



**TÜRKÇE KULLANICI YORUMLARINDAN HEDEF TEMELLİ DUYGU
ANALİZİ GERÇEKLEŞTİRİMİ**

Kıvanç BAYRAKTAR

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

KASIM 2019

Kıvanç BAYRAKTAR tarafından hazırlanan “TÜRKÇE KULLANICI YORUMLARINDAN HEDEF TEMELLİ DUYGU ANALİZİ GERÇEKLEŞTİRİMİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Dr. Öğr. Üyesi Uraz YAVANOĞLU

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. M. Ali AKCAYOL

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Üye: Dr. Öğr. Üyesi Adnan ÖZSOY

Bilgisayar Mühendisliği Ana Bilim Dalı, Hacettepe Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 07/11/2019

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Sena YAŞYERLİ
Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

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

- 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 çalışmada 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.

Kıvanç BAYRAKTAR

07/11/2019

TÜRKÇE KULLANICI YORUMLARINDAN HEDEF TEMELLİ DUYGU ANALİZİ GERÇEKLEŞTİRİMİ

(Yüksek Lisans Tezi)

Kıvanç BAYRAKTAR

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Kasım 2019

ÖZET

Kişilerin yorumları başka insanların ve insanlara hizmet sağlayan kurumların karar verme sürecinde önemli bir yere sahiptir. İnternetin yaygınlaşması neticesinde kişilerin görüşlerini bu ortamda ifade etmesinden kaynaklı veri boyutu günden güne artmaktadır. Bu verinin insanlar tarafından herhangi bir araç kullanılmadan analiz edilmesi mümkün değildir. Bu sebeple duygu analizi kavramı ortaya çıkmıştır. Duygu analizi alanında yapılan çalışmaların çoğunda veri pozitif, negatif veya nötr olarak sınıflandırılır, kişilerin konular hakkındaki bakış açılarına göre sınıflandırma yapılmaz. Daha detaylı sonuç alınması için hedef temelli duygu analizi gereklidir. Bu çalışmada duygu analizinin alt seviyesi olan hedef temelli duygu analizinin Türkçe veri setinde gerçekleştirilmesi için bütüncül bir yaklaşım sunulmuştur. Hedef terimlerin çıkarılması için LDA konu modelleme tabanlı iki yaklaşım önerilmiştir. Önerilen yaklaşımlarda LDA'nın kelime çantası probleminin çözülmesi amaçlanmıştır. Belirlenen hedef terimleri kategorilere ayırmak için PMI yönteminden yararlanan bir metot geliştirilmiştir. Hedef terimler ve kategoriler belirlendikten sonra sınıflandırma işlemini gerçekleştirmek için kural ve sözlük tabanlı yaklaşımlardan yararlanılmıştır. Hedef terim çıkarma işleminde LDA temelli yöntem ile %70'in üzerinde f-skor elde edilmiştir. Hedef terim kategori eşleştirmesinde %59 oranında doğruluk elde edilirken, hedef terimlerinin polaritesinin belirlenmesinde ise %79 oranında doğruluk elde edilmiştir.

Bilim Kodu	: 92416
Anahtar Kelimeler	: Duygu analizi, hedef temelli duygu analizi, hedef terim çıkarımı
Sayfa Adedi	: 52
Danışman	: Dr. Öğr. Üyesi Uraz YAVANOĞLU

IMPLEMENTATION OF ASPECT-BASED SENTIMENT ANALYSIS FROM TURKISH USER REVIEWS

(M. Sc. Thesis)

Kıvanç BAYRAKTAR

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

November 2019

ABSTRACT

Opinions of people greatly impact the decision-making process of other people and institutions providing services to the society. Today, internet has become an integral part of our daily lives as it people can easily disseminate their thoughts through the internet. Along with the rapidly increasing data sizes, data processing has become even more challenging. As a result, the concept of sentiment analysis has emerged. Most studies involving sentiment analysis do not reflect the interpretations of individuals, but only classify data as positive, negative or neutral. Aspect-based sentiment analysis takes into account people's opinions to come up with more detailed results. In this study, a holistic approach has been proposed to perform an aspect-based sentiment analysis -a sub-level of sentiment analysis- on Turkish data. Two approaches based on LDA topic modeling have been proposed for the extraction of aspect terms. Proposed approaches aim to solve the LDA's bag of words problem. An algorithm that uses the PMI method is proposed to categorize the identified aspect terms. After determining the aspect terms and categories, rule and dictionary-based approaches have been used to perform the classification process. In the aspect extraction phase, an f-score of over 70% was obtained by LDA-based method. In the aspect term category matching stage, 59% accuracy was achieved. In determining the polarity of aspect terms 79% accuracy was attained.

Science Code : 92416
Key Words : Sentiment analysis, aspect-based sentiment analysis, aspect extraction
Page Number : 52
Supervisor : Asst. Prof. Dr. Uraz YAVANOĞLU

TEŞEKKÜR

Çalışmam boyunca değerli fikirlerini ve yardımını esirgemeyen danışman hocam sayın Dr. Uraz YAVANOĞLU'na teşekkürü borç bilirim. Bu süreçte bilgilerinden ve deneyimlerinden yararlandığım, bana yol gösteren Dr. Güçlü ONGUN'a, Dr. Alper ÇEVİK'e ve Sema AYGAR'a teşekkür ederim. Yüksek lisans eğitimi almamda beni cesaretlendiren Hemosoft Bilişim ve Eğitim Hizmetleri A.Ş.'deki işverenlerime ve tüm çalışma arkadaşlarıma teşekkür ederim.

Her koşulda beni destekleyen, her zaman yanımda olan annem Zuhal BAYRAKTAR'a, babam Bahattin BAYRAKTAR'a ve ağabeyim Ender BAYRAKTAR'a sonsuz teşekkür ederim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. LİTERATÜR TARAMASI / İLGİLİ ÇALIŞMALAR.....	3
2.1. Duygu Analizi Kullanım Alanları	3
2.2. Duygu Analizi Türleri	3
2.2.1. Doküman seviyesi duygu analizi	3
2.2.2. Cümle seviyesi duygu analizi	6
2.2.3. Hedef temelli duygu analizi	7
2.3. Sözcük Tabanlı Duygu Sınıflandırma	10
2.3.1. Sözlük tabanlı yaklaşım	10
2.3.2. Derlem tabanlı yaklaşım	11
3. BİLİMSEL ARKA PLAN.....	13
3.1. Latent Dirichlet Allocation (LDA).....	13
3.2. Pointwise Mutual Information (PMI).....	15
3.3. WSBFE	16
3.4. C-value	16
3.5. Bag of Words ve N-gram model	19
3.6. Yandex.XML	20

	Sayfa
4. PROBLEM TANIMI VE ÖNERİLEN METOT	21
4.1. Veri Seti Oluşturma.....	21
4.2. Veri Ön İşleme	22
4.2.1. Cümlelere ayırma	23
4.2.2. Metin düzeltme	23
4.3. Hedef Terimlerin Çıkarımı.....	23
4.4. Hedef Terimlerin Kategorisinin Bulunması	25
4.5. Hedef Terimlerin Duygu Sınıflandırması	25
5. DENEYSEL SONUÇLAR	29
5.1. Hedef Terimlerin Çıkarımı.....	29
5.2. Hedef Terimlerin Kategorisinin Bulunması	35
5.3. Hedef Terimlerin Duygu Sınıflandırılması	35
6. SONUÇ VE ÖNERİLER	39
KAYNAKLAR	41
EKLER.....	47
EK-1. Veri setinden örnek kısım	48
EK-2. Çok kelimeli terimler	49
ÖZGEÇMİŞ	52

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. C-value örnek sonuçlar	18
Çizelge 4.1. Hedef terim kategorileri.....	22
Çizelge 5.1. Hedef terimlerin çıkarımına ait hata matrisi.....	30
Çizelge 5.2. Hedef terimlerin çıkarımı ait başarı metrikleri	30
Çizelge 5.3. Metotların çok kelimeli terimleri keşfetme başarılarının karşılaştırılması hata matrisi	34
Çizelge 5.4. Metotların çok kelimeli terimleri keşfetme başarılarının karşılaştırılması.....	34
Çizelge 5.5. Başlangıç hedef terim listesi verilince elde edilen hata matrisi ve doğruluk oranı.....	35
Çizelge 5.6. Başlangıç hedef terim listesi verilmeden elde edilen hata matrisi ve doğruluk oranı.....	35
Çizelge 5.7. Hedef terim listesi verilerek sınıflandırılmış bazı hedef terimler	36
Çizelge 5.8. Hedef terim listesi verilmeden sınıflandırılmış bazı hedef terimler	36

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 3.1. LDA'nın grafiksel model gösterimi.....	14
Şekil 3.2. Yandex.XML örnek kullanımı.....	19
Şekil 4.1. Hedef temelli duygu analizi için önerilen yöntem.....	21
Şekil 4.2. Veri setinin sınıf diyagramı	22
Şekil 4.3. Ön işleme aşaması	23
Şekil 4.4. Hedef terim çıkarımı akış diyagramı	24
Şekil 4.5. Hedef terim çıkarım algoritması	24
Şekil 4.6. Hedef terim kategori eşleme algoritması	25
Şekil 4.7. Kural tabanlı yaklaşımın mantıksal ifadesi.....	26
Şekil 4.8. Hedef terim – görüş eşleştirme ve duygu sınıflandırma algoritması	27
Şekil 5.1. M-III ve M-VI numaralı metotların karşılaştırılması	31
Şekil 5.2. C-value yönteminden yararlanan metotlar ile LDA'nın karşılaştırılması	32
Şekil 5.3. WSBFE yönteminden yararlanan metotlar ile LDA'nın karşılaştırılması.....	32
Şekil 5.4. Hedef terim çıkarım yöntemlerinin karşılaştırılması	33

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

Σ

Genel toplam operatörü

Π

Genel çarpım operatörü

Kısaltmalar

Açıklamalar

ANTF

Augmented Normalized Term Frequency

API

Application Programming Interface

BoW

Bag of Words

CRF

Conditional Random Field

EM

Expectation Maximization

k-NN

k-Nearest Neighbors Algorithm

LDA

Latent Dirichlet Allocation

ME

Maximum Entropy

MI

Mutual Information

MNB

Multinomial Naive Bayes

NB

Naive Bayes

PMI

Pointwise Mutual Information

POS

Part of Speech

RF

Random Forest

SVM

Support Vector Machine

TF

Term Frequency

TF-IDF

Term Frequency – Inverse Document Frequency

1. GİRİŞ

Düşünceler insan hayatını etkileyen ve kurumların ya da işletmelerin gelecek planlarını şekillendirmesini sağlayan en temel unsurdur. İnsanlar karar verme ve seçim yapma aşamasında başka kişilerin düşüncelerini merak ederler ve onların deneyimlerinden yararlanmak isterler. Eskiden, insanlar yakın çevresindeki kişilerin tecrübelerinden faydalanırken internetin yaygınlaşması sonucu, zamanla bu durum değişmeye başlamıştır. Gündelik yaşamda internetin önemli bir konuma sahip olması ve insanların düşüncelerini internet aracılığı ile kolayca yayabilmesi neticesinde internet üzerindeki veriler karar verme sürecinde faydalanılabilecek etkin bir araç olmuştur. Fakat her geçen gün veri boyutunun artması, insanların bu verilerden tam anlamıyla yararlanabilmesinin önündeki en büyük engeldir. Örneğin, bir kişinin herhangi bir ürün veya hizmet almadan önce o ürün veya hizmet hakkında yapılmış yorumların hepsini okuyup değerlendirmesi, veri boyutunun büyüklüğü ile orantılı olarak oldukça zor bir süreç halini almaktadır. Bu süreci kolaylaştırarak zaman kaybını önlemek amacıyla “duygu analizi” kavramı ortaya çıkmıştır.

Duygu analizi ile kişilerin belirli bir konu hakkında yargıya varmadan önce başka insanların görüşlerini dikkate alarak karar sürecini kolaylaştırmak ve bu süreci otomatikleştirmek amaçlanmaktadır. Duygu analizi kısaca, verilen metin içindeki görüşleri bulup sunma işlemi olarak tanımlanabilir.

Duygu analizi 3’e ayrılmaktadır [1]:

- Doküman seviyesi duygu analizi
- Cümle seviyesi duygu analizi
- Hedef temelli duygu analizi

Doküman seviyesi duygu analizinde bütün dokümanın pozitif veya negatif duyguyu ifade edip etmediği belirlenmektedir. Cümle seviyesi duygu analizi ise doküman içindeki her bir cümlenin pozitif veya negatif duygu ifade edip etmediği bulmayı amaçlar. Ancak bu iki analiz de insanların neyi beğenip beğenmediğini bulamaz. Daha detaylı bir analiz için hedef temelli duygu analizi uygulanır. Hedef temelli duygu analizi hedef çıkarımı ve hedef duygu sınıflandırması aşamalarından oluşmaktadır [1]. Hedef çıkarımı aşamasında metindeki konular bulunurken, hedef sınıflandırma aşamasında çıkarılan hedefler pozitif, negatif veya nötr olarak sınıflandırılır.

Bu çalışmanın 3 temel amacı vardır. Birincisi yeni ve farklı metotlar kullanılarak hedef terim belirleme aşamasının performansını artırmaktır. İkinci olarak literatürde kullanılan güncel yaklaşımlardan birisi olan sözlük tabanlı sınıflandırma yönteminden ve kural tabanlı yaklaşımdan yararlanarak bütüncül bir hedef temelli duygu analizi çalışması gerçekleştirmektir. Çalışmanın bir diğer hedefi ise Türkçe çalışmalar için kullanılabilecek bir veri seti oluşturmaktır.

2. LİTERATÜR TARAMASI / İLGİLİ ÇALIŞMALAR

Duygu analizi insanların farklı markalar, bireyler ve gruplar hakkındaki düşüncelerini, tutumlarını ve duygularını çözmeye çalışan bir çalışma alanı olarak tanımlanmaktadır [1].

Bu bölümde sırasıyla duygu analizinin kullanım alanları, duygu analizi türleri ve sözcük tabanlı duygu sınıflandırma ile ilgili literatürde yer alan önemli çalışmalara yer verilmektedir.

2.1. Duygu Analizi Kullanım Alanları

Duygu analizi ile insanların belirli konular hakkında bilgi edinmelerinin kolaylaştırılması amaçlanmaktadır. İnsanlar bir ürünü veya hizmeti almadan önce diğer insanların o konu hakkındaki görüşlerini dikkate alarak tercihlerini gerçekleştirirler. Benzer şekilde şirketler ve kuruluşlar, müşteri yorumlarını analiz ederek iyi ve kötü taraflarını tespit ederler. Duygu analizi çalışmaları insanı ilgilendiren her alana uygulanabilir, bu nedenle bu alanda günlük hayatı ilgilendiren çok sayıda çalışma yapılmıştır.

Liu ve arkadaşları, blog verilerinden satış performansının tahmin edilmesi için Sentiment-PLSA isimli metodu önermişlerdir [2]. Pennacchiatti ve Popescu, Twitter kullanıcılarının davranış, çevre ve tweet gibi bilgilerinden yararlanarak kullanıcıların politik yönelim ve etnik kökenini tespit etmiştir [3]. Twitter üzerinden politik analiz yapan başka çalışmalar da mevcuttur [4, 5]. Bollen ve arkadaşları, tweetlerdeki ruh hali değişikliklerini ele alarak borsa tahmini yapmışlardır [6]. Şimşek ve Özdemir, Türkçe tweetler ile borsa endeksi arasındaki ilişkiyi araştırmışlardır [7]. Birjali ve arkadaşları, Twitter verisi kullanarak intihar düşüncesi kestirimi yapmıştır [8]. Baj-Rogowska, Facebook yorumlarını kullanarak Uber hakkında analiz yapmıştır [9]. Otel yorumları [10 - 12] ve film yorumları [13, 14] kullanılarak duygu analizi çalışmaları gerçekleştirilmiştir.

2.2. Duygu Analizi Türleri

2.2.1. Doküman seviyesi duygu analizi

Bu analiz tipinde görüş içeren doküman seti D olmak üzere, her dokümanın $d \in D'$ nin olumlu veya olumsuz bir görüş ifade edip etmediğini belirlenir [15]. Bu kısımda doküman seviyesi duygu analizinde yapılmış bazı çalışmalar incelenmiştir.

