplam kelimenin %2'si diyez etiketidir.

Koleksiyondaki 10 690 diyez etiketi, aynı köke sahip olan, çekim ekleri kullanılarak yazılanlar, hatalı ve eksik yazılmış şekilleri, yabancı dilde yazılmış olan kelimelerin her biri ayrı kelime olarak sayılmıştır. Söz konusu diyez etiketlerinden (yabancı dilde yazılmış olanlar dışında) genel olarak mantıksal, anlamsal ve aynı köke sahip olma gibi birbiri ile ilişkilendirilebilecek olanlar aynı grup içine dahil edilerek 1 537 gruba ayrılmıştır. Koleksiyondaki toplam kullanıcı sayısının 28 824 olduğu dikkate alınırsa, yaklaşık 2,7 kullanıcıya ancak bir diyez etiketi düşmektedir ki buna göre diyez etiketinin kullanım oranının çok düşük olduğu söylenebilir.

Koleksiyondaki diyez etiketleri listesi gruplandırılmış olarak, koleksiyon içindeki frekansları tespit edilmiştir. Koleksiyonda en yüksek frekansa sahip ilk 50 diyez etiketinin frekansı Çizelge 3.2'de listelenmiştir.

Çizelge 3.2. Tüm kullanıcıların tweetlerindeki diyez etiket frekansları

Diyez Etiket Frekansları			
Diyez Etiketi	Frekansı	Diyez Etiketi	Frekansı
#gelal	800	#et/etdöner	77
#eveservis/evesipariş	793	#restoran	73
#YemekSepeti	672	#burger	70
#siparis	547	#ankara	69
#yemeksiparisi/yiyecek	408	#kurye	65
#paketservis	266	#lahmacun	62
#evdehayatvar/EvdeKal	205	#getir	59
#corana	191	#online	59
#lezzet	172	#samsun	58
#tavuk	133	#food	56
#masterchecf	109	#HayatEveSığar	55
#YökErkenMezuniyeteDurDe	107	#kebap	54
#istanbul	100	#dürüm	53
#Pizza	99	#canik	52
#pide	92	#hamburger	52
#saglik	84	#ilkadım	52

Çizelge 3.2'de verilen kelimelerin kelime bulutu olarak gösterimi Şekil 3.7'de verilmiştir.



Şekil 3.7. Tüm kullanıcıların kullandıkları diyez etiket frekanslarının kelime bulutu

Daha önce belirtildiği gibi reklam içerikli tweetlerin analiz sonuçlarında yanıltıcı etkiye sahip olabileceği dikkate alınarak reklam içerikli tweetlerde işletmeler tarafından kullanıldığı tespit edilen diyez etiketleri, frekans sıralamasından çıkarılmış, yeni oluşan frekans sıralaması Çizelge 3.3'te, söz konusu sıralama için oluşturulan kelime bulutu Şekil 3.8'de verilmiştir. Şekil 3.7 ve Şekil 3.8'de öne çıkan konuların çoğunlukla birbirinden farklı olduğu, Şekil 3.7'de daha çok yemek sipariş hizmetine yönelik kelimelerin öne çıktığı, Şekil 3.8'de potansiyel müşteri gündeminde bir televizyon programının öne çıktığı görülmektedir.

Çizelge 3.3. Yalnızca potansiyel müşteriler tarafından kullanılan diyez etiket frekansları

Diyez Etiket Frekansları			
Diyez Etiketi	Frekansı	Diyez Etiketi	Frekansı
#masterchef	109	#ankara	69
#YökErkenMezuniyeteDurDe	107	#kurye	65
#istanbul	100	#lahmacun	62
#saglik	84	#getir	59
#et/etdöner	77	#online	59
#restoran	73	#HayatEveSığar	55
#burger	70	#kebap	52



Şekil 3.8. Yalnızca potansiyel müşterilerin kullandıkları diyez etiket frekanslarının kelime bulutu

Koleksiyondan alana özel reklam kelimeleri sözlüğü kullanılarak otomatik seçimle reklam içerikli tweetler ayıklanmış olmasına rağmen, ayıklanamamış reklam içerikli tweetlerde kullanılan diyez etiket frekanslarının tüm koleksiyonda kullanılan diyet etiketlerine baskın olduğu görülmektedir. Reklam içerikli tweetlerde çok sayıda diyez etiketi kullanılırken, ürün ve hizmet konusunda görüş yazan potansiyel müşteri konumundaki kullanıcıların tweetlerinde diyez etiket kullanımının düşük olduğu görülmektedir. Bu nedenle tanımlayıcı analiz safhasında diyez etiketlerine dayalı tweet koleksiyonun içeriğine yönelik çıkarımlar yapmak doğru sonuçlar vermeyebilir. Bu nedenle sıralamada yer alan diyez etiketlerinden hangilerinin gerçek kullanıcıya ait olduğunun tespit edilmesi gerekir. Örneğin Şekil 3.1'deki diyez etiket frekansına göre sıralamada *gelal* kelimesinin ilk sırada yer aldığı görülmektedir. Ancak frekansı 800 olarak gözüken bu kelimenin, tekrar ettiği tweetlerden biri dışında diğerlerinin tamamının reklam tweetleri içindeki tekrarlar olduğu görülmektedir. Ancak Şekil 3.1'de 11'inci sıradaki televizyon program isminin frekansının ise 109 olduğu, bunlardan sadece bir tekrarın reklam tweeti içinde kullanıldığı, diğerlerinin (108 tekrarın) ise potansiyel müşteri görüşlerinin olduğu tweet metinleri içinde yer aldığı görülmektedir. Bu duruma göre müşteri gündemini yansıtması açısından, frekans değeri olarak 13'üncü sıradaki diyez etiketinin 1'nci sırada yer alan diyez etiketinden daha öncelikli olduğu söylenebilir.

Koleksiyondaki tweetlerin yalnızca %9'unda diyez etiketi kullandığı, yaklaşık üç kullanıcıya ancak bir diyez etiketi düştüğü, diyez etiketlerinin büyük çoğunluğunun reklam içerikli olduğu dikkate alındığında; (i) Twitter (X) kullanıcılarının bu sektöre ilişkin

tweetlerinde diyez etiketi kullanmadıkları ya da çok az sayıda kullanıcının diyez etiketi kullanma eğilimi olduğu, (ii) ÇYS alanında diyez etiketi kullanarak yapılacak sorgulamalarda / veri çekme işlemlerinde yeterli sayıda ve istenen veriye ulaşılamayacağı, dolayısı ile sorgulamaların anahtar kelime[10] kullanılarak yapılmasının daha uygun olacağı, (iii) koleksiyon gündem konularının sadece diyez etiketi frekansları ile doğru öncelikler verilerek belirlenemeyeceği görülmektedir.

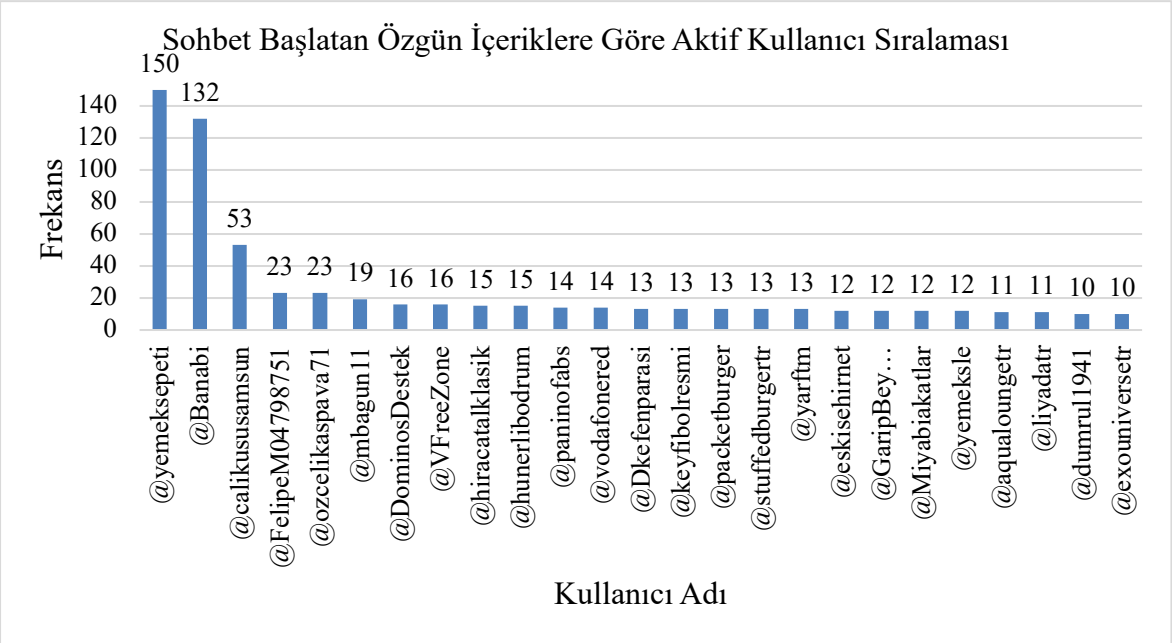
3.2.3. Tweetlerdeki kullanıcı istatistikleri

Twitter (X) kullanıcıları ile ilgili faydalı bilgilere ulaşmak için en aktif, en görünür kullanıcılar ve kullanıcı gruplarına yönelik analizler yapılır [3]. Bir yıl içinde bir kullanıcı başına ortalama 1,23 tweet düştüğü dikkate alındığında, bunun kullanıcıların tweet paylaşım yoğunluğu açısından yüksek bir paylaşım olmadığı söylenebilir.

Özgün içerik paylaşımı yapan (sohbet başlatan/orijinal) tweet sayılarına göre belirlenen *en aktif kullanıcılarından* ilk 25'i, Şekil 3.9'da gösterilmiştir. Şekil 3.9 incelendiğinde, ilk bakışta kullanıcı isimlerinden 1'inci, 2'nci, 5'inci, 7'nci, 8'inci, 12'nci, 15'nci, 16'ncı ve 21'inci sıradaki kullanıcıların ticari kuruluş, işletme olduğu anlaşılmaktadır ve doğrudan değerlendirme dışında tutulmalıdır. Geri kalan 16 kullanıcının aktif potansiyel müşteri olabilecekleri dikkate alınmalı ve işletmeler söz konusu 16 kullanıcıyı paylaştıkları tweetler bazında incelemelidir. Bu inceleme sonucunda işletmeler; hangi kullanıcıların hedef olarak seçilebileceği, mali kapasite çerçevesinde işletmenin hangi kullanıcılara odaklanması gerektiği konusunda karar vermelidir. Şekil 3.9'deki potansiyel 16 kullanıcı incelendiğinde üçüncü, dokuzuncu, 10'uncu, 11'inci, 18'inci, 20'nci, 22'nci, 23'üncü ve 24'üncü sıradaki

[10] Anahtar Kelime (Keyword): İçerikle ilgili esas bilgiyi veren, içeriğin ne olduğunu ifade eden kelimelerdir. Anahtar kelime yazının başlığı, yazının metin kısmı yani içerik bölümünde ve meta/üst etiketlerde (internet adresinde yer alan ve arama motorlarına sağlanan bilgidir ve İngilizcesi "meta tag" olarak ifade edilir) bulunur. Uzun süreli ve kalıcı kullanıma sahiptir [369].
Etiket (Tags): Aynı anlamdaki farklı kullanıma sahip kelimeleri birbiri ile ilişkilendirerek bir içeriğin/metnin kategorize edilmesini sağlayan, böylece web sayfasına trafiğin çekilmesini ve kullanıcının yaptığı aramada doğru sonuçların önüne getirilmesini sağlayan bir tanımlamadır. Web sayfasında işleve sahip olan etiket, örneğin "bisiklet kaskı" hakkındaki bilgiyi, ilgili kategoriler olarak "bisiklet güvenliği" ve "kask tasarımı" kategorilerine bağlar [69].
Diyez Etiketi (#Hashtag): Etiketin işlevine benzemekle birlikte, ilave olarak anlık konular, gündemdeki içeriklerle ilişki kurulmasını sağlar. Aynı diyez etiketini kullanarak paylaşım yapan kullanıcıların birbirlerine bağlanmasını sağlar. Diyez etiketi, konunun daha daraltılmış hali, ana konunun alt kategorisi olarak kabul edilebilir ve anahtar kelime ve etikete göre daha kısa ömürlüdür ve birkaç haftada diyez etiketli konu kullanım dışı kalabilir [369]. Türkçe'de "hashtag" karşılığı olarak "başlık etiketi", "etiket kelimesi", "haştag", "tıklanabilir etiket", "kare etiketi" ve "diyez etiketi" şeklindeki ifadelerin kullanıldığı görülmektedir.

kullanıcıların da reklam içerikli tweetler paylaştıkları ve bu kullanıcıların sektördeki işletmeler olduğu anlaşılmaktadır. Sonuç olarak en aktif 25 kullanıcıdan sadece dördüncü, altıncı, 13'üncü, 14'üncü, 17'nci, 19'uncu tweetlerin müşteri görüşü içeren tweetler olduğu, 25'inci sıradaki kullanıcının ise bu çalışmanın konu kapsamı dışında içeriğe sahip olduğu görülmektedir, bu nedenle sadece altı kullanıcının işletmeler tarafından dikkate alınması gerekir.

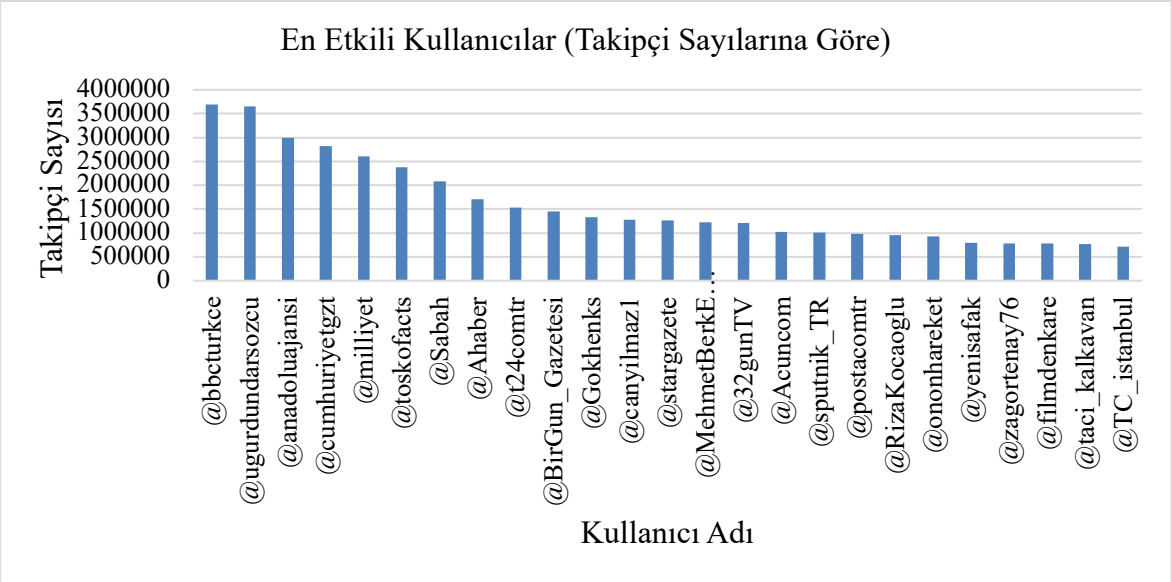


Şekil 3.9. Özgün içerikli tweet frekans sıralaması[11]

Koleksiyondan reklam twetlerinin (% 7 oranında) ayıklanmış olmasına rağmen Şekil 3.9'daki çoğunlukla ticari kullanıcı (koleksiyondan ayıklanamayan az sayıdaki reklam maksatlı Twitter (X) kullanıcıları) isimlerinin, üst sıralara çıkabildiği görülmektedir. Değerlendirme için altı kullanıcının görüşleri incelendiğinde ana gündem konularının; ürün teslim edilmediği halde teslim edildi mesajı gönderilmesi, ihtiyaç sahiplerine yemek gönderilerek yardım yapılabilmesi (askıda yemek gibi) isteği, ÇYS'nin restorana ulaşmaması, kuryelerin çalışma koşulları ve restorana hizmet veren kuryelerin restorana gönderilmesinde gecikme yaşanması, yoğun yemek siparişi veren bir kullanıcının yemek sipariş konusundaki duyguları, yüksek fiyat artışı ve hizmetin kötü olduğu yönünde şikayetleri olduğu görülmektedir.

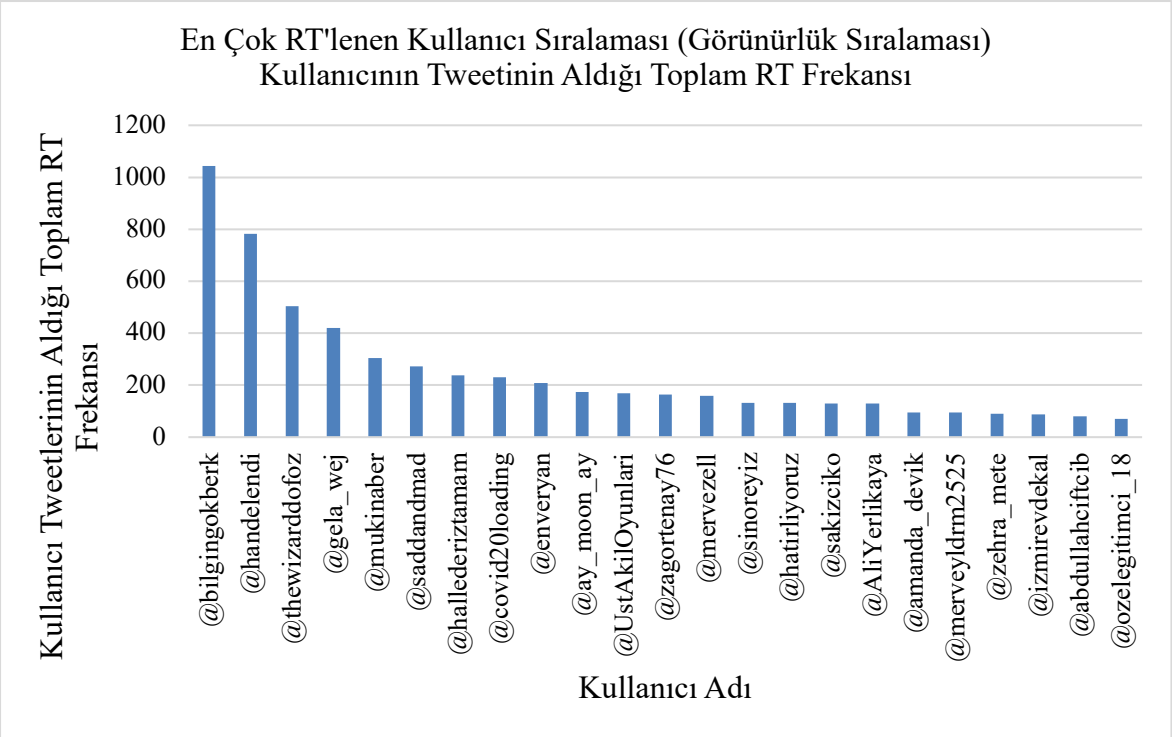
[11] Yemeksepeti isimli kullanıcının frekansı 2006'dır ancak görsel olarak grafik gösterimde diğer frekansların sütun görüntülerinin düzgün olması için frekansı 150 olarak gösterilmiştir.

Koleksiyondaki kullanıcıların takipçi sayılarına göre *en etkili kullanıcılardan* ilk 25 kullanıcı Şekil 3.10'da sıralanmıştır. Şekil 3.10 incelendiğinde en etkili 25 Twitter (X) kullanıcı isminin 15'inin haber kanal ve programlarına, dördünün ticari kuruluş ve resmi kuruma ait olduğu görülmektedir. Bu tür Twitter (X) kullanıcılarının takipçi sayıları fazla olduğundan, en etkili kullanıcılar olarak bu grafikte yer almaları doğaldır. Koleksiyon içinde analiz konusu olan ÇYS'ne ilişkin herhangi bir haberin yer alması böyle bir sonucu ortaya çıkarabilir. Olumlu ya da olumsuz bir haberin basında yer alması, sektörün kamuoyundaki algılamasına doğrudan etki eder. Detaylı incelendiğinde, haber kanallarında ÇYS ve bunun koronavirüs bulaştırmasına etkisi, motosikletli bir kurye kazası şeklinde geçen haberler nedeniyle bu haber kaynaklarının tweet paylaşımları koleksiyonda yer almıştır. Müşteri görüşleri dışında kalıyor olmasına rağmen bu tür konular işletmeler tarafından dikkate alınmalı ve değerlendirilmelidir. Burada geçen iki konu işletmeler için; sektörel olarak koronavirüs koruyucu tedbirlerin hayata geçirilmesine ihtiyaç olduğunu, kuryelerin emniyetinin sağlanmasına yönelik emniyet tedbirlerini geliştirici tedbirler alınması gerektiğini göstermektedir. Örneğin kazaların önlenmesi için üç tekerli motosiklet kullanımı yaygınlaştırılabilir ve bu kamuoyuna çeşitli reklam araçları ile ulaştırılabilirse, kamuoyunun algısında pozitif bir etki oluşturulabilir. Burada işletmelerin her türlü imkan kabiliyeti değerlendirilerek farklı çözüm önerileri geliştirilebilir. Sektöre yön verecek bu tür görüşlerin (müşteri görüşü olmasa da), en etkili kullanıcılar tarafından gündeme getirildiği için öncelikle dikkate alınması gerekir. Bunların dışında kalan altı kullanıcının (Şekil 3.10'daki 11'inci, 12'nci, 14'üncü, 20'nci, 23'üncü ve 24'üncü kullanıcılar) gündem konularının; ÇYS hizmetinin kolaylığı, sipariş yoğunluğu ve siparişteki gecikme, bir yemek programının ÇYS konusunda kullanıcıları sipariş vermeye motive etmesi, ÇYS alanında faaliyet gösteren bir işletmenin kampanya olarak müşterilere gönderdiği kitap hediyesinin olumlu karşılanması, koronavirüs salgını sonrası sipariş vermeyi bırakma eğilimi olduğu görülmektedir.



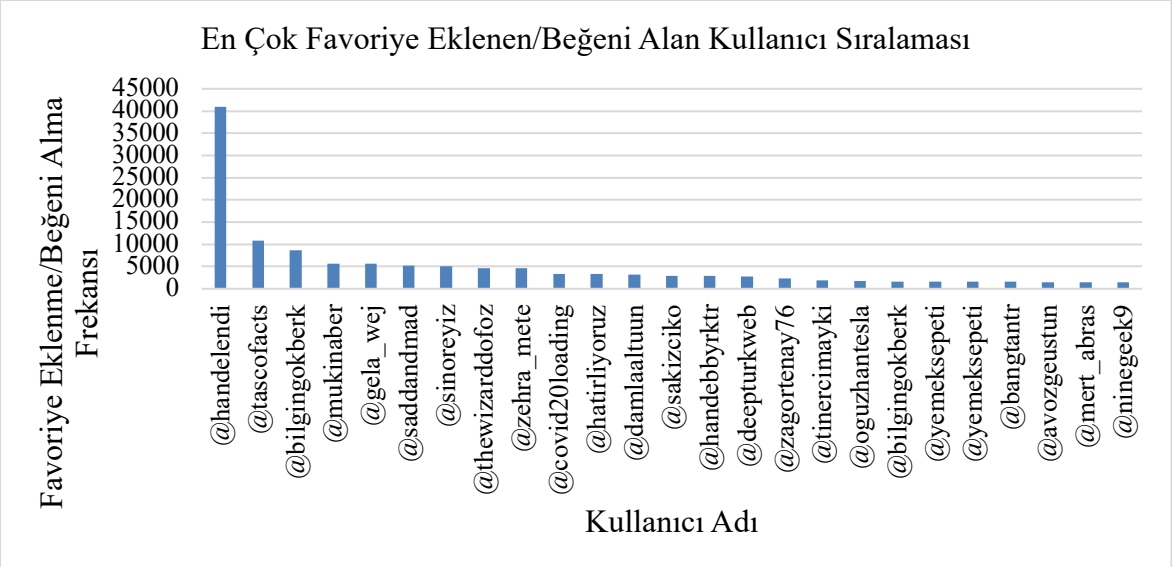
Şekil 3.10. En etkili kullanıcılar

En çok RT'lenerek *görünürlük* elde eden kullanıcılar, Şekil 3.11'de gösterilmiştir. Şekil 3.11'deki görünürlük sıralaması, kullanıcı tweetlerinin aldıkları RT sayıları dikkate alınarak yapılmıştır [67]. Şekil 3.11 incelendiğinde 14'üncü, 19'uncu ve 20'nci sıradakiler (işletme ve kurum kullanıcıları) dışındaki kullanıcıların ÇYS konusunda paylaşımda bulunan kullanıcılar oldukları görülmektedir. Şekil 3.11'deki kullanıcıların görünürlük sıralaması ile, en aktif (Şekil 3.9) ve en etkili (Şekil 3.10) kullanıcı sıralamalarındakine nazaran daha fazla sayıda gerçek müşterilere ulaşılabildiği görülmektedir. Kurum, kuruluş, haber kanalları, işletmelerin otomatik sistem ve uygulamalarla otomatik tweet gönderileri paylaşarak sosyal ağlarda en aktif kullanıcılar arasında yer alabildikleri, bu nedenle en aktif kullanıcı analizinin gerçek potansiyel müşteri konumundaki kullanıcılara ulaşmada yanıltıcı olabileceği, en etkili kullanıcı sayısının ise takipçi sayısı ile ölçülmesi nedeniyle kurum, kuruluş, işletme isimlerinin ön plana çıkmasına neden olabildiği görülmektedir. Bu nedenle kullanıcı analiz sonuçları üzerinde detaylı inceleme yaparak gerçek potansiyel müşteri dışındaki kullanıcıların ayıklanması gerekir.



Şekil 3.11. Tweetleri en çok RT alarak görünürlük elde etmiş kullanıcılar

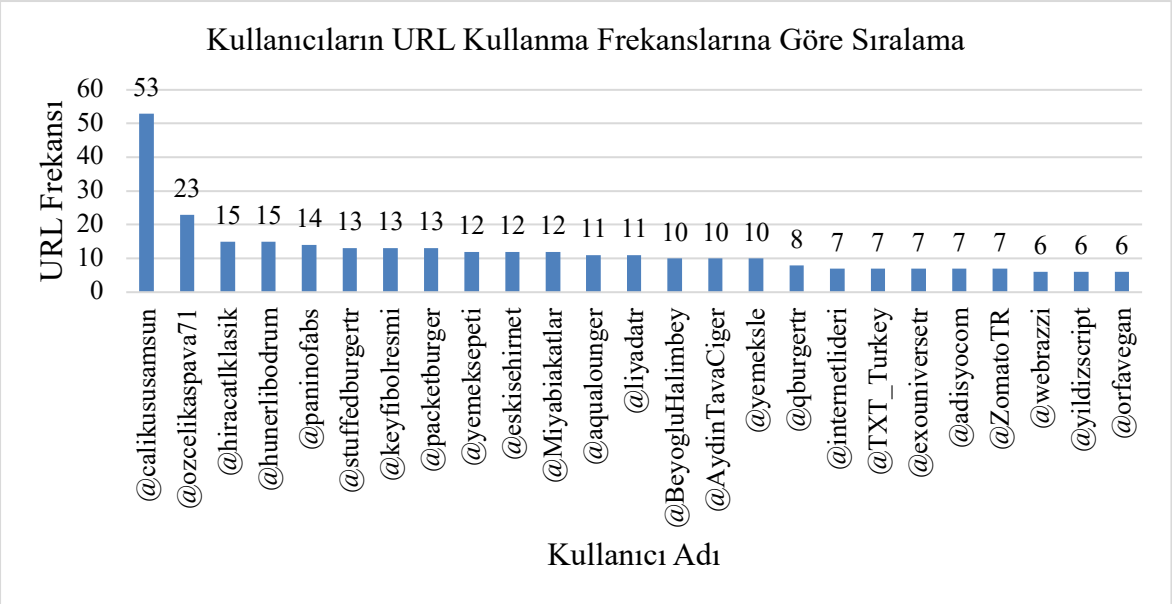
En çok beğeni alan kullanıcılar Şekil 3.12'de sıralanmıştır. Şekil 3.12'deki sıralama incelendiğinde 16'ncı, 20'nci ve 21'nci sıradaki kullanıcılar (sektördeki işletmelere ait kullanıcılar) dışındaki kullanıcıların gerçek potansiyel müşteriler olabileceği dikkate alınarak işletmeler tarafından öncelikli hedef kitle olarak değerlendirilmelidir.



Şekil 3.12. Favoriye eklenme/beğeni alma sayıları açısından kullanıcı sıralaması

3.2.4. Tweetlerin URL istatistikleri

Tüm koleksiyon içinde URL içeren toplam tweet sayısı 3 330'dur ve bu koleksiyondaki tweetlerin % 9'dur, yaklaşık 11 tweet başına bir URL düşmektedir. URL kullanımının da bu alana yönelik tweetlerde kullanımının düşük olduğu söylenebilir. Tweetlerinde en çok URL bulunduran kullanıcılardan ilk 25'i aşağıdaki Şekil 3.13'te gösterilmiştir. Şekil 3.13'teki kullanıcı isimleri incelendiğinde; 18'inci (sektöre ilişkin bilgilendirme yazıları), 19 ve 20'nci (alıntı konuşma metinleri ve müzik grubu görseli) sıradaki kullanıcılar dışında kalanların tamamının sektöre ilişkin reklam ve işletme tweetleri olduğu görülmektedir. Sonuç olarak, ÇYS konusundaki tweetlerde URL analizleri ile sektöre yönelik çıkarımlar yapmanın yanıltıcı olabileceği ya da yeterli olmayacağı öngörülmektedir.



Şekil 3.13. URL bulundurma sayılarına göre kullanıcı sıralaması

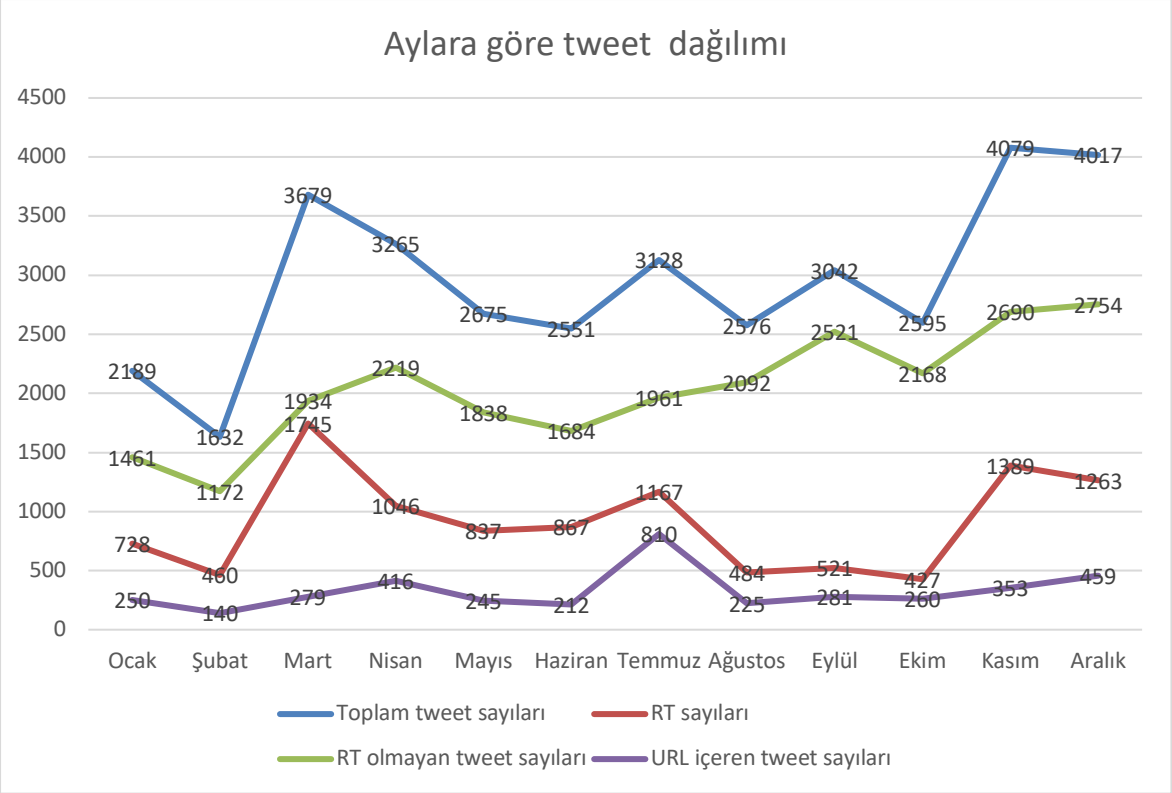
URL analizi kapsamında URL'lerin içerikleri açısından konularına göre dağılımı, URL'lerde kullanılan dokümanların formatlarına göre kullanım oranları, URL'lerin doküman formatına göre tasnif edilerek yüzdelerinin belirlenmesi yapılabilir. En çok RT'lenen tweetlerin URL bulundurma yüzdeleri ile URL'lerin görünürlük elde etme oranları, en çok RT'lenen Tweetlerdeki URL'lerin içeriklerine göre tasnifi ve içeriklere göre yüzdeleri, en çok favoriye alınan/beğenilen tweetlerin URL bulundurma yüzdeleri ve içeriklerine göre tasnifi ve içeriklere göre yüzdeleri gibi URL analiz araçları kullanılabilir. Bu çalışmada, koleksiyondaki tweetlerde URL kullanımının çok düşük olması ve

kullanılanların çoğunun sektördeki işletmelerin reklam tweetlerinde olması nedeniyle URL konusunda detaylı analizlere ihtiyaç duyulmamıştır.

3.2.5. Tweetlerin zaman grafikleri

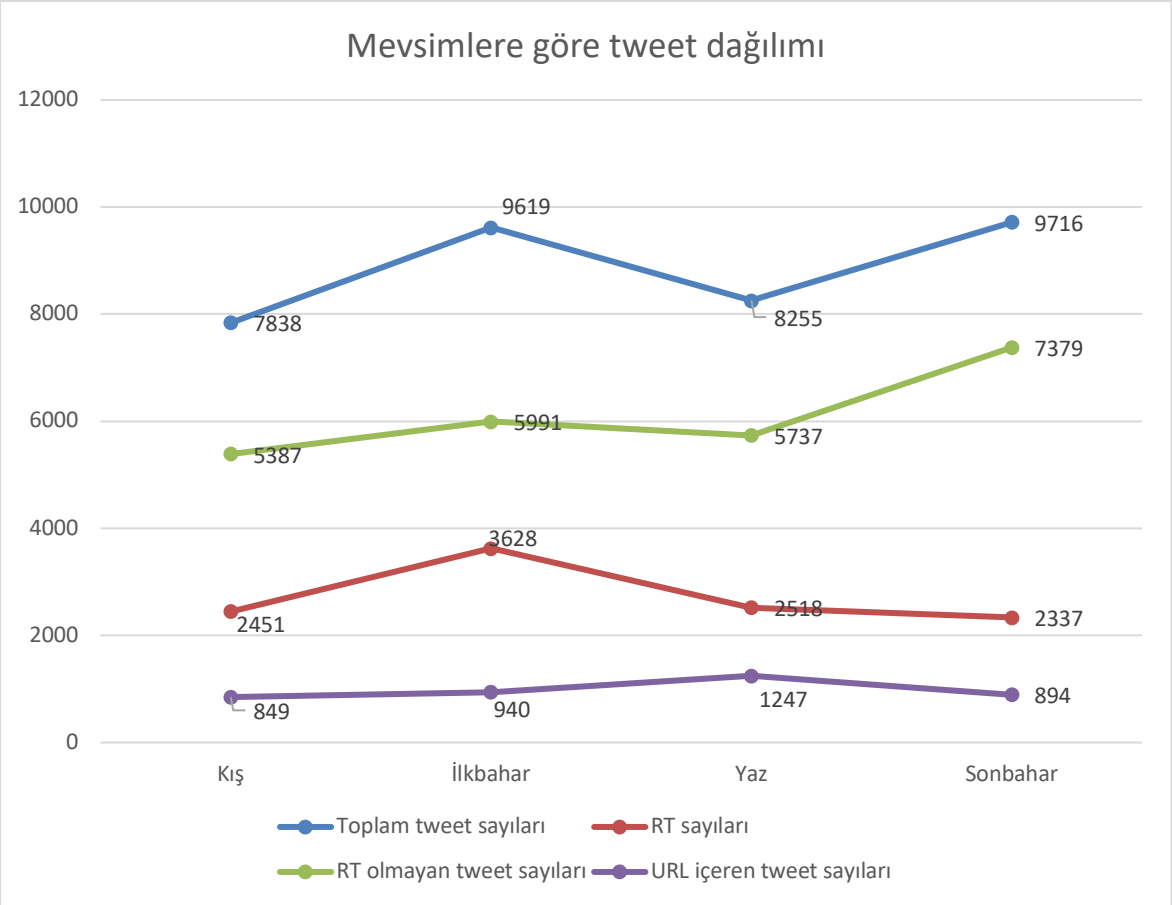
Buraya kadar kullanılan istatistiki değerlerin zaman birimleri ile ilişkilendirilmesi, işletmeler için kullanışlı ilave bilgilerin ortaya çıkarılmasını sağlayabilir. İleriki maddelerde tweet, RT, URL frekanslarının farklı zaman dilimleri ile ilişkilendirilmesinden çıkarılan sonuçlar verilmiştir. Tweetlerin aylara göre dağılımı Şekil 3.14'te verilmiştir.

Şekil 3.14'te toplam tweet (tweet+RT), RT olmayan tweet (sohbet başlatan tweet), URL içeren tweet sayıları olarak ayrı ayrı gösterilmiştir. ÇYS konusunda Twitter (X)'da paylaşımların en düşük olarak Şubat ayında olduğu, Mart ayında yıl içinde ikinci yüksek seviyeye ulaştığı, Nisan ayında düşüşe geçerek Kasım ayı ortasına kadar birbirine yakın düzeylerde devam ettiği, Kasım ayı ortasından sonra Aralık ayına kadar yılın en yüksek seviyesine ulaştığı görülmektedir. Şekil 3.14'te URL içeren tweetleri gösteren çizgi hariç diğerlerinin paralellik göstermesi olağandır. Çünkü en üst sıradaki mavi renkli çizgi diğer ikisinin toplamını göstermektedir. Mart, Kasım, Aralık aylarında ÇYS konusundaki paylaşımlardaki artış, bu dönemlerde işletmelerin satışlarında da artış olduğunu/olabileceğini göstermektedir. Bu bilgiler, işletmelerin kendi veri tabanlarındaki bilgiler ile karşılaştırılarak, söz konusu artışlardan ne kadar istifade edilip edilemediği ortaya çıkarılabilir, eğer istifade edilememişse işletmenin hangi konularda eksiklikleri olduğunun araştırılması gerekir.



Şekil 3.14. Tweetlerin aylara göre akışı-zaman grafiği

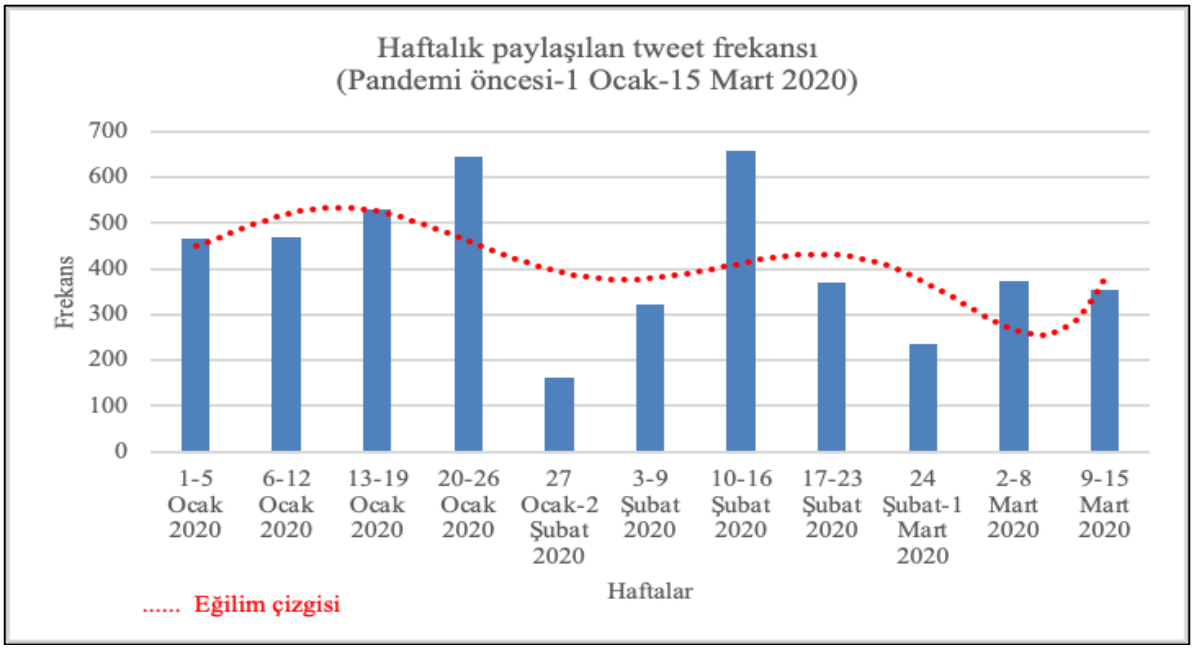
Şekil 3.15'te, Şekil 3.14'teki bilgilerin mevsimlere göre uyarlanmış hali verilmiştir. Mevsim olarak ÇYS konusundaki sohbet ve paylaşımların İlkbahar ve Sonbahar dönemlerinde yüksek seviyede gerçekleştiği görülmektedir. Şubat ayında ve yaz döneminde okul tatili, Mayıs ayının Ramazan ayı olması nedeniyle paylaşımlarda düşüş olabileceği yönünde yorum yapılabilir ancak bunun kesin olarak ifade edilebilmesi için ayrıca koleksiyonun içerik analizi yapılması gerekir.



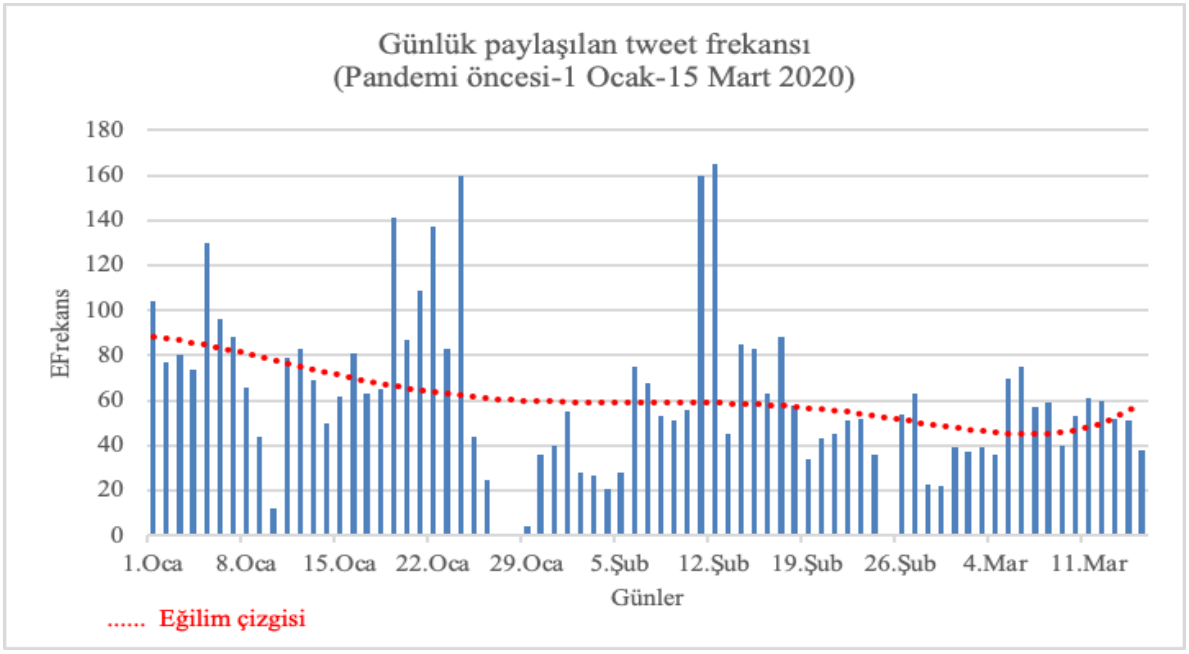
Şekil 3.15. Tweetlerin mevsimlere göre akışı-zaman grafiği

Koleksiyonu oluşturan zaman dilimi içinde pandemi döneminin de yaşandığı dikkate alınarak yukarıdaki aylık ve mevsimsel incelemeler, pandemi etkisini yansıtabilecek zaman dilimlerine ayrılarak ileriki şekillerde gösterilmiştir. Pandemi açısından tarih olarak, gündelik hayata yönelik tedbirlerin alınmaya başladığı ve okulların tatil edildiği 16 Mart 2020 tarihi esas alınmış; 1 Ocak-15 Mart 2020 tarih aralığı pandemi öncesi, 16 Mart-31 Aralık 2020 tarih aralığı pandemi dönemi olarak incelenmiştir.

Pandemi öncesi döneme ait haftalık ve günlük tweet paylaşımları Şekil 3.16 ve Şekil 3.17'de gösterilmiştir. Her iki grafikten de paylaşımlarda düşüş eğilimi olduğu görülmektedir. Bu dönemde tweet paylaşımındaki haftalık ortalama 417, günlük ortalama 61'dir.

