



T.C.

BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI

Yüksek Lisans Tezi

**SOSYAL MEDYA MESAJLARI ÜZERİNDE
TELEKOMÜNİKASYON
SEKTÖRÜNE AİT MÜŞTERİ ŞİKAYETLERİNİN
SINIFLANDIRILMASI**

**Sevin KARAGÖZ
195052016**

Tez Danışmanı:

Dr. Öğr. Üyesi Zeynep ÖZER

Bandırma 2023

T.C.
BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI

Yüksek Lisans Tezi

SOSYAL MEDYA MESAJLARI ÜZERİNDE
TELEKOMÜNİKASYON
SEKTÖRÜNE AİT MÜŞTERİ ŞİKAYETLERİNİN
SINIFLANDIRILMASI

Sevin KARAGÖZ
195052016

Tez Danışmanı:
Dr. Öğr. Üyesi Zeynep ÖZER

Bandırma 2023

ETİK BEYAN SAYFASI

TÜRKİYE CUMHURİYETİ

BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Bu belge ile, bu tezdeki bütün bilgilerin akademik kurallara ve etik ilkelere uygun olarak toplanıp sunulduğunu beyan ederim. Bu kural ve ilkelerin gereği olarak, çalışmada bana ait olmayan tüm veri, düşünce ve sonuçları andığımı ve kaynağını gösterdiğimi ayrıca beyan ederim. (14/07/2023)

ÖZET

SOSYAL MEDYA MESAJLARI ÜZERİNDE TELEKOMÜNİKASYON SEKTÖRÜNE AİT MÜŞTERİ ŞİKAYETLERİNİN SINIFLANDIRILMASI

Sevin KARAGÖZ

Sosyal medya duyguların, düşüncelerin ve görüşlerin sıklıkla ifade edildiği bir iletişim ve yayın platformu haline gelmiştir. Sosyal medya aracılığıyla duyguları ifade edilmesi, hayatımızın önemli bir parçası olarak büyük bir dikkat gerektirmektedir. Fikir ve düşüncelerin paylaşılmasının yanı sıra, güzel anıları da paylaşmaktayız. Facebook, Twitter ve LinkedIn gibi sosyal medya, zengin fikirli metin verileri kaynakları olarak kabul edilmektedir. Hem kuruluşlar hem de bireyler, insanların fikirlerini analiz etmek ve duygu ve hisleri çıkarmak için sosyal medya platformlarından faydalanmaktadır. Sosyal medya analizi üzerine yapılan çoğu çalışma, duyguları olumlu, olumsuz veya tarafsız sınıflar olarak sınıflandırmaktadır. Duygu analizindeki zorluk, insanların bir veya birkaç duyguyu tek bir ifadeyle ifade edebilmesinden kaynaklanmaktadır. İnsan bu farklı duyguları iyi tanıyabilir; ancak yine de bir duygu analiz sistemi için bu kolay değildir.

Telekomünikasyon, iletişim teknolojilerinin kullanılarak bilgi, ses, görüntü ve veri gibi iletişim sinyallerinin iletimini sağlayan bir alanı ifade eder. Bu teknolojiler, sesli telefonlar, mobil telefonlar, internet, radyo, televizyon, uydu iletişimi ve diğer iletişim sistemleri gibi çeşitli iletişim araçlarını kapsar.

Telekomünikasyon, insanlar arasındaki uzaklığı aşarak bilgi paylaşımını kolaylaştırır. İletişim sinyalleri, ses, görüntü ve veri gibi farklı formatta iletim yapabilir. Örneğin, sesli telefonlar insanların birbirleriyle konuşmasını sağlarken, internet aracılığıyla metin, fotoğraf, video gibi veriler paylaşılabilir. Telekomünikasyon ayrıca ticari ve endüstriyel alanlarda da büyük öneme sahiptir. İşletmeler, telekomünikasyon altyapısı üzerinden müşterileriyle iletişim kurar, veri

transferi yapar ve iş süreçlerini yönetir. Telekomünikasyon ayrıca sağlık, ulaşım, medya, eğitim, finans gibi birçok sektörde de kullanılır.