Turney [16] çalışmasında, yorumları tavsiye edilebilir veya edilemez olmak üzere sınıflandırmak için basit denetimsiz öğrenme algoritmasından bahsetmiştir. Yorumların sınıflandırılması, yorumlardaki sıfat veya zarf içeren sözcük öbeklerinin ortalama anlamsal yönelim değeri kullanılarak gerçekleştirilmiştir. Sözcük öbeklerinin bulunması için çeşitli kurallardan yararlanılmıştır. Sözcük öbeğinin iyi yönde çağrışım yapması durumunda pozitif anlamsal yönelime sahip olduğu, kötü yönde çağrışım yapması durumunda ise negatif anlamsal yönelime sahip olduğu belirtilmiştir. Çalışmada, sözcük öbeğinin anlamsal yönelimi, verilen sözcük öbeği ile “excellent” kelimesi arasındaki PMI değerinden, sözcük öbeği ile “poor” kelimesi arasındaki PMI değerinin çıkarılması sonucu hesaplanmıştır. Sözcük öbeklerinin anlamsal yönelim değerlerinin toplamının 0’dan büyük olup olmamasına göre doküman tavsiye edilebilir veya edilemez olarak sınıflandırılmıştır. Bu algoritma ile otomobil, banka, film ve seyahat destinasyonlarına ait 4 farklı veri setinde %74’lük bir doğruluk oranı elde edilmiştir. Otomobil, banka, film ve seyahat destinasyonlarında sırasıyla %84, %80, %65,83 ve %70,53 oranlarında doğruluk elde edilmiştir.

Skuza ve Romanowski [17], Twitter verilerinden gelecekteki hisse senedi fiyatlarını farklı zaman aralıklarına (bir saat, yarım saat, 15 dakika ve 5 dakika) göre tahmin etmek için NB algoritmasının map-reduce versiyonundan yararlanan bir sistem tasarlamışlardır. İnsan değerlendirmesi sonucu etiketlenen veri seti kullanılarak tweet’lerin öznelliği ve tweet’lerin pozitiflik veya negatiflik durumunun belirlenmesi için iki sınıflandırıcı oluşturulmuştur. Hisse fiyatlarını tahmin edebilmek için sınıflandırma sonuçları geçmiş fiyatların lineer regresyonu ile birleştirilmiştir. En başarılı sonuç 5 dakikalık tahminlerde elde edilmiştir.

Mandal ve Gupta [18], duygu polaritesini hesaplamak için diğer sözlük tabanlı yaklaşımlardan farklı olarak kendi sözlüklerini oluşturmak için çeşitli web sitelerinden elde edilen veri setlerindeki kelimeleri öncelikle pozitif ve negatif olarak gruplandırmışlardır. Ardından bu iki gruptaki kelimeleri de pozitiflik, üstünlük ve en üstünlük dereceleri olmak üzere üç gruba daha ayırarak bunları veri tabanı tablosu haline getirmişlerdir. Her bir tabloya ait kelime için önceden belirlenmiş skorlar atanmıştır. Metnin anlamsal yönelim değeri hesaplanırken bu sözlük kullanılmıştır. Sözlük tabanlı yaklaşım NB ve SVM ile kıyaslanmıştır. NB ile %99, SVM ile %96 doğruluk oranı elde edilirken geliştirilen yöntemde %81 başarı elde edilmiştir.

Oğul ve Ercan [11], çalışmalarında otel yorumları içeren Türkçe veri kümesi oluşturup, MNB, RF, SVM ve sözlük tabanlı SentiTFIDF yöntemlerini kullanarak duygu analizini gerçekleştirip, performans karşılaştırması yapmışlardır. Random Forest (RF), %82 duyarlılık ile en başarılı yöntem olarak belirtilmiştir.

Çoban ve arkadaşları [19], Twitter'dan elde ettikleri Türkçe veri setini önceden belirledikleri his simgelerini kullanarak pozitif veya negatif olarak etiketlemişlerdir. Ön işleme adımlarından sonra Bag of Words (BoW) ve n-gram modellerini kullanılarak öznitelikler çıkarılmıştır. Çıkarılan özniteliklerin ağırlıklandırılması için Term Frequency (TF), Boolean ve Term Frequency – Inverse Document Frequency (TF-IDF) yöntemleri kullanılmış olup, CfsSubset algoritması ile öznitelik seçme işlemi gerçekleştirilmiştir. Sınıflandırma işlemi için ise NB ve MNB, SVM, k-NN kullanılmış ve elde edilen sonuçlar karşılaştırılmıştır. En başarılı sonucu BoW modelinde %62,48, n-gram modelinde ise %66,06 doğru sınıflandırma yüzdesi ile MNB vermiştir.

Çoban ve Özyer [20], daha önceki çalışmalarından [19] elde ettikleri Twitter veri setini Latent Dirichlet Allocation (LDA) temelli topik model uygulayarak etiketlemişlerdir. Ön işleme adımından sonra kelime torbası ve n-gram model kullanılarak öznitelik vektörü oluşturulmuş, öznitelikler ANTF kullanılarak ağırlıklandırılmıştır. SVM, k-NN, NB, MNB, ME algoritmaları ile sınıflandırma işlemi gerçekleştirilip elde edilen sonuçlar kıyaslanmıştır. BoW için en başarılı yöntem %92,50 doğru sınıflandırma başarısı ile MNB olmuştur. Bigram için %90,89 ile en başarılısı SVM olurken, trigram için %91,51 ve four-gram için %91,30 ile en başarılısı ME olmuştur. His simgeleri kullandıkları çalışmalarına [19] göre analiz başarısını yaklaşık %26 artırmışlardır.

Vural ve arkadaşları [21], film yorumlarını pozitif ve negatif olarak sınıflandırmak için sözlük tabanlı denetimsiz bir duygu analizi çerçevesi sunmuşlardır. Çalışmada kullanılan sözlük, SentiStrength kütüphanesinin Türkçeleştirilmesi sonucu oluşturulmuştur. Önerilen yaklaşım [22] tarafından oluşturulmuş veri seti üzerinde sınanmıştır ve %76 oranında doğruluk elde edilmiştir.

2.2.2. Cümle seviyesi duygu analizi

Bu analiz tipindeki doküman içindeki her bir cümlelerin anlamsal yönelimi bulunur. Cümleler kısa dokümanlar olduğu için bu analiz tipi ile doküman seviyesi arasında temel yöntemsel bir farklılık bulunmaz [1]. Cümle seviyesi duygu analizi öznellik sınıflandırma ve cümle duygu sınıflandırma aşaması olmak üzere ikiye ayrılmaktadır [15]. Öznellik sınıflandırma aşamasında verilen cümlelerin öznel veya nesnel olup olmadığına karar verilirken duygu sınıflandırma aşamasında öznel cümlelerin pozitif veya negatif görüş ifade edip etmediğine karar verilir.

Yu ve Hatzivassiloglou [23], cümlelerin öznellik sınıflandırması ve cümlelerin yönelimlerine karar vermek için bir çalışma yapmıştır. Öznellik sınıflandırması için NB, çoklu NB ve cümle benzerliklerinden yararlanmıştır. NB sınıflandırıcı kullanılarak öznellik sınıflandırılması yapıldığında yüksek duyarlılık ve kesinlik (%80-%90) değerleri elde edilmiştir. Tanımlanan her bir öznel cümlelerin duygu sınıflandırmasını yapmak için log-olabilirlik oranı kullanılmıştır.

Barbosa ve Feng [24], SVM metodundan yararlanarak Tweet'lere öznellik sınıflandırması yapan ve öznel tweet'leri de pozitif veya negatif olarak ayıran 2 adımlı bir metot önermiştir. Bu aşamalarda Twitter'a ait özellikler olan re-tweet, hashtag ve emojiler kullanılmıştır. Öznellik sınıflandırmasında %18,1 oranında hata yapılırken, polarite sınıflandırmada %18,7 oranında hata yapılmıştır.

Davidov ve arkadaşları [25], her biri basit cümle olan tweetlerde k-NN benzeri bir denetimli öğrenme yaklaşımı ile duygu sınıflandırması yaparak %86 oranında f-skor değeri elde etmiştir. Sınıflandırmada hashtag'ler, gülen suratlar, noktalama işaretleri ve sık geçen örüntüler kullanılmıştır.

Khan ve arkadaşları [26], yorumlar içindeki cümleleri öznel veya nesnel olarak sınıflandırmak için kural tabanlı alan bağımsız bir yöntem önermiştir. Her bir öznel cümlelerin polaritesi SentiWordNet kullanılarak hesaplanmıştır ve %86,6 oranından başarı elde edilmiştir.

2.2.3. Hedef temelli duygu analizi

Doküman ve cümle seviyesi duygu analizleri insanların neleri beğendiklerini ve beğenmediklerini tam olarak ortaya çıkaramaz [1]. Hedef temelli duygu analizi ile daha detaylı sonuçlar elde edilebilmektedir. Örneğin, “Restoranın manzarası ve yemekleri güzel fakat ulaşımı zor” cümlesinde “manzara” ve “yemek” açısından olumlu bir görüş ifade edilirken, “ulaşım” için negatif görüş belirtilmiştir.

Hu ve Liu [27] çalışmalarında, online ortamda satılan ürünler ile ilgili yorumlardan nitelik tabanlı özet oluşturma problemini ele almışlardır. Ürünlerin aday niteliklerini belirlemek için birliktelik kural madenciliği yöntemini kullanmışlardır ve gereksiz nitelikler elendikten sonra kalan nitelikler sık geçen nitelikler olarak dikkate alınmıştır. Sık geçen nitelikler yardımıyla duygu ifade içeren kelimeler çıkarılmıştır. Bu kelimeler ve önceden belirli bir kök liste yardımıyla duygu ifade eden kelimelerin listesi oluşturulmuştur. Bu kelimelerin zıt ve eş anlamlı kelimeleri de bu listeye eklenmiştir, listedeki kelimelerin anlamsal yönelimleri WordNet yardımıyla bulunmuştur. Birliktelik kural madenciliği ile bulunamayan nitelikler ise duygu ifade eden kelimelere yakın olan isimler bulunarak elde edilmiştir. Sonrasında, her bir görüş cümlesinin anlamsal yönelimi bulunarak nitelik bazında özet üretilmiştir. Önerilen yöntem 5 farklı elektronik ürüne ait yorumlar kullanılarak test edilmiştir. Nitelik belirleme aşamasında ortalama %72 oranında kesinlik ve %80 oranında duyarlılık elde edilmiştir.

Siqueira ve Barros [28], domain bağımsız nitelik çıkarımı için *WhatMatter* sistemini sunmuştur. Bu sistemde nitelik çıkarımı sık geçen isimlerin bulunması, ilgili isimlerin bulunması, nitelik belirteçlerinin eşleştirilmesi ve alakasız isimlerin silinmesi aşamalarından oluşmaktadır. Birinci aşamada metin içindeki isimler bulunur ve belirlenen eşik değerinin üzerindeki isimler sık geçen isimler olarak adlandırılır ve aday nitelikler olarak adlandırılır. İkinci aşamada aday niteliklerin yanındaki sıfatlar bulunur ve bu sıfatlara yakın olan isimler de aday niteliklere eklenir. Üçüncü aşamada imalı nitelikleri bulmak için alan bağımlı ve önceden tanımlanmış “fiyat-pahalı”, “fiyat-ucuz” gibi 20 adet nitelik-görüş çifti kullanılmıştır. Son aşamada alakasız isimlerin silinmesi için PMI tekniğinden yararlanılmıştır. Bu 4 aşama sırasıyla uygulandığında %77,24 oranında kesinlik, %90,94 oranında duyarlılık ve %83,54 oranında f-skor değeri elde edilmiştir.

Fei ve arkadaşları [29], imalı nitelik problemini çözmek için sözlük tabanlı bir yaklaşım önermiştir. Problem eş anlamlı, zıt anlamlı, alt anlamlı ve üst anlamlı kelimeleri dikkate alan kolektif sınıflandırma problemi olarak ele almıştır. İmalı nitelikleri sıfatlar aracılığı ile belirlemeyi amaçlamışlardır. Bunun için İngilizce sözlükler kullanılmıştır. Hui ve Liu'nun [27] oluşturduğu sıfat listesinden yola çıkılarak bu sıfatların İngilizce sözlüklerdeki örnek cümleleri bulunmuş ve POS etiketleyici yardımı ile her bir sıfat için bu sıfatın tanımlarında geçen isimler çıkarılarak aday nitelikler olarak adlandırılmıştır. Kolektif sınıflandırma kullanılarak her bir sıfat için ilgili olan ve olmayan nitelikler bulunmuştur. Test sonuçları neticesinde %76,3 oranında kesinlik, %77,3 değerinde duyarlılık ve %76,8 değerinde f-skor elde edilmiştir.

Qiu ve arkadaşları [30], düşünce sözlüğünü genişletme ve düşünce hedefi çıkarımı problemlerini ele almışlardır. Başlangıç sözlüğü olarak Hu ve Liu [27] tarafından oluşturulan sözlük kullanılmıştır. Görüş bildiren kelimeler ile hedefler arasındaki bazı söz dizimsel ilişkiler bağımlılık ayrıştırıcı yardımıyla bulunup, bu ilişkilerden başlangıçtaki düşünce sözlüğünü genişletmek ve yeni düşünce hedefleri çıkarmak için yararlanılmıştır. Çift yayılım yaklaşımı ile başlangıçtaki düşünce sözlüğü kullanılarak diğer yeni görüş kelimeleri ve hedefleri bulunmuştur. Bu döngü yeni hedef ve görüş kelimeleri bulunamayana kadar devam etmektedir. Oluşturulan sözlükteki kelimelerin yönelimi kural tabanlı yaklaşım ile hesaplanmıştır. Önerilen yöntemle %88 kesinlik, %83 duyarlılık, %86 f-skor elde edilmiştir.

Yauris ve Khodra [31], çalışmalarında oyun yorumlarından hedef temelli özetlemeyi amaçlamışlardır. Hedef ve görüş çiftlerini çıkarmak, hedefleri kategorize etmek ve bunları özetlemek için Qui ve arkadaşlarının [30] geliştirdiği çift yayılım yaklaşımından yararlanmışlardır ve %51,36 oranında f-skor elde etmişlerdir.

Bhoir ve Kolte [13], çalışmalarında filmlerin beğenilen veya beğenilmeyen özelliklerinin kolayca bulunabilmesi için film yorumlarını hedef seviyesinde özetlemeyi amaçlamışlardır. Önerilen sistemin öznellik analizini gerçekleştirmek için NB ve SentiWordNet'ten yararlanılmıştır. Öznellik analizi gerçekleştirildikten sonra kural tabanlı yaklaşım ile nitelik görüş çiftleri bulunmuş ve bu çiftler kullanılarak hedef temelli duygu analizi gerçekleştirilmiştir. Skorlama için SentiWordNet'ten ve kendi oluşturdukları pozitif ve negatif kelimeleri içeren sözlükten yararlanılmıştır. Çalışmanın sonunda öznellik analizinde kullandıkları yöntemleri karşılaştırmışlardır. NB ile %75 oranında f-skor elde edilirken

SentiWordNet ile %53,33 oranında f-skör elde edilmiştir. 5 farklı filme ait veri seti üzerinde nitelik görüş çifti elde etmede en düşük %62,23, en yüksek ise %75,38 f-skör değerleri elde edilmiştir.

Liu ve arkadaşları [32], hedef nitelik çıkarımı için otomatik kural seçme yöntemini önermişlerdir. Bu yöntem, görüş kelimeleri ve hedefler arasındaki dilbilgisi kurallarından yararlanılarak denetimsiz ve alan bağımsız çalışabilecek şekilde geliştirilmiştir. Deney sonuçları önerilen yöntemin hem tüm kural setinden hem de Conditional Random Field (CRF) tabanlı yöntemden daha iyi sonuç elde ettiğini göstermiştir.

Kama ve arkadaşları [33], Türkçe metinlerde hedef temelli duygu analizi için alan bağımsız ve denetimsiz nitelik çıkarma metodu önermiştir. Bu metot ile arama motoru kullanarak sıklık tabanlı nitelik çıkarımının performansını artırmak amaçlanmıştır. İlk olarak metindeki isim ve isim grupları bulunur, belirli bir değeri geçenler bir listeye alınır. Bu listedeki her bir eleman ve ilgili alan için Web üzerinde *Yandex.XML* Application Programming Interface (API) kullanılarak sorgulama yapılır ve eşik değerini geçen elamanlar önerilen nitelikler olarak adlandırılır. Web üzerinde sık geçmeyen isim ve isim grupları için her bir gruptaki kelimelerden biri nitelik listesinde varsa nitelik listesine eklenir ve hedef çıkarımı tamamlanır. %59,24 oranında kesinlik, %84,90 oranında duyarlılık ve %69,79 oranından f-skör elde edilmiştir.