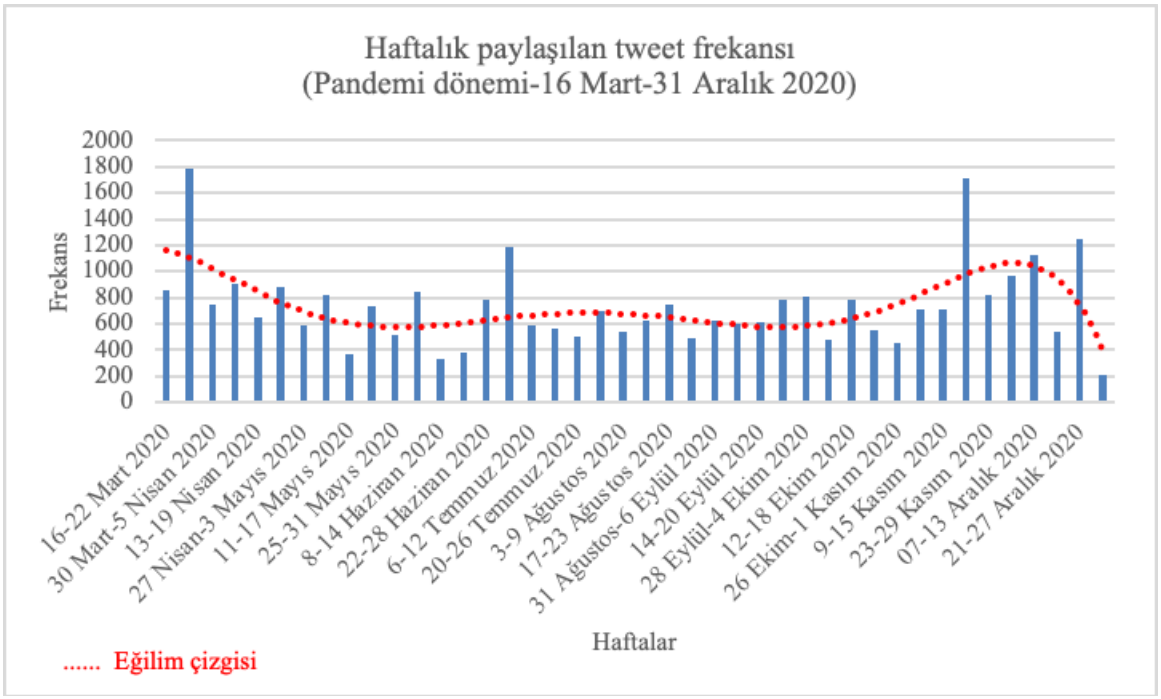


Şekil 3.16. Haftalık tweet paylaşım frekansları-pandemi öncesi

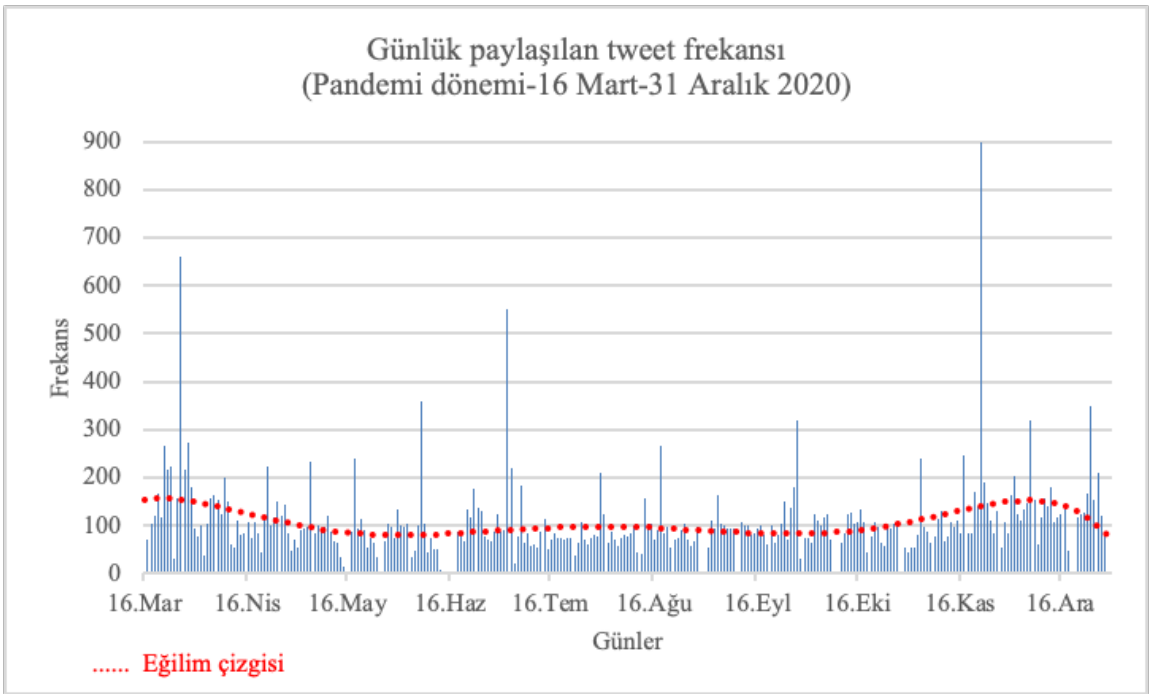


Şekil 3.17. Günlük tweet paylaşım frekansları-pandemi öncesi

Pandemi dönemine ait haftalık ve günlük tweet paylaşımları Şekil 3.18 ve Şekil 3.19'da gösterilmiştir. Bu dönemde tweet paylaşımındaki haftalık ortalama 734, günlük ortalama 106'dır. Tweet paylaşımındaki ortalama değerin, pandemi öncesindeki ortalamaya göre % 74 arttığı, pandemi dönemindeki tweet paylaşımlarındaki değişkenliğin pandemi öncesindeki döneme göre daha kararlı bir yapıda olduğu görülmektedir.



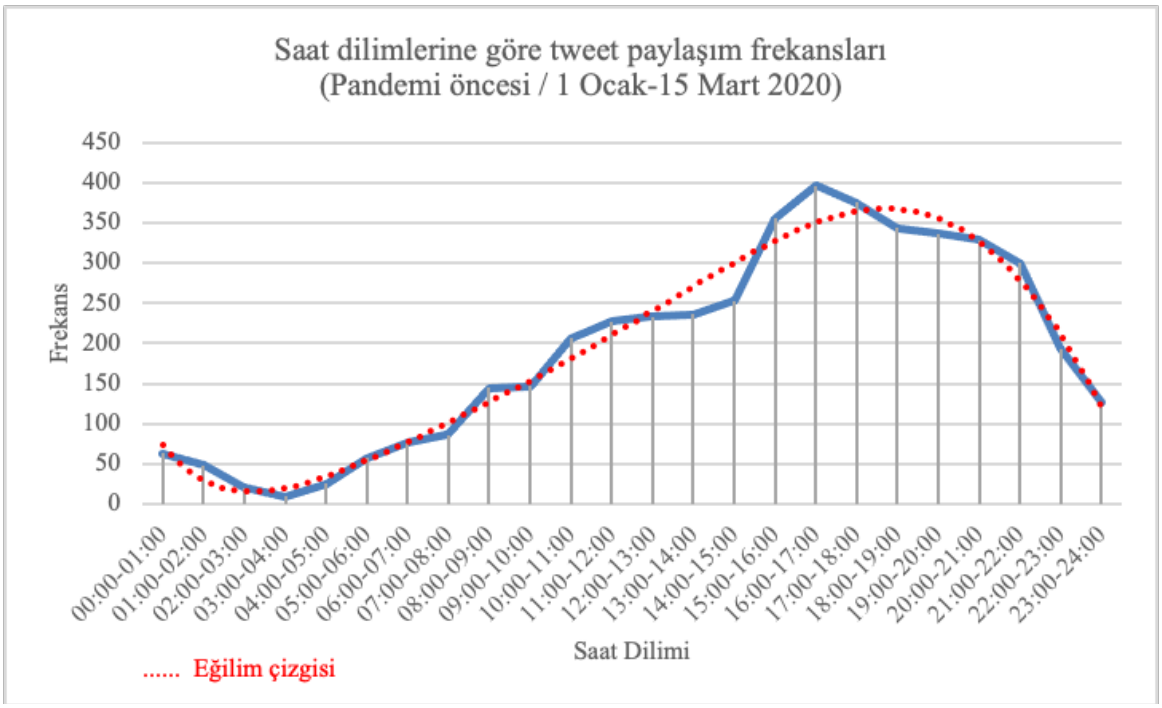
Şekil 3.18. Haftalık tweet paylaşım frekansları-pandemi dönemi



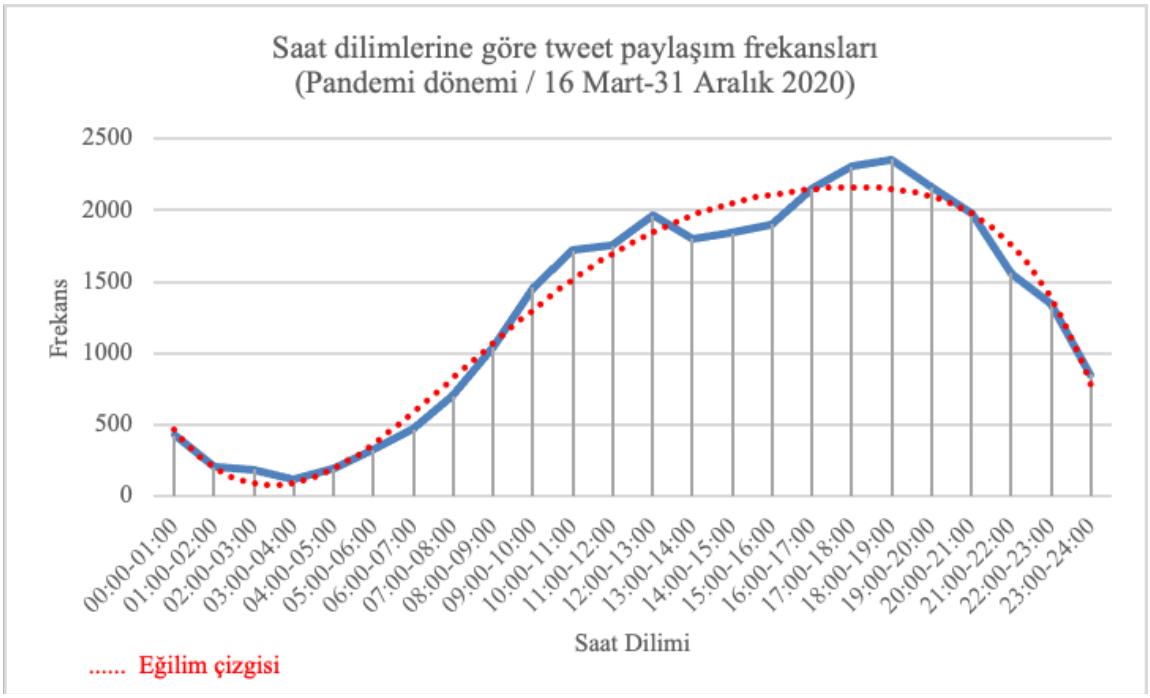
Şekil 3.19. Günlük tweet paylaşım frekansları-pandemi dönemi

Şekil 3.20 ve Şekil 3.21'de paylaşım saatleri esas alınarak tespit edilen tweet frekansları, pandemi öncesi ve pandemi dönemi şeklinde gösterilmiştir. Her iki şekilde de yemek siparişi verme konusunda paylaşımların saat 11:00-12:00'den itibaren yükselişe geçtiği, yüksek

seviyenin 21:00-22:00'a kadar korunduğu görülmektedir. Gün içindeki en yüksek seviyenin; pandemi öncesi dönemde saat 16:00-17:00'da, pandemi döneminde saat 18:00-19:00'da ortaya çıktığı görülmektedir.



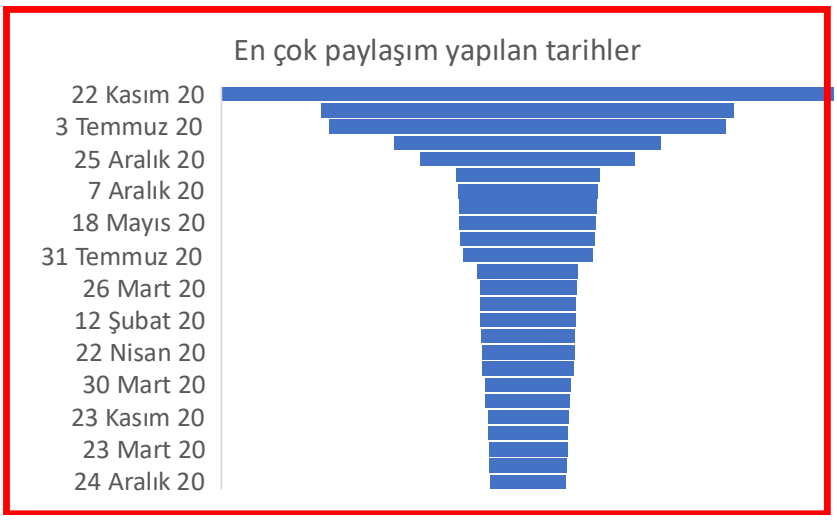
Şekil 3.20. Saat dilimlerine göre tweet paylaşım frekansları-pandemi öncesi



Şekil 3.21. Saat dilimlerine göre tweet paylaşım frekansları-pandemi dönemi

Her iki dönemde ortaya çıkan tweet paylaşım saatlerindeki artış eğiliminin öğle ve akşam yemek saatleri öncesi ve sonrasında gerçekleştiği görülmektedir. İşletmeler, paylaşımların arttığı bu saatleri, sipariş yoğunluğunun artış gösterebileceği/gösterdiği saatler kabul ederek satış politikalarında düzenlemeler yapabilirler. Örneğin müşteri tercihlerinin etkilenmesi için yapılacak reklamların, indirim ve joker gibi uygulamaların bu saatler dikkate alınarak yapılması daha etkili sonuçlar ortaya çıkarabilir.

Yıl içinde en çok paylaşım yapılan tarih sıralaması Şekil 3.22'de gösterilmiştir. Sıralama incelendiğinde ilk üç sırada yer alan tarihlerin, işletmeler tarafından çok fazla sayıda reklam içerikli tweet paylaşımı yapılması nedeniyle öne çıktığı, bu üç tarihte potansiyel müşterilerin tweet paylaşımının çok az sayıda gerçekleştiği görülmektedir. Dördüncü sıradan itibaren gösterilen tarihlerde ise potansiyel müşteri paylaşımlarının yoğun olduğu görülmektedir.

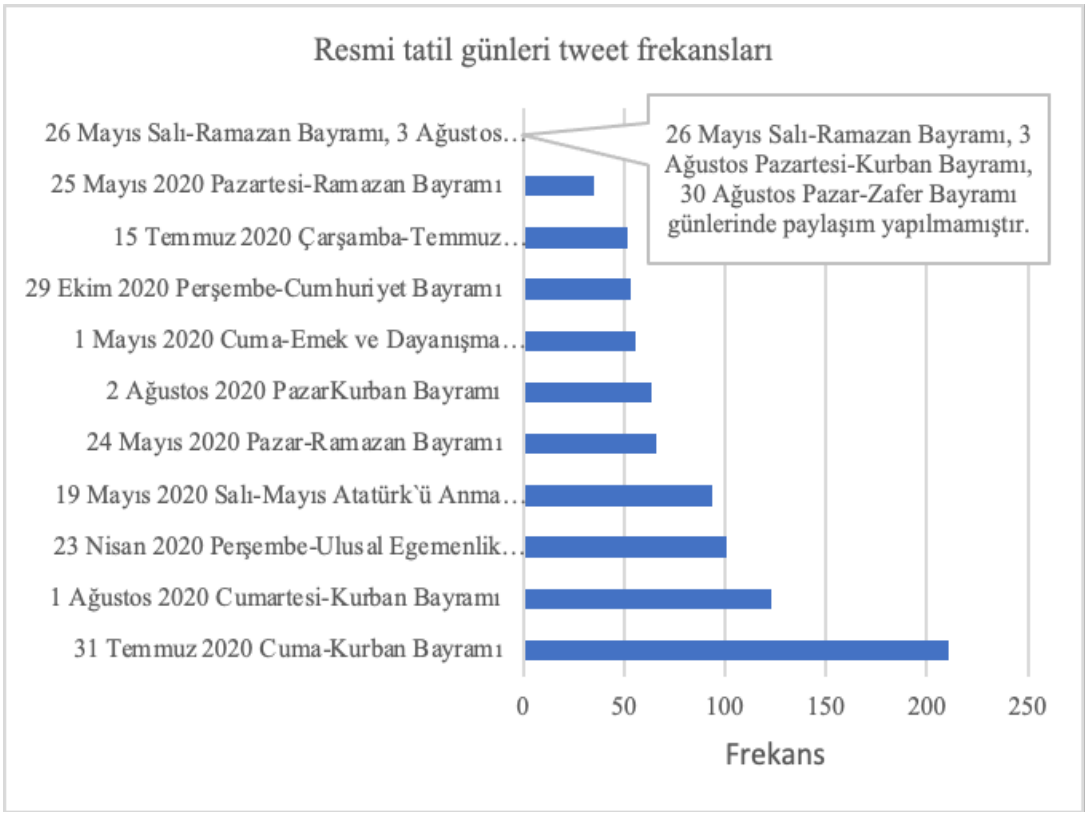


Şekil 3.22. En çok tweet paylaşımı yapılan tarihler

Şekil 3.22'de herhangi bir özelliği nedeniyle öne çıkan bir tarihin olmadığı görülmektedir. Dördüncü sıradaki 07 Aralık 2020 tarihindeki, kişisel güvenlik endişesi ile çevrimiçi yemek sipariş konusunun ilişkilendirildiği bir tweetin çok fazla sayıda RT yapılarak gün içinde yayıldığı görülmektedir. Gündem belirleyici nitelikteki bu paylaşım, ilk bakışta ÇYS ile doğrudan ilişkili gibi gözükmeyebilir. Ancak tweet içeriği incelendiğinde, yemek siparişini eve teslim eden kuryeler nedeniyle böyle bir ilişkilendirmenin yapıldığı görülmektedir. "Kadınların kişisel güvenlik endişesi" ile ilgili konunun yukarıda yapılan diğer analizlerde de koleksiyonun ana gündem konuları arasında yer aldığı görülmektedir. Bu konuda işletmelerin doğrudan uygulamaya sokabileceği bir tedbir olmasa da çalışanların davranış

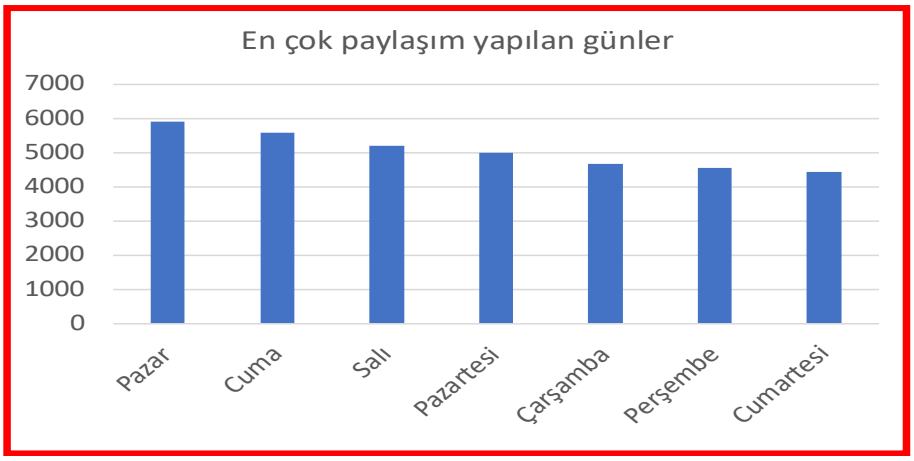
şekilleri ve eğitimine yönelik uygulamalar yapılabilir. Örneğin pandemi döneminde ortaya çıkan temassız teslimat uygulaması, bu konuda alınmış bir tedbir olmamakla birlikte, tweetlerde ifade edilen kişisel güvenlik endişesinin azaltılmasına katkı sağlayacak bir uygulama olarak kabul edilebilir.

Resmî tatil günlerinde yapılan tweet paylaşım frekansları Şekil 3.23'te gösterilmiştir. Bu günlerden, 23 Nisan, 1,19, 24-26 Mayıs tarihlerinde korona virüsü nedeniyle sokağa çıkma yasağı uygulanmıştır. Koleksiyonun tümü için günlük ortalama tweet paylaşımının 97 iken, resmi tatil günlerinde bu ortalamanın 66 olduğu tespit edilmiştir. Resmî tatil günlerindeki gündem maddelerinin ise kuryelerin uygun olmayan davranışları, siparişin yerine ulaşmaması ve siparişte yaşanan gecikmeler, ürünlerde yabancı madde tespit edilmesi, yemek siparişine duyulan özlem, temassız teslimat şeklinde konuların olduğu görülmektedir.



Şekil 3.23. Resmî tatil günleri tweet frekansları

Yıl içinde en çok paylaşım yapılan gün sıralaması Şekil 3.24'te gösterilmiştir.



Şekil 3.24. En çok tweet paylaşımı yapılan günler

Şekil 3.24 incelendiğinde, günler arasında çok yüksek seviyelerde fark ortaya çıkmadığı, Cuma ve Pazar gününün en yüksek ilk iki sırada yer aldığı, Cumartesi gününün ise en az paylaşım yapılan gün olduğu görülmektedir. Şekil 3.24'te görüldüğü gibi tweet paylaşımlarındaki Cuma ve Pazar günlerinde ortaya çıkan artışın, Cumartesi günündeki azalışın çevrimiçi yemek siparişlerindeki artış ve azalış ile paralel olacağı/olduğu öngörülmektedir.

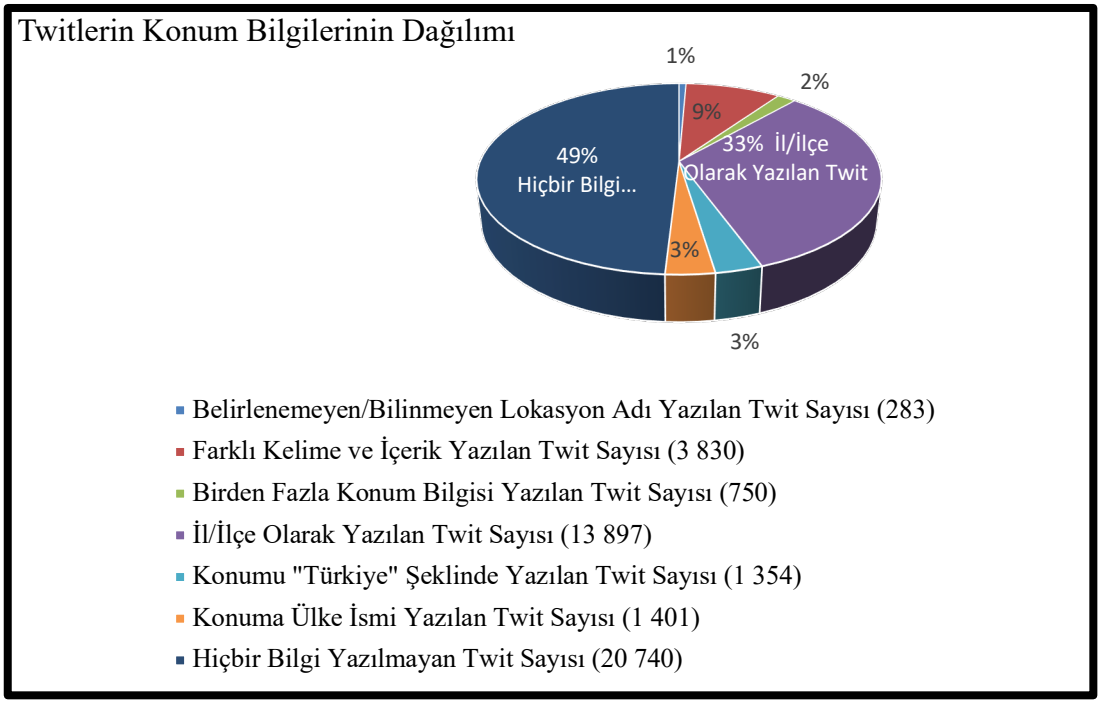
3.2.6. Tweetlerin coğrafi konumlarına göre dağılımı

İşletmeler için müşterilerin temel demografik özelliklerinin neler olduğunun tespit edilmesi müşteri profilinin belirlenmesinde kullanılan unsurlardan biridir [368]. Sosyal medya araçlarında kullanıcılar, zorunlu alanlar dışında kullanıcı profil bilgilerini genelde eksik bırakmaktadırlar ya da profil bilgisinde bulunan alana ilgisiz bilgiler yazmaktadırlar. Profil bilgilerindeki tüm alanlar ayrı ayrı istatiksel olarak faydalı bilgilere dönüştürülebilir. Tweet gönderilerinin profil alanından faydalı bilgi elde edilebilecek bir alan konum/adres bilgisidir. Koleksiyondaki tweetlerin demografik bilgiler açısından sınırlı ölçüde bilgi içerdikleri, kullanıcıların konum bilgileri yerine farklı formatta (sadece mahalle, semt ismi, sadece ilçe ismi, ilçe ve il ismi birlikte gibi) konum bilgisi yazdıkları ya da konum bilgisi dışında farklı ifadeler (espirili, siyasi görüş içerikli gibi) yazdıkları tespit edilmiştir. Bu şekilde yazılmış farklı formattaki bilgilerin ise otomatik tasnif işleminin zor olacağı görülmektedir. Bununla birlikte koleksiyondaki tweetler, kullanıcı profilindeki CKB açısından analiz edilmiş ve ne tür bilgilere ulaşılabileceği gösterilmiştir. Öte yandan birçok alanda olduğu gibi, ÇYS alanında faaliyet gösteren işletmelerde uygulamalarına/platformlarına kullanıcının onayı ile

kullanıcı konum bilgisine ulaşılmasını sağlayan özellik eklemiştir. Bu sayede bir yandan gönderilen ürün hareketi müşteriye iletilerek müşteri memnuniyeti sağlanmakta, diğer yandan müşteri konum bilgisi elde edilerek konuya ilişkin analizler için veri tabanı oluşturulabilmektedir.

CKB analizi için çalışmanın diğer bölümlerindeki analizlerden farklı olarak; koleksiyondan hiçbir ayıklama işlemi yapılmadan 42 255 tweetin tamamı incelenmiştir. Tweetlerin kullanıcı profil bilgisinde yer alan konum bilgisi ifade çeşitleri ve detayları EK-6'da verilmiştir.

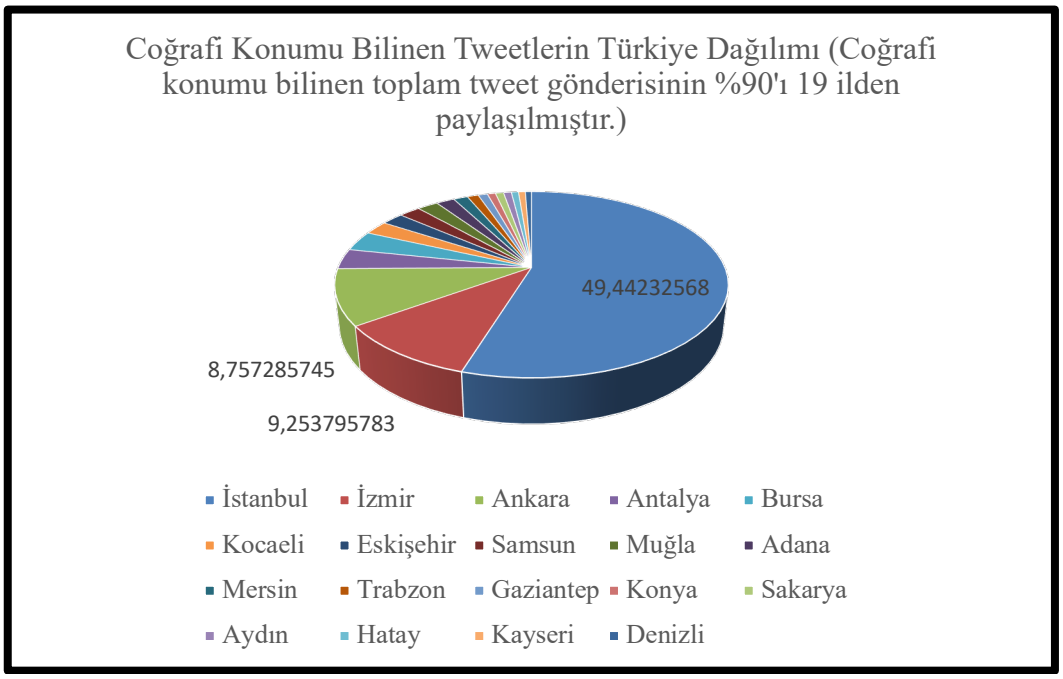
Koleksiyonda kullanıcı profilinde CKB olarak yazılmış ifadeler yedi grup altında sınıflandırılarak Şekil 3.25'te gösterilmiştir. Kullanıcıların %49'unun konum bilgisini hiç girmedikleri, geri kalanların girdikleri verilerin ise ancak %33'nün kullanılabilir bilgi içerdiği görülmektedir.



Şekil 3.25. Tweetlerin CKB'ne göre dağılımı

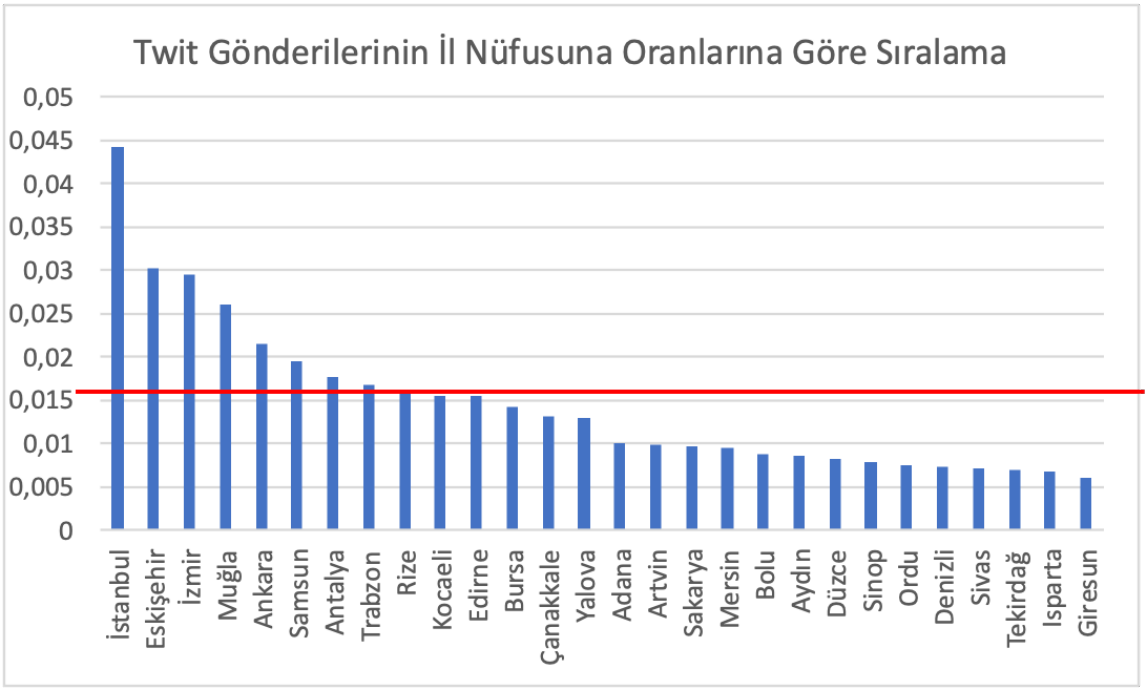
CKB olarak kullanılabilir bilgi içeren koleksiyonun %33'ündeki konumların illere dağılımı Şekil 3.26'da gösterilmiştir. İl bazında konum bilgisi tespit edilebilen bu tweetlerden %90'ının (diğer %10'u oluşturan il sayısı çok fazla sayıda olduğu için gösterilmemiştir)

Şekil 3.26'da listelenen 19 ilde olduğu görülmektedir. Bu (%90) oranın % 67.4' ünü ise İstanbul, İzmir ve Ankara illeri oluştururken, geri kalan % 22.6'sını diğer 16 ildeki kullanıcılar oluşturmaktadır.



Şekil 3.26. CKB olan tweetlerin illere göre dağılımı

Coğrafi konumu il/ilçe bazında tespit edilen toplam tweetlerin il nüfuslarına [369] göre oranları tespit edilerek Şekil 3.27'de gösterilmiştir. 2019 yılı il nüfusuna göre oranlama yapıldığında sırasıyla İstanbul, Eskişehir, İzmir, Muğla, Ankara, Samsun, Antalya ve Trabzon şehirlerinin profil bilgisinde CKB paylaşımı açısından Türkiye ortalamasının (coğrafi konumu bilinen il/ilçe bazındaki toplam tweet sayısının, Türkiye 2019 yılı nüfusuna oranı 0, 016'dır ve Şekil 3.27'de kırmızı yatay çizgi ile gösterilmiştir.) üzerinde bir değere sahip olduğu görülmektedir.



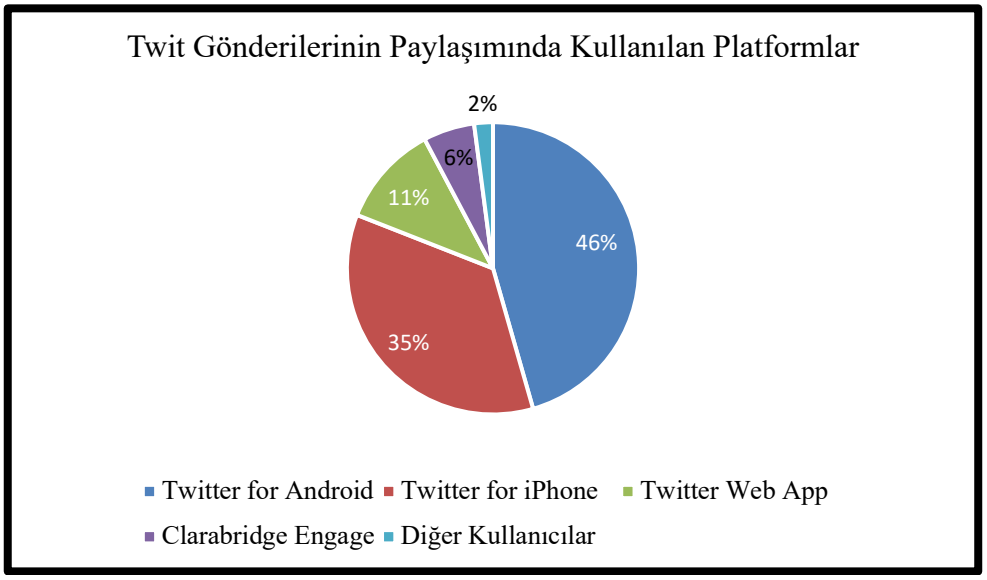
Şekil 3.27. Tweet gönderilerinin il nüfuslarına göre oranları

Twitter (X) kullanıcı profilinden çıkarılan bu bilgilerin, işletme kararlarına yön verebilecek yeterli içerik ve sayıda olmadığı, bu bilgilerin ancak başka kaynaklardan elde edilecek bilgilerle desteklenmesi halinde kullanılabileceği öngörülmektedir.

3.2.7. Tweetlerde kullanılan platform bilgilerinin dağılımı

Şekil 3.28'de tweet gönderilerinin paylaşımında kullanılan platformların dağılımı gösterilmiştir. Bu bilgi, işletmeler açısından uygulamaların, web sayfalarının geliştirilmesine, sosyal medya üzerinde kullanılabilecek araç ve uygulamalara yönelik İG faaliyetlerinde hangi platformlara yönelmeleri gerektiğine yönelik fikirler verebilir. Şekil 3.28 incelendiğinde platformların %91'inin Android, iPhone ve Web uygulamalarından meydana geldiği görülmektedir. Dördüncü sırada ise %6'lık dilimle "Clarabridge Engage" isimli platformun yer aldığı görülmektedir. Bu dördüncü platform, müşteri ilişkileri yönetimi amacıyla [370] kullanılan bir yazılımdır, dolayısı ile bu platformun ticari/kurumsal işletmeler tarafından kullanılıyor olabileceği değerlendirilebilir. Son dilimde görülen %2'lik dilimde ise farklı 60 platformun kullanıldığı, bunların içinde frekansı en yüksek olan platformun 180 kez kullanılmış olan "Twitter (X) for iPad" olduğu görülmektedir. Sonuç

olarak Twitter (X) for Android, iPhone, Web APP ve iPad platformlarının ÇYS alanında yaygın olarak kullanıldığı görülmektedir.



Şekil 3.28. Tweet gönderilerinin paylaşımında kullanılan platformların dağılımı

3.2.8. Tanımlayıcı analizle elde edilen bulgular

Kullanılan model ile 2020 takvim yılının tamamını kapsayan koleksiyondan elde edilen sonuçlar incelendiğinde, sektörel olarak işletmelere yön verebilecek nitelikte fikirlere ulaşılabildiği görülmüş ve amaca uygun sonuçlar için yapılması gereken işlemler tespit edilmiştir. Bu çerçevede bu tür ticari ürün/hizmet konusu sektörlere ilişkin Twitter (X) koleksiyon analizlerinde reklam içerikli tweetlerin ayıklanması işleminin gerekli bir adım olduğu öngörülmektedir. Metin önişlemleri aşamasında yapılacak reklam ayıklama işleminin analiz sonuçlarının yanıltıcılığını azaltabileceği, bununla birlikte reklam içerikli/kaynaklı bilgilerin analiz sonuçlarında tekrar ortaya çıkabileceği için bu aşamada son bir ayıklama işlemine ihtiyaç duyulabileceği görülmüştür.

Her konuda sohbetin gerçekleşebildiği Twitter (X), sektörel olarak yeni fikirlere kaynaklık etme potansiyeline sahiptir. Söz konusu potansiyelin, geliştirilen yeni modeller ve alana özel çalışmalarla işlenmesi ve sektörde faaliyet gösteren işletmelerin hizmetine sunulması gerekir. Özellikle alana özel yapılacak çalışmalar, analizlerde kullanılan algoritma ve modellerin etkinliğini, performansını artıracaktır. Bu tür çalışmalarda modellerin etkinliğini etkileyen önemli bir nokta ise, sosyal medya ortamlarında yoğunluğunu artıran reklamların

veri setleri içinden ayıklanmasının gerekliliğidir. Bu çalışmada önerilen model ile tanımlayıcı analiz araçları kullanılarak, Twitter (X) üst verilerinden ne tür bilgiler elde edilebileceği, bu bilgilerle ürün/hizmetlerin iyileştirilmesine kaynaklık edebilecek Twitter (X) kullanıcıları arasındaki esas sohbet, gündem konuları ve yeni fikirlere nasıl ulaşılabileceğinin adımları alana özel bir çalışma ve önerilen TVİTA modeli ile gösterilmiştir. Model ile detaylı metin önişlemlerine, DDİ metotlarına ihtiyaç duymaksızın koleksiyonun içeriğinin ve ürün/hizmetin olumlu ve olumsuz yönlerinden en etkili ve gündemde olanların tespit edilmesi sağlanmıştır. Modelde veri setleri içinden reklam tweetlerinin, çalışma kapsamında oluşturulan alana özel reklam sözlüğü kullanarak ayıklaması yapılmıştır. Çalışma kapsamında yemek alanına özel bir sözlük hazırlanmıştır.

Bu alanda, potansiyel müşterilerin diyez etiketi ve URL kullanma eğiliminin düşük olduğu, diyez etiketi ve URL bulunan tweetlerin çoğunluğunun sektördeki işletmelere ait olduğu görülmüştür. Benzer durumun diğer ticari ürün/hizmeti konu alan tweet paylaşımlarında da ortaya çıkabileceği, bu nedenle Twitter (X)'dan bu tür konularda yapılacak veri çekmede, diyez etiketi yerine doğru seçilmiş anahtar kelimelerin daha yüksek sayıda potansiyel müşteri tweetlerine ulaşılmasını sağlayabileceği öngörülmektedir.

Sonuç olarak yapılan analizlerle ÇYS konusunda faaliyet gösteren işletmelere, faaliyetlerini ne şekilde İG gerektiği konusunda yön verebilecek potansiyel müşterilerden elde edilen sohbet konuları olarak; (i) bazen gece geç saatlerde yemek siparişi verme, bazen yeni başlamış oldukları diyeti bozacak olma, bazen evde yemek olduğu halde sipariş verme nedeniyle yaşanan pişmanlık ve kararsızlık hissi, (ii) ÇYS verme konusunda özellikle koronavirüs salgını nedeniyle ürün ve hizmete güven konusunda yaşanan endişe ve buna bağlı olarak yemek siparişi verme davranışını terk etme eğilimi, (iii) evinde yalnız iken yemek siparişi veren kadınların yaşadıkları kişisel güvenlik endişesi (bu endişeyi gidermek için evde yalnız olmadıklarını kuryeye hissettirmek maksadıyla, ürün teslimatı esnasında evin içine doğru seslenme hareketi sergileme), (iv) hastane çalışanları ve hastaların hastane yemeklerini beğenmemeleri nedeniyle çevrimiçi yemek sipariş hizmetine yönelmeleri, (v) aynı evi/ortamı paylaşan kişilerin başlangıçta sadece kendileri için sipariş ettikleri yiyecekleri arkadaşları ile paylaşmak zorunda kalmaları, (vi) kuryelere emniyetli ve güvenli çalışma koşullarının sağlanması gerektiği, kuryelerin karıştıkları trafik kazaları, (vi) sipariş edilen ürün sunumunun beklenen şekilde gerçekleşmemesi, (vii) ÇYS uygulamaları üzerinden ihtiyaç sahiplerine yemek yardımı (pandemi nedeniyle), (vii) sektörde çalışan

işletmelerin indirimli/joker uygulamaları, (viii) televizyonda yayımlanan bir yemek programının, kişileri etkileyerek (programda kullanılan yemek görselinin de katkısıyla) yemek siparişi verme isteklerinin artırması, (ix) yemek sipariş eden müşteriye ürünün ulaşmamış olmasına rağmen, cep telefonuna "ürün teslim edildi" mesajının gönderilmesi, (xi) verilen yemek siparişinin ilgili restorana ulaşmaması, (xii) restorana hizmet veren kuryenin, restorana gönderilmesinde gecikme yaşanması, (xiii) yemek fiyatlarındaki artış, (xiv) ÇYS hizmetinin hayatı kolaylaştırması, (xv) yemek siparişi konusunda yaşanan yoğunluk ve siparişlerdeki gecikmeler, (xvi) sektörde faaliyet gösteren bir işletmenin yemek siparişi yanında ücretsiz kitap hediyesi kampanyasının beğenilmesi tespit edilmiştir.

Öte yandan tanımlayıcı analiz sonucunda, veri setinin oluşturulması aşamasına ilişkin şöyle bir husus tespit edilmiştir. En aktif kullanıcı istatistiklerinde sektördeki işletme isimlerinin ya da konuya ilişkin bir haber nedeniyle, en etkili kullanıcı isimlerinde haber kanal ve program isimlerinin ön plana çıktığı görülmektedir. Bu kullanıcılar, sektöre ilişkin bir potansiyel müşteriyi temsil etmemekle birlikte, etkinlik açısından potansiyel müşteri görüşlerini etkileme ve yönlendirme potansiyeline sahiptirler. Bu nedenle bu tür kullanıcıların tweetlerinin koleksiyondan ayıklanıp ayıklanmaması hususu yapılan çalışma konusuna bağlı olarak değerlendirilmeli ve karar verilmelidir.

ÇYS konusundaki paylaşımların; (i) Mart, Kasım, Aralık aylarında arttığı ve Şubat, Mayıs ayları ve yaz döneminde düştüğü, (ii) gün içinde 12:00-22:00 saat aralıklarında yükseliş olduğu, 10:00-12:00 saat aralığı ile saat 22:00'dan sonraki zaman dilimlerinde düşüş olduğu görülmektedir. İşletmeler bu bilgiler çerçevesinde pazarlama, reklam politikaları ve diğer faaliyetlerine yön verebilirler. Çalışma dönemi içindeki tweet paylaşımlarının günlük ortalamasında, pandemi döneminde, pandemi öncesi döneme göre %74 oranında artış gözlenmiş ve pandemi dönemi boyunca paylaşımlardaki günlük ortalamanın istikrarlı ve kararlı bir şekilde devam ettiği gözlenmiştir.

Twitter (X) paylaşım platformu olarak en çok kullanılan platformların sırasıyla; Android, iPhone, Web, iPad olduğu tespit edilmiştir. Bu bilgi alanda faaliyet gösteren bir işletmenin uygulama/web geliştirme çalışmalarında önceliklendirme yapmasına imkan sağlayabilir.

Çalışmanın tanımlayıcı analizi ile elde edilen sonuçlar üzerinden çıkarılabilecek sektörel öneri ve İG'ye ilişkin fikirler aşağıda sıralanmıştır. Bu kapsamda işletmeler;

- Kullanıcının verdiği siparişlerin frekansının tespit edilmesi ve en çok sipariş verdiği yemek ya da menünün kısayol uygulaması şeklinde uygulamaya "her zamankinden" şeklinde bir tercih şeklinde eklenmesi,

- Ürün filtrelemelerine "sağlıklı ürünler", "diyet menüleri", "vegan menüleri", "vejeteryan menüleri" şeklinde alternatiflerin eklenmesi,

- Günlük olarak en çok tüketilen yemek çeşidinin tespit edilerek "günün menüsü" şeklinde bir menü oluşturulması ve bunun uygulamaya tercih butonu ile eklenmesi,

- Uygulama içine "rastgele menü" şeklinde günlük olarak değişen menü alternatiflerinin sunulmasını sağlayan buton eklenmesi,

- Restoranların kapanma saatleri esas alınarak, kapanma saatine yakın verilen siparişlerin işletme tarafından iptal edilmesini önlemek maksadı ile platform uygulamalarına "restoranın son sipariş alma saati" butonu eklenmesi,

- Ev yemeği şeklinde üretim yapan restoranların platformdaki görünürlüğünün artırılması, kuryelerin çalışma koşullarının iyileştirilmesi, özellikle motosiklet kazalarının önlenmesi maksadıyla, kuryelerin yaralanmalarını ve hava koşullarından etkilenmelerini önleyici, üç tekerlekli motosiklet kullanılması gibi koruyucu ekipmanlarının artırılması,

- Müşteri şikayetlerine geri dönüş konusunda özellikle geri dönüş süresini de içine alan standardizasyon sağlanması,
- ÇYS konusunda platformlarda müşteri memnuniyeti için uygulanan promosyon, ödül, ücretsiz uygulamaların sonuçlarının müşteri ile etkileşimli olarak sonuçlandırılmasını sağlayan özellikler kazandırılması,

- Çalışanlara özellikle kuryelere standart bir davranış biçimi kazandırılması, müşterilerin memnuniyetini sağlamak için çalışan memnuniyetinin sağlanması, ürünlerin içerik konfigürasyonu için alternatiflerin genişletilmesi konusunda restoranların teşvik edilmesi,

- Bir platform ile birleştirilen işletmeler ile platform sağlayıcı işletmenin (PSİ) yetki ve sorumluluklarının kesin sınırlarla tanımlanması, müşterilerin problem çözümünde tek nokta ile muhatap olmasının sağlanması, PSİ'nin, bu platformu kullanan işletmelerden aldığı komisyon oranları konusunda işbirliği yapılması ve bu işbirliğinin kamuoyu ile paylaşılarak işletmelerin komisyonlarla ezildiği algısının ortadan kaldırılması, iş ortaklığı anlayışının geliştirilmesi,

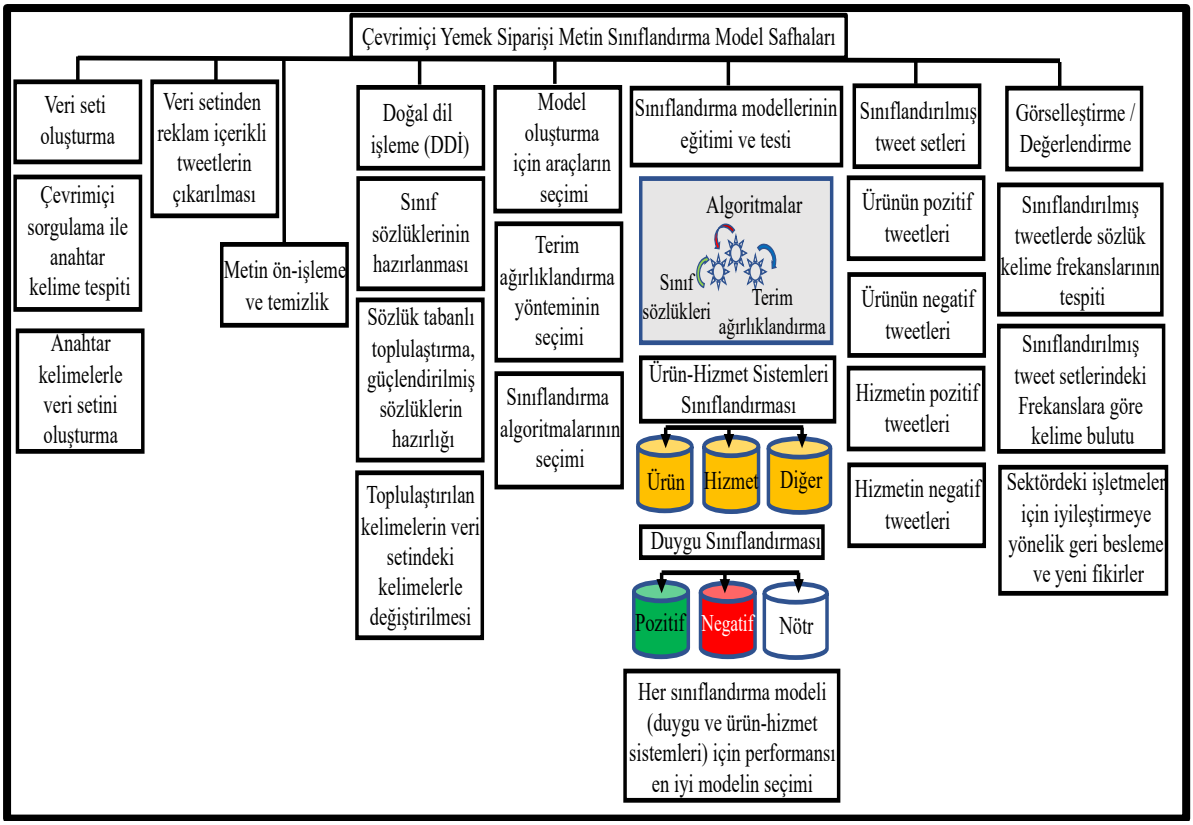
- Müşteriler tarafından restoranlar hakkında yapılan olumsuz yorumların yayınlanmadığı yönündeki algının önüne geçecek politikalar belirlenmesi, bunun kontrol mekanizması konusunda şeffaflık sağlayıcı uygulamalar geliştirilmesi, şikayette bulunan müşterilerin şikayet sonucuna ilişkin bilgilendirilmesi, televizyonlarda yayımlanan yemek programlarının, ÇYS konusunda reklamlar için kullanılması,

- Platformlar tarafından uygulanan joker menü, indirimli menü uygulamalarının restoranlar tarafından iptal edilmesinin önüne geçebilecek müşteri kontrollü kontrol mekanizması oluşturulması,

- Takvimdeki önemli gün, bayram, tatil, kişiye özel doğum günü gibi çok sık olmayacak şekilde sürpriz etkisi yaratacak hediye verilmesi konularını değerlendirerek işletme faaliyetlerini iyileştirebilirler.