Gelişen teknolojilerle birlikte telekomünikasyon alanında da sürekli yenilikler yaşanmaktadır. Örneğin, kablosuz iletişim teknolojileri (Wi-Fi, Bluetooth), fiber optik altyapı, 5G ağları gibi yeni teknolojiler telekomünikasyonun daha hızlı, güvenilir ve geniş kapsamlı hale gelmesini sağlamıştır. Bu gelişim ile tüketicinin artması ve telekomünikasyon sektörleri arasında en iyi hizmeti sunma durumu değerlendirilmektedir. Tüketici gereken hizmeti alamadığında, sorunlar yaşadığında bunu sosyal medya platformları üzerinden dile getirerek sosyal medya platformları üzerinden bir veri derlemi oluşturmaktadır.

Bu çalışmada, veri seti Twitter'dan toplanan telekomünikasyon sektörüne ait şikâyet derlemine içermektedir. Elde edilen üç milyon tweet içerisinde beş bin tweetin etiketlenmesi işlemi olmuştur. Etiketlenmesi sonucu oluşturulan verilerin ön işleme ve temizlenme süreci tamamlandıktan sonra GRU, LSTM, N-Gram GRU-LSTM modelleri ile sınıflandırma işlemi yapılmıştır. Sınıflandırma sonucunda elde edilen bulgular tablo ve şekiller ile tartışılıp değerlendirilmiştir. Ortaya çıkan sonuçlar üzerinden genel değerlendirme yapıldığında, hata değerleri bakımından LSTM ve N-gram LSTM modellerinin GRU tabanlı modellere kıyasla daha iyi performans gösterdiği tespit edilmiştir. Bunun durumun, LSTM tabanlı ağların uzun süreli bağımlılıkları yakalama noktasındaki başarısından kaynaklandığı değerlendirilmektedir.

Anahtar Sözcükler: LSTM, GRU, Twitter, Telekomünikasyon, Sınıflandırma

ABSTRACT

CLASSIFICATION of CUSTOMER COMPLAINTS of THE TELECOMMUNICATIONS INDUSTRY on SOCIAL MEDIA MESSAGES

Sevin KARAGÖZ

Social media has become a communication and broadcasting platform where emotions, thoughts and opinions are frequently expressed. Expressing emotions through social media requires great care as an important part of our lives. In addition to sharing ideas and thoughts, we also share good memories. Social media such as Facebook, Twitter and LinkedIn are recognized as rich sources of opinion-rich text data. Both organizations and individuals use social media platforms to analyze people's opinions and extract emotions and feelings. Most studies on social media analytics classify sentiments into positive, negative or neutral classes. The difficulty in sentiment analysis lies in the fact that people can express one or several emotions in a single phrase. One can recognize these different emotions well, but it is still not easy for a sentiment analysis system.

Telecommunications refers to a field that uses communication technologies to transmit communication signals such as information, voice, images and data. These technologies cover a variety of communication tools such as voice telephones, mobile phones, the internet, radio, television, satellite communications and other communication systems.

Telecommunications facilitates the sharing of information across distances between people. Communication signals can transmit in different formats, such as voice, images and data. For example, voice telephones allow people to talk to each other, while data such as text, photos and videos can be shared via the internet. Telecommunications are also of great importance in commercial and industrial areas. Businesses communicate with their customers over the telecommunications infrastructure, share data transfer and manage business processes. Telecommunications are also used in many sectors such as health, transportation, media, education and finance.

With the developing technologies, there are continuous innovations in the field of telecommunications. For example, new technologies such as wireless communication technologies (Wi-Fi, Bluetooth), fiber optic infrastructure, 5G networks have made telecommunications faster, more reliable and more comprehensive. With this development, consumption increases and the situation of providing the best service among telecommunication sectors is evaluated. When consumers do not receive the required service and experience problems, they express this on social media platforms and create a data corpus on social media platforms.

