



**TÜRKÇE METİNLERDE DUYGU ANALİZİ YAPILARAK SALDIRGAN
İÇERİKLERİN TESPİT EDİLMESİ**

Şeyma ŞAHİNER YILMAZ

**YÜKSEK LİSANS TEZİ
YÖNETİM BİLİŞİM SİSTEMLERİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ**

HAZİRAN 2022

ETİK BEYAN

Gazi Üniversitesi Bilişim 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.

Şeyma ŞAHİNER YILMAZ

15/06/2022

TÜRKÇE METİNLERDE DUYGU ANALİZİ YAPILARAK SALDIRGAN İÇERİKLERİN TESPİT EDİLMESİ

(Yüksek Lisans Tezi)

Şeyma ŞAHİNER YILMAZ

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Haziran 2022

ÖZET

Sosyal medya mecralarının hızlı büyümesiyle birlikte daha fazla kullanıcı, görüşlerini ve fikirlerini çevrimiçi olarak paylaşmaktadır. Bu büyümenin ardından insanların varlıklara yönelik fikirlerini, duygularını, değerlendirmelerini, tutumlarını ve bunların yazılı metinde ifade edilen niteliklerini analiz eden duygu analizi çalışmaları popüler bir çalışma alanı haline gelmiştir. Bununla birlikte popüler sosyal medya platformlarından biri olan Twitter, sunduğu konu çeşitliliği sebebiyle duygu analizi çalışmalarında yaygın olarak kullanılmaktadır. Çeşitli içerik formlarını kullanarak fikirlerin kolayca paylaşılmasını sağlayan bu sosyal medya platformunda kullanıcıların paylaşımları arasında saldırgan dil barındıran içeriklerin önemli oranda arttığı gözlemlenmiştir. Bu durum saldırgan dile maruz kalan kullanıcıların psikolojik iyi oluşlarını olumsuz yönde etkiler. Bununla birlikte sadece insanlar değil kurumsal firmalar da bu saldırgan dile maruz kalmaktadırlar. Bu şiddetin tekrarlanması, bu durumun normalleşme ve daha da artması sonucunu doğurmaktadır. Çalışma Türkçe dilinde bu sorunun otomatik çözümüne katkı sağlamayı amaçlamaktadır. Bu çalışmada Twitter platformundan elde edilen veri seti oluşturulmuştur. 14752 Türkçe tweet metninden oluşan bu veri seti etiketleyiciler tarafından manuel olarak etiketlenmiştir. Etiketlenen verileri sınıflandırma da en popüler derin öğrenme yöntemleri kullanılmıştır. Bu derin öğrenme yöntemleri CNN, LSTM, BiLSTM, GRU, BiGRU ve CNN-LSTM' dir. Bu modellerin sınıflandırma performansları karşılaştırılmıştır. Bununla birlikte makine öğrenmesi yöntemi olan SVM modelinin sınıflandırma performansı da karşılaştırılmıştır. Çalışmada ikili ve çoklu sınıflandırma yapılmıştır. Çalışma saldırgan dil tanımlama ile ilgili Türkçe için çoklu sınıflandırma yapılan ilk çalışmadır. Bunlara ek olarak Twitter platformundan elde edilen 1 milyon 860 bin tweet metninden oluşan genişletilmiş derlem elde edilmiştir. Burada Word2vec yöntemi ile kelime temsilleri elde edilmiştir. Böylelikle genişletilmiş derlem kullanımının sınıflandırma performanslarına katkısı karşılaştırılmıştır. Çalışmada yapılan ikili sınıflandırma da genişletilmiş derlem kullanımıyla en yüksek performans GRU modeli F1-skor değeri %96,65'tir. Bu sebeple çoklu sınıflandırma yapılırken GRU modeli kullanılmıştır. Çoklu sınıflandırmada elde edilen sınıflandırma performans değerleri genişletilmiş derlemin katkısıyla GRU F1-makro değeri %71,97 ve %54,10'dur.

Bilim Kodu : 114608
Anahtar Kelimeler : Duygu analizi, derin öğrenme, saldırgan dil, sosyal medya uygulaması
Sayfa Adedi : 93
Danışman : Prof. Dr. Hadi GÖKÇEN
İkinci Danışman : Dr. Öğr. Üyesi İlyas ÖZER

DETECTING OFFENSIVE CONTENT BY PERFORMING SENTIMENT ANALYSIS IN TURKISH TEXTS

(M. Sc. Thesis)

Şeyma ŞAHİNER YILMAZ

GAZI UNIVERSITY
INFORMATICS INSTITUTE

June 2022

ABSTRACT

Social media channels is growing rapidly, more users share their views and ideas online. After this growth, sentiment analysis studies, which analyze people's ideas, feelings, evaluations, attitudes towards entities and their qualities expressed in written text, have become a popular field of study. However, Twitter, one of the popular social media platforms, is widely used in sentiment analysis studies due to the variety of topics it offers. It has been observed that content containing offensive language has increased significantly among users' posts on this social media platform, which enables easy sharing of ideas by using various content forms. This situation negatively affects the psychological well-being of users who are exposed to offensive language. However, not only people but also corporate companies are exposed to this offensive language. The repetition of this violence results in the normalization and further escalation of this situation. Our study aims to contribute to the automatic solution of this problem in Turkish language. In this study, the data set obtained from the Twitter platform was created. This dataset, consisting of 14752 Turkish tweet texts, was tagged manually by the taggers. The tagged data used the most popular deep learning methods. These deep learning methods are CNN, LSTM, BiLSTM, GRU, BiGRU and CNN-LSTM. The classification performances of these models were compared. In addition the classification performance of SVM model which is a machine learning model was also compared. In our study, binary and multiple classification was made. Our study is the first multi-classification study on offensive language identification. In addition to these, an expanded corpus of 1 million 860 thousand tweets obtained from the Twitter platform was obtained. Here, word representations were obtained with the Word2vec method. Thus, the contribution of the use of extended corpus to the classification performances was compared. Thus, the highest performance GRU model F1-score value is 96.65% with the use of extended corpus in binary classification. For this reason, the GRU model was used while performing multiple classification. The classification performance values obtained, with the contribution of the expanded corpus, are 71,97 % and 54,10 % for the GRU F1-macro value.

Science Code	: 114608
Key Words	: Sentiment analysis, deep learning, offensive language, social media
Page Number	: 93
Adviser	: Prof. Dr. Hadi GÖKÇEN
Co-Adviser	: Assist. Prof. Dr. İlyas ÖZER

TEŞEKKÜR

Tez çalışmam sırasında kıymetli bilgi, birikim ve tecrübeleri ile bana yol gösterici ve destek olan değerli danışman hocam sayın Prof. Dr. Hadi GÖKÇEN'e ve çalışmamın yürütülmesinde ilgi ve desteğini esirgemeyen değerli bilgi ve tecrübelerinden yararlandığım sayın hocam Dr. Öğr. Üyesi İlyas Özer'e sonsuz teşekkür ve saygılarımı sunarım.

Her daim desteğini ve değerli motivasyonlarını yanımda hissettiğim hayat arkadaşım Gökhan YILMAZ'a teşekkürlerimi sunarım. Hayatıma girdiği günden beri her anımı anlamlı kılan neşe kaynağım sevgili oğlum Metehan Baybars YILMAZ'a sevgilerimi sunarım.

Hayatım boyunca maddi ve manevi destekleriyle beni hiçbir zaman yalnız bırakmayan, her anda beni cesaretlendiren, kendimi şanslı hissettiren çok değerli annem Zeynep ŞAHİNER ve babam Hasan ŞAHİNER'e sonsuz teşekkürlerimi sunmaktan mutluluk duyarım.

Son olarak başarılı bir araştırmacı olmam için beni yönlendiren ve motive eden merhum Saygıdeğer hocam, Prof. Dr. Emin ACAR'a sonsuz saygılarımı sunuyorum.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. DUYGU ANALİZİ.....	5
2.1. Duygu Analizine Genel Bir Bakış.....	5
2.2. Duygu Analizi için Yaklaşımlar.....	7
3. DERİN ÖĞRENME KULLANILARAK DUYGU ANALİZİ	13
3.1. Evrişimsel Sinir Ağları (CNN).....	13
3.2. Yinelemeli Sinir Ağı (RNN)	14
3.3. Kapılı Yinelemeli Sinir Ağları (GRNN)	16
3.3.1. Uzun kısa dönem bellek (LSTM) ve kapılı sinir ağı (GRU).....	16
3.4. Performans Ölçüm Yöntemleri	19
4. KELİME TEMSİL YÖNTEMİ.....	23
5. SALDIRGAN DİL VE DUYGU ANALİZİ	27
6. LİTERATÜR TARAMASI	33
7. UYGULAMA.....	43
7.1. Veri Toplama	45
7.1.1. Veri setleri.....	45
7.2. Veri Ön işleme	53

	Sayfa
7.2.1. Kelime temsil yöntemi	54
7.3. Kullanılan Sınıflandırma Algoritmaları	56
8. BULGULAR VE DEĞERLENDİRME	59
8.1. İki Sınıflı Verilerin Sınıflandırma Başarımlarının Değerlendirilmesi	59
8.1.1. İki Sınıflı Verilerde SVM Modeli Başarım Değerlendirilmesi.....	65
8.2. Çok Sınıflı Verilerin Sınıflandırma Başarımlarının Değerlendirilmesi	66
8.2.1. İkinci aşama alt sınıflar bazında başarımların değerlendirilmesi	66
8.2.2. Üçüncü aşama alt sınıflar bazında başarımların değerlendirilmesi	69
9. SONUÇ VE ÖNERİLER	75
KAYNAKLAR	79
ÖZGEÇMİŞ	93

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Hata matris gösterimi.....	20
Çizelge 3.2. Sınıflandırma performansı ölçüm yöntemleri.....	20
Çizelge 7.1. Veri setinde yer alan eşsesli sözcükler	47
Çizelge 7.2. Bağlama bağlı anlam farklılığı gösteren tweetler	48
Çizelge 7.3. Tweet metnini etiketleme örnekleri	52
Çizelge 7.4. Veri setinde yer alan tweet metnlerinin kelime uzunluklarının dağılımı ..	55
Çizelge 7.5. Veri setinde kullanılan 3 kelimenin 25 boyutlu vektör gösterimi	55
Çizelge 8.1. Veri setinden elde edilen Word2vec kelime temsillerini kullanarak iki sınıflı verilerin sınıflandırma başarımları	60
Çizelge 8.2. Genişletilmiş derlemde elde edilen Word2vec kelime temsilleri kullanarak iki sınıflı verilerin sınıflandırma başarımları	61
Çizelge 8.3. Genişletilmiş derlem kullanımının iki sınıflı verileri sınıflandırma performansına etkisi	61
Çizelge 8.4. Genişletilmiş derlem kullanımının modellerin iki sınıflı verilerde kayıp değerlerine etkisi	62
Çizelge 8.5. Genişletilmiş derlemde elde edilen Word2vec kelime temsilleri kullanılarak iki sınıflı verilerde GRU modeli sınıflandırma başarımları ...	65
Çizelge 8.6. SVM modeli sınıflandırma başarımları	66
Çizelge 8.7. İkinci aşama alt sınıflar bazında eğitim ve test verilerinin dağılımı	67
Çizelge 8.8. Kullanılan derin öğrenme GRU modelinin ikinci aşama alt sınıflar bazında çoklu sınıflandırma başarımları	68
Çizelge 8.9. Üçüncü aşama alt sınıflar bazında eğitim ve test verilerinin dağılımı	70
Çizelge 8.10. Kullanılan derin öğrenme GRU modelinin üçüncü aşama alt sınıflar bazında çoklu sınıflandırması.....	72

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Duygu analizi çalışmaları için benimsenen yaklaşımlar	7
Şekil 3.1. Cümle sınıflandırma için kullanılan CNN mimarisi örneği	13
Şekil 3.2. RNN mimarisi.....	15
Şekil 3.3. LSTM modeli.....	18
Şekil 3.4. GRU modeli.....	18
Şekil 4.1. CBOW ve Skip-gram mimarisi (Özer, 2019)	23
Şekil 4.2. Kosinüs benzerliğine göre anlamsal yakınlık	24
Şekil 6.1. "Google Trends" "Sentiment Analysis" çalışma alanını gösteren veriler grafiği	33
Şekil 6.2. (a)Duygu analizi ile ilgili makalelerin başlıklarından oluşan kelime bulutu, (b)yıllara göre makale başlıklarının karşılaştırılması.....	37
Şekil 7.1. Uygulama akış şeması	44
Şekil 7.2. Genişletilmiş derlemde elde edilen kelime bulutu.....	46
Şekil 7.3. Genişletilmiş derlemde elde edilen kelime ağacı modeli	46
Şekil 7.4. 'hakem' kelimesinin anlamsal yakınlıklarının gösterimi.....	48
Şekil 7.5. Tweet metninin etiketlenmesi işlemini gösteren akış şeması	51
Şekil 7.6. Veri setinde etiketleme yapılan tweet sayıları	53
Şekil 8.1. Kullanılan derin öğrenme modellerinin iki sınıflı verilerde hata matris gösterimi (a veri setinden elde edilen hata matrisi, b genişletilmiş derlem kullanılarak elde edilen hata matrisi)	63
Şekil 8.2. (devam) Kullanılan derin öğrenme modellerinin iki sınıflı verilerde hata matris gösterimi (a veri setinden elde edilen hata matrisi, b genişletilmiş derlem kullanılarak elde edilen hata matrisi)	64
Şekil 8.3. Kullanılan derin öğrenme GRU modelinin ikinci aşama alt sınıflar bazında çoklu sınıflandırma için hata matris gösterimi (a veri setinden elde edilen hata matrisi, a' genişletilmiş derlem kullanılarak elde edilen hata matrisi)...	69
Şekil 8.4. Kullanılan derin öğrenme GRU modelinin üçüncü aşama alt sınıflar bazında çoklu sınıflandırma için hata matris gösterimi	73

KISALTMALAR

Bu çalışmada kullanılmış bazı kısaltmalar, açıklamaları ile aşağıda sunulmuştur.

Kısaltmalar

Açıklamalar

Bi-GRU

Çift yönlü kapılı yinelemeli ağ (Bidirectional gated recurrent units)

Bi-LSTM

Çift yönlü uzun kısa dönem bellek (Bidirectional long-short term memory)

CBOW

Kelime çantası (Continuous bag of words)

CNN

Evrişimli sinir ağları (Convolutional neural network)

DDİ

Doğal Dil İşleme

FN

Yanlış negatif (False negative)

FP

Yanlış pozitif (False Positive)

GRNN

Kapılı yinelemeli yapay sinir ağları (Gated recurrent neural network)

GRU

Kapılı yinelemeli ağ (Gated recurrent units)

LSTM

Uzun kısa dönem bellek (Long-short term memory)

TN

Doğru negatif (True negative)

TP

Doğru pozitif (True positive)

1. GİRİŞ

Bireylerin başkalarıyla bağlantı ve iletişim kurması için çevrimiçi sosyal ağ platformları olarak tanımlanan sosyal medya (ör. Twitter) milyarlarca kullanıcıyı cezbetmiştir. Dünyada 7 milyardan fazla insan vardır. Bunların neredeyse 3,5 milyarı aktif sosyal medya kullanıcısıdır (We are Social ve Hootsuite, 2020). Bu sosyal medya platformlarından biri de Twitter’ dır. Kullanıcılara belirli büyüklükteki kelime kullanımı ile paylaşım yapma imkânı sunan bu mikroblog fikri, SMS olarak bilinen mesajlaşma yönteminden esinlenerek Twitter’ ın kurucularından olan Jack Dorsey tarafından ortaya atılmıştır. 2006’ da ortaya atılan bu fikir günümüzde oldukça popüler bir platform olan Twitter platformunun temellerini oluşturmuştur (Sagolla, 2009). Kullanıcılar, Twitter’da (Crannell, Clark, Jones, James ve Moore, 2016) günde ortalama 500 milyon tweet atmaktadır ve atılan bu tweetlerin %80’i bir mobil cihazdan (Carley, Malik, Landwehr, Pfeffer ve Kowalchuck, 2016) paylaşılmıştır. Ayrıca aktif Twitter kullanıcılarının sayısı dünyadaki internet kullanıcılarının %22’ sini aşmaktadır (Kayser ve Bierwisch, 2016). Her yıl sosyal medya istatistikleri yayınlayan We Are Social ve Hootsuite şirketinin hazırladığı “Digital in 2020” raporuna göre Twitter aktif kullanıcı sayısı, 339 milyon 600 bin kişidir. Kullanıcı sayısına göre ülkeler sıralandığında Türkiye, aktif kullanıcı sayısı 11 milyon 800 bin ile 6. sırada yer almaktadır. Raporda Twitter’ ın 13 yaşından başlayarak 50 yaşın üstü dahil her yaş grubundan kullanıcısı olduğu belirtilmiştir (We are Social ve Hootsuite, 2020). Ayrıca bu platformda 31 Aralık 2019 itibarıyla, insanların altı dilde takip edebileceği 1.700’den fazla konu bulunmaktadır (Aslam, 2020). Bu 2020 yılında %150 gibi bir artış göstermiştir. Böylelikle 2020 yılının 3. çeyreğinde takip edilebilecek 5100 konu olduğu görülmüştür (Newberry, 2021). Bu istatistikler, Twitter’ın küresel erişimini ve potansiyel etkisini yansıtmaktadır. Twitter, küresel bir konu kapsamına sahip olmanın yanı sıra, metinler, resimler, bağlantılar da dâhil olmak üzere çeşitli içerik formlarını kullanarak fikirlerin kolayca paylaşılmasını sağlayan bir sosyal medya platformu oluşturmaktadır. Nitekim Twitter, son on yılda on binden fazla araştırma makalesinde kullanılan büyük bir paylaşım platformu haline gelmiştir (Antonakaki, 2020).

Twitter gibi sosyal medya mecralarının hızlı büyümesiyle birlikte daha fazla kullanıcı, görüşlerini ve fikirlerini çevrimiçi olarak paylaşmaktadır. Bu büyümenin ardından, bilgisayar bilimlerinde en verimli araştırma alanlarından biri olan kullanıcı fikirlerini

belirlemeyi ve çıkarmayı amaçlayan duygu analizidir (Cambria, Poria, Gelbukh ve Thelwall, 2017). Duygu analizi çalışmalarında Twitter verileri yaygın olarak kullanılmıştır. Duygu analizi, insanların varlıklara yönelik fikirlerini, duygularını, değerlendirmelerini, tutumlarını ve bunların yazılı metinde ifade edilen niteliklerini analiz eden çalışma alanıdır (Liu, 2015). Fikir, düşünce (opinion) madenciliği olarak da adlandırılan duygu analizi, 2000'li yılların başından beri DDİ (Doğal Dil İşleme)' de en aktif araştırma alanlarından biri olmuştur (Liu, 2012).

Sosyal medya insanlara görüşlerini, bilgilerini, deneyimlerini ve duygularını ifade etmeleri için çevrimiçi bir platform sağlamaktadır. Ancak sosyal medya etkileşimleri taciz edici, küfürlü ve hakaret içeren saldırgan içerikli yorumlar yapılan bir platform haline geldiğinde büyük bir sorun ortaya çıkmaktadır. Bu durumda kullanıcılar saldırgan içeriklere maruz bırakılmaktadır. Dolayısıyla kullanıcılar bu durumdan rahatsız olmaktadır. Saldırgan içeriklere maruz kalan kullanıcılar arasında 11-15 ergen yaş grubu azımsanmayacak sayıda bulunmaktadır (Craig ve diğerleri, 2020). Ergen yaş grubu için saldırgan içerikler psikolojik açıdan geriye dönülmesi mümkün olmayan zararlar vermektedir. Aynı zamanda şirketlerde sosyal medya platformlarında bu tür saldırgan dil ile karşılaşmaktadır. Bu gibi sebepler ile saldırgan içeriklerin tespit edilmesine yönelik çalışmalar, gelecekte bu içeriklerin engellenebilmesi hususunda önem arz etmektedir. Bu çalışmada derin öğrenme yöntemleri kullanarak sosyal medyada saldırgan içeriklerin tespiti alanına katkı sağlamayı planlanmıştır.

Duygu analizinin köklerinin, 20. yüzyılın başlarında kamuoyu analizi üzerine yapılan çalışmalarda ve 1990'larda hesaplamalı dilbilim topluluğu tarafından gerçekleştirilen öznel metin analizi olduğu bilinmektedir (Mäntylä, Graziotin ve Kuuttila, 2018). Bununla birlikte, bilgisayar tabanlı duygu analizi çalışmaları yalnızca Web' deki öznel metinlerin mevcudiyetiyle gerçekleşti ve bu makalelerin %99'u 2004'ten sonra yayınlanmıştır (Mäntylä ve diğerleri, 2018). Böylelikle öznel metinler barındıran Twitter gibi sosyal medya uygulamalarının popülerleşmesi ile çalışmalar hem oldukça hız kazanmış hem de bu uygulamaların sunduğu konu bolluğu sayesinde birçok dalda duygu analizi çalışması yapılmıştır. Bu alandaki çalışmalara örnek verecek olursak: Hisse senedi piyasaları, seçimler, afetler, tıp, yazılım mühendisliği ve siber zorbalık gibi alanların da ötesinde çalışmalar mevcuttur. Bugün Google Akademi'de "sentiment analysis" olarak çalışmaları arattığımızda 2004 yılından itibaren günümüze kadar bu başlıkta 1 milyon 290 bin çalışma

olduğu görülmektedir. Bu çalışmalarının çoğunluğu İngilizce dilinde yapılmıştır. İngilizce, SentiWordNet (Esuli ve Sebastiani 2006) ve SenticNet (Cambria, Olsher ve Rajagopal, 2014) gibi en zengin duygu analizi kaynaklarına sahiptir. Bununla birlikte, sosyal medya farklı dillerin konuşulduğu diğer birçok yerde çoğaldıkça, bu dillerde duygu analizi yapılması güçlü bir talebe sebep olmuştur. Ancak Twitter aktif kullanıcı sayısı itibarıyla dünyada 6. Sırada olan Türkçe dilinde duygu analizi üzerine yapılan çalışma İngilizce diline nispeten oldukça azdır. Bu çalışmaların çoğu ikili (pozitif ve negatif) sınıflandırma üzerine odaklanmıştır.

Türkçe için Senti-TurkNet (Dehkharghani, Saygin, Yanikoglu ve Oflazer, 2016) gibi kaynaklar ve 2.000 kelimeden oluşan bir kelime kümesi (Dehkharghani, Yanikoglu, Tapucu ve Saygin, 2012) oluşturulmuştur. Ancak Türkçe'nin zengin morfolojik yapısı ve çok sayıda kelime oluşumunu sağlayan ek yapısı sebebiyle bu yaklaşımlar önerilmemektedir (Özer, 2019). Ayrıca sosyal medyada kullanılan ifadeler, dil bilgisi kurallarından uzak, yazım hataları, sözlükte yer almayan kelimeler ve kısaltmalar da oldukça fazladır. Bu durum duygu analizinde sınıflandırma problemlerine sebep olmuştur. Önerilen yöntem ise alana özgü verilerin kullanılması ile problemi çözmektir. Böylelikle Türkçe metinlerin sınıflandırılmasında başarılı sonuçlar elde edilmiştir (Özer, 2019). Biz de elde edilen bu sonuçlardan yola çıkarak araştırmalarımıza bağlı kalacak şekilde alana özgü verilerin olduğu iki ayrı veri seti kullandık. Bunlardan biri alana özgü verilerden oluşan genişletilmiş derlemidir. Diğeri ise sınıflandırma performansını ölçmek amacıyla kullanılan etiketlenmiş veri setidir. Çalışmada bu veri setleri kullanılarak word2vec-CBOW ile her bir tweet metninde yer alan kelimenin vektörel temsili elde edilmiştir. Bununla birlikte CNN, LSTM, Bi-LSTM, GRU, Bi-GRU gibi güncel etkinliği kanıtlanmış derin öğrenme yöntemleri ile sınıflandırma işlemi yapılmıştır. Bu yöntemlere ek olarak CNN-LSTM hibrit modeli kullanılmıştır. Ayrıca her birinin sınıflandırma performansları karşılaştırılmıştır. Sınıflandırma performansını değerlendirme ölçümleri kesinlik, doğruluk, duyarlılık ve F-skor olarak karar verilmiştir. Bu performans ölçümleri literatürde bu alanda kanıtlanmış ölçüm yöntemleridir.

Bu çalışmada, saldırgan dili etiketlemek için Zampieri ve arkadaşlarına benzer genel bir yaklaşım izlenmiştir. (Zampieri ve diğerleri, 2019; Wiegand, Siegel ve Ruppenhofer, 2018). Tanımlamalarımız her iki çalışmadan farklılıklar gösterse de her iki çalışmada olduğu gibi sınıf etiketleri için hiyerarşik bir şema izlenmiştir. Bu şemanın ilk aşamasında tweet metni

“saldırgan” ya da “saldırgan değil” olarak etiketlenmiştir. Eğer metin “saldırgan” olarak tanımlanırsa, “hedefli” ve “hedefsiz” olarak etiketlenmektedir. Çalışmada “hedefsiz” olarak tanımlananlar net bir hedef olmaksızın her türlü saldırgan dil kullanımı içermektedir. Bu durum özellikle çocuklar için uygunsuz olarak nitelendirilmektedir. Eğer metin “hedefli” olarak tanımlanırlar ise grup, birey ve diğer olmak üzere etiketleme yapılmıştır. Gruba yönelik saldırgan dil toplulukları hedef almıştır. Grup tanımına uymayan bir kişiye yönelik işlenen suç, bireyleri hedef alan bir suç olarak kabul edilmektedir. Bireye karşı yapılan saldırgan dil ise genellikle siber zorbalık, cinsel taciz suçunu içermektedir. Siber zorbalık genellikle çocuklara karşı işlenmekte ve birey üzerindeki etkileri de oldukça olumsuz olmaktadır. Cinsel taciz suçu ise cinsel arzu ve isteklere yönelik rahatsız edici sözleri kapsar (Sarmaşık, 2009). Suçun hedefinin bu kategorilerden hiçbirine uymadığı durumlar vardır. Bir organizasyon veya olay gibi insan olmayan bir varlığa yönelik suç buna örnek verilebilir. Bu gibi durumlarda hedefi “diğer” olarak kabul edilmiştir. Kısaca saldırgan ve saldırgan değil olarak yapılan ikili sınıflandırma, ardından tweet metni saldırgan ise hedefli ve hedefsiz olarak ikili sınıflandırma, en son hedefli sınıflandırmalar Birey, Grup ve diğer olarak üçlü sınıflandırma şeklinde veri seti etiketlenmiştir. Yapılan etiketleme hakkında bilgiler bölüm 7.’de ayrıntılı olarak işlenmiştir.

2. DUYGU ANALİZİ

Bu başlık altında duygu analizine genel bir bakış ve duygu analizinde kullanılan yaklaşımlar incelenecektir.

2.1. Duygu Analizine Genel Bir Bakış

Sosyal medya araçlarının hayatımızda etkin kullanımıyla beraber Web' deki verilerin yoğunluğu artmıştır. Bu artış veri tabanları ve veri ambarlarının kullanım kapasitesinin ötesine geçmiştir. Web üzerindeki hacimli bilgilerin erişilebilirliği düşünüldüğünde, bir insanın verileri manuel olarak analiz etmesi zor ve hataya açık bir şekilde olmaktadır. Bu nedenle web ortamında oluşan içeriği otomatik olarak analiz etmeye ihtiyaç duyulmuştur. Yapılan bu analiz, verilerdeki kalıpları keşfederek büyük miktarda veriden bilginin çıkarılması olan veri madenciliği kavramını ortaya çıkartmıştır (Manda, 2019). Veri Madenciliği, işletmelerin öngörüyle harekete geçmesini sağlamaktadır. Böylelikle bilgi odaklı kararlar almasına olanak tanıyan gelecekteki eğilimler ve davranışlar, veri madenciliği araçları kullanılarak tahmin edilmektedir. Veri madenciliği tarafından sunulan ileriye dönük analizler, otomatik bir şekilde yapılmaktadır. Buna ek olarak veri madenciliği ile yapılan analizler karar destek sistemlerinin geriye dönük araçları tarafından sağlanan geçmiş olayların analizlerinin ötesine geçmektedir. Veri madenciliği araçları, geleneksel olarak çözülmesi zaman alan iş sorularını yanıtlamamıza yardımcı olmaktadır. Böylelikle çoğu şirket büyük miktarda veriyi önceden toplayıp iyileştirmektedir (Kroeze, Matthee ve Bothma, 2004). Metin madenciliği ise, veri madenciliğinden biraz farklı olarak bilinmeyen bilgileri keşfeden veri madenciliğinin bir varyasyonu olmaktadır. Metin madenciliği, kalıpların doğal dil metninden çıkarılmasını içermektedir (Kroeze ve diğerleri, 2004). Duygu analizi ise, metin madenciliğinin bir alt alanı olan fikir madenciliği olarak da bilinir. İnsanların bir varlığa veya bir konuya karşı duygularını tahmin etmenin bir yolu olarak tanımlanmaktadır (Liu ve Zhang, 2012). Tahmin analizleri farklı yöntemler ile yapılabilmektedir. DDİ yöntemlerini (sözlük uygulamaları ve bazı makine öğrenimi tabanlı yaklaşımları) içermektedir. Web' de kullanıcı tarafından oluşturulan içeriğin çoğalması, muhtemelen 2000 yılından itibaren aktif araştırmalar ile başlamaktadır (Liu ve Zhang, 2012). Duygu analizi çalışmaları ile ilk olarak, hesaplamalı dilbilimde anketler ve öznellik analizi yoluyla 20. yüzyılın başlarında kamuoyunu anlama girişimlerinde bulunulmuştur. Ancak Web' de bulunan büyük miktarda fikir, modern duygu analizini doğurmuştur. Modern duygu

analizi çalışmalarının 2005 ile 2016 yılları arasında 50 kat büyüdüğü görülmektedir (Mäntylä ve diğerleri, 2018).