Kama ve arkadaşları, diğer çalışmalarında [34] imalı nitelikleri ve polaritesi hedefe bağlı olan görüş kelimelerini sürece dahil ederek hedef görüş eşleştirmesinin performansını artırmışlardır. Yalnızca imalı olmayan nitelikler kullanılarak yapılan eşleştirme işleminde %84,95 oranında f-skör elde edilirken, imalı nitelikler de sürece dahil edildiğinde %85,36 oranında f-skör elde edilmiştir. Son adımda polaritesi hedef terime göre değişen ifadeler de bu sürece eklendiğinde f-skör değeri %86,75'e yükselmiştir.

Ekinci ve arkadaşları [35], çalışmalarında n-gram model, sonlu durum makinesi ve PMI kullanarak Türkçe kullanıcı yorumlarından çok kelimeli hedefleri çıkarmayı amaçlamışlardır. N-gram model ile aday hedefler bulunduktan sonra frekans değeri 3'ün altında kalanlar aday hedeflerden çıkarılmıştır. Türkçe dil bilgisi kuralları kullanılarak oluşturulan sonlu durum makinesi ve PMI ölçütü kullanılarak eliminasyon işlemi

sürdürülmüş ve çok kelimeli hedefler elde edilmiştir. Cep telefonlarına ait veri setinde %82, otel ait veri setinde ise %79 kesinlik değeri elde edilmiştir.

2.3. Sözcük Tabanlı Duygu Sınıflandırma

Sözcük tabanlı yaklaşım, bilinen ve önceden derlenmiş referans duygu terimlerinin bir koleksiyonu olan duygu sözlüğüne dayanır. Görüş ifade eden kelimeler duygu sınıflandırma görevinde kullanılmaktadır. Pozitif duygu kelimeleri istenen durumları ifade ederken, negatif duygu kelimeleri istenilmeyen durumları ifade eder. Bunlar aynı zamanda sözcük öbekleri ve deyimler şeklinde olabilir. Bu kelime listeleri manuel, derlem tabanlı ve sözlük tabanlı yaklaşımlar kullanılarak oluşturulabilir. İnsan gücü ile yani manuel olarak bu listeleri oluşturmak zaman alıcı bir teknik olup tek başına kullanılmaz. Diğer iki yöntem aracılığıyla otomatik olarak oluşturulan listelerdeki hataları gidermek için kullanılır [36]. Bu kısımda sözlük ve derlem tabanlı yaklaşımlar detaylandırılmıştır.

2.3.1. Sözlük tabanlı yaklaşım

Bu yaklaşımda polaritesi bilinen duygu kelimeleri ile bir kök liste oluşturulur. Sözlükler kullanılarak bu listedeki her bir kelimenin eş anlamlı ve zıt anlamlı karşılıkları bulunur. Bulunan kelimeler kök listeye aktarılır. Bu işlem, döngüsel biçimde yeni kelime bulunmayıncaya kadar devam eder, sonrasında hatalar düzeltilir ya da silinir [36].

Wordnet [37], İngilizce isim, fiil, sıfat ve zarfları ve bunların eş anlamlı kelimelerini içeren elektronik veri setidir. SentiWordNet [38], duygu analizi için kullanılan sözcüksel bir kaynaktır. WordNet'in her bir anlamdaş kümesi için pozitif, negatif ve nötr puanı tanımlanmıştır.

Thelwall ve arkadaşları [39], görüş ifade eden kelimeleri, görüş kelimelerinin anlamını artıran veya azaltan kelimeleri, olumsuzluk eklerini, emojileri ve deyimleri kullanarak SentiStrength duygu sözlüğünü oluşturmuştur.

Bilgin ve arkadaşları [40], Balkanet projesi kapsamında WordNet'teki eş anlamlı kümeleri çevirerek Turkish WordNet'i geliştirmiştir. Dehkharghani ve arkadaşları [41], Turkish WordNet'te bulunan her eş anlamlı kümeye üç kutup skoru atayarak, literatürdeki ilk Türkçe kutupluluk kaynağı olan SentiTurkNet'i oluşturmuştur.

Ucan ve arkadaşları [42], İngilizce kaynaklar kullanarak yeni diller için kısa zamanda duygu sözlüğü inşa etmek amacıyla otomatik çeviriye dayanan bir yaklaşım önermiştir. Çalışmalarının sonunda Türkçe için 3 farklı sözlük elde etmişlerdir. Oluşturulan sözlüklerle cümle seviyesi sınıflandırma yapıldığında önerilen TSDp yöntemiyle elde edilen sözlükle en iyi sonucu almışlardır, film veri seti için %70,35, otel veri için %80,68 başarı elde etmişlerdir.

2.3.2. Derlem tabanlı yaklaşım

Bu yaklaşım ile derlem içindeki söz dizimsel kalıplar kullanılarak belirli bir alandaki duygu kelimelerini bulma probleminin ortadan kaldırılması amaçlanır.

Ding ve arkadaşları [43], cümle içindeki ve cümleler arasındaki bağlaçları kullanan kural tabanlı bir yaklaşım ile hedef – görüş kelimeleri arasındaki ilişkileri dikkate alarak duygu sözlüğü oluşturmuştur. Bu yaklaşım ile görüş ifade eden kelimenin farklı hedef terimler ile olan anlamsal yönelimi hesaba katılmıştır.

Parkhe ve Biswas [14], çalışmalarında film yorumlarından duygu analizi gerçekleştirmek için türe (aksiyon, komedi, drama, suç ve korku) özel sözlük oluşturmuşlardır. Bunun için günlük hayatta sık kullanılan 500 sıfat içeren liste oluşturulmuştur. Her bir tür ile bu 500 kelimelik listede bulunan her bir sıfat arasındaki PMI değeri hesaplanmıştır. Değerler -1 ile 1 arasında normalize edilmiştir. Metinden elde edilen sıfatların skorları bu sözlük kullanılarak bulunmuş ve sözlükte olmayan sıfatların skorları ise SentiWordNet'ten elde edilmiştir.

Han ve arkadaşları [44], etiketlenmemiş veri seti kullanarak alana özgü duygu sözlüğü oluşturmuşlardır. SentiWordNet temelli duygu sınıflandırıcı ile yorumlar pozitif veya negatif olarak sınıflandırıldıktan her bir yorumdaki isim, sıfat, zarf ve fiil olan terimler bulunur ve sözlükte yoksa sözlüğe eklenir. Bulunan terimlerin pozitif ve negatif yorumda geçme sayıları hesaplandıktan sonra MI kullanılarak bu terimlerin pozitiflik ve negatiflik skorları belirlenmiştir.

Feng ve arkadaşları [45], alan bazlı duygu sözlüğü oluşturmak için mobil alışveriş yorumlarını kullanarak duygu kelimeleri ve ürün nitelikleri arasındaki ilişkiyi dikkate alan

otomatik bir yaklaşıml önermiştir. Öncelikle duygu kelimeleri ile ürün nitelikleri arasındaki ilişki PMI algoritması ile elde edilir. Sonrasında duygu kelimelerinin mutluluk, üzüntü, korku, sürpriz, öfke ve kıskançlık kategorilerinden hangilerine ait olduğu bulunmuştur.

Yapılan incelemeler sonucunda Türkçe için gerçekleştirilen duygu analizi çalışmalarının mevcut olduğu gözlemlenmiştir. Bu çalışmaların çoğunda [11, 19 - 22] doküman temelli duygu analizi üzerine odaklanıldığı dikkat çekmiştir. Son zamanlarda ise hedef temelli duygu analizi çalışmaları gerçekleştirilmiştir [33 - 35] fakat sayıca az miktardadır. Bununla birlikte hedef temelli çalışmalarda kullanılabilecek veri seti eksikliği de gözlemlenmiştir. Gerçekleştirilen bu çalışmada bahsedilen bu eksikliklerin giderilmesi amaçlanmıştır. Ayrıca bu çalışma C-value yaklaşımını kullanan Türkçe dili için yapılmış bilinen ilk çalışma olma özelliğine sahiptir.

3. BİLİMSEL ARKA PLAN

Bu bölümde, çalışma kapsamında temel alınan metotlar hakkında bilgi verilmektedir.

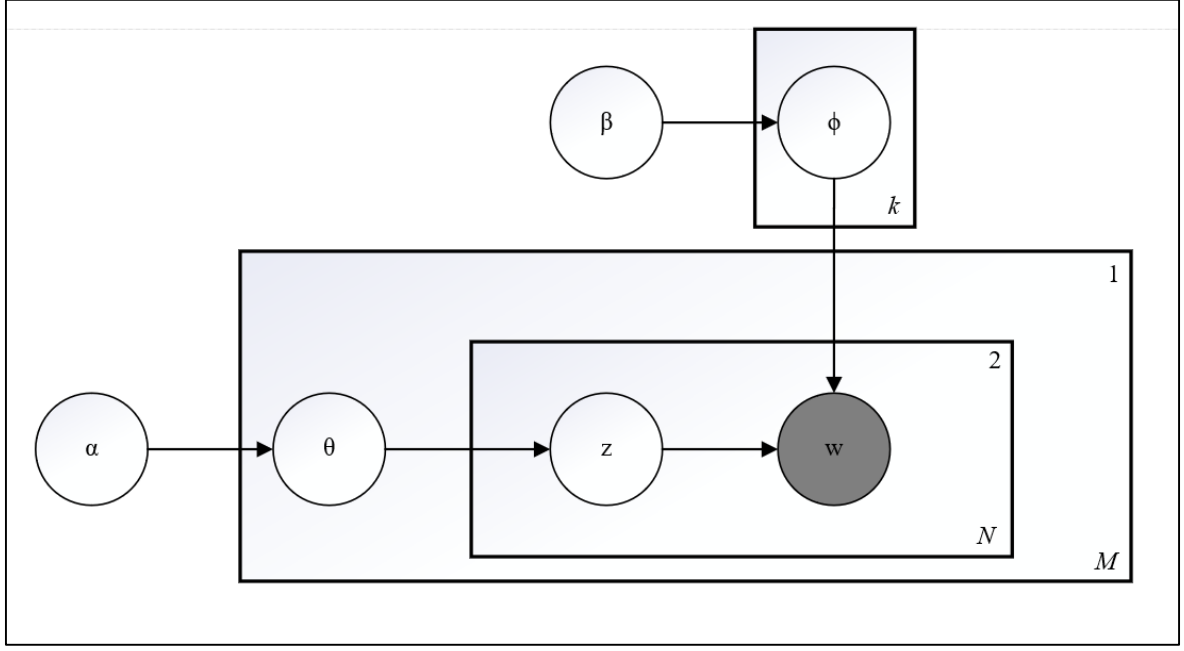
3.1. Latent Dirichlet Allocation (LDA)

LDA, üç seviyeli bir hiyerarşik Bayes modeli olan ayrık verilerin toplanması için üretici bir olasılıksal modeldir. LDA, dokümanların gizli konular üzerinde rastgele karışımlar olarak temsil edilirken her bir konunun kelimeler üzerine dağılımı ile karakterize edilmesine dayanır [46].

LDA kelime olasılıklarına göre konuları temsil eder. Her konudaki en yüksek olasılığa sahip kelimeler konunun ne anlama gelebileceği hakkında iyi bir fikir verir [47]. Bir derlemi modellemek için denetlenmemiş bir üretici olasılık yöntemi olan LDA, en sık kullanılan konu modelleme yöntemidir. LDA, her belgenin gizli konular üzerinde olasılıklı bir dağılım olarak temsil edilebileceğini ve tüm belgelerdeki konu dağıtımının ortak bir Dirichlet'i paylaştığını varsayar. LDA modelindeki her bir gizli konu, aynı zamanda kelimeler üzerinde olasılıksal bir dağılım olarak temsil edilir ve konuların kelime dağılımları da ortak bir Dirichlet'i paylaşır [47].

M doküman sayısını, N her bir dokümandaki kelime sayısını, k konu sayısını, z konuyu, w ise kelimeyi ifade etmektedir. Bir doküman N adet kelime dizisi olarak tanımlanır ve $d = (w_1, w_2, \dots, w_N)$ olarak ifade edilir. Derlem ise M adet doküman topluluğundan oluşur ve $C = \{d_1, d_2, \dots, d_M\}$ olarak ifade edilir. Şekil 3.1.'de LDA plaka notasyonu ile ifade edilmiştir. Bu notasyonda numaralandırılmış plakalar tekrarlı döngüleri temsil etmektedir. 1 numaralı plaka dokümanları, 2 numaralı plaka bir doküman içindeki tekrarlanan konu ve kelimelerin seçimini temsil eder.

- α parametresi kullanılarak her bir doküman için çok terimli θ_d dağılımı seçilir,
- β parametresi kullanılarak her bir konu için çok terimli Φ_z dağılımı seçilir,
- Her bir doküman içindeki kelimeler için:
 - ❖ Çok terimli θ_d dağılımı kullanılarak konu (z_{dn}) seçilir,
 - ❖ Çok terimli $\Phi_{z_{dn}}$ dağılımı kullanılarak kelimeler (w_{dn}) seçilir.



Şekil 3.1. LDA'nın grafiksel model gösterimi

Derlemin olasılığı Eş. 3.1'de verilen formül ile hesaplanır.

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d \quad (3.1)$$

α ve β parametreleri derlem seviyesinde parametrelerdir, derlem oluşturma süresince bir kez örneklendikleri varsayılmaktadır. θ_d doküman seviyesinde değişkendir, her doküman için örneklenir. z_{dn} ve w_{dn} ise kelime seviyesi değişkenlerdir ve her bir dokümandaki her bir kelime için örneklenir.

LDA veya LDA'yı temel alan yöntemlerin hedef temelli duygu analizin hedef terim çıkarma aşamasında kullanıldığı çalışmalar mevcuttur [48 - 54]. Titov ve McDonald [48], LDA gibi konu modellerinin ilgili hedefler yerine verilerdeki genel konuları ele alma eğiliminde oldukları için hedefleri tespit etme görevine uygun olmadığını belirtmiştir ve bu sorunu çözmek için (MG-LDA) modelini önermişlerdir. Brody ve Elhadad [49], LDA yöntemini kullanarak hedef terimleri belirledikten sonra hedefe ait görüşleri bulmak için sıfatlardan yararlanmıştır. Zhao ve arkadaşları [50], hedef terimleri ve hedef terime özel duygu kelimelerini keşfetmek için söz dizimsel özelliklerden yararlanan ME ve LDA'yı temel alan MaxEnt-LDA hibrid modelini önermiştir. Bagheri ve arkadaşları [51], yorumlardaki hedefleri saptamak için LDA'nın genişletilmiş hali olan ADM-LDA isimli modeli önermiştir. Bu model LDA'nın aksine dokümanları BoW olarak temsil etmez. ADM-LDA,

bir belgedeki sözcüklerin bir Markov zincirindeki konularının ve daha sonraki kelimelerin aynı konuya sahip olma ihtimalinin daha yüksek olduğunu varsaymaktadır. Poria ve arkadaşları [52], LDA'daki söz dizimsel yaklaşım yerine anlamsal yaklaşımı kullanan Sentic-LDA'yı önermiştir. Shams ve Baraani-Dastjerdi [53], hedef çıkarma için önceki alan bilgisinin LDA konu modeline dahil edilmesine dayanan bir yöntem önermiştir. Önerilen yöntemde, LDA konu modeli kelimelerin birlikte oluşması ile birleştirilmiştir. Bu şekilde, her bir döngüde eş-oluşumuna dayanan yinelemeli bir algorithmada, benzer yönlerden önceki bilgiler çıkarılır ve bu bilgi, bilgi çifti setlerinde LDA modeline eklenir. Ekinci ve Omurca [54], Türkçe yorumlardan hedef – görüş çiftlerini çıkarmak için LDA kullanmıştır. Çok kelimeli terimleri de bu sürece dahil etmek için Babelfy¹ kullanılmıştır.

Bu çalışma kapsamında LDA konu modelleme yöntemi hedef terim çıkarma aşamasında kullanılmıştır.

3.2. Pointwise Mutual Information (PMI)

İki terim arasındaki istatistiksel bağımlılık derecesini ölçmek için kullanılır [55]. Eş. 3.2'de verilen formül ile hesaplanır.

$$PMI(word_1, word_2) = \log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)} \quad (3.2)$$

$P(word_1, word_2)$ $word_1$ ve $word_2$ beraber bulunma olasılığını ifade ederken, $P(word_1)P(word_2)$ iki terimin istatistiksel olarak bağımsız olduğunda beraber bulunma olasılığını ifade eder.