In this study, the dataset includes a corpus of complaints about the telecommunications sector collected from Twitter. Out of three million tweets, five thousand tweets were labeled. After the preprocessing and cleaning process of the labeled data was completed, classification was performed with GRU, LSTM, N-Gram GRU-LSTM models. The findings obtained as a result of classification were discussed and evaluated with tables and figures. In terms of error values, it was found that LSTM and N-gram LSTM models performed better than GRU-based models. This is attributed to the success of LSTM-based networks in capturing long-term dependencies.

Keywords: LSTM, GRU, Twitter, Telecommunications, Classification

ÖNSÖZ

Bu tez çalışmasının planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteğini esirgemeyen, engin bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren Sayın hocam Dr. Öğr. Üyesi Zeynep ÖZER'e teşekkür ederim.

Engin bilgi, tecrübe, uzman görüşleriyle ve değerli katkılarıyla çalışmama destek veren, eğitim hayatımda bilgi ve desteklerini esirgemeyen çok kıymetli Sayın hocalarım Prof. Dr. Mehmet TEKTAŞ, Doç. Dr. Necla TEKTAŞ ve Doç. Dr. İlyas ÖZER'e sonsuz teşekkür ve saygılarımı sunarım.

Her zaman desteklerini, maddi manevi hiçbir yardımı esirgemedi yanımda olan ve bu konuma gelmemde destek olan sevgili babam Kadir KARAGÖZ'e, sevgili annem Fehime KARAGÖZ'e ve sevgili kardeşlerime sonsuz teşekkürlerimi sunarım.

Sevin KARAGÖZ

Bandırma, 14.07.2023

İÇİNDEKİLER

ETİK BEYAN SAYFASI.....	ii
ÖZET.....	iii
ABSTRACT	v
ÖNSÖZ.....	vii
İÇİNDEKİLER	viii
TABLolar	x
ŞEKİLLER	xii
KISALTMALAR	xiv
GİRİŞ	1
BİRİNCİ BÖLÜM.....	3
1. GENEL TANIMLAR.....	3
1.1. Telekomünikasyon Sektörü	3
1.1.1. Telekomünikasyon Sektörünün Temel Hizmetleri ve Durum Analizi	4
1.1.2. Telekomünikasyon Sektörünün Mevcut Sorunları	6
1.1.3. Dünya’ da Telekomünikasyon.....	8
1.1.4. Türkiye’de Telekomünikasyon.....	13
1.2. Sosyal Medya Kavramı.....	17
1.2.1. Sosyal Medyanın Özellikleri	26
İKİNCİ BÖLÜM	29
2. TWİTTER VERİLERİ	29
2.1. Twitter Jargonu	29
2.2. Verilerin Toplanması ve Etiketlenmesi.....	30
2.3. Verinin Ön İşlemesi	33
2.4. Kelime Temsili (Word Embedding)	34
2.5. N- gram	37
ÜÇÜNCÜ BÖLÜM	39
3. LİTERATÜR ÇALIŞMASI	39
DÖRDÜNCÜ BÖLÜM	48
4. METEDOLOJİ.....	48
4.1. Yapay Zekâ (Artificial Intelligence)	48

4.1.1.Yapay Sinir Ağları (Artificial Intelligence Networks).....	49
4.2. Makine Öğrenmesi (Machine Learning).....	52
4.2.1. Makine Öğrenimi Türleri	54
4.2.1.1.Denetimli Öğrenme (Supervised Learning).....	54
4.2.1.2.Denetimsiz Öğrenme (Unsupervised Learning).....	55
4.2.1.3. Yarı Denetimli Öğrenme (Semi - Supervised Learning)	56
4.2.1.4.Takviyeli Öğrenme (Reinforcement Learning).....	57
4.3. Makine Öğrenmesi Ölçekleri	57
4.3.1. Accuracy (Doğruluk)	57
4.3.2. Sensitivity (Duyarlılık)	59
4.3.3. Specificity (Belirginlik)	59
4.3.4. Precision (Hassaslık).....	60
4.3.5. F-Score	60
4.4. Derin Öğrenme (Deep Learning)	61
4.4.1. Konvansiyonel Sinir Ağları (Convolutional Neural Network-CNN)	64
4.4.2. Tekrarlayan sinir Ağı (Recurrent Neural Network-RNN)	66
4.4.2.1. Kapılı Yinelemeli Sinir Ağları (GRNN).....	70
4.4.3. Uzun Kısa Dönem Bellek (Long-Short Term Memory-LSTM).....	70
4.4.4. Kapılı Sinir Ağı (Gated Recurrent Unit-GRU).....	72
BEŞİNCİ BÖLÜM	74
5. DENEYSEL KIYASLAMALAR VE SONUÇLAR	74
5.1. Deneysel Kıyaslamalar	74
5.1.1. GRU İki Sınıflı Verilerin Sınıflandırma Başarımlarının Değerlendirilmesi.....	75
5.1.2. LSTM Modeli ile Verilerin Sınıflandırma Başarımlarının Değerlendirilmesi.....	78
5.2. Sonuçlar ve Öneriler	83
KAYNAKLAR	86
ÖZGEÇMİŞ.....	Error! Bookmark not defined.