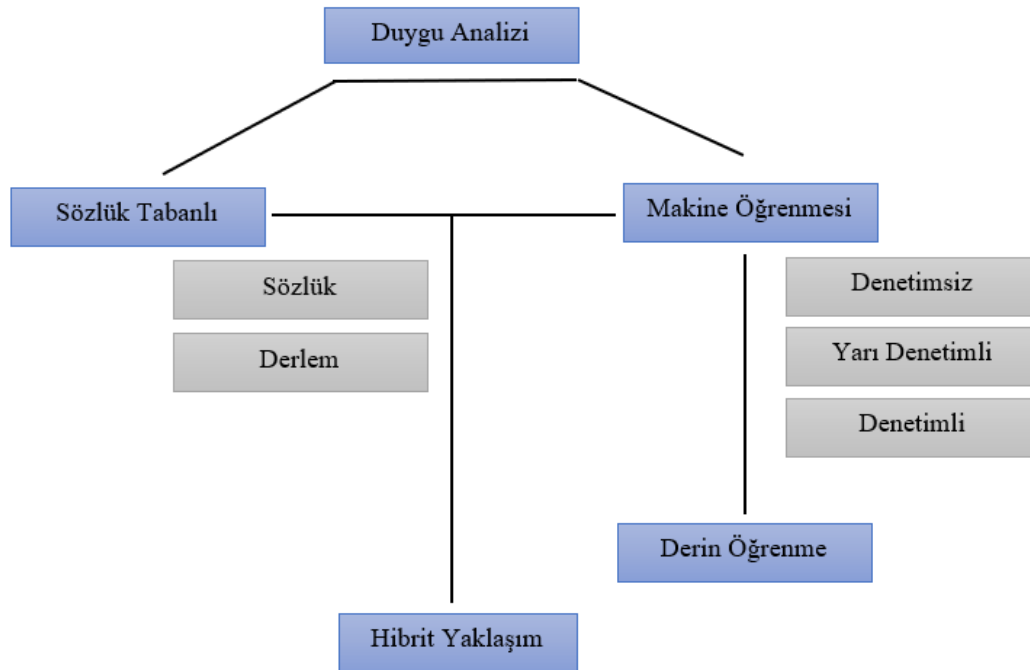
Standart duygu analizi, bir metindeki her kelimeyi ayrı ayrı okuyarak anlamsal yönelimine göre sınıflandırır. Daha sonra her kelime kutupluluğunun birbirini nasıl etkilediğini analiz eder. Yapılan bu analiz sonucunda ise metnin tamamının bulunduğu kelime kutupluluğu toplanarak duygu skoru elde edilmiştir (Pandey, 2018). Duygu analizi bilgisayarlar için çözülmesi zor bir sorun olarak kabul edilmektedir. Çünkü metinler temelde yatan birkaç düşünceden oluşabilir ve bu nedenden ötürü bilgisayarın anlamsal yönelimini başarılı bir şekilde sınıflandırmak için amaçlanan mesajı anlaması gerekmektedir (Pandey, 2018). Bir duygu analizi gerçekleştirilirken kural tabanlı veya otomatik yaklaşım olmak üzere iki yaklaşım kullanılmaktadır. Kurala dayalı yaklaşım biçiminde, metin verilerindeki kutupluluğu belirlemek için bir dizi dil kuralı kullanılmaktadır. Bu yaklaşım biçimi genellikle, anlamsal yönelimine göre olumlu, olumsuz veya nötr olarak etiketlenmiş bir sözcük listesinden oluşan bir sözlük aracılığıyla uygulanmaktadır. Bir diğer yaklaşım olan otomatik yaklaşım ise tam tersine, kurallara dayanmaz, bunun yerine makine öğrenimi tekniklerini kullanmaktadır. Otomatik bir yaklaşım olan duygu analizi, bir sınıflandırma problemi olarak modellenmektedir. Duygu analizi modeli sınıflandırılmış metin verileriyle beslenerek çalışmakta ve böylece modeli, belirli bir girdiye karşılık gelen bir duyguyla ilişkilendirmek için eğitmektedir (Feldman, 2013).

Genel olarak duygu analizi; belge seviyesi, cümle seviyesi ve özellik seviyesi olmak üzere üç farklı seviyede gerçekleştirilmektedir. Twitter verilerinin incelendiği duygu analizinde, tweet metninin uzunluk kısıtı nedeniyle, cümle düzeyinde analiz-yapılmaktadır. Belge düzeyindeki duygu analizi sırasında, bulgular, belgenin tamamında bulunan duyguları temel almaktadır. Ana görev, belgenin olumlu mu yoksa olumsuz mu olduğunu bulmak olacaktır. Genel olarak, belge düzeyinde duygu analizinde temel varsayım, görüş sahibinin bir seferde yalnızca bir hedef hakkında görüş bildirmesidir. Bu tür bir sınıflandırma, bazı veri setleri için uygun ve geçerli olmayabilir. Her zaman tek bir fikir hedefimiz olabileceği doğru değildir (Liu ve Zhang, 2012). Duygu analizinin cümle düzeyinde yapılan araştırmalarda; her cümle pozitif, negatif veya nötr olarak sınıflandırılır. Bu tür analizde, cümleler kutupluluğuna göre etiketlenir. Genel olarak, cümleler pozitif veya negatif olarak etiketlenirken, nötr cümlelerin kutupluluğunun olmadığı varsayılmaktadır (Liu ve Zhang, 2012). Duygu analizinin özellik seviyesinde ise, farklı varlıklar hakkında düşünceler içeren

bir cümle olasılığı vardır. Örneğin, şu ifadeyi ele alalım: "Bu filmi seviyorum ama müzik kötü!" Bu açıklamadan filmle ilgili hissiyatın olumlu ama müziğe karşı olumsuz olduğu söylenebilir. Film ve müzik, duyguya dahil olan iki varlık / yöndür. Bu durumu, daha ince duygu analizinde kullanılan "aspect-based sentiment analysis" olarak düşünebiliriz (Liu ve Zhang, 2012). Duygu analizi çalışmasında amaçlanan şey, duyguyu iletirken kelimelerin birbirini nasıl etkilediğini anlamaktır. Bu da duygu analizinin zorluklarından biridir. Bir diğer zorluk ise gayri resmi dil kullanımıdır. Sosyal medyadaki tüm yazılı metinlerin yaklaşık yüzde 97'si gayri resmi kelimeler içermektedir (Thelwall, 2009). Bu sebeple duygu analizi çalışmaları yapan şirketlerin, kurumların ve araştırmacıların gayri resmi kullanılan dil gibi yapılandırılmamış metin verilerini otomatik olarak analiz etmelerini sağlamaktadır ve bu alandaki gelişimin sürekliliği büyük önem taşımaktadır (Pandey, 2018).

2.2. Duygu Analizi için Yaklaşımlar

Sözlük tabanlı metot ve makine öğrenimi metodu, duygu analizine yönelik iki ana yaklaşımdır (Feldman, 2013). Bu başlıkta Şekil 2.1' de gösterilen yaklaşımlar incelenmiştir.



Şekil 2.1. Duygu analizi çalışmaları için benimsenen yaklaşımlar

2.2.1. Sözlük tabanlı yaklaşım

Sözlük tabanlı yaklaşım, bir belgenin duygusunu yansıtan kelimelerin, cümlelerin veya metinlerin polaritesinden ve gücünden hesaplanmasını içermektedir (Taboada, Brooke, Tofiloski, Voll ve Stede, 2011). Araştırmacılar, genellikle kelimelerin önceden polariteye sahip olduğunu varsaymaktadır. Yani kelimelerin, kullanıldıkları bağlam içindeki kullanımlarından bağımsız olarak içsel bir polariteye sahip olduğu varsayılır. Sınıflandırma, önceden tanımlanmış metin ve cümlelerin örneklerinden sınıflandırıcılar oluşturmayı içermektedir (Turney, 2002). Bu örnekler, mevcut kelime listesini genişletmek için önceden polariteye sahip olduğu varsayılan kelimelerin (pozitif veya negatif ilişkilere sahip kelime grupları) kullanılmasıyla manuel veya otomatik olarak oluşturulmuş sözlükleri içermektedir (Turney ve Littman, 2003). Çeşitli araştırmacılar tarafından kullanılan popüler sözlükler sıfatlara, zarflara, isimlere ve fiillere dayanmaktadır. Kullanılan popüler duygu sözlüklerinden bazıları Word Net ve General Enquirer' dır (Aldawi, 2015). Ancak bir kelime veya cümlenin içsel bir polariteye sahip olduğu genel varsayımı, bu yaklaşımın bir dezavantajıdır. Çünkü sosyal ağdaki gönderilerin çoğu gayri resmidir ve kişinin kendi yazım hatasından kaynaklı veya kendi kısaltmalarını oluşturması muhtemel olan bu alanda sözlükte olmayan kelimeler ile karşılaşma ihtimali oldukça yüksektir. Buna ek olarak kullanıldıkları bağlama göre sözcükler içsel polaritesinden ayrı farklı kutuplara sahip olabilmektedir. Bu sebeple bağlama dayalı yaklaşımlar bazı çalışmalarda ihtiyaç halini almıştır. Bu çalışmalara, morfolojik açıdan zengin bir dil ile yapılan çalışmalar örnek verilebilir. Bu durumda derlem tabanlı yaklaşım önerilmektedir. Derlem tabanlı yaklaşım ile sözcükler belirli bir alana uyarlanmaktadır. Bu yaklaşım ile genel amaçlı duygu sözcüklerinin listesi oluşturulur. Bu listeye uygun verilerden oluşan derlem elde edilmektedir. Böylelikle birlikte meydana gelen sözcük kalıplarına dayalı bir etki alanı oluşmaktadır ve derlem sayesinde diğer duygu sözcükleri keşfedilmektedir.

Türkçe için sözlük tabanlı yaklaşım ile İngilizce kadar gelişmiş olmasa da bu alanda çalışmalar yapılmıştır. Ancak Türkçe gibi eklemeli ve zengin morfolojik yapısı olan diller için, sözlük tabanlı yaklaşımlar zordur. Çünkü oldukça hacimli sözlüğe ihtiyaç duyar. Ayrıca morfolojik analiz sonucunda uygun olan kelimeler, daha önce hiç kullanmadığımız kelimelerden oluşabilmektedir (Örneğin; koltuklulukçuluk) (Özer, 2019). Bununla birlikte sosyal medyada kullanılan dil, Türkçe dil bilgisi kurallarından uzak bir şekilde ve yazım hataları da azımsanmayacak oranda fazla olmaktadır. Ayrıca Türkçe sözlükte yer almayan

Mite-Baidal ve diğerleri, 2018) kelimeler ve kısaltmalar mevcuttur. Örneğin; Türkçe’ de yanlış yazılmış “anadın” kelimesinin 20’den fazla kelime ile benzerliği görülmüştür (Özer, 2019). Sonuç olarak uygun kelimeyi bulmak adına bağlama bakmak bir ihtiyaç olmaktadır. Bu sebeple çalışmada derlem tabanlı yaklaşım kullanılmıştır.

2.2.2. Makine öğrenmesi yaklaşımı

Yapay Zekanın (Artificial Intelligence (AI)) öğrenmeye odaklanan bir dalı olan makine öğrenimi, birçok özelliğe dayalı olarak daha iyi tahmin sağlayabilen pratik bir yaklaşımdır (Klompenburg, Kassahun ve Catal, 2020). Makine öğrenimi, kalıpları ve korelasyonları belirleyebilmektedir. Buna ek olarak veri kümelerinden bilgileri keşfedebilmektedir (Klompenburg ve diğerleri, 2020). Bu modellerin, geçmiş deneyime dayalı olarak temsil edildiği veri kümeleri kullanılarak eğitilmesi gerekmektedir.

Makine öğrenimi yöntemleri genellikle denetimli ve denetimsiz öğrenim şeklinde sınıflandırılır. Denetimsiz öğrenme yöntemleri, sınıflandırma için eğitim veri setini kullanmazken, denetimli öğrenme yöntemleri algoritmalarını kullanan sınıflandırma ise, eğitim aşaması ve doğrulama için kullanılan bir test aşaması olan iki ana aşamayı içermektedir (Mejova ve Srinivasan, 2011). Mevcut araştırmalar hem denetimli hem de denetimsiz yöntemleri içeren çeşitli duygu analizi görevleri için çok sayıda teknik üretmiştir. Araştırmacılar her tür denetimli makine öğrenimi yöntemlerini (Örn. SVM, Naive Bayes vb.) ve bu yöntemlerin özellik kombinasyonlarını kullanmıştır. Denetimsiz yöntemler, duygu sözlüklerinden, dil bilgisel analizden ve söz dizimsel kalıplardan yararlanan çeşitli yöntemleri içermektedir. Bu yöntemleri ve uygulamaları kapsamlı bir şekilde ele alan birkaç makale yayınlanmıştır (Liu, 2012; Liu, 2015; Pang ve Lee, 2008).

Denetimsiz Öğrenme

Denetimsiz yöntemler, makine öğrenimine veya sözlüğe dayalı olabilir. Etiketlenmiş veri seti, denetimsiz öğrenmede gerekli değildir. Bu yöntem kullanılarak makineye sadece girdi veri seti sağlanır. Model, etiketleme gerektirmez. Duygu analizine yönelik denetimsiz bir yaklaşım genellikle bir duygu Sözlüğüne dayanmaktadır (Kharde ve Sonawane, 2016). Bu yöntemde metin sınıflandırması, çeşitli algoritmaların bir cümlelinin anlamsal yönelimini

tahmin etmek için sıfatlar veya zarflar içeren cümleleri çıkardığı semantik yönelim yaklaşımı kullanılarak yapılmıştır (Manda, 2019).

Denetimli Öğrenme

Sözlük bazlı yaklaşımların dilbilimsel olarak kabul görmesine ve uygulama kolaylığına rağmen bu tür yaklaşımların eleştirildiği taraflar olmuştur. Eleştirilen kısımlardan bir tanesi esnek sınıflandırma biçimi sağlayamamasından kaynaklıdır. Çünkü sözcük girişleri, önceden hesaplanmış sabit katsayıları baz almaktadır. Ancak bağlama göre birçok terim farklı yoğunluk ve polarite değerlerine sahip olmaktadır ve bu durum analiz sonuçlarının optimalin altında kalmasına yol açmaktadır (Yang ve Eisenstein, 2017). Sözlük bazlı yaklaşımın eleştiri almasının bir diğer sebebi ise, sözcüksel olmayan durumları karara dâhil edememesinden kaynaklanmaktadır. Oysaki ünlem işareti, yazım varyasyonları cümlelerin duygusu ile doğrudan ilişkili olabileceği düşünülmektedir. Ayrıca Türkçe gibi dillerde sözlük boyutunun oldukça fazla olması gerektiği sebebiyle kullanımı zor olmaktadır. Bunun bir sonucu olarak, manuel olarak etiketlenmiş veri setinin ortaya çıkmasıyla birlikte, sözlüğe dayalı sistemlerin yerini yavaş yavaş denetimli makine öğrenimi teknikleri almıştır. Denetimli öğrenme, bir sınıflandırıcıyı eğitmek için iyi etiketlenmiş ve yeterli büyüklükte bir veri setine ihtiyaç duyar, aksi takdirde eğitim ve testlerin gerçekleştirilmesi zor olmaktadır.

2.2.3. Derin öğrenme metodu

Yaklaşık on yıl önce, derin öğrenme güçlü bir makine öğrenimi tekniği olarak ortaya çıkmıştır (Goodfellow, Bengio ve Courville, 2016). Duygu analizine derin öğrenmenin uygulanması da son zamanlarda oldukça popüler hale gelmiştir. Derin öğrenme, yapay sinir ağlarının birden çok katmandan oluşan ağları kullanarak öğrenme görevlerine uygulanmasıdır (Zhang, Wang ve Liu, 2018). Biyolojik beynin yapısından esinlenen sinir ağları, birlikte çalışan katmanlar halinde organize edilmiş çok sayıda bilgi işlem biriminden (nöronlardan) oluşmaktadır (Zhang, Wang ve Liu, 2018). Sinir ağları biyolojik bir beynin öğrenme sürecine benzer şekilde nöronlar arasındaki bağlantı ağırlıklarını ayarlayarak görevleri (örneğin, sınıflandırma) gerçekleştirmeyi öğrenebilmektedir. Derin öğrenme, makine öğreniminin bir alanıdır ve derin sinir ağları, özyinelemeli sinir ağları RecNN, evrişimli sinir ağları CNN, tekrarlayan sinir ağları RNN gibi birçok modele sahiptir (Sarkar

ve Sharma, 2018). . Derin öğrenme teknikleri sensör tabanlı bilgisayarlı görü, DDİ, ses tanıma, konuşma tanıma gibi alanlarda uygulanmıştır. Bu tür uygulamalar, uzmanlara kıyasla çarpıcı ve bazı durumlarda onlardan daha iyi başarılar elde etmiştir (Shahin ve Hamsa, 2020).

DDİ ve derin öğrenme arasındaki iş birliği, birçok önemli görevde olağanüstü başarılar göstermiştir. Geleneksel makine öğrenimi tekniklerinde, özellikler insanlar tarafından tasarlanır ve bu teknik insan deneyimi gerektirdiği için derin öğrenmeye göre daha zor olmaktadır. Derin öğrenme, derin katmanlı bir model yapısı kullanarak bu sorunu ele almıştır (Interian ve diğerleri, 2018).

Derin öğrenme modelinde iki tür veri seti bulunur, bu veri setlerinden biri olan eğitim veri seti, modeli eğitmek için kullanılmaktadır. Girdiye yönelik çözümler zaten bilindiği için, model bu örneklerden öğrenmektedir. Böylece beklenen çıktıları verebilir. Diğer bir veri seti olan test veri seti, modeli test etmek ve modelin eğitim veri kümesinden sınıflandırmayı ne kadar iyi öğrendiğini kontrol etmek için kullanılan örnekler kümesidir (Manda, 2019). Bu girdilere yönelik çözümler zaten bilindiği için model, eğitim veri setinden farklı yeni veriler üzerinde test edilirken istenen tahminleri verip vermediğini kontrol etmek için bu örnekler kümesi kullanılmaktadır. Doğru sınıflandırılmış örneklerin oranı, test verileri olarak kullanılan örneklerin toplam sayısı test doğruluğunu ortaya koymaktadır. Böylelikle uygulanan derin öğrenme modellerinin veri seti üzerindeki performansı ölçülmektedir.

2.2.4. Hibrit yaklaşım

Literatürde farklı hibrit yaklaşımlar bulunmaktadır. Bazıları, sözlük tabanlı bilgi ile makine öğrenimi modellerini genişletmeyi amaçlamaktadır (Behera, Roy, Dash, 2016). Amaç hem sözlük hem de makine öğrenimi tabanlı tekniklerin etkili bir özellik kümesini kullanarak en iyi sonuçları elde etmek için her iki yöntemi birleştirmektir (Ahmad, Aftab, Ali, Hameed, 2017). Bu şekilde, her iki yaklaşımın eksiklikleri ve sınırlamaları giderilmektedir. Bununla birlikte farklı modellerin birlikte kullanımı da hibrit yaklaşıma örnek olmaktadır. Örneğin; CNN-LSTM algoritmalarının birlikte kullanımında bireysel modellerden daha iyi performans sağladığı görülmüştür. Ayrıca bu algoritmanın cümle düzeyinde duygu sınıflandırmasında başarısı kanıtlanmıştır (Ligthart, Catal ve Tekinerdogan, 2021).

Duygu analizi yaklaşımlarının karşılaştırılması

Sözlük tabanlı yaklaşımlar gibi geleneksel yaklaşımlar, zaman alıcı ve sıkıcı bir süreç olan el yapımı özellikleri kullanmaktadır. Ayrıca, bazı alanlar için iyi genelleme yapamazlar. Geleneksel makine öğrenimi yaklaşımlarında özellik mühendisliği ile özellik çıkarma çok zaman alan işlemlerdir. Bu nedenle, derin öğrenme, ağ öğrenirken özellik tasarımının yükünü azaltmaktadır, sınıflandırma işlemi için gerekli özellikleri otomatik olarak oluşturmaktadır.

Şu anda, sosyal ağlarda büyük miktarda veri üretilmektedir. Twitter’da, saniyede ortalama 6000 tweet üretilmektedir. Bu da yılda yaklaşık 200 milyar tweet üretildiği anlamına gelmektedir. Bu nedenle, bu kadar büyük miktarda veriyle, geleneksel makine öğrenimi tabanlı yaklaşımlar performans gösterememektedir (Yadav ve Vishwakarma, 2020). Buna karşın derin öğrenme modelleri ise büyük veri kümeleriyle daha fazla özellik öğrenmek amacıyla eğitilebildikleri için makine öğrenimi yaklaşımlarından daha iyi performans göstermektedir (Ligthart ve diğerleri, 2021). Bu, derin öğrenme modellerinin geleneksel makine öğrenimi modellerinden daha iyi performans gösterdiğini doğrulamaktadır.

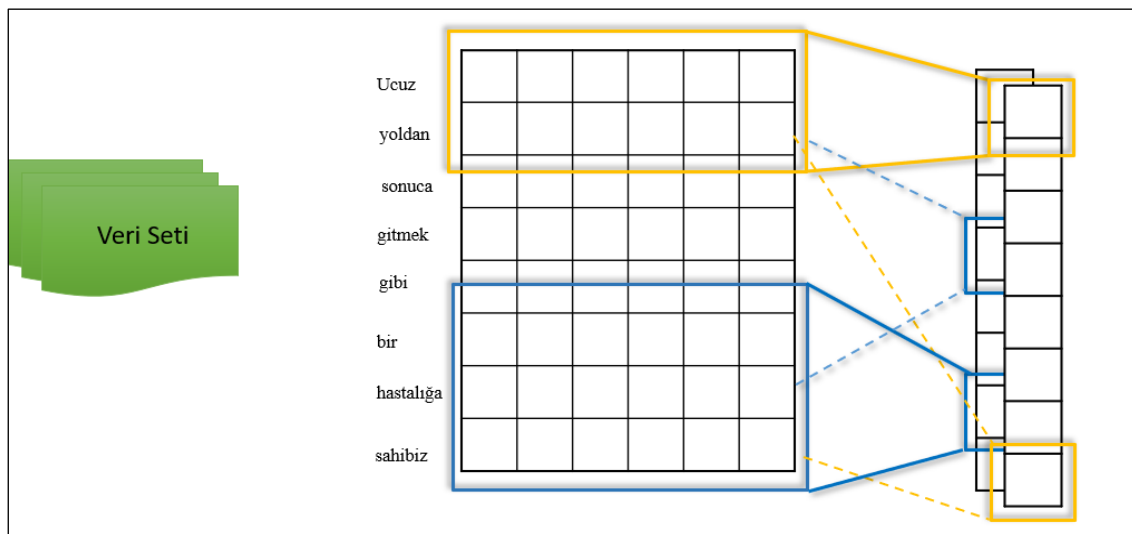
3. DERİN ÖĞRENME KULLANILARAK DUYGU ANALİZİ

Bu başlıkta Evrişimli Sinir Ağları, Yinelemeli Sinir ağları ve Kapılı Yinelemeli Sinir Ağları başlıkları incelenecektir. Buna ek olarak bahsedilen derin öğrenme yöntemlerinin karşılaştırılmasında kullanılan performans ölçüm yöntemlerinden bahsedilecektir.

3.1. Evrişimsel Sinir Ağları (CNN)

Evrişimsel Sinir Ağları (CNN) başlangıçta görüntü işleme alanında kullanılan özel ileri beslemeli sinir ağı türüdür (Zhang ve diğerleri, 2018). Tasarımı, beyindeki görsel bir mekanizma olan görsel korteksten esinlenmiştir. Görsel korteks, alıcı alanlar olarak adlandırılan, görsel alanların küçük ve üst üste binen alt bölgelerinde ışığı tespit etmekten sorumlu birçok hücre içermektedir. Bu hücreler, giriş alanı üzerindeki yerel filtrelerdir. Böylelikle CNN, her biri görsel kortekste hücreler tarafından işlenen işlevi yerine getiren çok sayıda evrişimli katmandan oluşmaktadır (Zhang ve diğerleri, 2018).

En popüler derin mimari türlerinden biri evrişimsel sinir ağlarıdır (CNN) (Shahin, Nassif ve Hamsa, 2019). Duygu analizi gibi NLP görevlerinde CNN'nin uygulanması etkili sonuçlar göstermiştir. Öncü olan Collobert ve diğerlerinin (Collobert ve diğerleri, 2011) ve Kim' in (Kim, 2014) yaptığı çalışmalar buna örnektir. Buna ek olarak CNN'in girdiden "bilgilendirici gizli anlamsal temsil" oluşturmaya olanak sağlayan n-gram özelliği ile etkin sonuçlar elde edilmiştir (Zhang ve diğerleri, 2018; Rojas-Barahona, 2016).

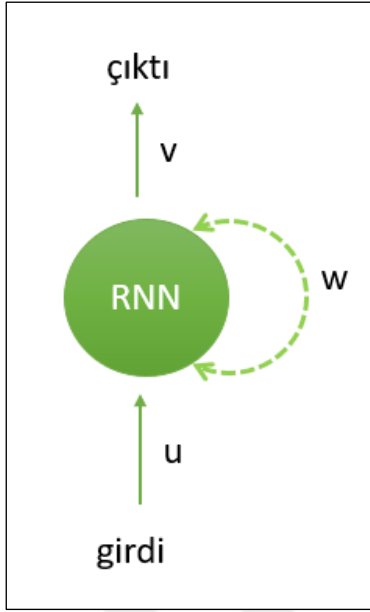


Şekil 3.1. Cümle sınıflandırma için kullanılan CNN mimarisi örneği

CNN, başlangıçta görüntü işleme çalışmalarında kullanılsa da duygu analizi çalışmalarında da oldukça başarılı sonuçlar elde etmiştir (Kim, 2014). CNN de bu görev için, girdi genellikle word2vec-CBOW gibi kelime gömme kullanılmaktadır. Böylelikle girdi görüntü pikselleri yerine bir matris olarak temsil edilmektedir. Şekil 3.1' de cümle sınıflandırması için CNN mimarisinin bir örneği gösterilmiştir. Ayrıca CNN önceden etiketlenmiş verilere ihtiyaç duymaktadır. Buna göre bir sınıflandırma yapmaktadır. Buna ek olarak Lighthart ve arkadaşları tarafından duygu analizine yönelik yapılan literatür çalışmasında en çok alıntı yapılan makaleler arasında CNN ile ilgili yapılan çalışmanın ilk sırada olduğu görülmüştür. Bununla birlikte aynı çalışmada 2020 yılında derin öğrenme alanında yapılan makalelere bakıldığında CNN'in LSTM den sonra en çok kullanılan model olduğu da görülmüştür (Lighthart ve diğerleri, 2021).

3.2. Yinelemeli Sinir Ağı (RNN)

Yinelemeli Sinir Ağı (RNN), nöronlardan oluşan derin öğrenme modelidir. Her bir nöron, önceki girişle ilgili bilgileri depolamak için kendi dahili belleğini kullanabildiğinden, sıralı verilerde model olarak oldukça faydalıdır. Aynı zamanda RNN, tüm hesaplamaları yakalayan veya şimdiye kadar olan her şeyi ezberleyen bir ağıdır denilebilmektedir. Bu sebeple etkin bir yöntemdir ancak uzun vadeli durumları tahmin etmekte problem yaşamaktadır. Diğer sinir ağları gibi RNN giriş, gizli ve çıktı katmanından oluşmaktadır (Şekil 3.2).



Şekil 3.2. RNN mimarisi

Metinsel verilerin analizi için derin sinir ağları önerilmiştir. Bunlar temel olarak ya kelime yerleştirmeyi öğrenmeye ya da öğrenilen özellik vektörleri üzerinde sınıflandırma ve kümeleme gibi makine öğrenimi görevlerini yerine getirmeye odaklanmaktadır (Zhang ve diğerleri, 2018). Geniş derin ağ türleri arasında, evrimsel sinir ağı (CNN) ve yinelemeli sinir ağı (RNN) metin işlemeyle ilgili çalışmalarda daha yaygın iki modeldir (Zhang ve diğerleri, 2018). Bu popülerliğin nedeni, CNN'lerin yerel kalıpları öğrenme yeteneği ve boyutsallık azaltma özelliğidir. RNN (LSTM)' in ise sıralı modellemedeki ve kelime bağımlılıklarını korumadaki gücü olmaktadır. RNN, birçok metin işleme uygulaması için uygun olsa da giriş verilerinde uzun vadeli bağımlılıklar olduğunda kaybolan gradyan problemi yaşamaktadırlar (Song, Park ve Shin, 2019). Bu tür bağımlılıkların çoğu NLP uygulamasında ve özellikle duygu analizinde yaygındır. Bu sorunu çözmek için “Kapılı Yinelemeli Sinir Ağları (GRNN)” modelleri önerilmiştir. Bunlar “Uzun Kısa-Dönem Belleği (LSTM)” ve “Kapılı Yinelemeli Ünite (GRU)” ağlarıdır. LSTM, sorunu giriş, unut ve çıkış kapıları aracılığıyla ele alırken, ikincisi bir sıfırlama geçidi ve bir güncelleme geçidi kullanılmaktadır. Standart RNN'lerin sorunlarını çözme potansiyellerinden dolayı hem LSTM hem de GRU, DDİ araştırmacılarının dikkatini çekmiştir (Cambria ve diğerleri, 2014). RNN' nin en popüler modeli, bir giriş kapısı, bir bellek hücresi, bir çıkış geçidi ve bir unutma geçidinden oluşan Uzun-Kısa Dönem Belleği (LSTM)' dir (Basiri, Nemati, Abdar, Cambria ve Acharya, 2020). Bellek hücresinin görevi, önceki durumu hatırlamak, kapıların görevi ise açığa çıkacak bellek miktarını kontrol etmektir (Sarkar, Bali ve Sharma, 2018). LSTM modelinin küçük bir varyasyonu, "unut kapılarını ve giriş kapılarını tek bir

güncelleme geçidinde" birleştiren "Kapılı Yinelemeli Ünite (GRU)" dır. Ayrıca hem önceki hem de sonraki bağlamları dikkate almak için, çift yönlü LSTM (Bi-LSTM) ve çift yönlü GRU (Bi-GRU) önerilmiştir. İleri ve geri gizli katmanları birleştirerek bu modeller sıralı modelleme problemini daha iyi çözebilmektedir (Nassif, Elnagar, Shahin ve Henno, 2020).

3.3. Kapılı Yinelemeli Sinir Ağları (GRNN)

CNN (Kalchbrenner, Grefenstette ve Blunsom, 2014; Kim, 2014) ve RNN (Tang, Qin ve Liu, 2015; Wang, Can, Kazemzadeh, Bar ve Narayanan, 2012) gibi modeller duygu analizi çalışmalarında etkin sonuçlar elde etmiştir. Bu yaklaşım manuel tekniklerin rolünü azaltmış, araştırmacıların çalışmalarını kolaylaştırmıştır. RNN, duygu sınıflandırma yaparken giriş sırasını kelime modeller ve bağımlılıklara bağlı ağırlıkları tekrar tekrar günceller. Yakalanacak bağımlılıkların süresi arttıkça RNN eğitilmesi zorlaşmaktadır. Bu durum gradyan değerinin çok artmasına veya gradyan kaybolmasına sebep olmaktadır. Bu da öğrenmeye engel bir durum ortaya çıkarmaktadır. Uzun vadede RNN pratik zorluklara sahip bir modeldir (Bengio, Simard ve Frasconi, 1994). Bu zorlukların üstesinden gelmek için, GRNN modelleri olan Long Short Term Memory (LSTM) (Hochreiter ve Schmidhuber, 1997) ve Gated Recurrent Units (GRU) (Cho ve diğerleri, 2014) önerilmiş ve üstün performans elde edilmiştir. GRNN' lerde, yeni bir giriş kelimesinin sınıflandırılması, yeni giriş ve önceki kodlanmış bilginin eklenmesiyle hesaplanır (Chen, Zhuo ve Ren, 2019). Başka bir deyişle, yeni duygu girdisi, tarihsel bilgiler tarafından değişen bir şekilde değiştirilir. Bu yaklaşımın metotları arasındaki LSTM ve GRU çeşitli uzunlukta ve uzun dönemli anlam bilim yakalamada gösterdikleri performanstan ötürü en yaygın kullanılan tekniklerden biridir (Chen ve diğerleri, 2019). LSTM geçmişten mevcut duruma bilgi akışını kontrol etmek için giriş kapısı, unut kapısı, çıkış kapısına sahiptir. GRU ise çalışma modeli olarak LSTM' ye oldukça benzerdir. Sıfırlama ve Güncelleme olarak iki kapısı vardır. LSTM ve GRU yöntemi arasında bir sınıflandırma performansı üstünlüğü literatürde belirtilmemiştir. Ayrıca birbirleri yerine kullanıldıkları da gözlemlenmiştir (Goularas ve Kamis, 2019).