Turney, PMI skorunu hesaplamak Eş. 3.3 ve Eş. 3.4 kullanmıştır [16].

$$P(word) \equiv hits(word) \quad (3.3)$$

$$P(word_1, word_2) \equiv hits(near(word_1, word_2)) \quad (3.4)$$

¹ <http://babelfy.org/>

$hits(X)$ fonksiyonu verilen X kelimesinin geme sayısını, $near(Y, Z)$ ise Y ile bařlayıp Z ile biten, aralarında belirli sayıda kelime olan kelime gruplarını ifade eder.

PMI teknięi hedef terimlerin kategorilerin belirlenmesi ařamasında kullanılmaktadır.

3.3. WSBFE

Kama ve arkadaşları, Trke metinlerde hedef temelli duygu analizi iin alan baęımsız bir yntem nermiřlerdir [33]. Bu yntemle, arama motoru kullanılarak frekans bazlı nitelik ıkarımının performansının artırılması amalanmıřtır. Bu yntem  ařamadan oluřmaktadır.

Ařama 1

- Metindeki isim ve isim grupları bulunur ve belirli bir eřik deęerinin zerinde olanlar aday nitelikler listesine eklenir.
- Listedeki her bir eleman iin arama sorgusu oluřturulur ve Yandex arama motorunun API'si kullanılarak her bir sorgunun geme sayısı bulunur ve depolanır.

Ařama 2

- Arama motorundan dnen sonulara gre isim ve isim grupları sıralanır ve eřik deęerinin zerinde olanlar hedef terim olarak kabul edilir.

Ařama 3

- Hedef terim listesine dahil olmayan isim grupları tespit edilir. Her bir isim grubunda bulunan kelimelerden en azından biri hedef terim listesindeyse ilgili isim grubu da hedef terim listesine eklenir.

Bu yntem ile bulunan isim grupları LDA modeline girdi olarak verilen veri setini gncellemek iin kullanılır. LDA'nın bulduęu hedef terimlerin geerlemesi bu yntemle yapılmaktadır.

3.4. C-value

C-value, dilbilimsel ve istatistiksel bilgileri birleřtirerek i ie gemiř terim ıkarma iřlemini geliřtirmeyi amalayan ok kelimeli hedef terim tanıma iin kullanılan alan baęımsız bir

metottur [56]. Bu metot, çeşitli filtreleme kurallarından yararlanarak veri setinden aday çok kelimeli terimlerin çıkartılmasını amaçlar.

C – value

$$= \begin{cases} \log_2 |a| \cdot f(a), & a \text{ iç içe geçmiş terim değil ise} \\ \log_2 |a| \cdot (f(a) - \frac{1}{P(T_a)} \sum_{b \in T_a} f(b)), & a \text{ iç içe geçmiş terim ise} \end{cases} \quad (3.5)$$

Eş. 3.5'te a aday ifadeyi, $|a|$ a ifadesindeki kelime sayısını, $f(a)$ a 'nın veri setindeki frekansını, T_a , a ifadesini içeren daha uzun kelime sayısına sahip aday ifadeleri, $P(T_a)$ farklı T_a sayısını, $\sum_{b \in T_a} f(b)$ a 'nın geçtiği daha uzun kelime sayısına sahip aday ifadelerin toplam frekansını ifade etmektedir.

C-value dil bilimsel ve istatistiksel kısımlardan oluşur. Birinci kısım etiketleme, dil bilimsel filtreleme ve durak kelimelerden (stopwords) arındırma aşamalarını içerir. Etiketleme aşamasında derlemdeki her bir kelime isim, sıfat, zarf, fiil vb. olmak üzere etiketlenir. Dilbilimsel filtreleme ile istenmeyen ifadelerin bulunmasının önüne geçilmektedir. Bulunan çok kelimeli terimler içindeki herhangi bir kelime bile durak kelime listesi içinde bulunuyorsa kesinlik değerinin artırılması amacıyla o terim silinir. İkinci kısımda ise her bir aday terimin C-value değerini hesaplanır ve terimler bu değere göre sıralanır. C-value değeri belirli eşik değerinin üzerinde olanlar hedef terim olarak kabul edilir. C-value hesaplanırken aday terimin frekansı, aday terimin daha uzun terimlerin içerisinde bulunma sıklığı, aday kelimeyi içeren daha uzun aday terimlerin sayısı, aday terimdeki kelime sayısı özellikleri dikkate alınır.

C-value orijinal olarak İngilizce yazılmış dokümanlardaki çok kelimeli terimleri bulmak için önerilmiş bir yaklaşımdır. Barrón ve arkadaşları İspanyolca için otomatik terim çıkarımı için C-value metodundan yararlanan bir yaklaşım önermiştir [57]. Çalışmalarında, çok kelimeli terimleri keşfedebilmek için orijinal C-value hesaplanmasına yenilik getirmişlerdir. C-value yönteminin tek kelimeli terimleri bulamama kısıtını ortadan kaldırmak için $c = i + \log_2 |a|$ eşitliği kullanılmıştır (Bknz. Eş. 3.6). Başlangıçta, formül sonucundaki değişimi minimize etmek için i değeri olarak 0,1 seçilmiştir. Deneyler sonucunda $i = 1$ değerinde en iyi sonucu verdiği gözlemlenmiştir. Orijinal metotta aday terimdeki herhangi bir kelime durak kelime listesinde bulununca o aday terim silinmektedir. Bu çalışmada ise aday terim yerine durak

kelime listesinde bulunan kelimenin silinmesi (selective stopwords deletion) yaklaşımı benimsenmiştir. Benzer hedef terimlerin bulunmasının önüne geçmek için derlemdeki kelimelerin çekim ekleri kaldırılmıştır. Bilgi kaybının az olması adına herhangi bir eşik değeri kullanılmamıştır.

$C - value$

$$= \begin{cases} c \times f(a), & a \text{ i\c{c}e ge\c{c}mi\c{s} terim de\c{g}il ise} \\ c \times (f(a) - \frac{1}{P(T_a)} \sum_{b \in T_a} f(b)), & a \text{ i\c{c}e ge\c{c}mi\c{s} terim ise} \end{cases} \quad (3.6)$$

Ventura ve arkadaşları [58] İngilizce ve Fransızca serbest metinlerden biyomedikal terimleri çıkarmak ve sıralamak için c-value ile keyword extraction metotlarından yararlanan bir yaklaşım geliştirmiştir. Okapi BM25 ve TF-IDF anahtar kelime çıkarımı için kullanılmıştır. C-value ile hem tek kelimeli hem de çok kelimeli terimleri bulabilmek için logaritma değeri hesaplanmadan önce 1 ekleyerek tek kelimeli terimler için sıfır (0) katsayı probleminin ortadan kaldırılmasını amaçlanmıştır (Bknz. Eş. 3.7). Durak kelimelerden arınma yaklaşımı ve eşik değeri kullanılmamıştır.

$C - value$

$$= \begin{cases} \log_2 |a + 1| \cdot f(a), & a \text{ i\c{c}e ge\c{c}mi\c{s} terim de\c{g}il ise} \\ \log_2 |a + 1| \cdot (f(a) - \frac{1}{P(T_a)} \sum_{b \in T_a} f(b)), & a \text{ i\c{c}e ge\c{c}mi\c{s} terim ise} \end{cases} \quad (3.7)$$

Çizelge 3.1. C-value örnek sonuçlar

Aday Terim	Kelime Sayısı	Frekans	$P(T_a)$	$\sum_{b \in T_a} f(b)$	C-value
akşam yemek seçeneği	3	1	0	0	2
yemek çeşitliliği	2	2	0	0	3,17
öğle yemeği	2	4	1	1	4,75
yemek kalitesi	2	4	0	0	6,34
akşam yemeği	2	8	4	4	11,094
yemek	1	339	48	70	337,54

Bu çalışmada, tek kelimeli terimler için C-value hesabının yapılabilmesi adına Ventura ve arkadaşlarının [58] önerdikleri metottan yararlanılmıştır. Çizelge 3.1.'de deneysel çalışmalarda kullanılan veri setinden elde edilen C-value sonuçları örnek olarak verilmektedir.

Bu yöntem ile bulunan isim grupları LDA modeline girdi olarak verilen veri setini güncellemek için kullanılır. LDA'nın bulduğu hedef terimlerin geçerlenmesi de bu yöntemle yapılmaktadır.

3.5. Bag of Words ve N-gram model

Kelime torbası yöntemi doküman sınıflandırma sürecini basitleştirmek için kullanılır. Bu yöntemde metin içindeki kelimelerin birbirleri ile olan ilişkileri dikkate alınmadan, sadece kelimelerin kullanım sıklığı dikkate alınır. LDA modeli kelime torbası yönteminden faydalanır.

N-gram ise bir metin veya karakter dizisi içinde bulunan, ardışık olarak sıralanmış, n uzunluktaki alt kümedir. İsim gruplarını ve görüş içeren sıfat ve fiil gruplarını bulmak için kullanılmıştır.

```
@org.junit.Test
public void didYouMean() {
    String[] queryArr = new String[]{"tmm otel odasi duzensizdi"};
    GroupByCriteria groupByCriteria = new GroupByCriteria(GroupByMode.FLAT, 10, 3);
    SortByCriteria sortByCriteria = new SortByCriteria(SortType.RELEVANCE, OrderType.DESC);

    YandexCriteriaBuilder builder = YandexCriteriaBuilder.create()
        .setCredentials(username, password)
        .setLanguage(Long.valueOf(983))
        .setLocalization("tr")
        .setMaxPassages(5)
        .setPageNumber(0)
        .setGroupByCriteria(groupByCriteria)
        .setFilterCriteria(FilterCriteria.FAMILY)
        .setSortByCriteria(sortByCriteria);
    YandexCriteria criteria = builder.build();
    YandexXML xml = new YandexXML(criteria);
    for (String query : queryArr) {
        try {
            xml.executeQuery(query);
            String correctQuery = xml.getCorrectQuery();
            System.out.println(correctQuery);
        } catch (UnsupportedEncodingException | MalformedURLException | YandexException ex) {
            Logger.getLogger(YandexXMLTest.class.getName()).log(Level.SEVERE, null, ex);
        }
    }
}
```

Şekil 3.2. Yandex.XML örnek kullanımı

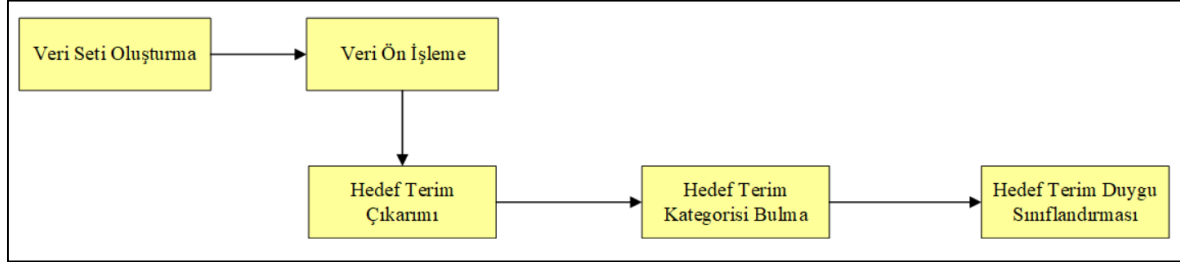
3.6. Yandex.XML

Yandex arama motorunun sunduğu bu API, kullanıcıların web ara yüzünden Yandex arama motoruna sorgu göndermelerine ve XML formatında sonuç almalarına izin veren bir hizmettir. Çalışma kapsamında, Kama ve arkadaşlarının [33] gerçekleştirdiği gibi web arama sorgusunun oluşturulmasında ve metin düzeltme işlemlerinde kullanılmıştır.

Kullanım kolaylığı sağlaması açısından Java programlama dili ile web ara yüzündeki işlevselliğe sahip bir program geliştirilmiştir. Şekil 3.2.'de programın kullanımına ait kod parçası verilmektedir.

4. PROBLEM TANIMI VE ÖNERİLEN METOT

Bu kısımda veri toplama ve ön işleme, hedef terim çıkarımı, hedef terim kategorisinin bulunması ve hedef terim duygu sınıflandırılması açıklanmaktadır (Bkz. Şekil 4.1.).



Şekil 4.1. Hedef temelli duygu analizi için önerilen yöntem

4.1. Veri Seti Oluşturma

Bu aşamada deneylerde kullanılacak veri seti elde edilmiştir. Bu amaçla TripAdvisor¹ sitesinden jsoup² kütüphanesi kullanılarak 5 farklı otele ait kullanıcı yorumları otomatik olarak toplanmıştır ve XML formatında kaydedilmiştir. Veri setinde 1691 yorum ve 8233 cümle bulunmaktadır. EK-1.'de veri setinden örnek bir kısım verilmektedir.

Her bir yorumunun polaritesi kullanıcı yorumlarındaki puanlara göre belirlenmiştir. Kullanıcı puanı 1 ve 2 ise negatif, 3 ise nötr, 4 veya 5 ise pozitif kabul edilmiştir. Yorumlar içindeki her bir cümlenin polaritesi, cümlelerdeki hedefler ve bu hedeflere ait görüşler kişisel değerlendirmeler sonucu tanımlanmıştır. Hedefler fiyat, hizmet, içecek, konum, ortam, restoran, yemek, yönetim ve genel olmak üzere 8'e ayrılmıştır. Kategorilerin bir kısmının da alt kategorileri bulunmaktadır (Bkz. Çizelge 4.1.).

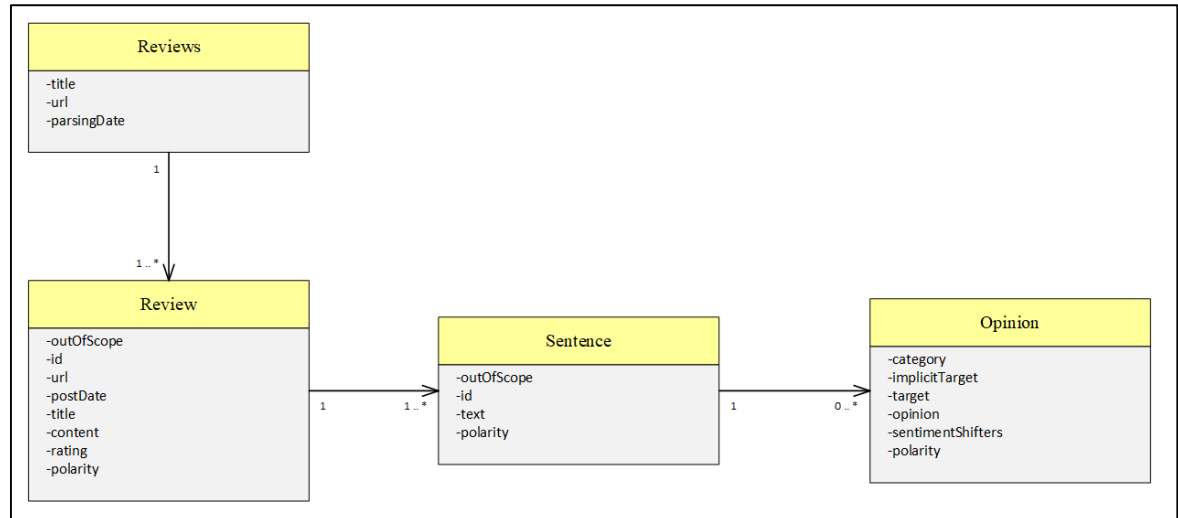
¹ <https://www.tripadvisor.com>

² <https://jsoup.org/>

Çizelge 4.1. Hedef terim kategorileri

Kategori	Alt Kategori
Fiyat	-
Hizmet	Çalışan, Kalite, Olanak, Temizlik
İçecek	Çeşitlilik, Fiyat, Kalite, Sunum
Konum	-
Ortam	-
Restoran	Fiyat, Çeşitlilik
Yemek	Çeşitlilik, Fiyat, Kalite, Sunum
Yönetim	Çalışan

Veri setini temsil eden sınıf diyagramı Şekil 4.2.'de sunulmaktadır. Model içerisinde yer alan “0..*” ifadesi sıfır veya daha fazla anlamına, “1..*” ifadesi ise bir veya daha fazla anlamına gelmektedir. Alan modeli sözel olarak ifade edilirse, her eleştiri grubu bir ya da daha fazla pozitif ya da negatif bir eleştiriden oluşurken, her eleştiri sıfır ya da daha fazla görüş içeren bir ya da daha fazla cümleden oluşmaktadır.