3.3. İçerik/Metin Analizi

Çalışmada duygu ve metin sınıflandırma için sözlük tabanlı yaklaşım ve makine öğrenme algoritmalarının birlikte kullanımıyla oluşturulan yöntemin safhaları Şekil 3.29'da gösterilmiştir.



Şekil 3.29. Çevrimiçi yemek siparişine ilişkin tweet metinlerinin sınıflandırılması için önerilen model

Metin sınıflandırmada, oluşturulan modellerin performans ölçümü, metnin gerçek sınıfı (değerlendirici tarafından verilen sınıf etiketi) ile model tarafından yapılan sınıflandırmanın karşılaştırması ile tespit edilir. Analizin bu safhasında 35 428 tweetten oluşan koleksiyon, etiketleme için rastgele olarak iki gruba ayrılmıştır. Birinci grup koleksiyon 21 039 tweetten, ikinci grup koleksiyon 14 389 tweetten oluşturulmuştur. Birinci grup koleksiyon modelin geliştirilmesi, eğitilmesi ve test edilmesi için kullanılmış ve performansı en yüksek model tespit edilmiştir. Söz konusu model, sektördeki işletmelerin faaliyet alanlarına yönelik ne tür bilgilere ulaşılabileceklerini ve İG faaliyetlerini hangi alanlara yönlendirebileceklerini göstermek maksadıyla ikinci koleksiyon üzerinde kullanılarak sonuçlar sunulmuştur.

Modellerin performans ölçümü için, koleksiyonun farklı değerlendiriciler tarafından etiketleme işlemi yapılmıştır. Koleksiyon etiketleme işlemi için altı değerlendirici kullanılmıştır. Koleksiyon, değerlendiriciler tarafından hedef özellikler olarak ürün, hizmet ve diğer şeklinde ve duygu sınıflandırması için pozitif, negatif ve nötr şeklinde etiketlenmiştir. Tweetlerin elle etiketlemesi zaman alıcı bir işlemdir ve bu işlemin doğru yapılabilmesi için analiz amacının anlaşılması gerekir [371]. Öznel değerlendirmeye

dayanan bir tweetin düşünme yöntemi farklı olan kişiler [206, 273] tarafından değerlendirilmesi farklı etiketleme sonuçları üretebilir. Yiyecek sektöründeki cümlelerin yorumlanmasındaki güçlükler, konuşma dilindeki muğlaklık ve belirsizlikler nedeniyle, değerlendiriciler etiketleme işlemi hakkında bilgilendirme notuyla ve sözlü olarak bilgilendirilmişlerdir. Etiketleme işleminde kullanmak üzere değerlendiricilere; çalışma amacı, koleksiyon özellikleri ve sınıflandırmanın hedef özelliklerini içeren EK-7'deki bilgilendirme notu ile yiyecek sektörüne yönelik sınıflandırma çalışmalarından [1, 18-21, 292-294] çıkarılan hedef özellik ve alt başlıkları Çizelge 3.4'teki liste verilmiştir.

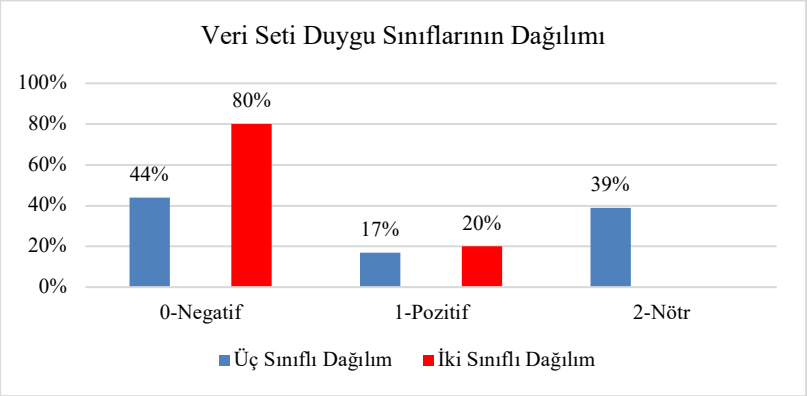
Çizelge 3.4. Ürün hizmet sistemleri sınıflandırması için alt konu başlıkları

Tweetlerin Ürün Hizmet Sistemleri Sınıflandırması İçin Alt Konu Başlıkları			
Ürün Kalitesi	Hizmet Kalitesi		
Tat ve lezzet	Hizmet koşulları	Çalışanların ilgi ve alakası/yardımseverliği	Şirket/işletme hakkındaki reklam yorumları
Sağlıklı alternatifler	Çalışan personel ve çalışma koşulları	Temizlik (Koronavirüs ve hijyen dahil)	Çevrimiçi hizmet
Ürün (menü) ve ürün çeşitliliği	Tutarlılık	Sektöre ilişkin reklamlar	Şikayetlere dönüş hızı, müşteri hizmetleri desteği
Tazelik	Paket	Servis hızı ve süresi, hava koşulları	İndirimler
Gıda/yiyecek güvenliği	Fiyat	Kurye ve kuryelerin çalışma koşulları	Promosyon uygulamaları
Olması gereken sıcaklık	Porsiyon yeterliliği	Müşteri hizmetleri personeli	Yemek kartları
Koku	Yabancı cisim varlığı	Sunum	---

Analizler için kullanılan koleksiyonun, sınıf dağılımlarında 1% ve altında olağanüstü dengesizlik oluşturacak bir sınıf bulunmamaktadır. Bu nedenle sınıflar arası dengeleme işlemi yapılmamıştır. Model oluşturmak için 21 039 tweetten oluşan birinci grup koleksiyon, çalışma kapsamında hazırlanan sınıflandırma sözlük ve algoritmaları birlikte kullanılarak üçlü (pozitif-negatif-neutral ve ürün-hizmet-diğer) ve ikili (pozitif-negatif ve ürün-hizmet) sınıflandırmalar yapılmıştır. Üçlü sınıflandırma modeli oluşturulurken 21 039 tweetin tamamı, ikili sınıflandırma modeli için 21 039 tweetten nötr ve diğer sınıf etiketine sahip tweetler çıkarıldıktan sonra kalanlar kullanılmıştır.

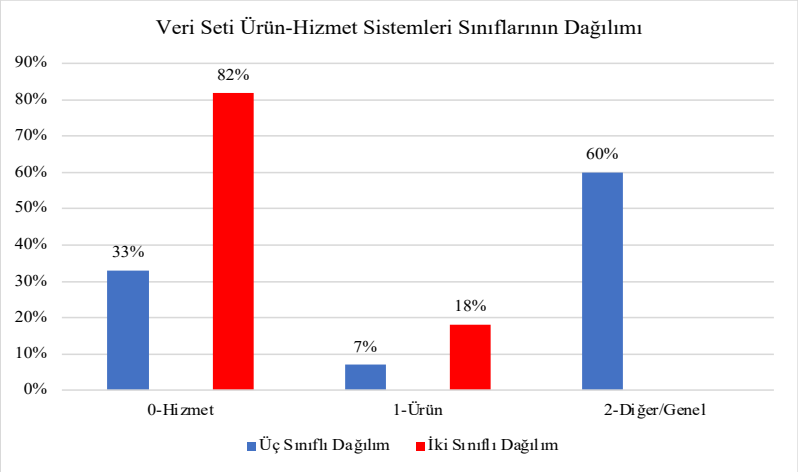
3.3.1. Koleksiyon içindeki sınıf dağılımları

Birinci koleksiyonda toplam 21 039 gözlem mevcuttur. Aşağıdaki grafiklerde bu gözlemlerin sınıflara dağılımları gösterilmiştir. Şekil 3.30 incelendiğinde analize tabi tutulacak koleksiyonun üç sınıflı duygu sınıflandırmasında, en düşük frekansa sahip pozitif sınıf oranının % 17 olduğu görülmektedir. Bu oranla koleksiyonun orta düzeyli dengesiz bir dağılıma sahip olduğu söylenebilir. Üçlü sınıflandırma içindeki nötr sınıfa ait veriler, koleksiyondan çıkarılarak iki sınıflı (pozitif ve negatif) bir yapıya dönüştürüldüğünde, düşük frekansa sahip pozitif sınıf oranı % 20 olmaktadır. Bu oranla koleksiyonun hafif düzeyli dengesiz bir dağılıma sahip olduğu söylenebilir [152].



Şekil 3.30. Veri seti duygu sınıfları dağılımı

Şekil 3.31 incelendiğinde analize tabi tutulacak koleksiyonun üç sınıflı ürün-hizmet sınıflandırmasında, en düşük frekansa sahip ürün sınıf oranının %7 olduğu görülmektedir. Bu oranla koleksiyonun orta düzeyli dengesiz bir dağılıma sahip olduğu söylenebilir. Üçlü sınıflandırma içindeki diğer sınıfa ait veriler, koleksiyondan çıkarılarak iki sınıflı (ürün ve hizmet) bir yapıya dönüştürüldüğünde, düşük frekansa sahip pozitif sınıf oranı %18 olmaktadır. Bu oranla koleksiyonun orta düzeyli dengesiz bir dağılıma sahip olduğu söylenebilir [152].



Şekil 3.31. Veri seti ürün-hizmet sistemleri sınıf dağılımı

3.3.2. Kelime frekans analizi

Reklam içerikli tweetler koleksiyondan ayıklandıktan sonra, EK-4'teki alana özel sözlükte yer alan kelimelerden tüm koleksiyondaki frekans sıralamasında ilk 75 sırada yer alanlar Çizelge 3.5'te ve Çizelge 3.5'in kelime bulutu olarak gösterimi Şekil 3.32'de verilmiştir.

Çizelge 3.5. Alana özel sözlük kelimelerinin koleksiyondaki frekansları

Kelime	Frekans	Kelime	Frekans	Kelime	Frekans	Kelime	Frekans
sipariş	14120	gelm	846	güzel	465	servis	575
yemek	13473	iste	821	dakika	457	kara kara düşün	432
yorum	4711	şimdi	570	sev	450	zorunda kal	431
saat	2159	adam	569	sokak	440	paket	412
yok	1840	online	548	telefon	816	kampanya	401
restoran	1436	iptal	540	zaman	812	hizmet	396
lahmacun	1222	dışarıdan sipariş	535	hasta	805	lira	383
kurye	1148	destek	533	döner	780	nöbet	380
burger	1078	pizza	528	öner	776	istem	372
tane	966	uygula	528	firma	772	müşteri	368
isti	912	pide	497	hamburger	692	verem	358
güven	895	tl	495	adres	606		
arkadaş	847	yemek yok	475	dk	600		

Şekil 3.32 incelendiğinde öne çıkan kelimelerin, yemek sipariş hizmetine ilişkin konular olmadığı görülmektedir.



Şekil 3.32. Alana özel sözlük kelimelerinin koleksiyondaki frekans kelime bulutu

Ancak koleksiyonu oluşturan tüm tweetlerde çok sık geçen kelimelerin bir ayırt ediciliği kalmayacağı, istatiksel açıdan yararı olmayacağı için koleksiyondan çıkarılması gerekir [372]. Ayrıca koleksiyon oluşturulurken kullanılan sorgulama kelimeleri de koleksiyondaki tweetlerin çoğunluğunda olacağı için ayırt edicilik niteliği olmayacaktır. Bu kapsamda Çizelge 3.3'te yer verilen kelimelerden *yemek, sipariş, döner, lahmacun, hamburger, burger, pide, dışarıdan sipariş, yemek yok, online sipariş* kelimeleri ayıklanmış ve bu ayıklama işleminden sonra hazırlanan kelime bulutu Şekil 3.33'te verilmiştir. Şekil 3.33'teki kelime bulutuna göre potansiyel müşterilerin gündeminde; (i) *restoran, kurye, firma, adres, sokak, servis, paket servis, hizmet, çalışan* gibi kelimelerle hizmete ilişkin konuların, (ii) *saat, zaman, dakika, şimdi, süre* gibi kelimelerle hizmetin süresine ilişkin konuların, (iii) *yok, iptal, hasta* gibi kelimelerle olumsuz öğelerin, (iv) *iste, sev, güven, arkadaş* gibi kelimelerle olumlu öğelerin öne çıktığı görülmektedir. Burada dikkat edilmesi gereken bir konu, kelime kökleri ile metin içinde kelime frekansları tespit edilirken kelime kök anlamının neden olabileceği yanlış yorumların önüne geçilmesi gerekir. Örneğin koleksiyon incelendiğinde *güven* kelimesinin işletme çalışanlarına duyulan güvensizlik anlamında kullanıldığı görülmektedir ve bu kelime sınıflandırmada tweetin negatif sınıfa ayrılmasına neden olacaktır.



Şekil 3.33. Alana özel sözlük kelimelerinin koleksiyondaki frekanslarına göre kelime bulutu-ayıklama sonrası

3.3.3. Sınıflandırmalarda uygulanan yöntem ve önerilen modeller

Koleksiyonun metin temizliği ve önişlemleri tamamlandıktan sonra model oluşturulması için şu adımlar takip edilmiştir. Çevrimiçi yemek sipariş alanına özel ürün ve hizmet sistemleri için 684 kelimeden, duygu sözlüğü için 769 pozitif ve negatif kelimeden oluşan iki ayrı sözlük hazırlanmıştır. Sözlük hazırlanmasında tohum kelime listesi ve veri seti derlemi şeklinde hibrit bir yöntem şu adımlarla kullanılmıştır: (i) Yemek siparişleri konusunda çeşitli sosyal ağ platformlarında ve sektördeki işletmelerin sayfaları incelenerek alana yönelik tohum kelime listesi hazırlanmıştır. (ii) Tohum kelime listesi, eş ve zıt anlamlılar sözlüğü kullanılarak yeni kelimelerle genişletilmiştir. (iii) Elde edilen kelime listesine jargon (motorcu, kuryeci, sepetçi gibi) ve argo kelimeler; kelimelerin en çok kullanılan çekim ve yapım eki almış halleri; kelimelerin çoğunlukla yanlış kullanılan hatalı yazım şekilleri dahil edilmiştir. (iv) Oluşan sözlüklerdeki kelimelerin veri seti içindeki frekansları tespit edildikten sonra frekansı sıfır olan kelimeler ve frekansı en yüksek olarak ortaya çıkan yemek ve sipariş kelimeleri sözlük dışına çıkarılmış, ancak *sipariş notu* gibi kelime grupları içinde yer alan sipariş ve yemek kelimeleri sözlükte yer almaya devam etmiştir.

Çalışmada için oluşturulan DTM'de başlangıç satırlarına metin önişleme ve temizlik işlemlerinden geçirilmiş her bir tweet cümlesi, başlangıç sütunlarına sınıflandırma için hazırlanmış sözlüklerdeki kelimeler, hücre içine kelimelerin terim ağırlıklandırma yöntemi ile elde edilmiş değerleri ve son sütuna tweet cümlesinin değerlendirici tarafından verilmiş sınıf etiketi yazılmıştır.

DTM'nin hücrelerine yazılacak sayısal değerlerin tespit edilmesi için, terim ağırlıklandırma metodu olarak CV, TF ve TF-IDF-ICF yöntemleri ile kullanılmıştır. Ağırlıklandırma yöntemlerinden CV, sözlükteki bir kelimenin analiz edilecek koleksiyondaki her tweet cümlesinde kaç kez kullanıldığını gösteren sayısal değerdir. TF olarak kullanılan değer ise sözlükteki bir kelimenin cümledeki tekrar sayısının cümlenin kelime sayısına bölünmesi ile elde edilen değerdir. TF-IDF-ICF yöntemi ise terimlerin sınıf frekanslarını da dikkate alan bir ağırlık hesaplama yöntemidir. Çalışmada koleksiyonun tamamı ürün, hizmet ve diğer şeklinde; duygu sınıflandırması için pozitif, negatif ve nötr şeklinde üç sınıflı olarak etiketlenmiştir. Etiketleme işlemi üçer sınıf olarak yapılmış olmakla birlikte sınıflandırmalardaki diğer ve nötr sınıfları için ayrıca sözlük oluşturulmamıştır. Bu nedenle algoritmalarla hem üçlü sınıflandırma sonuçları, hem ikili sınıflandırma sonuçları tespit edilerek üçlü ve ikili sınıflandırmaların başarı performansları tespit edilmiş ve karşılaştırılmıştır.

Algoritmaların çalıştırılması için koleksiyonun %30'u test ve %70'i eğitim seti olarak ayrılmış ve K-katmanlı doğrulama (K=10 olarak alınmıştır) yöntemi kullanılarak test hataları tespit edilmiştir. Algoritmaların çalıştırılması sonunda, her model için doğruluk oranları, F1 skorları, karışıklık matrisi tespit edilmiştir. Algoritmalar kullanılırken hiper parametre değerleri üzerinde farklı değer aralıkları ile optimizasyon işlemi de yapılmıştır.

Kelimelerin duygu puanlarının farklı düzeylerde bir araya getirilmesi genel bir yöntem olarak kullanılmaktadır. Toplulaştırma yöntemi aşağıda belirtilen şekilde, çalışma kapsamında hazırlanan temel sözlüklerde kullanılmış, temel sözlükler toplulaştırılan kelimelerin tek bir kelime ile temsil edilmesi şeklinde bir yöntem ile *güçlendirilmiş sözlük* adı verilen yeni bir sınıflandırma sözlüğüne dönüştürülmüştür. Sözlük düzeyinde yapılan bu toplulaştırma uygulamasıyla, sınıflandırmalarda etkisiz kalabilen jargon, argo ya da düşük değerli kelimelerin de hesaplamalara dahil edilmesi sağlanmıştır. (i) Veri setinde konuşma şekliyle ya da hatalı şekilde yazıya geçirilen kelimeler, kelimenin doğru yazımıyla grup yapılmıştır. Örneğin teşvik ve teşfik kelimesi aynı grupta yer almıştır, (ii) Eş anlamlı ya da metin içinde aynı anlamda kullanıldığı değerlendirilen kelimeler birlikte gruplandırılmıştır. Örneğin; alan, bölge, etraf, konum, civar, muhit, sokak, semt kelimeleri aynı grup içine alınmıştır, (iii) Aynı köke sahip kelimelerin farklı ekler almış halleri (ekler, ses düşmesi, sert harflerin yumuşaması ya da tersi gibi nedenlerle) tek bir grup olarak gruplandırılmıştır, (v) Grup haline getirilen kelimeleri temsilen, grup içinden bir kelime seçilmiştir. Temsilen

seçilen kelime, veri setinde temsil ettiği tüm kelimelerin yerine yazılmıştır. Benzer şekilde temsilen seçilen kelime, temel sözlüklerde yer almaya devam ederken, her grubun içinde yer alan diğer kelimeler temel sözlüklerden çıkarılmıştır. Böylece kelime sayısı daha az olan ve çalışma kapsamında güçlendirilmiş sözlük olarak isimlendirilen sözlükler elde edilmiştir. Sonuç olarak temel duygu sözlüğündeki kelime sayısı 105 azaltılarak 664 kelimelik güçlendirilmiş duygu sözlüğü; temel ürün-hizmet sistemleri sözlüğündeki kelime sayısı 93 azaltılarak 591 kelimelik güçlendirilmiş ürün-hizmet sistemleri sözlüğü elde edilmiştir. Boyut indirgemeyi sağlayan bu yöntemin bir kısıtlılığı, analiz öncesinde kelimelerin gruplandırılması ve grubu temsil eden kelimenin metinlerde değiştirilmesi şeklinde ilave bir işlem ortaya çıkarmasıdır. Güçlendirilmiş sözlük yapısı ile değişken görevindeki kelimelerin frekans ve ağırlık değerlerinin artırılması, kelime sayısının azaltılarak matris boyutunun küçültülmesi (boyut indirgeme işlemi) hedeflenmiştir. Geleneksel özellik seçme yöntemlerinde sınıflandırma için frekansı yüksek kelimeler daha etkili iken [130], frekansı düşük kelimelerin etkisi de düşük seviyede kalmaktadır. Güçlendirilmiş sözlük yapısı ile düşük frekanslı kelimelerin frekans değerleri yüksek frekanslı kelimelerle birleştirilerek aynı etki seviyesine çıkarılmaktadır. Özellik seçiminde önerilen kelime temsili ile güçlendirilmiş sözlük yapısı için, terim ağırlıklandırma formüllerinde kullanılacak kelime frekanslarının artırılması bir örnekle Çizelge 3.6'da gösterilmiştir.

Çizelge 3.6. Güçlendirilmiş sözlük için kelime temsili örneği

İşlem Sırası	Güçlendirilmiş Sözlük İçin Kelime Temsili Örneği		
1.	Temel duygu sözlüğündeki kelimeler	Toplulaştırılan kelimeler	Güçlendirilmiş duygu sözlüğünde toplulaştırılan kelimelerin yerini alan kelime
	Ödül Armağan Hediye	ödül, armağan, hediye	hediye

2.	Tweetler	Ön-işlemden geçirilmiş tweetler	Tweet-1'in güçlendirilmiş duygu sözlüğü ile uyumlaştırılması için kelime değiştirme işlemi
	Tweet-1	"dün gece etmeden aç aç yattığım kendimi *ödül* amaçlı kahvaltı pizza sipariş ettim"	"dün gece etmeden aç aç yattığım kendimi *hediye* amaçlı kahvaltı pizza sipariş ettim"
	Tweet-2	"kardeşim tatil *hediyesi* olarak etmiş aşko benzememeliydin"	"kardeşim tatil *hediyesi* olarak etmiş aşko benzememeliydin" (Tweet-2'deki kelime, toplulaştırılan kelimeleri temsil eden kelime olduğu için değiştirme işlemine ihtiyaç yoktur.)

3.	Temel duygu sözlüğü kelimelerinin koleksiyon içindeki frekansları	Güçlendirilmiş duygu sözlüğü kelimelerinin koleksiyon içindeki frekansları
	Ödül:47	---
	Armağan:2	---
	Hediye:160	Hediye: 209

Sınıflandırma modellerinde, terim ağırlıklandırma için Çizelge 3.7'de formülleri verilen CV, TF ve TF-IDF-ICF yöntemleri kullanılmıştır. Genellikle skor değerleri tweet metinlerinde diğer veri setlerine göre daha düşük gerçekleşir [77]. Kelimelerin gruplandırılması şeklinde uygulanan yöntemle, güçlendirilmiş sözlük yapısında kelimelerin ağırlık değerleri artırılmaktadır. Çizelge 3.6'da verilen kelimelerin güçlendirilmiş sözlük yapısı için kullanılan ağırlıklandırılmış değerlerinin hesaplaması bir örnek ile Çizelge 3.7'de gösterilmiştir.

Çizelge 3.7 Temel ve güçlendirilmiş sözlük kelimelerinin ağırlıklandırılmış değerleri arasındaki farkın bir örnek ile gösterimi

Terim Ağırlıklandırma Yöntemleri ve Temel ve Güçlendirilmiş Sözlük Terim Ağırlık Değerleri İçin Bir Örnek			
Terim Ağırlıklandırma Yöntemleri	Kelimeler	Temel Sözlük	Güçlendirilmiş Sözlük
CV $W_{CV}(t_i) = TF(t_i, d_j)$	hediye	160	209
	armağan	2	---
	ödül	47	---
TF $W_{TF}(t_i) = TF(t_i, d_j)/T_j$	hediye	0.00052	0.00068
	armağan	0.0000065	---
	ödül	0.00015	---
TF-IDF-ICF $W_{TF-IDF-ICF}(t_i) = \left(TF(t_i, d_j)/T_j\right) * \{1 + \log(\frac{D}{d(t_i)})\} * \{1 + \log(\frac{C}{c(t_i)})\}$	hediye	0.0016	0.0022
	armağan	0.000039	---
	ödül	0.000552	---
$TF(t_i, d_j)$: j dokümanındaki i teriminin frekansını gösterir.			
T_j: Koleksiyondaki toplam kelime sayısını gösterir.			
$d(t_i)$: t_i teriminin geçtiği doküman sayısını gösterir.			
$c(t_i)$: t_i teriminin geçtiği sınıf sayısını gösterir.			
D, koleksiyondaki toplam doküman sayısını; C, koleksiyondaki toplam sınıf sayısını gösterir.			
Formülde kullanılan kelimelere ait değerler: T_j=305 672; D=21 039; C=3'tür. Temel sözlük hesaplamasında "hediye" için $d(t_i)$ =147 ve $c(t_i)$ =3; "armağan" için $d(t_i)$ =2 ve $c(t_i)$ =1; "ödül" için $d(t_i)$ =44 ve $c(t_i)$:=3 değerleri kullanılmıştır. Güçlendirilmiş sözlük hesaplamasında "hediye" için $d(t_i)$ =193 ve $c(t_i)$ =3 değerleri kullanılmıştır. Bu hesaplama üçlü sınıflandırma için yapılmıştır.			

Başlangıçtaki temel sözlükler ile yapılan analizler, güçlendirilmiş sözlük yapıları ile tekrarlanmış ve tüm modellerde bu iki sözlük yapısının sonuçları karşılaştırılmıştır. Modellerin oluşturulması için LR, K-NN, doğrusal olmayan SVM, MLPC, GBM, XGBoost algoritmaları kullanılmıştır. Sınıflandırma algoritmaları, modellerin eğitimi için 10 katmanlı çapraz doğrulama yöntemi ile kullanılmış; modellerin en iyi performansının tespiti için hiper parametre iyileştirmesi yapılmıştır.

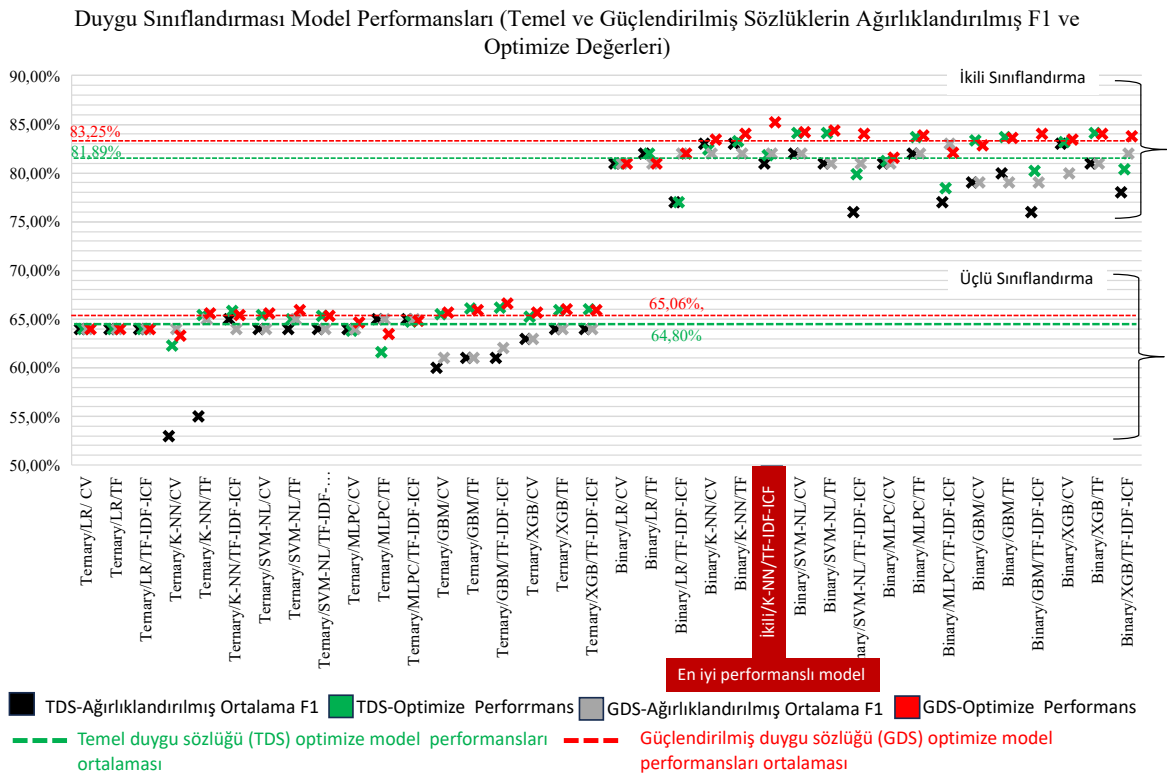
Duygu sınıflandırması ve ürün-hizmet sistemleri sınıflandırmaları için model geliştirmek maksadıyla; temel sözlük ve güçlendirilmiş sözlükler, üç terim ağırlıklandırma yöntemi, 6 sınıflandırma algoritması, ikili ve üçlü sınıflandırmalar ve model karşılaştırmaları yapılmıştır. Uygulama için MS Excel'in veri analiz özellikleri, Python programlama dili ve kütüphaneleri kullanılmıştır.

En iyi sınıflandırma modelinin seçimi için duygu sınıflandırmasında 72, ürün-hizmet sistemleri sınıflandırmasında 72 (üçlü ve ikili sınıflandırma * temel ve güçlendirilmiş duygu sözlükleri * 3 terim ağırlıklandırma yöntemi * 6 sınıflandırma algoritması) olmak üzere toplamda 144 model oluşturulmuştur.

Veri setindeki hafif düzeyli dengesizlikler de dikkate alınarak kurulan modellerin öncelikle ağırlıklandırılmış ortalama F1 değerleri ile performansları tespit edilmiştir. Hiper parametre iyileştirmeleri yapılmış modellerin performansları birbirleriyle karşılaştırılmıştır.

3.3.3.1. Duygu sınıflandırması

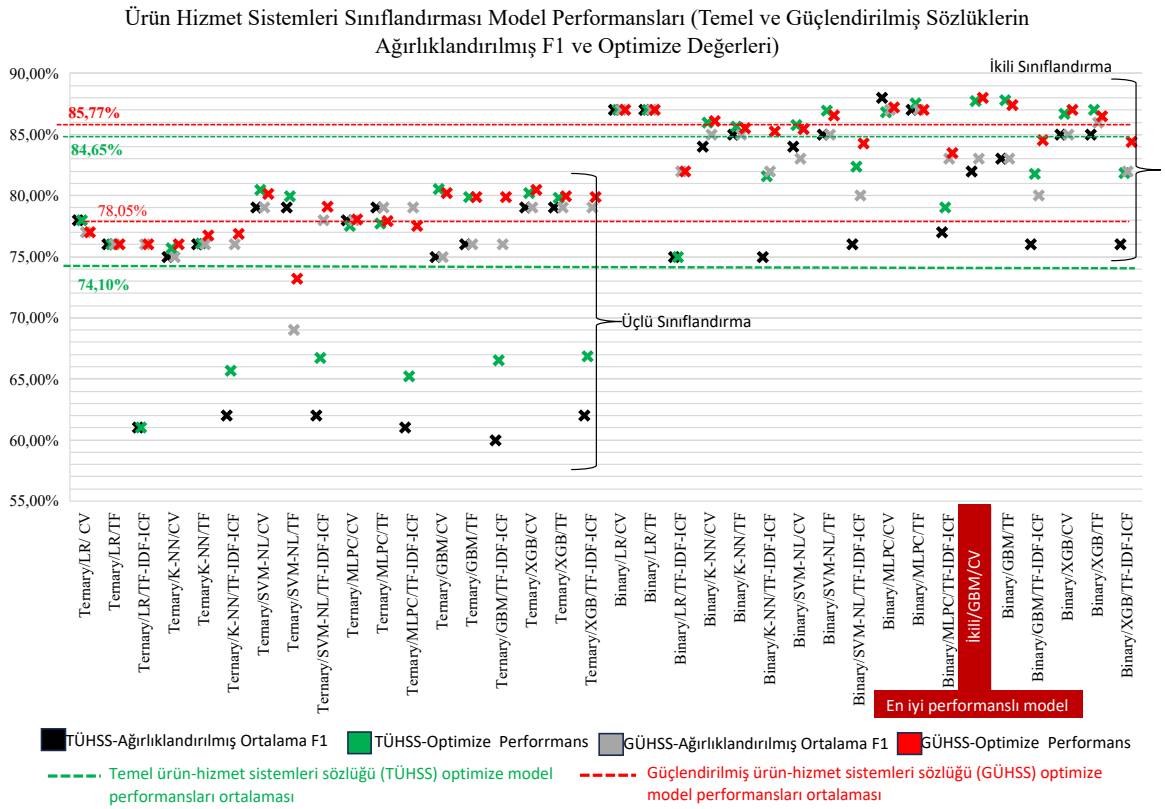
Duygu sınıflandırması için çalıştırılan modellerin performansı Şekil 3.34'te gösterilmiştir. Şekil 3.34'te gösterilen duygu modellerine ilişkin tespitler şunlardır: (i) Tüm ikili sınıflandırmalar, üçlü sınıflandırmalardan başarılıdır. (ii) Modellerin çoğunluğunda hiper parametre iyileştirmesi (hiper parametre iyileştirmesinden önceki performans siyah ve sonraki performans gri renkle gösterilmiştir) ile performansların artırılabildiği görülmüştür. (iii) Güçlendirilmiş duygu sözlüğü (ortalaması, kırmızı kesik çizgi ile gösterilmiştir) ile kurulan iki sınıflı modeller ortalama olarak, temel sözlük (ortalaması, yeşil kesik çizgi ile gösterilmiştir) ile kurulan iki sınıflı modellerden 1,36% daha üstün performans göstermiştir, bu üstünlüğün üç sınıflı modellerde 0.26% oranında devam ettiği görülmüştür. (iv) Duygu sınıflandırması için en yüksek performans 85,217% oranıyla, güçlendirilmiş duygu sözlüğü, ikili sınıflandırma, TF-IDF-ICF ağırlıklandırma yöntemi ve K-NN algoritması ile oluşturulan modelden elde edilmiştir.



Şekil 3.34. Temel ve güçlendirilmiş sözlüklerin duygu sınıflandırma performansları

3.3.3.2. Ürün ve hizmet sistemleri sınıflandırması

Ürün-hizmet sınıflandırması için çalıştırılan modellerin performansı Şekil 3.35'te gösterilmiştir. Şekil 3.35'te gösterilen ürün-hizmet sistemleri modellerine ilişkin tespitler şunlardır: (i) Tüm ikili sınıflandırmalar, üçlü sınıflandırmalardan başarılıdır. (ii) Duygu modellerinde olduğu gibi ürün-hizmet sistemleri modellerinin çoğunluğunda da hiper parametre iyileştirmesi ile performansların artırılabildiği görülmüştür. (iii) Güçlendirilmiş ürün-hizmet sistemleri sözlüğü (ortalaması, kırmızı kesik çizgi ile gösterilmiştir) ile kurulan iki sınıflı modeller ortalama olarak, temel sözlük (ortalaması, yeşil kesik çizgi ile gösterilmiştir) ile kurulan iki sınıflı modellerden 1,12% daha üstün performans göstermiştir, bu üstünlüğün üç sınıflı modellerde de 3.95% oranında devam ettiği görülmektedir. (iv) Ürün-hizmet sistemleri sınıflandırması için en yüksek performans 88% oranıyla, *güçlendirilmiş ürün-hizmet sistemleri sözlüğü, ikili sınıflandırma, CV ağırlıklandırma yöntemi ve GBM algoritması* ile oluşturulan modelden elde edilmiştir.



Şekil 3.35. Temel ve güçlendirilmiş sözlüklerin ürün-hizmet sistemleri sınıflandırma performansları

Farklı algoritma, terim ağırlıklandırma yöntemleri, ikili-üçlü sınıflandırma, duygu ve ürün-hizmet sistemleri sınıflandırmalarının eşleştirilmesi ile oluşturulan 144 model karşılaştırmasındaki en iyi sonuçlar Çizelge 3.6'da gösterilmiştir.

Kurulan modeller ağırlıklandırma yöntemi ve algoritmalar tüm sınıflandırmalardaki ortalama değerleri açısından incelendiğinde; (i) her üç terim ağırlıklandırma yönteminin en iyi sonuçları GBM algoritması ile elde ettikleri, (ii) TF yönteminin 79.3% ortalama ile en iyi ortalamaya sahip olduğu görülmüştür.

Çizelge 3.8'de kırmızı renkle gösterilen en iyi performansa sahip modelin yanı sıra, tablodaki diğer ikili sınıflandırma modellerinin de tatmin edici düzeyde performansa sahip oldukları ve sınıflandırma modeli olarak kullanılabileceği düşünülmektedir.

Çizelge 3.8. Duygu ve ürün-hizmet sistemleri sınıflandırması önerilen modellerin performans değerleri

Duygu ve Ürün-Hizmet Sistemleri Sınıflandırmaları Optimize Performansları			
Duygu Sınıflandırma Performansları			
Sınıf Sayıları	Terim Ağırlıklandırma Yöntemleri	Temel Güçlendirilmiş Sözlük (769 kelime)	Güçlendirilmiş Duygu Sözlüğü (664 kelime)
Üç Sınıf	CV	GBM: %65,53	GBM:%65,65
	TF	GBM: %66,11	XGB: %66,01
	TF-IDF-ICF	GBM: %66,18	GBM: %66,62
İki Sınıf	CV	NL-SVM: %84,15	NL-SVM: %84,19
	TF	NL-SVM/XGB: %84,15	NL-SVM: %84,39
	TF-IDF-ICF	KNN: %81,79	KNN: %85,22
Ürün-Hizmet Sistemleri Sınıflandırma Performansları			
Sınıf Sayıları	Term Weightening Method	Basic Product-Service Systems Lexicon (684 words)	Boosted Product-Service Systems Lexicon (591 words)
Üç Sınıf	CV	GBM: 80,52%	XGB: 80,49%
	TF	NL-SVM: 79,92%	XGB: 79,94%
	TF-IDF-ICF	XGB: 66,82%	GBM/XGB: 79,9%
İki Sınıf	CV	GBM: 87,77%	GBM: 88%
	TF	GBM: 87,82%	GBM: 87,43%
	TF-IDF-ICF	NL-SVM: 82,4%	KNN: 85,22%

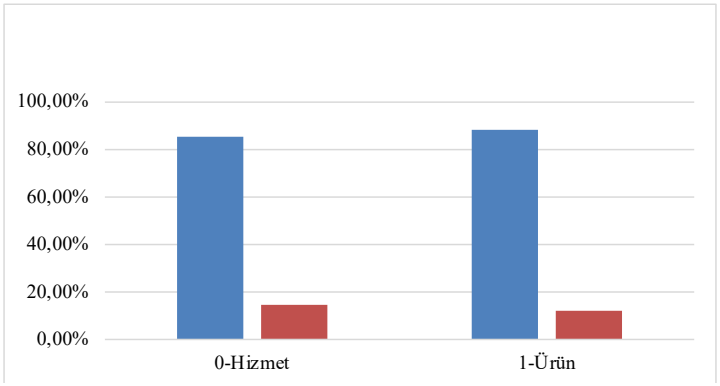
Çalışma kapsamında hazırlanan hem temel hem güçlendirilmiş sözlüklerin sınıflandırmalar için kullanılabileceği, güçlendirilmiş sözlüğün temel sözlükle elde edilen model performanslarını etkili şekilde artırdığı görülmüştür.

3.3.4. Önerilen sınıflandırma modellerin ikinci veri seti üzerinde kullanılması

Birinci grup veri seti ile uygulama sonucunda önerilen sınıflandırma modelleri; duygu sınıflandırmada, *güçlendirilmiş sözlük/ikili sınıflandırma/TF-IDF-ICF ağırlıklandırma yöntemi/K-NN algoritması* ve ürün-hizmet sınıflandırmada, *güçlendirilmiş sözlük/ikili sınıflandırma/CV ağırlıklandırma yöntemi/GBM algoritması* kombinasyonlarından oluşan modelleridir.

Önerilen söz konusu modeller kullanılarak sektörle ilgili ne tür çıkarımlar yapılabileceği, 14 389 tweetten oluşan ikinci grup veri seti ile gösterilmiştir. Bu maksatla ikinci veri seti, önerilen iki model ile sınıflandırılmış, sınıflandırma sonucunda ortaya çıkan dağılımlar,

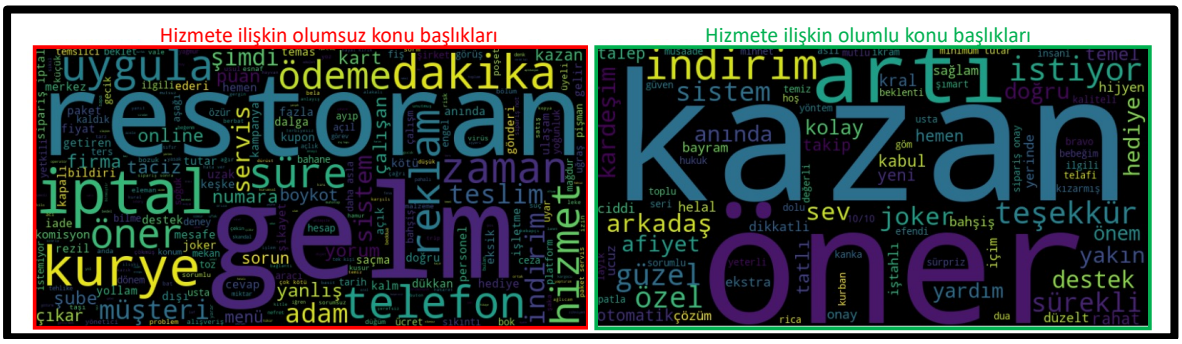
Şekil 3.36'da gösterilmiştir. Dağılım oranlarına göre ikinci veri setindeki tweetlerin 80% ve üzerinde bir oranla olumsuz görüşlerden oluştuğu görülmektedir.



Şekil 3.36. İkinci veri setinin en iyi modellerle sınıflandırması sonrasında ortaya çıkan sınıf oranları

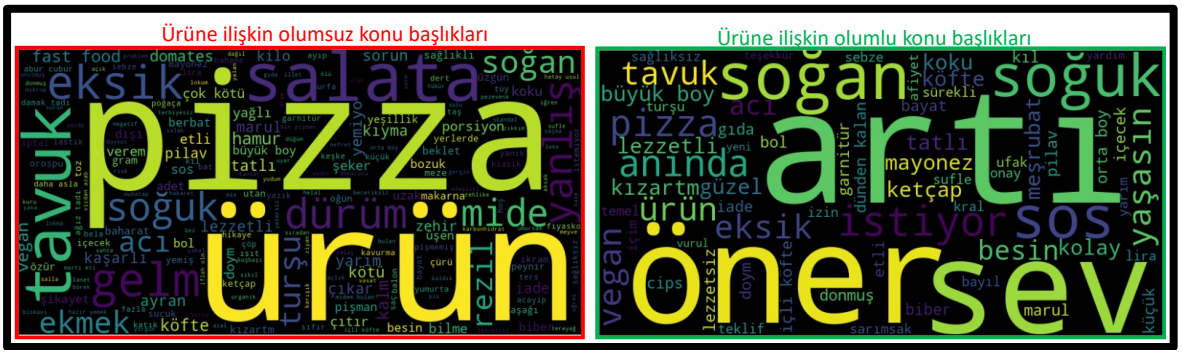
İkinci veri seti sınıflandırma işlemleri sonunda her sınıfta yer alan tweetlerdeki kelime frekansları tespit edilmiş ve bu frekanslara göre kelime bulutları oluşturulmuştur. Kelime bulutlarında öne çıkan konular, sektörde yer alan işletmelere, İG yönelik yol gösterici nitelikte sonuçlar sunabilmektedir.

Sınıflandırmalar sonucunda hizmet sınıfının negatif ve pozitif yönleri ile ilgili kullanıcılar arasında öne çıkan konular Şekil 3.37'de gösterilmiştir. Buna göre hizmet konusunda iyileştirilmesi gereken en önemli şikâyet konularının; restoranlar, kuryeler, reklamlar, iptal işlemleri, zaman-süre, ödeme ile ilgili işlemler, teslimat ve gelmeyen ürün ya da hizmet ile ilgili işlemler olduğu görülmektedir. Kullanıcıların hizmete ilişkin olumlu görüşlerinin ise kazanma, indirim, joker uygulamaları, öneriler, arkadaş konuları etrafında yoğunlaştığı görülmektedir.



Şekil 3.37. İkinci koleksiyonun sınıflandırılmasından sonra hizmet alanında öne çıkan negatif ve pozitif konular

Ürün sınıfının negatif ve pozitif yönleri ile ilgili kullanıcılar arasında öne çıkan konular, Şekil 3.38'de gösterilmiştir. Buna göre ürün kullanıcıların ürün konusundaki olumsuz görüşlerini, şikayetlerini; çoğunlukla pizza, ürün, tavuk, soğuk, acı, soğan, salata, eksik, yağlı, hamur, dürüm, kıyma, porsiyon, mide, çıtır, rezil kelimeleri ile ifade ettikleri görülmektedir. Bu konudaki olumlu görüşlerin ise artı, öneri, sevmek, tavuk, soğuk, büyük boy, lezzetli, anında, ürün, sos, ketçap, mayonez, kızarmış, besin, garnitür kelimeleri etrafında yoğunlaştığı görülmektedir.



Şekil 3.38. İkinci koleksiyonun sınıflandırılmasından sonra ürün alanında öne çıkan negatif ve pozitif konular

Sınıflandırmalar sonucunda öne çıkan olumlu ve olumsuz yönlerin, geliştirilmesine ihtiyaç duyulan konuların neler olduğuna, kelimeler üzerinden tweetler incelenerek karar verilebilir.

Boyut indirgeme için önerilen yöntemle oluşturulan güçlendirilmiş sözlükleri, sınıflandırma performansını artırmıştır. Ancak önerilen yöntemin bazı kısıtlılıkları da vardır. Yöntem, analiz işlemleri öncesinde ilave işlemler gerektirir. Kelime gruplarını temsil eden kelimelerin, veri setindeki grup kelimeleri ile değiştirilmesi gerekir.

Yöntem için kelimelerin gruplandırması, kelimelerin eş anlamları yanı sıra jargon, alana özel kullanımından kaynaklı anlam benzerlikleri ve imalı kullanımları da dikkate alınarak yapılmıştır. Dilin, doğal kullanımını etkili şekilde dikkate alan ve metin sınıflandırmaları için kullanışlı alan sözlükleri üretimine katkı sağlayabilecek bu yöntemin, elle gerçekleştirilmesi zaman alıcı bir işlemdir.

3.3.5. İçerik/metin analizi ile elde edilen bulgular

Çalışmanın ikinci aşamasında kullanılan içerik/metin analiz yöntemleri ile yapılan uygulamada elde edilen sonuçlar aşağıya çıkarılmıştır.

Ürün ve hizmet sınıflandırması konusunda başlangıç aşamasında ürün ve hizmet sistemleri sınıflandırması yerine ürün, hizmet, personel, fiyat, diyet ve sağlık başlıkları altında altı kategoride sınıflandırma yapılmış ancak tweet cümlelerinin kısa cümlelerden oluşması ve tweet cümlelerinde altı sınıf için yeterli sayıda ayırt edici kelime olmaması nedeniyle sağlıklı sınıflandırma yapılamadığı görülmüştür. Yine başlangıç aşamasında koleksiyonun analizi için ön hazırlık aşamasında ürün ve hizmet sistemleri kategorisi yerine "sipariş öncesi" ve "sipariş sonrası" şeklinde iki kategori şeklinde ayrım yapılması planlanmış ancak bu konuda da tweet cümlelerinin yeterli ayırt edici kelimeye sahip olmaması nedeniyle etiketlemenin yapılamadığı görülmüştür.