3.3.1. Uzun kısa dönem bellek (LSTM) ve kapılı sinir ağı (GRU)

Tarihle ilgili problemleri işlemek için, yapay zekâ alanında gizli katmanlar arasında tekrarlayan bağlantıları kullanan "Recurrent Neural Network (RNN)" önerilmiştir (Elhart,

Hinton ve Williams, 1986). Bununla birlikte, RNN' nin uzun vadeli geçmiş verilerini öğrenmede başarısız olduğuna dair bir sorun bulunmuştur ve bu sorun, 1997 yılında Hochreiter ve Schmidhuber (Hochreiter ve Schmidhuber, 1997) tarafından RNN' nin bir uzantısı olan LSTM ağını önererek çözülmek istenmiştir. LSTM, önceki verilerin hem uzun hem de kısa vadeli etkilerini öğrenebilmektedir. LSTM ağları, pozitif değerlerin geçmişini kısa bir süre için hatırlayarak tasarlanmış Tekrarlayan Sinir Ağı mimarilerinden biridir (Basiri ve diğerleri, 2020). Her LSTM ünitesi bir hafıza hücresinden oluşur. Her bir hafıza hücresi, girişte okumak için giriş kapısı, sonraki katmanlara çıktı yazmak için çıkış kapısı ve hangi verilerin hatırlanacağına ve hangi verilerin unutulacağına karar veren unut kapısı olan üç kapıdan oluşmaktadır (Şekil 4. (a)). Bu kapılar hafıza hücresine giren ve çıkan bilgi akışını düzenlemek için tasarlanmıştır. LSTM kullanılarak bellek üzerinde ayrıntılı denetim için çeşitli özellikler verilmiştir. Bunlar, yeni hafızayı oluşturmak için mevcut girdinin önem derecesini ve yeni hafızanın tasarımında önceki hafızaların ne kadar önemli olduğunu ve hafızanın hangi bölümlerinin gerekli olduğu ve çıktıyı üretmekte olduğunu kontrol etmemize yardımcı olmaktadır (Majumder, Poria, Gelbukh ve Cambria, 2017). 2020 yılında yayınlanan makalelere bakıldığında en çok kullanılan model olan LSTM' nin güncel bir model olduğu bilinmektedir (Ligthart ve diğerleri, 2021).

LSTM'lere oldukça benzer çalışan GRU modeli, Sıfırlama ve Güncelleme olarak iki kapısı vardır (Şekil 4. (b)). Sıfırlama geçidi, önceki gizli durumun ne zaman göz ardı edileceğine karar verir ve güncelleme geçidi, mevcut duruma geçirilmesi gereken bilgi miktarına karar verir (Basiri ve diğerleri, 2020). Bu model, standart LSTM modellerine kıyasla basitliği nedeniyle giderek daha yaygın kullanılmaya başlanmıştır.

Çift yönlü model olan “Bidirectional LSTM (Bi-LSTM)” ve “Bidirectional GRU (BiGRU)” hem geçmiş hem de gelecekteki bilgilere bakarak ağa ek bağlam sağlayabilir. Daha hızlı ve daha iyi öğrenmeyle sonuçlanmaktadır. Bi-LSTM ve Bi-GRU, NLP uygulamalarında yaygın olarak kullanılmasına rağmen metin işleme uygulamalarında yaygın olan yüksek boyutlu girdi alanı, modelin karmaşıklığını artırır ve optimize etmeyi zorlaştırır (Basiri ve diğerleri, 2020). Bu sorunu çözmek için literatürde birkaç yaklaşım önerilmiştir. Örneğin, CNN'ler özellik uzayının boyutunu azaltmak ve metinden anlamlı özellikler çıkarmak için kullanılmıştır (Mehta, Majumder, Gelbukh ve Cambria, 2019). Bununla birlikte CNN ve çift yönlü modellerin birlikte kullanan duygu analizi çalışmalarının başarısı kanıtlanmıştır (Basiri ve diğerleri, 2020; Minaee, Azimi ve Abdolrashidi, 2019).

3.4. Performans Ölçüm Yöntemleri

Olasılık matrisi olarak da bilinen hata matrisi, deneyler yoluyla elde edilen istatistiksel değerlerin görsel temsidir. Bu matris, sınıflandırılan metinde her bir sınıflandırılmış metnin gerçek ve tahmin edilen düzeyi hakkındaki istatistikleri gösterir. Bu istatistiki değerler, makine öğrenimi algoritmalarının çoğunun performansını değerlendirmek için kullanılmaktadır. Önerilen algoritmanın performansına, hata matrisinden ölçülen “doğruluk”, “kesinlik”, “duyarlılık” ve “F-skor” yöntemleri ile erişilmiştir (Çizelge 3.2). Çalışmada CNN, LSTM, Bi-LSTM, GRU, Bi-GRU ve CNN-LSTM modelleri kullanılacaktır. Bu modellerin performans sonuçlarının bulunması bu yöntemler ile yapılacaktır. Böylelikle modeller arasında karşılaştırma yapılarak etkin modele karar verilebilecektir. Örneğin; bu çalışmada, yorumların sınıflandırılması ilk aşamada saldırgan veya saldırgan olmayan duygular olarak etiketlenmiştir (Çizelge 3.1). Hata matrisinin, farklı performans parametrelerinin değerlendirilebileceği dört bileşeni vardır:

1. Doğru Pozitif (TP): Başlangıçta saldırgan olmayan olarak etiketlenen ve sınıflandırıcı tarafından saldırgan olmayan olarak tahmin edilen incelemeleri temsil eder.
2. Yanlış Pozitif (FP): Başlangıçta saldırgan olarak etiketlenen ancak sınıflandırıcı tarafından saldırgan olmayan olarak tahmin edilen incelemeleri temsil eder.

3. Doğru Negatif (TN): Başlangıçta negatif olarak etiketlenen ve sınıflandırıcı tarafından saldırgan olarak tahmin edilen incelemeleri temsil eder.

4. Yanlış Negatif (FN): Başlangıçta saldırgan olmayan olarak etiketlenen ancak sınıflandırıcı tarafından saldırgan olarak tahmin edilen incelemeleri temsil eder.

Çizelge 3.1. Hata matris gösterimi

		Gerçek Etiket	
		Saldırgan	Saldırgan Değil
Tahmin Edilen Etiket	Saldırgan	Doğru Pozitif (True Positive (TP))	Yanlış Pozitif (False Positive (FP))
	Saldırgan Değil	Yanlış Negatif (False Negative (FN))	Doğru Negatif (True Positive (TN))

Çizelge 3.2. Sınıflandırma performansı ölçüm yöntemleri

Doğruluk: doğru tahmin edilen örneklerin fraksiyonudur. $\text{Doğruluk} = (TP + TN) / (TP + FP + TN + FN)$
Kesinlik: Doğru pozitif tahminin toplam pozitif tahmine oranıdır. $\text{Kesinlik} = TP / (TP + FP)$
Duyarlılık: Doğru pozitif tahmin sayısının toplam örnek sayısına oranıdır. $\text{Duyarlılık} = TP / (TP + FN)$
F1-skor: “Kesinlik” ve “Duyarlılık” harmonik ortalamasıdır. $\text{F-skor} = 2 * \text{Kesinlik} * \text{Duyarlılık} / (\text{Kesinlik} + \text{Duyarlılık})$

Her sınıfa eşit ağırlıklar veren, sınıf başına puanların aritmetik ortalamasını alan F1 makro değeri çalışmada kullanılmıştır. Bununla birlikte roc-auc ve prc-auc değerleri de performans ölçüm yöntemlerinden biridir. Burada ROC eğrisi, her sınıflandırma eşiği için bir modelin performansını gösterir. AUC algoritması ise ROC eğrisinin altındaki alanı ölçer ve modelin rastgele seçilmiş bir pozitif örneği, rastgele seçilmiş bir negatif örnekten daha yüksek sıralama olasılığını temsil eden 0 ile 1 arasında bir puan sağlar. AUC, her iki sınıfın da eşit derecede önemli vermektedir. Bununla birlikte yüksek sınıf dengesizliği olan problemler için, azınlık sınıfına özel hassas geri çağırma eğrisi (PRC) altındaki alanı hesaplamaktadır.

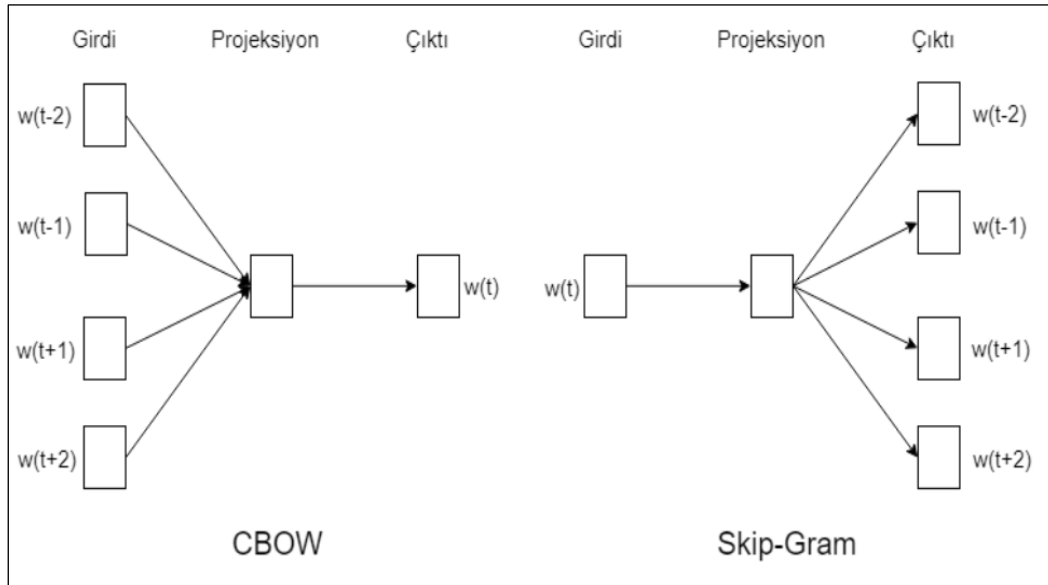
Ancak her iki ölçüm (ROC-AUC ve PRC-AUC), sınıflandırıcı karşılaştırması için iyi değildir (Massip, M., 2020).





4. KELİME TEMSİL YÖNTEMİ

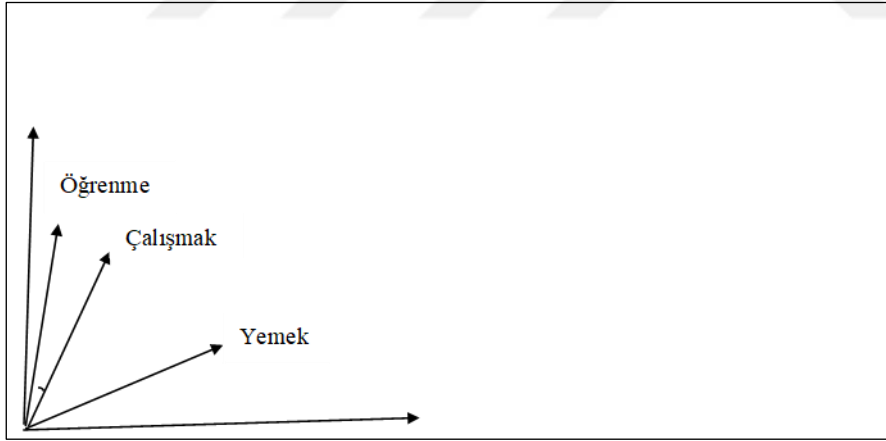
Bir metnin vektörel temsilini bulmada bilinen iki yöntem mevcuttur. Bunlar kelime çantası modeli (Bag of Words) ve TF-IDF yöntemleridir. Kelime çantası yaklaşımında tüm veriler ele alınarak kelimenin sadece tekrar sayısına odaklanılmaktadır. Böylece bu yaklaşım sayesinde, metinde yer alan kelimenin konum bilgisi tutulmamaktadır. TF-IDF modeli, konum bilgisini tutmaktadır ancak çok fazla sıfırdan oluşan büyük ve seyrek bir matristen oluşur. Bu da hesaplamayı zorlaştırmaktadır. Kelime temsil yaklaşımı bu dezavantajların üstesinden gelmek için önerilmiştir. Bu yaklaşım ile kelimelerin bilgisayar dilinin de anlayacağı ve hesaplamaya uygun hale getirilmesi için sayılardan oluşan vektörler kullanılmaktadır. Kelime temsil yaklaşımında her bir kelime bir vektör olarak temsil edilmektedir. Birbirine yakın kelimelerin vektörel temsili de yakın olmaktadır (Şekil. 4.1). DDİ çalışmalarında kelime temsil yönteminin performans artırıcı özelliğe sahip olduğu kanıtlanmıştır (Özer, 2019). Kelime temsil yaklaşımı, çalışmada veri setindeki her bir kelimenin vektörel temsilini sağlarken yapacağımız sınıflandırma işleminin performansını artırmak için kullandık. Bilinen kelime temsil modelleri “Word2vec”, “Glove” ve “Fasttext” tir.



Şekil 4.1. CBOW ve Skip-gram mimarisi (Özer, 2019)

Word2Vec modeli, derlemiden çıktı olarak kelime vektörleri üretir. Ayrıca bu modelde kelime vektörleri sabit uzunluktadır. Vektörlerden kelimelerin anlamsal bilgilerine ulaşılmaktadır (Wang, Lee ve Lee, 2018). Word2vec modelinde bilinen 2 yaklaşım vardır. Bunlar, “Skip-gram” ve “Continuous Bag Of Word (CBOW)” dur. Her ikisi de tek bir gizli katmana sahip basit sinir ağı modelidir. Ancak bu iki yaklaşım kelimeyi tahmin etmek için birbirine zıt çalışmaktadır (Şekil 4.1). CBOW mimarisi, geçmiş ve gelecek bağlamdan bir kelimeyi tahmin eder. Skip-gram ise, belirli bir kelimedenden bağlamsal kelimeleri tahmin eder (Sharma, Chaurasia ve Srivastava, 2020).

Glove modeli, bağlam ve kelime sayısı ile orantılı bir matris oluşturur. Burada matristeki her satır için karşılık gelen her bir kelimenin vektör temsili gösterilmektedir. Bu model, Word2Vec gibi kelimeleri vektör biçiminde temsil etmektedir. Anlamsal benzerliğe odaklanan bir diğer popüler yaklaşımdır. Ayrıca her iki yaklaşımda da kelimenin örnek uzayda yer alan konumu ile kosinüs benzerliğine bakılarak anlamsal yakınlık tespit edilebilir (Şekil 4.2). Ancak eğitim aşamasında Glove tüm derlemi ele alırken Word2Vec metindeki komşu kelimeleri baz almaktadır.



Şekil 4.2. Kosinüs benzerliğine göre anlamsal yakınlık

Fasttext Modeli, Facebook'un AI Arařtırma laboratuvarında Bojanowski ve arkadaşları tarafından oluşturulan sinir ağılarına dayanan popüler kelime temsili tekniklerinden biridir. (Bojanowski, Grave, Joulin ve Mikolov, 2017; Mikolov ve diğlerleri, 2018). Çalışmada model oluşturulurken Word2Vec CBOW kullanılmıştır. Morfolojiyi göz önünde bulundurarak kelimelerin yüksek kalite de temsillerini öğrenmek için uyarlanmıştır. Bu model, vektörlerin kelimeler arasında paylaşılmasını kolaylaştırarak nadir kelimelerin güvenilir temsiliini öğrenmeyi de sağlar (Bojanowski ve diğlerleri, 2017). Fasttext ise İngilizce, Almanca, İspanyolca, Fransızca, Çekçe ve Türkçe dahil 157 dilde çalışmak üzere tasarlanmış açık kaynaklı bir kütüphaneye sahiptir (Joulin ve diğlerleri, 2016).





5. SALDIRGAN DİL VE DUYGU ANALİZİ

Son on yılda, web de özellikle Facebook ve Twitter gibi popüler sosyal medya platformlarının ortaya çıkmasıyla birlikte kullanıcı tarafından oluşturulan içerikte katlanarak bir artış olmuştur (Reinsel, Gantz ve Rydning, 2018). Artık çevrimiçi herhangi bir bilgi milyarlarca insana saniyeler içinde ulaşma gücüne sahiptir. Bu, yalnızca olumlu fikir alışverişiyle sonuçlanmakla kalmamış, aynı zamanda saldırgan ve potansiyel olarak zararlı içeriğin web üzerinde yaygın bir şekilde yayılmasına da yol açmıştır. Zorbalık veya nefret söylemi gibi potansiyel olarak zararlı olayların çoğu internetten önce gelmiş olsa da internetin erişimi ve kapsamı bu olaylara milyarlarca insanın hayatını etkilemek için benzeri görülmemiş bir güç ve etki sağlamıştır. Bu olayların sadece web kullanıcıları için zihinsel ve psikolojik acı yaratmadığı, aslında insanları hayatları ile ilgili kararlar hakkında zorladığı ve aşırı durumlarda intihara meyillendirdiği bilinmektedir (Hinduja ve Patchin, 2010). Ayrıca sözlü sohbetlerde, üniversite öğrencilerinin söylediği kelimelerin %0,5'inin küfür olduğu belirtilmektedir (Mehl ve Pennebaker, 2003). Twitter'da bu oran iki kattan fazla, %1,2 ve %7,7 küfür içeren tweetler olarak görünmektedir (Wang, Chen, Thirunarayan ve Sheth, 2014). Saldırgan gönderilerin %9'unun daha büyük bir kısmının da küfür içermesi muhtemeldir (Xu, Jun, Zhu ve Bellmore, 2012).

Saldırgan içerik olayları küçük bir sorun olarak kalmamış, çok sayıda insanı etkileyen büyük bir suç faaliyeti biçimini kazanmıştır. Bu nedenle, çevrimiçi ortamda taciz edici davranış saldırganlığıyla başa çıkmak için önleyici tedbirlerin alınabilmesi önemlidir. Çevrimiçi agresif davranışlarla başa çıkma stratejilerinden biri, kullanıcı tarafından oluşturulan içeriği manuel olarak izlemek ve denetlemektir. Ancak, web 'de yeni verilerin oluşturulma miktarı ve hızı, manuel denetim ve müdahale yöntemlerini neredeyse tamamen kullanışsız hale getirmiştir. Bu nedenle, bu tür davranışları tanımlamak için (yarı) otomatik yöntemlerin kullanılması önemli hale gelmiştir ve son yıllarda araştırmacıların dikkatini çekmeye başlamıştır. (Davidson, Warmesley, Macy ve Weber, 2017; Malmasi ve Zampieri, 2017). Uygunsuz dilin dilbilimsel, psikolojik ve sosyolojik yönlerine olan akademik ilginin yanı sıra, uygunsuz dilin başarılı bir şekilde otomatik olarak tanımlanmasından yararlanabilecek pratik uygulamalar da vardır. Örneğin, sosyal medya platformlarının denetlenmesinde veya daha genel olarak kullanıcı içeriğine izin veren internet sitelerinde yararlı olabilir; yasadışı içeriği tespit etmek için kolluk kuvvetleri tarafından kullanılabilir; taciz edici dilin psikolojik etkilerinin incelenmesini kolaylaştırabilir.

Literatürdeki çalışmalarının çoğu, belirli bir saldırgan dil biçimini açıklar ve çoğu zaman, belirli bir uygulamada kullanılmak üzere tasarlanmıştır. Şimdiye kadar en yaygın uygulama, nefret söylemi algılamadır. Nefret söyleminin net bir tanımı olmamasına rağmen, tipik olarak ırk, etnik köken, cinsiyet, cinsel yönelim, sosyo-ekonomik sınıf, siyasi bağlantı veya din gibi özelliklere dayalı olarak bir grubu (veya bazen bir kişiyi) hedef alan saldırgan dili kapsar (Çöltekin, 2020). Bu çalışmalardan bazıları kapsamı belirli bir hedefe, genellikle ırk (Basile ve diğerleri, 2019; Kwok ve Wang, 2013), kadınlar (Basile ve diğerleri, 2019) ve mülteciler (Ross, Rist, Carbonell, Cabrera, Kurowsky ve Wojatzki, 2017), belirli bir ideolojiye sahip nefret söylemi (Jaki ve De Smedt, 2019) veya tek bir önemli olayla ilgili nefret söylemidir (Burnap ve Williams, 2014). Bir diğer yaygın saldırgan dil alt alanı da siber zorbalığın tespit edilmesidir (Dadvar, Trieschnigg, Ordeman, R., ve de Jong, 2013; Dadvar, Trieschnigg ve de Jong, 2014; Dinakar, Jones, Havasi, Lieberman ve Picard, 2012; Nitta ve diğerleri, 2013; Van Hee ve diğerleri, 2015; Xu ve diğerleri, 2012). Nefret söyleminin aksine, siber zorbalığın hedefi genellikle tek bir kişidir ve bu genellikle bir çocuktur. Zorbalığın çevrimiçi versiyonu siber zorbalık, ciddi bir sağlık sorunu olarak kabul edilmektedir (American Psychological Association, 2004). Bu nedenle, siber zorbalığın otomatik tespitinin tipik bir uygulaması, çocuklar için daha güvenli çevrimiçi iletişim sağlamaktır (Chen, Zhou, Zhu ve Xu, 2012). Uygulamalar farklı olsa da önemli bir örtüşme vardır. Örneğin, siber zorbalık genellikle nefret söylemi olarak kabul edilen ifadeler kullanır. Dahası, metinlerin hem dil özellikleri hem de bunları tespit etme yöntemleri benzerdir.

Genel yönergeler, saldırgan dil türlerini tanımlamaya yardımcı olsalar da belirli bir ifadenin saldırgan olup olmadığına dair karar, çoğunlukla öznel ve büyük ölçüde bağlama bağlıdır. Son zamanlarda yapılan birçok çalışmada sık sık gündeme getirilen bir nokta, saldırgan dilin tanımı ve alt kategorileri konusunda fikir birliğinin olmaması ve bunun sonucunda farklı derlemlerde etiketlemelerin uyumsuzluğudur. Birkaç istisna bir yana, saldırgan dille ilgili etiketleme görevlerinden herhangi biri üzerindeki yorumcuların aynı kararı almasındaki uyuma durumu nispeten düşüktür. Saldırgan dil etiketlemesinde kullanıcılar arasında Wiegand ve diğerleri (2018) rapor $\kappa = 0.66$ ve Zampieri ve diğerleri (2019) %60 uyuma bildirmiştir. (Wiegand ve diğerleri, 2018; Struß, Siegel, Ruppenhofer, Wiegand ve Klenner, 2019; Zampieri ve diğerleri, 2019). Ancak saldırgan dil de net tanımlar ve sınıflandırmalar için daha nesnel olmalarını sağlama amaçlı girişimler bulunmaktadır (Waseem, Davidson, Warmesley ve Weber, 2017; Ruppenhofer, Siegel ve Wiegand, 2018). Zampieri ve diğerleri, 2019 tarafından yapılan çalışmaya göre saldırgan dilde hedef, bir birey veya ırkına,

cinsiyetine, siyasi / ideolojik yakınlığına, dinine veya benzer bir mülke dayalı bir grup insandır. Eğer ki hedef bir bireyse; kategori genellikle siber zorbalık eylemlerini içerirken, hedef bir grup ise, muhtemelen bir nefret söylemi örneğidir. Ayrıca Zampieri ve arkadaşları (2019), hedefin bazen (açıkça) insanlar değil, örneğin bir organizasyon veya olay olduğunu da söylemektedir (Zampiere ve diğerleri 2019). Hedef alınmayan hakaretler de yaygındır. Hedeflenmemiş suç, küfürlü veya müstehcen dil veya genel olarak, bir bireyi veya bir grubu rencide etme niyeti veya etkisi olmaksızın kullanılan ifadeleri içerir. Bu durum tipik olarak bir suça karşılık gelmez (Zampiere ve diğerleri 2019). Bununla birlikte, çocuk kullanıcılar ve kurumsal şirket hesapları için bu dil biçimlerinden kaçınılmaktadır.

Öznel ve bağlama bağlı olduğu bilinen çeşitli saldırgan dil biçimlerinin otomatik olarak tanımlanması çalışmalarının yöntemleri ve başarı oranı da değişiklik gösterir. Genel olarak, başarı oranı da göreve ve veri setine bağlı olarak değişir. Çalışmaya en yakın otomatik tanımlama deneyi seti OffensEval-2019/2020 (Zampieri ve diğerleri, 2019; Zampieri ve diğerleri, 2020) ve GermEval yarışmalarındaki saldırgan dil tanımlama görevleridir. Her ikisinin de birbirini izleyen görevler mevcuttur. OffensEval 2019’ da ilk alt görev, saldırgan dili saldırgan olmayan dilden ayırmayı içerir; Saldırgan ve saldırgan olmayan arasındaki bu sınıflandırma, kullanıcı içeriğinin ebeveyn filtreleme gibi birçok uygulamada kullanışlı olduğu düşünülmektedir. Ayrıca şirketlerin web sayfalarında, e-ticaret şirketlerinde saldırgan içerik bulunduran yorumların otomatik tespit edilip engellenebilir. Buna ek olarak insanların yoğun ve şirketlerin etkin kullandığı Youtube gibi multimedya içerik bulunduran mecralarda tamamen yorumların kapanması yerine otomatik olarak saldırgan içeriklerin engellenmesi sağlanabilir. İkinci görev ise, verilen saldırgan belgenin hedeflenip hedeflenmediğini belirlemektir ve son olarak üçüncü alt görevler, hedef türü (grup, birey veya diğer) belirlemektir. Diğer bir benzer çalışma ise GermEval yarışmalarındaki saldırgan dil tanımlama görevidir. Bu çalışma da Struß ve arkadaşları tarafından yapılan Alman saldırgan dili tanımlama görevi, saldırgan dilin tanımlanması gibi ikili bir görevle başlar ve ardından saldırgan dilin türünü (küfür, taciz veya hakaret) tanımlamayı gerektiren ikinci bir görev izler. Alman saldırgan dili tanımlama görevinde çoklu etiket sınıflandırma söz konusudur (Struß ve diğerleri, 2019). Her iki yarışmanın birinci görevin de ulaşılan başarı oranı yüksektir (sırasıyla; %82,9 F1 ve %76,8 F1). Diğer alt görevler de ise başarı oranı da düşmektedir (sırasıyla; %66 F1 ve %53 F1).

Saldırgan dil veri seti içeren çalışmaların çoğu İngilizce üzerinde yürütülmekte, ardından Hint-Avrupa dil ailesinde nispeten iyi çalışılmış diğer diller (ağırlıklı olarak Avrupa dillerine eğilimli), örneğin Almanca (Jaki ve De Smedt, 2019), İspanyolca (Basile ve diğerleri, 2019), İtalyanca (Del Vigna¹², Cimino²³, Dell’Orletta, Petrocchi ve Tesconi, 2017) ve Hollandacadır (Van Hee ve diğerleri, 2015). Sadece Hint-Avrupa dışı dillerin saldırgan dilde derlem ve ek açıklamasını içeren çalışmalar, bildiğimiz kadarıyla Arapça (Mubarak, Darwish ve Magdy, 2017) ve Endonezyaca (Ibrohim ve Budi, 2018), Türkçe ’de ise tek çalışma Çöltekin tarafından yapılan çalışma olduğu görülmektedir (Çöltekin, 2020). Bu çalışmadan farklı olarak çoklu sınıflandırma yapılan veri seti çalışmada oluşturulmuştur. Bu veri setinin Türkçe dilinde tek olduğu görülmüştür.

Saldırgan dil çalışmalarında verileri etiketlemenin önemi; belirtildiği gibi, önceki çalışmaların çoğu, en yaygın olarak nefret söylemi ve siber zorbalık olmak üzere belirli saldırgan dil biçimlerine odaklanmaktadır. Bununla birlikte, farklı saldırgan dil türleri arasında önemli bir örtüşme vardır (dil özellikleri ve saldırgan dil tespit etme yöntemleri gibi). Saldırgan dili daha geniş bir şekilde kapsayan bir derlemin, farklı saldırgan dil türleri arasındaki benzerlikleri ve farklılıkları anlamaya ve / veya bunlardan yararlanmaya yardımcı olma olasılığı daha yüksektir. Göreve yönelik dar bir veri kümesi üzerinde eğitilen otomatik bir sistem, saldırgan dil örneklerinin diğer biçimlerini amaçlanan saldırgan dil türünden ayırt etmekte başarısız olabilir. Örneğin, nefret söyleminin tanımlanmasına yönelik bir sistem, küfür olumsuz durumlarda iyi temsil edilmeyen bir derlem ile eğitilmişse, küfür gibi diğer saldırgan dil biçimlerini nefret söylemi olarak karıştırabilir. Saldırgan ifadenin türüne bakılmaksızın, bu karışıklığın yaygın bir nedeni bağlam eksikliğidir. Bazı tweetler bağlam olmadan saldırgan görünebilirken, doğru bağlamda sadece arkadaşça bir alay konusu olabilir. Diğer durumlarda, tweetler rahatsız edici görünmeyebilir, ancak devam eden konuşmada veya hatta daha geniş bir sosyal bağlamda bir kişiye yönelik bir saldırgan dil suçu işleniyor olabilir. Örneğin; ironi olarak yapılan iltifat veya insan olmayı başarabilseydin gibi koşullu cümleler.

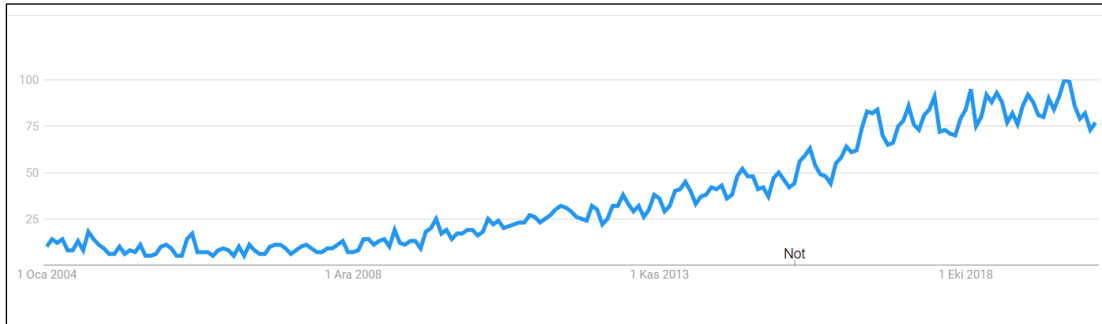
Açık Saldırgan Dil – Örtük Saldırgan Dil

Örtük saldırgan dil, açıkça veya örtük olarak verilen bir hedefe yönelik saldırgan dil kullanan kullanıcılar, nefretlerini açıkça ifade etmek yerine, başka şekilde ifade etmektedirler, yani tweetleri hedefin değersiz olduğunu ima ediyor, vb. İroni veya alay gibi mecazi bir dil kullanan saldırgan tweetler veya bir kelime oyunu da örtük dil olarak sayılmaktadır. Örtük saldırgan ifadeler bazen yalnızca kendi bağlamları içinde yorumlanabilir. Ayrıca, ciddi bir konuyu ele alırken uygunsuz gündelik dil, örtük saldırgan dil kapsamına alınır (Struß ve diğerleri, 2019). Örneğin; “Mülteciler vatanlarından kaçarken geride kalan akrabalarının ailelerinin ölümüne çok onurlu şekilde izin verirler.” Bu örnek saldırgan dildeki alay ve ironi durumlarıdır. Bunlar otomatik saldırgan dil tespitinde bağlamın etkisini göstermektedir. Bu alanda 2019 yılında Struß ve arkadaşları Almanca dilinde çalışma yapmıştır (Struß ve diğerleri, 2019).