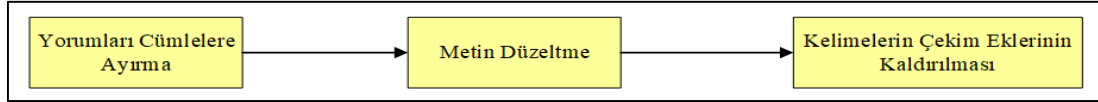
Şekil 4.2. Veri setinin sınıf diyagramı

4.2. Veri Ön İşleme

Veri seti cümlelere ayrıldıktan sonra yazım hataları doğal dil işleme kütüphanesi olan zemberek¹ ve Yandex.XML² yardımıyla giderilmiştir. Bu işlemin ardından kelimelerin çekim ekleri kaldırılmıştır (Bkz. Şekil 4.3.).

¹ <https://github.com/ahmetaa/zemberek-nlp>

² <https://xml.yandex.com.tr>



Şekil 4.3. Ön işleme aşaması

4.2.1. Cümlelere ayırma

Yorumları cümlelere ayırmak için doğal dil işleme kütüphanesi kullanılmıştır. Doğal dil işleme kütüphanesi cümlelere ayırırken his simgelerini (“:D”, “:”), “:(” de dikkate aldığı için bu aşamada ekstra bir işlem yapılmamıştır.

4.2.2. Metin düzeltme

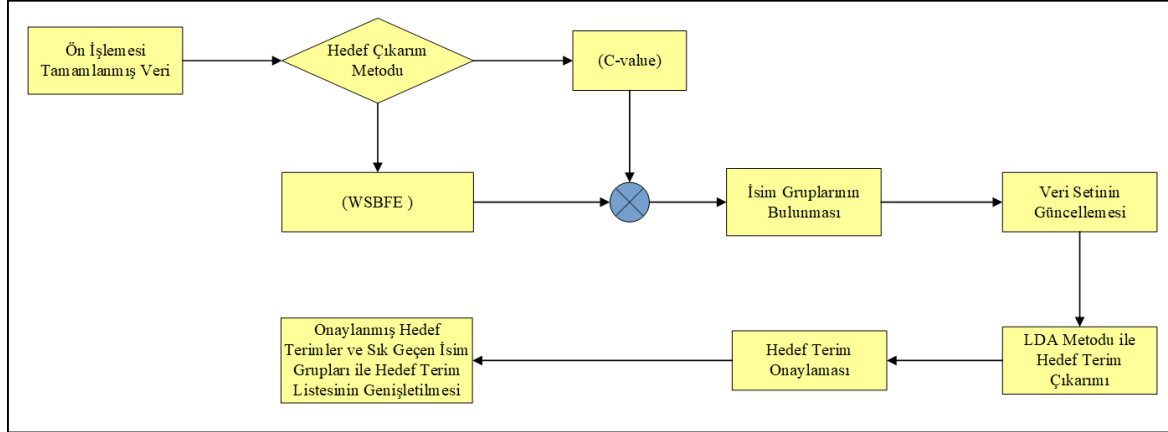
İnternet üzerinde yorum yazılırken resmi olmayan bir dil kullanıldığı için yazım hatalarının giderilmesi gerekmektedir. Bu hataların büyük bölümünü, Türkçe karakter problemi ve eksik harf problemi kaynaklı hatalar oluşturmaktadır. Yazım hatalarının giderilmesi için iki aşamalı bir sistem geliştirilmiştir. Öncelikle doğal dil işleme kütüphanesi yardımıyla her bir kelime için Türkçe karakter kontrolü yapılır ve eğer bir hata ile karşılaşılırsa düzeltme yapılır. Ardından her bir cümle için web arama sorgusu oluşturulur ve Yandex.XML kullanılarak, Yandex arama motoru üzerinde arama işlemi gerçekleştirilir. Eğer arama sonucunda, arama sorgusu düzeltilirse cümle arama motorunun önerdiği sorgu ile değiştirilmektedir.

Metin düzeltme işlemi tamamlandıktan sonra veri setindeki her bir kelime küçük harfe dönüştürülür ve kelimelerin çekim ekleri temizlenir. Kelimelerin köklerinin bulunması yerine yalnızca çekim eklerinin kaldırılmasının temel sebebi, kelimelerin anlamlarını kaybetmesini önlemektir.

4.3. Hedef Terimlerin Çıkarımı

Türkçe veri seti için WSBFE ve C-value tekniklerinden yararlanan LDA temelli iki farklı hedef çıkarım metodu önerilmiştir. Bu iki metottaki temel düşünce LDA metodundaki BoW problemini ortadan kaldırmak üzerinedir. WSBFE ve C-value metotları ile bulunan isim grupları, veri seti üzerinde değiştirilmiştir. Örneğin, “otel personel” isim grubu bulunduktan sonra veri setindeki “otel” ve “personel” ardışık ikilileri “otel_personel” haline getirilir. Veri

seti değiştirildikten sonra LDA ile hedefler bulunur. LDA ile bulunan hedefler WSBFE ve C-value teknikleri ile doğrulanır. Veri setindeki sık geçen isim grupları ile onaylanan hedefler kullanılarak hedef listesi genişletilir. İlgili akış Şekil 4.4.'te, akışa karşılık gelen algoritma Şekil 4.5.'te verilmektedir.



Şekil 4.4. Hedef terim çıkarımı akış diyagramı

```

procedure HEDEF_TERIM_BUL(doküman, metot_adı)
  metota_ait_hedef_terim_listesi = {}
  genişletilmiş_hedef_terim_listesi = {}
  doğrulanmış_hedef_terim_listesi = {}
  metota_ait_isim_grupları = {}
  sık_geçen_isim_grupları = {}
  if metot_adı == 'WSBFE' then
    metota_ait_isim_grupları ← WSBFE_isim_grupları()
    metota_ait_hedef_terim_listesi ← WSBFE_hedef_terim_listesi()
  else if metot_adı == 'C-value' then
    metota_ait_isim_grupları ← C-value_isim_grupları()
    metota_ait_hedef_terim_listesi ← C-value_hedef_terim_listesi()
  end if
  değiştirilmiş_doküman ← dokümanı_değiştir(doküman, metota_ait_isim_grupları)
  hedef_terim_listesi = LDA_hedef_terim_bul(değiştirilmiş_doküman)
  //hedef terim geçерleme
  for each terim in hedef_terim_listesi do
    if metota_ait_hedef_terim_listesi contains terim then
      doğrulanmış_hedef_terim_listesi'ne terim'i ekle
    end if
  end for
  //hedef terim listesi genişletme
  genişletilmiş_hedef_terim_listesi ← doğrulanmış_hedef_terim_listesi
  for isim_grubu in sık_geçen_isim_grupları do
    for each terim in doğrulanmış_hedef_terim_listesi do
      if terim kelime sayısı > 1 ve isim_grubu contains terim then
        genişletilmiş_hedef_terim_listesi'ne isim_grubu'nu ekle
      else if terim kelime sayısı == 1 then
        if isim_grubu içindeki herhangi bir kelime == terim then
          genişletilmiş_hedef_terim_listesi'ne isim_grubu'nu ekle
        end if
      end if
    end for
  end for
  Return genişletilmiş_hedef_terim_listesi
end procedure

```

Şekil 4.5. Hedef terim çıkarım algoritması

4.4. Hedef Terimlerin Kategorisinin Bulunması

Hedef terimlerin kategorisini bulmak için, önceden tanımlanmış kategori tanımları ile her bir hedef terimi arasındaki PMI değeri hesaplanmıştır. Hedef terimlerin en yüksek PMI skoruna sahip olduğu kategori tanımı, o hedefin kategorisi olarak kabul edilmiştir. Algoritma Şekil 4.6.'da verilmektedir.

```

procedure KATEGORI_BUL(kategori_listesi, hedef_terim_listesi)
  hedef_terim_kategori_eşlemesi = {}
  for each terim in hedef_terim_listesi do
    kategori_PMI_skoru_eşlemesi = {}
    for each kategori in kategori_listesi do
      PMI_kategori = PMI_hesapla(kategori, terim)
      kategori_PMI_skoru_eşlemesi'ne PMI_kategori'yi ekle
    end for
    kategori = en yüksek skora sahip olan kategori
    hedef_terim_kategori_eşlemesi'ne terim-kategori çiftini ekle
  end for
  Return hedef_terim_kategori_eşlemesi
end procedure

```

Şekil 4.6. Hedef terim kategori eşleme algoritması

4.5. Hedef Terimlerin Duygu Sınıflandırması

Hedef – görüş çiftlerini bulmak ve sınıflandırmak için bir önceki adımda belirlenmiş olan hedef terimler ve iki farklı duygu sözlüğünden yararlanan kural tabanlı bir yaklaşım önerilmektedir.

Birinci duygu sözlüğü, Hu ve Liu [27] tarafından oluşturulmuş İngilizce sıfatları içeren duygu sözlüğünün Tureng¹ Sözlüğü kullanılarak otomatik olarak Türkçeleştirilmesi sonucunda elde edilmiştir, bu sözlüğe görüş ifade eden fiiller, görüş ifade eden isimler de eklenerek sözlük zenginleştirilmiştir. İkinci sözlük ise Dehkharghani ve arkadaşlarının [41] geliştirdiği SentiTurkNet'tir. Duygu ifadeleri öncelikle ilk belirtilen sözlükte aranmaktadır, birinci sözlükte bulunamayan ifadeler SentiTurkNet üzerinde aranmaktadır. Böylelikle ilgili terimin polaritesinin belirlenmesi amaçlanmaktadır.

¹ <https://tureng.com/tr/turkce-ingilizce>

Algoritma dahilinde uygulanan kural seti bütüncül, kapalı bir mantıksal ifade olarak Şekil 4.7.'de sunulmaktadır. Kural tabanlı yaklaşım ile bu ifadenin çözümlenmesi beklenmektedir. Bu yaklaşımda “ama”, “fakat” bağlaçları dikkate alınmamıştır çünkü yalnızca sözcük öbekleri ve onlara ait görüşlerin bulunması amaçlanmıştır. Bu gösterimde İ: hedef terim olarak bulunan isim ve isim gruplarını, Z: zarfları, S: duygu sözlüğünde bulunan sıfatları, F: duygu sözlüğünde bulunan fiilleri temsil etmektedir. “*” ifadesi hiç bulunmama veya tekrarlı olarak bulunma, “?” ifadesi hiç bulunmama veya bir kere bulunma, “|” ifadesi veya anlamına gelmektedir.

$$\dot{I} (('ve' | 'ile' | ',') \dot{I})^* (Z)^* ((S | F) ('de' | 'da')? ('değil')?) (('ve' | ',') ((Z)^* (S | F) ('de' | 'da')? ('değil')?))^*$$

Şekil 4.7. Kural tabanlı yaklaşımın mantıksal ifadesi

Aşağıdaki kurallar ve bu kurallara ait cümleler, bu ifade ile çözümlenebilen örnek veriler olarak sunulmaktadır:

- Hedef terim olarak belirlenmiş isim veya isim grubundan sonra
 - ❖ Görüş ifade eden sıfatın yüklem olarak gelmesi durumu
 - Yemekler güzel.
 - Oda, banyo ve mobilyalar temizdi.
 - ❖ Yüklem olarak gelen sıfattan sonra değil kelimesinin gelebilme durumu
 - Yemekler güzel değil.
 - ❖ Görüş ifade eden olumlu ya da olumsuz fiil gelmesi durumu
 - Yemekleri beğendim.
 - Yemekleri beğenmedim.
 - ❖ Görüş ifade eden olumlu ya da olumsuz fiilden sonra “değil” kelimesinin tek başına veya “de/da” bağlacı ile beraber gelmesi durumu
 - Yemekleri beğenmedim değil.
 - Yemekleri beğenmedim de değil.
 - ❖ “var” veya “yok” ifadesi geliyorsa
 - Akşam yemeği var.
 - Akşam yemeği yok.

- Hedef terim içeren isim grubunda görüş ifade eden “problem”, “sıkıntı” gibi kelimelerin gelmesi durumu
- Otopark sıkıntısı var.
- Yemek problemi yaşadık.

```

procedure DUYGU_SINIFLANDIR(cümle_listesi)
  duygu_ifadeleri = {}
  hedef_terim_listesi = {}
  for each cümle in cümle_listesi do
    hedef_terim = ""
    duygu_ifadesi = ""
    anlam_değiştirici = {}
    belirteç_listesi = belirteçlere_ayır(cümle)
    for each belirteç in belirteç_listesi do
      bir_sonraki_belirteç = sonraki_belirteci_bul(belirteç)
      iki_sonraki_belirteç = sonraki_belirteci_bul(bir_sonraki_belirteç)
      if isimMi(belirteç) and !emojiMi(belirteç) then
        hedef_terim güncelle
        if yokKelimesiMi(bir_sonraki_belirteç) or varKelimesiMi(bir_sonraki_belirteç) then
          if yokKelimesiMi(bir_sonraki_belirteç) or değilKelimesiMi(iki_sonraki_belirteç) then
            anlam_değiştirici'yi güncelle
          end if
        end if
        if duygulçerirMi(hedef_terim) then
          hedef terime ait duygu ifadesini bul
          duygu_ifadeleri'ne duygu_ifadesi ekle
        else
          hedef_terim_listesi'ne hedef_terim'i ekle
        end if
      else if duygulçerirMi(belirteç) then
        duygu_ifadesi'ni güncelle
        if olumsuzlukEkiVarMi(belirteç) then
          anlam_değiştirici'yi güncelle
        end if
        if !duygulçerirMi(bir_sonraki_belirteç) then
          if değilKelimesiMi(bir_sonraki_belirteç) or (fiilMi(bir_sonraki_belirteç) and
            "olmak,kılmak,etmek,eylemek,kalmak".contains(bir_sonraki_belirteç)) then
            duygu_ifadesi'ni güncelle
            anlam_değiştirici'ni güncelle
          end if
        if !bosMu(hedef_terim_listesi) then
          for each terim in hedef_terim_listesi do
            hedef terime ait duygu ifadesini bul
            duygu_ifadeleri'ne duygu_ifadesi'ni ekle
          end for
        else
          aspect = Genel
          hedef terime ait duygu ifadesini bul
          duygu_ifadeleri'ne duygu_ifadesi'ni ekle
        end if
      end if
    end if
  end for
  end for
  Return duygu_ifadeleri
end procedure

```

Şekil 4.8. Hedef terim – görüş eşleştirme ve duygu sınıflandırma algoritması

Şekil 4.8.'de sunulan algoritmada görüleceği üzere, hedef belirleme ve hedef görüş eşleştirme algoritması iç içe iki döngüden oluşmaktadır. Dış döngüde parametre olarak verilen cümleler listesindeki her bir cümle kelime belirteçlerine ayrılır. Kelime belirteci, kelimelerin türünü ve aldığı ekleri ifade eder. İkinci döngüde ise o cümleye ait kelime belirteci listesi içindeki belirteçlere çeşitli kurallar uygulanır. Öncelikle sıradaki belirteç ve sonraki iki belirteç hafızaya alınır. Döngü içindeki ilk **if** cümlecigi içinde hafızaya alınan belirteçlerle hedef terimlerin bulunması amaçlanmıştır. Örneğin “otopark sıkıntısı var” cümlesinde olduğu gibi hedef terim kendi içinde olumlu veya olumsuz bir yönelim içermekte ise hedef terimin sınıflandırması yapılmaktadır. **Else if** içinde ise bir önceki **if** blokunda bulunan hedef terimlerin görüş sınıflandırması yapılır. Eğer hedef terim bulunamamışsa hedef terim “genel” olarak ifade edilir. Görüş ifade eden fiil ise olumsuzluk eki (me, ma) ve “değil” kelimesi dikkate alınarak sınıflandırma yapılır. Sıfat ise “değil” olumsuzluk kelimesi dikkate alınarak sınıflandırma yapılmaktadır.