Çalışmada ÇYS konusunda sözlük tabanlı yaklaşım kullanılarak metin analizi yapılmıştır. Duygu ve ürün-hizmet sistemleri konusunda olmak üzere iki kademeli sınıflandırma modeli geliştirilmiş ve model önerisi sunulmuştur. Model oluşturmak için her iki kademede yapılan sınıflandırma için 72'şer model sonucu birinci koleksiyon üzerinde karşılaştırılmıştır.

Yapılan çalışma ile Türkçe literatürde yiyecek alanına özel Temel Duygu Sözlüğü, Temel Ürün ve Hizmet Sistemleri Sözlüğü ve bu sözlüklerden çalışma kapsamında geliştirilen güçlendirilmiş sözlüklerle sınıflandırmalar yapılmıştır.

Güçlendirilmiş sözlük yapısı ile öznitelik sayısının azaltılması, aynı zamanda özniteliklerin ayırt edici özelliklerini ortaya koyan terim ağırlıklarının frekans değerleri artırılarak yükseltilmesi hedeflenmiş ve modellerin çoğunluğunda temel sözlüklerden daha başarılı sonuçlar elde edilmiştir. Türkçe metin sınıflandırma sorunlarından biri, yapılan alan çalışmalarının yetersizliği ve bunun sonucundaki metin sınıflandırma başarılarındaki performansların düşük olmasıdır. Bu çalışma ile üretilen alana özel farklı sözlük yapıları ile sınıflandırma başarısının artırılması, çok boyutluluk ve seyreklik sorununun etkilerinin azaltılması sağlanmıştır.

Duygu sınıflandırmasında önerilen model; *ikili sınıflandırma, TF-IDF-ICF ağırlıklandırma yöntemi, kelime temsili yöntemi ile hazırlanmış güçlendirilmiş sözlük yapısı ve KNN algoritması* ile oluşturulan modeldir. Ürün hizmet sistemleri sınıflandırmasında önerilen model; *ikili sınıflandırma, CV ağırlıklandırma yöntemi, kelime temsili yöntemi ile hazırlanmış Güçlendirilmiş Ürün ve Hizmet Sistemleri Sözlüğü ve GBM algoritmasıdır.*
Önerilen bu modellerin birlikte kullanımıyla, çevrimiçi yemek sipariş alanına özel bir tweetin ya da metinsel başka bir ifadenin, ürün ya da hizmet konusunda memnuniyetlerinin otomatik olarak tespit edilmesi mümkündür. Bu modellerle bir tweetin öncelikle ürün, hizmet sınıflandırması yapılır. Ürün ve hizmet olarak ayrılmış aynı tweetlerin pozitif ve negatif şeklinde ikinci bir sınıflandırılması ile müşterilerin sektöre ilişkin memnuniyetleri tespit edilmiş olur. Bu alandaki bir işletme yapılan bu tür bir tespitle, faaliyetlerinin hangi konuda geliştirmeye muhtaç olduğunu, hangi konuda iyi olduğu için bu yöndeki faaliyetlerinin seviyesini koruması gerektiğini belirleyebilir.

Önerilen sınıflandırma modelleri, geliştirilen alan sözlükleri, çevrimiçi yemek siparişleri konusunda müşteri geri bildirimlerinin ürün ve hizmet hedef özellikleri çerçevesinde duygu analizi yapılarak müşteri memnuniyetinin ölçülmesine, elde edilen sonuçlarla sektöre yön verebilecek geliştirmeye muhtaç olan alanların tespitine imkân sağlar.

Çalışma kapsamında yemek sipariş alanına özel, biri temel diğeri güçlendirilmiş duygu ve ürün-hizmet sistemleri sınıflandırma sözlükleri olmak üzere dört adet sözlük hazırlanmıştır. Güçlendirilmiş sözlüğün temel sözlüğe; ürün-hizmet sistemleri sınıflandırmalarının duygu sınıflandırmalarına oranla hem üçlü hem ikili sınıflandırmalarda daha iyi performans sergilediği görülmüştür.

Duygu sınıflandırmasında, çalışma kapsamında hazırlanan her iki sözlükten kelime temsiline dayalı güçlendirilmiş sözlük yapısının hem üçlü hem ikili sınıflandırmada en başarılı sonuçları ürettiği görülmüştür. Güçlendirilmiş sözlük yapısının ürün hizmet konusundaki ikili sınıflandırmada genel olarak üstün olduğu, üçlü sınıflandırmada ise TF-IDF-ICF ile kullanıldığında üstün, CV ve TF ile kullanıldığında temel sözlük yapısı ile çok yakın sonuçlara ulaşıldığı görülmüştür.

Terim ağırlıklandırma yöntemlerinden, TF'nin en iyi ortalama performansa sahip olduğu görülmüştür. GBM, XGB, KNN, Doğrusal Olmayan SVM algoritmalarının genellikle

yüksek performans sergilediği görülmüştür. Literatürde temel sınıflandırma algoritmaları olarak kabul edilen KNN, SVM gibi algoritmalardan [373] farklı olarak, son dönem çalışmalarında yüksek performans sağlayan GBM'in bu çalışma kapsamında kurulan tüm modellerde en iyi ortalama performansa sahip olduğu görülmüştür. Genel olarak ikili duygu sınıflandırmaları, üçlü duygu sınıflandırmasına göre daha yüksek performans sergilemiştir. Aynı durum ürün-hizmet sistemleri sınıflandırmasında da görülmüştür. Ayrıca ikili ürün-hizmet sistemleri sınıflandırması, ikili duygu sınıflandırmasına göre de daha yüksek performans göstermektedir.

Önerilen sınıflandırma modellerinin, ikinci koleksiyon üzerinde kullanılması ile işletmeler için doğrudan İG yönelik fikirler tespit edilmiştir.

3.4. Ağ Analizi

Sosyal ağ analizleri, analizin amacına göre farklı aktörlerin ağ içindeki konumlarının tespit edilmesi için yapılabilir. Örneğin sosyal medya fenomenini araştırmaya yönelik analizler, ağdaki aktörleri merkeze alarak inceleme yaparken; ticari bir ürün ya da hizmete ilişkin analizler ağ içindeki müşteri gruplarının tespitini analizin odağı olarak belirleyebilir. Buna başka bir örnek, politik konulara ilişkin bir çalışmada, toplum algısını etkilemek için hangi aktörlerin kullanılabileceğinin tespiti ya da kamuoyunda ortaya çıkarılmış olan bir algının hangi aktörler tarafından sosyal medyada yaratıldığının tespit edilmesi ağ analizine konu olabilir. Diğer bir örnekte ise sosyal medyadaki çevrimiçi etkileyicilerin takipçileri ile birlikte oluşturduğu ağlardan hangilerinin daha güçlü ve etkili bir yapıya sahip olduklarının tespit edilmesi ve bu tespit sonucuna göre ticari bir ürün ya da hizmet için reklam verilebilecek çevrimiçi etkileyicilerin tespit edilmesi amaçlanabilir.

Twitter (X) ya da diğer sosyal medya platformlarında yer alan etkili aktörlerin, takipçi sayıları ya da paylaşımları ile etkinlikleri bilinmektedir. Ancak, sosyal ağlarda anahtar kelimelerle yapılmış arama sonuçları ile oluşturulmuş veri setlerindeki aktörlerin durumu koleksiyona özel olarak ayrıca analiz edilmelidir. Çünkü bu tür veri setleri, aktörler arasındaki ilişkiler üzerinden gidilerek oluşturulmuş ilişkisel veri setleri değildir ve aktörler açısından durumun tespit edilmesi gerekir.

Literatürde Twitter (X) içeriklerinden elde edilen verilerin analizine yönelik çalışmalar yapılmakla birlikte, Twitter (X) verisinin ağ ilişkileri açısından incelenmesi konusunda az sayıda çalışma vardır. Bu çalışma ile Twitter (X)'dan elde edilmiş olan koleksiyondaki Twitter (X) kullanıcıları arasında ne tür ağ ilişkilerinin olduğu, bu ağ ilişkilerinin hangi yöntem ve algoritmalarla analiz edilerek görselleştirilebileceği ve söz konusu analizlerle ne tür sonuçlar elde edilebileceği gösterilmiştir. Uygulama için, metin analiz bölümünde kullanılan iki koleksiyonun tamamı tek bir koleksiyon olarak kullanılmış, bu veri seti içinden seçilen kullanıcılar üzerinde ağ analizi yapılmıştır.

Genelde tüm ağ seviyesinde yapılacak analizlerde, ilişkisel verilerin kullanılması hedeflenir İlişkisel veriler, aktörler arası temas ve bağlantıları gösterir [374] ve bu ilişkiler üzerindeki ağ analizleri ilişkilerin açıklanmasına imkan verir. Çalışmada kullanılan veri seti ilişkisel veri toplama yöntemiyle elde edilmeyip, anahtar kelime kullanılarak oluşturulmuştur. Ancak bu tür veri setlerinde seçilen aktörler üzerinden yapılacak ağ analizinin ve seçilen aktörlerin ağ içindeki ilişkileri açıklamaya ne derece etki edeceği, ilişkilerin açıklanmasında ne derece katkı sağlayabileceği bilinememektedir [295].

Anahtar kelimelerle bir araya getirilen tweetlerin, sektörün etkili kullanıcılarını bir araya getirip getiremediği ya da aktörler arasında bir ilişkinin varlığından söz edilemediği için, çalışmada öncelikle veri setindeki etkili aktörler belirlenmiştir. Sonrasında bu aktörlerin düğüm merkezli analizi yapılmış, aralarındaki ilişkilerin gücü, etkisi ve topluluk oluşturma eğilimleri ile ağ yapılarının ortaya çıkarılması hedeflenmiştir. Etkili aktörler belirlenirken, merkezilik derecelerinin dengeli şekilde hesaplanması için yöntem belirlenmiştir. Bu maksatla yerel ve küresel merkezilik değerleri hibrit bir yaklaşımla kullanılmış, merkezilik derecelerinin faklı özellikleri arasında dengeli bir hesaplama yapılmıştır. Altı merkezilik derecesi arasında dengeli bir hesaplama ağırlıklı ortalama ile yapılmıştır. Ağırlıklı ortalama ile belirlenen etkili kullanıcı sıralaması, ayrıca kullanıcılar için hesaplanan ortalama ve medyan değerlerine göre yapılan sıralamalarla karşılaştırılmaktadır.

Veri seti içindeki etkili aktörlerin tespiti için, merkezilik dereceleri belirlenmiştir. Etkinlik tespit edilirken aktörün yakınlık ve arasındalık gibi yerel bilgileri ya da özvektör ve Katz merkezilikleri gibi ağın küresel bilgileri kullanılır. Etkili bir yöntem olmakla birlikte merkezilik derecelerinin bazı sorunları ve çelişkileri de bulunmaktadır. Örneğin derece merkeziliği ve küme sıralaması gibi yerel merkezilik ölçütlerinin, kullanıcının genel

(küresel) ağ içindeki bağlantısını göz ardı etmesi, elde edilecek sonucu zayıflatabilir [306]. Derece, yakınlık, arasındalık merkeziliklerinin, bazen aynı düğüm için farklı merkezilik derecesi üretebildiği de görülmektedir. Bir ölçümde güçlü çıkan düğümün farklı ağ yapılarında zayıf kalabildiği, zayıf ölçülen bir düğümün aradan çekilmesi ile ağın başka bir alt grupla olan bağı kopardığı görülebilmektedir [311, 314]. Nitekim etkili düğüm tespitinde farklı topolojik ağ yapılarının da dikkate alınması gerektiği vurgulanmaktadır [331]. Mevcut durumuyla veri setinde, konuya ilişkin görüşlerini ifade eden kullanıcılar arasında herhangi bir ilişkinin varlığı, sohbet ortamını yönlendirici etkiye sahip kullanıcıların olup olmadığı, paylaşılan görüşlerin gerçek müşterilere ait olup olmadığı bilinmemektedir. Bu bilgilerin tespit edilmesi, konuyla ilgili kullanıcıları etkileme ve eğilimleri yönlendirme gücüne sahip Twitter (X) kullanıcı ve topluluklarının tespit edilmesi, bu kişi ve toplulukların promosyon, pazarlama, sektörel imajın güçlendirilmesi gibi maksatlarla kullanılması ve gerçek kullanıcılara ait geri bildirimlere ulaşılması açısından önemlidir [375].

Çalışmada, Twitter (X)'dan ÇYS konusunda 1 Ocak-31 Aralık 2021 tarihleri arasında yapılmış olan sohbetlerden tanımlayıcı analizlerle tespit edilen kullanıcılar arasındaki sohbet ağının, farklı algoritmalarla ağ analizi ve görselleştirilmesi yapılmış, kullanılan algoritmalar arasındaki benzerlik ve farklılıklar uygulama sonuçları ile incelenmiş ve koleksiyondaki düğüm noktasını oluşturan aktörlere ve aralarındaki iletişime yönelik çıkarımlar yapılmıştır. Söz konusu koleksiyondan ağ analizi için kullanılacak alt veri seti, Twitter (X) görünürlük ve etkinlik ölçütleri esas alınarak oluşturulmuş ve belirlenen 66 kullanıcı ismi Çizelge 3.9'da verilmiştir. Çalışmanın ileriki bölümlerinde kullanıcı isimleri yerine, Çizelge 3.9'daki kullanıcılara verilen numaralar kullanılmıştır.

Çizelge 3.9. Ağ analizinde kullanılan Twitter (X) kullanıcı numara ve isimleri

Ağ Analizinde Kullanılan Twitter (X) Kullanıcıları
0:@666lehce, 1:@abdulkadir06161, 2: @abdullahciftcib, 3: @adilomera1, 4: @afmsudan, 5: @AliYerlikaya, 6: @amanda_devik, 7: @artdbkglu, 8:@Asibelblonde, 9: @avozgeustun, 10: @ay_moon_ay, 11: @ayse_cng, 12: @bangtantr, 13: @besisko, 14: @bilgingokberk, 15: @burcincigerim, 16: @Can_Y_Can, 17: @cemreallly, 18: @CEngven, 19: @chabyhan, 20: @covid20loading, 21: @ctetikog, 22: @damlaaltuun, 23: @deepturkweb, 24: @delykalamis, 25: @dominoeffect34, 26: @enveryan, 27: @er_maannn, 28: @gela_wej, 29: @gtufekci, 30: @hallederiztamam, 31: @handebbyrktr, 32:@handelendi, 33: @hasankarsavuran, 34: @hatirliyoruz, 35: @ISTANist, 36: @ibrahimbagyapan, 37: @ihsanerkahraman, 38: @izmirevdekal, 39: @kagansaydan, 40: @KOXALBEY, 41: @lordsinov, 42: @mbagun11, 43: @mert_abras, 44: @merveyldrm2525, 45: @mervezell, 46: @mizantropii, 47: @mukinaber, 48: @ninegeek9, 49: @oguzhantesla, 50: @ozelegitimci_18, 51: @pillreder, 52: @saddandmad, 53: @sakizciko, 54: @sinoreyiz, 55: @srfzmr, 56: @thewizarddofoz, 57: @tinercimayki, 58: @toskofacts, 59: @UstAkilOyunlari, 60: @Wotah_, 61: @yarftm, 62: @yemeksepeti, 63: @yokmaalesef, 64: @zagortenay76, 65: @zehra_mete

Çalışma ile sosyal medya üzerinden elde edilecek veri setleri ile yapılacak analizlerde, veri setlerinin doğru yorumlanabilmesine ağ analizi ile yapılabilecek katkıların ortaya konulması hedeflenmiştir. Bu maksatla çalışmada tespit edilen 66 kullanıcı için düğüm seviyesi ve topluluk seviyesi analizler yapılmış ve farklı algoritmalar kullanılarak görselleştirilmiştir. Kullanılan algoritmalar arasındaki benzerlik ve farklılıklar uygulama sonuçları ile incelenmiş ve veri setindeki düğüm noktasını oluşturan kullanıcılara ve aralarındaki iletişime yönelik çıkarımlar yapılmıştır.

Ağ analizi için kullanıcılar, tanımlayıcı analiz ile tespit edilen aktör listelerinden yapılan seçimle belirlenmiştir. Analize konu edilen aktörler arasında, koleksiyonun iletişim kanalı olarak RT-bahsetme, favoriye alma özellikleri kullanılmıştır. Çalışmada, düğüm seviyesi ve topluluk seviyesi analizi yapılmıştır [3].

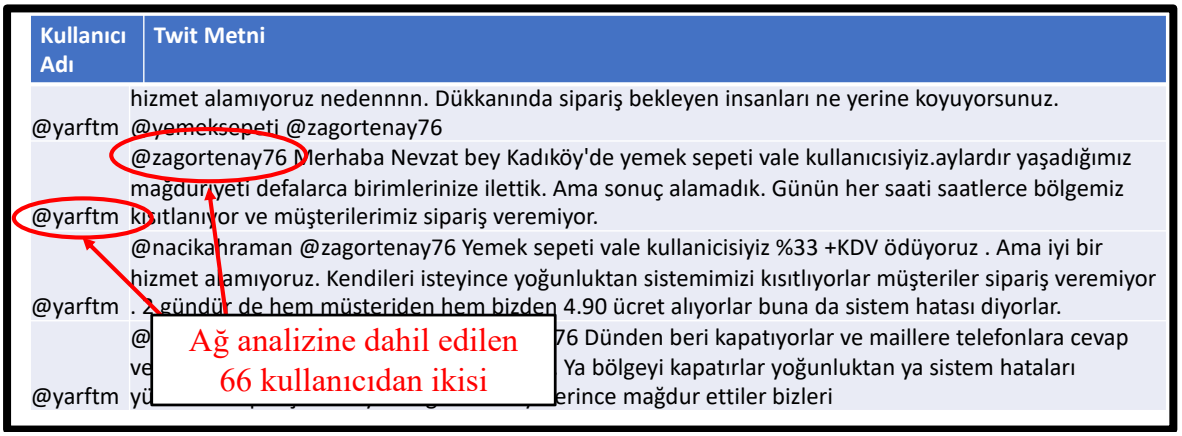
Ağ analizi için yaygın olarak; derece, arasındalık, yakınlık, özvektör ve Katz merkezilikleri, sayfa sıralaması algoritması, Louvian metodu, Girvan-Newman algoritması, etiket yayılım algoritması, grafik oluşturucular, mesafe algoritmaları, bağlantı algoritması, kümeleme algoritmaları, bağlantı öngörü algoritması, Fruchterman-Reingold algoritması, Harel-Koren hızlı-çok ölçekli grafik algoritmaları kullanılmaktadır [340].

Çalışma kapsamında ağ analizleri için derece, arasındalık, yakınlık ve özvektör, sayfa sıralaması, katz merkezilikleri [376], Harel-Koren [377], Fruchterman-Reingold [378], Girvan-Newman, etiket yayılım, bağlı bileşenler, kümeleme katsayısı, Jaccard katsayısı algoritmaları kullanılmıştır.

Algoritmaların ağ analizine yönelik kullanılabilmesi için matris yapının elde edilmesi gerekir. Komşu matrisler (adjacency matrices) olarak adlandırılan yapı, aktör sayısının karesi büyüklüğündedir. Matris yapının satır ve sütun başlıkları aktörlerin isimlerinden, matrisin hücre içinde ise aktörler arasında ilişkinin varlığı ve kuvvetini göstermektedir [295].

Ağ analizi için matrisler, Twitter (X) kullanıcıları arasındaki retweet ve bahsetme ilişkileri üzerinden oluşturulmuştur. Çalışmada ağ analizine yönelik matris yapının oluşturulması maksadıyla sırasıyla şu işlemler gerçekleştirilmiştir: (i) En çok RT alan ilk 25 tweetin kullanıcı isimleri, en çok RT'leme yapan 25 kullanıcı ismi ve 1000'den daha fazla favoriye alınan/beğeni alan tweetlerin kullanıcı isimleri birleştirilerek bir liste hazırlanmış, (ii) birden

fazla tekrar eden kullanıcı isimlerinin tekrarları listeden ayıklanmış ve 66 kullanıcı ismi belirlenmiş, (iii) Kullanıcı adı ve Tweet metni şeklindeki (35 428 tweetten oluşan) Excel tablosu her iki sütununda aynı anda 66 isimden birini içerecek şekilde (@kullanıcıismi şeklinde yazılarak) filtrelenmiş, böylece 66 kullanıcının birbirleri hakkındaki/arasındaki sohbeti içeren tweetlerden oluşan alt koleksiyon oluşturulmuş, alt koleksiyonun dört satırlık örneği aşağıda Şekil 3.41'de gösterilmiş, (iv) satır ve sütun isimleri 66 kullanıcı isminden oluşan 66x66'lık bir matris yapı oluşturulmuş, (v) matristeki hücrelere Şekil 3.39'da örneği gösterilen alt koleksiyondan kullanıcılar arasında tespit edilen iletişim frekansları (RT/bahsetme sayıları) yazılmıştır. Matris yapı için alt koleksiyonu, 66 kullanıcının 4 728 tweeti oluşturmuştur.



Şekil 3.39. Kullanıcılar arasındaki iletişim-bahsetme tweeti

Matris yapıda hücre içine yazılan değer, şu şekilde bir örnekle açıklanabilir. Şekil 3.39'da kesiti verilen koleksiyon satırlarında @yarftm kullanıcısının @zagortenay76 kullanıcısından bahsettiği bir tweeti görülmektedir. Tüm koleksiyon içinde bu şekilde iki kullanıcı arasında toplam 8 bahsetme iletişimi tespit edilmiştir. Sonuç olarak matris yapıda iki kullanıcı isminin kesişimine 8 rakamı yazılmıştır. Söz konusu işlemlerden sonra ağ analizi için elde edilen komşuluk matrisinden bir kesit Çizelge 3.10'da gösterilmiştir.

Çizelge 3.10. Komşuluk matrisinden bir kesit

	@UstAkilOyunlari	@Wotah_	@yarftm	@yemeksepeti	@yokmaalesef	@zagortenay76	@zehra_mete
@UstAkilOyunlari	0	0	0	0	0	0	0
@Wotah_	0	0	0	0	0	0	0
@yarftm	0	0	0	0	0	8	0
@yemeksepeti	0	0	1	0	0	11	1
@yokmaalesef	0	0	0	0	0	0	0
@zagortenay76	0	0	0	0	0	0	0
@zehra_mete	0	0	0	2	0	0	0

Ağ ilişkisi hesaplamalarının yapılması ve görselleştirilebilmesi maksadıyla yazılım ile uygulama yapabilmek için, düğümler arasındaki ilişkileri gösteren Çizelge 3.8'deki gibi bir matris yapıya ihtiyaç vardır. 66 Twitter (X) kullanıcısının oluşturduğu ağ yapısının özellikleri, merkezilik ve topluluk oluşturma durumları ile kullanıcılara ilişkin diğer özellikleri incelenerek takip eden maddelerde sıralanmıştır.

3.4.1. Twitter (X) kullanıcılarının merkezilik durumları

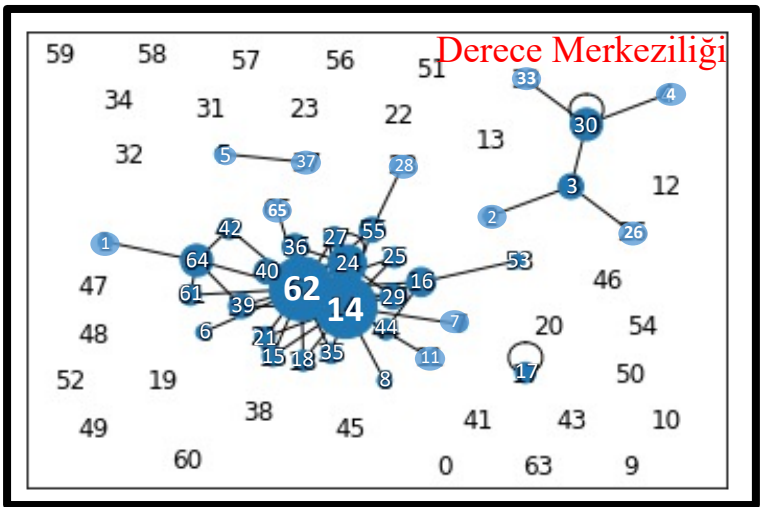
Merkezilik dereceleri, düğümlerin göreceli olarak önemini belirleyen bir ölçüttür. Ağdaki merkezi düğümler, ağdaki kritik düğümlerin belirlenmesi için kullanılır. Kritik düğümler ise ağ yapısından çıkarıldığında ağ içindeki iletişimi önemli derecede olumsuz etkileyen düğümlerdir. Merkezi düğümlerin tespit edilmesi için en çok kullanılan merkezilik ölçütleri derece merkeziliği, arasındalık ve yakınlık merkeziliğidir. Bu ölçütlerin farklı avantaj ve dezavantajları vardır. Derece merkeziliği grafikteki sadece yerel bilgileri kullanır ve hesaplamalar açısından karmaşıklığı düşüktür; diğer yandan yakınlık ve arasındalık merkeziliği grafiğin genel yapısına ait bilgileri kullanır ve hesaplamalar açısından daha karmaşıktır [379]. Ayrıca özvektör merkeziliği ve sayfa sıralaması, düğümün bağlandığı düğümlerin de önemini hesaplamalara dahil eden merkezilik ölçütlerindendir. Hesaplamalarda ağ yapısının farklı değerlerini kullanarak farklı özelliklerini öne çıkaran merkezilik ölçütlerinden birine göre kritik olan düğüm, diğer bir ölçüt tarafından tespit edilmeyebilir.

Sosyal ağ analiziyle, ağlarda etkin olan ana aktörler ve bu ana aktörlerin rollerinin tespit edilmesi sağlanmaktadır. Sosyal medya ağlarında gündemin birden fazla belirleyicisi olabilmektedir [317]. Ağ içinde güçlü aktörlerin etkisi bilinmekle birlikte, zayıf bağların da güçlü bağlardan daha fazla sayıda aktöre ulaşarak etkili olabildiği belirtilmektedir [380]. Bu

durumun Twitter (X)'daki zayıf bağlarda da görülebildiği, zayıf bağlara sahip aktörlerin bilgi yayılımı açısından yüksek etkiye sahip olabildikleri belirtilmektedir [262]. Tüm bu aktörler farklı özellikleri nedeniyle ağ içinde etkinlik gösterebilirler ve bu aktörler farklı merkezilik ölçütlerinin kullanılması ile ortaya çıkarılabilir.

Derece merkeziliği

En temel merkezilik ölçütlerinden biri olan derece merkeziliği, hesaplaması en kolay ölçütlerdendir. Düğümün sahip olduğu bağlantı sayısı, düğümün derecesini verir, derece ne kadar yüksekse düğüm o kadar merkezi konumdadır. Derece merkeziliği etkili bir ölçüm aracıdır, çünkü yüksek dereceye sahip birçok düğüm, diğer ölçütlere göre de yüksek merkeziliğe sahiptir [381]. Şekil 3.40'ta (algoritma grafik düzeninin kodlama ismi:neato) kullanıcıların ağ içindeki derece merkezilikleri, merkezilik değerlerinin yüksekliği ile orantılı büyüklükte gösterilmiştir. 62, 14, 24, 30, 64, 16, 3, 29, 36 ve 39 numaralı kullanıcılar, ağdaki en yüksek derece merkeziliğine sahip kullanıcılardır. Bu kullanıcılar, bilgiye ulaşma ve kendine ulaşan bilgiyi daha kolay yayabilme güçleri nedeniyle ağdaki en etkili kullanıcılardır.

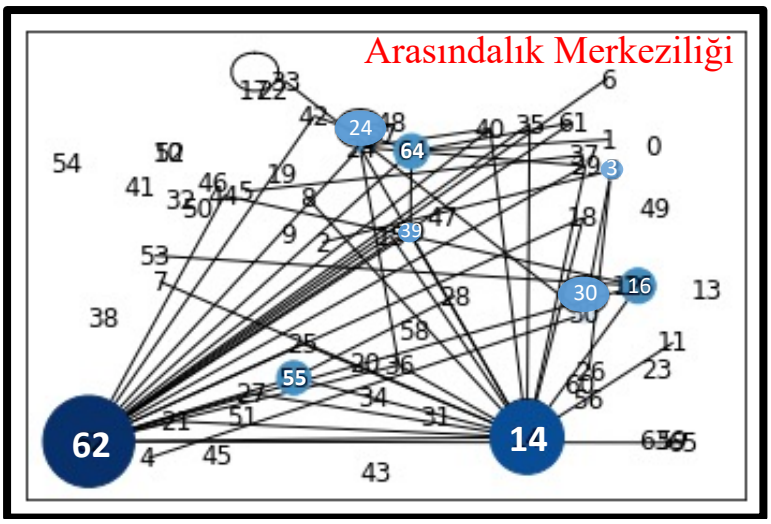


Şekil 3.40. 66 Twitter (X) kullanıcısının derece merkeziliği

Arasındalık merkeziliği

Arasındalık merkeziliği, birbiri ile bağlantısı olmayan düğümlerin birbirine bağlanmasını sağlar, bağlantısız düğümlerin arasında köprü görevi görür. Böylece bağlantısız düğümler arasında bilgi akışı sağlanmış olur ve bu bilgi akışı üzerinde kontrol bu bağlantıyı sağlayan düğümdedir [382]. Bu bağlantıları sağlayan en kısa yol üzerinde yer alan düğümlerin arasındalık merkeziliği yüksektir. Arasındalık merkeziliği ağ yapısı içinde dar boğaz etkisi gösteren düğümlerin tespit edilmesini sağlar.

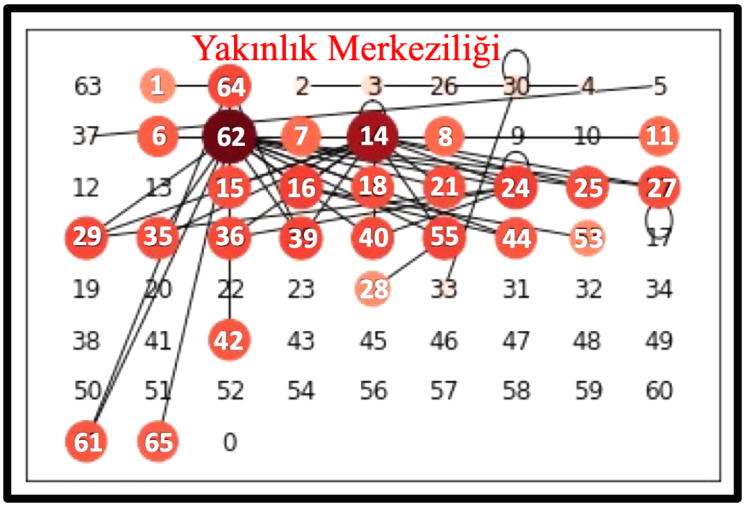
Kullanıcıların, arasındalık merkezilik değerlerine göre ağ içinde konumlanışları Şekil 3.41'de (algoritma grafik düzeninin kodlama ismi: random) verilmiştir. Arasındalık merkezilik değerlerine göre kullanıcılar; 62, 14, 16, 64, 55, 3, 30, 39 ve 24 şeklinde sıralanmaktadır. Bu kullanıcılardan yalnızca beş tanesi ağ düzeninde görülmekte, bunlardan 62 ile 14 diğerlerine göre daha büyük boyutlu olarak görülmektedir. Ağ düzeninde gözükmeyen diğer dört kullanıcının (3, 30, 39 ve 24 numaralı) değerlerinin ise sıfıra yakın derecede düşük olduğu görülmektedir. Derece merkeziliği grafiğinde olduğu gibi arasındalık merkeziliği değerlerine göre yerleşimde de değeri en yüksek olan bu iki kullanıcının, ağın merkezine en yakın kullanıcı olması beklenen bir durumdur. Burada sıralanan kullanıcılar, ağdaki en fazla sayıda kısa yola sahip kullanıcılardır ve diğer kullanıcılar arasında geçiş noktası olma özelliğine sahiptirler. Söz konusu kullanıcıların bu konumları, onlara diğer ikili kullanıcı grupları arasında bilgi akışını engelleyebilme gücü vermektedir ve bu nedenle de ağ içinde önemli bir konumdadırlar. Ağdaki diğer tüm kullanıcıların arasındalık değerleri sıfırdır. Bu kullanıcılarda Şekil 3.41'de görülen hiçbir başka kullanıcı ile bağı olmayan izole durumdaki kullanıcılardır.



Şekil 3.41. 66 Twitter (X) kullanıcısının arasındalık merkeziliği

Yakınlık merkeziliği

Yakınlık merkeziliği, düğümün diğer tüm düğümlerle arasındaki en kısa yol ortalamasının bir ölçüsüdür. Düğümün diğer düğümlere ne kadar yakın olduğunu gösterir ve yakınlık ise o düğümden diğer düğümlere olan iletişimin hızı hakkında bilgi verir. Bu açıdan örneğin bir hastalığın yayılma hızında yakınlık merkezilik ölçüsünün kullanılması tercih edilir. Ağdaki kullanıcıların yakınlık merkeziliği değerlerine göre ağdaki konumları Şekil 3.42'de (algoritma grafik düzeninin kodlama ismi:osage) gösterilmiştir. Ağdaki diğer tüm kullanıcılara olan yakınlığı ifade eden yakınlık merkezilik değerlerine göre kullanıcılar; 62, 14, 16, 24, 39, 55, 29 ve 36 şeklinde sıralanmaktadır.

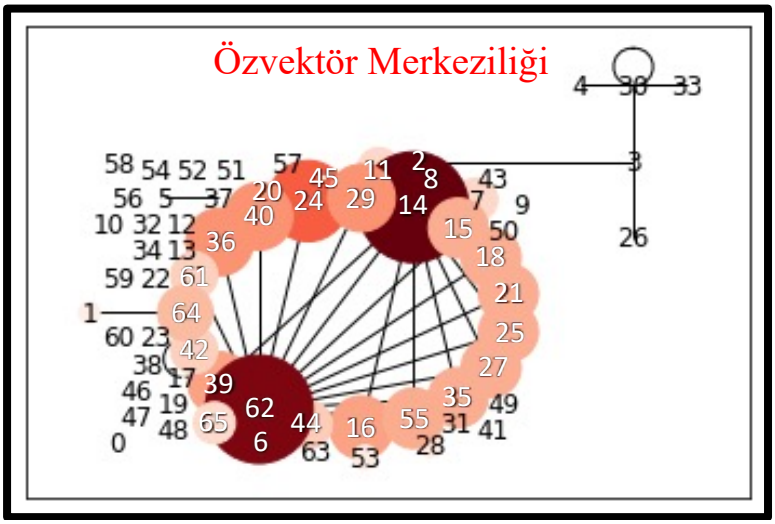


Şekil 3.42. 66 Twitter (X) kullanıcısının yakınlık merkeziliği

Özvektör merkeziliği

Özvektör merkeziliğine göre ağ yapısı içindeki bir düğümün sahip olduğu bağlantılar eşdeğer değildir. Bir düğüm, bağlandığı düğümün değerinin yüksekliği seviyesinde değer kazanır. Az sayıdaki yüksek değerli bağlantıya sahip olmak, çok sayıdaki düşük değerli bağlantıdan daha önemli olabilir [383]. Bu açıdan özvektör merkeziliği ile Twitter (X)'da oluşturulan bir ağda, düğümlerin ne kadar kaliteli bağlantıya ve etkileşime sahip oldukları tespit edilir ve değeri düşük bağlantıların etkisi azaltılır. Bir tepe noktasında yer alan bağlantıların sayısına ve kalitesine bağlı bir merkeziyet kazandıran özvektör merkeziliği, sosyal ağ analizlerinde tercih edilen bir merkezilik ölçüsüdür ve ağ analizlerinde sahte hesapların etkisinin azaltılmasını sağlar [384].

Şekil 3.43'te (algoritma grafik düzeninin kodlama ismi:circo) özvektör merkezilik derecelerine kullanıcıların konumlanmaları gösterilmiştir. Kullanıcılardan özvektör merkezilik değerleri yüksek olanlar, diğerlerine göre daha koyu renkle gösterilmiştir. Ağ içinde kurulan ilişkilerde en önemli ilk on kullanıcı 14, 62, 24, 40, 29, 36, 39, 16, 55 ve 35 numaralı kullanıcılardır. Önemli komşulara da sahip olmaları nedeniyle kendileri de ağ içinde önemli konumdaki kullanıcılardır.

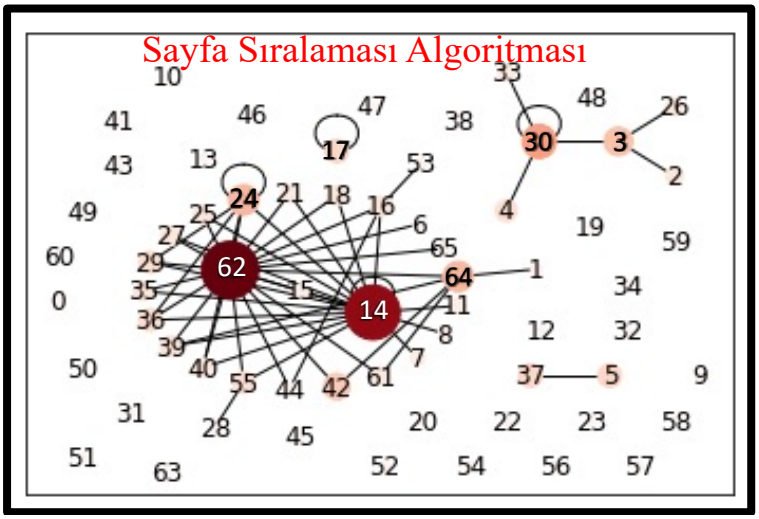


Şekil 3.43. 66 Twitter (X) kullanıcısının özvektör merkeziliği

Sayfa sıralaması

Internette arama sonuçlarını ağırlıklandırma yaparak sıralayan sayfa sıralama algoritması, merkezilik ölçümlerinde de kullanılmaktadır [385, 386]. Sayfa sıralaması, bir internet sayfasının, aldığı link sayısını ve bu linklerin kalitesini hesaba katarak değerlendirir ve buna göre bir sıralama yapar. Bu hesaplama bir ağ yapısı içinde düğümlerin önem sıralamalarının tespit edilmesi içinde kullanılabilir. Sayfa sıralama algoritması, özvektör merkeziliğinin bir çeşididir, özvektör merkeziliğinden en önemli farkı bağlantı yönünü ve düğüme gelen bağlantıları hesaba katmasıdır [387].

Düğüm noktalarının önem derecesini belirleyen bu algoritma sonuçları Şekil 3.44'te (algoritma grafik düzeninin kodlama ismi:twopi) ve ağdaki en yüksek ağırlık puanına sahip kullanıcılar diğerlerine göre daha koyu renkle olacak şekilde gösterilmiştir. Bu algoritmaya göre önem derecesi en yüksek ilk on sıralaması; 62, 14, 30, 64, 24, 3, 42, 5, 17 ve 37 numaralı kullanıcılar şeklinde ortaya çıkmıştır.

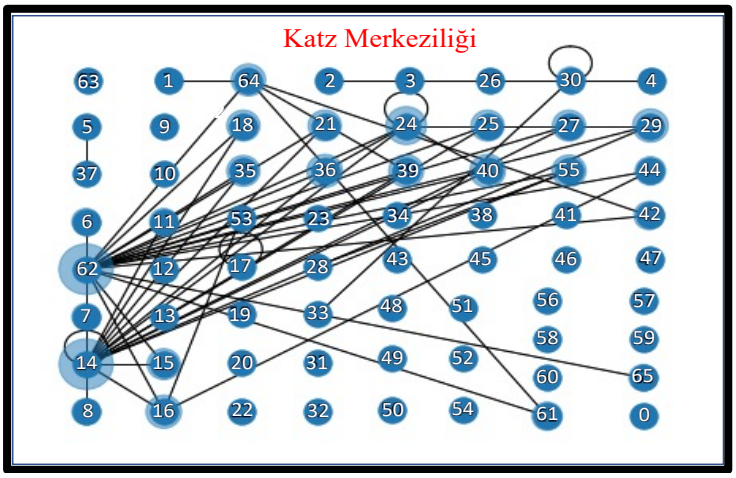


Şekil 3.44. 66 Twitter (X) kullanıcısının sayfa sıralaması

Katz merkeziliği

Katz merkezilik algoritması, düğümlerin komşu düğümlerle sahip oldukları hem doğrudan ve hem dolaylı bağlantıları hesaba katarak merkezilik hesaplaması yapar. Katz merkezilik değeri yüksek olan düğümün bulunduğu ağ içinde, bilgi ve kaynak akışının sağlanması açısından daha yüksek önemde olduğu kabul edilir. Katz merkeziliği doğrudan ve dolaylı

bağların hepsini hesaplamaya dahil ettiği için, sadece doğrudan bağlantıları dikkate alan derece merkeziliği hesaplamasına göre daha kapsamlıdır. Katz algoritması ile yapılan merkezilik hesaplaması sonucunda elde edilen merkezilik önem dereceleri Şekil 3.45'de gösterilmiş, önem derecesi en yüksek ilk on kullanıcı sıralaması; 62, 14, 24, 29, 36, 40, 16, 39, 64 ve 55 numaralı kullanıcılar şeklinde ortaya çıkmıştır. Şekil 3.45'teki (algoritma grafik düzeninin kodlama ismi:patchwork) gösterimde kullanıcılar, Katz merkezilik değerlerinin yükseklikleri ile orantılı olacak şekilde gösterilmiştir.



Şekil 3.45. 66 Twitter (X) kullanıcısının Katz merkeziliği

Yukarıda incelenen altı merkezilik ölçüsüne göre elde edilen merkezilik değerlerinden bir kesit Çizelge 3.11'de verilmiştir. Altı merkezilik ölçütüne göre 66 Twitter (X) kullanıcısı için elde edilen merkezilik değerlerinin tamamı EK-8'de listelenmiştir. 66 kullanıcıdan 6 merkezilik değerinin kendi içindeki ilk on kullanıcı sıralaması Çizelge 3.12'de verilmiştir.

Çizelge 3.11. Twitter(X) (66 kullanıcı)kullanıcılarının merkezilik değerler tablosundan kesit

Kullanıcı No.	Kullanıcı Adı	Derece Mrk.	Arasındalık Mrk.	Yakınlık Mrk.	Özvektör Mrk.	Katz Mrk.	Sayfa Sıralaması
0	@666lehce	0.0	0.0	0.0	0.03	0.0688	0.0037
1	@abdulkadir06161	0.0154	0.0	0.135	0.0193	0.0845	0.006
2	@abdullahciftcib	0.0154	0.0	0.0349	0.21	0.0783	0.0139
3	@adilomeral	0.0462	0.0034	0.0549	-0.13	0.0949	0.036
4	@afmsudan	0.0154	0.0	0.0349	-0.11	0.0793	0.021

Çizelge 3.12. Merkezilik değerlerine göre ilk on kullanıcı sıralaması

Merkezilik Ölçümlerindeki Kullanıcı Sıralamaları						
S.	Derece Mrk.	Arasındalık Mrk.	Yakınlık Mrk.	Sayfa Sıralaması	Özvektör Mrk.	Katz Mrk.
1	62	62	62	62	14	62
2	14	14	14	14	62	14
3	24	64	16	30	24	24
4	30	55	24	64	2	29
5	64	16	39	24	26	36
6	16	3	55	3	29	40
7	3	30	29	42	36	16
8	29	39	36	5	40	39
9	36	24	40	17	54	64
10	39	0	64	37	39	55

Çalışma kapsamında kullanıcıların etkinliklerinin belirlenmesi için, sırasıyla şu işlemlerle ağırlıklı ortalama puan hesabı yapılmış ve sıralama belirlenmiştir: (i) Her bir merkezilik sıralaması için kullanıcılara bulundukları sıraya göre, birinci sıradan son sıraya doğru 66'dan (burada 66 kullanıcıya yönelik sıralama yapıldığı için) bire kadar ağırlık değeri verilmiş ve bu değer kullanıcının o merkezilik derecesindeki sıra numarası ile çarpılmıştır. (ii) Elde edilen değer altı ile bölünerek kullanıcının ortalama ağırlıklı puanı tespit edilmiştir. Kullanıcıların etkinlik sıralaması için, Fei vd. [307]'nin çalışmasındaki minimum, maksimum ve medyan değerlerinin kullanımına benzer şekilde ortalama ve medyan değerleri ile de sıralama yapılmıştır. Buradaki ortalama ve medyan değerleri, her bir kullanıcının altı merkezilik derecesinden elde edilen ortalama ve medyan değeridir. Her üç yöntem ile elde edilen kullanıcıların etkinlik sıralamaları Çizelge 3.13'te verilmiştir.

Çizelge 3.13. Kullanıcıların ağırlıklı puan, ortalama ve medyan değerlerine göre sıralaması

Kullanıcıların Ağırlıklı Puan-Ortalama-Medyan Değerlerine Göre Sıralaması
62:1-1-1 / 14:2-2-2 / 24:3-3-3 / 29:4-5-5 / 16:5-4-4 / 36:6-6-6 / 40:7-7-7 / 39:8-9-8 / 64:9-8-9 / 55:10-10-10 / 15: 11-16-11 / 18:12-11-12 / 21:13-15-13 / 25:14-14-14 / 27:15-13-15 / 35:16-14-16 / 42:17-17-17 / 44:18-19-18 / 61:19-18-20 / 30:20-22-17 / 2:21-26-28 / 26:22-27-29 / 6:23-21-25 / 65:24-20-24 / 7:25-23-21 / 8:26-25-22 / 11:27-24-23 / 53:28-28-30 / 3: 29-37-26 / 54:30-31-46 / 28:31-29-31 / 1:32-30-33 / 41:33-32-42 / 17:34-33-27 / 4:35-56-32 / 33:36-35-36 / 5:37-66-34 / 45:38-34-43 / 0:39-38-37 / 19:40-36-40 / 10:41-40-38 / 37: 42-65-35 / 13:43-42-39 / 12:44-50-53 / 9:45-53-52 / 31:46-44-41 / 23:47-51-56 / 32:48-46-49 / 20:49-58-54 / 49:50-41-45 / 43:51-47-50 / 48:52-43-44 / 38:53-54-58 / 22:54-60-55 / 47:55-49-60 / 58:56-39-47 / 34:57-64-57 / 56:58-48-51 / 46:59-61-59 / 59:60-45-48 / 50:61-59-61 / 60:62-52-65 / 52:63-63-63 / 51:64-62-62 / 57:65-57-64 / 63:66-55-66
Not: Siyah renk, kullanıcı numaralarını; kırmızı renk kullanıcıların ağırlıklı puanlarına göre sıralamasını; mavi renk, kullanıcıların ortalama değerlerine göre sıralamasını ve yeşil renk, kullanıcıların medyan değerlerine göre sıralamalarını göstermek için göstermek için kullanılmıştır.

Söz konusu sıralamaların birbirleri ile uyum ve benzerliklerinin değerlendirmesi için; bu maksatla kullanılan Kendal Tau sıralama korelasyon katsayısı (Kendall Tau), Spearman

sıralama korelasyonu (Spearman Ranking Correlation), Spearman Footrule uzaklık ölçüsü (Spearman Footrule Distance), sınıflar arası korelasyon katsayısı (Interclass Correlation Coefficient-ICC) ölçütleri kullanılmıştır (314, 392). Ağırlıklandırılmış puanla oluşturulan etkin kullanıcı sıralamasının, söz konusu ölçütlere göre ortalama ve medyan değerleri ile yapılan sıralamalarla benzerlikleri Çizelge 3.14'te verilmiştir.

Çizelge 3.14'te ICC dışındaki üç karşılaştırmanın ortalaması alındığında, ağırlıklandırılmış ortalama ile yapılan kullanıcı etkinlik sıralamasının; ortalama değerlere göre yapılan sıralama ile 86,5% ve medyan değerine göre yapılan sıralama ile 92.81% benzerliği olduğu görülmüştür. Bu benzerliğin ICC karşılaştırmasından elde edilen değerle de uyumlu olduğu görülmektedir.

Çizelge 3.14. Kullanıcıların ortalama ve medyan değerlere göre benzerlik karşılaştırması

Ağırlıklandırılmış Ortalama Sıralamasının Ortalama ve Medyan Değerlere Göre Yapılmış Sıralamalarla Benzerlik Karşılaştırması		
	Ortalama Değerlere Göre	Medyan Değerlerine Göre
Kendall Tau	80.76 %	87.95 %
Spearman Sıralama Korelasyonu	92.74%;	97.13%
Spearman Footrule Uzaklık Ölçüsü	86.02%	93.35%
ICC	292	194
Not: ICC değeri, sıralı iki liste arasındaki benzersizlikleri hesaplar, değer küçüldükçe benzerlik artar.		