6. LİTERATÜR TARAMASI

Başkalarının ne düşündüğü karar verme sürecinde çoğumuz için her zaman önemli bir bilgi parçası olmuştur (Pang ve Lee, 2008). Duygu analizi çalışmaları bu ihtiyacımıza yönelik olarak ortaya çıkmıştır. Son yıllarda duygu analizi ve fikir madenciliğine odaklanan makale sayısında büyük bir artış görülmektedir (Şekil 6.1). Bu konuyla ilgili yaklaşık 7000 makale yayınlanmıştır (Mäntylä ve diğerleri 2018). Duygu analizi ile ilgili makalelerin çoğunluğu duygu kutuplaşmasıyla ilgilidir. Yani bir kişinin bir şeye karşı olumlu, tarafsız veya olumsuz bir görüşü olup olmadığıyla ilgilenir. (Dave ve diğerleri, 2003).



Şekil 6.1. "Google Trends" "Sentiment Analysis" çalışma alanını gösteren veriler grafiği

Bu alandaki ilk akademik çalışmalar İkinci Dünya Savaşı sırasında ve sonrasında politika alanında yapılmıştır (Stagner, 1940). Duygu analizi çalışmaları 2000'lerin ortalarında ürün incelemelerine odaklanmıştır. Örneğin; Dave ve diğerleri, 2003' te yaptığı çalışmada ürün özelliklerinin listesi ve her biri hakkındaki görüşleri bir araya getiren Web aramalarından toplanan bireysel cümlelerden olumlu ve olumsuz görüşleri otomatik olarak ayırt etmek için bir yöntem geliştirmiştir (Dave ve diğerleri, 2003). Bununla birlikte, duygu analizinin kullanımı, finansal piyasaların tahmini (Nassirtoussi ve diğerleri, 2014), terörist saldırılara verilen tepkiler (Bernap ve diğerleri, 2014) gibi geniş bir alana yayılmıştır. Buna ek olarak ironi tespiti (Reyes ve diğerleri, 2014), çok dilli destek (Hogenboom ve diğerleri, 2014) ve basit kutupluluk tespitinden daha karmaşık olan olumsuz duyguların öfke, keder gibi ayrılması (Munezero, 2014) gibi duygu analizinin uygulanabilirliğine katkıda bulunmak adına çalışmalar da mevcuttur. Duygu analizi çalışma alanı, yıllar içinde içerik açısından da değişmiştir. Bu çalışma alanının tarihçesine baktığımızda çevrimiçi olarak büyük miktarda metin ve görüş bulunmadan önce, araştırmalar esas olarak anket tabanlı yöntemlere dayanıyordu ve kullanıcıların veya müşterilerin görüşlerinden ziyade kamuya açık veya uzman görüşlerine odaklanılmıştır. Buna göre ilk makale 1940 yılında yayınlanmıştır ve bu

çalışma "The Cross-Out Technique as a Method in Public Opinion Analysis' " başlığını taşımaktaydı (Stagner, 1940). 1945 ve 1947 yılında üç bildiri, İkinci Dünya Savaşı'nda Japonya (Knutson, 1945), İtalya (Fegiz, 1947) ve Çekoslovakya (Adamec ve Viden, 1947) savaşa katılan askerlerin ve savaşın mağduriyetlerini yaşayan insanların görüşlerini ölçmüştür ve hepsi "Public Opinion Quarterly" dergisinde yayınlanmıştır. 90'lı yılların ortalarında bilgisayar tabanlı sistemler ortaya çıkmaya başlamıştır. Bilgi işlem devrimi, araştırmada da yansıtıcı rol üstlenmiştir. Örneğin, "Elicitation, Assessment and Pooling of Expert Judgments Using Possibility Theory" başlıklı bir makale 1995 yılında yayınlanmıştır. Makalede, oluşturulan duygu havuzu bilgisayar sistemlerin de yardımıyla endüstriyel güvenlik alanında uzman görüşü analizinde kullanılmıştır (Sandri ve diğerleri, 1995). Yine de modern duygu analizinin örneklerinin çoğalmasa neredeyse on yıl sonrasına kalmıştır. Modern duygu analizinin ortaya çıkmasındaki en eski çalışmalardan biri, 1962'de kurulan "Computational Linguistic" derneğinde yapılan çalışmadır. Dernek'te çalışan Wiebe, 1990'da yaptığı çalışmasında öznel cümlelerin tespiti için yöntemler önermiştir (Wiebe, 1991). 1999'da ise çalışmalarını geliştirerek bu çalışmalarda kullanılacak bir standart temel sunmuştur (Wiebe ve diğerleri, 1999). Modern duygu analizi çalışmalarıyla ilgili "Scopus" ve "Google Akademik" de görülen alıntı sayılarının çokluğuna göre bu alandaki çalışmalara etki edildiği düşünülmüştür. Buna göre 1997'de Hatzivassiloglou ve McKeown tarafından yapılan "Predicting the Semantic Orientation of Adjectives" adlı çalışmada 2 milyon kelime bulunan "1987 Wall Street Journal" sözlüğünden rastgele elde edilen sıfatlardan oluşan verilerin ilk olarak yönelimine ve (Örneğin; "ucuz" kelimesi ilk anlamıyla pozitif sayılırken düşük kaliteyi ifade ettiğinde negatif sayılmaktadır.) daha sonra da bu sıfatların negatif ve pozitif olarak ayrılmasını sağlayan bir çalışmadır (Hatzivassiloglou ve McKeown, 1997). 2002 yılında ise Pang ve diğerleri tarafından yapılan "Thumbs up?" adlı çalışmada "Internet Movie Database (IMDb)" den elde edilen film yorumlarından sadece kullanıcı derecelendirmesinin yıldızlarla veya bazı sayısal değerlerle ifade edildiği incelemelerinin seçildiği bir veri setinden negatif, pozitif veya nötr olarak duygu sınıflandırması yapılırken "Naive Bayes" ve "SVM" gibi makine öğrenmesi teknikleri de kullanılmıştır. Çalışma da web derecelendirmesi kullanan veri setlerinin daha iyi performans gösterdiği görülmüştür (Pang ve diğerleri, 2002). Benzer olarak "Thumbs up or thumbs down?" adlı çalışmada otomobiller, bankalar, filmler ve seyahat verileri ile denetimsiz sınıflandırma kullanılarak %74 doğruluğa ulaşmış olan bir çalışmadır (Turney, 2002). 2003'de ise Turney ve Litmann'ın yaptığı çalışma, olumlu semantik yönelim övgü (örneğin; dürüst, cesur ...) olumsuz semantik yönelim eleştiri (örneğin; rahatsız edici, gereksiz) olan anlamsal yönelimi

kelimenin istatistiksel bağlamından otomatik olarak çıkaran bir yöntem sunmuştur (Turney ve Litmann, 2003). 2003’ te yapılan “Mining the peanut gallery: Opinion extraction and semantic classification of product reviews” adlı çalışma da duygu analizi, Web ‘de bulunan ürün incelemeleri verileri ile yapılmıştır. Yazarlar, bu tür incelemelerin yüksek hacimli olması nedeniyle bu tür incelemelerin otomatik analize ihtiyaç duyduğunu savunmuştur (Dave, Lawrence ve Pennock, 2003). Pang ve Lee tarafından 2008’de yapılan duygu analizi ile ilgili bir literatür taraması çalışması mevcuttur (Pang ve Lee, 2008). Duygu analizinin temellerine ve temel uygulamalarına odaklanan bu çalışma, “Google Scholar” ve “Scopus” da bu alanda en çok alıntıya sahip makaledir. 2013’de yapılan bir çalışmada, duygu analizi ve bunların son gelişmeleriyle ilgili farklı tekniklere geniş girişler vermektedir. Yazarlar tarafından duygu analizi için video ve sesin gelecekteki veri kaynakları olacağı tahmin edilmektedir (Cambria, Schuller, Xia ve Havasi, 2013).

2011’de Taboada ve diğerleri tarafından yapılan “Lexicon-based methods for sentiment analysis” adlı makalede önceden var olan bir sözlük tabanlı yöntemle yoğunlaştırma ve olumsuzlama gibi kuralları ekleyerek mevcut analiz aracı "Semantic Orientation CALculator (SO-CAL)"ı geliştirdiler. Bu yeni yaklaşım daha sonra, çeşitli veri setlerinin sonuçlara varmak için tüm yaklaşımlarla analiz edildiği bir deneyde, birkaç mevcut duygu analizi araçları ve yaklaşımları ile doğrulanır, değerlendirilir ve karşılaştırılır. Yazarlar, aracı değerlendirmek için, ürün incelemeleri, sosyal medya yorumları, haber makaleleri ve manşetler olarak birkaç veri seti kullanmıştır (Taboada ve diğerleri, 2011).

Metin madenciliği ve duygu analizi yapılandırılmamış verilerle ilgilenir. Bu yönüyle klasik veri madenciliği tekniklerinden farklıdır. (Oza ve Naik, 2016). Baskın bir duygu analizi tekniği olarak, sözlük yaklaşımı, metin verilerinin önceden tanımlanmış bir duygu sınıfları kümesine sınıflandırıldığı denetimsiz bir yöntemdir . Metnin duygu puanları, kelimelerden ve bunlara karşılık gelen duygu puanlarından oluşan bir sözlük olan duygu sözlüğü temel alınarak hesaplanır (Sun, Luo ve Chen, 2017). Gönderen veri modelleme, sözlük temelli yöntemler metin verilerinin kutupluluğunu belirlemeye yönelik modeller olarak görülebilir. Sözlük tabanlı yaklaşımlar, metin verilerindeki terimler sözlükte belirtilenden farklı bir kutupluluk taşıyabileceğinden yüksek performanslı sonuçlar sağlamayabilir (Muhammad, Wiratunga ve Lothian, 2016). Bu problem, alana özgü sözlükler oluşturularak önlenir, böylece kutupluluk problemi terimi en aza indirilebilir. 2012’de yapılan bir çalışma da çeşitli makine öğrenimi algoritmalarıyla birlikte kullanılan sözlük tabanlı bir

sınıflandırıcı olan "SentiStrength" adlı bir aracı değerlendirdiler. Bu araç, YouTube yorumları ve tweet 'ler gibi kısa metinler içeren çeşitli veri kümeleriyle değerlendirilir. Araç, sosyal ağda bulunan kısa metinleri iyi analiz eder (Thelwall, Buckley ve Paltoglou, 2012). Ayrıca "SentiStrength" Türkçe 'ye de uyarlanmıştır (Vural, 2013). Özellikle duygu analizinde kullanılmak üzere Baccianella ve arkadaşları tarafından geliştirilen Senti-Word Net, otomatik olarak oluşturulmuş bir sözlüktür. Sözlük, yarı denetimli bir öğrenme algoritması kullanılarak oluşturulmuş ve kelimelerin oluşturulan anlamsal değerlerinin manuel olarak oluşturulmuş olan eşanlamlı kelime karşılaştırmaları ile değerlendirilmiştir (Baccianella, Esuli ve Sebastiani, 2010). 2013'te yapılan çalışma ise duygu analizini belge ve cümle bazlı, bakış açısı odaklı, karşılaştırmalı olarak ayırmıştır. Aynı zamanda duygu analizinde sözlük ediniminin öneminden bahsetmiştir (Feldman, 2013). Buna ek olarak "Google Akademik" de en çok alıntı yapılan araştırmalar arasında sözlük tabanlı yaklaşımlar mevcuttur. Ancak günümüzde makine öğrenimi yaklaşımları kullanan çalışmaların sayısı sözlük bazlı yaklaşımların önüne geçmiştir (Pirani, Madhavi ve Singh, 2017).

R Paketi 'wordcloud' ile çizilmiş olan Şekil 6.2, "Google Akademik" ve "Scopus" da yayınlanan makalelerin başlıkları derlenerek araştırma konularını ve bu konuların yıllar içindeki farklılıklarını göstermektedir. Şekil 6.2 (a) duygu analizinin temel kavramlarının "social", "online", "reviews", "media", "product", "twitter" olduğunu göstermektedir. En yaygın kullanılan diller Çince ve Arapça'dır. İş alanlarında ise "movie", "news", "political", "stock" ve "financial" konuları yaygındır. Farklı analiz yöntemleri ise "neural", "fuzzy and "supervised" olduğu görülmektedir (Mäntylä ve diğerleri, 2018). Şekil 6.2. (b) ise bu konuların yılların içindeki farklılıklarını göstermektedir. Bir kelime karşılaştırma bulutudur ve her kelime önceki yıllardaki kullanım sayısına göre meydana gelen sapmalardır. Örneğin; "review" kelimesi küçülmüştür. Bu da dönemler arasındaki farkın daha az olduğu anlamına gelir. 2013 öncesine kadar Çince daha yaygın iken sonraki yıllarda Arapça ve Hintçe yaygın kullanılmaya başlanmıştır. 2013'ten önce yapılan çalışmalar çevrimiçi ürün incelemelerini analiz ederken, 2014-2016 yılları arasında sosyal medya araçları kullanılarak yapılan çalışmalar artmıştır.

Şekil 6. 2. (b) ' de vurgulandığı gibi, Twitter ve sosyal medya son yıllarda duygu analizinde popülerlik kazanmıştır. Pak ve Paroubek yaptıkları çalışmada, mikrobloglardan bir derlemin otomatik olarak toplanması için bir yöntem sunar ve bunu bir duygu sınıflandırıcı oluşturmak için kullanır. Bu örnekte, veriler Twitter dan toplanmıştır. Yazarlar, yaklaşımın

2016). Kanser hastalarının, hastalıkları hakkındaki duygularını Twitter'da açık ve samimi bir şekilde anlattıklarını bulmuşlardır.

Twitter' dan toplanan 36 milyon tweet' e dayanarak, Wang ve arkadaşları, 2012 ABD başkanlık seçimleri sırasında siyasi tweetlerin sınıflandırılması için gerçek zamanlı bir duygu analizi sistemi önerdiler. Onların modeli, siyasi tweetlerin duygularını tahmin etmede %59 doğruluk elde etmiştir (Wang ve diğerleri, 2012). Başka bir araştırma çalışmasında, Park ve diğerleri 2012 Güney Kore cumhurbaşkanlığı seçiminde TV tartışmaları sırasında tartışılan konuların Twitter'daki konuşmalara ve siyasi tartışmalara nasıl yansındığını araştırmışlardır (Park, Park, Lim ve Park, 2016). Yakın zamanda yapılan bir araştırma, genel seçimler sırasında Hindistan'daki siyasi kampanyaları analiz etti ve internetin ilk kez seçmenler üzerindeki etkilerini ölçmeyi amaçlamıştır (Ahmed, Jaidka ve Cho, 2016).

Öte yandan Cheong ve diğerleri, Twitter' da toplanan kamuoyu duygu verilerine dayalı olarak potansiyel terörizm senaryoları hakkında yararlı grafiksel görselleştirmeler sağlayan bir duygu analizi modeli sundular (Cheong ve Lee, 2011). Bir vaka çalışmasında Yates ve Paquette, 2010 Haiti Depremi sırasında sosyal medya teknolojilerinin kullanımını ve bunların bilgi paylaşımı ve etkili karar verme üzerindeki etkilerini araştırdılar (Yates ve Paquette, 2011). Vaka çalışması, sosyal medya ve bilgi yönetimi teknolojilerinin acil durum bilgilerini paylaşarak afetlere etkili müdahalede kritik bir rol oynayabileceğine bir örnektir. Yakın zamanda yapılan bir başka çalışmada Kim ve Hastak, Louisiana'daki 2016 selinden sonra sosyal medya tepkilerini araştırarak bir afet sonrası sosyal ağların özelliklerini araştırdılar (Kim ve Hastak, 2018). Acil durumlarda afetle ilgili bilgilerin sosyal ağlarda farklı şekilde yayıldığını gözlemlediler. Benzer şekilde Neppalli ve diğerleri acil durumlarda bilgi elde etmek için Twitter'da duygu analizinin kullanımını araştırdı. Kasırga sırasında Twitter kullanıcılarının etkilenen yerlerdeki duygularını analiz ettiler. Ayrıca konuma dayalı duygu değişiklikleri sundular (Neppalli, Caragea, Squicciarini, Tapia ve Stehle, 2017).

Halk sağlığı açısından, bulaşıcı hastalıkların yayılması da sosyal medya verileri kullanılarak çeşitli çalışmalarla analiz edilmiştir (Lampos, De Bie ve Cristianini, 2010 ; Broniatowski, Paul ve Dredze, 2013 ; Elkin, Topal ve Bebek, 2017). Sağlıkla ilgili bir konuda duygu analizi yapan başka bir çalışmada, Twitter kullanıcılarının nargile içiciliği hakkındaki görüşlerini Twitter verilerini kullanarak incelemişler ve tweetlerin %59' unun

nargile içimi konusunda olumlu düşüncelere sahip olduğunu bulmuşlardır (Grant ve O'Mahoney, 2016).

Wu ve Shen, Twitter'daki haberlerin popülerliğini tahmin etmek için bir duygu analizi modeli önerdiler (Wu ve Shen, 2015). Twitter'daki haber yayılımının özelliklerini araştırdılar ve haberin popülerliği ile “retweet” lerin haber kaynağıyla etkileşim sıklığı arasında bir ilişki olduğunu bulmuşlardır. Japon Twitter kullanıcılarına odaklanan yeni bir araştırma, aşırı bilgi yüklemesini ve kullanıcı tepkilerini araştırdı. Yazarlar, kullanıcıların zaman tünellerinde tüm tweet beslemelerine bakmama eğiliminde olduklarını, ancak yine de yeni kullanıcıları takip etmeye devam ettiklerini bulmuşlardır (Sasaki, Kawai ve Kitamura, 2016). Kullanıcıları daha iyi anlamak için video oyunu sohbet mesajlarındaki duyguları analiz eden bir başka ilginç araştırma çalışması da yakın zamanda yayınlandı (Thompson, Leung, Blair ve Taboada, 2017). Yazarlar, kullanıcılar tarafından oluşturulan uydurma bir dilin sözlükte dikkate alındığı sözlük tabanlı bir yaklaşım önerdiler. Örneğin "gg", sözlükte "iyi oyun" anlamına gelir. Sözlük tabanlı yaklaşımların taşınabilir olduğunu ve modeli diğer veri kümelerine uygulamak için gerektiğinde sözlüklerin kolayca güncellenebileceğini öne sürmüşlerdir. Sözlükler, analiz edilen metin verilerine göre oluşturulur. Böylelikle, bu yaklaşımlar, veri seti için iyi hazırlanmış bir sözlük kullanımı gerektiğini belirtmişlerdir.

Mikroblog platformlarındaki duygu analizi, farklı dillere odaklanan araştırmacıların da dikkatini çekiyor. Tellez vd. İspanyolca dilinde bir duygu analizi çalışması gerçekleştirmiş ve sınıflandırıcı olarak Support Vector makinelerini kullanarak metin dönüşümlerini analiz etmiştir (Tellez ve diğerleri, 2017). Metin sınıflandırması açısından Liao ve diğerleri, 30.000 Çinli mikro blog üzerinde duygu analizi yaptılar ve mesajları negatif, nötr ve pozitif kategorilere ayırmışlardır (Liao, Feng, Yang ve Huang, 2016). Film verilerini kullanan Khan ve diğerleri, Hintçe 50.000 film yorumunu incelemiştir (Khan, Qamar, ve Bashir, 2016). Öztürk ve Ayvaz ise Twitter da Suriyeli mülteci krizine yönelik duyguları karşılaştırmalı bir şekilde araştırmaya yönelik ilk çalışma sunmuştur (Öztürk ve Ayvaz, 2018)

We Are Social ve Hootsuite şirketinin hazırladığı “Digital in 2020” raporuna göre, Twitter da yer alan aktif kullanıcı sayısı ile Türkiye, dünyada 6. sırada bulunmaktadır. Ancak Duygu Analizi için Türkçe dilinde yapılan çalışmalar oldukça sınırlıdır. “Science Direct” veri tabanında yayınlanan bu alandaki çalışmalara bakıldığında, Öztürk ve Ayvaz 2018 yılında

Türkçe ve İngilizce tweetlerden Suriyeli mültecilere yönelik olumlu ve olumsuz duyguların karşılaştırılmasını yapmıştır. Çalışma da İngilizce cümle düzeyinde duyguları analiz etmeye yardımcı olan ve bir duygu puanı sağlayan kapsamlı bir duygu sözlüğü içeren R Sentiment Paketi kullanılmıştır. Ancak Türkçe için böyle bir sözlük bulunmamasından dolayı bir alana özgü bir sözlük oluşturulmuştur (Öztürk ve Ayvaz, 2018). Diğer bir çalışma ise Türkiye'deki üniversitelerin yenilikçilik girişimcilik endeksine göre sıralanmasında Twitter verileri kullanılmaktadır. Alana özgü sözlük oluşturularak üniversiteler arasında karşılaştırma yapılmıştır (İskender ve Batı, 2015).

Saldırgan Dil ve Sosyal Medya Etkileri ile İlgili Çalışmalar

Yakın zamanda ortaya çıkan aktif bir araştırma alanı olan duygu analizi, DDİ yöntemlerini kullanarak kullanıcıların metin verilerinden insanların duygu, görüş, düşünce ve davranışlarını çıkaran bir disiplindir (Heimann ve Danneman, 2014: 2). Dahası, duygu analizi, metin sınıflandırma problemine vurgu yaparak fikir madenciliği olarak da bilinir. Web ölçekli metin verilerinden duyguları hakkında bilgi çıkarmak, büyük miktarda veri nedeniyle çok zor ve pahalı bir görev olabilmektedir (Fernández-Gavilanes, Álvarez-López, Juncal-Martínez, Costa-Montenegro ve González-Castaño, 2016). Tarihsel olarak, Web çok fazla öznel veriye sahip değildi. 2000'li yılların başından sonra sosyal ağların patlamasıyla insanlar duygularını ve görüşlerini sosyal medya üzerinden paylaşmaya başladı. Artan küresel kapsam ve sosyal medyanın yüksek etkisi nedeniyle, insanların sosyal medya üzerinden paylaşılan bazı konulardaki görüşleri önemli ölçüde etkileyebilir. Ancak sosyal medyanın sunduğu korunan özellikleri nedeniyle bireyleri hedef alan nefret ve saldırgan söylemin giderek arttığına tanık oluyoruz.

Son on yılda, nefret ve saldırgan dilin sistematik olarak tespiti, sosyal ağ sitelerinde artan zararlı ve saldırgan içeriğin artması nedeniyle birçok akademisyenin dikkatini çekmiştir. Bu sitelerin çoğunun, ırk, etnik köken, din, cinsiyet ve milliyet gibi korunan özelliklerine göre bireylere veya gruplara saldıran veya tehdit eden gönderileri yasaklayan nefret ve saldırgan dil söylemi politikaları vardır. Nefret ve Saldırgan dil söylemi tespiti alanında çeşitli araştırma çalışmaları ve TRAC 2019-2020 gibi yarışmalar ve atölyeler ortaya çıkmıştır. (Kumar, Ojha, Malmasi ve Zampieri, 2018), SemEval 2019-2020 (Basile ve diğerleri, 2019) ve OffenseEval 2019-2020 (Zampiere ve diğerleri, 2019), bu konunun artan önemini açıkça vurgulamaktadır.

İngilizce saldırgan dil söyleminin tespiti konusunda bir miktar ilerleme kaydedilmiştir. Çalışmalar saldırgan dil söyleminin türüne odaklanmıştır. Çalışmaların yalnızca İngilizce dilince olmasının eksikliğini giderilmesi adına yukarıda bahsedilen yarışmalar da çok dilli saldırgan dil tespitine yönelik duygu analiz çalışmaları istenmiştir. SemEval yarışması, 2020 yılı için 12. görevi olarak sosyal medya da çok dilli saldırgan dil tespiti yarışması düzenlemiştir.

Yarışma da görevler:

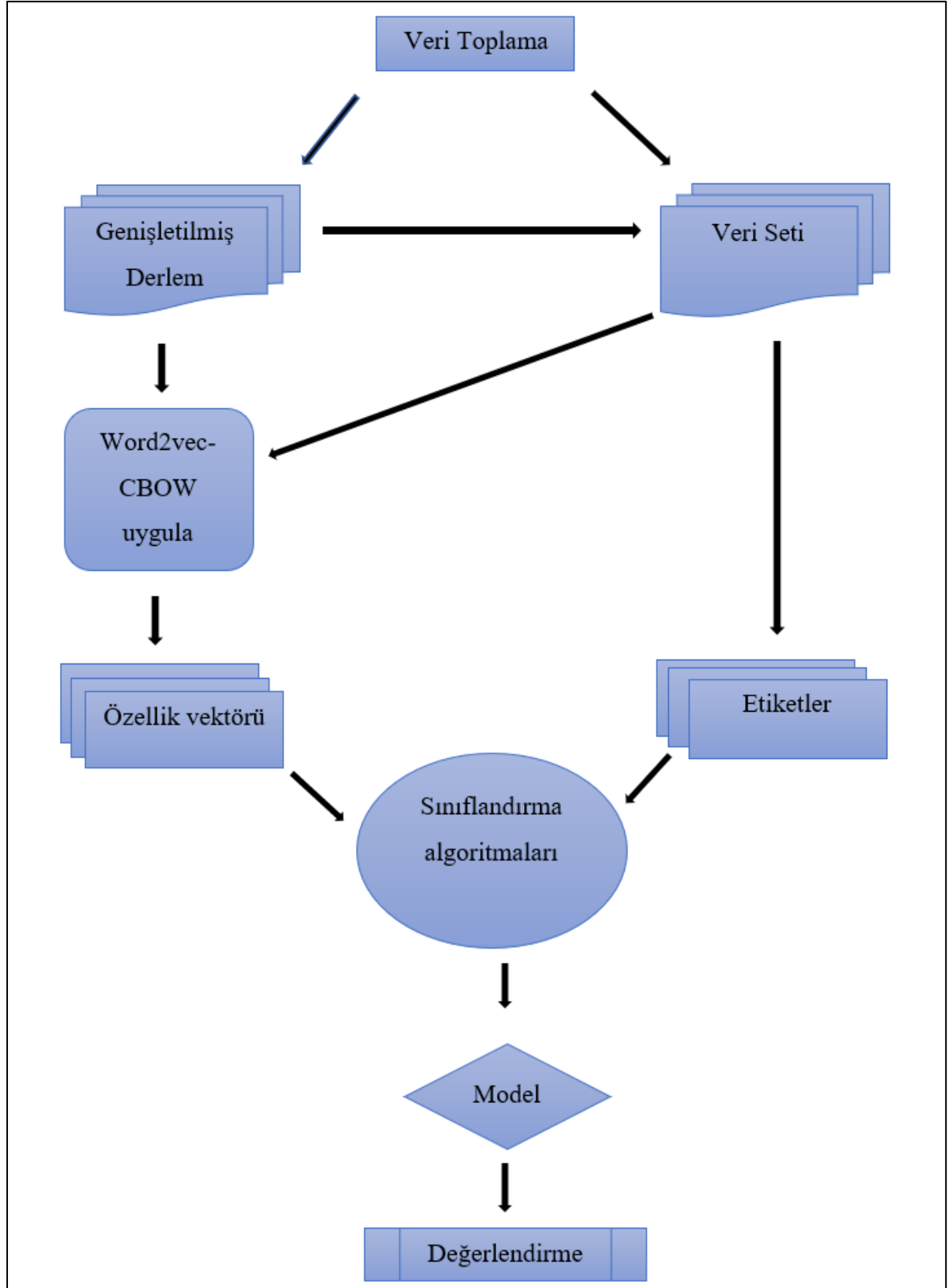
- a. Yarışma saldırgan dilin tanımlanması
- b. Saldırgan tiplerinin otomatik olarak kategorilendirilmesi (kişi veya gruba yöneltilen ya da hiç kimseye ithaf etmeyen)
- c. Eğer kişi veya gruba yönlendirilen bir saldırgan dil söz konusu ise bunları kategorilendirme (Grup, Kişi ve diğerleri olarak)

Yarışma veri seti Twitter'dan alınmıştır. Yarışma da istenen diller arasında İngilizce, Yunanca, Almanca, Danimarka dili, Arapça ve Türkçe vardır. Galileo Türkçe dili dahil tüm diller de ilk sıradadır (Zampiere ve diğerleri, 2020). Sonraki sıradaki ise Türk araştırmacıdır. Yaptığı çalışmada en yüksek performanslar, Word2Vec ile CNN-LSTM (F1: 0,773), Word2Vec ile CNN-Attention (F1: 0.760), BERTurk (F1: 0,808) modelleri ile sağlanmıştır (Özdemir ve Yeniterzi, 2020).



7. UYGULAMA

Uygulama için kullanılan yöntem Şekil 7.1.' de gösterilmektedir. İlk aşama verilerin Twitter uygulamasından toplanma aşamasıdır. İkinci aşama, toplanan verilerden iki ayrı veri seti elde ettiğimiz aşamadır. Bunlar 'genişletilmiş derlem' ve 'veri seti' olarak adlandırılmıştır. Üçüncü aşama da veri setinde yer alan tweet metinlerini etiketleme işlemi yapılmıştır. Etiketleme yapılırken belirli kriterler kullanılmış ve 4 ayrı kişi tarafından etiketleme yapılmıştır. Dördüncü aşama da kelime temsil yöntemlerinden word2vec-CBOW uygulanmıştır. Burada tweetler de yer alan boşluk gözetilerek her bir kelime veya ifade içeren varlık için vektörel temsil elde edilmiştir. Beşinci aşama da karşılaştırılacak olan genişletilmiş derlem ve veri setinden elde edilen iki vektörel temsil ayrı ayrı kullanılarak CNN, LSTM, Bi-LSTM, GRU, Bi-GRU ve CNN-LSTM sınıflandırma algoritmaları uygulanmıştır. Her bir algoritmaya uygun parametreler belirlenip model seçilmiştir. En son aşamada ise, her modelin performans değerlendirme ölçütleri elde edilmiştir. Bu ölçütler, genişletilmiş derlemin sınıflandırma algoritmalarının performansına etkisini göstermiştir. Buna ek olarak sınıflandırma algoritmalarını da karşılaştırabilmemizi sağlamıştır.



Şekil 7.1. Uygulama akış şeması

7.1. Veri Toplama

Bu başlıkta oluşturulan iki ayrı veri setinden bahsedilecektir. Sınıflandırma için oluşturulan veri setini etiketleme kriterleri örnekleriyle açıklanmıştır.