TABLÖLAR

Tablo 1:	Dünya Bilgi ve İletişim Sektörü	10
Tablo 2:	İnternet Kullanan Bireyler	12
Tablo 3:	Türkiye Sabit Telefon, Cep Telefonu ve İnternet Abone Sayısı.....	15
Tablo 4:	Dünya Geneline En Çok Kullanılan Sosyal Medya Platformları ve Aylık Aktif Kullanıcı Sayıları	23
Tablo 5:	Türkiye’de En Popüler Sosyal Medya Platformları.....	24
Tablo 6:	Telekomünikasyon ile İlgili Tweetlerin Elde Edilmesinde Kullanılan Anahtar Kelimeler.....	31
Tablo 7:	Konu Bazlı Tweet Sayıları.....	32
Tablo 8:	GRU Word2vec Kelime Temsillerini Kullanarak Verilerin Sınıflandırma Başarımları	76
Tablo 9:	GRU Doğruluk Başarım Oranları	76
Tablo 10:	GRU N-Gram TF-IDF Yöntemi Kullanarak Verilerin Sınıflandırma Başarımları	77
Tablo 11:	GRU N-Gram TF-IDF Yöntemi Kullanarak Alt Sınıflar Bazında Verilerin Sınıflandırma Başarımları	78
Tablo 12:	LSTM Word2vec Kelime Temsillerini Kullanarak Verilerin Sınıflandırma Başarımları	79
Tablo 13:	LSTM Doğruluk Başarım Oranları.....	79
Tablo 14:	LSTM N-Gram TF-IDF Yöntemi Kullanarak Verilerin Sınıflandırma Başarımları	80
Tablo 15:	LSTM N-Gram TF-IDF Yöntemi Kullanarak Alt Sınıflar Bazında Verilerin Sınıflandırma Başarımları	81
Tablo 16:	Word2vec Kelime Temsillerini Kullanarak Karşılaştırmalı Doğruluk(ACC) Sınıflandırma Sonuçları	81
Tablo 17:	Word2vec Kelime Temsillerini Kullanarak Karşılaştırmalı Alt Sınıflar Bazında Sınıflandırma Sonuçları	82
Tablo 18:	N-Gram TF-IDF Yöntemi Kullanılarak Karşılaştırmalı Doğruluk(ACC) Sınıflandırma Sonuçları	82

Tablo 19: N-Gram TF-IDF Yöntemi Kullanılarak Karşılaştırmalı Alt Sınıflar Bazında Sınıflandırma Sonuçları	83
--	----