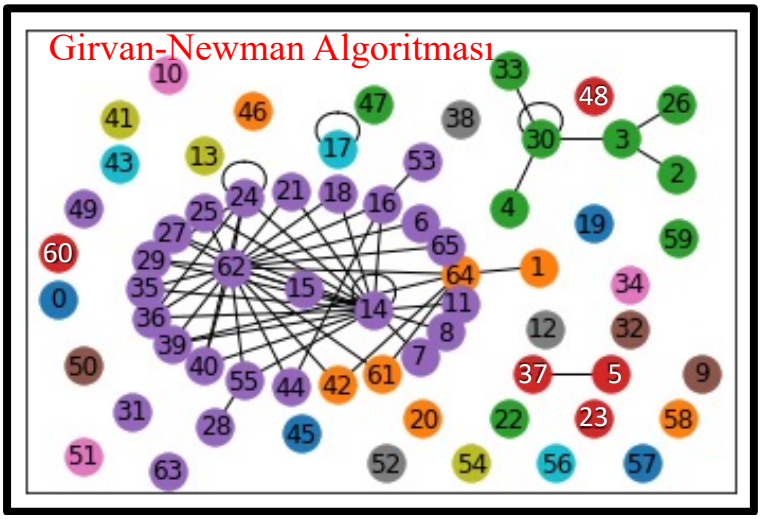
Ağırlıklı ortalama puan hesaplamasıyla, ortalama ve medyan değerlerinde olduğu gibi etkin bir sıralama yapılabildiği, bu sıralamanın güçlü bir şekilde medyan yöntemi ile benzer sonuçlar ürettiği görülmektedir.

3.4.2. Ağ içindeki toplulukların tespiti

Girvan-Newman algoritması

Girvan Newman algoritması, bir ağ içindeki topluluğu, kümelenmeyi tespit etmek maksadıyla kullanılan bir algoritmadır. Algoritma arasındalık değerini kullanarak çalışır, başlangıçta tüm düğümler için arasındalık değerleri hesaplanır. Arasındalık değeri en yüksek olan kenar başlangıç ağından çıkarılır. Bu çıkarma işleminden etkilenen tüm kenarlar için

arasındalık değerleri yeniden hesaplanır. Bu hesaplama işlemi, ağ içinde kenar kalmayana, ağ başka bir düğümle bağlanamaz hale gelinceye kadar tekrarlanır ve sonuçta algoritma bir ağaç diyagramı elde ederek toplulukları tespit eder. Algoritma, küçük ve orta seviyeli ağ analizleri için daha uygundur [388]. Algoritma ile ağ içinde oluşan topluluklar farklı renklerle, Şekil 3.46'da gösterilmiştir.



Şekil 3.46. 66 Twitter (X) kullanıcısının Girvan-Newman algoritmasına göre oluşturduğu topluluklar

Şekil 3.46'da gösterilen topluluk içinde yer alan kullanıcıların listesi Çizelge 3.15'te verilmiştir.

Çizelge 3.15. 66 Twitter (X) kullanıcısından Girvan-Newman algoritması ile tespit edilen topluluklar

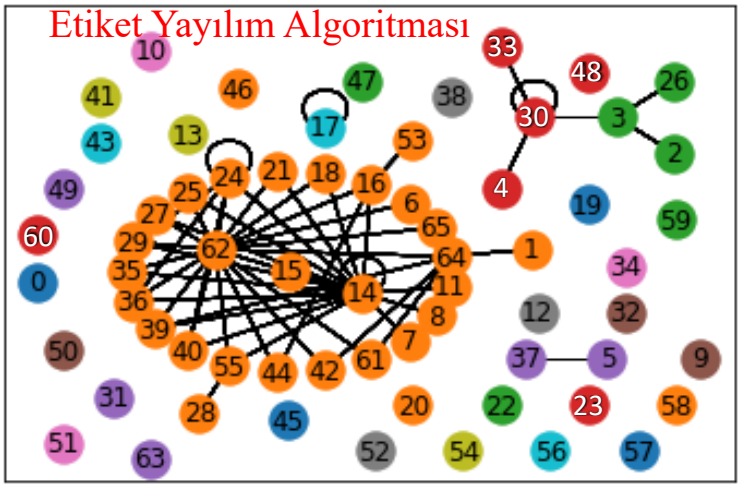
Topluluklar	Topluluk Elemanları
Topluluk-1 (Eflatun renk)	{6, 7, 8, 11, 14, 15, 18, 21, 24, 25, 27, 28, 29, 35, 36, 39, 40, 55, 62, 65}
Topluluk-2 (Yeşil renk)	{33, 2, 3, 4, 26, 30}
Topluluk-3 (Turuncu renk)	{64, 1, 42, 61}
Topluluk-4	{16, 44, 53}
Topluluk-5 (Kırmızı renk)	{37, 5}
Topluluk-Diğer	Ağdaki diğer tüm kullanıcılar (31 kullanıcı) tek üyeli birer topluluk olarak belirlemiştir.

Etiket yayılım algoritması

Etiket yayılım algoritması sınıflandırma ve kümeleme için kullanılan yarı gözetimli bir algoritmadır. Algoritma, etiketli bir düğümden etiketsiz düğüme, düğümler arasındaki

benzerlikleri kullanarak ve etiketleyerek çalışır. Algoritma, iteratif olarak her düğümdeki etiketi, düğümün komşularıyla bağlantılı olarak günceller. Algoritma yayılım için düğümler arasındaki benzerliğe göre etiketleme yapar [389].

Twitter (X) kullanıcıları arasındaki bağlar, etiket yayılım algoritması ile elde edilen ağ yapısı Şekil 3.47'de gösterilmiştir.



Şekil 3.47. 66 Twitter (X) kullanıcısının etiket yayılım algoritmasına göre oluşturduğu topluluklar

Bu algoritma ile 34 topluluk tespit edilmiş ve ağdaki birden fazla üyesi olan topluluklar Çizelge 3.16'da gösterilmiştir. İlk dört topluluk dışında toplam 31 tekil kullanıcı vardır ve bunlar diğer kullanıcılardan hiçbiri ile irtibatlı değildir.

Çizelge 3.16. 66 Twitter (X) kullanıcısının etiket yayılım algoritmasına göre yer aldığı topluluklar

Topluluk Numarası	Topluluk Elemanları
Topluluk-1 (Turuncu)	{1, 6, 7, 8, 11, 14, 15, 16, 18, 21, 24, 25, 27, 28, 29, 35, 36, 39, 40, 42, 44, 53, 55, 61, 62, 64, 65}
Topluluk-2	{26, 2, 3}
Topluluk-3	{33, 4, 30}
Topluluk-4	{37,5}
Topluluk-Diğer	Ağdaki diğer tüm kullanıcılar tek üyeli birer topluluk olarak belirlenmiştir.

Girvan-Newman algoritması ile etiket yayılım algoritmasının tespit ettikleri en büyük topluluklar karşılaştırıldığında; (i) 1'nci topluluklar açısından etiket yayılım algoritmasının

Girvan-Newman algoritmasındaki elemanların hepsini kapsadığı ve ilave olarak 7 elemanı (1, 16, 42, 44, 53, 61, 64 numaralı kullanıcılar) daha içerdiği, (ii) Etiket yayılım algoritmasının 2'nci ve 3'ncü topluluklarının, Girvan-Newman algoritmasında tek bir topluluk olarak (Topluluk 2) oluşturulduğu görülmektedir. Her iki algoritmanın verdiği ortak sonuç dikkate alınırsa; ağ yapısındaki kullanıcılar arasında esas topluluğu; @amanda_devik (6), @artdbkglu (7), @Asibelblonde (8), @ayse_cng (11), @bilgingokberk (14), @burcincigerim (15), @CEngven (18), @ctetikog (21), @delykalamis (24), @dominoeffect34 (25), @er_maannn (27), @gela_wej (28), @gftufekci (29), @ISTANist (35), @ibrahimbagyapan (36), @kagansaydan (39), @KOXALBEY (40), @srfzmr (55), @yemeksepeti (62), @zehra_mete (65) isim ve numaralı kullanıcılar oluşturmaktadır.

3.4.3. Ağ yapısı için kullanılan diğer algoritmalar

Kümeleme katsayısı algoritması

Analiz edilen 66 kullanıcıdan 47'sinin kümeleme katsayıları sıfırdır, bu 47 kullanıcı ağdaki diğer komşuları ile üçgen bir yapı ya da kapalı bir döngü oluşturacak ilişki içinde değildir, bunlar arasında daha sıkı bağlarla alt grup oluşturma eğilimi yoktur. Sıfırdan farklı 19 kullanıcının kümeleme katsayıları aşağıdaki Çizelge 3.17'de verilmiştir. Kümeleme katsayısının 1 olması, o kullanıcının tüm komşularının doğrudan birbirine bağlı olduklarını ve bu durum ağ içinde birbirine sıkı bağlarla bağlı alt grupların varlığını göstermektedir. Buna göre 15, 18, 21, 25, 27, 29, 35, 36, 40, 42, 44, 61 numaralı kullanıcılar, tüm komşuları ile birbirlerine bağlıdırlar ve komşuları ile ağ içinde alt grup oluştururlar. Ayrıca 39 numaralı kullanıcının da 1'e yakın değerleri ile alt küme oluşturma eğilimleri yüksektir. Burada sayılan 13 kullanıcının, Girvan-Newman ve etiket yayılım algoritmaları ile tespit edilen Topluluk-1 içindeki elemanlar arasında oldukları görülmektedir.

Çizelge 3.17. Kullanıcıların kümeleme katsayıları

Kullanıcı Numarası	Kümeleme Katsayısı	Kullanıcı Numarası	Kümeleme Katsayısı
14	0.1176	39	0.6666
15	1.0	40	1.0
16	0.3333	42	1.0
18	1.0	44	1.0
21	1.0	55	0.3333
24	1.0	61	1.0
25	1.0	62	0.1053
27	1.0	64	0.3
29	1.0	---	---
35	1.0	---	---
36	1.0	---	---

Jaccard katsayısı algoritması

Jaccard algoritması, karşılaştırılan iki grup arasındaki benzerliklerin ölçüsüdür. Benzerlik ölçüsü, karşılaştırılan iki grup arasındaki kesişim kümesinin birleşim kümesine oranı ile hesaplanır. Metin madenciliği ve bilgi çıkarımı için kullanılan Jaccard algoritması, bir komşuluk matrisinde düğümler arasında benzerlik oranını hesaplar. Burada kullanıcılar arasındaki benzerliklerin ölçülmesi için kullanılmıştır. Jaccard katsayısı 0 ile 1 arasında değişir ve 1 sayısı benzerliğin tam olduğunu gösterir [390].

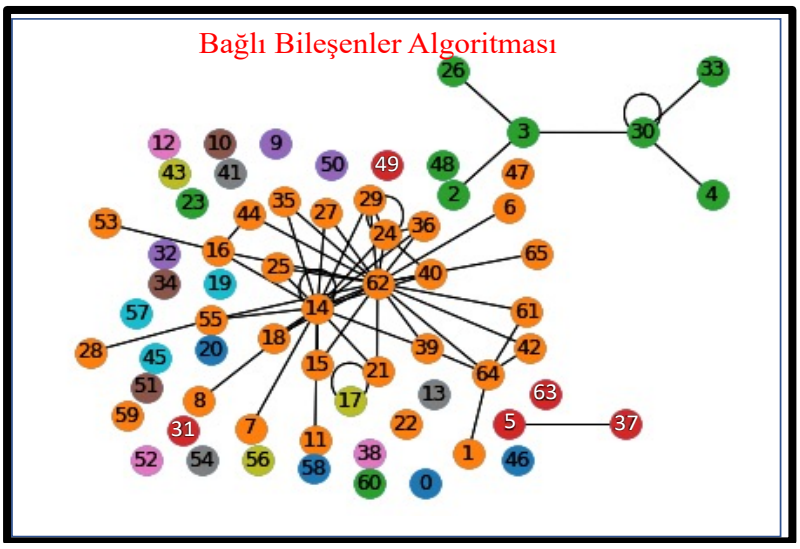
Jaccard katsayıları bir olan 40 kullanıcı çifti Çizelge 3.18'de listelenmiştir. Tabloda verilen tüm ikili kullanıcıların arasındaki benzerlik oranları birdir. Arasındaki benzerlikler dikkate alındığında 40 karşılaştırma işlemi sonucunda algoritma tarafından, 15, 18, 21, 2, 25, 27, 35; 2, 26; 4, 33; 6, 65; 7, 8, 11; 29, 40; 29,36 şeklinde aynı özelliklere sahip 7 grubun olduğu görülmektedir. 7 gruptan ilk sıradaki grubun, kümeleme katsayı algoritması ile elde edilen sonuçtaki alt grup elemanları ile aynı kullanıcılar olduğu görülmektedir. Jaccard katsayısının bir olması, iki kullanıcının birebir aynı olması, bu iki kullanıcının ağ içindeki diğer tüm kullanıcılarla olan ilişkilerinin, ilişki derecesi dahil aynı olduğu anlamına gelmektedir. Örneğin Çizelge 3.18'de ilk sırada yer alan 2 ve 26 numaralı kullanıcılar, ağ içinde kullanıcılardan sadece 3 numaralı kullanıcı ile irtibat halindedirler ve her ikisinin de diğer hiçbir kullanıcı ile irtibatları yoktur. Bu nedenle 2 ve 26 numaralı kullanıcılar birbirleri ile tam benzerlik göstermektedirler ve Jaccard katsayısı birdir.

Çizelge 3.18. Twitter (X) kullanıcılarının Jaccard katsayıları

Benzerlikleri Nedeniyle Eşleştirilen İkili Twitter (X) Kullanıcıları	Jaccard Katsayısı
2 ve 26 / 4 ve 33 / 6 ve 65 / 7 ve 8 / 7 ve 11 / 8 ve 7 / 8 ve 11 / 11 ve 7 / 11 ve 8 / 15 ve 18 / 15 ve 21 / 15 ve 25 / 15 ve 27 / 15 ve 35 / 18 ve 15 / 21 / 25 / 27 / 40 / 15 / 18 ve 35 / 21 ve 15 / 21 ve 18 / 21 ve 25 / 21 ve 27 / 21 ve 35 / 25 ve 15 / 25 ve 18 / 25 ve 21/ 25 ve 27 / 25 ve 35 / 26 ve 2 / 27 ve 15 / 27 ve 18 / 27 ve 21 / 27 ve 25 / 27 ve 35 / 29 ve 36 / 33 ve 4 / 35 ve 18	1.0

Bağlı bileşenler algoritması

Ağdaki 66 kullanıcının 27'si birbirine bağlı en geniş ağı oluşturmaktadır ve ağ analizi yapılan kullanıcı topluluğunun oluşturduğu grafik tümüyle birbirine bağlı değildir. Birbirlerine bağlı kullanıcılar Şekil 3.48'de gösterilmiştir. Turuncu renkle görülen en geniş bağlı bileşen grubunun 1, 6, 7, 8, 11, 14, 15, 16, 18, 21, 24, 25, 27, 28, 29, 35, 36, 39, 40, 42, 44, 53, 55, 61, 62, 64, 65 kullanıcılarından; ikinci olarak yeşil renkle gösterilen 2, 3, 30, 4, 33, 26 kullanıcılarından meydana geldiği görülmektedir. Bağlı bileşen algoritması ağdaki en geniş bağlı elemanlar grubunu tespit eder, aynı zamanda bağlı olmasa da benzer özellikleri taşıyan düğümleri de tespit eder. Yukarıda belirtilen en geniş bağlı bileşen grubunun elemanlarından olmamakla birlikte Şekil 3.48'te 59, 22, 47 numaralı kullanıcılar da turuncu renkle gösterilmiştir. Çünkü bu kullanıcılar da turuncu renkle gösterilen diğerleriyle benzer özellikleri göstermektedirler.



Şekil 3.48. 66 Twitter (X) kullanıcısının bağlı bileşenler algoritmasına göre birbirleri ile kurdukları bağlar

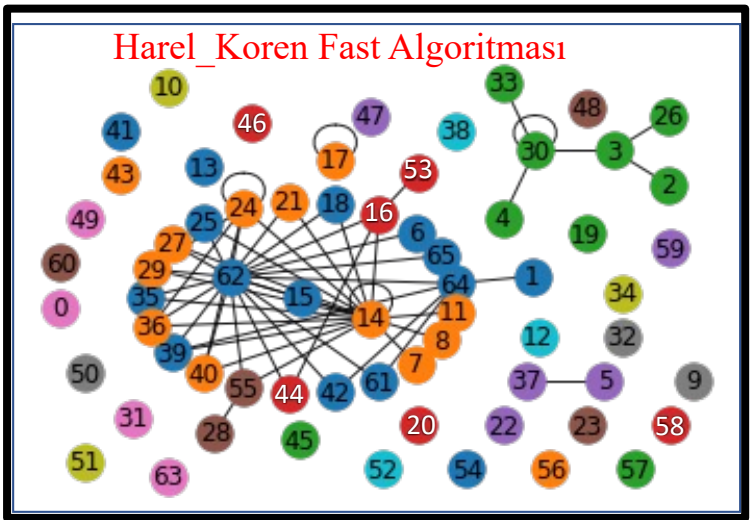
Twitter (X) kullanıcılarının ağ yapısı ileriki maddelerde ağ yapıları için kullanılan, ağ içinde kümelemelere ve daha estetik görselleştirmeye odaklanan, kuvvet yönlendirmeli iki ve üç boyutlu görselleştirmeyi sağlayan düzen algoritmaları ile incelenmiş ve görselleştirilmiştir.

Harel-Koren algoritması

Kuvvet yönlendirmeli bir algoritmadır. Algoritma, özellikle döngüsel olmayan grafiklerin (başlangıç noktasına dönmeyen) iki boyutlu uzayda görselleştirilmesi için kullanılır. Bu algoritma, özellikle yönlendirilmiş döngüsel olmayan grafikler için kullanılır ve kenarlar arasındaki geçişleri minimize eder.

Ağ içinde kullanıcı sayısı en fazla olan ilk dört topluluk mavi, turuncu, kırmızı ve yeşil renklerle Şekil 3.49'da (algoritma grafik düzeninin kodlama ismi:twopi) gösterilmiştir. Ağdaki benzer özellikleri nedeniyle kümeleşen en büyük grupta, 62, 65, 39, 64, 61, 42, 6, 25, 15, 18, 35, 1, 41, 54, 13 numaralı mavi renkli kullanıcılar vardır. Turuncu (7, 8, 11, 14, 17, 21, 24, 27, 29, 36, 40, 43, 56) ve yeşil (2, 3, 4, 19, 26, 30, 33, 45, 57) gruplar sırasıyla en fazla kümeleşen gruplardır.

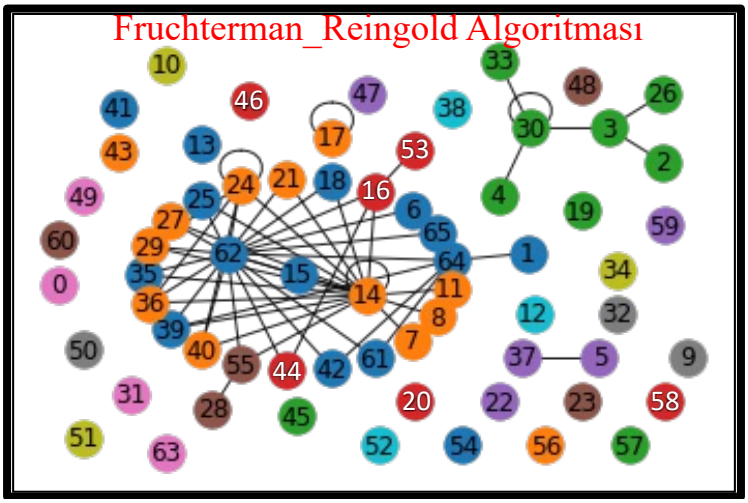
Harel-Koren algoritması, hangi kullanıcılar arasında kümeleşme oluştuğunu ortaya çıkarmak için açgözlü modülerlik algoritmasını kullanır.



Şekil 3.49. 66 Twitter (X) kullanıcısından Harel-Koren algoritmasına göre ortaya çıkan topluluklar

Fruchterman-Reingold algoritması

Kuvvet yönlendirme algoritması olarak bilinen Fruchterman-Reingold algoritması orta büyüklükteki ağların görselleştirilmesi için uygundur. Algoritma, düğümleri yüklü parçacıklara ve kenarları ise yay olan bir sistemin çalışmasına benzeterek çalışır. Algoritma düğüm ve kenarlardan oluşan ağı, iki düğüm birbirine yakın olduğunda birbirini iten, uzak olduklarında birbirini çeken güç olarak kabul eder. İki düğüm arasındaki yay gücü, iki düğüm arasındaki mesafeyi korumaya çalışan çekim gücü olarak kabul edilir. İki düğüm arasındaki potansiyel enerji aralarındaki mesafe kullanılarak hesaplanır. Her iterasyonda, düğüm, sistemdeki toplam enerjiyi minimize edecek küçük mesafelerle hareket ettirilir. İterasyon, sistemdeki toplam enerjiyi minimize eden bir denge durumuna ulaşana kadar devam ettirilir [391]. Algoritma ile oluşturulan genel ağ yapısı Şekil 3.50'de (algoritma grafik düzeninin kodlama ismi:twopi) verilmiştir. Şekil 3.49 ve Şekil 3.50 karşılaştırıldığında Fruchterman-Reingold ve Harel-Korel algoritmalarının aynı sonucu ürettiği görülmektedir.



Şekil 3.50. 66 Twitter (X) kullanıcısından Fruchterman-Reingold algoritmasına göre ortaya çıkan topluluklar

3.4.4. Koleksiyonun ağ analizinden elde edilen bulgular

ÇYS konusunda toplanan koleksiyon üzerinde yapılan ağ analizi ile aşağıdaki sonuçlar elde edilmiştir.

Biri dışında tüm merkezilik ölçütlerinde 62 numaralı kullanıcının, en etkili aktör konumunda, 14 numaralı kullanıcın ise ikinci sırada olduğu, 14 numaralı kullanıcının sadece daha önemli komşulara sahip olması nedeniyle özvektör merkeziliğinde birinci sırada yer aldığı tespit edilmiştir.

Merkezilik ölçütlerine göre diğer Twitter (X) kullanıcıları için genel bir değerlendirme yapıldığında; en önemli ve en güçlü kullanıcılar sırasıyla, 24, 16, 64, 29, 30, 39, 36, 40, 3, 55 numaralı kullanıcılardır. Bu kullanıcılar, ağ içinde köprü görevi görmeleri ve bilginin yayılmasına etkileri nedeniyle konumları nedeniyle güç sahibidirler.

Merkezilik ölçütlerinin ilk on sıralaması incelendiğinde, genelde birbirine benzer sıralamaların yapıldığı, ancak sayfa sıralaması algoritmasının diğer merkezilik ölçülerinin ilk on sıralamasında gözükmeyen 42, 5, 17, 37 numaralı kullanıcıları listeye soktuğu görülmektedir. Yukarıda tartışıldığı gibi en kısa yolları esas alan yöntemlerle komşuluk ilişki ve önem derecelerine dayanan yöntemler birbirinden farklı kullanıcıları ön plana çıkarabilmektedir. Saxena ve Iyengar [392] çalışmasında da görüldüğü gibi farklı merkezilik dereceleri farklı veri seti ve ağ yapılarında diğerlerine göre daha iyi sonuçlar verebilir. Burada uyguladığımız birleştirilmiş merkezilik derece sıralaması ile farklı yöntemlerin ön plana çıkardığı Twitter (X) kullanıcılarının hepsi dikkate alınmıştır.

Etkili aktörlerin belirlenmesi için, ağ içindeki aktör ve ağ yapılarının farklı özelliklerini öne çıkaran altı merkezilik puanı, bu puanlara göre aktörlerin sırası ve belirlenecek aktör sayısına göre aktörlere verilecek ağırlık puanı kullanılarak etkili ve dengeli bir yöntemle sıralama yapılmıştır. Bu yöntemin ortalama ve medyan değerleri kullanılarak yapılan sıralamayla güçlü benzerlikle sonuçlar üretebildiği görülmüştür [306, 307].

Çizelge 3.11 ve EK-8'deki merkezilik derecelerinin değerleri incelendiğinde genel olarak çok düşük olduğu, diğer bir ifade ile kritik kullanıcılar da dahil olmak üzere, kullanıcılar arasındaki bağların zayıf olduğu görülmektedir. Zhang ve Luo [393] arasındalık, yakınlık ve derece merkeziliklerinin arasındaki korelasyona vurgu yapmaktadırlar. Çalışmamızda ortaya çıkan ağ yapısında da birkaç aktör akışın sağlanması açısından kritik derecede önem kazanırken, diğer aktörler arasında nispeten bireysel ilişkilerin öne çıktığı görülmüştür.

Girman-Newman ve etiket yayılım algoritmaları ile yapılan topluluk analizi sonucunda, her iki algoritma tarafından da tespit edilen topluluk içinde; 6, 7, 8, 11, 14, 15, 18, 21, 24, 25, 27, 28, 29, 35, 36, 39, 40, 55, 62, 65 numaralı kullanıcıların yer aldığı, topluluk içindeki kullanıcı sayısının ihtiyaca göre etiket yayılım algoritması ile tespit edilen 1, 16, 42, 44, 53, 61, 64 numaralı kullanıcıların dahil edilmesi ile genişletilebileceği, kullanıcılar arasında tespit edilen ikinci topluluğun 33, 2, 3, 4, 26, 30 numaralı kullanıcılardan meydana geldiği görülmektedir. Her iki algoritmanın analiz edilen ağ büyüklüğüne bağlı olarak hız açısından etkinlikleri değişmekle birlikte bizim çalışmamızda kullanılan ağ büyüklüğü açısından herhangi bir fark ortaya çıkmamıştır [394]. Ancak her iki algoritma ile tespit edilen kullanıcı topluluklarının farklı olduğu görülmüştür.

Kümeleme katsayısı değerleri incelendiğinde, söz konusu topluluklardaki 15, 18, 21, 25, 27, 29, 35, 36, 40, 42, 44, 61 numaralı kullanıcılar, kendi aralarında daha güçlü bir alt grup meydana getirmiştir. Aynı alt grup, Jaccard katsayısı ile benzerlikler üzerinden tespit edilen 15, 18, 21, 2, 25, 27, 35 kullanıcı grubu ile teyit edilmiştir. Söz konusu kullanıcıların, derece merkeziliği ve arasındalık merkezilikleri arasında Zhang ve Luo [393]'nun ifade ettikleri korelasyonun mevcut olduğu tespit edilmiştir.

Bağlı bileşenler algoritması sonuçları incelendiğinde, kullanıcılardan 31'inin başka hiçbir kullanıcı ile bağının olmadığı, dolayısıyla bu kullanıcılara başka kullanıcılar üzerinden ulaşılamayacağı, etki edilemeyeceği ve yine bağımsız olarak ulaşılabileceği, 27 kullanıcının ise birbiri ile kesintisiz olarak birbirine bağlı olduğu tespit edilmiştir.

Harel-Koren ve Fruchterman-Reingold algoritmaları ile tespit edilen alt kullanıcı grupları tamamen aynıdır. Buradan ortaya çıkan en geniş alt grup ise 62, 65, 39, 64, 61, 42, 6, 25, 15, 18, 35, 1, 41, 54, 13 numaralı kullanıcılardan meydana gelmektedir.

Analiz edilen kullanıcılar arasında kuvvetli ilişkiler ortaya çıkmadığı, merkezileşmenin zayıf ve dolayısı ile içerik analizi ile elde edilen sonuçların yönlendirilmiş etkilerden uzak olduğu görülmektedir. Çalışmada uygulanan ağ analizinden elde edilen sonuçlar aşağıdaki şekilde özetlenebilir.

Twitter (X) üzerinden elde edilen veriler üzerinde yapılacak ağ analizi ile kullanıcılar arasında en etkili aktörler ve topluluklar tespit edilebilir. Tespit edilen bu aktör ve

toplulukları, işletmeler kendi amaçları doğrultusunda hedef kitle olarak seçebilir ve bu hedef kitleyi yönlendirmek için işletmenin bütçe kısıtlılıkları çerçevesinde stratejiler belirleyebilir.

Ağ içindeki topluluk ve merkezi durumdaki kritik aktörlerin tespiti iki açıdan önemlidir. Birincisi; eğer gündeme ilişkin müşteri görüşlerinin tespit edilmesi, gündem konularının belirlenmesi ve bu görüşlerden İG'ye yönelik fikirler çıkarılması hedefleniyorsa, bu görüşlerin, gerçek müşterilere ait olduğu teyit edilmelidir. Gündem konuları ağ içindeki etkili ve merkezi durumdaki aktör ve topluluklar tarafından belirleneceği için, bu aktörlerin kimler olduğunun ve ona göre hangi görüşlerin dikkate alınacağının tespit edilmesi önemlidir. İkincisi; gündemi belirlemek ve yönlendirmek isteyen işletme öncelikle etkilemek istediği aktör ve toplulukları ağ analizi ile tespit ederek öncelikli hedef kitlelerini belirleyebilir.

Ağ içindeki aktör ve toplulukları, ağ içindeki ilişkilerin farklı yönlerini öne çıkararak tespit eden farklı ölçüt ve algoritmalar, analiz amaçlarına göre belirlenebilir. Söz konusu ölçüt ve algoritmalar, birbirinden farklı aktörleri öne çıkarabildikleri için, tek bir ölçüt ve algoritma yerine birbirini tamamlayacak şekilde birleştirilebilir. Böylece farklı özellikleri nedeniyle kritikliği bulunan tüm aktörler tespit edilebilir. Özellikle zayıf ilişkilerin olduğu veri setlerinde, birleştirilmiş ölçütler kullanarak daha fazla sayıda etkili kullanıcıya ve sektörel hedef kitleye ulaşmak mümkün olabilir. Yerel ve küresel merkezilik ölçütlerini hibrit olarak kullanan yaklaşımların, merkezilik değerlerinin dengeli bir yaklaşımla hesaplanmasını ve kullanıcı etkinliğinin tüm yönleri ile dikkate alınmasını sağlayacağı için daha etkili sonuçlar üretebileceği düşünülmektedir.

Çalışma ÇYS alanında, anahtar kelime kullanılarak oluşturulmuş Twitter (X) koleksiyonu üzerinde, liste yöntemi, yani önceden belirlenmiş kullanıcı isimleri kullanılarak yapılan bir ağ analiz çalışmasıdır. Çalışmada farklı ölçütler, algoritmalar birleştirilmiş olarak kullanılarak merkezi durumdaki aktör ve topluluklar tespit edilmiş, ancak koleksiyondaki merkezileşmenin düşük olduğu görülmüştür.

4. SONUÇ VE ÖNERİLER

Bu çalışmada ÇYS'ne ilişkin ürün ve hizmetler konusunda Twitter (X) verilerinin tanımlayıcı, içerik/metin ve ağ analiz türü şeklinde üç farklı analiz türü ile analizi yapılmış ve her üç analiz türü için yeni yöntem ve araçlar geliştirilmiştir. Önerilen analiz yöntem ve araçlar; eve/işyerine yemek siparişi hizmetine yönelik kullanıcı eğilim ve beklentilerini tespit etmek, siparişin başlangıç noktasından (siparişin işyerine iletilmesinden) siparişin müşteriye teslimi aşamasına kadar olan (perakendeci/ restoran tarafından verilen hizmetin) hizmetin iyileştirilmesine yönelik (problem sahalarının tespit edilerek, bunlara yönelik düzeltici/düzenleyici) olan sürece ilişkin öneriler geliştirmek için kullanılmıştır.

4.1. Sonuçlar

Çalışmada elde edilen sonuçlar aşağıya çıkarılmıştır.

Çalışma, yemek siparişi alanında üç farklı analiz türünün bir arada kullanıldığı Türkçe literatürdeki ilk çalışmadır. Literatürdeki çalışmalarda söz konusu analizler, ayrı ayrı uygulanmakta, bu analizlerin birbirlerini bütünleyici etkileri göz ardı edilmekte, veri setinin yönlendirilmiş olup olmadığı konusu incelenmemekte ve analiz sonuçları doğrudan gerçek müşterilere ait görüşler olarak kabul edilmektedir. Bu çalışmada veri setlerinin reklam, propaganda ve yönlendirmeye açık olduğu gösterilmiş, tanımlayıcı analiz aşamasında reklam içeriklerinden ayıklama yöntemleri gösterilmiş, veri setinde etkinliği güçlü kullanıcı ve toplulukların varlığı araştırılmış, bu kullanıcıların ağ içindeki merkezilikleri ve topluluk oluşturma eğilimleri ağ analizleri ile tespit edilmiştir. Çalışmada içerik analizi ile elde edilen analiz sonuçlarının gerçekçi ya da yönlendirilmiş olup olmadığı tanımlayıcı analiz ve ağ analizi ile yapılan işlemlerle ortaya konulmuştur.

Kullanıcıların diyez etiket kelimesi kullanma alışkanlığının çok düşük seviyede kalması nedeniyle, bu alana yönelik analizler için araştırmacı ya da sektörel analiz yapacak işletmelerin doğru veri setlerine ulaşmak için uygulamalarda *diyez etiket kelimesi* yerine *anahtar kelime* kullanmaları önerilmektedir.

Twitter (X) gibi paylaşım platformlarından elde edilen veri setlerinden, tanımlayıcı analiz özelliklerini kullanarak sektörel olarak iyileştirme/geliştirmeye yönelik fikirler çıkarmak, veri setinin gündem konularını tespit etmek, sektörel faydalı bilgilere ulaşmak mümkündür.

Twitter (X) veri analizlerine ilişkin literatürdeki çalışmalarda, reklam içerikli tweetlerin veri analiz sonuçlarını olumsuz etkilemeyeceği kabul edilmektedir. Bu çalışmada literatürden farklı olarak reklam ve propaganda (gürültü) içerikli tweetlerin analiz sonuçlarını doğrudan etkilediği, elde edilen sonuçların değiştiği tespit edilmiş ve doğru analiz sonuçlarına ulaşmak için içeriklerin gürültüden ayıklanmasını da kapsayan ve TVİTA olarak isimlendirilen bir yöntem önerilmiştir. Önerilen yöntemle birlikte, yemek sektörüne ilişkin veri setlerinde gürültünün ayıklanması için kullanılabilecek bir reklam sözlüğü oluşturulmuştur.

Çalışmada metin sınıflandırma ve duygu analizi için kullanılacak alana özel modeller geliştirilmiş, metin sınıflandırmanın önemli problemlerinden özellik seçimi, boyut indirgeme konusunda kelimelerin gruplandırılmasını esas alan bir yöntem önerilmiş, önerilen yöntemlerle alana özel duygu (pozitif ve negatif) ve ürün-hizmet sistemleri sınıflandırmalarında kullanılabilecek dört adet sözlük geliştirilmiştir. Bu kapsamda çalışmanın literatüre önemli bir katkısı, duygu ve ürün-hizmet sınıflandırmalarında kullanılmak üzere bir güçlendirilmiş sözlük önerilmesi ve bu sözlük yapısının temel sözlüğe göre hem daha iyi sonuçlar vermesi hem de boyut olarak daha küçük olması nedeniyle problemin karmaşıklığını azaltmasıdır. Uygulanan yöntemin hem duygu hem ürün-hizmet sistemleri sınıflandırmasında performansı artırdığı görülmüştür. Önerilen yöntem ve sınıflandırma modelleriyle, 85% ve üzerinde bir sınıflandırma performansı ile literatürde yer alan birçok duygu analizi ve metin sınıflandırma çalışmasından daha iyi performans elde edilmiştir. Kullanıcıların ağ içindeki merkeziliğini tespit etmek maksadıyla, yerel ve genel merkezilik ölçütlerini ortalama ve ağırlıklandırılmış ortalama ile bir araya getiren bir merkezilik ölçütü önerilmiştir. Önerilen ölçüt ile, merkeziliklerin medyan değerinin alınması şeklinde uygulanan literatürdeki yöntemle benzer sonuçlar elde edildiği görülmüştür.

4.2. Müteakip çalışmalar için öneriler

Twitter analitiği için önerilen TVİTA modeli farklı sektörlere uyarlanarak yeni metotlarla genişletilerek kullanılabilir.

Tanımlayıcı analiz özellikleri ve çalışma kapsamında önerilen model ile ticari ürün/hizmet konulu bu çalışmanın bir benzeri Facebook, Instagram, Youtube gibi diğer sosyal ağlarda yapılan paylaşımlardan oluşan veri setleri üzerinde kullanılabilir. Bu çalışmada önerilen tanımlayıcı analiz modeli, Twitter (X) kullanıcı profillerinde yer alan diğer üst verilerin kullanımı ve demografik analizlerle (cinsiyet analizi, yaş grupları, coğrafi konum bilgileri) genişletilerek Twitter (X)'den işletmelerin kullanabileceği ne tür bilgiler elde edilebileceğine ışık tutacak yeni çalışmalar yapılabilir. Ayrıca sosyal medyadan elde edilecek veri setlerinde yapılacak tanımlayıcı analiz sonuçları ile işletmelerin kendi havuzlarındaki müşterilerine ilişkin veri setlerinin analizinden elde edilecek sonuçlar karşılaştırmalı olarak incelenebilir, analiz sonuçlarında paralellik olup olmadığı test edilebilir.

Çalışma kapsamında geliştirilen alan sözlükleri, dilde kullanıma yeni giren kelimelerle güncelleştirilerek yemek alanına özel yeni çalışmalarda kullanılabilir. Çalışmada geliştirilen uygulanan iki kademeli özellik tabanlı sınıflandırma modeli, ürün ve hizmet sınıfına eklenecek fiyat ve personel sınıfları gibi yeni sınıflarla genişletilerek alana özel yeni çalışmalarla kullanılabilir. Çalışmada oluşturulan temel sözlüklerin diğer kelimeleri üzerinde de güçlendirme işlemi uygulanarak, sözlük performanslarında ne tür değişimler ortaya çıkacağı tespit edilebilir ve performans artışının sağlanabileceği optimum kelime/değişken sayısı belirlenebilir.

Metin sınıflandırma problemlerinde önemli bir problem sahası olan değişken sayısının azaltılması, boyut indirgeme problemine çözüm olarak önerilen güçlendirilmiş sözlük yapısı; kelime sayısı tweetlere göre daha fazla olan diğer metin türleri üzerindeki sınıflandırmalarda kullanılabilir. Güçlendirilmiş sözlüklerle elde edilen performanslarda metin uzunluklarının etkisi ölçülebilir.

Sektörlere yönelik ağ analiz çalışmaları için, Twitter (X)'da liste yöntemi yerine ilişkisel veri toplama yöntemi kullanılarak elde edilmiş veri setleri üzerinde analizler gerçekleştirilebilir. Alana özel bu çalışmadaki ağ analizinin, sektörel özgün çalışmalar yapılması konusunda ağ analizinin kullanılmasını destekleyici bir kaynak sağlayacağı, bu tür çalışmaların sektörel yapıların anlaşılmasına önemli katkı vereceği öngörülmektedir.

Çalışma kapsamında hazırlanan alana özel reklam sözlüğü, tohum kelime listesi olarak kabul edilip, genişletilebilir, genişletilen reklam sözlüğün sınıflandırma algoritmaları ile kullanılarak performans ölçümü yapılabilir. Elde edilen analiz sonuçları, sektörde faaliyet gösteren işletmelerin veri tabanındaki müşteri verilerinin analizinden elde edilecek sonuçlar ile karşılaştırmalı olarak test edilebilir. Böylece Twitter (X) kullanıcılarından elde edilen müşteri görüşlerinin gerçeği ne derecede yansıttığı ortaya konulabilir. Buradan hareketle, işletmelere Twitter (X) verileri ile işletmelere hızlı geri besleme bilgilerinin sağlanıp sağlanamayacağı, hangi zaman aralıklarını kapsayan Twitter (X) verilerinin gerçek verilere daha yakın şekilde gerçekleştiği konusunda öneriler geliştirilebilir.

Bu çalışmada yemek konusunda diyet ve diyeti bozma nedeniyle yaşanan pişmanlığın, tweetlerde önemli gündem maddelerinden biri olduğu, kişilerin yemek siparişi verme ile diyet yapma/spor konusunda pişmanlık, kararsızlık arasında duygusal gidiş-gelişler yaşadıkları görülmektedir. Bu konuda pişmanlık kategorisini de içine alan duygu sınıflandırması şeklinde alana özel bir çalışma yapılabilir. Çalışma kapsamında yapılan saat, gün gibi zamanla ilişkilendirilmiş analizler, insanın biyolojik saati ve yapısı, beslenme ihtiyacı gibi konularla ilişkilendirilerek aralarında ne derece uyumluluk/uyumsuzluk olduğu tespit edilebilir ve çalışma bu alanda genişletilebilir. Koleksiyonun ağ yapısı kullanıcılar arasındaki tüm ilişkileri dikkate alan birleşik ağ ilişkileri analizi şeklinde ve düğüm sayısı artırılarak yapılabilir ve sonuçlar karşılaştırılarak, ağ yapısı için optimum kullanıcı sayısının tespitine yönelik ağ analizi yapılabilir. Müteakip dönemde sektörlere yönelik ağ analiz çalışmaları için, Twitter (X)'da liste yöntemi yerine ilişkisel veri toplama yöntemi kullanılarak elde edilmiş veri setleri üzerinde analizler gerçekleştirilebilir. Alana özel bu çalışmanın, sektörel özgün çalışmalar yapılması konusunda ağ analizinin kullanılmasını destekleyici bir kaynak sağlayacağı, bu tür çalışmaların sektörel yapıların anlaşılmasına önemli katkı vereceği öngörülmektedir.

KAYNAKLAR

1. Mishra, N. ve Singh, A. (2018). Use of Twitter data for waste minimisation in beef supply chain. *Annals of Opeations Research*, 2018(270), 337–359.

2. Dixon, S. (2022). Number of Twitter Users Worldwide from 2019 to 2024. URL: https://www.statista.com/statistics/303681/twitter-users-worldwide/. Son Erişim Tarihi: 25.04.2023.

3. Chae, B. K. (2015). Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics,* 2015(165), 247-259.

4. Choudhary, A., Hendrix, W., Lee, K., Palsetia, D., ve Liao, W. K. (2012). Social media evolution of the Egyptian revolution. *Communications of the Association for Computing Machinery*, 55(5), 74-80.

5. Çetin, M. ve Amasyalı, M.F. (2013). *Supervised and traditional term weighting methods for sentiment analysis*. Signal Processing and Communications Applications Conference, Girne, 1-4.

6. Li. J., Vishwanath, A. ve Rao, H.R. (2014). Retweeting the Fukushima nuclear radiation disaster. *Communications of the Association for Computing Machinery*, 57(1), 78-85.

7. Meral, M. ve Diri, B. (2014). *Twitter üzerinde duygu analizi. Institute of Electrical and Electronics Engineers* 22nd Signal Processing and Communications Applications Conference, Trabzon, 690-693.

8. Çevik, M., Öztürkcan, S. ve Kasap, N. (2015, Haziran 10-13). *Sosyal medya analitiği: Twitter için büyük veri yaklaşımı*. 20'nci Ulusal Pazarlama Kongresi, Eskişehir, 145-148.

9. Akgül, E. S., Ertano, C. ve DİRİ, B. (2016). Twitter verileri ile duygu analizi. *Pamukkale Üniversitesi Mühendislik Bilim Dergisi*, 22(2), 106-110.

10. Kocak, B., Polat, İ. ve Kocak, C. B. (2016). *Twitter kullanıcılarının havayolu pazarına yönelik duygu kutuplarının belirlenmesi: Bir fikir madenciliği örneği*. Global Business Research Congress , Istanbul, 684-691.

11. Onan, A. (2017). Twitter mesajları üzerinde makine öğrenmesi yöntemlerine dayalı duygu analizi. *Yönetim Bilişim Sistemleri Dergisi*, 3(2), 1-14.

12. Albayrak, M. Topal, K. ve Altıntaş, V. (2017). Sosyal medya üzerinde veri analizi: Twitter. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Kayfor15 Özel Sayısı(22), 1991-1998.

13. Çoban, Ö. ve Tümüklü-Özyer, G. (2018). Twitter duygu analizinde terim ağırlıklandırma yönteminin etkisi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 24(2), 283-291.

14. Taşdelen, B. (2020). Twitter'da Suriyeli mültecilere karşı çevrimiçi nefret söylemi, *Gümüşhane Üniversitesi Sosyal Bilimler Enstitüsü Elektronik Dergisi*, 11(2), 562-575.

15. Akar, E., Bayar, Ö. ve Koçtürk, N. (2021). İstanbul Sözleşmesi: Twitter üzerinden sözleşmeye karşı çıkanların görüşlerine yönelik bir söylem analizi. *Uluslararası İnsan Çalışmaları Dergisi*, 4(7), 88-106. DOI: https://doi.org/10.35235/uicd.893945.

16. Ainapure, B.S., Pise, R. N., Reddy, P., Appasani, B., Srinivasulu, A., Khan, M. S. ve Bizon, N. (2023). Sentiment Analysis of COVID-19 Tweets Using Deep Learning and Lexicon-Based Approaches. *Sustaintability*, 15 (2573), 1-21.

17. Qi, Y., Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon and machine-learning-based approach. *Social Network Analysis and Mining*, 13 (31), 1-14. DOI: https://doi.org/10.1007/s13278-023-01030-x.

18. Hingle, M., Yoon, D., Fowler, J., Kobourov, S., Schneider, M. L., Falk, D., Burd, R., (2013). Collection and visuliation of dietary behavior and reasons for eating using Twitter. *Journal of Medical Internet Research*, 15(6), e125.

19. Park, B. S., Jang, J., Chihyung Ok, C. M., (2016). Analyzing Twitter to explore perception of Asian restaurants. *Journal of Hospitality and Tourism Technology*, 7(4), 405-422.

20. Singh, A., Shukla N. ve Mishra N. (2018). Social media data analytics to improve supply chain management in food industry. *Transportion Research*,114, 398-415.

21. El-Khchine, R., Amar, A., Guennoun, Z. E., Bensouda C. ve Laaroussi, Y., (2018). *Machine learning for supply chain's big data: State of the art and application to social network's data.* Materials Science, Engineering and Chemistry Web of Conferences 200 (00015), 7. DOI: https://doi.org/10.1051/matecconf/201820000015.

22. Boyd, D. ve Crawford, K. (2013). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomeno. *Information, Communication and Society*, 15(5), 662–679.

23. İnternet: Anomaly detection. URL: https://en.wikipedia.org/wiki/Anomaly_detection, Son Erişim Tarihi: 16.01.2023.

24. Internet: Ramos, L.P. Creating MongoDB databases with the Monge Shell. URL: https://realpython.com/introduction-to-mongodb-and-python/#creating-collections-and-documents, Son Erişim Tarihi: 25.05.2023.

25. Internet: Hypersense. URL: https://www.hypersenseai.com/aiglossary/corpus/#:~:text=A%20corpus%20is%20a%20collection,shows%2C%20movies%2C%20and%20tweets, Son Erişim Tarihi: 09.07.2023.

26. İnternet: Glossary. URL:https://cloud.ibm.com/docs/watson-knowledge-studio-data?topic= watson-knowledge-studio-data-glossary#gloss_C, Son Erişim Tarihi: 27.05.2023.

27. Internet: Social network analysis. URL: https://en.wikipedia.org/wiki/Social_network_analysis, Son Erişim Tarihi: 25.05.2023.

28. İnternet: API nedir? API açılımı ve API kullanımı. URL: wmaraci.com/nedir/api, Son Erişim Tarihi: 21 Ekim 2020.

29. İnternet: Aksan, C. E. API (Uygulama Programlama Arayüzü) nedir, Ne işe yarar? URL: ceaksan.com/tr/api-nedir, Son Erişim Tarihi: 21 Ekim 2020.

30. İnternet: Library (Computing). URL: wikipedia.org/. Son Erişim Tarihi: 21 Ekim 2020.

31. İnternet: En popüler yazılım dilleri nelerdir? URL: recelinteractive.com/, Son Erişim Tarihi: 21 Ekim 2020.