5. DENEYSEL SONUÇLAR

5.1. Hedef Terimlerin Çıkarımı

Önerilen metotları karşılaştırmak için izlenen akış şu şekildedir. Metotların buldukları tüm terimler bir listeye tekrarlı olmayacak şekilde aktarılmıştır. Metotların tahmin ettiği ve gerçek hedef terim listesinde bulunan terimler gerçek pozitif olarak kabul edilirken, hedef terim listesinde bulunmayanlar yanlış pozitif olarak kabul edilmiştir. Benzer şekilde metotların tahmin etmediği ve gerçek hedef terim listesinde bulunan terimler yanlış negatif olarak değerlendirilirken, gerçek hedef terim listesinde bulunmayanlar gerçek negatif olarak değerlendirilmiştir. Kesinlik (precision), duyarlılık (recall) ve f-skor metrikleri bu yaklaşıma göre hesaplanmıştır. Hedef terim çıkarımı için önerilen metotların kesinlik, duyarlılık ve f-skor metrikleri bakımından klasik LDA metodundan daha başarılı sonuçlar verdiği görülmüştür.

Kama ve arkadaşları [33], kesinlik ve duyarlılık hesabını hata matrisi sunmadan gerçekleştirmişlerdir. Kesinlik oranının hesaplanması için doğru tahmin edilen terim sayısı toplam tahmin sayısına bölünmüştür, duyarlılık oranının hesaplanması içinse doğru tahmin edilen terim sayısı bulunması gereken terim sayısına bölünmüştür. Bu yaklaşım kullanılarak elde edilen sonuçlar ile çalışma kapsamında benimsenen yaklaşımın sonuçları bire bir örtüşmektedir. Hedef terim tespitine ait kesinlik ve duyarlılık oranlarının hata matrisine ihtiyaç duyulmadan hesaplanabilmesinin mümkün olduğu görülmüştür.

LDA gerçekleştirmek için gerekli olan α , β , iterasyon ve konu sayısı değişkenleri sırasıyla 0,1; 0,01; 50 ve 20 olarak belirlenmiştir. WSBFE metodunda kelime frekansı eşik değeri olarak 3, arama motoru sorgu sonuç eşik değeri olarak ise 1500 kabul edilmiştir. C-value metodunda tek kelimeler için 5, çok kelimeli ifadeler için ise 2,5 eşik değeri olarak ayarlanmıştır. Verilen algoritma parametreleri yapılan ön deneylere göre ampirik olarak belirlenmiştir.

Çizelge 5.1. Hedef terimlerin çıkarımına ait hata matrisi

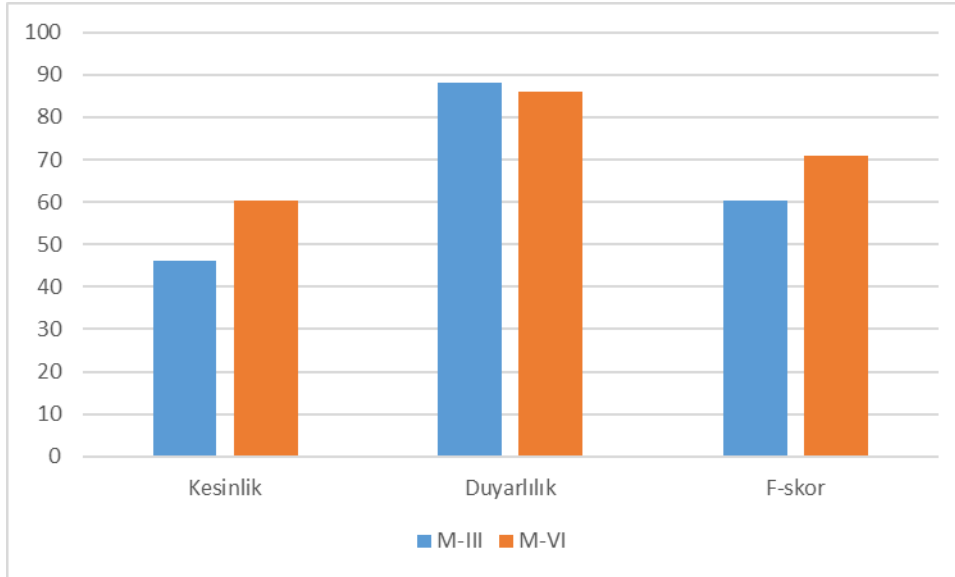
Numara	Metot	Gerçek Pozitif	Yanlış Pozitif	Yanlış Negatif	Gerçek Negatif
M-I	LDA	160	330	303	442
M-II	C-value	269	273	194	499
M-III	WSBFE	408	479	55	293
M-IV	LDA with replacing C-value	239	257	224	515
M-V	LDA with Replacing & Verifying C-value Not Extending NGs	242	200	221	572
M-VI	LDA with Replacing & Verifying C-value Extending NGs	398	262	65	510
M-VII	LDA with Replacing WSBFE	276	239	187	533
M-VIII	LDA with Replacing & Verifying WSBFE Not Extending NGs	247	172	216	600
M-IX	LDA with Replacing & Verifying WSBFE Extending NGs	380	240	83	532

Çizelge 5.2. Hedef terimlerin çıkarımı ait başarı metrikleri

Numara	Metot	Kesinlik (%)	Hassasiyet (%)	F-skor (%)
M-I	LDA	32,65	34,55	33,57
M-II	C-value	49,63	58,10	53,53
M-III	WSBFE	45,99	88,12	60,44
M-IV	LDA with replacing C-value	48,18	51,62	49,84
M-V	LDA with Replacing & Verifying C-value Not Extending NGs	54,75	52,26	53,48
M-VI	LDA with Replacing & Verifying C-value Extending NGs	60,30	85,96	70,88
M-VII	LDA with Replacing WSBFE	53,59	59,61	56,44
M-VIII	LDA with Replacing & Verifying WSBFE Not Extending NGs	58,94	53,34	56,01
M-IX	LDA with Replacing & Verifying WSBFE Extending NGs	61,29	82,07	70,17

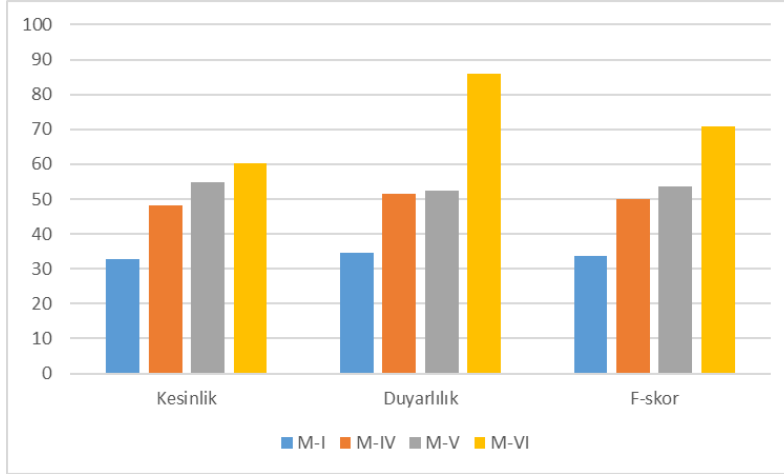
Çizelge 5.1.'de tüm hedef terimlerin çıkarımına ait hata matrisi sunulmaktadır. Sonuçların yorumlanmasında kolaylık sağlamak amacıyla metotlara numaralar verilmiştir ve sonuçlar

bu numaralara göre yorumlanmıştır. Çizelge 5.2.'de ise tüm hedef terimlerin çıkarımına ait başarı metrikleri verilmektedir. Çizelge 5.2.'den görüleceği üzere, çalışmanın amacı olarak belirtilen LDA temelli hedef terim çıkarımı için önerilen M-VI ve M-IX yöntemleri ile M-I metodu f-skor metriği bazında karşılaştırıldığında %100'ün üzerinde bir başarı artışı elde edilmiştir. Literatürde bahsedilen ilk üç metodun f-skor cinsinden en başarılısı 60,44 skoruyla M-III metodudur. Bu çalışmada gerçekleştirilen yöntemler arasında ise 70,88 skoruyla M-VI numaralı metot birinci sıradadır. İki grup karşılaştırıldığında başarıda yaklaşık %10'luk bir artış sağlanmıştır. Bunun yanında M-VI numaralı metot kesinlik skorunda yaklaşık 15 puanlık bir artış sağlamaktadır. Bunun sebebi, M-VI numaralı metodun daha az hedef terim tahmini yapmasıdır. Şekil 5.1.'de bu iki yöntemi karşılaştıran grafik verilmektedir.



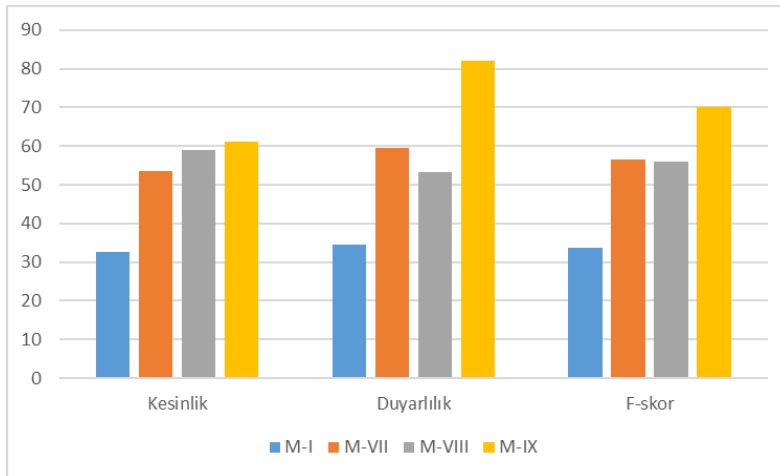
Şekil 5.1. M-III ve M-VI numaralı metotların karşılaştırılması

Metotların, dayandığı temel yaklaşımlara göre gruplanarak karşılaştırılması yapılırsa, C-value yönteminden yararlanarak LDA temelli hedef çıkarım metotları için Şekil 5.2.'deki grafik elde edilmektedir. C-value yöntemi kullanılarak önerilen metotlar 3 aşamadan oluşmaktadır. Birinci aşamada C-value ile bulunan isim grupları kullanılarak veri seti güncellenir ve LDA ile hedef terimler bulunur. Bu aşamada f-skor oranında %16'dan fazla artış gözlemlenmiştir. İkinci aşamada, bir önceki adımda elde edilen hedef terimler C-value yöntemi ile geçerlenir. Bu aşamada, bir önceki aşamaya göre f-skor değeri %3'ten fazla artmıştır. Üçüncü aşamada, ikinci aşamada elde edilen hedef terimler, sık geçen isim grupları ile genişletilmiştir ve f-skor değeri %17'nin üzerinde artış göstermektedir.



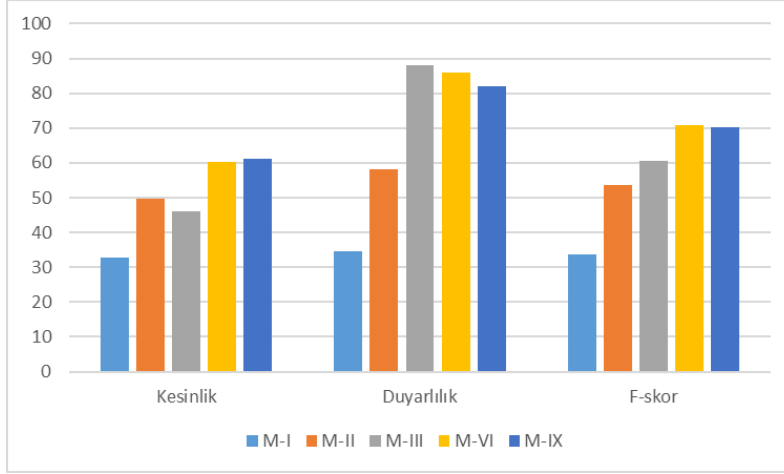
Şekil 5.2. C-value yönteminden yararlanılan metotlar ile LDA'nın karşılaştırılması

Benzer karşılaştırma WSBFE yönteminden yararlanılan metotlar için yapıldığında Şekil 5.3.'teki grafik elde edilmektedir. WSBFE yöntemi kullanılarak önerilen metotlar 3 aşamadan oluşmaktadır. Birinci aşamada WSBFE ile bulunan isim grupları kullanılarak veri seti güncellendikten sonra LDA ile hedef terimler bulunur. Bu aşamada f-skor oranında %23'e yakın artış gözlemlenmiştir. İkinci aşamada, bir önceki adımda elde edilen hedef terimler WSBFE yöntemi ile geçerlenir. Bu aşamada, bir önceki aşamaya göre f-skor değeri %0,4 civarında azalmıştır. Bu durum geçerleme sırasında doğru tahmin edilen bazı hedef terimlerin elimine edilmesinden kaynaklanmaktadır. Üçüncü aşamada, ikinci aşamada elde edilen hedef terimler, sık geçen isim grupları ile genişletilmiştir ve f-skor değeri %14'ün üzerinde artış göstermiştir.



Şekil 5.3. WSBFE yönteminden yararlanılan metotlar ile LDA'nın karşılaştırılması

Karşılaştırmayı genelleylebilmek için önerilen metotlardan en yüksek f-skoruna sahip iki metot ile literatürdeki yöntemler Şekil 5.4'teki grafikte karşılaştırılmaktadır.



Şekil 5.4. Hedef terim çıkarım yöntemlerinin karşılaştırılması

M-VI ve M-IX metotları f-skor bakımından birbirlerine yakın sonuçlar üretmiştir. Duyarlılık bakımından en başarılı yöntem M-III olarak görülmektedir. Çalışma kapsamında önerilen iki yöntemin LDA metoduna kıyasla büyük ölçüde başarılı olduğu görülmektedir.

221 adet çok kelimeli hedef terim bulunan veri setinde uygulanan yöntemlerin çok kelimeli terimleri keşfetme başarısına ait hata matrisi Çizelge 5.3.'te verilmektedir. Çizelge 5.4.'te ise başarı metrikleri sunulmaktadır. Çalışmanın amacında belirtildiği üzere hedef terim çıkarmada önerilen yöntemler ile LDA metodunun çok kelimeli terim bulamama kısıtının ortadan kaldırılması amaçlanmıştır. M-VI ve M-IX numaralı metotlar ile M-III metodu karşılaştırıldığı zaman bu üç yöntemin de yakın f-skor değerlerine sahip oldukları görülmektedir. Ayrıca M-IV ve M-V metotları ile M-VII ve M-VIII metotları hedef terim listesini isim grupları kullanarak genişletmemesine rağmen bu metotlarda da dikkate değer sonuç alınmaktadır. EK-2.'de M-III, M-VI ve M-IX metotlarının tespit ettiği çok kelimeli terimler sunulmaktadır.

Çizelge 5.3. Metotların çok kelimeli terimleri keşfetme başarılarının karşılaştırılması hata matrisi

Numara	Metot	Gerçek Pozitif	Yanlış Pozitif	Yanlış Negatif	Gerçek Negatif
M-I	LDA	-	-	-	-
M-II	C-value	52	39	169	69
M-III	WSBFE	193	75	28	33
M-IV	LDA with replacing C-value	36	11	185	97
M-V	LDA with Replacing & Verifying C-value Not Extending NGs	31	7	190	101
M-VI	LDA with Replacing & Verifying C-value Extending NGs	192	76	29	32
M-VII	LDA with Replacing WSBFE	73	15	148	93
M-VIII	LDA with Replacing & Verifying WSBFE Not Extending NGs	58	6	163	102
M-IX	LDA with Replacing & Verifying WSBFE Extending NGs	191	71	30	37

Çizelge 5.4. Metotların çok kelimeli terimleri keşfetme başarılarının karşılaştırılması

Numara	Metot	Kesinlik (%)	Hassasiyet (%)	F-skor (%)
M-I	LDA	-	-	-
M-II	C-value	57,14	23,53	33,33
M-III	WSBFE	72,01	87,33	78,93
M-IV	LDA with replacing C-value	76,59	16,29	26,86
M-V	LDA with Replacing & Verifying C-value Not Extending NGs	81,57	14,02	23,94
M-VI	LDA with Replacing & Verifying C-value Extending NGs	71,64	86,87	78,52
M-VII	LDA with Replacing WSBFE	82,95	33,03	47,25
M-VIII	LDA with Replacing & Verifying WSBFE Not Extending NGs	90,625	26,24	40,7
M-IX	LDA with Replacing & Verifying WSBFE Extending NGs	72,9	86,42	79,08

5.2. Hedef Terimlerin Kategorisinin Bulunması

Hedef terimlerin kategorilerinin bulunması için Turney'in [16] önerdiği eşitlik kullanılmıştır. Veri setinde önceden belirlenmiş hedef terim listesindeki her bir terim ile *fiyat, hizmet, içecek, konum, ortam, restoran, yemek, yönetim* kategorileri için ayrı ayrı hesaplama yapılmıştır. En yüksek değerin elde edildiği kategori o hedef terimin kategorisi olarak kabul edilmiştir ve %59'luk bir doğruluk oranına ulaşılmıştır.