7.1.1. Veri setleri

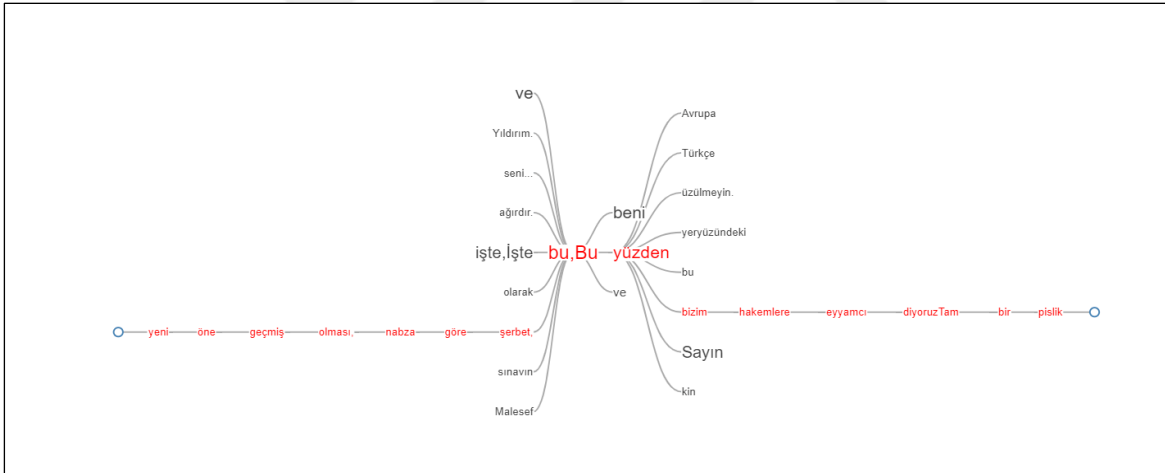
Bir veri setinin dört temel özelliği; gerçek dünyaya uygun, nispeten büyük, elektronik yollarla elde edilmiş ve alana özgü kriterleri barındırıyor olmasıdır (Bowker ve Pearson, 2002). Akademik dil, konuşma dili, sosyal medyada kullanılan gayri resmi dil, gazete yazıları, romanlar- hikayeler ve yasal davalar gibi kısa metinlerin yanı sıra metin türleri olarak bilinen çeşitli dil dizgesi içeren veri setleri oluşturulmaktadır. Öte yandan her bir dilin kendi modeline ve keskin dil kalıbına sahip olduğu görülmüştür (Römer, 2009:140). Buna ek olarak veri setinin boyutuyla ilgili belirli kural yoktur. Ancak veri seti kullanılan hedef dil kalıplarının sistematik bir analizini yapacak kadar büyük olması gerekmektedir. Böylelikle bir kelimenin yalnızca kendisinin metne yön veremediği aynı zamanda çevresindeki kelimelerin öğrenmeye katkısı olduğu bulunmuştur. Bu da bir kelimenin anlamını ve faydasını artıran bir durumdur.

Bu çalışma da kendi oluşturduğumuz veri setini kullanacağız. Bunun için Python programlama dili kullanılmıştır. Türkçe’ de bilinen 150’den fazla saldırgan dil içeren kelime ve kelime öbekleri aratılarak veri setimiz oluşturulmuştur. 14.02.2020- 17.06.2020 tarihleri arasında filtreleme yapılarak 1 milyon 860 bin tweetten oluşan bir genişletilmiş derlem elde edilmiştir. Ancak veri toplama periyodunun kısa olması tweet çeşitliliğinin yetersiz olmasına yol açabilmektedir. Bu dezavantajı gidermek için Python da geçmişe dayalı arama kodu kullanılarak filtreleme yapılmıştır. Böylelikle genişletilmiş derlemde kullanılan tweetlerin tarihleri 5.09.2010- 17.06.2020 tarihleri arasındadır ve bu derlemde 3 milyon 859 bin 322 adet eşsiz ifade bulunmaktadır. Her bir ifade için çeşitli boyutlarda vektörel temsiller elde edilmiştir. Derlem için word2vec olarak bilinen kelime temsil yöntemi kullanılmıştır. Böylelikle vektör uzayında anlamsal olarak yakın kelimeler bilinmektedir. Derlem kullanımının öğrenmeye bariz fayda sağladığı görülmüştür (Yılmaz, Özer ve Gökçen, 2021). Çünkü derlem kullanımı, sınırlı verilere dayanmak yerine derlemde yer alan kelimelerin frekans dağılım özelliklerinin ve ilgili konunun gerçek dünyayı yansıtan dil kullanımının

gözlemlenmesine ve incelenmesine izin vermektedir. Bununla birlikte belirli bir alana uygun oluşturulan derlem sözcük kalıplarına dayalı yeni duygu sözcüklerini keşfetmektedir.



Şekil 7.2. Genişletilmiş derlemde elde edilen kelime bulutu



Şekil 7.3. Genişletilmiş derlemde elde edilen kelime ağacı modeli

Şekil 7. 2’de genişletilmiş derlemde yer alan tweetler de en çok geçen 500 kelimenin kelime kullanım sıklığına göre boyutunun belirlendiği bir kelime bulutu elde edilmiştir. Örneğin bu-bir-ve-ama-için-kadar kelimeleri tweetler de oldukça fazla geçmektedir. Bu kelimeler veri temizleme işlemi yapılarak derlemde çıkarılabilir. Yapılan bazı çalışmalarda belirli kullanım sıklığının üzerinde olan kelimelerin veri ön süreç aşamasında temizlendiği görülmüştür. Ancak Çalışmada bu kelimelerden saldırgan içerik barındıran tweetler belirlenmiştir. Örneğin; veri setimizde de yer alan “bu kadar mı insansınız” tweet’i için ‘bu-kadar’ kelimeleri anlam ifade etmektedir. Bununla birlikte saldırgan içerikleri barındıran

metinlerin dil kalıbının beklenildiği gibi gerçek dünyayı yansıtması için veri temizleme işlemi yapılmamıştır. Şekil 7.3’ de derlemden elde edilen kelime ağacı modeli bu dil kalıbına örnektir. Burada birbirleriyle en çok kullanılan kelimelerin sırası gözetilerek bir kelime ağacı oluşturulmuştur. Şekil de de görüldüğü gibi anlamlı bir cümle elde edilebilmiştir.

Şekil 7. 3. Genişletilmiş derlemden elde edilen kelime ağacı modeli etiketlenerek eğitimde kullanılacak olan veri seti, genişletilmiş derlemden ve Google Colab’ dan 14 bin 752 tweet seçilerek elde edilmiştir. Seçilmiş olan tweetler genişletilmiş derlemden silinmiştir. Seçme işlemi yapılırken dikkat edilen durumlar:

- Birçok harfin bir araya gelerek anlamsız ifade oluşturduğu tweetler saldırgan dil için önemi belirtilmiştir ve diğer çalışmaların aksine veri setine dahil edilmiştir. Çünkü bu ifadeler ironi ve alaycı tutuma sahip metinlerde kullanılmaktadır (Filatova, 2012).
- Veri çeşitliliğini sağlamak için derlem de yer alan 150’ den fazla saldırgan içerik veri setine dahil edilmiştir.
- Eşsesli sözcüklerin farklı kullanımları veri setine dahil edilmeye çalışılmıştır. Örneğin; yollu ve alçak kelimesi gibi (Çizelge 7. 1).

Çizelge 7.1. Veri setinde yer alan eşsesli sözcükler

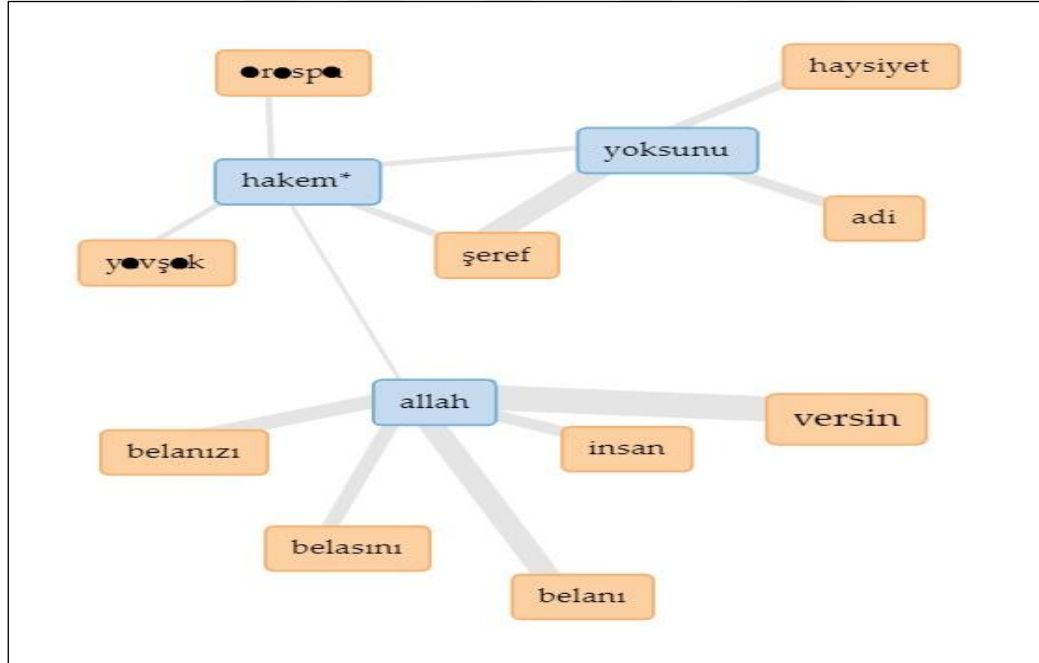
Sen yollu bir bi olduğundan mütevellit çabuk beğeniyor olabilirsin tabii, normal	Saldırgan
Ucuz yoldan sonuca gitmek gibi bir hastalığa sahibiz%9x ımız	Saldırgan Değil
Telefonun elimdeki konumunu hiç değiştirmedigimden dolayı parmak uçlarım uyuşuyor sanki ve hafif yollu şarjım dibini çekiyor ayy ayy #TurkishArmysAlwaysLoveYouBTS	Saldırgan Değil
Ben de şaka yollu şecttim zaten	Saldırgan Değil
Orta yollu bir şey yaparlar	Saldırgan Değil
Pislik konuşmaya değmez aşağılık alçak	Saldırgan
Ebru hanım, çok alçak gönüllü gerçekten	Saldırgan Değil
Yapılıyor, ancak zaman gerekmekte. Şu anda denenmiş sadece Hisar A var, o da alçak irtifa 15-20 km menzil	Saldırgan Değil

- Bağlama bağlı anlam farklılığı gösteren tweetler dâhil edilmeye çalışılmıştır (Çizelge 7.2).

Çizelge 7.2. Bağlama bağlı anlam farklılığı gösteren tweetler

Hayvan gibi güzel oynadı herif	Saldırgan Değil
“Çık dışarı hayvan herif.” B.K	Saldırgan
Sen de az k*lt*k değilsin he	Saldırgan

Etiketli veri setimizden elde edilen belirli kelimelerin anlamsal yakınlık göstermesi bazı durumlarda saldırgan dile maruz kalan grupları da göstermektedir. Örneğin Şekil 7.3, ‘hakem’ kelimesiyle birlikte en çok kullanılan 5 kelimedenden oluşturulmuştur. Buna göre ülkemizde hakemler saldırgan dile maruz kalan bir gruptur.



Şekil 7.4. 'hakem' kelimesinin anlamsal yakınlıklarının gösterimi

Veri Setinin Etiketlenmesi

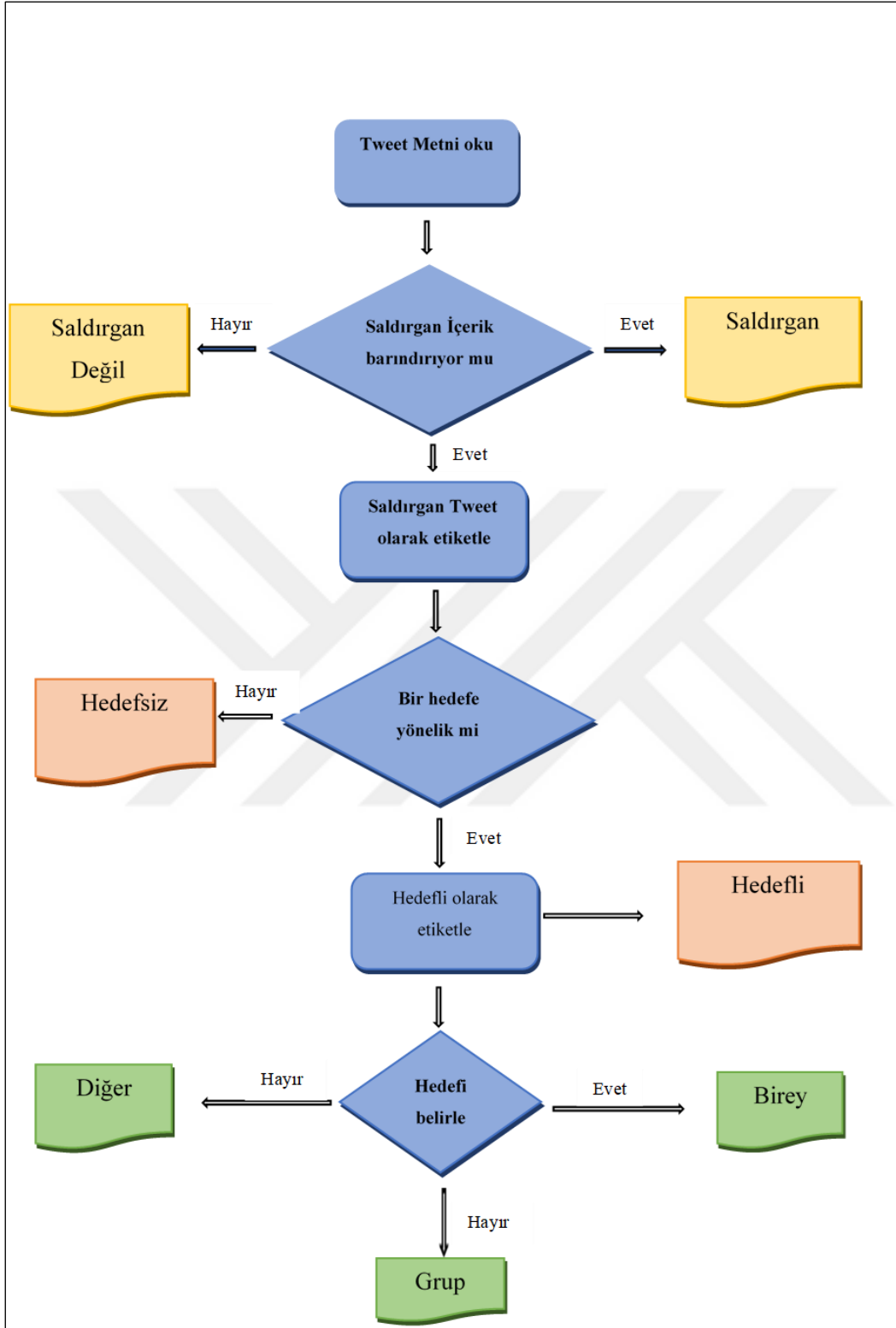
Zampiere ve diğerleri tarafından benzer şekilde etiketleme yöntemi izliyoruz (Zampiere ve diğerleri, 2019). Tanımlarımız bu çalışmadan farklılıklar göstermektedir. Sınıflama etiketleri için Şekil 7.4’ de gösterilen akış şemasını kullanmaktayız. Burada gösterilen sarı, mavi ve yeşil çıktıların her biri ayrı bir sütuna ait etiketleme çıktısını temsil etmektedir.

Böylelikle çalışmada çoklu sınıflandırma modeline uygun etiketleme yapıldığı görülmektedir. İlk olarak bir tweet metni saldırgan veya saldırgan değil olarak ayrılmıştır. Daha sonra saldırgan olan tweet metninin bir hedefe yönelik bir içerik barındırıp barındırmadığına bakılmıştır. Bir hedef belirtmeyen tweet metni hedefsiz olarak etiketlenmiştir. Hedefsiz tweetler saldırgan içerik barındırmaktadır. Bu sebeple çocuklar için filtreleme görevinde kullanılabilir. Bununla birlikte bu içerikler şirketlerin kurumsal yapısına da zarar vermektedir. Tweet metni belirli bir hedefe yönelik ise, hedefli olarak etiketlenmiştir. Hedefli olarak etiketlenen tweetler, birey, grup ve diğer olarak tweet metninin hedef aldığı türlere göre ayrılmaktadır. Buradaki “grup” tanımı, herhangi bireyler topluluğunu (örn. mülteci, futbol takımı, grup gibi) ya da birden çok kişiyi (örn. hainler) ifade etmektedir. Birey ise, grup tanımına uymayan kişiyi hedef alan suçlardır. Suçun hedefinin bu durumlardan herhangi birine uymadığı durumlar vardır, tipik olarak bir organizasyon veya olay gibi insan olmayan bir varlığa yönelik suçları kapsamaktadır. Bu gibi durumlarda, hedefi diğer olarak işaretliyoruz.

Saldırgan içeriklerin belirlenmesi öznel yargı içermektedir. Bu sebeple etiketleyiciler tarafından bir metnin saldırganlığı belirlenirken kişiye göre değişiklik göstermektedir. Zampiere ve diğerleri tarafından yapılan çalışma da etiketlemelerde kişiler arasında %60 uyum olduğu gösterilmiştir (Zampiere ve diğerleri, 2020). Bu alanda yapılan çalışmalar da bu durumun dezavantajından bahsetmiştir. Örneğin, Waseem ve diğerleri, tanımlardaki fikir birliği eksikliğinin nasıl çelişkili yönergeler ve açıklamalarla sonuçlandığını açıkça belirtmektedir (Waseem ve diğerleri, 2017). Başka bir çalışma da ise, 100 metin saldırganlıklarına göre etiketlenmiştir. 3 farklı etiketlemeci arasındaki uyum %61 görülmüştür. Ancak aynı kişilerden oluşan 2 farklı kişideki uyum ise %88 e ulaşmıştır Buna göre kişinin karakteri, hayat görüşü gibi özelliklerin benzerliği uyumu artırmaktadır (Caselli, Basile, Mitrović, Kartoziya ve Granitzer, 2020). Benzer olarak bu çalışmada da etiketlemecilerin saldırgan metinleri etiketlerken uyum sorunu yaşadığı görülmüştür. Örneğin; “*Devleti ve milleti hedef aldı! İşte alçak saldırının faturası*”. Bu tweet metninde yer alan “alçak” kelimesi bir olayı nitelemesi sebebiyle saldırgandır. Ancak aynı zamanda terör olaylarını ifade etmek adına kamuoyunda kullanılan genel bir ifadedir. Bu tweet metni genel bir ifadeyi içermesi sebebiyle saldırgan değil olarak etiketlenmesine yorumlayıcılar tarafından karar verilmiştir. Böylelikle etiket tanımlamaları yapılırken Çöltekin’ in çalışmasında da önemini belirttiği gibi netlik kazandırılmak açısından her durumun ele alındığı etiket tanımlama ilkesi benimsenmelidir.

Çalışmada 4 farklı kişi etiketleme işlemini yapmıştır. Bu kişiler en az lisans düzeyinde öğrenim görmüştür ve anadilleri Türkçe'dir. Etiketlemeler çoğunluğun kararına uygun olarak yapılmıştır. Etiketlemeciler arasında eşit oy söz konusu olduğunda ise farklı 3 etiketleyici oylamaya dahil edilerek karar verilmiştir. Bununla birlikte çalışmada yukarıda belirtildiği gibi daha net ve basit etiketleme kriterleri koyulmuştur. Böylelikle tweet metnini etiketlerken daha tutarlı olmak ve genel yargı içeren etiketlenmiş veri seti elde etmek hedeflenmiştir.





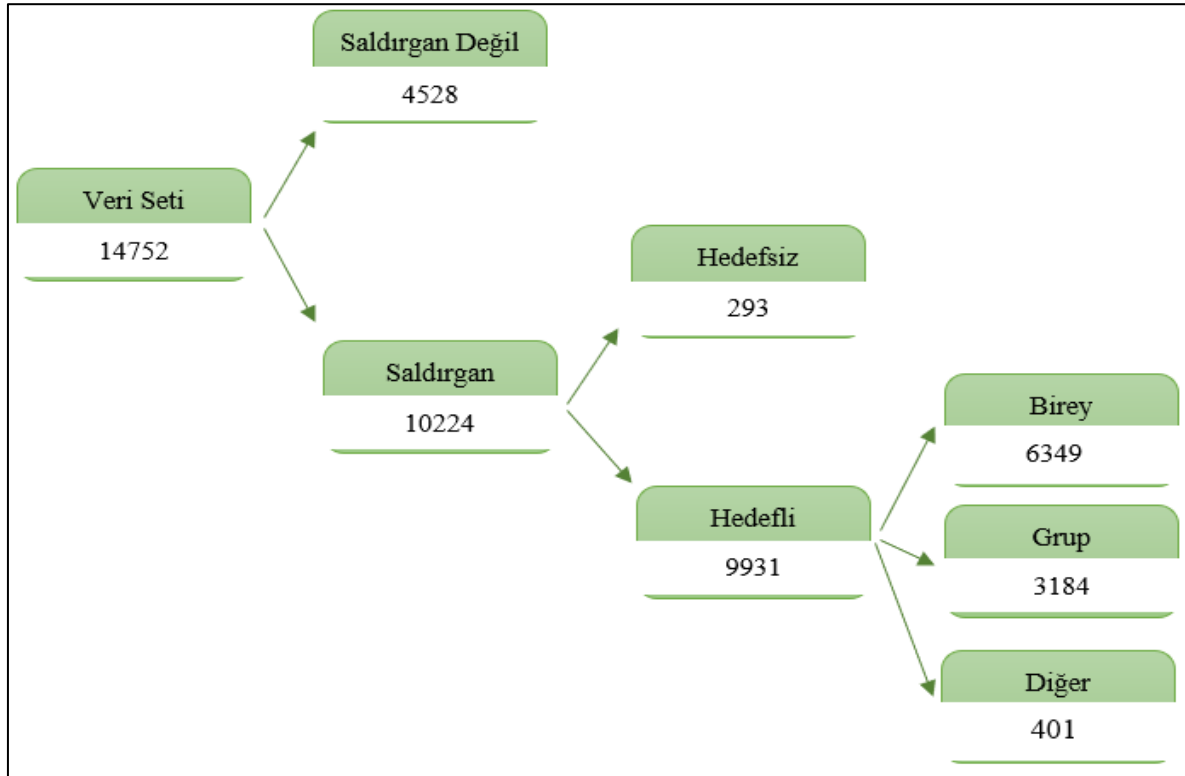
Şekil 7.5. Tweet metninin etiketlenmesi işlemini gösteren akış şeması

Çalışmada yer alan Tweet metnini etiketlemeye örnekler çizelgede gösterilmiştir. Çizelge 7.3’ de gösterilen tweet metnindeki bazı ifadelerin tezdeki gösterimi için sansürlenmek amacıyla ‘*’ kullanılmıştır.

Çizelge 7.3. Tweet metnini etiketleme örnekleri

Dimessiz bir günüm geçmiyordu evde dimes bitmiş  @dimestr	Saldırgan Değil		
Çok şey yapmıyorsunuz s*ktir edin amk. pic.twitter.com/RF7XfqVW1L – at bizimevler 3	Saldırgan	Hedefsiz	
B*k çuvalı gibi hissediyorum.	Saldırgan	Hedefsiz	
bitch or*sp* fah*ş* s*rt*k bundan başka söyleyecek bir şeyiniz var mı????	Saldırgan	Hedefsiz	
Muratbey taze kaşar peynirinin hammadresi seni	Saldırgan	Hedefli	Birey
yasak elma Leyla az pislik değilsin kuduruk karı	Saldırgan	Hedefli	Birey
Ben gerçek gönülden yıllarımı tribünlere adanmış, Fenerbahçe için kan dökmüş saatlerce yolculuk, deplasman yapmış bir adam olarak bu hastalara karşıyım. Yense de yenilse de s<<kimde değil o ayrı. Sonuçta Türk takımı bir 6s değil.! maksat spor olsun... #BaşarılarAjax	Saldırgan	Hedefli	Grup
hepsi bir birinden pislik ve aşağılık . hep yasakçı rezil zihniyet	Saldırgan	Hedefli	Grup
Ya siz nasıl şeref ve haysiyet yoksunu bir firmasınız? 13.12.2019'a iade için kargo randevusu alıyorsunuz, 23.12.2019 oldu Ne gelen var ne giden. Dahası hiçbir mesajıma cevap dahi verilmiyor. @* ve @* dolandırıcıdır. 3 kuruşa çökmek için yapmadıkları rezillik yok	Saldırgan	Hedefli	Grup
# suriyeli demek ;Suriye isimli bir ülkenin vatandaşı olduğu halde benim güzel ülkemın sahibi gibi ortada dolaşan ıpsız sapsız , beleşçi , insanlıktan nasibini alamamış grup demektir !!!	Saldırgan	Hedefli	Grup
# ** kaçırılmış ki bu girme btk da bıkmıştır bunların dolandırıcılığından evde internet bağlattım modemın yanında bile internete bağlanamadım iptal ettirdim sonuç cayma bedeli 1156 tı haram zıkkımın zehir olsun iğrenç bir firma	Saldırgan	Hedefli	Grup
bok gibi bir gün	Saldırgan	Hedefli	Diğer
ragnarok malesef elendin kardeşim y**r*k gibi dizi yapmışsınız sırf norveççe kinkim var diye izliyorum hatta dinliyorum sadece	Saldırgan	Hedefli	Diğer
2020 sen ne karaktersiz bir sene çıktın *m*na koyim ya	Saldırgan	Hedefli	Diğer

Veri setinin boyutuyla ilgili belirli kural yoktur. Ancak çalışmada kullanılan dil kalıbını öğrenmesi beklenmektedir. Çalışmada Türkçe tweetler kullanılmıştır. Veri setimizdeki tweet sayısı 14752'dir. Burada 'Saldırgan Değil' etiketli tweet sayısı 4528 iken 'Saldırgan' etiketli tweet sayısı 10224'dür. Saldırgan etiketli tweetlerin 9931'i 'Hedefli', 293'ü 'Hedefsiz' olarak etiketlenmiştir. Hedefli olarak etiketli tweetlerin 6346'sı 'Birey', 3184'ü 'Grup' ve 401'i 'Diğer' olarak etiketlenmiştir (Şekil 7.5).



Şekil 7.6. Veri setinde etiketleme yapılan tweet sayıları

7.2. Veri Ön işleme

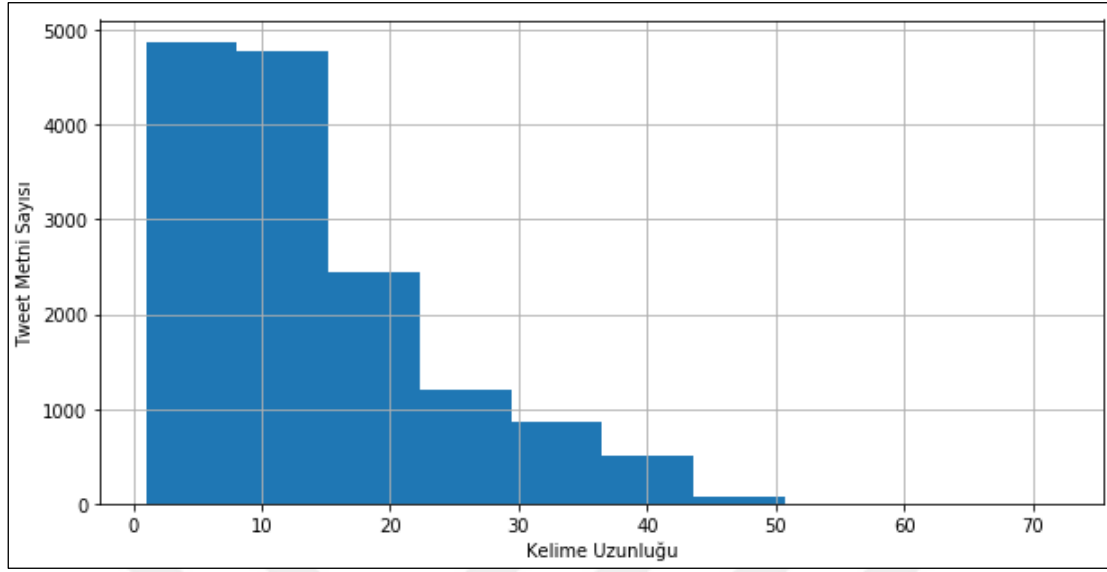
Verileri kodlarken sınıflandırma performansını iyileştirmek için bazı ön işleme adımlarını gerçekleştirmek önemlidir. Çalışmada yer alan veri seti, doğrudan Twitter' dan çıkarılan tamamen ham verilerden oluşmaktadır. Bunlardan faydalı bilgiler çıkarmak için veri ön işleme aşamasına ihtiyaç duyulmaktadır. Ancak ünlem işareti gibi noktalama işaretleri metnin duygusu ile doğrudan ilişkili olabileceği düşünülmektedir. Bu sebeple çalışmada noktalama işaretleri veri setinden kaldırılmamıştır. Buna ek olarak 'hashtag', 'tamamı büyük harf', 'uzatılmış', 'tekrarlanan', 'vurgu', 'sansürlü' 'emoji' içeren tweet metinleri herhangi bir ön işlem uygulamasına dahil edilmemiştir. Çünkü bunlar güçlü bir duygu veya alaycılık ve ironi ifadelerini temsil etmek için kullanılmaktadır (Filatova, 2012).

Daha sonra metin, sınıflandırma algoritmaları tarafından daha kolay öğrenilebilir hale getirmek için sayısal forma dönüştürülmüştür. Bunun için word2vec-CBOW'u kullanarak kelime dizileri yerine gerçek sayı vektör dizileri elde edilmiştir.

7.2.1. Kelime temsil yöntemi

Birçok makine öğrenimi algoritması, girdinin sabit uzunlukta bir özellik vektörü olarak temsil edilmesini gerektirir. Metinler söz konusu olduğunda, en yaygın sabit uzunluk özelliklerinden biri kelime torbasıdır (word2vec-CBOW) (Le ve Mikolov, 2014). Bunun sebebi, basitliği, etkinliği ve sınıflandırma performansına katkısı olarak bilinmektedir. Burada metinlerden oluşan diziler yerine sabit uzunluklu vektörlerden oluşan sayısal ifadeler yer almaktadır. Bununla birlikte çalışmada kullanacağımız LSTM, BiLSTM, GRU, BiGRU, CNN ve CNN-LSTM modelleri eğitim aşamasında sabit uzunluğa sahip verilere ihtiyaç duymaktadırlar (Le ve Mikolov, 2014). Ancak elde edilen tweetler çeşitli uzunluktadır. Bu sebeple belirlenen sabit uzunluklu vektöre uygun olarak uzun tweetler belirli bir yerden sonra alınmamaktadır. Kısa tweetler de ise tweet bittikten sonraki kısmın vektörel temsili "0" olarak yazılmaktadır. Bu çalışma da veri setinde yer alan her bir tweet metninin kelime uzunlukları karşılaştırılmıştır. En fazla 71 kelimeden oluşan veri setimizin yaklaşık olarak %92'i 30 kelimeden, %99'i ise 46 kelimeden az cümle içermektedir (Grafik 7.1). Bu belirtilen kelime uzunlukları dikkate alınarak modellerin girdi uzunluklarına karar verilmiştir. Böylelikle sınıflandırma yapılırken kullanılan kelime uzunluğu 46 olarak seçilmiştir. Buna ek olarak elde edilen genişletilmiş derlemde 25, 50, 75 ve 100 boyutlu vektörel temsiller elde edilmiştir. Bu boyutların büyüklüğünün sınıflandırma performansı daha iyi olan vektörel boyut çalışmaya dahil edilmiştir. Böylelikle sınıflandırma performansına katkı sağlayan en iyi vektörel boyut 25 bulunmuştur. Çizelge 7.5, kelimelerin 25 boyutlu vektörel temsillerine örnektir. Çalışmada tweet metninde yer alan kelimenin birlikte kullanılan kelimeler arasında benzerlik ilişkisi kurularak vektörel temsil elde edilmiştir.