32. Kaisler, S., Armour, F., Espinosa, J. A., Money, W. (2013). *Big data: Issues and challenges moving forward*. 46th Hawaii International Conference on System Sciences-Wailea-HI, 995-1004.

33. Köseoğlu, Ö., Demirci, Y. (2017). Türkiye'de büyük veri ve veri madenciliğine ilişkin politika ve stratejiler: Ulusal politika belgelerinin içerik analizi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Kayfor 15 Özel Sayısı (2), 2223-223.

34. Russom, P. (2011). Big data analytics. *TDWI Best Practices*, Fourth Quarter (19), 1-34, 2011.

35. Tiwan, S., Wee, H. M., Daryanto Y. (2018). Big data analytics in supply chain mangement between 2010 and 2016: Insights to industries. *Computers and Industrial Engineering*, 2018(115), 219-230.

36. İnternet: Rouse, M. Big data analytics. URL: https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics, Son Erişim Tarihi: 21 Mart 2019.

37. Çakırel, Y. (2016). İşletmelerde büyük veri. *Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(1), 2146-3417.

38. Gökalp, M. O., Kayabay, K., Çoban, S., Yandık, Y. B. ve Eren, P. E. (2018). *Büyük veri çağında işletmelerde veri bilimi*. 5th International Management Information Systems Conference, Ankara, 94-97.

39. Zhong, R. Y., Newman, S. T., Huang, G. Q., Lan, S. (2016). Big data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers and Industrial Engineering*, (2016) (101), 572-591.

40. Addo-Tenkorang, R., Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literatüre review. *Computers and Industrial Engineering*, 2016(101), 528-543.

41. Nguyen, T., Zhou, L., Spiegler, V., Ieromonachou, P. ve Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. *Computers and Operations Research*, 2018 (98), 254-264.

42. Barton, D. ve Court, D. (2012). Making advanced analytics work for you, a practical guide to capitalizing on big data. *Harvard Business Review*, 90 (10), 78-83.

43. Marchand, D. A. ve Peppard, J. (2013). Why IT fumbles analytics, tech projects should focus less on technology and more on information. *Harvard Business Review*, 91, 104-112.

44. İnternet: Economist Intelligence Unit (EIU) (2012). Deciding factor: Big data and deciding factor. Capgemini, Consulting Technology, Outsourcing, Business Analytic, URL: http://www.capgemini.com/resources/the-deciding-factor-big-data-decision-making, Son Erişim Tarihi: 22.04.2023.

45. Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *Massachusetts Institute of Technology Sloan Management Review*, 52(2), 21-32.

46. Manyika, J, Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. ve Byers, A. H. (2011). Big data: The next frontier for innovation, competetion, productivity. *McKinsey Global Institute*.

47. İnternet: Veri analizi. URL: https://tr.wikipedia.org/wiki/Veri_analizi#:~:text=Veri%20analizi%2C%20faydalı%20bilgiler%20bulma,temizleme%2C%20dönüştürme%20ve%20modelleme%20işlemidir, Son Erişim Tarihi: 27.05.2023.

48. Silahtaroğlu, G., (2013). *Veri Madenciliği Kavram ve Algoritmaları*. Papatya Yayıncılık Eğitim, 2. Baskı, İstanbul, Türkiye.

49. Tunalı, V. (2011). *Metin madenciliği için iyileştirilmiş bir kümeleme yapısının tasarımı ve uygulaması*. Doktora Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

50. Demir, Ö., Baban Chawai, A. İ. ve Doğan, B. (2019). Türkçe metinlerde sözlük tabanlı yaklaşımla duygu analizi. *International Periodical of Recent Technologies in Applied Sciences*, 2019(2), 58-66.

51. Hotho, A., Nürnberger, A., Paass, G. (2005). A brief survey of text mining. *Journal for Computational Linguistics and Language Technology*, 20, 19-62.

52. Oğuzlar, A. (2003). Veri önişleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (21), 67-76.

53. Larose, D. T. ve Larose C. D. (2005). *Discovering Knowledge In Data: An Introduction to Data Mining.* John Wiley & Sons, Data 2nd Edition, Hoboken, New Jersey, USA.

54. Ralha, C.G. ve Silva, C.V. (2012). A multi-agent data mining system for cartel detection in Brazilian government procurement. *Expert Systems with Applications*, 39(14), 11642-11656).

55. İnternet: Herkese açık ve korumalı tweetler hakkında. URL: https://help.twitter.com/tr/safety-and-security/public-and-protected-tweets, Son Erişim Tarihi: 10.10.2019.

56. İnternet: Yanıtlar ve bahsedenler hakkında. URL: https://help.twitter.com/tr/using-twitter/mentions-and-replies, Son Erişim Tarihi: 10.10.2019.

57. İnternet: Retweetleme. URL: https://help.twitter.com/tr/using-twitter/how-to-retweet, Son Erişim Tarihi: 10.10.2019.

58. İnternet: Yeni kullanıcılar için Twitter terimleri. URL: https://sosyalmedya.co/yeni-kullanicilar-icin-twitter-terimleri/, Son Erişim Tarihi: 10.10.2019.

59. İnternet: İnternet: Etiketleri kullanma. URL: https://help.twitter.com/tr/using-twitter/how-to-use-hashtags, Son Erişim Tarihi: 01.10.2019.

60. İnternet: Twitter'ın API'leri hakkında. URL: https://help.twitter.com/tr/rules-and-policies/twitter-api, Son Erişim Tarihi: 01.10.2019.

61. Internet: Twitter developer. URL: https://twitter.com/TwitterDev/status/1621026986784337922, Son Erişim Tarihi: 05.07.2023.

62. İnternet: Developer Platform-happening to build what's next? Get immediate access to the Twitter API and unlock the potential of the Twitter data. URL: https://developer.twitter.com/en, Son Erişim Tarihi: 05.07.2023.

63. Atan, S. ve Çınar, Y. (2019). Borsa İstanbul'da finansal haberler ile piyasa değeri ilişkisinin metin madenciliği ve duygu (Sentiment) analizi ile incelenmesi. *Ankara Üniversitesi Siyasal Bilgiler Fakültesi Dergisi*, 74(1), 1-34.

64. Bruns, A. ve Stieglitz, S. (2013). Towards more systematic Twitter analysis: Metrics for tweeting activities. *International Journal of Social Research Methodology*, 16(2), 91-108.

65. Karabulut, Y. E., Küçüksille, E.U. (2018). Twitter profesyonel izleme ve analiz aracı, *Teknik Bilimleri Dergisi*, 8(2), 17-24.

66. Tutar, K., Ünalır, M., O. ve Toker, L. (2015). Sosyal ağlar üzerinde ontoloji tabanlı sezgi analizi için bir uygulama çerçevesinin geliştirilmesi. *Pamukkale Universitesi Mühendislik Bilimleri Dergisi*, 21(5), 194-202.

67. Delen, D., ve Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, 34(3), 1707-1720.

68. Feldman, R. ve Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data.* Cambridge University Press, 34(1), Cambridge, England.

69. İnternet: Saura, J. R. ve Bennet, D. R. (2019). A Three-stage method for data text mining: Using UGC in Business Intelligence Analysis. MDPI, 11(4), 519. URL: https://doi.org/10.3390/sym11040519. Son Erişim Tarihi: 26.04.2023

70. Bach, M. P., Krstic, Z., Seljan, S. ve Turulja, L. (2019). Text mining for big data Analysis in financial sector: *A literature review*. 11(5), 1277.

71. Zohar, E. Y. (2002). *Introduction to text mining, supercomputing 2002*. Automated Learning Group National Center for Supercomputing Applications, University of Illinois.

72. Ravi, K., ve Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 2015(89), 14-46.

73. Boumans, J. W., Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques. *Digital Journalism*, 4(1), 8-23.

74. Uçan, A. (2014). *Otomatik duygu sözlüğü çevirimi ve duygu analizinde kullanımı*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Bilgisayar Mühendisliği Ana Bilim Dalı, Ankara.

75. Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the Association for Computing Machinery*, 2013(56), 82-89, 2019.

76. Pang, B. ve Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2008(2), 1-135.

77. Hussein, D.M.E. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences,* 2018(30), 330-338.

78. Sun, S., Luo, C. ve Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, 2017(36), 10-25.

79. Chaturvedi, I., Cambria, E., Welsch, R. E. ve Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion*, 2018(44), 65-77.

80. Korayem M., Aljadda K., and D. Crandall (2016). Sentiment/subjectivity analysis survey for languages other than English. *Social Network Analysis and Mining*, 6(1), 1-17.

81. Aggarwal, C. C. ve Zhai, C. (2012). *Mining text data, Science and Business Media*-Springer, New York, USA.

82. Tetik, S. (2017). *Twitter verisi ile R'de metin analizi uygulaması*. Lisans Bitirme Tezi, Yıldız Teknik Üniversitesi, Fen-Edebiyat Fakültesi, İstanbul.

83. El-Din, D. M. (2016). Enhancement bag-of-words model for solving the challenges of sentiment analysis. *International Journal of Advanced Computer Science and Applications*, 2016(7), 244-247.

84. Zhang, Z. (2008). Weighing stars: Aggregating online product reviews for intelligent ecommerce applications. *Institute of Electrical and Electronics Engineers Intelligent Sysems*, 2008(23), 42-49.

85. Pang, B. ve Lee, L. (2005, Temmuz 25-30). *Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating sales.* The Association for Computational Linguistics 43rd Annual Meeting (ACL, 115-124), Ann Arbor, Michigan, New Brunswick.

86. Sağlam, F. (2019). *Otomatik duygu sözlüğü geliştirilmesi ve haberlerin duygu analizi*. Doktora Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Ankara.

87. Uysal, A. K. ve Günal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management*, 50(1), 104-112.

88. Akpınar, H. (2000). Veri tabanlarında bilgi keşfi ve veri madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29(I), 1-22.

89. İnternet: Durmaz, O. Metin önişleme. URL: http://www.osmandurmaz.com/makale/metinonisleme/makale.aspx, Son Erişim Tarihi: 17.05.2020.

90. Çoban, Ö., Özyer, B. ve Özyer, G. T. (2015). *Sentiment analysis for Turkish Twitter feeds*, Signal Processing and Communications Applications Conference (SIU. IEEE-2015(23), 2388-2391.

91. Şeker, S. E. (2015). Doğal dil işleme (natural language processing) *İstanbul Medeniyet Üniversitesi YBS Ansiklopedi*, 2015(2), 14-22.

92. Türkmenoğlu, C. (2015). *Türkçe metinlerde duygu analizi*. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

93. Oflazer, K. (2012). Türkçe ve doğal dil işleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5 (2), 1-12.

94. Aytekin, Ç, Sütçü, C. S. ve Özfidan, U. (2018). Karar ağacı algoritması ile metin sınıflandırma: Müşteri yorumları örneği. *Uluslararası Sosyal Araştırmalar Dergisi*, 11(55), 782-792.

95. Kurdi, M. Z. (2017). *Natural Language Processing and Computational Linguistics 2: Semantics, Discourse and Applications.* John Wiley & Sons, New Jersey, USA.

96. Liddy, E. D. (2001). *Natural language processing. Encyclopedia of Library and Information Science*, M. Drake (Eds), 2, New York, USA.

97. Çetin, F.S. ve Eryiğit, G. (2018). Türkçe hedef tabanlı duygu analizi için alt görevlerin incelenmesi-hedef terim, hedef kategori ve duygu sınıfı belirleme. *Bilişim Teknolojileri Dergisi*, 11(1), 43-56.

98. Xia, R., Xu, F., Yu, J., Qi, Y. ve Cambria, E. (2016). Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis, *Information Processing and Management*, 52 (1), 36-45.

99. Polanyi, L. ve Zaenen, A. (2006). Contextual valence shifters, Computing attitude and affect in Text. *Theory and applications*. 20, 1-10.

100. Agarwal, B. ve Mittal N. (2016). Prominent Feature Extraction for Sentiment Analysis. *Springer International Publishing*, 21-45.

101. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. ve Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 1990(3), 235-244.

102. Hu, M. ve Liu, B. (2004, Ağustos 22-25). *Mining and summarizing customer reviews*. Tenth Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) International Conference on Knowledge Discovery and Data Mining, ACM-168-177, Seattle, Washington, USA.

103. Alqasemi, F., Abdelwahab, A. ve Abd Elkader, H. A. (2017). Constructing automatic domain-specific sentiment lexicon using KNN search via terms discrimination vectors. *International Journal of Computers and Applications*, 41(3), 1-11.

104.Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G. ve Reynar, J. (2008, Nisan 22-24). *Building a sentiment summarizer for local service reviews*. Natural Language Processing Challenges in the Information Explosion Era (NLPIX-2008), Beijing, China.

105.Remus, R., Quasthoff, U., Heyer, G. ve Calzolari N. (Eds), (2010, Mayıs 17-23). SentiWS-A, *German-language resource for sentiment analysis*. International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta.

106.Oliveira, N., Cortez, P. ve Areal, N. (2016). Stock market sentiment lexicon acquisition using microblogging data and statistical measures. *Decision Support Systems*, 85, 62-73.

107.Baccianella, S., Esuli, A. ve Sebastiani, F. (2010). *Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining*. The Seventh International Conference on Language Resources and Evaluation, Valletta, Malta.

108.Vishnu, K.S., Apoorva, T. ve Gupta, D. (2014). *Learning domain-specific and domain independent opinion oriented lexicons using multiple domain knowledge*. Seventh International Conference on Contemporary Computing (IC3-2014), Noida, India.

109.Arslantürk, G. (2021). Anti-immigrant attitudes in social media: An examination within the frame of integrated threat theory. *Psychology Researches*, 1(1):06-16.

110.Zaabar, L.S., Yaakub, M.R., Abu Latiffi, M.I. (2022). Combination of lexicon based and machine learning techniques in the development of political tweet sentiment analysis model. *International Journal of Synergy in Engineering and Technology* (IJSET), 3(2), 72-83.

111.Dey, A, Jenamani, M., ve Thakkar, J.J. (2018). Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*. 103. 10.1016/j.eswa.2018.03.004.

112.Hota, H.S., Sharma, D.K. ve Verma, N. (2021). *Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad*. Data Science for COVID-19. (2021), 275–95. 10.1016/B978-0-12-824536-1.00015-0.

113.Muppuru, B., Vamsi, S., Jayashree, J., Sohan, L., Krishna, D.S. (2023). Real time sentimental analysis of chat data using deep learning. *Journal of Data Acquisition and Processing,* 38 (2), 4545-4556.

114.Bilgiç, E. ve Koçak, A. (2019). Duygu ve görüş madenciliği: Son gelişmeler ve 3. Havalimanı için bir uygulama. *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(17), 327-338.

115.Vural, G., Cambazoglu, B., Tokgoz, O. Z. ve Karagöz P. (2012). A framework for Sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish. *Computer and Information Sciences*, 437-445, 2012.

116.Türkmenoğlu, C. ve Tantuğ, A. C. (2014, Haziran 22-24). *Sentiment analysis in Turkish media*. International Conference on Machine Learning (ICML), Beijing, China.

117.Dehkharghani, R., Saygin, Y., Yanikoglu, B. ve Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*, 50(3), 667-685.

118.Güldal, H. ve Çakıcı, Y. (2017). Eğitsel Veri Madenciliği. Balkan Eğitim Araştırmaları-2017, *Trakya Üniversitesi Eğitim Fakültesi*, 135-143.

119.Tunalı, V., Bilgin, T. T. (2012). *Türkçe metinlerin kümelenmesinde farklı kök bulma yöntemlerinin etkisinin araştırılması*. Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu, Bursa, 598-602.

120.Göker, H. ve Tekedere, H. (2017). Fatih projesine yönelik görüşlerin metin madenciliği yöntemleri ile otomatik değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(3), 291-299.

121.Gürcan, F. (2017, Aralık 07-10). *Denetimli öğrenme yordamlarıyla Türkçe metinler üzerinde konu sınıflandırma analizi*. 1st International Turkish World Engineering and Science Congress, Antalya, Türkiye.

122.Torunoğlu, D., Çakırman, E., Ganiz, M. C., Akyokuş, S. ve Gürbüz, M. Z. (2011). *Analysis of preprocessing methods on classification of Turkish* texts. 2011 International Symposium on Innovations in Intelligent Systems and Applications (INISTA), 112-117, Piscataway, New Jersey, USA.

123.Oflazer, K. (1994). Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2), 137-148.

124.İnternet: Akın, A. A., Zemberek-nlp. URL: https://github.com/ahmetaa/zemberek-nlp, Son Erişim Tarihi: 05.07.2023.

125.İnternet: Wordnet-A lexical database for English, What is wordnet. URL: http://wordnet.princeton.edu, Wordnet Resmi Ağ Adresi, Son Erişim Tarihi: 05.07.2023.

126.İnternet: Topuz, M. Veri madenciliği, n-gram nedir? Neden kullanılır? URL: https://merttopuz.com/diger/veri-madenciligi/n-gram-nedir-neden-kullanilir, Son Erişim Tarihi: 22.12.2019.

127.Tang, B., Shepherd, M., Milios, E. ve Heywood, M. I. (2005). *Comparing and combining dimension reduction techniques for efficient text clustering*. Society for Industrial and Applied Mathematics International Workshop on Feature Selection for Data Mining, Newport Beach, CA, USA.

128.Sebastiani, F. (2002). Machine learning in automated text categorization. *Association for Computing Machinery Computing Surveys*, 34(1), 1-47.

129.Bordoloi M. ve Biswas, S.K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. *Artificial* Intelligence Review. https://doi.org/10.1007/s10462-023-10442-2.

130.Chang, J.R., Liang, H.Y., Chen, L.S.ve Chang, C.W. (2020). Novel feature selection approaches for improving the performance of sentiment classification. *Journal of Ambient Intelligence and Humanized Computing*. 10.1007/s12652-020-02468-z.

131.Chen, P.I. ve Lin, S.J. (2010). Automatic keyword prediction using google similarity distance. *Expert Systems Appllications*, 37(3):1928-1938.

132.Rana, M.R.R., Rehman, S.U., Nawaz, A., Ali, T., Imran, A, Alzahrani, A.ve Almuhaimeed, A. (2023). Aspect-based sentiment analysis for social multimedia: A hybrid computational framework. *Computer Systems and Engineering*, 46(2):2415-2428. 10.32604/csse.2023.035149.

133.Sun, T., Jing, L., Wei, Y., Song, X., Cheng, Z. ve Nie, L. (2023). Dual consistency-enhanced semi-supervised sentiment analysis towards COVID-19 tweets. *IEEE Transactions on Knowledge and Data Engineering*, 1-13. 10.1109/TKDE.2023.3270940.

134.Barreto, S., Moura, R., Carvalho, J., Paes, A.ve Plastino, A. (2023). Sentiment analysis in tweets: an assessment study from classical to modern word representation models. *Data Mining Knowledge Discovery* 37, 318–380. https://doi.org/10.1007/s10618-022-00853-0.

135.Yulia, D. ve Bilashova, E.A. (2021). *Learning analytics of MOOCs based on natural language processing*. CEUR Workshop Proceeding, 3017(15), 187-197.

136.Doğan, T. ve Uysal, A. K. (2018, Kasım 09-10). *Tıbbi metin dokümanlarının sınıflandırılmasında terim ağırlıklandırma yöntemlerinin başarımlarının kıyaslanması*. 6th International Symposium on Innovative Technologies in Engineering and Science (ISITES2018), Antalya, Türkiye.

137.Akba, F. (2014). *Duygu Analizinde Öznitelik Seçme Metriklerinin Değerlendirilmesi: Türkçe Film Eleştirileri*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, Ankara.

138.Değerli, O. (2012). *Naive Bayes yöntemi ile blog içeriklerinin* sınıflandırılması. Yüksek Lisans Tezi, Gazi Üniversitesi, Bilişim Enstitüsü, Ankara.

139.Diao, R. (2014). *Feature selection with harmony search and its application*. Doktora Tezi, Aberystwyth University, Computer Science, Penglais.

140.İnternet: URL: https://verimadencisi.blogspot.com/2011/06/oznitelikleri-secsek-de-mi-veri.html, Son Erişim Tarihi: 18.05.2020.

141.İnternet: Faggella, D., What is machine learning. URL: https://emerj.com/ai-glossary-terms/what-is-machine-learning/, Son Erişim Tarihi: 4.07.2023.

142.Bayer, H. ve Çoban, T. (2015). Web istatistiklerinde makine öğrenmesi algoritmaları ile kritik parametre tespiti. *Electronic Journal of Vocational Colleges, Special Issue: The Latest Trends in Engineering*, 5(4), 23-41.

143.Yılmaz, Ş.K. ve Patır, S. (2011). Kümeleme analizi ve pazarlama, *Akademik Yaklaşımlar Dergisi,* 2(1), 91-113.

144.Mahesh, B. (2018). Machine learning algorithms - A review. *International Journal of Science and Research (IJSR)*,9(1), 381-386.

145.İnternet: Torbalama (Bootstrap toplama), genel bakış, nasıl çalışır, avantajlar. URL: https://tr.pharoskc.com/1529-what-is-bagging-bootstrap-aggregation, Son Erişim Tarihi: 30.05.2023.

146.İnternet: Keskin, M. V. URL: https://www.veribilimiokulu.com/agaca-dayali-yontemlerde-bagging-ve-boosting-arasinda-ne-fark-var/, Son Erişim Tarihi: 28.01.2023.

147.İnternet: Bootstrap aggregating. URL: https://en.wikipedia.org/wiki/Bootstrap _aggregating, Son Erişim Tarihi: 27.01.2023.

148.İnternet: Singh, H. Understanding gradient boosting machines. URL: https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab, Son Erişim Tarihi: 28.01.2023.

149.İnternet: Keskin, M. V. Ağaca dayalı yöntemlerde bagging ve boosting arasında ne fark var? URL: https://www.veribilimiokulu.com/agaca-dayali-yontemlerde-bagging-ve-boosting-arasinda-ne-fark-var/, Son Erişim Tarihi: 27.01.2023.

150.İnternet: Ersoy, C. Extreme Gradient Boosting (XGBoost). URL: https://www.datasciencearth.com/extreme-gradient-boosting-xgboost/, Son Erişim Tarihi: 28.01.2023.

151.Kartal, E. ve Ozen, Z. Torkul, O., Gülseçen, S., Uyaroğlu, Y., Çağıl, G. ve Uçar, M.K. (Editörler) (2017). *Dengesiz veri setlerinde sınıflandırma. Mühendislikte Yapay Zeka ve Uygulamaları*, Sakarya Üniversitesi Kütüphanesi Yayınevi, 109-131.

152.İnternet: Imbalanced data. URL: https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data, Son Erişim Tarihi: 25.04.2023.

153.Chawla, N.V., Bowyer, K.W., Hall, L.O. ve Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 2002(16), 321–357.

154.İnternet: Anomaly detection. URL: https://en.wikipedia.org/wiki/Anomaly_detection, Son Erişim Tarihi: 16.01.2023.

155.Internet: Kasturi, S. N. Undefitting and overfitting in machine learning and how to deal with it? URL: https://towardsdatascience.com/underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6fe4a8a49dbf, Son Erişim Tarihi: 20.04.2023.

156.İnternet: Kocakafa, T. Aşırı öğrenme (Overfitting). URL: https://www.veribilimiokulu. com/ overfitting/, Son Erişim Tarihi: 23.01.2023.

157.İnternet: Çıtır E. Underfitting and overfitting. URL: https://www.datasciencearth.com/ underfitting-ve-overfitting/, Son Erişim Tarihi: 28.01.2023.

158.İnternet: Modeli hiper parametre ayarlama (v2). URL: https://learn.microsoft.com/tr-tr/azure/machine-learning/how-to-tune-hyperparameters, Son Erişim Tarihi: 28.01.2023.

159.İnternet: URL: https://learn.microsoft.com/tr-tr/azure/machine-learning/how-to-tune-hyperparameters, Son Erişim Tarihi: 28.01.2023.

160.Tran, N., Schneider, J.G., Weber, I ve Qin, A.K. (2020). Hyper-parameter optimization in classification: To-do or not-to-do. *Pattern Recognition*, 103 (107245).

161.Elgeldawi, E., Sayed, A., Galal, A.R., Zaki, A.M. (2021). Hyperparameter tuning for machine learning algorithms used for Arabic sentiment analysis. *Informatics* 8 (4), 79.

162.Atasever, Ü.H. (2011). *Uydu görüntülerinin sınıflandırılmasında hızlandırma (Boostıng), destek vektör makineleri, Rastgele Orman (Random Forest) ve Regresyon Ağaçları yöntemlerinin kullanılması*. Yüksek Lisans Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri.

163.Akçetin, E., ve Turgut, H. (2015). Eksik verili işletme veri tabanlarında veri analizi ve veri madenciliği uygulamaları. *Route Educational and Social Science Journal*, 2(2), 181-199.

164.Turgut, H. (2012). *Veri madenciliği süreci kullanılarak alzheimer hastalığı teşhisine yönelik bir uygulama*. Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.

165.Tantuğ, A. C. (2012). Metin sınıflandırma (text classification). *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5(2).

166.İnternet: What does AUC stand for and what is it? URL: https://stats.stackexchange.com/questions/132777/what-does-auc-stand-for-and-what-is-it, Son Erişim Tarihi: 16.01.2023.

167.Kleinbaum, D. G. (1994). *Logistic regression-A self-learning text. Springer Science and Business Media*, Edition Number:1, New York, USA.

168.İnternet: Samur, O. Boosting Algoritmaları. URL: https://www.datasciencearth.com/boosting-algoritmalari/, Son Erişim Tarihi: 28.01.2023.

169.Deconinck, E., Hancock, T., Coomans, D., Massart, D. L. ve Vander Heyden Y. (2005). Classification of drugs in absorption classes using the classification and regression trees methodology. *Journal of Pharmaceutical and Biomedical Analysis*, 39(1), 91-103.

170.Kurt, İ., Ture, M. ve Turhan Kurum, A. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Systems with Applications*, 34(1), 366-374.

171.Güner, Z. B. (2014). Veri madenciliğinde CART ve lojistik regresyon analizinin yeri: İlaç provizyon sistemi verileri üzerinde örnek bir uygulama. *Sosyal Güvenlik Uzmanları Derneği Sosyal Güvence Dergisi*, 2014(6), 53-99.

172.Teng, J. H., Lin, K. C. ve Ho, B. S. (2007). Application of classification tree and logistic regression for the management and health intervention plans in a community-based study. *Journal of Evaluation in Clinical Practice*, 13, 741-748.

173.Lee, C. H., Wang, Y. H. ve Trappey, A. J. C. (2015). Ontology-based reasoning fort he intelligent handling of customer complaints. *Computers and Industrial Engineering*, 2015(84) 144-155.

174.İnternet: Taşan, H. S. Machine learning dersleri-sınıflandırma. URL: https://www.kaggle.com/code/hasansezertaan/machine-learning-dersleri-s-n-fland-rma, Son Erişim Tarihi: 28.01.2023.

175.Ritha, N., Hayaty, N., Matulatan, T., Uperiati, A., Rathomi, M., Bettiza, M ve Farasalsabila, F. (2022). *Sentiment analysis of health protocol policy using K-nearest neighbor and cosine similarity*. ICSETDI-22 Proceedings of 1st International Conference Sustainable Engineering Development and Technological Innovation, 195-203, Indenosia.

176.Amrani, Y.A., Lazaar, M. ve Kadiri, K.E.E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 127, 511-520.

177.Patil, G., Galande, V., Kekan, V. ve Dange, K. (2014). Sentiment analysis using support vector machine. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(1), 2607-2612.

178.Arı, A. ve Berberler, M. E. (2017). Yapay sinir ağları ile tahmin ve sınıflandırma problemlerinin çözümü için ara yüz tasarımı. *Acta Infologica*, 1(2), 55-73.

179.İnternet: What are neural networks? URL: https://www.ibm.com/topics/neural-networks#:~:text=Neural%20networks%2C%20also%20known%20as,neurons%20signal%20to%20one%20another, Son Erişim Tarihi: 11.07.2023.

180.Kaynar, O. ve Taştan, S. (2009). Zaman serisi analizinde MLP yapay sinir ağları ve ARIMA modelinin karşılaştırılması. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 33, 2009, 161-172.

181.İnternet: Yiu, T. Understanding random forest-how the algorithm works and why it is so effective. https://towardsdatascience.com/understanding-random-forest-58381e0602d2, Son Erişim Tarihi: 28.01.2023.

182.İnternet: Şimşek, H. K. Makine öğrenmesi dersleri-random forest (sınıflandırma). URL: https://medium.com/data-science-tr/makine-öğrenmesi-dersleri-5-bagging-ve-random-forest-2f803cf21e07, Son Erişim Tarihi: 28.01.2023.

183.İnternet: Donges, N. Random forest: Complete guide for machine learning. URL: https://builtin.com/data-science/random-forest-algorithm#, Son Erişim Tarihi: 11.07.2023.

184.Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.

185.İnternet: Browniee, Jason, A. Gentle Introduction to XGBoost for Applied Machine Learning. URL: https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/, Son Erişim Tarihi: 11.07.2023.

186.İnternet: Muratlar, E. R. XGBoost nasıl çalışır? Neden iyi performans gösterir? URL: https://www.veribilimiokulu.com/xgboost-nasil-calisir/, Son Erişim Tarihi: 28.01.2023.

187.İnternet: Morde, V. XGBoost Algorithm: Long May She Reign. URL: https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d, Son Erişim Tarihi: 11.07.2023.

188.Chen, T. ve Guestrin, C. (Ağustos 2016). *XGBoost: A scalable tree boosting system.* KDD '16: Proceedings of the 22nd Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) International Conference on Knowledge Discovery and Data Mining, 785–794. DOI:10.1145/2939672.2939785.

189.Internet: Şeker, Ş. E. ID3 Tekrarlı ikilikçi ağaç, Iterative Dichotomiser 3. URL: https://bilgisayarkavramlari.com/2012/11/09/id3-tekrarli-ikilikci-agac-iterative-dichotomiser-3/, Son Erişim Tarihi: 30.05.2023.

190.Internet: Pektaş, M. Makine öğrenmesi algoritmaları-ID3 Algoritması. URL: https://medium.com/deep-learning-turkiye/makine-öğrenmesi-algoritmaları-id3-algoritması-71983b3e3b77, Son Erişim Tarihi: 30.05.2023.

191.Yıldırım, S. (2003). *Tümevarım Öğrenme Tekniklerinin C4.5'in İncelenmesi*, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

192.Özkan, Y. (2008). *Veri Madenciliği Yöntemleri*, Papatya Yayıncılık, 1.Baskı, İstanbul.

193.Çalış, A., Kayapınar, S. ve Çetinyokuş, T. (2014). Veri madenciliğinde karar ağacı algoritmaları ile bilgisayar ve internet güvenliği üzerine bir uygulama. *Endüstri Mühendisliği Dergisi*, 25(3-4), 2-19.

194.İnternet: Watson natural language understanding. URL: https://www.ibm.com/cloud/watson-natural-language-understanding, Son Erişim Tarihi: 31.05.2023.

195.İnternet: Documentation-cloud natural language API. URL: https://cloud.google.com/natural-language#section-5, Son Erişim Tarihi: 31.05.2023.

196.İnternet: Text analytics. URL: https://azure.microsoft.com/en-au/products/cognitive-services/text-analytics, Son Erişim Tarihi: 31.05.2023.

197.İnternet: Hutto, C. J. Documentation-natural language toolkit. URL: https://github.com/cjhutto/vaderSentiment, Son Erişim Tarihi: 19.04.2023.

198.Internet: Pandas. URL: https://pandas.pydata.org/, Son Erişim Tarihi: 19.04.2023.

199.İnternet: Text mining: Entity extraction. URL: https://rapidminer.com/products/text-and-sentiment-analysis/, Son Erişim Tarihi: 19.04.2023.

200.İnternet: What is sentiment analysis. URL: https://www.lexalytics.com/products/salience/sentiment-analysis, Son Erişim Tarihi: 19.04.2023.

201. Kızılkaya, Y., M. ve Oğuzlar A. (2017). Tweetlerin etkinliğinin sosyal ağlar ile irdelenmesi. *Uludağ Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 135-150, 2017.

202. Çelik, S. (2017). Web günlük dosyalarının analizi için Web kullanım madenciliğinin uygulanması. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 46(1), 62-75.

203. Krosuri, L.R. ve Aravapalli, R.S. (2023). Feature level fine grained sentiment analysis using boosted long short-term memory with improvised local search whale optimization. *Peer Journal Computer Science* 9:e1336. https://doi.org/10.7717/peerj-cs.1336.

204. Can, U. ve Alatas, B. (2017). Duygu analizi ve fikir madenciliği algoritmalarının incelenmesi. *International Journal of Pure Applied Sciences*, 3(1), 75-111.

205. Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y. ve Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2, 1-11. https://doi.org/10.1016/j.nlp.2022.100003.

206. Medhat, W., Hassan, A. ve Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*. 5, 1093-1113.

207. Singh, V., Rajesh, P., Ashraf, U. ve Pranav, W. (2013). *Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches*. Institute of Electrical and Electronics Engineers 5th International Conference on Knowledge and Smart Technology-KST-2013, 122-127. 10.1109/KST.2013.6512800.

208. Appel, O., Chiclana, F., Carter, J. ve Fujita, H. (2017). *IOWA & Cross-ratio uninorm operators as aggregation tools in sentiment analysis and ensemble methods*. Institute of Electrical and Electronics Engineers International Conference on Fuzzy Systems, 1-6. 10.1109/FUZZ-IEEE.2017.8015375.

209. Mirtalaie, M.A. ve Hussain, O.K. (2020). Sentiment aggregation of targeted features by capturing their dependencies: Making sense from customer reviews. International *Journal of Information Management,* 2020 (53). https://doi.org/10.1016/j.ijinfomgt.2020.102097.

210. Basiri, M.E., Kabiri, A., Abdar, M., Mashwani, W.K., Yen, N.Y. ve Hung J.C. (2019). The effect of aggregation methods on sentiment classification in Persian reviews. *Enterprise Information Systems*, 2020 (14), 1394-142. https://doi.org/10.1080/17517575.2019.1669829.

211. Li, Y., Pan, Q., Yang, T., Wang, S., Tang, J. ve Cambria, E. (2017). Learning word representations for sentiment analysis. *Cognitive Computing*, 9(6), 843–851.

212. Benghuzzi, H. ve Elsheh, M.M. (2020). An investigation of keywords extraction from textual documents using word2vec and decision tree. *International Journal of Computer Science and Information Security (IJCSIS)*, 18(5), 13-18.

213. Suneetha, S. ve Row, V. (2023). Aspect-based sentiment analysis: A comprehensive survey of techniques and applications. *Journal of Data Acqusition and Processing-JCST*, 38 (3), 177-203. https://sjcjycl.cn/ DOI: 10.5281/zenodo.777648.

214.Ounacer, S., Mhamdi, D., Ardchir, S., Daif, A. ve Azzouazi, M. (2023). Customer sentiment analysis in hotel reviews through natural language processing techniques. *International Journal of Advanced Computer Science and Applications*, 14(1), 569-579. 0.14569/IJACSA.2023.0140162.

215.Du, K., Xing, F., Cambria, E. (2023). Incorporating multiple knowledge sources for targeted aspect-based financial sentiment analysis. *ACM Transactions on Management Information Systems*, 4(3), 1-24. https://doi.org/10.1145/3580480.

216.Rodrigues, A.P., Fernandes, R., Shetty, A., Lakshmanna, K. ve Shafi, R.M. (2022). Real-time Twitter spam detection and sentiment analysis using machine learning and deep learning techniques. *Computational Intelligence and Neuroscience*, 2022:5211949, 14 pages. https://doi.org/10.1155/2022/5211949.

217.O'Keefe, T. ve Koprinska, I . (2009). *Feature selection and weighting methods in sentiment analysis.* Proceedings of the 14th Australasian Document Computing Symposium, Sydney, 67-74.

218.Giachanou, A. ve Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys* (CSUR), 49(2), 1-41.

219.Taj, S., Meghji, A.F. ve Shaikh, B.B. (2019). *Sentiment Analysis of News Articles: A Lexicon based approach.* 2nd International Conference on Computing Mathematics & Engineering Technologies-2019 (iCoMET). 10.1109/ICOMET.2019.8673428.

220.Ghasemi, R., Ashrafi Asli S.A. ve Momtazi, S. (2020). Deep Persian sentiment analysis: Cross-lingual training for low-resource languages. *Journal of Information Science*, 48(1).

221.Kılıçer, S. ve Samlı, R. (2022). Farklı dillerdeki metinler üzerinde yapılan duygu analizi çalışmalarının incelenmesi. *DÜMF Mühendislik Dergisi*. 13(3), 493-509. 9509. 10.24012/dumf.1143479.

222.İnternet: Faggella, D. (2018). What is artificial intelligence, an informed definition. https://emerj.com/ai-glossary-terms/what-is-artificial-intelligence-an-informed-definition/. Son Erişim Tarihi: 28.04.2023.

223.Sağlam, M. (2022). Key themes in brand reputation research: A bibliometric analysis with VOSviewer software. *Research Journal of Business and Management*, 9(1), 1-12. 10.17261/Pressacademia.2022.1536.

224.Alantari, H.J., Currim, I.S., Deng, Y. ve Singh, S. (2021). An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. *International Journal of Research in Marketing*, 39 (1), 1-19. https://doi.org/10.1016/j.ijresmar.2021.10.011.

225.Monika, P., Kulkarni, C., Harish Kumar, N., Shruthi, S. ve Vani, V. (2022). Machine learning approaches for sentiment analysis: A survey. *International Journal of Health Sciences*, 6 (S4), 1286-1300.

226.Saifullah, S., Drężewski, R., Dwiyanto, F. A, Aribowo, A.S. ve Fauziah, Y. (2023). Sentiment analysis using machine learning approach based on feature extraction for anxiety detection. *Computational Science*-ICCS, 2023(6), 365-372. 10.1007/978-3-031-36021-3_38.

227.Gulati, K., Kumar, S.S., Kumar Boddu, R.S., Sarvakar, K., Sharma, D.K., Nomani, M.Z.M. (2021). Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to COVID-19 pandemic. *Materials Today: Proceedings*, 51(1),38-41. https://doi.org/10.1016/j.matpr.2021.04.364.

228.Chen, Y., Rao, Y., Chen, S., Lei, Z., Xie, H., Lau, R.Y. K, ve Yin, J. (2023). Semi-supervised sentiment classification and emotion distribution learning across domains. *ACM Transactions on Knowledge Discovery from Data*, 17 (5), 1-30. https://doi.org/10.1145/3571736.

229.Shahzad, M., Freeman, C., Rahimi, M. ve Alhoori, H. (2023). Predicting Facebook sentiments towards research. *Natural Language Processing Journal*, 3 (100010).

230.Vatambeti, R., Mantena, S.V., Kiran, K.V.D. ve Manohar, M. (2023). Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique. *Cluster Computing*. https://doi.org/10.1007/s10586-023-03970-7.

231.Wang, Y., Guo, J., Yuan, C. ve Li, B. (2022). Sentiment analysis of Twitter data. *Applied Sciences,* 12 (22). 10.3390/app122211775. https://doi.org/10.3390/.

232.Al Amrani, Y., Lazaar, M. ve El Kadiri, K.E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science* 127:511-520.

233.Thelwall, M, Buckley, K, Paltoglou, G., Cai, D. ve Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

234.Bai, X. (2011). Predicting consumer sentiments from online text. *Decision Support Systems,* 50(4), 732-742.

235.Kaur, C., Boush, M.S.A., Hassen, S.M., Hakami, W.A., Ab-dalraheem, M.H.O., Galam, N.M., Hadi, N.A., Hadi NA ve Benjeed, A.O.S. (2022). Incorporating sentimental analysis into development of a hybrid classification model. *International Journal of Health Science,* 6 (S1), 1709–1720.

236.Tiwari, D., Nagpal, B., Bhati, B.S., Mishra, A. ve Kumar, M. (2023). A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-023-10472-w.

237.Khaleghparast, S., Maleki, M., Hajianfar, G., Soumari, E., Ovveisi, M., Golandouz, H.M., Noohi, F., Dehaki, M.G., Golpira, R., Mazloomzadeh, S., Arabian, M. ve Kalayinia, S. (2023). Development of a patients' satisfaction analysis system using machine learning and lexicon-based methods. *BMC Health Services Research*, 23, 280. https://doi.org/10.1186/s12913-023-09260-7.

238.Wankhade, M., Rao, A.C.S. ve Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55, 5731–5780. https://doi.org/10.1007/s10462-022-10144-1.

239.Kristina, M., Marian, M. ve Miroslava, H. (2020). Classification of special web reviewers based on various regression methods. *Acta Polytechnica Hungarica*. 17, 229-248. 10.12700/APH.17.3.2020.3.12.

240.Dey, R.K. ve Das, A.K. (2023). Modified term frequency-inverse document frequency based deep hybrid framework for sentiment analysis. *Multimedia Tools and Applications*, 82, 32967–32990.

241.Erşahin, B., Aktaş, Ö. Kılınç, D. ve Erşahin, M. (2019). A hybrid sentiment analysis method for Turkish, *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(3), 1780-1793. 10.3906/elk-1808-189.

242.Madani, Y., Erritali, M. ve Bouikhalene, B. (2023). A new sentiment analysis method to detect and analysis sentiments of Covid-19 Moroccan tweets using a recommender approach. *Multimedia Tools and Applications* 82, 27819–27838.

243.Gopalachari, M.V., Gupta, S., Rakesh, S., Jayaram, D. ve Venkateswara Rao, P. (2023). Aspect-based sentiment analysis on multi-domain reviews through word embedding. *Journal of Intelligent Systems*, 32(1),

244.Alexandrovna, M. M. ve Ali Mohsin A-L.M. (2023). Tweet sentiment analysis with CNN and XG-BOOST. *Sustainability*, 21(4), 1-9.

245.Bandhakavi, A. S., Nirmalie, W., Deepak, P. ve Stewart, M. (2016). Lexicon based feature extraction for emotion text classification. *Pattern Recognition Letters*. 93.10.1016/j.patrec.2016.12.009.

246.Saranya, S. ve Usha, G. (2023). A Machine learning-based technique with IntelligentWordNet lemmatize for Twitter sentiment analysis. *Intelligent Automation and Soft Computing*, 36(1), 339- 352.

247.Kaur, G. ve Sharma, A. (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data*, 10 (5).

248.Alshehri, A. ve Algarni, A. (2023). TF-TDA: A novel supervised term weighting scheme for sentiment analysis. *Electronics* 12(7), 1632-1642.

249.Sharma, A. ve Kumar, S. (2023). Machine learning and ontology-based novel semantic document indexing for information retrieval. *Computers and Industrial Engineering* 176, 108940, https://doi.org/10.1016/j.cie.2022.108940.

250.Eroğul, U. (2009). *Sentiment analysis in Turkish*, Yüksek Lisans Tezi, Ortadoğu Teknik Üniversitesi Bilgisayar Bilimleri, Ankara.

251.Kaya, M., Fidan, G., Toroslu, I. H. (2012, Aralık 04-07). *Sentiment analysis of Turkish political news.* International Conferences on Web Intelligence (WI-2012), Institute of Electrical and Electronics Engineers Computer Society, 174-180, Macau, China.

252.Boynukalın, Z. (2012). *Makine öğrenimi teknikleriyle Türkçe metinlerde duygu analizi.* Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

253.Akbaş, E. (2012). *Aspect based opinion mining on Turkish tweets*. Yüksek Lisans Tezi, Bilkent Üniversitesi, Mühendislik ve Fen Bilimleri Enstitüsü, Ankara.

254.Tokcaer, S. (2021). Türkçe metinlerde duygu analizi sentiment analysis in Turkish texts. *Journal of Yasar University*, 16(63), 1514-1534.

255.Aytekin, Ç. (2013). An opinion mining task in Turkish language: A model for assigning opinions in Turkish blogs to the polarities. *Journalism and Mass Communication*, 2013(3) 179-198, 2013.

256.Thelwall, M., Buckley, K., Paltoglou, G., Cai , D. ve Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

257.Özsert, C. M. ve Özgür, A. (2013, Mart 24-30). *Word polarity detection using a multilingual approach*. Computational Linguistics and Intelligent Text Processing, 14th International Conference (CICLing-2013), Samos, Greece.

258.Dehkharghani, R. (2015). *Sentiment analysis in Turkish: Resources and techniques*. Doktora Tezi, Sabancı Universitesi, Mühendislik ve Doğa Bilimleri, İstanbul.

259.Baykara, M.; Göktürk, U. (2017). *Classification of social media shares using sentiment analysis.* UBMK'17, 2nd International Conference on Computer Science and Engineering, 911-916. DOI: 10.1109/UBMK.2017.8093536.

260.Fan, W., ve Gordon, M. D. (2014). The power of social media analytics. *Communications of the Association for Computing Machinery*, 57(6), 74-81.

261.He, W., Zhang, W., Tian, X., Tao, R. ve Akula, V. (2019). Identifying customer knowledge on social media through data analytics. *Journal of Enterprise Information Management* 32(1), 152-169.

262.Shi, Z., Rui, H. ve Whinston, A., B. (2014). Content sharing in a social broadcasting environment: Evidence from Twitter. *Management Information Systems Quarterly*, 38(1), 123-142, 2014.

263.Oh, O., Aggrawal, M. ve Rao, H.R. (2013), Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises, *Management Information Systems Quarterly*, 37(2), 407-426.

264.Rui, H., Liu, Y. ve Whinston, A. (2013). Whose and what chatter matters? The effects of tweets on movie sales. *Decision Support Systems*, 55(4), 863-870.

265.İnternet: Çalışma kitabını metin biçimine kaydetme (.txt veya .csv). URL: https://support.microsoft.com/tr-tr/office/çalışma-kitabını-metin-biçimine-kaydetme-txt-veya-csv-3e9a9d6c-70da-4255-aa28-fcacf1f081e6, Son Erişim Tarihi: 11.07.2019.

266.Morstatter, F., Fred, M., Pfeffer, J., Liu, H. ve Carley, K.M. (2013). Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose. *Association for the Advancement of Artificial Intelligence*, 7(1), 400-408.

267.İnternet: About different types of tweets. URL: https://help.twitter.com/en/using-twitter/types-of-tweets, Son Erişim Tarihi: 13.08.2019.

268.İnternet: URL: https://wearesocial.com/global-digital-report-2019, Son Erişim Tarihi: 5.08.2019.

269.İnternet: Aslam, S. Twitter by the numbers: Stats, demographic and fun facts. URL: https://www.omnicoreagency.com/twitter-statistics/, Son Erişim Tarihi: 05.12.2019.

270.Knoke, D. ve Kuklinski, J. H. (1982). *Network Analysis: Quantitative Application in the Social Sciences.* SAGE Publications, London, England.

271.Kleinberg, J. (2008). The Convergence of social and technological networks. *Communications of the Association for Computing Machinery*, 51(11), 66-72.

272.Savage, N. (2011). Twitter as medium and message. *Communications of the Association for Computing Machinery*, 54(3), 18-20, 2011.

273.Adak, A., Pradhan, B. ve Shukla, N. (2022). Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review. *Foods*, 11(10), 1500 https://doi.org/10.3390/foods11101500.

274.Ahmed, K., Nadeem, M.I., Zheng, Z., Li, D., Ullah, I., Assam, M., Ghadi, Y.Y., ve Mohamed, H.G. (2023). Breaking down linguistic complexities: A structured approach to aspect-based sentiment analysis. *Journal of King Saud University-Computer and Information Sciences*, 35 (8), 1-21.

275.Aktaş, Ö., Coskuner, B. ve Soner, İ., (2021). Turkish Sentiment Analysis Using Machine Learning Methods: Application on Online Food Order Site Reviews. *Journal of Artificial Intelligence and Data Science*, 1(1), 1-10.

276.Girgin, A. B. A. ve Şahin, S. (2023). Improving the performance of sentiment analysis by ensemble hybrid learning algorithm with NLP and cascaded feature extraction. *International Journal of Advances in Engineering and Pure Sciences*, 35 (1), 125-141.

277.Burt, R. S., Kilduff M. ve Tasselli, S. (2013). Social network analysis: Foundation and frontiers on advantage. *Annual Review of Psychology*, 64, 527-547, 2013.

278.Wasserman, S., ve Faust, K. (1994). *Social network analysis, methods and applications*. Cambridge University Press, 8, Cambridge, USA.

279.Pantelidis, İ. S. (2010). Electronic meal experience: A Content analysis of online restaurant comments, food service management. *Cornell Hospitality Quarterly* 51(4) 483–491.