ŞEKİLLER

Şekil 1	: Dünya Bilgi ve İletişim Sektörü Yaygınlık Oranları (100 Kişi Başına)	11
Şekil 2	: ITU'ye Göre Dünya Kişisel İnternet Kullanımı.....	13
Şekil 3	: Türkiye Yıllara Göre Sabit- Cep telefonu ve İnternet Kullanımı.....	16
Şekil 4	: 1996-2020 Yılları Arasında Türkiye Çizgi Grafiği.....	17
Şekil 5	: 2004-2019 Sosyal Medya Platformlarını Kullanan Kişi Sayısı	24
Şekil 6	: Sosyal Medya Platformlarını Kullanan Kişi Sayısı, 2019.....	25
Şekil 7	: Çevrimiçi Sosyal Ağlara Katılan Gençlerin Yüzdesi.....	26
Şekil 8	: Duygu Sınıf Grubu	31
Şekil 9	: Oluşturulan Derlemden Elde Edilen Kelime Bulutu.....	33
Şekil 10	: Basit Bir CBOW Modeli	35
Şekil 11	: CBOW ve Skip-Gram Modellerinin Gösterimi	36
Şekil 12	: 1'den 5'e Kadar Kelime Seviyesi N-Gram Örnekleri.....	38
Şekil 13	: Biyolojik Sinir Ağı Hücresi.....	50
Şekil 14	: Yapay Sinir Hücresi Yapısı.....	50
Şekil 15	: Basit Sinir Ağı Modeli	51
Şekil 16	: Yapay Sinir Ağı Kullanım Alanları.....	52
Şekil 17	: Makine Öğrenmesi Şeması.....	54
Şekil 18	: Denetimli Öğrenme	55
Şekil 19	: Denetimsiz Öğrenme	56
Şekil 20	: Duyarlılık ve Belirginlik	60
Şekil 21	: Basit Sinir Ağı Modeli	62
Şekil 22	: Derin Sinir Ağı Modeli	63
Şekil 23	: Ortak Derin Sinir Ağlarının Mimarisi	64
Şekil 24	: Evrişimli Katmanlar	65
Şekil 25	: CNN Mimarisi.....	66
Şekil 26	: Yinelenen Sinir Ağı.....	67
Şekil 27	: Basit Yinelenen Sinir Ağı	69
Şekil 28	: Uzun Kısa Dönem Bellek Ağı.....	71
Şekil 29	: RNN ve LSTM Model.....	71
Şekil 30	: GRU Model Yapısı.....	72

Şekil 31 : GRU Bellek Hücresinin Yapısı	73
--	----



KISALTMALAR

AT&T	: Dünyanın En Büyük Telekomünikasyon Şirketi
AI	: Artificial Intelligence- Yapay Zekâ
BPTT	: Back Propagation in The Time-Zaman İçinde Geri Yayılım
BTK	: Bilgi Teknolojileri ve İletişim Kurumu
Cloud-RAN	: Bulut Radyo Erişim Şebekesi
CNN	: Convolutional Neural Network-Konvolüsyonel Sinir Ağları
DL	: Deep Learning-Derin Öğrenme
GRU	: Gated Recurrent Unit-Kapılı Sinir Ağı
GRNN	: Gated Recurrent Neural Networks-Kapılı Yinelemeli Sinir Ağları
FN	: Değer Negatif Yanlış
FP	: Değer Pozitif Yanlış
IMDb	: Internet Movie Database-İnternet Film Veritabanı
ITU	: International Telecommunication Union- Uluslararası Telekomünikasyon Birliği
LSTM	: Long Short-Term Memory-Uzun Kısa Süreli Bellek
ML	: Machine Learning – Makine Öğrenmesi
NLP	: Neuro Linguistic Programming-Doğal Dil İşleme
NLTK	: Natural Language Toolkit-Doğal Dil Araç Seti
OECD	: Organisation for Economic Co-Operation and Development- Ekonomik İşbirliği ve Kalkınma Örgütü
PTT	: T.C. Posta ve Telgraf Teşkilatı Genel Müdürlüğü
RNN	: Recurrent Neural Network-Tekrarlayan Sinir Ağı
RT	: Retweet

TF-IDF	: Term Frequency(TF) - Inverse Dense Frequency(IDF) - Terim Frekansı(TF) - Ters Yoğun Frekans(IDF)
TN	: Değer Negatif Doğru
TÜİK	: Türkiye İstatistik Kurumu
TP	: Değer Pozitif Doğru
TTNET	: Türk Telekom Net
YSA	: Yapay Sinir Ağları
Wireless PON	: Kablosuz Pasif Optik Ağ
WAS	: We Are Social-Biz Sosyaliz