Çizelge 7.4. Veri setinde yer alan tweet metnlerinin kelime uzunluklarının dağılımı



Çizelge 7.5. Veri setinde kullanılan 3 kelimenin 25 boyutlu vektör gösterimi

Fenerbahçe	3.0332024 0.723972 3.46077 -3.4732566 -2.3561437 2.7074313 2.1959655 -4.978826 -1.9381531 -1.6330372 8.622788 3.7074513 2.9207187 -3.498927 -0.6142568 -4.600532 11.249953 -6.4922366 0.18945235 4.5617967 3.1848278 -11.555731 3.209911 -3.0202086 5.7745175
iki	4.643571 0.51053613 -4.363012 -3.9832335 0.5638066 7.306147 -4.1599326 -0.63093364 -2.59678 -4.2963805 -2.7082624 -2.6001842 -2.949042 3.0574667 -4.5959425 -6.9270444 1.8467431 -4.121654 -6.6174483 2.155678 1.7817645 -12.0553665 -6.820993 2.803715 -3.65197
olsa	-3.2525935 -0.6794258 -6.840006 -2.159432 0.23544416 4.833356 0.49306995 -0.3514422 -5.029789 -0.6224053 -1.4755503 -1.2938479 0.1056764 1.2905974 -1.1743172 2.7471185 3.5308106 -7.3458986 -7.2660027 -0.3244698 5.3333807 -7.280303 2.9028053 -3.8608403 -7.650578

Kelime temsil yöntemlerinin çalışma prensibi ise kelimelerin birlikte kullanma sıklığını ele almaktadır. Sosyal medya da gayri resmi ve sözlükte yer almayan ifadeler kullanılmaktadır. Böylelikle Türkçe gibi morfolojik açıdan zengin dillerin kullanıldığı DDİ çalışmalarında herhangi bir dil bilgisi kuralı kullanılarak normalizasyon işlemi yapılmadan sınıflandırma algoritmalarından başarı elde edilmiştir. DDİ, makine çevirisi, duygu analizi, dil modelleme, ad-varlık tanıma gibi çok çeşitli görevleri içermektedir. Son birkaç yılda derin öğrenme, DDİ alanında popüler hale gelmiştir. Derin öğrenme modelleri, farklı DDİ görevlerinde umut verici sonuçlar göstermiştir. Derin öğrenme modelleriyle ilgili bir sorun, veri yoğun çalışma modelini benimsemeleridir. Bu modeller de denetimli öğrenme kullanılırken çok büyük miktarda veri gerektirir. Çoğu DDİ görevi için veri setleri yalnızca birkaç bin insan etiketli örnek içermektedir. Bu, veri kümelerinde derin öğrenme modellerinin sınıflandırma

performansını kısıtlamıştır. Bu sorunun çözümü, önceden eğitilmiş olarak bilinen bir yöntemle gelmiştir (Qiu, Zeng, Zhang, Jiang ve Qiu, 2020). Önceden eğitilmiş yönteminin arkasındaki fikir, metnin genel temsilini öğrenmek ve çok miktarda etiketlenmemiş metinden oluşan genişletilmiş derlemi eğitmektir. Böylelikle alana özel söz dizimsel ve anlamsal özellikleri anlamaktadır. Etiketlenmemiş tümceler kolayca erişilebilir durumdadır ve bu nedenle daha uzun dönemler içeren ve daha kapsamlı verileri eğitmekte kullanılmaktadır. Etiketlenmemiş metinden oluşan derlemin birçok DDİ görevine katkısı görülmüştür (Bojanowski ve diğerleri, 2017). Çalışmada etiketlenmemiş verilerden oluşan genişletilmiş derlem kullanımıyla elde edilen 1 milyon 860 bin tweetten 3 milyon 859 bin 322 eşsiz ifade elde edilmiştir. Burada genişletilmiş derlem kullanımının sınıflandırma performansına katkısının görülmesi amaçlanmıştır.

7.3. Kullanılan Sınıflandırma Algoritmaları

2020 yılında duygu analizi alanında yapılan çalışmalara bakıldığında en çok kullanılan sınıflandırma algoritması sırasıyla LSTM, CNN, GRU, BERT olduğu görülmüştür (Ligthart ve diğerleri, 2021). Ayrıca bu alanda en çok alıntı yapılan makaleler CNN, LSTM ve Bi-LSTM' dir (Ligthart ve diğerleri, 2021). Minaee ve diğerleri ise LSTM ve CNN algoritmasını kullanarak bir hibrit model geliştirmiştir. Bu modelin bireysel modellerden daha iyi performans sağladığını göstermiştir (Minaee, Azimi ve Abdolrashidi, 2019). Bununla birlikte LSTM ve CNN modellerinin, mevcut çalışmaya benzer olan nefret söylemi sınıflandırması için iyi performans gösterdiği belirtilmiştir (Beddiar, Jahan ve Oussalah, 2021). Çalışmada Twitter' dan elde edilen veriler kullanılmıştır. Twitter verilerini kullanan çalışmalarda ise en çok CNN modeli tercih edilmiştir (Ligthart ve diğerleri, 2021). Bunun sebebi Twitter' dan elde edilen metinlerin kısa metinlerden oluşmasıdır. Şekil 7.5' de eğitim modellerinin çalışma prensipleri karşılaştırılmıştır. Çalışmada sınıflandırma algoritması seçilirken, bu alanda yaygın kullanımı ve sınıflandırma performansları dikkate alınmıştır. Bununla birlikte zaman ve doğruluk parametreleri arasındaki dengeye bakılmıştır. Böylelikle çalışmada yukarıda bahsedilen başarısı kanıtlanmış CNN, LSTM, BiLSTM, GRU, BiGRU VE CNN-LSTM modelleri kullanılacaktır. Bu modellerin sınıflandırma performansları karşılaştırılacaktır. Burada kullanılan modellerin çıktısı çoklu sınıflandırmadır. Buna ek olarak makine öğrenmesi yöntemi olan ve duygu analizi çalışmalarında sıklıkla kullanılan SVM modeli kullanılacaktır. Böylelikle derin öğrenme modellerinin makine öğrenmesi modeline göre performansı karşılaştırılacaktır.

Kullanılan Hiperparametreler

Derin öğrenmede hiper-parametreler veri kümemizdeki algoritma davranışını kontrol etmemize izin vermektedir. Hiper-parametre optimizasyonu, derin öğrenme modeli performansını optimize etmeye yönelik bir öğrenme algoritması elde etmeyi amaçlamaktadır. Böylelikle algoritma da kullanılan hiperparametre seti ayarlanmaktadır. Her derin öğrenme algoritması, farklı hiperparametre türleri nedeniyle çok fazla hiperparametre ayarlama sürecine sahiptir (Ali, Mohd ve Fauzi, 2021). Derin öğrenme modellerinin dezavantajlarından biri de modellerin performansının hiper-parametrelerden etkilenmesidir. Çünkü hiper-parametrelerin seçimi çok zordur. Bu seçim, modeli tasarlayan kişiye bırakılmıştır. Probleme ve her bir veri setine özgü olarak değişiklik göstermektedir. Hiper-parametrelerin eğitimde başarı göstermesi için veri setinin büyüklüğü ve çeşitliliğinin önemli olduğu görülmüştür. Böylelikle çalışmada eğitimde kullanılan veri setinin %20'si test veri seti %80'i ise eğitim veri seti olarak ayrılmıştır.

Derin öğrenme modellerine bakıldığında bir optimizasyon problemi olduğu görülmektedir. Bu sebeple modellerde optimizasyon algoritmaları kullanılmaktadır. Burada amacımız yitim değerinin azalmasıyla en kısa eğitim sayısında en iyi öğrenme sağlayabilmek olmuştur. Çalışmada kullanılan optimizasyon algoritması 'Adam' dır. Bu optimizasyon algoritması öğrenme hızı sebebiyle kullanılmıştır. Buna ek olarak 0,2 oranında bir seyreltme işlemi yapılmıştır. Böylelikle negatif değerli katmanlar çıkartılmıştır. Bu durum hem işlem yükünü azaltmıştır hem de modelimize çok fazla güvenmekten kaçınmamızı sağlayarak aşırı uydurma ve az uydurma problemlerine çözüm oluşturmuştur. Ayrıca Srivastava ve diğerleri yaptığı çalışmada seyreltme işleminin çeşitli veri setlerindeki sınıflandırma performansına katkısını göstermiştir (Srivastava, Hinton, Krizhevsky, Sutskever ve Salakhutdinov, 2014). Bir diğer kullanılan parametre ise aktivasyon fonksiyonudur. Sınıflandırma performansındaki başarısı ve yaygın kullanımı sebebiyle 'Sigmoid' aktivasyon fonksiyonu kullanılmıştır.

Çalışmada sınıflandırma yapılırken ikili ve çoklu sınıfların performansları değerlendirilmiştir. İkili sınıflandırmalarda tek bir çıkış nöronu kullanıldığında kayıp fonksiyonu olarak 'binary cross entropy' kullanılmıştır. Çoklu sınıflandırma yapıp iki veya daha fazla çıkış nöronu kullanıldığında ise kayıp fonksiyonu olarak 'categorical-cross-

entropy' kullanılmıştır (Hassan, Âmin, Al Azad, Mohammed, 2016). Böylelikle çalışmada hiper-parametreler belirlenirken benzer çalışmalardaki yaygın kullanımı dikkate alınmıştır.



8. BULGULAR VE DEĞERLENDİRME

Saldırgan dilin tanımlanması ve metinsel verilerden saldırgan dil söyleminin otomatik olarak tespit edilmesi DDI' nin alt alanıdır. Bu alt alanları anlamak, özellikle sosyal medyada, kullanıcı tarafından oluşturulan içerikte saldırgan dil söyleminin belirlenmesine ve azaltılmasına yardımcı olmaktadır. Araştırma için en popüler sosyal medya platformlarından biri, insanların kısa gönderiler kullanmasına izin veren Twitter' dır. Burada, insanların siyasi olaylar, tanınmış kişiler veya ürünlerle ilgili görüşlerini paylaştığı görülmüştür. Dolayısıyla Twitter hem endüstri hem de akademi için önemli bir veri kaynağı haline gelmiştir. Ancak bu platform da paylaşılan içeriklerin çok fazla gürültülü veri barındırması bu alandaki çalışmalar için bir zorluktur. Özellikle Türkçe gibi zengin morfolojik dil yapısına sahip olmak, bu alanda yapılan çalışmaların sayısını da etkilemiştir. Öyle ki Türkçe dilinde saldırgan içeriklerin tespit edilmesine yönelik çalışmanın oldukça az olduğu görülmüştür. Bu tez çalışmasında bu alana katkı sağlamak amaçlanmıştır.

Çalışma da alana özgü verilerden oluşan bir genişletilmiş derlem elde ettik. Çalışmanın bu bölümünde ise bu derlemde oluşturulan kelime temsillerinin başarıya katkısını gözlemledik. Buna ek olarak çalışmada hem ikili hem de çoklu sınıflandırma işlemleri yapılmıştır. Çalışma saldırgan içeriğin tanımlanmasına yönelik Türkçe dilinde ilk çoklu sınıflandırma yapan çalışma olmuştur. Burada CNN, LSTM, GRU, BiLSTM, BiGRU, CNN-LSTM ve SVM modellerinin Türkçe dilinde sınıflandırma başarıları karşılaştırılmıştır. Bu sınıflandırma işlemleri, metin sınıflandırma çalışmalarında en çok kullanılan derin öğrenme modelleridir.

8.1. İki Sınıflı Verilerin Sınıflandırma Başarılarının Değerlendirilmesi

Bu bölümde ilk olarak saldırgan ve saldırgan değil olarak sırasıyla '1' ve '0' olarak etiketlenmiş iki sınıflı verilerin sınıflandırma başarıları değerlendirilmiştir. Burada 14752 tweetten oluşan veri setinin 11802'si eğitim veri seti iken 2950'si test veri setidir. Rastgele seçilmiş olan test veri setinin 913'ü saldırgan olmayan, 2037'si saldırgan içerik barındırmaktadır. Yaptığımız bu çalışmada saldırgan içeriği tespit etmeyi amaçlamıştır. Böylelikle Çizelge 8.1 ve Çizelge 8.2'de görülen başarımların değerleri, saldırgan içeriğin tespit edilebilme performansını göstermektedir. Grafik 8.1'de ise, modeller eğitilirken alana özgü elde edilen genişletilmiş derlem kullanımının modellerin sınıflandırma performansına etkisi

gösterilmiştir. Burada sınıflandırma performans değerlendirme ölçütlerinden F1-skor değeri baz alınmıştır.

Veri setinden elde edilen word2vec-CBOW kelime temsilleri kullanılarak iki sınıflı verilerin sınıflandırma başarımları %90 ile %96 arasındadır (Çizelge 8.1). CNN ve CNN-LSTM modelleri diğerlerine nazaran yaklaşık %6 oranında daha yüksek bir performans göstermiştir. Buna karşın genişletilmiş derlem kullanımının başarımları %96 ile %97 arasındadır. Genişletilmiş derlem kullanımının her durumda sınıflandırma performansını artırdığı görülmüştür (Grafik 8.1). Bu artış %0,3 ile %7 arasında değişmektedir. Öyle ki, genişletilmiş derlem kullanıldığında %96,65 ile en yüksek performans gösteren GRU modelinde yaklaşık %7 oranında iyileşme görülmüştür. Buna ek olarak genişletilmiş derlem kullanılan bütün modeller de kesinlik ve duyarlılık değerleri arasında bir denge söz konusudur. Buna karşın veri seti kullanılan modellerde kesinlik ve duyarlılık arasında fark yaklaşık olarak %9 a kadar çıkmıştır. Ayrıca GRU ve LSTM arasında sınıflandırma performanslarına bakıldığında belirgin bir üstünlük görülmemiştir.

Çizelge 8.1 ve 8.2’ deki pcr-auc değeri saldırgan olan içerikleri tespit etmedeki başarısını dikkate alırken roc-auc değeri her iki sınıfı da tespit etmedeki başarısını dikkate almıştır. Genişletilmiş derlem kullanımı prc-auc değerinde %0,6 ile %4 oranında iyileşme sağlarken roc-auc değerinde %1 ile %10 arasında iyileşme sağladığı görülmüştür. Burada genişletilmiş derlem kullanımının azınlık sınıfı da dahil ederek iki sınıfı tespit etmeye olan katkısının(roc-auc) daha yüksek olduğu görülmüştür. Bu durum, genişletilmiş derlem kullanımının daha iyi öğrenme gerçekleştirdiğini göstermiştir.

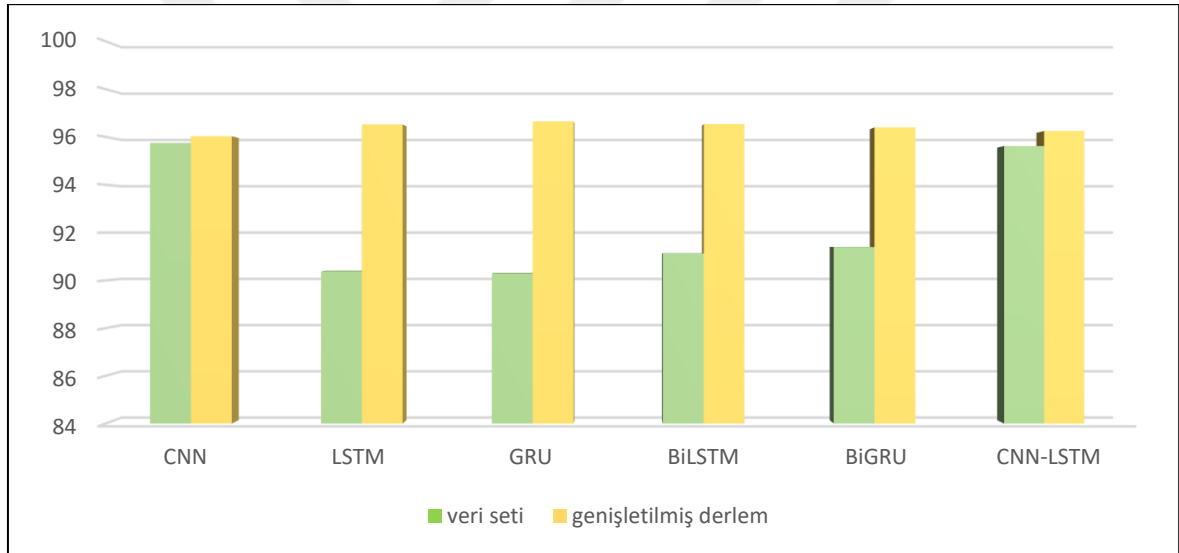
Çizelge 8.1. Veri setinden elde edilen Word2vec kelime temsillerini kullanarak iki sınıflı verilerin sınıflandırma başarımları

Değerlendirme Ölçütleri	Doğruluk %	Kesinlik %	Duyarlılık %	F1-skor %	pcr-auc %	roc-auc %
CNN	94,17	96,09	95,43	95,75	97,34	93,39
LSTM	86,44	90,79	89,99	90,38	93,84	84,81
GRU	88,71	94,19	86,70	90,29	95,03	87,38
BiLSTM	88,07	93,67	88,71	91,12	95,09	87,67
BiGRU	87,59	95,38	87,72	91,39	93,15	82,80
CNN-LSTM	93,93	95,32	95,92	95,62	97,03	92,70

Çizelge 8.2. Genişletilmiş derlemden elde edilen Word2vec kelime temsilleri kullanarak iki sınıflı verilerin sınıflandırma başarımları

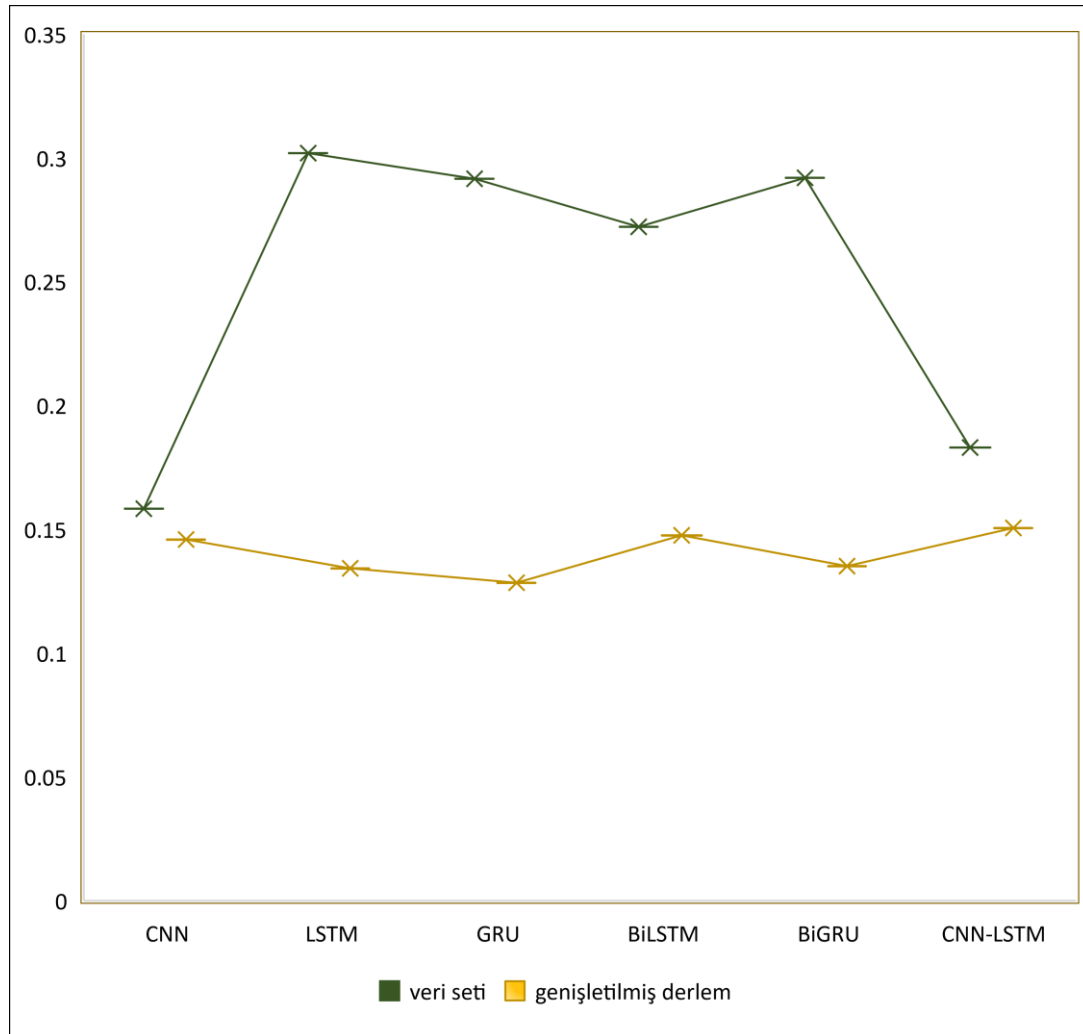
Değerlendirme Ölçütleri	Doğruluk %	Kesinlik %	Duyarlılık %	F1-skor %	pcr-auc %	roc-auc %
CNN	94,61	97,47	94,65	96,03	97,91	94,59
LSTM	95,19	96,47	96,56	96,52	97,70	94,34
GRU	95,39	96,84	96,46	96,65	97,88	94,43
BiLSTM	95,42	96,53	96,86	96,54	97,78	94,54
BiGRU	95,12	96,19	96,61	96,40	97,57	94,03
CNN-LSTM	94,85	96,77	95,73	96,25	97,73	94,30

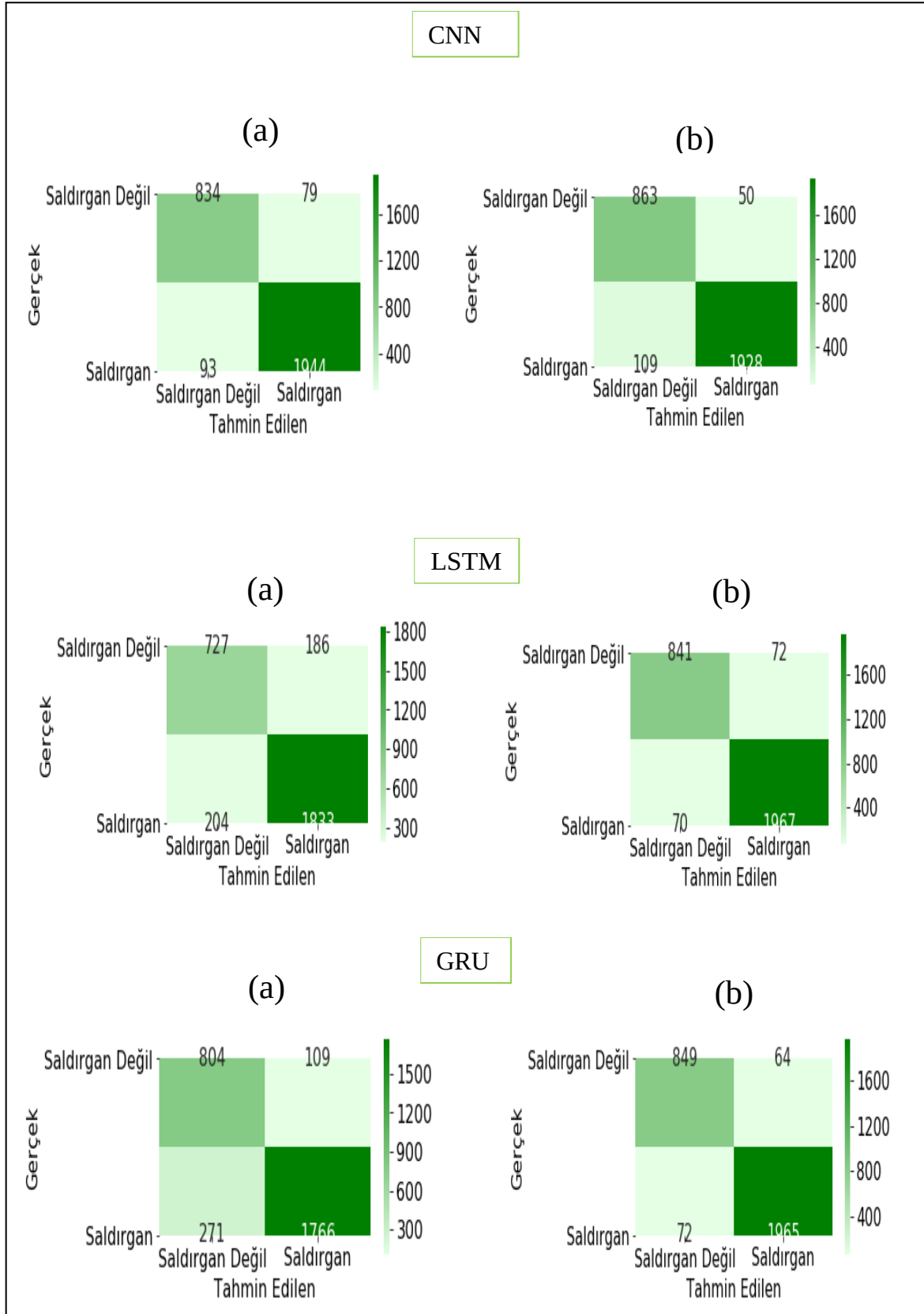
Çizelge 8.3. Genişletilmiş derlem kullanımının iki sınıflı verileri sınıflandırma performansına etkisi



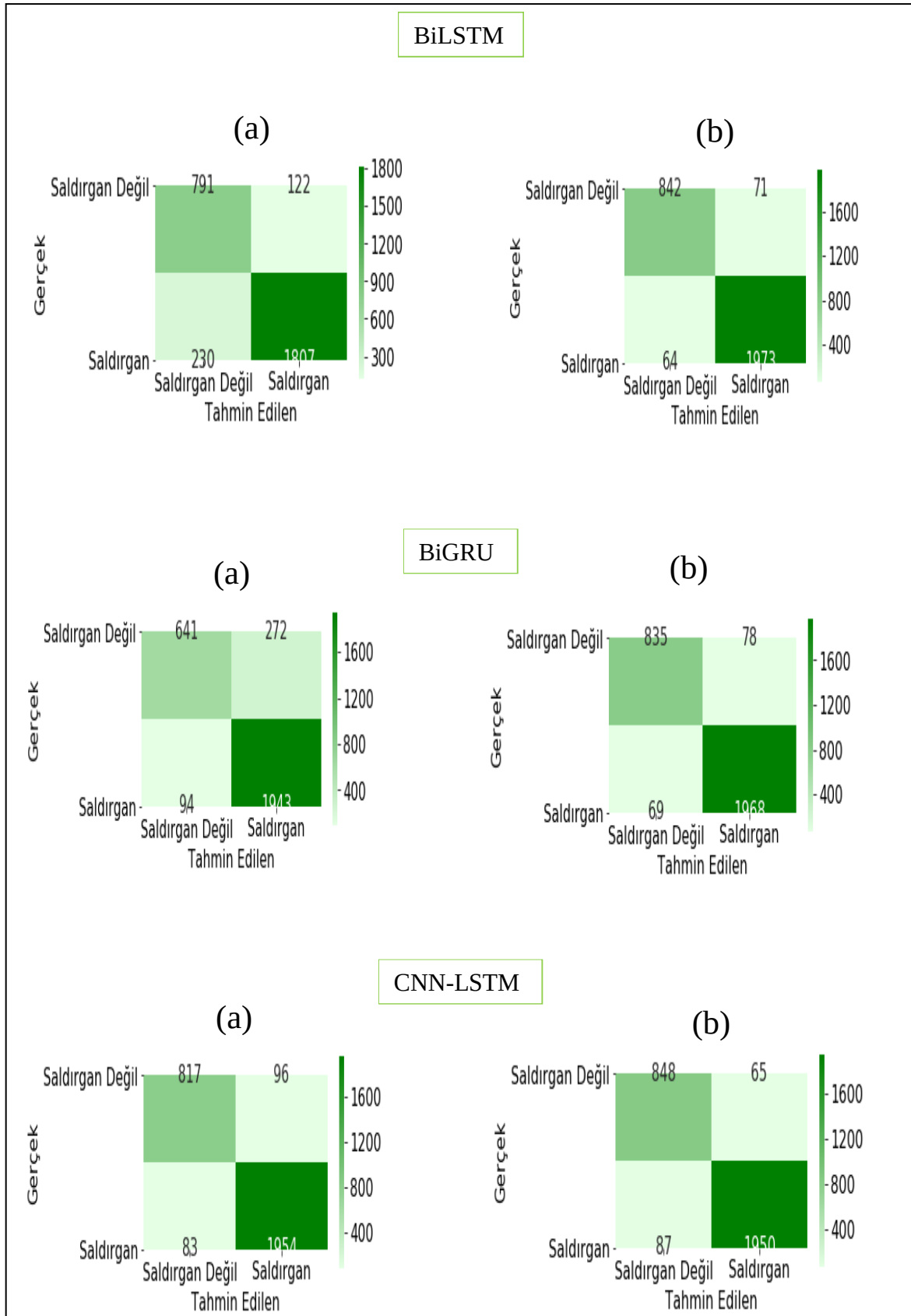
Eğitim sürecimiz sırasında minimum kayıp oranına sahip modellerimizin test setinde gösterdiği sınıflandırma başarımları paylaşılmıştır. Çalışmada genişletilmiş derlem kullanımının kayıp oranlarına etkisi Grafik 8.2’de görülmektedir. Burada GRU modeli en az kayıp oranına sahip model olmuştur. Ancak bu modelde genişletilmiş derlem kullanılmamış olsaydı kayıp oranı en yüksek ikinci model olarak görülecekti. CNN ve CNN-LSTM modeli ise genişletilmiş derlem kullanımından en az etkilenen modeller olmuştur. Buna ek olarak her bir modelin hata matrisleri Şekil 8.1’de gösterilmiştir. Burada genişletilmiş derlem kullanımının hatalı olarak yapılan tahminleri düşürdüğü görülmüştür. Twitter gibi büyük bir platform için her bir hatalı tahmin veri boyutu düşünüldüğünde ciddi oranlara karşılık gelebilmektedir.

Çizelge 8.4. Genişletilmiş derlem kullanımının modellerin iki sınıflı verilerde kayıp değerlerine etkisi





Şekil 8.1. Kullanılan derin öğrenme modellerinin iki sınıflı verilerde hata matris gösterimi (a veri setinden elde edilen hata matrisi, b genişletilmiş derlem kullanılarak elde edilen hata matrisi)



Şekil 8.2. (devam) Kullanılan derin öğrenme modellerinin iki sınıflı verilerde hata matris gösterimi (a veri setinden elde edilen hata matrisi, b genişletilmiş derlem kullanılarak elde edilen hata matrisi)

Karşılaştırılan derin öğrenme modelleri arasında GRU, %96,65 F1 skor değeri ile en iyi performansa sahiptir. Genişletilmiş derlemden elde edilen word2vec kelime temsilleri kullanılarak iki sınıflı verilerin GRU modelinin sınıflandırma başarımları Çizelge 8.3’ de gösterilmektedir. Burada bu modelin saldırgan ve saldırgan olmayan içeriği tespit etmesindeki performans başarımları farkı %4’ dür. Buna ek olarak makro ortalama F1-skor değeri %2 oranında değişiklik göstermiştir. Böylelikle bu model, saldırgan olmayan içeriği saldırgan olarak görme eğilimi göstermiştir. Şekil 8.1 Öyle ki, GRU modeli, 2037 adet saldırgan içerik barındıran tweet’in 57’sini saldırgan değil olarak yanlış etiketlemiştir. Buna karşın 913 adet saldırgan içerik barındırmayan tweet’in 67 tanesini saldırgan olarak yanlış etiketlemiştir (Şekil 8.1). Derin öğrenme modellerinde bir problem olan çoğunluğa yönelim davranışı, mevcut çalışmada da görülmüştür.