280.İnternet: ScraperWiki has two name, quickcode and the sensible code company. URL: https://scraperwiki.com., Son Erişim Tarihi:06.07.2023.

281.Gebert, H., Geib, M., Kolbe, L. ve Reimpp, G. (2002, Aralık 10-13). *Towards customer knowledge management: Integrating customer relationship management and knowledge management concepts*. The Second International Conference on Electronic Business, Taipei, Taiwan.

282.Pang, B., Lee, L. ve Vaithyanathan, S. (2002). *Thumbs up?: Sentiment classification using machine learning techniques*. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002),79–86, (Association for Computational Linguistics. DOI: 10.3115/1118693.1118704.

283.İnternet: Custumor retention reimaged. URL: www.voziq.com., Son Erişim Tarihi: 06.07.2023.

284.Özaslan, Y. ve Uygur, S., M. (2014). Negatif ağızdan ağıza iletişim ve elektronik ağızdan ağıza iletişim (E-WOM): Yiyecek-içecek işletmelerine yönelik bir araştırma. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 28(3), 69-88.

285.Costache, M. (2010). Qualman, Eric. (2009) Socialnomics: how social media transforms the way we live and do business. *Journal of Comparative Research in Anthropology and Sociology*, 1 (2), 225-228.

286.Allsop, D. T., Bassett, B. R. ve Hoskins, J. A. (2007). Word-of-mouth research. *Principles and Applications*, 47, 2007. DOI:10.2501/S0021849907070419.

287.Bovée, C.L., Housten, M.J. and Thill J.V. (1995). *Marketing.* McGraw-Hill, 2nd edition, London, England.

288.Sundaram, D. S., Mitra, K., ve Webster, C. (1998). Word-of-mouth communications: a motivational analysis. *Association for Consumer Research*, 25, 527-531.

289.Armağan, E. ve Eskici, Y. (2019). Tüketicilerin online yemek servislerine karşı tutum, davranış ve satın alma niyetleri. *Erzurum Kültür Eğitim Vakfı Akademi Dergisi, International Conference on Applied Economics and Finance Özel Sayısı*, 39-75.

290.Namkung, Y. ve Jang, S. (2007). Does food quality really matter in restaurants? Its impact on customer satisfaction and behavioral intentions. *Journal of Hospitality ve Tourism Research*, 31, 387-409.

291.Liu, Y. ve Jang, S. (2009). Perceptions of Chinese restaurants in the U.S.: What affects customer satisfaction and behavioral intentions? *International Journal of Hospitality Management*, 28, 338-348.

292.Zahoor, K., Bawany, N.Z. ve Hamid, S. (2020). *Sentiment analysis and classification of restaurant reviews using machine learning*. 21st International Arab Conference on Information Technology, Giza, 1-6, Giza.

293.Alamoudi, E.S. ve Alghamdi ,N.S. (2021). Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings. *Journal of Decision Systems*, 30(2-3), 59-281. 10.1080/12460125.2020.1864106.

294.Liapakis, A., Tsiligiridis, T., Yialouris, C. ve Darren, S. (2020). A Sentiment lexicon-based analysis for food and beverage industry reviews, the Greek language paradigm. *Natural Language Engineering*. 2020 (9), 21-42.

295.Ağcasulu, H. (2018). Sosyal Bilimlerde İlişkileri İnceleyen Bir Yöntem: Sosyal Ağ Analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 22(Özel Sayı), 1915-1933.

296.Dujin, M. A. J. V. ve Vermunt, J. K. (2006). What is special about social network analysis. *Methodology*, 2(1), 2-6.

297.Codal, K. S. ve Coşkun, E. (2016). Sosyal ağ türlerinin karşılaştırılmasına ilişkin bir ağ analizi. *AİBÜ Sosyal Bilimler Enstitüsü Dergisi*, 16(1), 143-158.

298.Borgatti, S. P., Everett, M. G.ve Jonhson, J. C. (2013). *Analyzing Social Networks*. New Delhi, Sage Publications.

299.Chau, M. ve Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *Management Information Systems Quarterly*, 36(4), 1189-1216.

300.Girvan, M. ve Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*. 99, 7821–7826.

301.İnternet: Jayawickrama, T.D. Community detection algorithms. https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae, Son Erişim Tarihi:29 Ocak 2021.

302.İnternet: Anderson, J. ve Rainie, L. (2017). The Future of truth and misinformation online. Pew Research Center. URL: https://www.pewresearch.org/internet /2017/10/19/the-future-of-truth-and-misinformation-online/.

303.Garcia, M.F., Alan, J.D. ve Sánchez-Cabezudo, S.S. (2016). Identifying the new influences in the Internet era: Social media and social network analysis. *Revista Española de Investigaciones Sociologicas*, 2016 (153), 23-40.

304.Huang, X., Chen, D., Ren, T. ve Wang, D. (2020). A survey of community detection methods in multilayer networks. *Data Mining and Knowledge Discovery*, 35, 1–45.

305.Luo, L., Liu, K., Guo, B. ve Ma, J. (2020). User interaction-oriented community detection based on cascading analysis. *Information Sciences*, 2020(510), 70-88.

306.Ishfaq, U., Khan, H.U. ve Iqbal, S. (2022). Identifying the influential nodes in complex social networks using centrality-based approach. *Journal of King Saud University Computer and Information Sciences*, (2022 (34), 9376–9392.

307.Fei, L., Mo, H. ve Deng, Y. (2017). A new method to identify influential nodes based on combining of existing centrality measures. *Modern Physics Letters B* 2017 (31), 257-267..

308.Zareie, A., Sheikhahmadi, A., Jalili, M. ve Fasaei, M.S.K. (2020). Finding influential nodes in social networks based on neighborhood correlation coefficient. *Knowledge-Based Systems*, 194-204.

309.Yang, X.H., Xiong, Z., Ma, F., Chen, X., Ruan, Z., Jiang, P. ve Xu, X. (2021). Identifying influential spreaders in complex networks based on network embedding and node local centrality. *Physica A: Statistical Mechanics and its Applications*, 2021 (573).

310.Ullah, A., wang, B., Sheng, J., Long, J., Khan, N., Gambuzza, L.V. (2021). Identification of influential nodes via effective distance-based centrality mechanism in complex networks. *Complexity*, 2021 (1–16).

311.Wan, Y.P., Wang, J., Zhang, D.G., Dong, H.Y. ve Ren, Q.H. (2018). Ranking the spreading capability of nodes in complex networks based on link significance. *Physica A: Statistical Mechanics and its Applications*, 2018(503), 929-937.

312.Zhong, L.F., Liu, Q.H., Wang, W. ve Cai, S.M. (2018). Comprehensive influence of local and global characteristics on identifying the influential nodes. *Physica A: Statistical Mechanics and its Applications*, 2018 (511), 78-84.

313.Rehman, A.U., Jiang, A., Rehman, A., Paul, A., Din, S. ve Sadiq, M.T. (2020). Identification and role of opinion leaders in information diffusion for online discussion network. *Journal of Ambient Intelligence and Humanized Computing*, 2020 (1–13).

314.Zhao, J., Wang, Y. ve Deng, Y. (2020). Identifying influential nodes in complex networks from global perspective. *Chaos, Solitons Fractals*, 133, 1777–1787.

315.İspir, N. B. ve Deniz, K. (2017). Kasım 2015 genel seçimlerinde köşe yazarlarının Twitter gündemine yönelik bir sosyal ağ analizi uygulaması. *Kurgu*, 25-1, 77-83.

316.Arslan, C., Gümüş B. ve Öztürk, İ. D. (2019). Türkiye'de ekoloji hareketlerinin sınırlı sosyal medya kullanımı: Ekoloji birliği Twitter ağı analizi örneği. *Connectist: İstanbul University of Journal Communication Sciences*, 56, 31-66.

317.Demir Y. ve Ayhan B. (2020). Sosyal medyanın gündem belirleyicileri: Twitter'da gündem belirleme süreci üzerine bir sosyal ağ analizi. *Ankara Hacı Bayram Veli Üniversitesi İletişim Kuram ve Araştırma Dergisi*, 2020(51), 1-18.

318.Bakan, U. (2020). Sanat okullarının Twitter kullanım karakteristiklerine ilişkin bir sosyal ağ analizi perspektifi. *Aydın Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 7(1), 138-155.

319.Kobak, K. (2022). #TikTokkapansın hareketi: Twitter'da sosyal ağ analizi. *MANAS Sosyal Araştırmalar Dergisi*, 11(1), 309-319.

320.Yee, T.P, Ismail, H.N. ve Jafaar, S.M.R.S. (2021). A Systematic literature review of social network analysis in tourism flows. *Journal of Tourism Hospitality and Environment Management*, 6 (26), 143-154.

321.Es'haghi, S. R. ve Karamidehkordi, E. (2023) Understanding the structure of stakeholders-projects network in endangered lakes restoration programs using social network analysis. *Environmental Science and Policy*, 140, 172-188.

322.Rodda, I. ve Bhavani, D. (2023). *Visualition of the dynamics in character networks using social network analysis*. CODS-COMAD'23: Proceedings of the 6th Joint International Conference on Data Science and Management of Data, the Association for Computing Machinery Digital Library, 284-288.

323.Costa, A. R., Ralha, C. G. (2023). AC2CD: An Actor-critic architecture for community detection in dynamic social networks. *Knowledge-Based Systems*, 261 2023(110202). DOI: https://doi.org/10.1016/j.knosys.2022.110202.

324.Hastuti, H., Maulana, H., Tompo, A. ve Ferizka, Z. (2023). Analysis of social media opinion on the representation of the 2024 presidential election on Twitter: A social network analysis. *Jurnal Studi Ilmu Pemerintahan*, 4-1, 117-128.

325.Aragon, E. (2023). Connecting sunken actors: Social network analysis in maritime archaeology. *Springer Nature, Historical Archieves* (2023), 57(5916). DOI:10.1007/s41636-023-00385-4.

326.Pang, N., Sun, M. ve Zhu, H. (2023). Louise or Ferdinand? Exploring the protagonists of love and intrigue using social network analysis. *Digital Scholarship in the Humanities.* 38(3), 1214-1226. DOI:10.1093/llc/fqad007.

327.Purbasari, R., Munajat, E. ve Fauzan, F. (2023). Digital innovation ecosystem on digital entrepreneur: Social network analysis approach. *International Journal of E-Enterpreneurship and Innovation*, 13(1), 21.

328.Kim, Y., Choi, T. Y., Yan, T. ve Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. *Journal of Operations Management*, 2011(29), 194-211.

329.Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1, 201-233.

330.Everett, M. G. ve Borgatti, S. P. (2013). The Dual-projection approach for two-mode networks. *Social Networks*, 35(2), 204-210.

331.Al-Garadi, M.A., Varathan, K.D., Ravana, S.D., Ahmed, E. ve Chang, V. (2016). Identifying the influential spreaders in multilayer interactions of online social networks. *Journal of Intelligent and Fuzzy Systems*, 31, 2721–2735.

332.Prell, C. (2012). *Social network analysis: History. Theory and Methodology*, SAGE Publications, Londra, İngilter.

333.Özdemir, A. A. (2008). Sosyal ağ özellikleri bakış açısıyla sosyal sermaye ve bilgi yaratma ilişkisi: Akademisyenler üzerinde yapılan bir alan araştırması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 8(2), 81-102.

334.Demiral, D.G. (2020). Ağ bilimi ve modeller. *Uygulamalı Ekonomi ve Sosyal Bilimler Dergisi*, 2(2), 36-55.

335.Britta Ruhnau, B. (2000). Eigenvector-centrality-a node-centrality. *Social Networks*, 22(4), 357-365. URL: https://doi.org/10.1016/S0378-8733(00)00031-9.

336.Gürsakal, N. (2009). *Sosyal Ağ Analizi (Pajek, UCINET ve Gmine Uygulamalı)*, Dora Yayıncılık, Bursa.

337.llabas, B., Janson S. ve Riordan, O. (2007). The Phas Transition in Inhomogeneous Random Graphs. *Random Structures and Algorithms*, 2007 (31), 3-122.

338.Watts, D. J., Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature-Macmillan Publishers*, 393 (6684): 440–442.

339.Barabasi, A. L. ve Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286 (5439), 509–512.

340.Internet: NetworkX Developers NetworkX-network analysis in Python. URL: https://networkx.org, Son Erişim Tarihi: 02.04.2023.

341.Internet: Centrality. URL: https://networkx.org/documentation/stable/reference/algorithms/centrality.html, Son Erişim Tarihi: 02.04.2023.

342.Internet: Louvain_communities. URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.louvain.louvain_communities.html, Son Erişim Tarihi: 02.04.2023.

343.Internet: Girvan_newman. URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.centrality.girvan_newman.html, Son Erişim Tarihi: 02.04.2023.

344.Internet: Label_propagation.URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.label_propagation.label_propagation_communities.html, Son Erişim Tarihi: 02.04.2023.

345.Internet: Erdos_Renyi_graph. URL: https://networkx.org/documentation/stable/reference/generated/networkx.generators.random_graphs.erdos_renyi_graph.html, Son Erişim Tarihi: 02.04.2023.

346.Internet: Watts_Strogatz_graph. URL: https://networkx.org/documentation/stable/reference/generated/networkx.generators.random_graphs.watts_strogatz_graph.html, Son Erişim Tarihi: 02.04.2023.

347.Internet: Shortest paths. URL: https://networkx.org/documentation/stable/reference/algorithms/shortest_paths.htm,Son Erişim Tarihi: 02.04.2023.

348.Internet: Connected_components. URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.components.connected_components.html, Son Erişim Tarihi: 02.04.2023.

349.Internet: Clustering. URL: https://networkx.org/documentation/stable/reference/algorithms/clustering.html, Son Erişim Tarihi: 02.04.2023.

350.Internet: Adamic_Adar_index. URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.link_prediction.adamic_adar_index.html, Son Erişim Tarihi: 02.04.2023.

351. Internet: Jaccard_coefficient. URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.link_prediction.jaccard_coefficient.html, Son Erişim Tarihi: 02.04.2023.

352. Internet: Preferential_attachment.URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.link_prediction.preferential_attachment.html, Son Erişim Tarihi: 02.04.2023.

353. İnternet: Schuhmacher, A. Fruchterman-Reingold algorithm, drawing graphs with circular vertices. Schneide blog, Public Business Secrets. URL: https://schneide.blog/tag/fruchterman-reingold/, Son Erişim Tarihi: 25.05.2020.

354. İnternet: Harel, D. ve Koren, Y. (2000) A Fast multi-scale method for drawing large graphs. Proceedings of the working conference on Advanced visual interfaces, 282-285. DOI: https://doi.org/10.1145/345513.345353.

355. İnternet: Diagram layout references. URL: https://pro.arcgis.com/en/pro-app/latest/help/data/network-diagrams/force-directed-layout-reference.htm, Son Erişim Tarihi: 03.04.2023.

356. İnternet: Drawing-graph_layout. URL: https://networkx.org/documentation/stable/reference/drawing.html , Son Erişim Tarihi: 02.04.2023.

357. İnternet: Graphviz-layout engines. URL: https://graphviz.org/docs/layouts/, Son Erişim Tarihi: 02.04.2023.

358. İnternet: Yemek siparişinde piyasaya yeni giren Glovo'yu, Yemeksepeti ile kıyasladık. URL: https://webrazzi.com/2018/12/21/yemek-siparisinde-piyasaya-yeni-giren-glovoyu-yemeksepeti-ile-kiyasladik/, Son Erişim Tarihi: 23.10.2019.

359. İnternet: Şikayetvar. URL: https://www.sikayetvar.com/yemek-sepeti?page=5, Son Erişim Tarihi: 10.08.2019.

360. İnternet: Şikayetvar. URL: https://www.sikayetvar.com/dominos-pizza/ memnuniyet/ siparis, Son Erişim Tarihi: 10.08.2019.

361. İnternet: Şikayetvar. URL: https://www.sikayetvar.com/hot-doner/siparis/yemek, Son Erişim Tarihi: 10.08.2019.

362. İnternet: Şikayetvar. URL: https://www.sikayetvar.com/bereket-doner, Son Erişim Tarihi: 10.08.2019.

363. İnternet: Şikayetvar. URL: https://www.sikayetvar.com/yemek-sepeti/lahmacun, Son Erişim Tarihi: 10.08.2019.

364. İnternet: Şikayetvar. URL: https://www.sikayetvar.com/hacioglu-restaurant, Son Erişim Tarihi: 10.08.2019.

365. İnternet: Şikayetvar. URL: https://www.sikayetvar.com/yemek-sepeti/firma/hamburger, Son Erişim Tarihi: 10.08.2019.

366.İnternet: Şikayetvar. URL: https://www.sikayetvar.com/yemek-sepeti/pide, Son Erişim Tarihi: 10.08.2019.

367.Bruns, A. ve Burgess, J. (2011). #ausvotes: How Twitter covered the 2010 Australian federal election. *Communication, Politics & Culture,* 44(2), 37-56.

368.Aydoğan, E. K., Gencer, C. ve Akbulut, S. (2008). Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri bölümlenmesi. *Sigma*, 26(1), 43-57.

369.İnternet: İstatistik veri portalı. https://data.tuik.gov.tr/Search/Search?text=adrese%20dayalı%20nüfus&dil=1, Son Erişim Tarihi: 18.01.2021.

370.İnternet: URL: https://www.clarabridge.com/, Son Erişim Tarihi: 19.01.2021.

371.Liu, Z., Yang, C., Rüdian, S., Liu, S., Zhao, L. ve Wang, T .(2019). Temporal emotion-aspect modeling for discovering what students are concerned about in online course forums. *Interactive Learning Environment*. 27 (5–6), 598–627. http://dx.doi.org/10.1080/ 10494820.2019.1610449.

372.İnternet: Marr, B. (2013). The Awesome ways big data is used today to change our world. URL: https://www.linkedin.com/pulse/20131113065157-64875646-the-awesome-ways-big-data-is-used-today-to-change-our-world/, Son Erişim Tarihi: 21 Mart 2019.

373.Muhammed, M., Rustam, F., Alasim, F., Siddique, M. ve Ashraf, I. (2023). What people think about fast food: Opinions analysis and LDA modeling on fast food restaurants using unstructured tweets. *PeerJ Computer Science* 9:e1193. https://doi.org/10.7717/peerj-cs.1193.

374.İnternet: Colombia University, Social Network Analysis. URL: https://www.publichealth.columbia.edu/research/population-health-methods/social-network-analysis, Son Erişim Tarihi: 17 Eylül 2023.

375.Dwivedi, Y.K., Ismagilova, E., Y., Hughes, D.L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A.S., Kumar, V., Rahman, M.M., Raman, R., Rauschnabel, P.A., Rowley, J., Salo, J., Tran, G.A. ve Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, 2021 (59).

376.Internet: Katz_centrality. URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.katz_centrality.html, Son Erişim Tarihi: 02.04.2023.

377.Internet: Algorithm. URL: https://networkx.org/documentation/stable/reference/algorithms/index.html, Son Erişim Tarihi: 02.04.2023.

378.Internet: Fruchterman_Reingold_layout. URL: https://networkx.org/documentation/networkx-1.11/reference/generated/networkx.drawing.layout.fruchterman_reingold_layout.html, Son Erişim Tarihi: 02.04.2023.

379.Uğurlu, O. (2022). Comparative analysis of centrality measures for identifying critical nodes in complex networks. *Journal of Computational Science*, 62.

380.Granovetter, M.S. (2015). The Strength of Weak Ties. *American Journal of Sociology*, 78 (6), 1360-1380.

381.Kaufmann, M. (2013). *Network Structure and Measures,* Chapter 3-Network Structure and Measures, 23-44.

382.Balkundi, P. ve Kilduff, M. (2006). The ties that lead: A social network approach to Leadership. *The Leadership Quarterly*, 17-4, 419-439.

383.İnternet: Newman, M. E. (2008). The mathematics of networks. The new Palgrave Encyclopedia of Economics, (2008)(2), 1-12. URL: http://www-personal.umich.edu/~mejn/papers/palgrave.pdf. Son Erişim Tarihi: 02.07.2023.

384.İnternet: Disney, A. PageRank centrality and EigenCentrality. URL: https://cambridge-intelligence.com/eigencentrality-pagerank/#:~:text=PageRank%20centrality%3A%20the%20Google%20algorithm,any%20kind%20of%20network%2C%20though, Son Erişim Tarihi: 01.05.2023.

385.Internet: PageRank algoritması nedir? URL: https://www.seohocasi.com/pagerank-algoritmasi-nedir/, Son Erişim Tarihi: 29.03.2023.

386.Internet: Günaçar, O. Pagerank nedir, SEO için neden önemlidir? URL: https://www.dijitalzade.com/pagerank/, Son Erişim Tarihi: 29.03.2023.

387.İnternet: PageRank. URL: https://en.wikipedia.org/wiki/PageRank#:~:text=PageRank%20works%20by%20counting%20the,more%20links%20from%20other%20websites, Son Erişim Tarihi: 01.05.2023.

388.Internet: Girvan-Newman algorithm. URL: https://en.wikipedia.org/wiki/Girvan–Newman_algorithm, Son Erişim Tarihi: 04.04 2023.

389.Internet: Label propagation. URL: https://neo4j.com/docs/graph-data-science/current/algorithms/label-propagation/, Son Erişim Tarihi: 04.04.2023.

390.Internet: Jaccard Index. URL: https://en.wikipedia.org/wiki/Jaccard_index, Son Erişim Tarihi: 04.04.2023.

391.Internet: Sébastien, H. Fruchterman Reingold. URL: https://github.com/gephi/gephi/wiki/Fruchterman-Reingold, Son Erişim Tarihi: 04.04.2023.

392.İnternet: Saxena, A. ve Iyengar, S. (2020). Centrality measures in complex networks: A Survey. 2020 (abs/2011.07190). Son Erişim Tarihi:30.04.2023

393.Zhang, J. ve Luo, Y. (2017). *Degree centrality, betweenness centrality and closeness centrality in social network.* Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017), 2017 (03), 300-303. DOI: 10.2991/msam-17.2017.68.

394.İnternet: Raghavan, U. N., Albert, R. ve Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. Son Erişim Tarihi: 30.04.2023.

EKLER

EK-1. Durak kelimeleri listesi

Aşağıdaki durak kelimeler listesi, yaygın olarak kullanılan durak kelimelerine, bu çalışmadaki koleksiyon içinde tespit edilen durak kelimeleri eklenmesi ile genişletilerek hazırlanmış bir listedir. Listeye kelimelerin hatalı, yanlış yazım şekilleri, konuşma dilinde kullanılan formları dahil edilmiştir. Listede ayrıca aynı kelimenin farklı çekim ekleri almış şekilleri de listeye ayrı ayrı yazılmıştır. Bu listede yer alan durak kelimeleri, her çalışmaya özel uyarlanabilir. Listede iki ya da üç harfli kelimeleri de yer almaktadır. Bu kelimelerin metinler içinden doğrudan temizlenmesi, başka kelime içinde hece olarak yer alan kısımların silinmesine yol açabileceği için iki ve üç harfli durak kelimelerinin temizleme için kullanılmaması uygun olacaktır.

Durak Kelime Listesi
acaba, akabinde, akebinde, altı, altına, altında, altta, ama, ancak, aralarında, arasında, arasından, arkada, arkasında, artık, asla, aslında, aşağı, aşağıdan, aşağısı, aşağıya, aynen, aynı, ayrıca, az, azıcık, bana, başka, bazen, bazı, bazıları, bazısı, belki, ben, beni, benim, benzer, benzeri, beş, bi, bide, bile, binaenaleyh, bir, bir defa, bir hayli, bir kere, bir kere daha, bir kerecik, bir kimse, bir miktar, bir şey, bir şeyi, bir takım, bir vakitler, bir zamanlar, biraz, biraz önce, birbirinden, birçoğu, birçok, birçokları, birde, biri, birisi, birkaç, birkaçı, birşey, birşeyi, bişey, bişi, biz, bizatihi, bize, bizi, bizim, bizimki, bizzat, bizzat ben, bizzat kendileri, bizzat kendimiz, bizzat kendisi, boyunca, böyle, böylece, böylelikle, böylesine, bu, bu gibi, bu kadar, bu noktada, bu suretle, bu şekilde, bu türlü, bugünlerde, buna, buna benzer, bundan, bundan başka, bunlar, bunu, bunun, bunun gibi, bununla birlikte, burada, buraya, bütün, civarında, çevresinde, çoğu, çoğuna, çoğunu, çok, çok az, çünkü, da, daha, daha çok, daha evvel, daha fazla, daha önce, daha ziyade, dahi, de, dedi, dedik, dediler, dedim, dedin, dediniz, değin, demek, demek ki, demi, dışarda, dışarı, dışarıda, dışarıya, diğer, diğeri, diğerleri, dimi, diye, diyor, dokuz, dolayı, dört, ediyor, eğer, ek olarak, elbette, en, epeyce, eski, eskiden, esnasında, etmek, etrafında, evvelce, evvelki, fakat, falan, felan, filan, gene, gibi, hala, halbuki, halinde, hangi, hangisi, hangisini, hani, hatta, hayır, hem, hemen, hemen sonra, henüz, hep, hepsi, hepsine, hepsini, her, her biri, her ikisi, her ikisini, her ne, her ne kadar, her tarafa, herbiri, herhangi bir, herhangibir, herkes, herkese, herkesi, hiç, hiç birine, hiç birini, hiç kimse, hiçbirine, hiçbirini, içerisinde, içerisine, içi, için, içinde, içinden, içine, iken, iki, ikisi, ikisini, ilaveten, ile, ileri, ileride, ileriye, ilk, ise, ise de, işbu, işte, itibarıyla, itibarıyle, itibariyla, itibariyle, iyi, kaç, kadar, karşı, kendi, kendi kendine, kendi kendini, kendi kendinize, kendi kendisine, kendi kendisini, kendilerinde, kendilerine, kendilerini, kendiliğinden, kendim, kendin, kendine, kendini, kendinin, kendiniz, kendinizde, kendinize, kendisi, kendisinin, keza, ki, kim, kime, kimi, kimin, kimisi, kimse, lakin, madem, mı, mıdır, mısın, mısınız, mıydı, mıyım, mi, midir, misin, misiniz, miydi, miyim, mu, mudur, musun, musunuz, muydu, muyum, mü, müddetince, müdür, müsün, müsünüz, müydü, müyüm, nasıl, ne, ne kadar, ne sebeple, ne vakit, ne zaman, neden, nedeniyle, nedir, nerede, nereden, neredeyse, nereye, nesi, netice olarak, neyse, niçin, niye, o anda, o halde, o hususta, o kadar, o noktada, o türlü, o vakit, o yer, o yere, o zaman, o zamanın, o zamanki, oldukça, olmak, olmakla beraber, olur olmaz, on, ona, ondan, ondan sonra, onlar, onlara, onlardan, onları, onların, onlarınki, onu, onun, onunki, ora, orada, oradaki, orası, orasında, oraya, oysa, oysaki, öbür, öbürü, ön, önce, önceden, önceki, önünde, ötede, öteki, öteye, ötürü, öyle, öyle ise, öylesine, özellikle, pek çok, rağmen, sadece, sana, sanki, sebebiyle, sebep, sekiz, sen, senden, seni, senin, seninki, sırf, siz, sizden, size, sizi, sizin, son, son derece, sonra, sonuç olarak, sözcük, süresince, şahıs, şahsı, şayet, şey, şeyden, şeye, şeyi, şeyler, şimdi, şöyle, şu, şu anda, şu halde, şu kadar, şu sırada, şuna, şunda, şundan, şunlar, şunu, şunun, şurada, şuraya, ta kendisi, ta ki, takdirde, takriben, tam, tamamen, tamamı, tamı tamına, tastamam, tekrar, tıpkı, tıpkısı, tüm, tümü, üç, üstelik, üstü, üstünde, üstüne, üzere, üzerinde, vaktiyle, var, vasıtasıyla, vb, ve, veya, veyahut, vs, ya, ya da, yada, yahu, yahut, yakınında, yaklaşık, yalnız, yanında, yani, yalnız, yapar, yapıyor, yapmak, yeniden, yerine, yıl, yine, yoksa, yukarı, yukarısı, yukarısında, yukarıya, yüzünden, zarfında, zaten, zira, ziyade

EK-2. Tohum kelime listesi

Duygu ve Sınıflandırma Sözlükleri İçin Tohum Kelime Listesi
abartma, abla, acı, acılı, adalet, adres, alakasız, açık paket, aç kaldık, açım, adet, adres, ağbi, ağız tadı, aksaklık, akşam, alakasız, alış veriş, alışveriş, amatör, anne, aracılık, artık yeter, asla, aşırı, aşırı yağlı, aynı hata, ayran, az, az yağlı.
baba, bafra pidesi, bahane, bakmak, bayat, bayat ürün, bekle, berbat, besleyici, bıçak, bıkmak, bilet, bimutlulukgetir, bir daha asla, boş, boş köfte, bozuk, bozuk salata, bölge, bulantı, burger, buz, buz gibi, büfe, büyük boy, büzüşmüş.
cafe, canım çekti, canlı yardım, cevap.
çatal, çeşit azlığı, çiğ köfte, çocuk, çok, çok az, çok iyi, çok kızarmış, çok kötü, çöp, çözüm.
dağılmış, dakika, değer, deli olmak, deneyim, dışarıdan sipariş, dikkatsiz, dip, diş kovuğu, diyet, doğru bir nokta, domates, double, doymadım, doymayan, döner, duble, dükkan, dürüm, düzeltme.
eğitim, eğitimsiz personel, ekmek, ekmek arası, eksik, eksik geldi, eksik ürün, en beğendiğim, en güzel, en güzel yanı, en kötü, en kötü yemek, et et döner, etraf, extra, ekstra, ev, evde yemek, ev yemeği.
fark, fast food, fast food zinciri, fazla fiyat, fındık lahmacun, fıstık lahmacun, firma, fiyasko, fiyat, fiyat politikası.
gaflet, gıda, gram, gece, gece yarısı, gecikme, geç gelen, gönderi, gönderme, güzel.
haber ver, hak, hak etmiyor, hamur, hamur gibi, hamburger, hata, hayal kırıklığı, hazır yemek, hediye, helal etmiyorum, helal olsun, hesap, hızlı dönüş, hızlı yemek, hijyen, hizmet, hizmet kalitesi, hizmet sıfır, hizmet verme, homeburger.
ıslak, ıslak hamburger.
iade, iade talebi, içecek, içi boş, iftar, iletişim, iletişim sorunu, ilgisiz, indirim, indirim kodu, insaf, insan sağlığı, internet sitesi, internetten yemek siparişi, ishal, itham, itibar, iptal, işte yemek, İtalyanpizza, İtalyan lezzetleri, iyi fikir.
joker, joker indirim, kaba, kampanya, kardeş, karadeniz pidesi, karışık pide, kart, kasa, kaşar, kaşarlı, kaşık, kavurmalı, kavurmalı kaşarlı, kayış gibi, kazanma, ketçap, keyif, kıkırdak, kıl, kıral gibi, kıralsın, kıymalı, kıymalı pide, kızarmış, kızartma, king burger, koku, kola, köfte, köpek gibi açım, kötü, kötü puan, kredi kartı, kurumsal, kurye, kusma, kuşbaşılı, kuşbaşılı kaşarlı, kuş tüyü, küçük, küçük boy,.
lahmacun, latte, lanet, lanet olsun, lanet ediyorum, leziz, lezzetli, lezzetsiz, limit
mağdur, mağduriyet, mahal, mail adresi, malzeme, malzeme eksikliği, manipülasyon, mayhoş, mayonez, memnuniyet, memnuniyetsizlik, menü, mönü, merkez, mesaj, mide, minimum, minimum tutar, mis, mis gibi, mobil uygulama, muamma, muhatap, multinet, mutfak, mükemmel, mükemmel ürün, müşteri, müşteri hizmetleri, müşteri memnuniyeti, müşteri memnuniyetsizliği.
nakit, ne yesem, niyet, numara.
objektif, olumlu, olumlu yorum, olumsuz, olumsuz yorum, onay, online, online ödeme, online sipariş, orta boy, otomatik ulaşmak, oyala.
ödeme, ödeme yöntemi, öğle, öğrenci, öğün, öncelik, öneri, özen, özensiz, özür.
paket, paketleme, paket servisi, paket sipariş, para, para iadesi, patates, personel, personel ilgisizliği, peynirli, peynirli pide, pide, pilav üstü döner, pişmanlık, pipet, pişmiş, pizza, pizzasipariş, pizza siparişi, poşet, problem, promosyon, puan.
resmi tatil, restoran, restoran zinciri, restorant, restorant zinciri, restaurant, rezalet.
saat, saç, saçmalık, sağlıklı, sağlıksız, salata, Samsun pidesi, Samsun pidecisi, sanki, sayfa, saygısızlık, servis, servis elemanı, servis sıfır, set card, setcard, setkart, severek, seviliyorsun, seviliosun, seviyorum, sıkıntı, sinek, sipariş, sipariş hattı, sipariş iptali, sipariş notu, sipariş onayı, sipariş öncesi, sipariş sonrası, sistem, sodexo, soğuk, soğumuş, son sipariş, sonuç, sorumlu, sorumsuzluk, sorun, sos, sübjektif, suç, suçlama, sufle, süre, şımartmak, şikayet, şube, şüpheli.
taahhüt, takip, talep, talep etmek, tam zamanında, tat, tavuk, tavuk burger, tavuk döner, tavuk hamburger, tecrübe, tehlike, telafi, telefon, temiz, terleme, teslim, teslim etmek, teslimat, teslimat süreci, teslimat süresi, teşekkür, ticket, tövbe, trend, tutar, tuz, Türk mutfağı, tüy.
uğraşıyorum, ulaşamamak, umursamaz, umursuz, unutmuş, urfa, uygulama, uygun lokasyon, uzak.
ücret, ücret iadesi, ürün, üstü boş, üşenmek, üye.
vıcık, vıcık vıcık, vurdum duymaz.
Web sitesi
yağlı, yalan, yanlış, yalnış, yanlış sipariş, yanlış sipariş, yanmış, yapışmış, yaptırım, yardım, yarım, yaşasın, yazık, yemek, yemek arası, yemeksepeti, yemek sepeti, yemek yapma, yemek yok, yeter artık, yetersiz, yetki, yetkili, yettim, yoğunluk, yol, yorum, yumurtalı, yüzde.
zaman, zamanında, zehirlendim, zehirlenme, zevk veren.

EK-3. Alana özel (ÇYS) sözlüklerin hazırlığında kullanılan yöntemler

1. Müşteri şikâyet ve memnuniyetine ilişkin görüşlerin paylaşıldığı forumlardan EK 1'deki tohum kelime listesi oluşturulmuştur.

2. Pozitif ve negatif sözlükler oluşturmak için tohum kelime listelerine Afinn pozitif ve negatif sözlükten yapılan çeviri ile kelime eklemesi yapılmıştır. Türkçe'ye çeviri için Zargan sözlük kullanılmıştır. Çeviri yapılırken, İngilizce kelimenin karşılığında yazılı tüm Türkçe kelimelerin anlamı kontrol edildikten sonra, kelime listesine dahil edilmiştir. Zargan sözlükte karşılığı bulunmayan kelimeler için Tuareg sözlüğü kullanılmıştır. Çeviri işlemi esnasında, eğer mevcutsa Türkçe karşılık olarak kelime grupları, argo ifadeler, 2-3 kelimelik deyimler, aynı köke sahip kelimenin yapım eki ile türetilmiş şekilleri de hazırlanan kelime listesine dahil edilmiştir.

3. İnternet üzerinde kelimelerin eş ve zıt anlamları bulunarak tohum kelime listesi genişletilmiştir [327, 328]. Sözlük ile genişletme işlemi için ayrıca Office 365'in Excel uygulamasında yer alan eş anlamlar sözlüğü de kullanılmıştır. Eş anlamlar, eş anlamı olduğu kelime ile aynı sözlüğe, pozitif kelimelerin zıt anlamları negatif sözlüğe, negatif kelimelerin zıt anlamları pozitif sözlüğe dahil edilmiştir. Ürün ve hizmet sistemleri sözlüğü için yalnızca eş anlamlı kelimelerle genişletme işlemi yapılmıştır.

Eş anlamlı kelimelerle genişletme şu şekilde yapılmıştır. Bir kelimenin eşanlamları birinci kaynak sözlükten bulunmuş, bulunan yeni kelimelerin eş anlamları da aynı kaynak sözlükten bulunmuş ve hazırlanan sözlüğe dahil edilmiştir [327]. Bu işlem iteratif şekilde bulunan tüm kelimeler için aynı kelimeler tekrar edene ya da bulunan kelimenin başlangıçtaki kelimenin anlamı dışına çıktığı tespit edilene kadar devam ettirilmiştir. Birinci kaynaktan yapılan eş ve zıt anlamların yazılması işleminden sonra ortaya çıkan listelerdeki tüm kelimelerin eş ve zıt anlamlarına ikinci kaynak [328] ve üçüncü kaynaktan (Office 365'in Excel eş anlamlar sözlüğü) bakılarak oluşturulan listelerde mevcut olmayan kelimeler, listelere dahil edilmiştir, bu aşamada ikinci ve üçüncü kaynak sözlüklerde iteratif yöntem uygulanmamıştır. Sonuçta oluşan listedeki kelimeler tekrar kontrol edilmiş, listelerde bulunduğu sınıf ile ilgili olmadığı değerlendirilen kelimeler listelerden çıkarılmıştır.

EK-3. (devam) Alana özel (ÇYS) sözlüklerin hazırlığında kullanılan yöntemler

4. Sözlüklere bazı kelimelerin (sık yapılan yazım hataları dikkate alınarak) hatalı yazım şekilleri de dahil edilmiştir [329].

5. Zıt anlam ile sözlük genişletme yapılırken, zıt anlam karşılığı bulunmayan kelimeler için olumlu / olumsuzluk ekleri (-li, -lı, -lu, -lü, -sız, -siz, -suz, -süz ve -me, -ma) kullanılarak kelimeler olumsuzluğa dönüştürülmüş ve dönüştürülen bu kelimeler de sözlüğe dahil edilmişlerdir.

6. Yukarıdaki işlemler sonunda alana özel başlangıç sözlüklerindeki kelime sayıları şu şekilde oluşmuştur:

(i) Pozitif kelime sayısı: 513 (Başlangıçta 5 879 kelimelik bir pozitif kelime listesi hazırlanmış alana özel hale getirildikten sonra bu sayıya ulaşılmıştır.),

(ii) Negatif kelime sayısı: 636 (Başlangıçta 11 727 kelimelik bir negatif kelime listesi hazırlanmış alana özel hale getirildikten sonra bu sayıya ulaşılmıştır.),

(iii) Ürün ile ilgili kelime sayısı: 278,

(iv) Hizmet ile ilgili kelime sayısı: 723.

EK-4. Kelimelerin temsili için gruplandırılan kelimeler ve güçlendirilmiş sözlükler

Duygu Sözlüğü (Pozitif-Negatif) İçin Gruplandırılan ve Grubu Temsil Eden Kelimeler
{"acele":["acele", "acil"], "ağzımın su":["ağzımın su", "azımın su"], "aksiyon":["aksiyon","ekşın"], "aralıksız":["aralıksız", "durmadan"], "hediye":["armağan", "ödül", "hediye"], " aşık":["asko", "aşko", "aşkito", " aşık"], "beğeni":["beğeni","beğend"], "canı çek":["canım çek", "canı çek"], "çık hayat":["çık aklı", "çık hayat"], "eline sağlık":["eline sağlık", "ellerinize sağlık"]," esas ":[" elit"," esas "], "gönül":["gönül", "gönlü"], "hakikaten":["hakikaten","hakkaten"], "hapır hupur": ["hapır hupur", "hapur hupur"], "hayırlısı":["hayır ol","hayırlısı"], "hastası":["hastası", "hasta ol"], "hemen":["hemen", "hızla", "hızlı", "ivedi", "çabuk"], "inşallah":["inşallah", "inş "], " istiyor": [" istiyor"," ister"," istey"," istemiş"," isted"], "iyilik":["iyice", "iyilik"], "kalp": ["kalp", "kalb"], "kardeşim":["kardeşim", "kardeşlerim"], "keyif":["keyf","keyif"], "lezzet":["leziz", "lezzetli", "nefis", "lezzet"], "müsaade":["müsade","müsaade"], "otomatik":["otomasyon", "otomatik"], "özenli":["özendir", "özenli"], "saygılı":["saygılı", "saygın"], "şükür":["şükür", "şükran"], "taktir":["taktir", "takdir"],"uygun":["uygun"," uyum"],"usta":["usta"," piri "], "teşvik":["teşfik","teşvik"], "yakışıklı":["yakışıklı", "yakışır"], "abuk sabuk":["abidik gubidik", "abuk sabuk", "antin kuntin"], "acayip":["acayip", "absürt"],"acemi":["acemi","toy"], "açlık": ["aç kald", " açım"," açız", "açlıktan bayıl"], "ağlıcam":["ağlıcam","ağlicam"], "sövmek":[" amk","amq","aq"," mk ", "skm","skt"," söv"], "aşırı yağlı":["aşırı yağ", "çok yağlı"], "bağlanmak": ["bağlanam","bağlanm"], "bomboş":["bomboş"," boş "], "yasak":["banla", "yasak"], "bela": ["belanı versin", "belanızı versin"], "bıkmak":["bıkkın", "bıktım"],"boşuna":["boş yere", "boşuna"], "bozuk":["bozuk", "bozulmuş"], "donmuş":[" buz ","donmuş"],"pişmemiş":[" çiğ ", "pişmemiş"], "çökmüş":["çökmüş", "çöktü"], "dağınık":["dağınık", "dağıl"], "dağ başı":["dağ baş", "dağbaş","dağın baş"], "deli ol":["deli ol", "delir"], "duygu sömür":["duyar kas", "duygu sömür"], "gecik":["gecik", "geç gelen"], "gergin":["gergin", "geril"], "getirmek": ["getirem", "getirm"], " uzak": [" ırak "," uzak"], " istemiyor":[" isteme"," istemiyor"],"kafayı ye": ["kafayı ye", "kafayı yi"], "kahrolsun":["kahretsin","kahrolsun"], "kaldık":["kaldık","kaldım"], "kirli": ["kirlenmiş","kirli"], "lanet":["lanet","nalet"],"negatif":["negatif"," eksi "], "orospu":["or*spu çocuğu","or*spuçocuğu","orospu"], "rezil":["rezalet", "rezil"], "soğuk":["soğuk", "soğumuş"], "trip":["trib","trip"], "ulaşam":["ulaşam","ulaşmam", "ulaşm"], "umrunda değil":["umrumda değil", "umrunda değil"], "üzgün":["üzgün","üzül", "üzüyo", "üzücü"], "yanık":["yanık", "yanmış"], "yeter artık":["yeter artık", "yeter yahu"], "yetersiz":["yetersiz", "yok"], "zarar ":["zarar ","zararlı"], "zoraki":["zor bela", "zoraki", "zorla", "zorunda bırak", "zorunda kal"]}
Köşeli parantez içindeki kelimeler kelime gruplarını, köşeli parantezden önceki kırmızı renkle yazılmış kelimeler o grubu temsil eden kelimeleri göstermektedir. Güçlendirilmiş duygu sözlüğüne kelime gruplarını temsil eden kırmızı renkle yazılmış kelimeler dahil edilmiştir. Ayrıca koleksiyonda kelime gruplarında yer alan kelimeler, grubu temsil eden kırmızı renkle yazılmış olan kelimelerle değiştirilmiştir. Koleksiyon üzerinden frekans değerlerine bağlı tüm hesaplamalar bu değişim işleminden sonra yapılmıştır.