GİRİŞ

Günümüz dönemin en önemli endüstrilerinden birisi olan bilgi teknolojileri ve iletişim sektörü hızla büyüyerek gelişmektedir. Telekomünikasyon sektörünün gelişimi, elektronik ticaret ve tele pazarlamada yeni teknolojilerin birçok farklı telekomünikasyon şebekeleri yardımıyla tüketicilerin hizmetine sunumu sağlanarak; maliyetlerin azalmasıyla birlikte gelir seviyesi, etkinlik ve şeffaflık gibi alanlarda önemli ilerlemeler sağlanmaktadır.

Telekomünikasyon sektöründe talep, artan nüfus, hızla büyüyen dijitalleşme ve yükselen şehirleşme trendi gibi faktörlerle birlikte sürekli olarak artmaktadır. Bu sektöre yönelik talep, uzun vadeli olarak hız kesmesi beklenmeyen bir ivmeyle artmaya devam etmektedir, ancak alt kalemlerde bazı azalmalar gözlemlenebilir. Covid-19 pandemisi sürecinde birçok şirketin uzaktan çalışmaya geçmesi ve eğitimin sanal ortamda devam etmesi, güçlü altyapı ve kesintisiz hizmet ihtiyacını artırmıştır. Bu durum, telekomünikasyon şirketlerine olağanüstü bir talep getirmiştir. Pandemi döneminde, dünya genelinde izolasyon ve karantina önlemleriyle birlikte veri trafiğinde ciddi artışlar gözlemlenmiştir. Telekomünikasyon şirketleri, artık sadece iletişim sağlayıcısı olmaktan çıkarak, teknoloji üreten ve müşteri ihtiyaçlarına önceden cevap veren projeler geliştiren kurumlara dönüşmektedir. Bu şirketler, yeni teknolojileri benimseyerek, hızlı ve güvenilir iletişim hizmetleri sunmanın yanı sıra, müşterilere dijitalleşme, bulut bilişim, yapay zekâ gibi alanlarda da çözümler sunmaktadır. Özetle, telekomünikasyon sektörü, artan nüfus, dijitalleşme ve şehirleşme trendleriyle birlikte talebin sürekli olarak arttığı bir sektördür. Covid-19 pandemisi de bu talebi daha da artırmış ve telekomünikasyon şirketlerini teknoloji odaklı hizmet sağlayıcıları haline getirmiştir.(KPMG, 2020).

Sosyal medya, dijital iletişim çağının ulaşılmasını kolaylaştıran sanal ortam araçlarını temsil etmektedir. Sosyal medya platformları aracılığı ile gerçek zamanlı tartışmalara katılmak, aynı ve farklı toplumdaki bireyler arasındaki ekonomik, kültürel, demografik, coğrafi ve kültürel farklılıkların, sınırların kaldırılmasına olanak sağlamaktadır. Sosyal içerikli bu platformlarda bireylerin fikir düşünceleri, deneyimleri üzerinde durularak ilgili konuşma ortamlarına zemin hazırlayan çevrimiçi

platformlar olarak değerdendirilir. Ortak bir noktada birleşerek daha iyi karar verebilmeye yönelik olarak daha çok bilgi edinme olanağı taşımasıdır.

Bununla birlikte teknolojinin, mobil iletişimin ve internetin gelişimi ile birlikte aktif kullanılan sosyal medya platformlarının tüketicilerin hizmetlere olan memnuniyet ve şikayetlerini ifade etmede öncülük etmektedir. Şebeke, mobil, internet ve diğer telekomünikasyon hizmetlerinin kullanıcıların çoğunun çoklu hizmet anlayışı ile kullanımı ve bir arada elde etmesine olanak sağlayan teknolojinin gelişiminde büyük öncülük sağladığı ve faydalandığı görülmektedir. Bununla birlikte sektörün takip edilebilmesi gelişimine odaklanması zorunlu hale getirilmektedir.