Çizelge 8.5. Genişletilmiş derlemden elde edilen Word2vec kelime temsilleri kullanılarak iki sınıflı verilerde GRU modeli sınıflandırma başarımları

	Saldırgan Değil	Saldırgan	Makro Ort.
Kesinlik (%)	93,42	96,84	95,13
Duyarlılık (%)	94,42	96,46	95,44
F1-skor (%)	93,39	96,65	95,02

8.1.1. İki Sınıflı Verilerde SVM Modeli Başarım Değerlendirilmesi

Bu bölümde ilk olarak saldırgan ve saldırgan değil olarak sırasıyla ‘1’ ve ‘0’ olarak etiketlenmiş iki sınıflı verilerin sınıflandırma başarımları değerlendirilmiştir. Burada 14752 tweetten oluşan veri setinin 11802’si eğitim veri seti iken 2950’si test veri setidir. Burada duygu analizi yöntemlerinde sıklıkla kullanılan SVM modelinin derin öğrenme yöntemlerine göre performansı görülmüştür. Böylelikle word2vec-CBOW yöntemiyle SVM modeli sınıflandırma başarımları Çizelge 8.6’da görülmektedir. Burada Derin öğrenme modellerinde en yüksek F1-makro değeri 95,02 iken SVM modelinde 84,84 bulunmuştur. Buna göre saldırgan dili tespit etme başarımında derin öğrenme modellerinin SVM modeline göre bariz bir üstünlük sağladığı görülmüştür. Literatüre bakıldığında derin öğrenme modellerinin sığ modellere göre daha iyi sonuç verdiği bilinmektedir.

Çizelge 8.6. SVM modeli sınıflandırma başarımlarının değerleri

	Saldırgan Değil	Saldırgan	Makro Ort.
Kesinlik (%)	81,76	89,34	85,55
Duyarlılık (%)	76,44	92,05	84,25
F1-skor (%)	79,01	90,68	84,84
Doğruluk (%)	87,09		

8.2. Çok Sınıflı Verilerin Sınıflandırma Başarımlarının Değerlendirilmesi

Çalışmada sınıflandırma işlemi üç aşamalıdır. İlk aşamada veri seti saldırgan ve saldırgan değil olarak etiketlenmiştir. Burada ikili sınıflandırma söz konusudur. İkinci aşamada ise saldırgan olarak etiketlenen veriler, hedefli ve hedefsiz olarak alt sınıflara ayrılmıştır. Bu aşamada saldırgan olan tweet metninin bir hedefe yönelik olup olmadığına bakılarak etiketleme işlemi yapılmıştır. Üçüncü aşama da ise saldırgan olup hedefli olarak etiketlenen veriler birey, grup ve diğer olarak alt sınıflara ayrılmıştır. Burada saldırgan içeriğin yöneldiği hedefe göre etiketleme yapılmıştır. Bu ek olarak çoklu sınıflandırma yapılırken üstün sınıflandırma başarımları sebebiyle GRU modelinin performansına bakılmıştır.

8.2.1. İkinci aşama alt sınıflar bazında başarımların değerlendirilmesi

Bu bölümde ilk olarak hedefsiz, hedefli ve saldırgan değil sırasıyla ‘0’, ‘1’ ve ‘2’ olarak etiketlenmiş çok sınıflı verilerin sınıflandırma başarımları değerlendirilmiştir. Burada 14752 tweetten oluşan veri setinin 11802’si eğitim veri seti iken 2950’si test veri setidir. Rastgele seçilmiş olan test veri setinin 913’ü saldırgan olmayan, 2037’si saldırgan içerik barındırmaktadır. 2037 saldırgan içerik barındıran tweetten 66’sı hedefsiz, 1971’i hedefli olarak etiketlenmiş verilerdir (Çizelge 8.4). Buna ek olarak çalışmada hedefli ve hedefsiz olarak ikili sınıflandırma da yapılmıştır. Burada eğitim veri seti 2044, test veri seti ise 2044 tweetten oluşmuştur. Eğitim veri setinin 54’ü hedefsiz 1990’ı ise hedefli olarak etiketlenmiş verilerdir. Böylelikle ikili ve çoklu sınıflandırma başarımları karşılaştırılmıştır (Çizelge 8.7). Ancak ikili sınıflandırma yapıldığında dengesiz veri problemi yaşanmıştır ve öğrenme sağlanamamıştır. Çoklu sınıflandırmayla ise F1 makro değeri %56,17’e ulaşmıştır. Genişletilmiş derlem kullanımıyla F1 makro değeri %15 oranında artış göstererek %71,97’e ulaşmıştır. Buna ek olarak çoklu sınıflandırma da genişletilmiş derlem kullanımı bütün değerlendirme ölçütlerini %6 ile %30 arasında iyileştirmiştir.

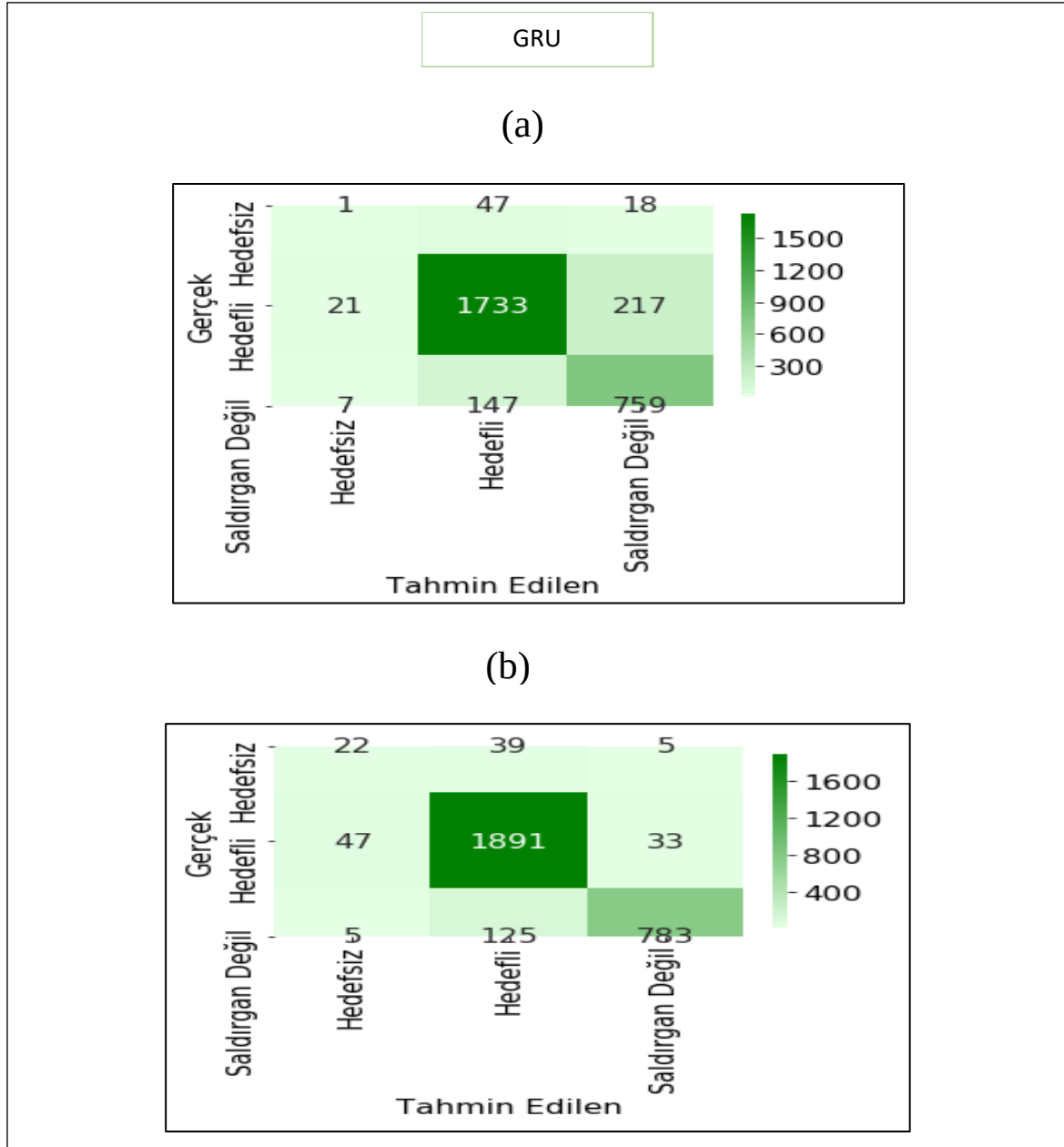
Çizelge 8.7. İkinci aşama alt sınıflar bazında eğitim ve test verilerinin dağılımı

Etiketlenmiş Veri Seti	Eğitim Veri Seti		Test Veri Seti	
	Saldırgan Değil	Saldırgan	Saldırgan Değil	Saldırgan
Saldırgan Değil	3615		913	
Hedefsiz		227		66
Hedefli		7960		1971
Veri Seti	11802		2950	

Çizelge 8.8’ de ikinci aşama alt sınıflandırma performans değerleri gösterilmektedir. Burada sınıflandırma performansının değerlendirilmesinde sadece doğruluk oranlarına bakmanın yanıltıcı olduğu görülmüştür. Öyle ki, ikili sınıflandırma en yüksek doğruluk oranına ve makul kayıp değerine sahiptir. Ancak model öğrenememiş ve tahminlerinin hepsi ‘hedefli’ dir. Burada hedefsiz olarak etiketlenen verileri tahmin edebilmesi çoklu sınıflandırma da genişletilmiş derlem kullanımı ile sağlanmıştır. Böylelikle hedefsiz olarak etiketlenen verileri sınıflandırma performansının ise F1-skor değeri sadece %30 olduğu görülmüştür. Böylelikle modelin performansını etkileyen bir başka faktör de sınıf dengesi olduğu görülmüştür. Sınıf dengesi problemi için veri artırma ve aşırı örnekleme teknikleri önerilmiştir (Massip, M., 2020). Buna ek olarak, saldırgan dil tespitine katkı için düzenlenen OffenseEval yarışma verilerine bakıldığında sınıflar eşit olarak temsil edilmemektedir. Alt sınıflandırmalardaki verilere bakıldığında ise sınıf dengesi problemi daha da artmaktadır. Böylelikle saldırgan dil tespitine yönelik yapılan ileriki çalışmalar da bu problem ele alınacaktır.

Çizelge 8.8. Kullanılan derin öğrenme GRU modelinin ikinci aşama alt sınıflar bazında çoklu sınıflandırma başarımları

<i>GRU</i>		Çoklu Sınıflandırma				İkili Sınıflandırma			
		Veri Seti		Genişletilmiş Derlem		Veri Seti		Genişletilmiş Derlem	
Doğruluk %		85,08		93,83		97,36		97,36	
roc-auc %		73,28		82,93		50,00		50,00	
Test – Eğitim Kayıp Değeri %		39,45	39,25	20,68	16,40	11,89	16,18	10,50	13,34
F1 Makro Ort %		56,17		71,97		49,33		49,33	
Kesinlik %	Hedefli	89,93		91,49		97,36		97,36	
	Hedefsiz	0		30,30		0		0	
	Saldırgan Değil	76,36		86,46		-		-	
Duyarlılık %	Hedefli	87,92		98,12		1		1	
	Hedefsiz	0		30,30		0		0	
	Saldırgan Değil	83,13		95,84		-		-	
F1-skor %	Hedefli	88,92		94,69		98,66		98,66	
	Hedefsiz	0		30,30		0		0	
	Saldırgan Değil	79,60		90,91		-		-	



Şekil 8.3. Kullanılan derin öğrenme GRU modelinin ikinci aşama alt sınıflar bazında çoklu sınıflandırma için hata matris gösterimi (a veri setinden elde edilen hata matrisi, a' genişletilmiş derlem kullanılarak elde edilen hata matrisi)

8.2.2. Üçüncü aşama alt sınıflar bazında başarımlar değerlendirilmesi

Bu bölümde ilk olarak hedefsiz, birey, grup, diğer ve saldırgan değil sırasıyla '0', '1', '2', '3' ve '4' olarak etiketlenmiş çok sınıflı verilerin sınıflandırma başarımları değerlendirilmiştir. Burada 14752 tweetten oluşan veri setinin 11802'si eğitim veri seti iken 2950'si test veri setidir. Rastgele seçilmiş olan test veri setinin 913'ü saldırgan olmayan, 2037'si saldırgan içerik barındırmaktadır. 2037 saldırgan içerik barındıran tweetten 66' sı

hedefsiz, 1971'i hedefli olarak etiketlenmiş verilerdir. 1971 hedefli içeriğin 1233'ü birey, 646 grup ve 92'si diğer olarak etiketlenmiş verilerdir (Çizelge 8. 9).

Çizelge 8.9. Üçüncü aşama alt sınıflar bazında eğitim ve test verilerinin dağılımı

Etiketlenmiş Veri Seti	Eğitim Veri Seti		Test Veri Seti	
	Saldırgan Değil	Saldırgan	Saldırgan Değil	Saldırgan
Saldırgan Değil	3615		913	
Hedefsiz		227		66
Birey		5116		1233
Grup		2538		646
Diğer		309		92
Veri Seti	11802		2950	

Çizelge 8.10' de üçüncü aşama çoklu sınıflandırma başarımları karşılaştırılmıştır. Burada genişletilmiş derlem kullanımıyla F1 makro değerinde %14 oranında iyileşme sağlayarak %54,10 başarımları elde edilmiştir. Bu değer modelin yeterince iyi olmadığını göstermektedir. Burada dengesiz veri problemine yönelik yöntemlerin değerlendirme performansını artıracakları düşünülmektedir. Ancak bu yöntemlerin değerlendirilmesi bu çalışmanın dışındadır.

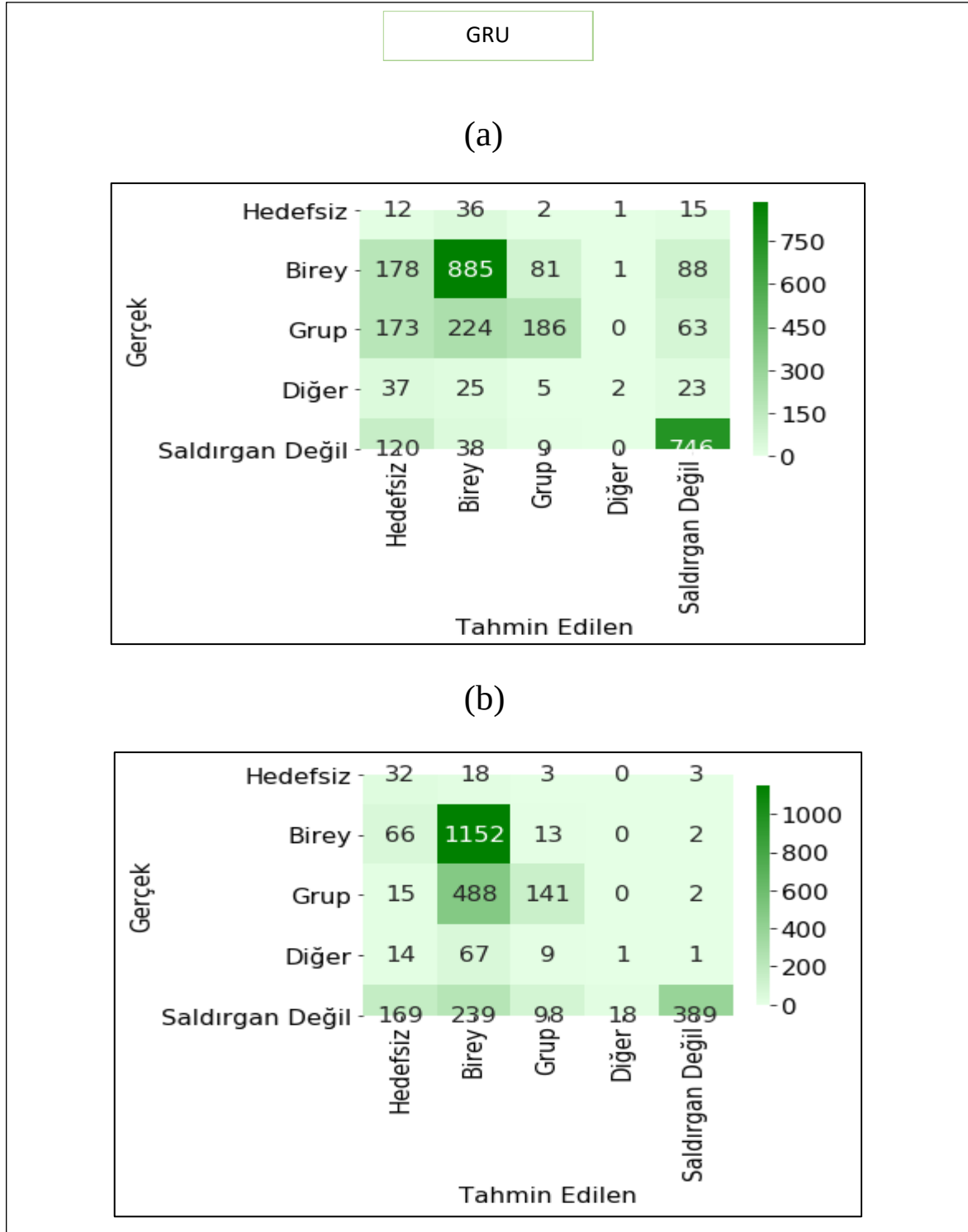
GRU modelinin diğer ve hedefsiz verileri tahmin etmedeki performansı oldukça düşüktür. Genişletilmiş derlemde elde edilen kelime temsillerini kullanan modelin bu değerleri iyileştirdiği görülmüştür. Burada diğer olarak etiketlenen verileri tahmin etme F1-skor değeri %4,17 iken genişletilmiş derlem ile %32,22'ye ulaşmıştır. Hedefsiz olarak etiketlenen verilerde ise F1-skor değeri %0 iken %17,68'e ulaşmıştır. Buna ek olarak çalışmada genişletilmiş derlem kullanımının düşük performans değerine sahip öğrenememiş ağlarda daha yüksek iyileştirmeler sağladığı görülmüştür. Örneğin genişletilmiş derlem kullanımı diğer olarak etiketlenen verilerde F1-skor değeri %28 iyileşme sağlarken birey olarak etiketlenen daha çok örnekleme sahip verilerde sadece F1-skor %2'lik iyileşme sağlayabilmiştir.

Şekil 8.2’ de çizilen hata matrisine göre, GRU modeli genişletilmiş derlem kullanımı ile birey etiketli verilerin doğru tahmin edilmesinde artış görülmüştür. Ancak bu verilerin grup olarak yanlış etiketlenmesinde kayda değer bir artış olmuştur. Burada grup etiketli tweetlerin birey olarak tahmin edildiği görülmüştür. Bu iki etiket türü arasında öğrenmede güçlük yaşandığı söylenebilir.



Çizelge 8.10. Kullanılan derin öğrenme GRU modelinin üçüncü aşama alt sınıflar bazında çoklu sınıflandırması

GRU		Çoklu Sınıflandırma			
		Veri Seti		Genişletilmiş Derlem	
Doğruluk %		68,61		83,86	
roc-auc %		65,21		78,87	
Test – Eğitim Kayıp Değeri %		79,01	77,75	47,90	39,32
F1 Makro Ort. %		39,49		54,10	
Kesinlik %	Birey	73,26		58,09	
	Grup	65,72		40,94	
	Diğer	50,00		21,64	
	Hedefsiz	0		10,81	
	Saldırgan Değil	79,79		83,89	
Duyarlılık %	Birey	71,78		98,70	
	Grup	28,79		96,56	
	Diğer	2,17		63,04	
	Hedefsiz	0		48,48	
	Saldırgan Değil	81,71		96,93	
F1-skor %	Birey	72,51		73,14	
	Grup	40,04		57,51	
	Diğer	4,17		32,22	
	Hedefsiz	0		17,68	
	Saldırgan Değil	80,74		89,94	



Şekil 8 4. Kullanılan derin öğrenme GRU modelinin üçüncü aşama alt sınıflar bazında çoklu sınıflandırma için hata matris gösterimi



9. SONUÇ VE ÖNERİLER

Sosyal medya platformları insanların duygu ve düşüncelerini paylaştığı bir alandır. Bu platformları dünyada bulunan insanların yarısının aktif kullandığı görülmüştür. Bu sebeple burada kullanıcılar tarafından yapılan paylaşımlar akademik alanda oldukça ilgi görmüştür. Bunlardan biri olan Twitter uygulaması kısa metin paylaşımına izin vermektedir. Günümüz toplumlarında, bu platformlarda görüşlerini ifade etmekte özgür olan çok sayıda sosyal medya kullanıcısı vardır. Kullanıcılar arasındaki saldırgan dil kullanan online tartışmaların önemli oranda arttığı gözlemlenmiştir. Bu durum mağdurların psikolojik iyi oluşlarını olumsuz yönde etkilemektedir. Bununla birlikte sadece insanlar değil kurumsal firmalar da bu saldırgan dile maruz kalmaktadırlar. Bu tür şiddetlerin tekrarlanması, bu durumun normalleşme ve daha da artması sonucunu doğurmaktadır.

Son yıllarda Twitter gibi sosyal ağ kurucularının saldırgan dil barındıran zarar verici paylaşımların engellenmesinde aktif rol alması beklenmektedir. Ancak sosyal ağ kurucuları, kullanıcıların konuşma özgürlüğü hakkından ödün vermeden saldırgan dil kullanımını cezalandırmak gibi zorlu bir göreve sahiptir. Ayrıca bu platformlar çevrimiçi topluluklar tarafından oluşturulan çok miktarda veriye sahiptir. Bu sebeple Twitter gibi sosyal ağ kurucuları bu soruna otomatik çözüm bulmak zorunda kalmıştır. Bu sorunun çözümünde insan istihdamı, maliyetli ve çok zaman alıcı olduğu için artık bir seçenek değildir. Buna ek olarak çalışmada saldırgan dil içeren veri seti oluşturma ve etiketleme işlemi yapılırken uzun süren yoğun uğraş göz önüne alındığında psikolojik iyi oluşu olumsuz yönde etkilediği görülmüştür. Bu durum saldırgan dil tanımlamada insan istihdamı kullanmanın sakıncalı yönünü ortaya koymaktadır. Ayrıca, çevrimiçi kanallarda saldırgan dil söylemi hızlı yayılmaktadır. Bu durum zamanında yanıt veren çözümlerin gerekli olduğunu göstermektedir. Tüm bunlar, otomatik yöntemlerle rahatsız edici dil algılama sorununu çözen doğru ve ölçeklenebilir bir çözüm bulma ihtiyacını doğurmaktadır. Sonuç olarak, son yıllarda Doğal Dil İşleme topluluğu saldırgan dil ve nefret dili ile ilgili alanlar ile daha fazla ilgilenmeye başladı. Bu alanlarda yapılan çalışmalar özellikle sosyal medya da yer alan içeriklerdeki saldırgan dili tanımlamaya ve azaltmaya yardımcı olacağı düşünülmektedir. Ancak bu alanda Türkçe dilinde yapılan çalışmaların oldukça az olduğu görülmüştür. Böylelikle tez çalışmasında alana dikkat çekmek ve yapılacak olan çalışmalara katkı sağlamak amaçlanmıştır.

Çalışmada Twitter platformundan elde edilen metin verileri sınıflandırılmıştır. Sınıflandırma işlemi üç aşamalıdır. Birinci aşamada saldırgan ve saldırgan değil olarak ikili sınıflandırma yapılmıştır. İkinci aşamada saldırgan olan tweet metni hedefli ve hedefsiz olarak ayrılmıştır. Böylelikle bu aşamada hedefli, hedefsiz ve saldırgan değil olarak çoklu sınıflandırma yapılmıştır. Üçüncü aşamada ise, hedefli olan tweet metni birey, grup, diğer olarak ayrılmıştır. Bu aşamada ise saldırgan değil, hedefsiz, birey, grup ve diğer olarak çoklu sınıflandırma yapılmıştır. Ayrıca saldırgan dil tespitine yönelik bu alanda yapılan çalışmalara bakıldığında bu çalışma Türkçe dilinde çoklu sınıflandırma yapılan ilk çalışmadır. Buna ek olarak çalışmada Türkçe’de bilinen 150’den fazla saldırgan dil içeren kelime ve kelime öbekleri aratılarak veri seti ve genişletilmiş derlem oluşturulmuştur. Burada veri çeşitliliği sağlamak için geçmişe dayalı arama yapılmıştır. Veri setimizde yer alan 14752 tweet metni manuel olarak yukarıda belirtilen sınıflandırma kriterlerine göre manuel olarak etiketlenmiştir. Burada saldırgan dil etiketleme işleminin öznel yargılar barındırmasından kaynaklı etiketleyiciler arasında uyum problemi ile karşılaştık. Bu problemin çözümünde oy sistemi ile çoğunluğun kararına bakılmıştır. Ayrıca 1 milyon 860 bin alana özel tweetten oluşan genişletilmiş derlem elde edilmiştir. Böylelikle genişletilmiş derlemde word2vec-CBOW yöntemi kullanılarak elde edilen kelime temsillerinin sınıflandırma performansına katkısı görülmüştür.

Çalışmada üç aşamalı sınıflandırma yapılmıştır. Birinci aşamalı sınıflandırma olan saldırgan ve saldırgan değil olan ikili sınıflandırmadır. Burada kullanılan derin öğrenme yöntemleri; CNN, LSTM, BiLSTM, GRU, BiGRU ve CNN-LSTM’dir. Bu derin öğrenme yöntemleri literatürde en çok atıf alan popüler modellerdir. Böylelikle bu modellerin Türkçe dilinde performansları karşılaştırılmıştır. Burada veri setinden word2vec yöntemi kullanılarak elde edilen kelime temsilleri ile yapılan sınıflandırmada CNN ve CNN-LSTM modeli diğer modellerden F1 skor değeri yaklaşık %4-%5 daha iyi performans göstermiştir. Ancak genişletilmiş derlem kullanımıyla GRU saldırgan içeriği tespit etmesindeki F1 skor değeri %97 ile en yüksek performansa sahip model olmuştur. Burada bu modelin saldırgan olmayan içeriği tespit etmesindeki F1 skor değeri ise %93’tür. Buna ek olarak GRU ve LSTM arasında sınıflandırma performanslarına bakıldığında belirgin bir üstünlük görülmemiştir. Sonraki aşamalarda çoklu sınıflandırma yapılırken üstün sınıflandırma başarımı sebebiyle GRU modelinin performansına bakılmıştır. Buna ek olarak genişletilmiş derlem kullanımının en az etkilediği modeller F1-skor değeri %0,3 ve %0,6 ile sırasıyla CNN ve CNN-LSTM olmuştur. Böylelikle genişletilmiş derlem kullanımı CNN ve CNN-LSTM

modellerinde belirgin deęişikliğe sebep olmamıştır. Diğer modeller de genişletilmiş derlem kullanımının modellerin performansına katkısının yaklaşık olarak F1-skor deęeri %6 ile %7 arasındadır. Sonuç olarak genişletilmiş derlem ile çalışılması durumunda GRU modeli kullanılabileceęi gibi küçük boyutlu etiketli veri setinden elde edilen kelime temsilleri kullanılacağı zaman CNN ve CNN-LSTM modeli başarılı sonuçlar elde etmiştir. Bununla birlikte makine öğrenmesi yöntemi olan SVM modelinin sınıflandırma performansı da karşılaştırılmıştır. Buna göre saldırgan dili tespit etme başarımında derin öğrenme modellerinin SVM modeline göre belirgin bir üstünlük sağlayamadığı görülmüştür. Problemin doğası ve iki sınıflı verilerde kullanılan etiketleme kriterlerinin bu duruma sebep olduğu söylenebilir.

Çalışmada yapılan üç aşamalı sınıflandırmanın ikinci aşamasında hedefli, hedefsiz ve saldırgan deęil olarak çoklu sınıflandırma yapılmıştır. Burada üstün sınıflandırma başarımına sahip GRU modelinin sınıflandırma performansı karşılaştırılmıştır. Ayrıca çalışmada çoklu sınıflandırmanın yanında sadece hedefli ve hedefsiz verileri içeren ikili sınıflandırma da yapılmıştır. Böylelikle ikili ve çoklu sınıflandırma performans başarımları karşılaştırılmıştır. Ancak GRU modeli ikili sınıflandırma yapıldığında dengesiz veri sorunu ile karşılaşmıştır. Dolayısıyla model öğrenememiştir. GRU modelinin çoklu sınıflandırma F1 makro deęeri %71,97'dir. Genişletilmiş derlem kullanıldığında bu deęer %14 artış göstermiştir. Buna ek olarak daha az örneęe sahip olan hedefsiz verileri tahmin etme F1 skor deęeri %0 iken genişletilmiş derlem kullanımıyla bu deęer %30,30 olmuştur. Ayrıca hedefli ve saldırgan deęil olan verileri tahmin etme F1 skor deęeri sırasıyla %88,92 ve %79,6'tür. Genişletilmiş derlem kullanımıyla bu deęerler sırasıyla %94,69 ve %90,91 olmuştur. Sonuç olarak genişletilmiş derlem kullanımı modelin öğrenemeyen az örneęe sahip sınıfının sınıflandırma başarımını da artırmıştır. Ayrıca modelin performansını etkileyen bir başka faktöründe sınıf dengesi olduğu görülmüştür.

Çalışmada yapılan üç aşamalı sınıflandırmanın üçüncü aşamasında birey, grup, diğer, hedefsiz ve saldırgan deęil olarak çoklu sınıflandırma yapılmıştır. Burada üstün sınıflandırma başarımına sahip GRU modelinin sınıflandırma performansı karşılaştırılmıştır. GRU modelinin çoklu sınıflandırma F1 makro deęeri %39,49'dur. Genişletilmiş derlem kullanımıyla F1 makro deęerinde %14 oranında iyileşme sağlayarak %54,10 başarımla elde edilmiştir. GRU modelinin örnek sayısının azlığı sebebiyle diğer ve hedefsiz verileri tahmin etmedeki performansı oldukça düşüktür. Genişletilmiş derlemden elde edilen kelime

temsillerini kullanan modelin bu deęerleri iyileřtirdięi g r lm řt r. Burada dięer olarak etiketlenen verileri tahmin etme F1-skor deęeri %4,17 iken geniřletilmiř derlem ile %32,22'ye ulařmıřtır. Hedefsiz olarak etiketlenen verilerde ise F1-skor deęeri %0 iken %17,68'e ulařmıřtır. Geniřletilmiř derlem kullanımı dięer olarak etiketlenen verilerde F1-skor deęeri %28 iyileřme saęlarken birey olarak etiketlenen daha  ok  rnekleme sahip verilerde sadece F1-skor %2'lik iyileřme saęlayabilmiřtir. Dolayısıyla  alıřmada geniřletilmiř derlem kullanımının d ř k performans deęerine sahip  ęrenememiř aęlarda daha y ksek iyileřtirmeler saęladığı g r lm řt r.

Sonuç olarak alana  zg  verilerden oluřan geniřletilmiř derlem kullanımının sınıflandırma performansını  nemli  l  de artırdığı g r lm řt r. Bununla birlikte k   k bir veri setinden oluřan kelime temsilleri kullanıldığında CNN ve CNN-LSTM modellerinin dięer modellere oranla geniřletilmiř derleme yakın performans saęladığı g r lm řt r. Buna ek olarak  alıřmada kullanılan derin  ęrenme modellerinin literat re g re en pop ler modeller olduęu s ylenebilir. Bu modeller T rk e dilinde  st n sınıflandırma bařarım deęerleri elde etmiřtir. Ayrıca  alıřma da kullanılan veri setinin etiketlenmesi iřleminde uyumsuzluk problemi yařandığı g r lm řt r. Gelecek  alıřmalar da bu problemin   z m ne y nelik etiketleyiciler i in kapsamlı tanımlamalar yapılabilir.  te yandan bazı az  rneęe sahip sınıflarda dengesiz veri problemi yařanmıř, sonraki  alıřmalar da bu problemin   z m nde kullanılan veri artırma ve geri  eviri gibi y ntemlerin T rk e metinlerde etkisine y nelik  alıřmalar yapılabilir.