EK-4. (devam) Kelimelerin temsili için gruplandırılan kelimeler ve güçlendirilmiş sözlükler

Güçlendirilmiş Duygu Sözlüğü
["10 puan","number one", "10/10", "6/10", "8/10", "acar", "acele", "adalet", "aferin", "afiyet", "ağzımın su", "ak pak", "aksiyon", "alakalı", " alfa", "anında", "anlamlı", "aralıksız", "arkadaş", " artı", "arzu"," asıl "," aşık", "aşer"," aşkın", "avantajlı", "bağımlı", "bahşiş", "başarılı", "başlıca", "bayıl", "bayram", " baz ", "bebeğim", "beğeni", "beklenti", "beyefendi", "bilerek", "bilgili", "bilinçli", "boğazımdan geçm"," bol ","bravo", "buruk", "canı çek", "cansın", "centilmen", "ciddi", "çare", "çekiliş", "çeşni", "çık hayat", "çılgın", "çok iyi", "çözüm", "daima", "dayanışma", "değerli", "demlenmiş", "dengeli", "derle", "derman", "destek", "devamlı", "dikkatli", "dilek", "doğal", "doğru","dolu","doyam","dua", "duyarlı", "dürüst", "düşünceli", "düzelt", "düzenli", "efendi", "eğitim", "eksiksiz", "ekstra", "ev yemeği", "eline sağlık"," esas ", " esen "," eşit"," etkili", "faydalı", "fazilet", "gayet", "gerçek ","gerekli"," göm", "gönül", "görev", "görgülü", "gurur", "güven", "güzel", "hakikaten ", "hakkıyla", "halis", "hapır hupur", "hasret kal", "hassas", "hastası", "havalı", "hayati", "hayırlısı", "hayran", "hediye", "helal", "hemen", "heves", "hijyen"," hoş ","hukuk", "huzur", "içim", "ikram", "ilave"," ilgili", "iltifat", "incelik", "indirim", "insan sev", "insani", "inşallah"," istiyor", "işbirli", "iştahlı", "itibar", "iyi fikir", "iyi puan", "iyi yemek", "iyilik"," izin", "jest", "joker", "kabul", "kalp", "kalıcı", "kaliteli", "kanka", "kardeşim", "karlı", "kazan", "kefil", "kesintisiz", "keyif", "kızarmış", "kibar", "kolay", "kral", "kurban", "latif", "layık", "lokum", "makbul", "mantıklı", "medeni", "memnuniyet", "merhamet", "mesut", "meşhur", "minimum tutar", "minnet", " mis ", "motivasyon", "muhteşem", "mutlu", "mükemmel", "müsaade", "müsait ","nazik", "nitelikli", "nizam", "olağanüstü", "olumlu", "onay", "onur", "optimum", "otomatik", "öncelik", "önem", "öner", "özel", "özenli", "patla", "pişkin", "pişmiş"," pls ", "pozitif", "prestij", "profesyonel", "promosyon", "rahat", "rica", "rüzgar", "safa", "sağlam", "sağlıklı", "sakin", "salim", "saygılı", "sefa"," seri", "sev", "sıcacık", "sipariş onay","sistem","sorumlu","stalk","sürekli", "sürpriz", "şahane", "şerefine", "şevk", "şımart", "şükür", "taahhüt", "tadı güzel", "takip", "taktir", "talep", "tarafsız", "tatlı", "tavla", "taze", "tecrübe", "teklif", "telafi", "temel", "temiz", "terbiyeli", "teşekkür", "teşvik"," tez ","titiz", "tolerans", "toparla", "toplu", "tutul", "ucuz", "umut", "usta", "uyanık"," uygun", "ücretsiz", "üstlen", "üstün", "vaktinde", "verimli", "vurul", "yakın", "yakışıklı", "yarar", "yardım", "yaşasın", "yemek video", "yeni ","yerinde", "yeterli", "yoğunlaş", "yöntem", "yüce", "zahmet olmazsa", "zevkli", "abart", "abes", "abuk sabuk", "acayip", "acemi"," acı ","acitasyon", " açık ", "açlık", " adi ", " ağır","ağlıcam", "ahlaksız", " aksak"," aksi ","alakasız", "alt tarafı", "sövmek", "andaval", "anksiyete", "anlamsız", "arıza", "asılsız", "aşağı", "aşınmış", "aşırı yağlı", "ayar ol"," ayıp"," ayı "," azal", "azar ye", " azarla", "bağlanmak", "bahane", "balık hafıza", "balon", "yasak", "basit", "başarısız"," bat", "bayağı", "bayat", "baygın", "beceriksiz", "beddua", "beğenm", "beklet", "bela", "bencil", "bitmiştir", "berbat", "bereket", "beter", "beyhude", "beyinsiz", "bez", "bıkmak", "bilgisiz", "bilinçsiz"," bilme", "bin pişman", "daha asla", "bitkin", "blokl", " bok ", "bomboş", "boşuna", "botla", "boykot", "bozuk", "böcek"," bulan", "donmuş", "büyütme", "cahil", "camış","cansız", "cenabet", "cennet", "cereme", "cesaretim yok", "ceza"," cılk", "crash","çakal","çamur","çekin","çelişki","çıkar","çıkış","çıldır","pişmemiş","çirkin", "çok kötü", "çökmüş"," çöp ","çürü", "dağ başı", "dağınık", "dalgın", "dandik"," dar ", "dayanma sınır", "dedikodu", "değersiz", "deli ol", "dengesiz", "dert"," dışı", "diken üst", "doym", "dökülmüş", "dönek", "dram", "duygu sömür", "duyarsız", "düğüm", "düşman", "düşük", "düşüncesiz", "düzenbaz", "düzensiz", "edepsiz"," eksik ", "eleştiri", "enayi"," engel", "erimiş"," ertele"," eşek ","eyvah"," ezik","fake","fani", "fasa fiso", "faydasız", "fazla", "felaket", "feleğim şaş","fena","fırsatç","fiyasko", "gaflet", "gazabına uğra", "geber", "gecik", "geçici", "gelm", "gereksiz", "gergin", "getirmek", "gıcık", "gına gel", "görgüsüz", "gözü karart", "gudubet", "hadsiz", "hak etmiyor", "hakaret", "haksız", "halt", "hamur", "haram", "hatalı", "hayal kırık", "haysiyetsiz", "hayvan", "hazetm","hikaye"," uzak", "ıslanm","ıssız","iade","ibne","iflah olm", "iflas", "iğren", "ihmal", "ihtar", " ilgisiz"," illet", "insaf", "iptal", "israf"," istemiyor", "istismar", "isyan", "işkence", "işsiz", "iştahsız", " itici", "iyi değil", "kaba ","kafayı ye", "kahrolsun", "kalas", "kaldık", " kalm","kan emici", "kanser", "kapısının önü", "kara kara düşün","kara liste", "karışık", "kaybet", "kayış", "kazık", "kendimi tut", "keşke"," kını", "kırıl", "kıtlık", "kifayetsiz", "kilo ", "kirli", "kitle", "kokan", "kokm", "kopya", "korkunç", "köle", "kömür", "kötü ","köylü"," kuru ", "kusm", "kusur", "küçük", "küflü", "küstah"," laf ", "lakayıt","lanet", "lastik", "leke"," leş ", "lezzetsiz", "lüzumsuz", "mağdur", "mahsur", "mantıksız", "manyak", "maraz doğ", "mesele", "midem bulan", "mikrop","minik","muhatap","mutsuz","nefret","negatif","olacaksa ol", "odun"," oha ", "olumsuz", "orospu", "ortadan kaldır", "oyala", "öküz"," ölü ","özensiz", "özür", "pahalı", "panik", "paranoyak", "pezevenk"," pis ", "pişman", "problem", "psikopat", "rasgele", "reddet", "rezil", "risk", "rötar", "ruh hasta", "saçma"," saç ", "sağlıksız", "sahte", "sakat", "sakınca", "salak", "salla", "saman", "sası", "savsak", "saygısız", "sebepsiz", "serseri", "sert", "ses seda yok", "sıfır", "sıkıl", "sıkıntı", "sıradan", "sızlan", "sinek", "sipariş hata", "sipariş iptal", "sitem", "sivrisinek", "skandal", "soğukta", "soğuk", "sorm", "sorumsuz", "sorun", "sömür", "sözde"," suç", "sürün", "şaka", "şans gülm", "şerefsiz", "şeytan", "şımarık", "şikayet", "şişir", "şişko patates", "şopar", "şüpheli", "taciz", "tadı yok", "takat", "talan et", "talihsiz", "tasa ", "taş ","tatava yap", "tatsız", "tehlike", "tekel", "telaş", "tembel", "terbiyesiz", "ters", "tırs", "toz", "trip", "tuzsuz","tüketme","tükür","tüy","uçurum","ufak","uğraş","ukala"," ulan", "ulaşam", "umrunda değil", "umursam", "unutmuş", "unutul"," usan"," utan", "uyar", "uydur", "ürkütücü"," üşen", "üzgün", "vahim", "vasat", "vazgeç"," verem", "verimsiz", "vıcık","vicdan azab","vicdansız", "vizyonsuz", "yağı don", "yağsız", "yalaka", "yalan", "yanlış", "yamuk", "yanık", "yanıl", "yapışmış", "yaram", "yaratık", "yaşlan", "yavaş", "yavşak", "yazık", "yemek yok","yemiyo","yerlerde", "yeter artık", "yetersiz", "yorgun", "zarar ","zehir", "zevksiz", "zıkkım", "ziyan", "zoraki"]

EK-4. (devam) Kelimelerin temsili için gruplandırılan kelimeler ve güçlendirilmiş sözlükler

Ürün-Hizmet Sistemleri Sözlüğü (Ürün-Hizmet) İçin Gruplandırılan ve Grubu Temsil Eden Kelimeler
{"adet":["adet", "tane"],"çok yağlı":["aşırı yağ", "çok yağlı"],"bozuk":["bozuk", "bozulmuş"], "donmuş":[" buz ","donmuş"], "canı çek":["canım çek", "canı çek"],"pişmemiş":[" çiğ ", "pişmemiş"], "damak tadı":["damak tadı", "damak zevk"], "duble":["double","duble"], "doyurucu":["doyum", "doyurucu"], "ekşimiş":["ekşim"," ekşi"], "ev yapımı":["el yapım", "ev yapım ", "ev yemeği"],"fast food":["fast food","fastfood"],"lezzetli":["leziz", "lezzetli", "nefis"], "soğuk":["soğuk", "soğumuş"], "vejetaryan":["vejetaryan", "vejeteryan"], "lira":["₺","lira", "nakit", " para ", " tl"], "acemi":["acemi", "toy"], "konum":[" alan ","adres", "bölge", "civar"," etraf", "konum", "mahal", "mevki", "muhit", "semt", "sokak"], "alışveriş":["alış veriş", "alışveriş"], "alt sınır":["alt limit", "alt sınır"], "anasayfa":["anasayfa","ana sayfa"], "aplikasyon": ["aplikasyon", " app","mobil uygulama"], "boşuna":["boş yere", "boşuna"], "callcenter": ["callcenter", "call center","canlı destek", "canlı yardım"], "çökmüş":["çökmüş", "çöktü"], "dağın baş":["dağ baş", "dağbaş","dağın baş"], "dağınık":["dağınık", "dağıtıcı", "dağıtım ağı"], "dakika": ["dakika","dk","saat"], "davranış":["davranış","davranm"], "debit":["debit","setcard","sodexo", "multinet"], "dışardan sipariş":["dışardan sipariş", "dışarıdan sipariş"], "entegrasyon": ["entegrasyon","entegre"], "fiyat":["fiat","fiyat"], "gecik":["gecik", "geç gelen"], "gel al":["gel al", "gel-al"], "gelen abi":["gelen abi", "gelen arkadaş"], "gönderi": ["gönderi","gönderm"], "hemen": ["hemen", "hızlı"], "hes cod": ["hes cod", "hes kod"], " uzak":[" ırak ", " uzak"], "iletm":[" ileti", "iletm"], "kapalı":["kapalı", "kapanmış", "kapanış", "kapatmış", "kapatmak"], "kara liste":["kara liste","karaliste"], "kokorecci":["kokoreççi", "kokorecci"], "kurye":["kuriye", "kurye"], "posta":["mail", "posta"], "min tutar":["min paket tutar", "min sipariş tutar", "minimum sipariş tutar", "minimum tutar"], "motokurye":["motokurye", "motosiklet", "motor"],"nerde kal":["nerde kal", "nerden gel"], "otomasyon":["otomasyon", "otomatik"], "özenli":["özenli", "özen göster"], "paket servis":["paket servis", "paket sipariş"], "rasgele": ["rasgele", "rastgele"], "saygılı":["saygılı", "saygın"], "sipariş iptal":["sipariş hata", "sipariş iptal"], "soğuk hava": ["soğukta", "soğuk hava"], "telefon":["telefon","mesaj","sms"], "yemekçi": ["sepetçi", "yemekçi"],"trip":["trib","trip"], "ulaşmam":["ulaşam", "ulaşmam", "ulaşm", "ulaşım"]}
Köşeli parantez içindeki kelimeler kelime gruplarını, köşeli parantezden önceki kırmızı renkle yazılmış kelimeler o grubu temsil eden kelimeleri göstermektedir. Güçlendirilmiş ürün-hizmet sistemleri sözlüğüne, kelime gruplarını temsil eden kırmızı renkle yazılmış kelimeler dahil edilmiştir. Ayrıca koleksiyonda kelime gruplarında yer alan kelimeler, grubu temsil eden kelimelerle değiştirilmiştir. Koleksiyon üzerinden frekans değerlerine bağlı tüm hesaplamalar bu değişim işleminden sonra yapılmıştır.

EK-4. (devam) Kelimelerin temsili için gruplandırılan kelimeler ve güçlendirilmiş sözlükler

Güçlendirilmiş Ürün-Hizmet Sistemleri Sözlüğü
["abur cubur", " acı ", "adet", "ağız tadı", "altın günü yiyecek", "ana yemek", "aperatif", "çok yağlı", "ayran", "azıcık", "baharat", " bal ", "bayat", "besin", "biber", "bisküvi", " bol ", "bozuk", "böcek", "cips", "börek", "burger", "buruk", "donmuş", "büyük boy", "canı çek", "çeşni", "çevirm", "çıtır", "pişmemiş", "çoban", " çöp ", "çörek", "dağıl", "damak tadı", "dolma", "domates", "duble", "doym", "doyurucu", "döner", "dünden kalan", "dürüm"," ekle ","ekşimiş", " ekmek ", " eksik ", "ev yapımı", " erzak", "gıda", "eşantiyon", " etli", " etsiz", "fast food", "futuristik", "garnitür", "gevrek", "gluten", "gram", "gurme", "hafif yemek", "hamburger", "hamur", "haram", "hastası", "hatay usul", "havyar", "hazır yemek", "helal"," hoş "," ısıt", "ıspanak", "iade", "içecek"," içki ", "içli köfte", "ikram", "kadınbudu", "kalın hamur", "kayış", "karbonhidrat", "karışık", "kaşarlı", "katık", "kavurma", "köpek yem", "kestane", "ketçap", " kıl ", "kıtır"," kıyma", "kızarmış", "kızartm","kilo ", "kişi başı", "klasik", "kokan", "kokm", "koku", "konserve", "köfte", "kömür", "kurt", " kuru ", "kusm", "kuşbaşı", "küçük", "küflü", "lahmacun","lastik", "latif","latte", "lüfer", "lezzetli", "lezzetsiz", "lokma", "lokum", "makarna", "martı eti", "marul", " maya", "mayonez", "meyve", "meşrubat", "meze ","mide", "minik"," mis ","nane", "organik", "orta boy", "öğün", "ölçü", "patates", "patlıcan", "peynir", "pide", "pilav", "pişmiş", "pizza", "poğaça", "porsiyon", "pörsüm", "reçel", " saç ", "sağlıklı", "sağlıksız", "salata", "salça", "saman", "sarımsak", "sası", "sebze", "sıcacık", "sinek", "soğan", "soğuk", "son kullanma tarih"," sos ", "sucuk", "sufle", "sulu", "şalgam", "şeker", "şerbet", "taraftar", "taş ", "tatlı", "tatsız", "tavuk", "taze", "tereyağ", "turşu", "tuzlu", "tuzsuz", "ufak", "urfa", "ürün", "vegan", "vejetaryan", "vıcık", "yağı don", "yağlı", "yağsız", "yanık", "yanmış", "yapışmış", "yarım", "yemiş", "yeşillik", "yöresel", "yudum", "yumurta", "zehir","zevkli", "lira","3d secure", "abone", "acele", "acemi"," açık "," açıl", "adalet"," adam", "adisyon"," adi ", "ahlaksız", "ak pak", " aksi ", "aktarm", "alakalı", "alakasız", "konum", "alçal", " algı", "alışveriş", "alt sınır", "altyap", "anasayfa", "anında", "aplikasyon", "anlamsız","anlaşılm", "anlayış", "aracı", " asıl ", "asistan", "aşağı"," ayı "," azarla", "bağlantı", "bahşiş", "bakım", "basit", "başlıca", "bayağı", " baz ", "beceriksiz", "bedel", "beklet", "belgeli", "bencil", "beyhude", "beyinsiz", "bıçak", "bildiri", "bilet", "bilgisiz"," anda ", "blokl", "boşuna", "bölüm", "buton", "cahil", "callcenter", "cevap", "crash", "cüzdan", "çaba", "ivedi", "çağrı", "çakal", "çalışan", "çalışm", "çatal", "çelişki", "çevre", "çıkış", "çiçek", "çirkin", "çökmüş", "dağın baş", "dağınık", "dakika", "dalga", "dalgın", "danışma", "davranış", "debit", "değerli", "değersiz", "deney", "denk", "dert", "destek", "devre", "dezenfekte", "dış kapı", "dışardan sipariş", "dikkate alm", "diyet", "doğru", "dönem", "duyarsız", "dükkan", "dürüst", "düşük", "düşüncesiz", "düşünceli", "düzensiz", "edepsiz", " ederi"," eğitim", "eksiği", "eleman", "emekçi", "entegrasyon", " esas ", " esnaf "," eşek ", " eşit", "fatura", "faydasız", "fazilet", "fazla", "fiyat", "filtre", "firma", " fiş", "gamsız", "gayret", "gece yarısı", "gecik", "gel al", "gelen abi", "gelir", "gerçek ", "gereksiz", "getiren", "getirtm", "gönderi", "görev", "görgülü", "görgüsüz", "görmem", "görüş", "götürm", "haberleş", "hack", "hain", "hakaret", "hata ver", "havalı", "hayvan", "hediye", "hemen", "hes cod", "hesap", "hizmet", "hukuk", " uzak", "ısmarlam", " icra", "ihmal", "ihtar", "iletm", " ilgili", "ilgisiz", " ilişki", " ilkel", "indirim", "influencer", "insan sağlığı", "internet site", "internetten yemek siparişi", " işçi", "işlem", "işletme", "işyeri", " izin", "joker", "kaba ", "kaide", "kalas", "kampanya", "kapalı", "kapıda öde", "kara liste", "kargocu", "karşılı", "kasiyer", "kaşık", "katır", "kaytar", "kazan", "kazık", "kdv", "kebapçı", "kısır", "kıymetli", "kifayetsiz", "kokorecci", "komisyon", "konsept", "kota", "kart", "kullanıcı dost", "kupon", "kural", "kurye", "kurumsal", "küçümseme", "küstah", "lakayıt", "legal", "posta", "maliye", "malzeme", "mekan", "memur", "mendil", "menü", "merkez", "mesafe", "mesele", "meslek", "mevsim", "mezun", "miktar", "min tutar", "misli", "motokurye", "muamele", "muhatap", "mukabil", "mutfak", "mücadele", "müdavim", "müddet", "müşteri", "naçiz", "nazik", "negatif", "nerde kal", "nizam", "nöbet", "numara", "odun", "online", "operatör", "ortak", "otomasyon", "ödeme", "ödev", "öküz", "öner", "örgüt", "özenli", "özensiz", "pahalı", "paket servis", "paket", " pay ", "peçete", "perhiz", "personel", "pipet", "platform", "portör", "pos cihazı", "poşet", "pratik", "promosyon", "prosedür", "puan", "range", "rasgele", "rejim", "reklam", "restoran", "rozet", "rötar", "ruhsat", "rut dışı", "sağanak", "salah", "sapa", "satış", "savsak", "saygılı", "saygısız", "yemekçi", " seri", "server", "servis", "sezon", "sıradan", "sırılsıklam", "sipariş hattı", "sipariş iptal", "sipariş not", "sipariş onay", "sipariş önce", "sipariş sonra", "sistem", "sitem", "soğuk hava", "sorumlu", "sorumsuz", "sorun", "stil", "story", "sunucu", " süre", "şimdi", "şirket", "şube", "taciz", "tahsil", "tahsis", "taksit", "tarih", "tarz", "taşı", "tatbik", "tayfa", "tecrübe", "tehlike", "tek kişi", "teklif", "teknoloji", "telefon", "temas", "temel", "temiz", "temsilci", "terbiyesiz", "ters", "teslim", " tez ","ticket", "titiz", "token", "tolerans", "toplam", "trafik", "trip", "tutar", "tutum", "tüketici", "tükür", "türe", "tüy", "uğraş", "ukala", "ulaşmam", "umursam", "unutmuş", "usta"," usul", "uyar", "uygula", " uyum", "ücret ", "üyeli", "üzücü", "vakit"," vale", "vasıta", "verimli", "verimsiz", "viral", "virüs", "web site", "webchat", "yağmur", "yakışıklı", "yalın", "yanlış", "yanıt", "yaptırım", "yarar", "yasal", "yatkın", "yavşak", "yazılım", "yerel", "yetersiz", "yetkili", "yoğunluk", "yollam", " yol", " yorum", "yoz", "yönetici", "yöntem", "yürütm", "zaman", " zam ", "zararlı", "zihniyet"]

EK-5. Reklam içerikli tweetlerin ayıklanması için kullanılan reklam kelimeleri sözlüğü

Koleksiyondan reklam içerikli tweetlerin ayıklanması maksadıyla reklam kelimeleri sözlüğü hazırlanmıştır. Reklam kelimeleri sözlüğünün hazırlanması için; rastgele bir şekilde veri havuzundaki tweetlerin % 10'luk bir bölümü seçilmiş, bunlar içindeki reklam içerikli tweetler elle tek tek incelenmiş ve bu tweetlerden seçilen kelimeler reklam kelimeleri sözlüğü için tohum kelime listesi olarak kullanılmıştır. Diğer sözlüklerde olduğu şekilde tohum kelime listesi genişletilerek alana özel *reklam kelimeleri sözlüğü* hazırlanmıştır. Bu sözlük, koleksiyonun ön temizliği aşamasında kullanılmış, sözlükteki herhangi bir kelimenin tweet cümlesinde mevcut olması halinde o tweetin reklam tweeti olduğu kabul edilerek koleksiyondan çıkarılması şeklinde ayıklama işlemi gerçekleştirilmiştir. Aşağıda liste olarak verilen kelime listesine, kelimelerin hatalı, gündelik kullanım şekilleri ve jargon kullanımları da dikkate alınarak dahil edilmiştir. Sözlük kelimeleri bir sınıflandırma sözlüğü şeklinde kullanılmamış, reklam içerikli metinleri ayıklama konusunda bir performans ölçümü yapılmamış, sadece sözlükte yer alan kelimelerle reklam içerikli tweetlerin ayıklanması ve ön eleme maksadıyla kullanılmıştır. Sözlükte yer alan kelimelerin, koleksiyonda yer alan tweet metinleri içinden rastgele seçilmiş kelimeler olduğu dikkate alınmalıdır. Bu sözlük yalnızca ön eleme maksadıyla kullanıldığı için tam bir sınıflandırma işleminin adımları yürütülmemiş ve kelimelerin bu alandaki reklam cümleleri açısından temsil gücü tespit edilmemiştir. Sözlük tohum kelime listesi olarak kullanılarak geliştirilebilir, performans ölçümü yapılabilir

EK-5. (devam) Reklam içerikli tweetlerin ayıklanması için kullanılan reklam kelimeleri sözlüğü

Alana özel reklam içerikli kelime/kelime grupları
.co, .com, .net.tr, 7/24 açığız, 7/24 açık, 7/24 hizmetinizdeyiz, 7/24 sipariş hattı, 7/24 sipariş ver, ahşabın büyülü dünyası, anında kapına gelsin, anında kapında, ayağına gelsin, bi tık'la burada, bi tık'la gelsin, bi tık'la kapınıza gelsin, bi tık'la kapınızda, bi tık'la sipariş ver, bi tıkla burada, bi tıkla gelsin, bi tıkla kapınıza gelsin, bi tıkla kapınızda, bi tıkla sipariş ver, birbirinden enfes, birbirinden lezzetli, birbirinden nefis, bit.ly/2Ze1GIH, çeşitlerimiz, denemek ister misin, denemek istermisin, dijital pazarlama, dijitalpazarlama, dm üzerinden, doyadoyakazan, emeksiz yemek olmaz, enfes lezzetlerimiz, evinize getiriyoruz, facebook.com, gönderim bölgemiz, hemen ara gelsin, hemen sipariş et, hemen sipariş ver, hemen sipariş verebilirsiniz, hemen sipariş verin, hemen yemek sepetine sipariş, hemen yemeksepetine sipariş, http:, indirimli menü yemek sepeti'nde, indirimli menü yemek sepetinde, indirimli menü yemeksepeti'nde, indirimli menü yemeksepetinde, indirimli yemekler yemek sepeti'nden, indirimli yemekler yemek sepetinden, indirimli yemekler yemeksepeti'nden, indirimli yemekler yemeksepetinden, instagram.com, kapına gelelim, kapına gelsin, kapına getirelim, lezzet ajandası, lezzeti kapıda, lezzeti kapınıza getiriyoruz, lezzeti kapınızda, lezzetleri kapıda, lezzetleri kapınızda, mağazalarımızı arayarak, mağzalarımızı arayarak, meyvel pastalarımız, nefis lezzetler, nefis lezzetlerimiz, no.lu telefonumuzu, numaralı telfonumuzu, oluşan öğretmendim ben, parça yemek takımı, pizzavalue, pratik ve lezzetli, restoranlarımız, sipariş alımına açık, sipariş hattı, sipariş numaranızı, Sipariş numaranı, sipariş için dm, sipariş için tel, sipariş için telefon, sipariş tel, sipariş telefon, sipariş üzerine, sipariş verebilir siniz, şubelerimiz, şubemiz, tatlılar bizden, tek tık'la burada, tek tık'la gelsin, tek tık'la kapınıza gelsin, tek tık'la kapınızda, tek tık'la sipariş ver, tek tıkla burada, tek tıkla gelsin, tek tıkla kapınıza gelsin, tek tıkla kapınızda, tek tıkla sipariş ver, tıkla gelsin, twitter.com, ücretsiz kargo, ürün sipariş platformu, ürünleri sipariş platformu, üzerinden vereceğiniz, vistamart, wat's app hattı, wat's app sipariş, wat's upp hattı, wat's upp sipariş, wats upp hattı, wats upp sipariş, watsapp, watsapp hattı, watsapp iletişim, watsapp sipariş, web sitemiz, websitemiz, what's app, what's app hattı, what's app iletişim, what's app sipariş, what's upp, what's upp hattı, what's upp iletişim, what's upp sipariş, whats upp, whats upp hattı, whats upp iletişim, whats upp sipariş, whatsapp, whatsapp hattı, whatsapp iletişim, whatsapp sipariş, yemeğin kapına gelsin, yemeğin kapıya gelsin, yemek masası, yemek masası takımı, yemek sepeti'nde seni bekliyor, yemek sepetinde seni bekliyor, yemek sepetinden sipariş verebilirsin, yemek sipariş etmenize, yemek sipariş şeklimiz, yemek yapıyorum sipariş üzerine, yemekgelir.com'a, yemekgelir.com'da, yemekgetir.com'a, yemekkgetir.com'da, yemeksepeti banabi'den, yemeksepeti banabiden, yemeksepeti vale'den sipariş ver, yemeksepeti valeden sipariş ver, yemeksepeti.com'da, yemeksepeti.com'dan, yemeksepeti.com'u, yemeksepeti.comdan, yemeksepeti.comu, yemeksepeti'nde seni bekliyor, yemeksepetinde seni bekliyor, yemeksepetinden sipariş verebilirsin, yemeksepetivale'den sipariş ver, yemeksepetivaleden sipariş ver, yemeksiparişmotoru, yemekumbara'da, yemekumbarada, yılbaşıürünleri, yilbasiurunleri, yüzlerce indirimli menü, yüzlerce indirimli ürün, zencefilli yılbaşı kurabiyesi

EK-6. Twitter (X) kullanıcılarının konum bilgisi yerine yazdıkları ifadeler ve özellikleri

Twitter (X) kullanıcılarının tweet gönderilerinde elle yapılan incelemede, konum bilgisi yazma oranlarının genel olarak düşük olduğu, konum bilgisi yazanların ise farklı format, içerik ya da şekillerde yazdıkları tespit edilmiştir. Tweet veri havuzu oluşturulması esnasında Twitonomy aracılığı ile çekilen tweetlerin coğrafi konum bilgileri manuel olarak incelenmiş, uygulama tarafından harita üzerine aktarılamayan konum bilgileri üç grup olarak tasnif edilerek frekansları tespit edilmiştir. Üç grup altında manuel yapılan tasnif işlemi sonucunda elde edilen bilgiler aşağıya çıkarılmıştır. Konum bilgileri yerine Twitter (X) kullanıcıları tarafından yazılan ifadeler genel olarak üç gruba ayrılmıştır. Kelime, anahtar kelime, diyez etiketleri ve bu kelimelerin frekansları, metin madenciliğinin odak noktasını oluşturmaktadır. Bu nedenle konum bilgilerine yazılan söz konusu ifadelerin tespit edilmesi önem arz etmektedir.

(i) *Bilinmeyen:* Bu sınıf içine, …… mahallesi gibi hangi il ve ilçede olduğu belirlenemeyen isimler, Saray gibi birden çok isme sahip ilçeler ya da Mezopotomya gibi genel bölge isimleri ve Türkçe dışındaki bir dilde yazıldığı ve alfabesi nedeniyle anlaşılamayan (Rusça, Arapça gibi) yazılar dahil edilmiştir.

(ii) *Birden fazla konum bilgisi:* Birden fazla konum bilgisi yazılmış olanlar bu gruba dahil edilmiştir.

(iii) *Konum bilgisi dışında kullanılan ifadeler*: Yukarıdaki iki grup dışında kalan ve her türlü ifadeler bu gruba dahil edilmiştir.

Konum bilgisine yönelik yapılacak analizlerde kullanıcıların adres, konum bilgilerini aşağıda örnekleri verildiği şekilde farklı form ve yapıda yazdıkları dikkate alınmalıdır.

EK-6. (devam) Twitter (X) kullanıcılarının konum bilgisi yerine yazdıkları ifadeler ve özellikleri

Konum bilgisi yerine yazılan ifadeler	Açıklama
İstanbul	---
İstanbul, Türkiye	---
Türkiye	Türkiye içine yönelik çalışmalarda bu konum bilgisi kullanılamayacaktır.
Karşıyaka, İzmir	
trabzon	Büyük, küçük harf kullanımı ile farklı yazımlar mevcuttur.
Türkiye / Eskişehir	İçerik olarak aynı bilgi farklı formatlarda yazılmıştır.
Eskişehir, Türkiye	
saray	Hem Tekirdağ hem Van'ın Saray ilçesi olması nedeniyle yalnızca bu yer isminin yazılmış olması konum bilgisi olarak bir anlam taşımaz.
…… mahallesi	Yalnızca mahalle adı yazıldığında aynı isimle çok sayıda mahalle olabileceği için neresi olduğunun belirlenmesi mümkün değildir.
evren	Konum bilgisi yerine girilen kelime, cümle, siyasi içerikli sözler, bina, tesis, kurum adı gibi örnekler mevcuttur.
kırık türküler diyarı	
Ho	Karşılığı olmayan yer isimleri (hayali ya da filmlerde geçen hayali yer isimleri gibi)
Kocaeli / Sinop	Birden fazla konum bilgisi, kişilerin bir şekilde bağlantısı olan yer isimlerini birlikte yazmaktadırlar.
Berlin / Konya	
İzmit	Merkez ilçe isimleri, il ismi olarak kullanılmaktadır.
Kocaeli	
Istanbul	Noktalı Türkçe karakterlerin kullanılmadığı yazımlar
34	Muhtemelen il plaka bilgisi
34750	Muhtemelen posta kodu bilgisi
İstanbul konumundayım	Kullanıcılar bazen cümle şeklinde ifadelerle nerede oldukları bilgisini yazmaktadırlar.
Turkey	
Angara	Bilinçli ya da hatalı yazım
Fenerbahçe, beşiktaş	Takım adı mı yoksa yer adı mı olduğu anlaşılmayan isimler
T.C.	Ülke kısaltması
Ank, ist	Muhtemelen şehrin ilk üç harfini kullanarak yapılan kişiye özel kısaltmalar
USA, VA, CA, FL, TR	Yer isimlerinin uluslararası kısaltmaları
Başkent	Ankara anlamında kullanıldığı değerlendirilmiştir.
Türkistan, mezopotomya	Noktasal yer ismi yerine genel bölge isimleri
Ağasar, konstanipol	Yabancı dildeki yer ismi karşılıkları ya da ideolojik, siyasi maksatla kullanılan isimler

EK-6. (devam) Twitter (X) kullanıcılarının konum bilgisi yerine yazdıkları ifadeler ve özellikleri

İzmir'li	konum olarak nerede olduğu bilgisi çıkarılamayacak ama yer ismi içeren ifadeler
İstanbul, adanalı	Bu ifadelerden bu kişinin ikametgahının İstanbul ama memleketinin adana olduğu değerlendirilebilir.
samsun, #herseyguzelolacak	Siyasi mesaj içerikli ifadelerle birlikte yer ismi yazılması
Urfa, maraş	Birleşik kelimelerden oluşan yer isimlerinin sadece bir kelimesinin yazılması.
…. Vakfı İstanbul Şubesi	Yer ismi de içerecek şekilde yazılmış vakıf, kurum, dernek isimleri.
Bir sevdadır urfa	Yer isminin de yer aldığı cümle ya da ifadeler.
kayseripresshaber	Yer ismi de kapsayan internet sitesine ait isimler.
İzmirin göçmen kızı	Yer simi ile birlikte kullanılan ifadeler. Bu ifadeden kişinin İzmir ile bir bağlantısı olduğu değerlendirilebilir ancak konumunneresi olduğu konusunda kesin bir yargıya ulaşmak mümkün gözükmemektedir. Belki kişinin tweeti de ayrıca incelense belki konumu ile ilgili bir destekleyici başka bir ifade varsa konum hakkında yargıya varmak mümkün olabilir.
Değirmendere ama konum düzce	Cümle şeklinde konum bilgisi
05……..	Genelde reklam tweetlerinde konum bilgisi yerine internet adresi ya da telefon numarası yazılmaktadır.
AİBÜ, ODTÜ, KTÜ	Üniversite isimlerinin kısaltmaları
……. Genel Başkanı	Siyasi parti liderleri ya da milletvekili ve bakanların isimleri yazılmaktadır.
İstanbul ♡	Yer isimleri ile birlikte duygu ifadeleri belirten emojiler.
…….. et ürünleri	Reklam maksatlı marka ve ürün isimleri
ي….	Arapça ve Rusça gibi alfabesi farklı dillerde yazıldığı için neresi olduğu ya da ne yazıldığı anlaşılmayan ifadeler

Twitter (X) kullanıcıları tarafından, CKB yerine; birbiri ile ilgisiz birden fazla konum bilgisinin girildiği, yalnızca mahalle ya da cadde ismi gibi eksik bilgi girişi yapıldığı, konum bilgisi yerine yer adı dışında yazı, karakter, simge girilmesi, yazım hataları, konuşma dilindeki ifadelerin (üniversite adı, futbol takımı, simge-🍔, e-posta adresi, gibi esprili ifadeler) yazıldığı, sadece ülke adı ile konum bilgisi girildiği, konum bilgilerinin farklı formatlarda yazıldığı (Antalya, Turkey; Antalya/Alanya ve Antalya gibi) görülmüştür.

EK-7. Koleksiyon sınıflandırması için değerlendiricilere ön bilgilendirme notu

Çalışmada kullanılan koleksiyon, *çevrimiçi/telefonla ev/işyerinden yemek siparişi* veren kişilerin konuyla ilgili Twitter (X) paylaşımlarındaki cümlelerinden meydana gelmektedir. Çalışma herhangi bir şirket ya da işletme ismine yönelik yapılmamış, tez metninde belirtildiği şekilde anahtar kelime belirleme yöntemleri kullanılarak belirlenen anahtar kelimeler çerçevesinde koleksiyon oluşturulmuştur. Koleksiyondaki cümlelerin çoğunluğu ÇYS konusu ile ilgili olmakla birlikte, koleksiyon içinde az sayıda da olsa konu kapsamı dışında tweet cümleleri de mevcuttur.

Değerlendiriciler, tweet cümlelerinin her birini müstakilen değerlendirecek ve cümlenin müteakip maddelerde belirtilen sınıflardan hangisine dahil olması gerektiğine karar vereceklerdir. Reklam sınıfına ayrılmayan her bir tweet cümlesi için, aşağıda belirtilen ikinci ve üçüncü grup sınıflandırma içindeki sınıflardan birine atama işlemi yapılacaktır.

- Birinci grup sınıflandırma: Reklam (Eğer tweet cümlesi reklam olarak sınıflandırılmışsa o cümle için iki ve üçüncü gruptaki sınıflandırma yapılmayacaktır.),

- İkinci grup sınıflandırma: Pozitif-negatif-nötr,

- Üçüncü grup sınıflandırma: Ürün-hizmet-diğer.

Her tweet cümlesinin karşısında tabloda iki tane sınıflandırma sonucunu gösterecek şekilde 1 sayısı yazılır. Bir tanesi duygu sınıfını (pozitif, negatif, nötr), diğeri ürün-hizmet sistemleri (ürün, hizmet, diğer) sınıfını gösterir.

Her cümle müstakil olarak değerlendirilecektir, cümlenin yarım bırakılmış olması, eksik kelimelerle cümle kurulması dikkate alınmaksızın edinilen kanaat çerçevesinde sınıflandırma yapılacaktır. Tweet cümlesinin ürün, hizmet ve genel şeklindeki sınıflandırması için Çizelge 3-4'te verilen konu başlıklarının esas alınarak sınıflandırma yapılması, sınıflandırma işlemlerini kolaylaştıracaktır.

EK-7. (devam) Koleksiyon sınıflandırması için değerlendiricilere ön bilgilendirme notu

Özet olarak, ürüne ilişkin konular yiyeceğin lezzeti, kokusu, tadı, sosu, içindekileri, bayat, taze olması, eksik malzeme ya da istenen dışında bir ürün getirilmesi gibi konulardır. Hizmet ise işletme, lokanta, burada çalışanlar, yemek şirketinin çevrimiçi uygulamaları, müşteri temsilcileri, kurye, işletmelerin yemek ile ilgili promosyon, indirim, joker uygulamaları (mesela çok sipariş verenlere muhtar ünvanı verilmesi gibi bir uygulama var), şirketin televizyon/internette yayımladığı reklam hakkında yorumlar, restorantların varsa web sitesinin nasıl olduğu, çevrimiçi sistemlerden karttan para çekme işlemleri, şirketlerin yemek kartları, işletmeler, işletme sahipleri, ürünün pahalı olması, çalışanlar kurye, çevrimiçi hizmetler hakkındaki düşünceler, çalışan kuryelerin yağışlı havalarda çalışmaları, sipariş notu, kuryeler için bahşiş verilmesi gibi konuları kapsamaktadır. Duygu sınıflandırması konusunda değerlendiricilerin tamamen cümleden edindikleri duygu doğrultusunda sınıflandırma yapmaları beklendiğinden bir tablo hazırlanmamıştır.

Bir tweet cümlesi eğer bir firmanın/lokantanın reklamını yapıyorsa bu cümle reklam cümlesi olarak kabul edilecektir. Reklam içerikli tweet cümleleri Excel tablolarından ayıklanmış olmakla birlikte koleksiyon içinde az sayıda da olsa reklam cümlesi olması ihtimal dahilindedir. Örneğin, “harika pizza lezzetleri, tıkla/hemen ara siparişin kapına gelsin, tlf: 542 789 44 35” gibi bir tweet cümlesi, bir lokantanın reklamıdır. Eğer bir cümlenin reklam cümlesi olduğuna karar verilirse artık o cümle üzerinde başka bir değerlendirme yapılmasına gerek yoktur. Tabloda reklam kutucuğunun içine “1” rakamı yazılacaktır. Örneğin “……nın reklamları çok kötüydü” şeklinde bir ifade varsa, bu tweet cümlesi, tweeti yazan kişinin reklam hakkındaki bir görüşünü içerdiği için reklam cümlesi olarak kabul edilmeyecek, ikinci ve üçüncü gruptaki sınıflandırmalara dahil edilecek, tabloda seçildiği sınıf kutusu içine “1” yazılacaktır.

Bazı cümlelerde şöyle durumlarla karşılaşmak mümkündür: Tweeti yazan kişi konu hakkında (yemek şirketi, lokanta ile ilgili) hem olumlu yani beğendiği, hem olumsuz yani beğenmediği hususları ifade edebilir, iki farklı şirketin karşılaştırmasını yaparken birini överken diğerini eleştirebilir. Bu tür cümlelerde hangi duygunun (pozitif, negatif, nötr) daha ağır bastığı düşünülüyorsa o yönde değerlendirme yapılabilir. Benzer şekilde bir tweet içinde hem ürün hem hizmetle ilgili görüşler belirtiliyor olabilir. Bu tür durumlarda da yine (ürün,

EK-7. (devam) Koleksiyon sınıflandırması için değerlendiricilere ön bilgilendirme notu

hizmet ve diğer sınıflarından) ağır basan sınıf için etiketleme yapılabilir. Cümleye verilen sınıf etiketi için kutucuğun içine 1, diğerlerine 0 (sıfır) yazılır.

Bazı cümlelerde ise kararsızlık, pişmanlık ("gece yarısı yemek yemek" ya da "diyeti bozmak" gibi nedenlerle) ifadeleri olabilir. Değerlendirici tarafından ürün ya da hizmete ilişkin kişinin duygusu konusunda ne tür bir düşünce algılanmışsa o yönde sınıf etiketi verilmelidir.

Tweet cümlesine sınıf etiketi verilirken, şu hususa özellikle dikkat edilmesi gerekir. Twitter (X) kullanıcısının "bu akşam pide sipariş ettim" şeklinde bir cümlesi, kişinin sırf pide sipariş etmiş olması nedeniyle pozitif bir duygu olarak kabul edilmemelidir. Çünkü tüm koleksiyon zaten kullanıcıların yemek sipariş vermesi ile ilgilidir. Önemli olan sipariş verdiği ürün ya da aldığı hizmet ile ilgili edindiği olumlu ya da olumsuz duygunun ne olduğunun belirlenmesidir. Diğer bir ifade ile tweet cümlelerinde, kişilerin ürün ve hizmetle ilgili ("fena değilmiş", "pideyi geç getirdiler", "pideyi getirirken malzemeleri dağılmıştı", "güzel pideydi", "arkadaşlarla keyifle yedik" gibi) etkileşimleri sınıflandırma yapılmasında esas belirleyici olmalıdır.

Çalışma konusunun, sadece internetteki uygulamalar üzerinden verilen yemek siparişi olduğu düşünülmemelidir. Mahallesindeki lokantaya telefonla sipariş veren kişinin bu konuyla ilgili görüşlerini belirten cümleleri de çalışma kapsamında değerlendirilecektir. Öte yandan anahtar kelimeler nedeniyle koleksiyona girmiş bulunan "yemek masası", "yemek masa örtüsü" gibi siparişe konu olan ürünlere ilişkin tweetler çalışma kapsamı dışında olduğundan nötr ve diğer sınıfı içine ayrılmalıdır. Tweet cümleleri içinde bir iki kelimeden oluşan çok kısa cümlelerin yer aldığı dikkate alınırsa, bazı tweet cümleleri hakkında değerlendirme yapmak imkansız olabilir, bazı cümlelerde ise belirtilen görüşün olumlu mu olumsuz mu olduğu anlaşılamamaktadır. Bu tür cümlelerin nötr ve diğer olarak sınıflandırılması uygun olacaktır. Bazen konu doğrudan verilen yemek siparişi ile ilişkili olmasa bile, yemek siparişine ilişkin bir benzetme ifadesi kullanıldığı için anahtar kelimelere bağlı olarak koleksiyona girmiş olabilir. Bu durumda benzetme ifadeleri, eğer tweeti yazan kişinin yemek siparişlerine ilişkin geçmiş deneyimlerinin bir neticesi olarak ortaya çıkmış görünüyorsa, böyle bir cümlenin de duygu yönü tespit edilip sınıf etiketi verilebilir.

EK-7. (devam) Koleksiyon sınıflandırması için değerlendiricilere ön bilgilendirme notu

Bir tweet cümlesinin sınıfı konusunda eğer hiçbir şekilde karar verilemiyorsa, nötr ve diğer sınıfına ayrılması uygun olacaktır.

Koleksiyon içinde, ürün ve hizmete ilişkin görüşlerin yanı sıra, yemek şirketinin çalışanlarının çalışma koşullarına ilişkin paylaşımların da yapıldığı görülmektedir. Müşterilerin ya da çalışanların çalışma koşullarına ilişkin görüşlerinin de Twitter (X) platformunda paylaşılması, işletmelerin kamuoyu algılarına doğrudan etki edebilecek bir unsurdur. Çalışanların motivasyonlarının işletmelerin verimliliğini doğrudan artırabileceği ya da azaltabileceği dikkate alındığında bu alanda da iyileştirmeye ihtiyaç duyulabilecektir. Bu nedenle bu konularla ilgili paylaşımlarında analizlere dahil edilmesi için hizmet sınıfı şeklinde etiketlenmesi uygun olacaktır.

Eğer bir cümle; (i) tamamen inceleme konusu dışında bir konuyla ilgili ise, (ii) yemek konusunda genel bir ifade ise, (iii) ifade edilen hususlardan hiçbir şey anlaşılmıyorsa, ifadelere bir anlam vermek mümkün değilse, ifadelerin ne ile ilgili olduğu belli değilse diğer-sınıfına ayrılabilir.

Eğer, (i) tweetin taşıdığı duygu yönü anlaşılmıyorsa, (ii) tweet cümlesi konu dışında bir içerik taşıyorsa, (iii) ifadelere her iki yönlü (olumlu ve olumsuz) olarak anlam yüklemek mümkünse ya da her iki yönde de anlam yüklemek mümkün değilse nötr sınıfında olduğu kabul edilebilir.

Değerlendiricilerin birbirlerini etkilememesi için, tablolarda diğer değerlendiricilere ait değerlendirme gösterilmemiştir.

EK-8. 66 Twitter (X) kullanıcısı için altı merkezilik ölçütüne göre merkezilik değerleri

Kullanıcı İsmi	Derece Merkeziliği	Arasındalık Merkeziliği	Yakınlık Merkeziliği	Sayfa Sıralaması	Özvektör Merkeziliği	Katz Merkeziliği
@666lehce	0.0	0.0	0.0	0.0037	0,03	0.0688
@abdulkadir06161	0.0153	0.0	0.1350	0.0059	0.0192	0.0845
@abdullahciftcib	0.0153	0.0	0.0349	0.0139	0,21	0.0783
@adilomera1	0.0461	0.0033	0.0549	0.0360	-0,13	0.0949
@afmsudan	0.0153	0.0	0.0349	0.0210	-0,11	0.0792
@AliYerlikaya	0.0153	0.0	0.0153	0.0246	-1,16	0.0764
@amanda_devik	0.0153	0.0	0.1824	0.0047	0.0737	0.1079
@artdbkglu	0.0153	0.0	0.1704	0.0099	0.0774	0.1076
@Asibelblonde	0.0153	0.0	0.1704	0.0068	0.0774	0.1076
@avozgeustun	0.0	0.0	0.0	0.0037	-0,01	0.0688
@ay_moon_ay	0.0	0.0	0.0	0.0037	0,02	0.0688
@ayse_cng	0.0153	0.0	0.1704	0.0099	0.0774	0.1076
@bangtantr	0.0	0.0	0.0	0.0037	0,00	0.0688
@besisko	0.0	0.0	0.0	0.0037	0,01	0.0688
@bilgingokberk	0.2923	0.0584	0.2888	0.1314	0.5211	0.3881
@burcincigerim	0.0307	0.0	0.1962	0.0099	0.5211	0.1467
@Can_Y_Can	0.0615	0.0129	0.2080	0.0174	0.1696	0.1677
@cemreallly	0.0307	0.0	0.0	0.0246	0,03	0.0764
@CEngven	0.0307	0.0	0.1962	0.0161	0.1511	0.1467
@chabyhan	0.0	0.0	0.0	0.0037	0,03	0.0688
@covid20loading	0.0	0.0	0.0	0.0037	-0,04	0.0688
@ctetikog	0.0307	0.0	0.1962	0.0130	0.1511	0.1467
@damlaaltuun	0.0	0.0	0.0	0.0037	-0,08	0.0688
@deepturkweb	0.0	0.0	0.0	0.0037	0,00	0.0688
@delykalamis	0.1076	0.0004	0.2080	0.0403	0.2783	0.2192
@dominoeffect34	0.0307	0.0	0.1962	0.0161	0.1511	0.1467
@enveryan	0.0153	0.0	0.0349	0.0139	0,20	0.0783
@er_maannn	0.0307	0.0	0.1962	0.0161	0.1511	0.1467
@gela_wej	0.0153	0.0	0.1368	0.005	0.0229	0.0843
@gftufekci	0.0461	0.0	0.2	0.0221	0.1925	0.1686
@hallederiztamam	0.0769	0.0033	0.0549	0.0509	0,09	0.1046
@handebbyrktr	0.0	0.0	0.0	0.0037	0,00	0.0688
@handelendi	0.0	0.0	0.0	0.0037	0,00	0.0688
@hasankarsavuran	0.0153	0.0	0.0349	0.0123	-0,01	0.0792
@hatirliyoruz	0.0	0.0	0.0	0.0037	-0,23	0.0688
@ISTANist	0.0307	0.0	0.1962	0.0161	0.1511	0.1467
@ibrahimbagyapan	0.0461	0.0	0.2,	0.0221	0.1925	0.1686
@ihsanerkahraman	0.0153	0.0	0.0153	0.0246	-1,13	0.0764
@izmirevdekal	0.0	0.0	0.0	0.0037	-0,02	0.0688
@kagansaydan	0.0461	0.0019	0.2039	0.0186	0.1704	0.1624
@KOXALBEY	0.0461	0.0	0.2	0.016	0.1925	0.1686
@lordsinov	0.0,	0.0	0.0	0.0037	0,13	0.0688
@mbagun11	0.0307	0.0	0.189	0.0312	0.0930	0.1236
@mert_abras	0.0	0.0	0.0	0.0037	0,00	0.0688
@merveyldrm2525	0.0307	0.0	0.189	0.0065	0.0989	0.1247
@mervezell	0.0	0.0	0.0	0.0037	0,07	0.0688
@mizantropii	0.0	0.0	0.0	0.0037	-0,08	0.0688
@mukinaber	0.0	0.0	0.0	0.0037	0,00	0.0688
@ninegeek9	0.0	0.0	0.0	0.0037	0,01	0.0688
@oguzhantesla	0.0	0.0	0.0	0.0037	0,01	0.0688
@ozelegitimci_18	0.0	0.0	0.0	0.0037	-0,05	0.0688
@pillreder	0.0	0.0	0.0	0.0037	-0,13	0.0688
@saddandmad	0.0	0.0	0.0	0.0037	-0,16	0.0688
@sakizciko	0.0153	0.0	0.1386	0.0055	0.0252	0.0855
@sinoreyiz	0.0	0.0	0.0	0.0037	0,17	0.0688
@srfzmr	0.0461	0.0120	0.2039	0.0204	0.1546	0.1551
@thewizarddofoz	0.0	0.0	0.0	0.0037	0,00	0.0688
@tinercimayki	0.0	0.0	0.0	0.0037	-0,04	0.0688
@toskofacts	0.0	0.0	0.0	0.0037	0,02	0.0688
@UstAkilOyunlari	0.0	0.0	0.0	0.0037	0,00	0.0688
@Wotah_	0.0	0.0	0.0	0.0037	-0,01	0.0688
@yarftm	0.0307	0.0	0.189	0.0138	0.093	0.1236
@yemeksepeti	0.0307	0.0899	0.325	0.1440	0.4964	0.3912
@yokmaalesef	0.0	0.0	0.0	0.0037	-0,03	0.0688
@zagortenay76	0.0769	0.0127	0.2	0.0403	0.1295	0.1573
@zehra_mete	0.0153	0.0	0.1824	0.0057	0.0737	0.1079
@666lehce	0.0	0.0	0.0	0.0037	0,03	0.0688



Gazili olmak ayrıcalıktır