5.3. Hedef Terimlerin Duygu Sınıflandırılması

463 farklı hedef terim ve 1837 hedef terim – görüş çifti içeren etiketlenmiş test verisinde 185 negatif, 24 nötr ve 1628 pozitif çift bulunmaktadır, sınıflandırma yapılırken nötr çiftler kapsam dışı bırakılmıştır. Hedef terim çıkarımında en başarılı olan M-VI numaralı metot neticesinde elde edilen 398 hedef terim kural tabanlı hedef terim görüş eşleştirme metoduna girdi olarak verildiğinde, 213 farklı hedef terim için 637 eşleştirme yapılmıştır ve bunlardan 506'sı doğru sınıflandırılmaktadır (Bknz. Çizelge 5.5.). Kural tabanlı metoda girdi olarak hedef terim listesi verilmediğinde ise 237 hedef terim için 1568 hedef terim – görüş çifti bulunmuştur ve bunların 643'i doğru olarak sınıflandırılmaktadır (Bknz. Çizelge 5.6.).

Çizelge 5.5. Başlangıç hedef terim listesi verilince elde edilen hata matrisi ve doğruluk oranı

Gerçek Pozitif	Yanlış Pozitif	Yanlış Negatif	Gerçek Negatif	Doğruluk (%)
445	97	34	61	79,43

Çizelge 5.6. Başlangıç hedef terim listesi verilmeden elde edilen hata matrisi ve doğruluk oranı

Gerçek Pozitif	Yanlış Pozitif	Yanlış Negatif	Gerçek Negatif	Doğruluk (%)
551	767	158	92	41,007

Hedef terim çıkarım aşamasında elde edilen terimler kullanılarak hedef terim ile görüş sınıflandırma başarısı artırılmıştır. Başarılı ve başarısız sınıflandırma örnekleri Çizelge 5.7.'de verilmektedir. Hedef terim listesi verilmeden gerçekleştirilen eşleştirmede ise filtreleme yapılmadığı için düşük doğruluk oranı elde edilmektedir. Hedef terim listesi verilmeden yapılan sınıflandırma işleminde elde edilen başarılı ve başarısız sonuçlar Çizelge 5.8'te verilmektedir.

Çizelge 5.7. Hedef terim listesi verilerek sınıflandırılmış bazı hedef terimler

Hedef Terim	Görüş	Polarite	Başarılı mı?
yemek sunumu	iyi	pozitif	evet
kahvaltı çeşidi	az	negatif	evet
internet bağlantısı	güzel	pozitif	evet
resepsiyon personeli	çok yardımsever	pozitif	evet
oda servisi	fevkalade	pozitif	evet
oda servisi	çok lezzetli	pozitif	hayır
park sorunu	yok	pozitif	evet
havuz	küçük	negatif	evet
dekorasyon	eski	negatif	evet
fiyat	yüksek	pozitif	hayır
oda manzarası	harika	pozitif	evet
oda manzarası	çok güzel	pozitif	evet
pide	iyi değil	negatif	evet
kahvaltı	beğenmek	pozitif	evet

Çizelge 5.8. Hedef terim listesi verilmeden sınıflandırılmış bazı hedef terimler

Hedef Terim	Görüş	Polarite	Başarılı mı?
otel organizasyonu	harika	pozitif	evet
tatıl	geri	negatif	hayır
neden	güzel	pozitif	hayır
kart	geniş çok	pozitif	hayır
omlet	sıcak	pozitif	evet
omlet	taze	pozitif	evet
resepsiyon personeli	çok yardımsever	pozitif	evet
tv	küçük	negatif	evet
fiyat	yüksek	pozitif	hayır
banyo	kötü	negatif	evet
resepsiyon	çok yetersiz	negatif	evet
garson	çok kaba	negatif	evet
kahvaltı	beğenmek	pozitif	evet

Hedef terim listesi verildiği durumda %79'un üzerinde doğruluk oranı elde edilmesine rağmen hedef terim – görüş çiftlerinin büyük kısmı eşleştirilememiştir. Bunun sebepleri günlük hayatta kullanılan Türkçede dil bilgisi kurallarına uyulmaması, noktalama işaretlerine dikkat edilmemesi, çok uzun cümleler kurulması olarak gözlemlenmiştir. Hedef terimlerin yanlış görüşlerle eşleştirilmesi ve bazı görüş ifadelerinin hedef terimlere göre polaritesinin değişmesi (fiyat-yüksek, hizmet kalitesi-yüksek vb.) de başarıyı etkileyen diğer sebeplerdir. Bu sebepler kural tabanlı yaklaşımın performansını düşüren temel etkenlerdir ve bu sebeplere dayalı performans düşüklüğünün giderilmesi, başlı başına yürütülmesi gereken bir çalışma olarak kurgulanmalıdır.

6. SONUÇ VE ÖNERİLER

Çalışma kapsamında güncel bir konu olan hedef temelli duygu analizi için hedef terim çıkarma, hedef terim kategorisi belirleme ve hedef terim duygu sınıflandırması için çeşitli yöntemler önerilmiştir. Türkçe duygu analizi çalışmalarında kullanılabilecek, gerçek kullanıcı yorumlarını içeren bir veri seti oluşturulmuştur.

Konu modelleme yöntemi olan LDA'nın BoW kısıtını gidermek için iki farklı yöntem önerilmiştir. Bu iki yöntem de çok kelimeli terimleri sürece dahil etmeyi amaçlamaktadır. Literatürdeki metotların ve bu çalışma kapsamında önerilen yöntemlerin, gerçek kullanıcı verisi üzerinde uygulanmasıyla elde edilen karşılaştırmalı deneysel sonuçlar, önerilen yöntemlerin klasik LDA'ya oranla daha başarılı sonuçlar üretme kabiliyetine sahip olduğunu ortaya koymuştur.

PMI yönteminden yararlanarak hedef terim kategori sınıflandırılması yapılmıştır. Duygu sınıflandırma aşaması için kural ve sözlük tabanlı yaklaşımlardan yararlanan bir yaklaşım neticesinde kayda değer sonuçlar alınmıştır. Türkçe'nin doğru kullanılmaması, çok uzun cümleler kurulması gibi etkenler, kural tabanlı yaklaşımın performansını sınırlandıran olası engeller olarak görülmektedir.

Arama motoru kullanılarak PMI değeri hesaplamada karşılaşılan en büyük sorun büyük miktarda sorgulama yapıldığında çok fazla vakit almaktadır ve kullanılan API günlük olarak sınırlı sayıda sorguya izin vermektedir. Bu durumlar göz önüne alındığında, her bir sorgunun veri tabanı veya başka bir depolama alanında saklanması yoluyla hem deney hem de oluşturulabilecek son ürün performansı artırabileceği düşünülmektedir.

Bu tez çalışması kapsamında önerilen algoritmaların söz konusu özel durumları ele alacak şekilde geliştirilmesinin duygu analizi konusunda Türkçe gibi sondan eklemeli dillerde başarı oranlarını artırabileceği ve literatüre hem akademik hem de ticari anlamda katkıları olacağı düşünülmektedir.

KAYNAKLAR

1. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
2. Liu, Y., Huang, X., An, A., Yu, X. (2007, July). *ARSA: a sentiment-aware model for predicting sales performance using blogs*. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, Amsterdam, The Netherlands, 607-614.
3. Pennacchiotti, M., Popescu, A. M. (2011, July). *A machine learning approach to twitter user classification*. In Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 281-288.
4. Tumasjan, A., Sprenger, T. O., Sandner, P. G., Welp, I. M. (2010, May). *Predicting elections with twitter: What 140 characters reveal about political sentiment*. In *Fourth International AAAI Conference On Weblogs And Social Media*, Washington, DC, USA, 178-185.
5. Watts, D., George, K. M., Kumar, T. A., Arora, Z. (2016, December). *Tweet sentiment as proxy for political campaign momentum*. In 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 2475-2484.
6. Bollen, J., Mao, H., Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
7. Şimşek, M. U., Özdemir, S. (2012, October). *Analysis of the relation between Turkish twitter messages and stock market index*. In 2012 6th International Conference on Application of Information and Communication Technologies (AICT), Tbilisi, Georgia, 1-4.
8. Birjali, M., Beni-Hssane, A., Erritali, M. (2017). Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. *Procedia Computer Science*, 113, 65-72.
9. Baj-Rogowska, A. (2017, December). *Sentiment analysis of facebook posts: The uber case*. In 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 391-395.
10. Türkmen, H., Omurca, S. I., Ekin, E. (2016). An aspect based sentiment analysis on Turkish hotel reviews. *Girne American University Journal of Social and Applied Sciences*, 6, 9-15.
11. Oğul, B. B., Ercan, G. (2016, May). *Sentiment classification on Turkish hotel reviews*. In 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, Türkiye, 497-500.
12. Hu, Y. H., Chen, Y. L., Chou, H. L. (2017). Opinion mining from online hotel reviews—A text summarization approach. *Information Processing & Management*, 53(2), 436-449.

13. Bhoir, P., Kolte, S. (2015, December). *Sentiment analysis of movie reviews using lexicon approach*. In 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 1-6.
14. Parkhe, V., Biswas, B. (2015, August). *Genre specific aspect based sentiment analysis of movie reviews*. In 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India, 2418-2422.
15. Indurkha, N., Damerau, F. J. (Eds.). (2010). *Handbook of natural language processing* (Vol. 2). Florida: CRC Press.
16. Turney, P. D. (2002, July). *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*. In Proceedings of the 40th annual meeting on association for computational linguistics, Philadelphia, Pennsylvania, USA, 417-424.
17. Skuza, M., Romanowski, A. (2015, September). *Sentiment analysis of Twitter data within big data distributed environment for stock prediction*. In 2015 Federated Conference on Computer Science and Information Systems (FEDCSIS), Lodz, Poland, 1349-1354.
18. Mandal, S., Gupta, S. (2016, September). *A novel dictionary-based classification algorithm for opinion mining*. In 2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, 175-180.
19. Çoban, Ö., Özyer, B., Özyer, G. T. (2015, May). *Sentiment analysis for Turkish Twitter feeds*. In 2015 23rd Signal Processing and Communications Applications Conference (SIU), Malatya, Türkiye, 2388-2391.
20. Çoban, Ö., Özyer, G. T. (2016, May). *Sentiment classification for Turkish Twitter feeds using LDA*. In 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, Türkiye, 129-132.
21. Vural, A. G., Cambazoglu, B. B., Senkul, P., Tokgoz, Z. O. (2013). *A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish*. In Computer and Information Sciences III, London, England, 437-445.
22. Erogul, U. (2009). *Sentiment Analysis in Turkish*, Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
23. Yu, H., Hatzivassiloglou, V. (2003, July). *Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences*. In Proceedings Of The 2003 Conference On Empirical Methods in Natural Language Processing, Stroudsburg, PA, USA, 129-136.
24. Barbosa, L., Feng, J. (2010, August). *Robust sentiment detection on twitter from biased and noisy data*. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Beijing, China, 36-44.

25. Davidov, D., Tsur, O., Rappoport, A. (2010, August). *Enhanced sentiment learning using twitter hashtags and smileys*. In Proceedings of the 23rd International Conference On Computational Linguistics: Posters, Beijing, China, 241-249).
26. Khan, A., Baharudin, B., Khan, K. (2011). Sentiment classification using sentence-level lexical based. *Trends in Applied Sciences Research*, 6(10), 1141-1157.
27. Hu, M., Liu, B. (2004, August). *Mining and summarizing customer reviews*. In Proceedings of The Tenth ACM SIGKDD International Conference On Knowledge Discovery and Data Mining, ACM, 168-177.
28. Siqueira, H., Barros, F. (2010, October). A feature extraction process for sentiment analysis of opinions on services. In *Proceedings of International Workshop on Web and Text Intelligence*, Sao Paulo, Brazil, 404-413.
29. Fei, G., Liu, B., Hsu, M., Castellanos, M., Ghosh, R. (2012). *A dictionary-based approach to identifying aspects implied by adjectives for opinion mining*. Proceedings of COLING 2012: Posters, Mumbai, India, 309-318.
30. Qiu, G., Liu, B., Bu, J., Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1), 9-27.
31. Yauris, K., Khodra, M. L. (2017, August). *Aspect-based summarization for game review using double propagation*. In 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), Denpasar, Indonesia, 1-6.
32. Liu, Q., Gao, Z., Liu, B., Zhang, Y. (2015, June). *Automated rule selection for aspect extraction in opinion mining*. In Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 1291-1297
33. Kama, B., Ozturk, M., Karagoz, P., Toroslu, I. H., Ozay, O. (2016, September). *A web search enhanced feature extraction method for aspect-based sentiment analysis for Turkish informal texts*. In International Conference on Big Data Analytics and Knowledge Discovery, Cham, Springer.
34. Kama, B., Ozturk, M., Karagoz, P., Toroslu, I. H., Kalender, M. (2017, November). *Analyzing implicit aspects and aspect dependent sentiment polarity for aspect-based sentiment analysis on informal Turkish texts*. In Proceedings of the 9th International Conference on Management of Digital EcoSystems, Bangkok, Thailand, 134-141.
35. Ekinci, E., Türkmen, H., Omurca, S. İ. (2017). Multi-word aspect term extraction using turkish user reviews. *International Journal of Computer Engineering and Information Technology*, 9(1), 15.
36. Medhat, W., Hassan, A., Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
37. Fellbaum, C. (1998). WordNet: An electronic lexical database and some of its applications. MIT Press, Cambridge, Mass.

38. Baccianella, S., Esuli, A., Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. *In Lrec*, 10, 2200-2204.
39. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.
40. Bilgin, O., Çetinoğlu, Ö., Oflazer, K. (2004). Building a wordnet for Turkish. *Romanian Journal of Information Science and Technology*, 7(1-2), 163-172.
41. Dehkharghani, R., Saygin, Y., Yanikoglu, B., Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*, 50(3), 667-685.
42. Ucan, A., Naderalvojud, B., Sezer, E. A., Sever, H. (2016, November). *SentiWordNet for new language: Automatic translation approach*. In 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 308-315). IEEE.
43. Ding, X., Liu, B., Yu, P. S. (2008, February). *A holistic lexicon-based approach to opinion mining*. In Proceedings of the 2008 International Conference On Web Search And Data Mining, Palo Alto, California, USA, 231-240.
44. Han, H., Zhang, J., Yang, J., Shen, Y., Zhang, Y. (2018). Generate domain-specific sentiment lexicon for review sentiment analysis. *Multimedia Tools and Applications*, 77(16), 21265-21280.
45. Feng, J., Gong, C., Li, X., Lau, R. Y. (2018). Automatic approach of sentiment lexicon generation for mobile shopping reviews. *Wireless Communications and Mobile Computing*, 2018, 13.
46. Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning Research*, 3, 993-1022.
47. Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., Zhao, L. (2018). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 1-43.
48. Titov, I., McDonald, R. (2008, April). *Modeling online reviews with multi-grain topic models*. In Proceedings of the 17th international conference on World Wide Web, Beijing, China, 111-120.
49. Brody, S., Elhadad, N. (2010, June). *An unsupervised aspect-sentiment model for online reviews*. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, California, 804-812.
50. Zhao, W. X., Jiang, J., Yan, H., Li, X. (2010, October). *Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid*. In Proceedings of the 2010 conference on empirical methods in natural language processing, Cambridge, Massachusetts, 56-65.

51. Bagheri, A., Saraee, M., De Jong, F. (2014). ADM-LDA: An aspect detection model based on topic modelling using the structure of review sentences. *Journal of Information Science*, 40(5), 621-636.
52. Poria, S., Chaturvedi, I., Cambria, E., Bisio, F. (2016, July). Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In *2016 international joint conference on neural networks (IJCNN)*, Vancouver, Canada, 4465-4473.
53. Shams, M., Baraani-Dastjerdi, A. (2017). Enriched LDA (ELDA): Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction. *Expert Systems with Applications*, 80, 136-146.
54. Ekinci, E., Türkmen, H., Omurca, S. İ. (2017). Multi-word aspect term extraction using turkish user reviews. *International Journal of Computer Engineering and Information Technology*, 9(1), 15.
55. Church, K. W., Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22-29.
56. Frantzi, K., Ananiadou, S., Mima, H. (2000). Automatic recognition of multi-word terms: the c-value/nc-value method. *International Journal On Digital Libraries*, 3(2), 115-130.
57. Barrón-Cedeno, A., Sierra, G., Drouin, P., Ananiadou, S. (2009, March). An improved automatic term recognition method for Spanish. In *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, Berlin, Heidelberg, 125-136.
58. Lossio-Ventura, J. A., Jonquet, C., Roche, M., Teisseire, M. (2013, December). Combining c-value and keyword extraction methods for biomedical terms extraction. In *Proceedings of the Fifth International Symposium on Languages in Biology and Medicine (LBM'13)*, Tokyo, Japan, 45-49.