Bu amaç doğrultusunda bu tez çalışmasının birinci bölüm genel kısımlarında telekomünikasyon sektörünün ve Türkiye'deki telekomünikasyon sektörünün hizmet ve kalitesinin değerdendirme ve tanımı, tüketici memnuniyet ve şikayetlerinin telekomünikasyon sektörü kapsamında kavramsal tanımı, teknoloji ve iletişim çağının çevrimiçi platformları olan sosyal medya araçlarının önemi ve veri toplama yönteminin derin öğrenme yöntemleri ile sınıflandırma algoritma detaylarına yer verilmiştir. Bir sonraki bölümlerde Türkiye telekomünikasyon sektörünün sosyal medya platformları üzerinde aldığı şikâyet ve olumsuz yorumların derin öğrenme algoritmaları ile analizinin yapılması, sınıflandırma sonucunda performans ölçütlerinin karşılaştırılması yapılarak en iyi başarıyı sağlayan model belirlenmiştir. Yapılan bu tez çalışmasının diğer yapılan çalışmalardan farkı, Türkiye telekomünikasyon sektörünün eksik yönlerinin belirlenmesi ve firmalar bünyesinde geleceğe yönelik atılımlarında fayda sağlaması, elde edilen sonuçların literatüre katkısı, sonuçların tartışılması ve gelecek çalışmaların neler olabileceği konularına yer verilmiştir.

BİRİNCİ BÖLÜM

1. GENEL TANIMLAR

1.1. Telekomünikasyon Sektörü

Telekomünikasyon, bilgi toplumunun iletilmesi hususunda gereken zemini hazırlayan ve bu bilginin etkin bir şekilde aktarılmasına yardımcı olmaktadır. Telekomünikasyon sektörü geleceğe yön veren teknoloji ve iletişim en etkin ve doğru şekilde aktarılmasını sağlayan sürekli gelişen bir sektördür.

İletişim anlamına gelen telekomünikasyon kavramı genel olarak farklı iletim sistemleri ile birlikte her türlü verilerin iletilmesi, gönderilmesi ve alınması anlamına gelmektedir. Telekomünikasyon sektörü, elektronik sektörün bir dalı olarak kabul edilmekte olup telefon cihazı ve diğer ucuna bağlı kablo ağlarından oluşan ses iletimi sistemi ile özdeş kabul edilmektedir. Gelişen teknoloji ve iletişim çağıyla birlikte sesin yanında görüntü, metin resim, hareketli resimler(video) gibi görüntülerinde iletilmesi mümkün hale gelmiştir.

Telekomünikasyon sektörü ister telefon ister internet ister hava dalgaları ister kablolu veya kablosuz iletişim olsun, küresel ölçekte iletişimi mümkün kılan şirketlerden oluşmaktadır. Sözlü, sesli veya görüntülü verilerin dünyanın herhangi bir yerine gönderilmesini sağlayan altyapıyı bu şirketler oluşturmaktadır. Sektördeki en büyük şirketler telefon (kablolu ve kablosuz) operatörleri, uydu şirketleri, kablo şirketleri ve internet servis sağlayıcılarıdır.

Kısa bir süre önce telekomünikasyon sektörü, büyük ulusal ve bölgesel operatörlerden oluşan bir kulüpten oluşmaktaydı. 2000'li yılların başından bu yana, endüstri, hızlı kuralsızlaştırma ve inovasyona sürüklenmiştir. Dünyanın birçok ülkesinde, hükümet tekelleri ve özelleştirilmiş çok sayıda yeni rakip karşı karşıyadırlar. Mobil hizmetlerdeki büyüme sabit hattı geçtiğinden ve İnternet'in temel iş olarak sesin yerini almaya başlamasıyla geleneksel pazarlar harekete geçmiştir.

Önemli Bazı Çıkarımlar

Telekomünikasyon sektörü, dünya genelinde kelime, ses, ses veya video olarak veri ileten şirketlerden oluşmaktadır.