KAYNAKLAR

1. Adamec, Ć. and Viden, I. (1947). Polls come to Czechoslovakia. *Public Opinion Quarterly*, 11 (4), 548-552.
2. İnternet: Amidi A. and Amidi S. (2018). VIP Cheatsheet: Deep Learning. Deep Learning cheatsheets for Stanford's CS 230. URL:<https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-deep-learning-tips-and-tricks>, Son Eriřim Tarihi: 20.12.2020.
3. Ahmad M, Aftab S, Ali I. and Hameed N (2017).Hybrid tools and techniques for sentiment analysis: a review. *International Journal Multidiscipliner. Science Engineer*, 8 (4),7.
4. Ahmed, S., Jaidka, K. and Cho, J. (2016). The 2014 Indian elections on Twitter: A comparison of campaign strategies of political parties. *Telematics and Informatics*, 33 (4), 1071-1087.
5. Aldahawi, H. (2015). *Mining and analysing social network in the oil business: Twitter sentiment analysis and prediction approaches*, Doctoral dissertation, Cardiff University, Cardiff, UK.
6. Ali, W. N. H. W., Mohd, M. and Fauzi, F. (2021). Cyberbullying Predictiand Model: Implementation of Machine Learning Approach. In *2021 Fifth International Conference on Information Retrieval and Knowledge Management*, 65-69.
7. İnternet: American Psychological Association. (2004). APA resolution on bullying among children and youth. URL:<http://www.apa.org/about/goandrnance/council/policy/bullying.pdf>, Son Eriřim Tarihi: 20.07.2020
8. Antonakaki, D., Fragopoulou, P. and Loannidis, S. (2020). A surandy of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications*, 164.
9. İnternet: Aslam, S. (2020). Twitter by the numbers: Stats, demographics and fun facts URL:<https://www.omnicoreagency.com/twitter-statistics/>, Son Eriřim Tarihi: 20.07.2020
10. Baccianella, S., Esuli, A. and Sebastiani, F. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining, *In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 10, 2200-2204.
11. Bakshi, R. K., Kaur, N., Kaur, R. and Kaur, G. (2016). Opinion mining and sentiment analysis. In *2016 3rd International Conference on Computing for Sustainable Global Deandlopment (INDIACom)*, 452-455.

12. Basile, V., Bosco, C., Fersini, E., Debara, N., Patti, V., Pardo, F. M. R., ... and Sanguinetti, M. (2019). Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter. *In 13th International Workshop on Semantic Evaluation*, 54-63.
13. Basiri, M. E., Nemati, S., Abdar, M., Cambria, E. and Acharya, U. R. (2020). ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Future Generation Computer Systems*, 115, 279-294.
14. Behera, R. N., Roy, M. and Dash, S. (2016). Ensemble based hybrid machine learning approach for sentiment classification-a review. *International Journal of Computer Applications*, 146 (6), 31-36.
15. Bengio, Y., Simard, P. and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5 (2), 157-166.
16. Bojanowski, P., Graand, E., Joulin, A. and Mikolov, T. (2017). Enriching word andctors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135-146.
17. Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2 (1), 1-8.
18. Bourgonje, P., Moreno-Schneider, J., Srivastava, A. and Rehm, G. (2017, September). Automatic classification of abusiant language and personal attacks in various forms of online communication. *In International Conference of the German Society for Computational Linguistics and Language Technology*, 180-191.
19. Bowker, L. and Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. New York: Routledge,15-29.
20. Broniatowski, D. A., Paul, M. J. and Dredze, M. (2013). National and local influenza surandillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PloS one*, 8 (12).
21. Burnap, P. and Williams, M. L. (2014). Hate speech, machine classification and statistical modelling of information flows on Twitter: Interpretation and communication for policy decision making. *In Proceedings of Internet, Policy and Politics*, 1–18.
22. Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., ... and Voss, A. (2014). Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Social Network Analysis and Mining*, 4 (1), 1-14.
23. Cambria, E., Olsher D. andRajagopal D. (2014). SenticNet 3: a common and common-senseknowledge base for cognition-driandn sentiment analysis. *InProceedings of Twenty-eighthAAAI conference on Artificial Intelligence*, Quebec, Canada, 1515-1521.

24. Cambria, E., Poria, S., Gelbukh, A. and Thelwall, M. (2017). Sentiment Analysis Is a Big Suitcase. *IEEE Intelligent Systems*, 32 (6), 74–80.
25. Cambria, E., Schuller, B., Xia, Y. and Havasi, C. (2013). New advances in opinion mining and sentiment analysis. *IEEE Intelligent systems*, 28 (2), 15-21.
26. Cambria, E., Wang, H. and White, B. (2014). Guest editorial: Big social data analysis. *Knowledge-Based Systems*, 69 (1), 1-2.
27. Carley, K. M., Malik, M., Landwehr, P. M., Pfeffer, J. and Kowalchuck, M. (2016). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. *Safety science*, 90, 48-61.
28. Caselli, T., Basile, V., Mitrović, J., Kartoziya, I., & Granitzer, M. (2020, May). I feel offended, don't be abusive! implicit/explicit messages in offensive and abusive language. In *Proceedings of the 12th language resources and evaluation conference*, 6193-6202.
29. Chen, C., Zhuo, R. and Ren, J. (2019). Gated recurrent neural network with sentimental relations for sentiment classification. *Information Sciences*, 502, 268-278.
30. Chen, Y., Zhou, Y., Zhu, S. and Xu, H. (2012). Detecting offensive language in social media to protect adolescent online safety. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, 71–80.
31. Cheong, M. and Lee, V. C. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers*, 13 (1), 45-59.
32. Internet: Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv:1406.1078, URL:<https://arxiv.org/abs/1406.1078>, Son Erişim Tarihi: 20.11.2019
33. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of machine learning research*, 12, 2493-2537.
34. Craig, W., Boniel-Nissim, M., King, N., Walsh, S. D., Boer, M., Donnelly, P. D., ... and Van den Eijnden, R. (2020). Social media use and cyber-bullying: a cross-national analysis of young people in 42 countries. *Journal of Adolescent Health*, 66 (6), 100-108.
35. Crannell, W. C., Clark, E., Jones, C., James, T. A. and Moore, J. (2016). A pattern-matched Twitter analysis of US cancer-patient sentiments. *Journal of surgical research*, 206 (2), 536-542.

36. Çöltekin, Ç. (2020). A corpus of Turkish offensive language on social media. In *Proceedings of the 12th Language Resources and Evaluation Conference*, 6174-6184.
37. Dadvar, M., Trieschnigg, D. and de Jong, F. (2014). Experts and machines against bullies: A hybrid approach to detect cyberbullies. In *Canadian Conference on Artificial Intelligence*, 275–281.
38. Dadvar, M., Trieschnigg, D., Ordelman, R. and de Jong, F. (2013). Improving cyberbullying detection with user context. In *European Conference on Information Retrieval*, 693–696.
39. Danneman, N. and Heimann, R. (2014). *Social media mining with R*. Birmingham: Packt Publishing Ltd, 1-20.
40. Daand, K., Lawrence, S. and Pennock, D. M. (2003,). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, 519-528.
41. İnternet: Davidson, T., Warmesley, D., Macy, M. and Weber, I. (2017). Automated hate speech detection and the problem of offensive language, arXiv:1703.04009, URL:<https://arxiv.org/abs/1703.04009>, Son Erişim Tarihi: 20.11.2019
42. Dehkharghani, R., Yanikoglu, B., Tapucu, D. and Saygin, Y. (2012) Adaptation and use of sub-jectivity lexicons for domain dependent sentiment classification. In *Proceedings of DataMining Workshops (ICDMW) 2012 IEEE 12th International Conference on, Brussels Belgium*, 669–73.
43. Dehkharghani, R., Saygin, Y., Yanikoglu, B. and Oflazer, K. (2016) SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language resources and evaluation*, 50 (3), 667-85.
44. Dehkharghani, R., Yanikoglu, B., Saygin, Y. and Oflazer, K. (2017). Sentiment analysis in Turkish at different granularity levels. *Natural Language Engineering*, 23(4), 535-559.
45. Del Vigna¹², F., Cimino²³, A., Dell’Orletta, F., Petrocchi, M. and Tesconi, M. (2017). Hate me, hate me not: Hate speech detection on facebook. In *Proceedings of the First Italian Conference on Cybersecurity (ITASEC17)*, 86-95.
46. Demircan, Ö. (1994). İletişim açısından hakaret and küfür. *Dilbilim Araştırmaları Dergisi*, 5, 264-273.
47. İnternet: Devlin, J., Chang, M. W., Lee, K. and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv:1810.04805, URL:<https://arxiv.org/abs/1810.04805>, Son Erişim Tarihi: 20.11.2019
48. Dinakar, K., Jones, B., Havasi, C., Lieberman, H. and Picard, R. (2012). Common sense reasoning for detection, prevention and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2 (3),18.

49. Elkin L. S., Topal K., and Bebek G. (2017). Network based model of social media big data predicts contagious disease diffusion. *Information Discovery and Delivery*, 45 (3) : 110-120.
50. Eryiğit, G., Cetin, F. S., Yanık, M., Temel, T. and Çiçekli, I. (2013). Turksent: A sentiment annotation tool for social media. *In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, 131-134.
51. Esuli, A., Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. *In Proceedings of LREC*, Genoa, Italy, (6), 417–22.
52. Fegiz, P. L. (1947). Italian public opinion. *Public Opinion Quarterly*, 11(1), 92-96.
53. Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56 (4), 82-89.
54. Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E. and González-Castaño, F. J. (2016). Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications*, (58), 57-75.
55. Filatova, E. (2012). Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. *In Lrec*, 392-398.
56. Friginal, E., Peter, D. Y. E. and Nolen, M. (2020). Corpus-Based Approaches in Language Teaching: Outcomes, Observations and Teacher Perspectives. *Boğaziçi Üniversitesi Eğitim Dergisi*, 37 (1), 43-68.
57. González, J. Á., Hurtado, L. F. and Pla, F. (2020). Transformer based contextualization of pre-trained word embeddings for irony detection in Twitter. *Information Processing and Management*, 57 (4).
58. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. Cambridge: MIT press, 12-27.
59. İnternet: Google Trends (2020). URL:<https://trends.google.com/trends/explore?date=all&q=%2Fm%2F0g57xn>, Son Erişim Tarihi: 20.11.2020
60. Grant, A. and O'Mahoney, H. (2016). Portrayal of waterpipe (shisha, hookah, nargile) smoking on Twitter: a qualitative exploration. *Public Health*, 140, 128-135.
61. Greenwell, J. and Dengerink, H. A. (1973). The role of perceived and actual attack in human physical aggression. *Journal of Personality and Social Psychology*, 26(1), 66.
62. Hassan, A., Âmin, M. R., Al Azad, A. K., & Mohammed, N. (2016, December). Sentiment analysis on bangla and romanized bangla text using deep recurrent models. *In 2016 International Workshop on Computational Intelligence (IWCI)*, 51-56.

63. Hatzivassiloglou, V. and McKeown, K. (1997). Predicting the semantic orientation of adjectiands. In *35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, 174-181.
64. Hinduja, S. and Patchin, J. W. (2010). Bullying, cyberbullying and suicide. *Archands of suicide research*, 14 (3), 206-221.
65. Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9 (8), 1735-1780.
66. Hogenboom, A., Heerschop, B., Frasincar, F., Kaymak, U. and de Jong, F. (2014). Multi-lingual support for lexicon-based sentiment analysis guided by semantics. *Decision support systems*, 62, 43-53.
67. İbrohim, M. O. and Budi, I. (2018). A dataset and preliminaries study for abusiand language detection in Indonesian social media. *Procedia Computer Science*, 135, 222-229.
68. Interian, Y., Rideout, V., Kearney, V. P., Gennatas, E., Morin, O., Cheung, J., ... and Valdes, G. (2018). Deep nets vs expert designed features in medical physics: an IMRT QA case study. *Medical physics*, 45 (6), 2672-2680.
69. İskender, E. and Batı, G. B. (2015). Comparing Turkish Uniandrsities Entrepreneurship and Innovatiandness Index's Rankings with Sentiment Analysis Results on Social Media. *Procedia-Social and Behavioral Sciences*, 195, 1543-1552.
70. İnternet: Jaki, S. and De Smedt, T. (2019). Right-wing German hate speech on Twitter: Analysis and automatic detection, arXiv:1910.07518, URL:<https://arxiv.org/abs/1910.07518>, Son Erişim Tarihi: 20.11.2019
71. İnternet: Joulin, A., Graand, E., Bojanowski, P., Douze, M., Jégou, H. and Mikolov, T. (2016). Fasttext. Zip: Compressing text classification models, arXiv:1612.03651, URL:<https://arxiv.org/abs/1612.03651>, Son Erişim Tarihi: 20.11.2019
72. İnternet: Kalchbrenner, N., Grefenstette, E. and Blunsom, P. (2014). A convolutional neural network for modelling sentences, arXiv:1404.2188, URL:<https://arxiv.org/abs/1404.2188>, Son Erişim Tarihi: 20.11.2019
73. Kayany, J. M. (1998). Contexts of uninhibited online behavior: Flaming in social newsgroups on usenet. *Journal of the American Society for Information Science*, 49 (12),1135–1141.
74. Kayser, V. and Bierwisch, A. (2016). Using Twitter for foresight: An opportunity?. *Futures*, 84, 50-63.
75. Khan, F. H., Qamar, U. and Bashir, S. (2016). eSAP: A decision support framework for enhanced sentiment analysis and polarity classification. *Information Sciences*, 367, 862-873.

76. İnternet: Kharde, V. and Sonawane, P. (2016). Sentiment analysis of twitter data: a surandy of techniques, arXiv:1601.06971, URL:<https://arxiv.org/abs/1601.06971>, Son Erişim Tarihi: 20.11.2019.
77. Kim, J. and Hastak, M. (2018). Social network analysis: Characteristics of online social networks after a disaster. *International Journal of Information Management*, 38 (1), 86-96.
78. Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751, Doha, Qatar. Association for Computational Linguistics.
79. Knutson, A. L. (1945). Japanese opinion surandys: the special need and the special difficulties. *Public Opinion Quarterly*, 9 (3), 313-319.
80. Köksal Ö. (2021). "Enhancing Turkish Sentiment Analysis Using Pre-Trained Language Models," *29th Signal Processing and Communications Applications Conference (SIU)*, 1-4.
81. Kroeze, J. H., Matthee, M. C. and Bothma, T. J. (2004). Differentiating between data-mining and text-mining terminology. *SA Journal of Information Management*, 6 (4).
82. Kumar, R., Ojha, A. K., Malmasi, S. and Zampieri, M. (2018). Benchmarking aggression identification in social media. *In Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018)*, 1-11.
83. Kwok, I., & Wang, Y. (2013, Temmuz). Locate the hate: detecting tweets against blacks, *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 1621-1622.
84. Lampos, V., De Bie, T. and Cristianini, N. (2010). Flu detector-tracking epidemics on Twitter. *In Joint European conference on machine learning and knowledge discoandry in databases*, 599-602.
85. Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. *In International conference on machine learning*, 1188-1196.
86. Lea, M., O'Shea, T., Fung, P. and Spears, R., (1992). *Flaming'in computer-mediated communication: Observations, explanations, implications.*, Harvester Wheatsheaf, 89–112.
87. Lee, N. Y., Kim, Y. and Sang, Y. (2017). How do journalists leandrage Twitter? Expressiand and consumptiand use of Twitter. *The Social Science Journal*, 54 (2), 139-147.
88. Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D. and Chen, M. (2014). Medical image classification with convolutional neural network. *In 2014 13th International Conference on Control Automation Robotics and Vision (ICARCV)*, 844-848.

89. Liao, C., Feng, C., Yang, S. and Huang, H. (2016). Topic-related Chinese message sentiment analysis. *Neurocomputing*, 210, 237-246.
90. Ligthart, A., Catal, C. and Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artifitical Intelligence Revolution*, 54, 4997–5053.
91. Liu, B. (2009). *Handbook chapter: sentiment analysis and subjectivity*. Handbook of natural language processing. New York: Marcel Dekker, 467-562.
92. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5 (1), 1-167.
93. Liu, B. (2020). Sentiment analysis: Mining opinions, sentiments and emotions. *Computational Linguistics*, 42(3), 595-598.
94. Liu, B. and Zhang, L. (2012). A surandy of opinion mining and sentiment analysis. In *Mining text data*, 415-463.
95. İnternet: Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach, arXiv:1907.11692, URL:<https://arxiv.org/abs/1907.11692>, Son Erişim Tarihi: 20.11.2019
96. Majumder, N., Poria, S., Gelbukh, A. and Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32 (2), 74-79.
97. İnternet: Malmasi, S. and Zampieri, M. (2017). Detecting hate speech in social media, arXiv:1712.06427, URL:<https://arxiv.org/abs/1712.06427>, Son Erişim Tarihi: 20.11.2019
98. Manda, K. R. (2019). *Sentiment Analysis of Twitter Data Using Machine Learning and Deep Learning Methods*. Master Thesis, Blekinge Tekniska H ochskola University Blekinge Institute of Technology, Sweden, 16-20.
99. P amies Massip, M. (2020). *Multilingual identification of offensive content in social media*. Master Thesis, Department of Computer and Information Science, Sweden, 28-42.
100. M antyl , M. V., Graziotin, D. and Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics andnues and top cited papers. *Computer Science Review*, 27, 16-32.
101. Mehl, M. R. and Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conandrsations. *Journal of personality and social psychology*, 84 (4),857
102. Mehta, Y., Majumder, N., Gelbukh, A. and Cambria, E. (2019). Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, 1-27.

103. Mejova, Y. and Srinivasan, P. (2011). Exploring feature definition and selection for sentiment classifiers. In *Fifth international AAAI conference on weblogs and social media*, , 5(1), 546-549.
104. Mikolov T., Grave T, Bojanowski T, Puhersch C and Joulin A. (2018). Advances in Pre-Training Distributed Word Representations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
105. İnternet: S. Minaee, E. Azimi and A. Abdolrashidi (2019). Deep-sentiment: Sentiment analysis using ensemble of CNN and Bi-LSTM models, arXiv:1904.04206, URL:<https://arxiv.org/abs/1904.04206>, Son Erişim Tarihi: 20.11.2019
106. Mite-Baidal, K., Delgado-Andra, C., Solís-Avilés, E., Espinoza, A. H., Ortiz-Zambrano, J. and Varela-Tapia, E. (2018). Sentiment analysis in education domain: A systematic literature review. In *International Conference on Technologies and Innovation*, 285-297.
107. Mubarak, H., Darwish, K. and Magdy, W. (2017). Abusiand language detection on Arabic social media. In *Proceedings of the first workshop on abusiand language online*, 52-56.
108. Muhammad, A., Wiratunga, N. and Lothian, R. (2016). Contextual sentiment analysis for social media genres. *Knowledge-based systems*, 108, 92-101.
109. Munezero, M., Montero, C. S., Sutinen, E. and Pajunen, J. (2014). Are they different? Affect, feeling, emotion, sentiment and opinion detection in text. *IEEE transactions on affectiand computing*, 5 (2), 101-111.
110. Nassif, A. B., Elnagar, A., Shahin, I. and Henno, S. (2020). Deep learning for Arabic subjectiand sentiment analysis: Challenges and research opportunities. *Applied Soft Computing*, 98.
111. Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y. and Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, 41 (16), 7653-7670.
112. Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A. and Stehle, S. (2017). Sentiment analysis during Hurricane Sandy in emergency response. *International journal of disaster risk reduction*, 21, 213-222.
113. İnternet: Newberry, C. (2021). 36 Twitter Stats All Marketers Need to Know in 2021. URL:<https://blog.hootsuite.com/twitter-statistics/>, Son Erişim Tarihi: 07.12.2021.
114. Nitta, T., Masui, F., Ptaszynski, M., Kimura, Y., Rzepka, R., & Araki, K. (2013, October). Detecting cyberbullying entries on informal school websites based on category relevance maximization. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 579-586.

115. Oza, K. S. and Naik, P. G. (2016). Prediction of online lectures popularity: a text mining approach. *Procedia Computer Science*, 92, 468-474.
116. Ozdemir, A. and Yeniterzi, R. (2020). SU-NLP at SemEval-2020 Task 12: Offensive language identification in Turkish tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 2171-2176.
117. Özer Z. (2019). *Trafik yorumlarının sınıflandırılmasında normalizasyonun etkisi*. Doktora Tezi, Karabük Üniversitesi Fen Bilimleri Enstitüsü, Karabük, 15-23.
118. Öztürk, N. and Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, 35 (1), 136-147.
119. Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, 10, 1320-1326.
120. İnternet: Pandey, P. (2018). Simplifying sentiment analysis using VADER in Python (on social media text). Retrieved from Analytics Vidhya. URL:<https://medium.com/analytics-vidhya/simplifying-socialmedia-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f>, Son Erişim Tarihi: 20.11.2019
121. İnternet: Pang, B., Lee, L. and Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. ArXiv 0205070, URL:<https://arxiv.org/abs/0205070>, Son Erişim Tarihi: 20.11.2019
122. Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Computer Linguistic*, 35(2), 311-312.
123. Park, S. J., Park, J. Y., Lim, Y. S. and Park, H. W. (2016). Expanding the presidential debate by tweeting: The 2012 presidential election debate in South Korea. *Telematics and informatics*, 33 (2), 557-569.
124. Piryani, R., Madhavi, D. and Singh, V. K. (2017). Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Information Processing and Management*, 53 (1), 122-150.
125. İnternet: Reinsel, D., J. Gantz and J. Rydning (2018). Data Age 2025: The Evolution of Data to Life-Critical, IDC White Paper, URL:<https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WPDataAge2025-March-2017>, Son Erişim Tarihi: 12.08. 2020
126. Reyes, A. and Rosso, P. (2014). On the difficulty of automatically detecting irony: beyond a simple case of negation. *Knowledge and Information Systems*, 40(3), 595-614.
127. Rojas-Barahona, L. M. (2016). Deep learning for sentiment analysis. *Language and Linguistics Compass*, 10 (12), 701-719.
128. Ross Stagner (1940) The Cross-Out Technique as a Method in Public Opinion Analysis, *The Journal of Social Psychology*, 11(1), 79-90

129. Ross, M. Rist, G. Carbonell, B. Cabrera, N. Kurowsky and M. Wojatzki. 2016. Measuring the reliability of hate speech annotations: The case of the european refugee crisis. In *Proceedings of NLP4CMC III: 3rd Workshop on Natural Language Processing for Computer-Mediated Communication*, 17, 6–9.
130. Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536.
131. İnternet: Ruppenhofer, J., Siegel, M. and Wiegand, M. (2018). Guidelines for IGGSA shared task on the identification of offensiv language. URL:<https://github.com/uds-lsv/GermEval-2018-Data/blob/master/guidelines-iggsa-shared.pdf>, Son Erişim Tarihi: 20.11.2019
132. İnternet: Sagolla, D. (2009). How Twitter Was Born. Retrieved from. URL:<http://www.140characters.com/2009/01/30/how-twitter-was-born/>, Son Erişim Tarihi: 27.10.2020
133. Sandri, S. A., Dubois, D. and Kalfsbeek, H. W. (1995). Elicitation, assessment and pooling of expert judgments using possibility theory. *IEEE transactions on fuzzy systems*, 3 (3), 313-335.
134. Sanguinetti, M., Poletto, F., Bosco, C., Patti, V. and Stranisci, M. (2018). An italian twitter corpus of hate speech against immigrants. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2798–2895
135. Sarkar, D., Bali, R. and Sharma, T. (2018). Practical machine learning with Python. *A problem-solving guide to building real-world intelligent systems*. Apress, Berkely.
136. Sasaki, Y., Kawai, D. and Kitamura, S. (2016). Unfriend or ignore tweets? A time series analysis on Japanese Twitter users suffering from information overload. *Computers in Human Behavior*, 64, 914-922.
137. Shahin, I., Nassif, A. B. and Hamsa, S. (2019). Emotion recognition using hybrid Gaussian mixture model and deep neural network, *IEEE access*, 7, 26777-26787.
138. Shahin, I., Nassif, A. B. and Hamsa, S. (2020). A cascaded Gaussian mixture model-deep neural network classifier for speaker identification in emotional talking environments. *Neural Computing and Applications*, 32 (7), 2575-2587.
139. Sharma, A. K., Chaurasia, S. and Srivastava, D. K. (2020). Sentimental Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec. *Procedia Computer Science*, 167, 1139-1147.
140. Sharma, M., Kandasamy, I. and Kandasamy, V. (2021). Deep Learning for predicting neutralities in Offensiv Language Identification Dataset. *Expert Systems with Applications*, 185.
141. Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404.

142. Song, M., Park, H. and Shin, K. S. (2019). Attention-based long short-term memory network using sentiment lexicon embedding for aspect-leandl sentiment analysis in Korean. *Information Processing and Management*, 56 (3), 637-653.
143. Spertus, E. (1997). Smokey: Automatic recognition of hostile messages. In Proceedings of the Fourteenth National conference on Arirtificial Intelligence and Innovatiand Applications of Artificial Intelligence, 1058–1065.
144. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskeandr, I. and Salakhutdinov, R. (2014). Dropout: a simple way to preandnt neural networks from oandrfitting. *The journal of Machine learning research*, 15 (1), 1929-1958.
145. Stagner, R. (1940). The cross-out technique as a method in public opinion analysis. *The Journal of Social Psychology*, 11 (1), 79-90.
146. Struß, J. M., Siegel, M., Ruppenhofer, J., Wiegand, M. and Klenner, M. (2019). Oandrview of GermEval task 2, 2019 shared task on the identification of offensiand language. In Preliminary proceedings of the 15th Conference on Natural Language Processing (KONANDNS 2019), 352– 363.
147. Sun, S., Luo, C. and Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information fusion*, 36, 10-25.
148. Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37 (2), 267-307.
149. Tang, D., Qin, B. and Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, 1422-1432.
150. Tellez, E. S., Miranda-Jiménez, S., Graff, M., Moctezuma, D., Siordia, O. S. and Villaseñor, E. A. (2017). A case study of Spanish text transformations for twitter sentiment analysis. *Expert Systems with Applications*, 81, 457-471.
151. Thelwall, M. (2009). MySpace comments. *Online Information Review*, 33(1), 58-76
152. Thelwall, M., Buckley, K. and Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63 (1), 163-173.
153. Thompson, J. J., Leung, B. H., Blair, M. R. and Taboada, M. (2017). Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model. *Knowledge-Based Systems*, 137, 149-162.
154. Tumasjan, A., Sprenger, T. O., Sandner, P. G. and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reandal about political sentiment. In *Fourth international AAAI conference on weblogs and social media*, 4(1), 178-185.

155. P. Turney (2002). "Thumbs Up. Thumbs Down. Semantic Orientation Applied To Unsupervised Classification of Reviews, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 417-424.
156. Turney, P. D. and Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21 (4), 315-346.
157. Van Hee, C., Lefeandr, E. andrhoeandn, B., Mennes, J., Desmet, B., De Pauw, G., Daelemans, W. and Hoste, V. (2015). Detection and fine-grained classification of cyberbullying eandnts. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, Shoumen, Bulgaria, 672–680.
158. Van Klompenburg, T., Kassahun, A. and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709-105727.
159. Vural, A. G., Cambazoglu, B. B., Senkul, P. and 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*, 437-445.
160. Wang, H., Can, D., Kazemzadeh, A., Bar, F. and Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 system demonstrations*, 115-120.
161. Wang, W., Chen, L., Thirunarayan, K. and Sheth, A. P. (2014). Cursing in english on twitter. In *Proceedings of the 17th ACM conference on Computer supported cooperatiand work and social computing*, 415-425.
162. Wang, Y. H., Lee, H. Y. and Lee, L. S. (2018). Segmental audio word2andc: Representing utterances as sequences of andctors with applications in spoken term detection. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6269-6273.
163. Waseem, Z., Davidson, T., Warmesley, D. and Weber, I. (2017). Understanding abuse: A typology of abusian language detection subtasks. In *Proceedings of the First Workshop on Abusiand Language Online*, 78–84.
164. İnternet: We are Social and Hootsuite (2020). *Dijital in 2020*. URL: <https://wearesocial.com/digital-2020>, Son Erişim Tarihi: 08.06.2020
165. Wiebe, J. M. (1991). Recognizing subjectiand sentences: A computational inandstigation of narratiand text. *References in Narrative Text Noûs*, 25(4), 457-486.
166. Wiebe, J., Bruce, R. and O'Hara, T. P. (1999). Deandlopment and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*, 246-253.

167. Wiegand, M., Siegel, M. and Ruppenhofer, J. (2018). Overview of the GermEval 2018 shared task on the identification of offensive language. In *Proceedings of the GermEval 2018 Workshop at KONANLP 2018*, 1– 10.
168. Wu, B. and Shen, H. (2015). Analyzing and predicting news popularity on Twitter. *International Journal of Information Management*, 35 (6), 702-711.
169. Xiang, G., Fan, B., Wang, L., Hong, J. and Rose, C. (2012). Detecting offensive tweets via topical feature discovery and a large scale twitter corpus. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, 1980-1984.
170. Xu, J.-M., Jun, K.-S., Zhu, X. and Bellmore, A. (2012). Learning from bullying traces in social media. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 656–666.
171. Yadav, A. and Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53 (6), 4335-4385.
172. Yang, Y. and Eisenstein, J. (2017). Overcoming language variation in sentiment analysis with social attention. *Transactions of the Association for Computational Linguistics*, (5), 295-307.
173. Yates, D. and Paquette, S. (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International journal of information management*, 31 (1), 6-13.
174. Yuan, J., Wu, Y., Lu, X., Zhao, Y., Qin, B. and Liu, T. (2020). Recent advances in deep learning based sentiment analysis. *Science China Technological Sciences*, 1-24.
175. Zampieri M., Malmasi S., Nakov P., Rosenthal S., Farra N and Kumar R. (2019). SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffenseEval). In *Proceedings of the 13th International Workshop on Semantic Evaluation*, 75–86.
176. Zampieri M., Malmasi S., Nakov P., Rosenthal S., Atanasova P., Karadzhov G., Mubarak H., Derczynski L., Pitenis Z. and Çöltekin Ç. (2020). SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffenseEval 2020). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 1425–1447.
177. Zhang, L., Wang, S. and Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8 (4), 1-25.
178. Internet: Qiu, H., Zeng, Y., Zhang, T., Jiang, Y., & Qiu, M. (2020). FenceBox: A Platform for Defeating Adversarial Examples with Data Augmentation Techniques, ArXiv: 2012.01701, URL:<https://arxiv.org/abs/2012.01701>, Son Erişim Tarihi: 20.11.2020

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : YILMAZ ŞAHİNER Şeyma
Uyruğu : T.C.
E-mail :

Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üni.-Yönetim Bilişim Sistemleri	Devam Ediyor
Lisans	Bartın Üni.-Yönetim Bilişim Sistemleri	2015
Lise	Bahçelievler Anadolu Lisesi	2011

İş Deneyimi

Yıl	Yer	Görev
2019	Bandırma Onyediy Eylül Üniversitesi	Araştırma Görevlisi

Yabancı Dil

İngilizce: Orta Seviye

Tezden Üretilen Yayınlar

1. Yılmaz, Ş. Ş., Özer, İ. ve Gökçen, H. (2021). Türkçe Metinlerde Derin Öğrenme Yöntemleri Kullanılarak Duygu Analizi. In *International Symposium of Scientific Research and Innovative Studies*, 22, 971-982



GAZİ OLMAK AYRICALIKTIR...