EKLER

EK-1. Veri setinden örnek kısım

```

<reviews>
  <title>Türkçe Veri Seti</title>
  <review id="641235695" polarity="positive">
    <url>*****</url>
    <post-date>*****</post-date>
    <title>En güzeli</title>
    <content>Ankarada bu bölgede en güzel otellerden biri.Eğitim için bize ayrılan konaklama süresinde,sunumlar, çalışanların
    özverili yaklaşımları, yemekler, titizlik hepsi mükemmeldi. İyi bir yönetim, iyi bir iş disiplini hakim olduğu
    besbelli.Teşekkür ederim.
    </content>
    <rating>5.0</rating>
    <sentences>
      <sentence id="641235695:1" polarity="positive">
        <text>Ankarada bu bölgede en güzel otellerden biri.</text>
        <opinions>
          <opinion category="genel" implicitTarget="false" target="otel" opinion="güzel" sentimentShifters=""
            polarity="positive"/>
        </opinions>
      </sentence>
      <sentence id="641235695:2" polarity="positive">
        <text>
          Eğitim için bize ayrılan konaklama süresinde,sunumlar, çalışanların özverili yaklaşımları, yemekler, titizlik hepsi
          mükemmeldi.
        </text>
        <opinions>
          <opinion category="yemek#sunum" implicitTarget="false" target="sunumlar" opinion="mükemmel"
            sentimentShifters="" polarity="positive"/>
          <opinion category="yemek#kalite" implicitTarget="false" target="yemekler" opinion="mükemmel"
            sentimentShifters="" polarity="positive"/>
          <opinion category="hizmet#çalışan" implicitTarget="false" target="çalışanların
            yaklaşımı" opinion="mükemmel" sentimentShifters="" polarity="positive"/>
          <opinion category="hizmet#temizlik" implicitTarget="false" target="titizlik" opinion="mükemmel"
            sentimentShifters="" polarity="positive"/>
          <opinion category="hizmet#çalışan" implicitTarget="false" target="çalışanların
            yaklaşımı" opinion="özverili" sentimentShifters="" polarity="positive"/>
        </opinions>
      </sentence>
      <sentence id="641235695:3" polarity="positive">
        <text>
          İyi bir yönetim, iyi bir iş disiplini hakim olduğu besbelli.
        </text>
        <opinions>
          <opinion category="yönetim#genel" implicitTarget="false" target="yönetim" opinion="iyi"
            sentimentShifters="" polarity="positive"/>
          <opinion category="hizmet#çalışan" implicitTarget="false" target="iş
            disiplini" opinion="iyi" sentimentShifters="" polarity="positive"/>
        </opinions>
      </sentence>
      <sentence id="641235695:4" polarity="positive">
        <text>Teşekkür ederim.</text>
        <opinions>
          <opinion category="genel" implicitTarget="true" target="otel" opinion="teşekkür
            etmek" sentimentShifters="" polarity="positive"/>
        </opinions>
      </sentence>
    </sentences>
  </review>
</reviews>

```

Şekil 1.1. Veri setinden örnek kısım

EK-2. Çok kelimeli terimler

Çizelge 2.1. WSBFE ile bulunan çok kelimeli terimler – çekim eki olmadan

personel_kalite	dünya_mutfak	iş_bölge	otopark_sorun
kahvaltı_personel	lobi_alan	fiyat_kalite	şehir_iç
çay_kahve	toplantı_amaç	oda_manzara	oda_fiyat
yemek_lezzet	personel_hizmet	yol_bakan_oda	yatak_konfor
buhar_banyo	alışveriş_alan	peynir_çesit	gezi_sıra
tv_kanal	kahvaltı_servis	şehir_merkez_ulaşım	fitness_merkez
yol_üzeri	yemek_sunum	şehir_manzara	konum_itibar_ulaşım
fitness_center	vale_hizmet	resepsiyon_personel	oda_yemek
oda_bornoz	spor_tesis	seyahat_sebep	iş_gezi
restaurant_mevcut	yiyecek_kalite	yüzme_havuz	ortak_alan
çalışan_yaklaşım	yastık_kalite	oda_servis	kent_merkez
iş_seyahat_sebep	alt_yapı	oda_dekorasyon	sigara_koku
konum_açı	personel_iş	iş_amaç	oda_dizayn
hizmet_kalite	yemek_oda	iş_toplantı	ütü_masa
iç_yer	akşam_yemek	toplantı_oda	yol_kenar
merkez_ulaşım	balo_salon	arkadaş_tavsiye	yemek_menü
alışveriş_bölge	ses_izolasyon	kahvaltı_ortalama	konum_merkez
büfe_kahvaltı	fiyat_kalite_denge	oda_sigara	iş_alan
öğle_yemek	iş_arkadaş	ortalama_üzeri	personel_gülyüzlülük
tur_grup	servis_kalite	kalite_bakım	bar_alan
ulaşım_imkan	executive_salon	oda_kalite	kalite_denge
oda_düzen	hizmet_anlayış	ihtiyaç_cevap	oda_ses
oda_ücret	yerleşim_bölge	toplantı_organizasyon	kahvaltı_kalite
fitness_salon	standart_oda	imkan_sahip	kahvaltı_salon
bafra_pide	deluxe_oda	sağlık_merkez	toplantı_salon
bölge_iç_yer	yemek_seçenek	iş_salon	resepsiyon_hizmet
yatak_oda	ilgi_alaka	iş_ziyaret	kahvaltı_çeşitlilik
oda_sıcaklık	is_seyahat	taksi_yolculuk	aile_seyahat
şehir_gürültü	merkez_yer	yemek_servis	yardımcı_personel
yemek_çeşitlilik	iş_seyahat	diş_fırça	personel_kadro
uyku_kalite	yemek_salon	fiyat_aralık	su_basınç
kahve_ikram	havuz_alan	iç_dizayn	oda_konfor
oda_banyo	kahvaltı_hizmet	konum_itibari	kahvaltı_sıra
arkadaş_düğün	club_lounge	oda_kapı	meyve_su
personel_taraf	kat_görevli	yemek_standart	internet_hizmet
resepsiyon_görevli	şehir_merkez	spor_salon	executive_lounge
internet_bağlantı	executive_oda	oda_ihtiyaç	kahvaltı_dahil
personel_oda	yastık_seçim	portakal_su	kahvaltı_seçenek
yatak_kalite	kahvaltı_menü	servis_personel	oda_ferah
salon_servis	iş_saha	eskişehir_yol	salon_erişim
ev_sahiplik	hizmet_açı	oda_kart	bar_bölüm
konum_itibar	oda_iç	müşteri_memnuniyet	konum_yer
personel_ilgi	malzeme_kalite	yemek_kalite	ürün_kalite
alışveriş_merkez	tıraş_set	yastık_konfor	aile_oda
ulaşım_açı	park_yer	personel_gülyüz	buhar_oda
banyo_kapı	ses_yalıtım	kahvaltı_büfe	personel_yaklaşım
ulaşım_konu	executive_kat	park_sorun	kahvaltı_çesit
merkez_konum	kahvaltı_keyif	lobi_bar	düğün_organizasyon
iş_merkez			

EK-2. (devam) Çok kelimeli terimler

Çizelge 2.2. LDA with Replacing & Verifying C-value Extending NGs ile bulunan çok kelimeli terimler – çekim eki olmadan

personel kalite	dünya mutfak	iş bölge	şehir iç
kahvaltı personel	lobi alan	fiyat kalite	oda fiyat
çay kahve	toplantı amaç	oda manzara	yatak konfor
yemek lezzet	personel hizmet	yol bakan_oda	gezi sıra
buhar banyo	alışveriş alan	peynir çeşit	fitness merkez
tv kanal	kahvaltı servis	şehir merkez ulaşım	konum itibar ulaşım
yol üzeri	yemek sunum	şehir manzara	oda yemek
fitness_center	vale hizmet	resepsiyon personel	iş gezi
oda bornoz	spor tesis	seyahat sebep	ortak alan
restaurant mevcut	yiyecek kalite	yüzme havuz	kent merkez
çalışan yaklaşım	yastık kalite	oda servis	sigara koku
iş seyahat sebep	alt yapı	oda dekorasyon	oda dizayn
konum açısı	personel iş	iş amaç	ütü masa
hizmet kalite	yemek oda	iş toplantı	yol kenar
iç yer	akşam yemek	toplantı oda	yemek menü
merkez ulaşım	balo salon	arkadaş tavsiye	konum merkez
alışveriş bölge	ses izolasyon	kahvaltı ortalama	iş alan
büfe kahvaltı	fiyat kalite denge	oda sigara	personel güleryüzlülük
öğle yemek	iş arkadaş	ortalama üzeri	bar alan
tur grup	servis kalite	kalite bakım	kalite denge
ulaşım imkan	executive salon	oda kalite	oda ses
oda düzen	hizmet anlayış	ihtiyaç cevap	kahvaltı kalite
oda ücret	yerleşim bölge	toplantı organizasyon	kahvaltı salon
fitness salon	standart_oda	imkan sahip	toplantı salon
bafra pide	deluxe_oda	sağlık merkez	resepsiyon hizmet
bölge iç yer	yemek seçenek	iş salon	kahvaltı çeşitlilik
yatak oda	ilgi alaka	iş ziyaret	aile seyahat
oda sıcaklık	is seyahat	taksi yolculuk	yardımcı personel
şehir gürültü	merkez yer	yemek servis	personel kadro
yemek çeşitlilik	iş seyahat	fiyat aralık	su basınç
uyku kalite	yemek salon	iç dizayn	oda konfor
kahve ikram	havuz alan	konum itibari	kahvaltı sıra
oda banyo	kahvaltı hizmet	oda kapı	meyve su
arkadaş düğün	club lounge	yemek standart	internet hizmet
personel taraf	kat görevli	spor salon	executive lounge
resepsiyon görevli	şehir merkez	oda ihtiyaç	kahvaltı dahil
internet bağlantı	executive_oda	portakal su	kahvaltı seçenek
personel oda	yastık seçim	servis personel	oda ferah
yatak kalite	kahvaltı menü	eskişehir yol	salon erişim
salon servis	iş saha	oda kart	bar bölüm
ev sahipliği	hizmet açısı	müşteri memnuniyet	konum yer
konum itibar	oda iç	yemek kalite	ürün kalite
personel ilgi	malzeme kalite	yastık konfor	aile oda
alışveriş merkez	tıraş set	personel güleryüz	buhar oda
ulaşım açısı	park yer	kahvaltı büfe	personel yaklaşım
banyo kapı	ses yalıtım	park sorun	kahvaltı çeşit
ulaşım konu	executive kat	lobi bar	düğün organizasyon
merkez konum	kahvaltı keyif	otopark sorun	iş merkez

EK-2. (devam) Çok kelimeli terimler

Çizelge 2.3. LDA with Replacing & Verifying WSBFE Extending NGs ile bulunan çok kelimeli terimler – çekim eki olmadan

personel_kalite	dünya_mutfak	iş_bölge	oda_fiyat
kahvaltı_personel	lobi_alan	fiyat_kalite	yatak_konfor
çay_kahve	toplantı_amaç	oda_manzara	gezi_sıra
yemek_lezzet	personel_hizmet	yol_bakan_oda	fitness_merkez
buhar_banyo	alışveriş_alan	peynir_çesit	konum_itibar_ulaşım
tv_kanal	kahvaltı_servis	şehir_merkez_ulaşım	oda_yemek
yol_üzeri	yemek_sunum	şehir_manzara	iş_gezi
fitness_center	vale_hizmet	resepsiyon_personel	ortak_alan
oda_bornoz	spor_tesis	seyahat_sebep	kent_merkez
restaurant_mevcut	yiyecek_kalite	yüzme_havuz	sigara_koku
çalışan_yaklaşım	yastık_kalite	oda_servis	oda_dizayn
iş_seyahat_sebep	alt_yapı	oda_dekorasyon	ütü_masa
konum_açı	personel_iş	iş_amaç	yol_kenar
hizmet_kalite	yemek_oda	iş_toplantı	yemek_menü
iç_yer	akşam_yemek	toplantı_oda	konum_merkez
merkez_ulaşım	balo_salon	arkadaş_tavsiye	iş_alan
alışveriş_bölge	ses_izolasyon	kahvaltı_ortalama	personel_gülyüzlülük
büfe_kahvaltı	fiyat_kalite_denge	oda_sigara	bar_alan
öğle_yemek	iş_arkadaş	kalite_bakım	kalite_denge
tur_grup	servis_kalite	oda_kalite	oda_ses
ulaşım_imkan	executive_salon	ihtiyaç_cevap	kahvaltı_kalite
oda_düzen	hizmet_anlayış	toplantı_organizasyon	kahvaltı_salon
oda_ücret	yerleşim_bölge	imkan_sahip	toplantı_salon
fitness_salon	standart_oda	sağlık_merkez	resepsiyon_hizmet
bafra_pide	deluxe_oda	iş_salon	kahvaltı_çeşitlilik
bölge_iç_yer	yemek_seçenek	iş_ziyaret	aile_seyahat
yatak_oda	ilgi_alaka	taksi_yolculuk	yardımcı_personel
oda_sıcaklık	is_seyahat	yemek_servis	personel_kadro
şehir_gürültü	merkez_yer	fiyat_aralık	su_basınç
yemek_çeşitlilik	iş_seyahat	iç_dizayn	oda_konfor
uyku_kalite	yemek_salon	konum_itibari	kahvaltı_sıra
kahve_ikram	havuz_alan	oda_kapı	meyve_su
oda_banyo	kahvaltı_hizmet	yemek_standart	internet_hizmet
arkadaş_düğün	club_lounge	spor_salon	executive_lounge
personel_taraf	kat_görevli	oda_ihtiyaç	kahvaltı_dahil
resepsiyon_görevli	şehir_merkez	portakal_su	kahvaltı_seçenek
internet_bağlantı	executive_oda	servis_personel	oda_ferah
personel_oda	yastık_seçim	eskişehir_yol	salon_erişim
yatak_kalite	kahvaltı_menü	oda_kart	bar_bölüm
salon_servis	iş_saha	müşteri_memnuniyet	konum_yer
ev_sahiplik	hizmet_açı	yemek_kalite	ürün_kalite
konum_itibar	oda_iç	yastık_konfor	aile_oda
personel_ilgi	malzeme_kalite	personel_gülyüz	buhar_oda
alışveriş_merkez	tıraş_set	kahvaltı_büfe	personel_yaklaşım
ulaşım_açı	park_yer	park_sorun	kahvaltı_çesit
banyo_kapı	ses_yalıtım	lobi_bar	düğün_organizasyon
ulaşım_konu	executive_kat	otopark_sorun	iş_merkez
merkez_konum	kahvaltı_keyif	şehir_iç	

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : BAYRAKTAR, Kıvanç
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 04.10.1991, Ankara
 Medeni hali : Bekar
 Telefon : 0 (505) 580 10 35
 e-mail : bayraktarkivanc@gmail.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi / Bilgisayar Mühendisliği	Devam ediyor
Lisans	Gazi Üniversitesi / Bilgisayar Mühendisliği	2013
Lise	Çağrıbey Anadolu Lisesi	2009

İş Deneyimi

Yıl	Yer	Görev
2013-Halen	Hemosoft Bilişim ve Eğitim Hizmetleri A.Ş.	Yazılım Geliştirici

Yabancı Dil

İngilizce

Yayınlar

Bayraktar, K., Yavanoglu, U. and Ozbilen, A. (2019, December). *A rule-based holistic approach for turkish aspect-based sentiment analysis*. In 2019 IEEE International Conference on Big Data (IEEE BigData 2019), Los Angeles, CA, USA.

Bayraktar, K. (2019, June). *LDA-based aspect extraction from turkish hotel review data*. Data Science, Machine Learning and Statistics, Van Yüzüncü Yıl Üniversitesi, Van.

Hobiler

Kitap okumak, Spor yapmak, Müzik dinlemek.



GAZİ GELECEKTİR..