- Telekom ekipmanı, Telekom hizmetleri ve kablosuz iletişim, telekomünikasyonun üç temel alt sektörüdür.
- Telekom, sesin aksine giderek daha fazla video, metin ve veriye odaklanmaktadır.
- Telekomünikasyon şirketleri hem büyüme hem de gelir odaklı yatırımcılara hitap edebilmektedir.
- Bireysel hisse senetleri oldukça değişken olabilse de, telekomünikasyon iş çevrimlerinden etkilenmeyen, giderek daha önemli bir temel endüstri haline geldiğinden, Telekom sektörü genel olarak istikrarlı ve uzun vadeli bir büyüme sergilemektedir(Investopedia, 1999).

1.1.1. Telekomünikasyon Sektörünün Temel Hizmetleri ve Durum Analizi

Telekomünikasyon şebekesinin, santral noktasından hizmetin iletileceği noktalar olan ev, iş yeri gibi kadar ki kısmına “yerel şebeke ağı” denilmektedir. Yerel enerji dağıtım ağları, temel enerji hizmetlerinin en temel birimleridir ve nitelikleri itibariyle hala 'doğal tekel' olarak kabul edilmektedir. Bu ağlar, düşük kapasiteli ancak çok yaygın bir yapıya sahiptir ve hatların yatırım, bakım ve işletme maliyetleri yüksektir. Bu nedenle, sektörün en önemli segmentini oluşturmaya rağmen yerel enerji dağıtım ağları, ana enerji şirketleri için evrensel hizmet zorunluluğu gibi yükümlülükler nedeniyle sınırlı kâr potansiyeline sahiptir. Enerji ağlarının farklı bölgeleri birbirine bağlayan enerji iletim ağı bölümüne ise 'şehirlerarası ağ' denilmektedir. Ulusal enerji şebekesinin uluslararası bağlantısını sağlayan bölüme ise 'uluslararası ağ' denilmektedir. Bu bölümler genellikle yüksek kapasiteli kablo veya uydu iletim sistemlerinden oluşmaktadır. Hat başına yatırım, bakım ve işletme maliyetleri, yerel enerji dağıtım ağlarına göre daha düşüktür (Kent, 2012).

a) Güçlü Yönler

- Güçlü talep tabanına dayanan hızlı büyüme trendi
- Etkili ve sağlam iletişim altyapısı
- Akıllı telefon kullanımındaki yaygınlığın artması
- "Elektronik imza ve diğer mobil uygulamaların kullanımının artması

- Sektör şirketlerinin güçlü kurumsal, mali ve insan kaynakları altyapıları
- Sektörün tüketiciler nezdindeki etkileyici imajı
- 4.5G altyapısının 5G için büyük ölçüde uyumlu olması
- Sektör aktörlerinin, iletişim hizmetlerinin ötesine geçerek teknoloji şirketi olma vizyonu

b) Fırsatlar

- Endüstri 4.0 dönüşümü kapsamında 5G için yüksek talep beklentisi
- Operatörler için ağ ve servis hizmetinin yanı sıra 5G aracılığıyla geliştirilecek ileri teknoloji sunma imkânı
- Tüm dünyada teknolojinin çok hızlı gelişmesinin etkisiyle telekomünikasyon alanındaki hizmetlerin gelişmesi
- Teknolojiye düşkün ve yeni ürünleri hemen benimseyen, talep eden demografik yapı
- Teknolojik mobil cihazlara olan talebin canlılığı ve internete olan inelastik talep
- Özellikle mobil mecra da var olan penetrasyon potansiyeli
- İçerik ve uygulama sektörünün yüksek dinamizmi
- Elektronik hizmet ve alışveriş sektörünün büyümeye devam etmesi

c) Zayıf yönler

- Görece düşük sermaye birikimi sebebiyle yatırım harcamalarının yeterli miktarda büyümemesi
- Teşviklerin etkin işlememesi
- Altyapı sağlayıcılarının ileri teknoloji içeren ürünlerde İthalata bağımlılık oranının yüksekliği
- Özellikle vergi mevzuatının karmaşıklığı ve yaşanan hızlı değişimler
- Vergi ve yasal yükümlülüklerin yüksekliği

- Özellikle fiber altyapısının geliştirilmesi noktasında yaşanan hukuki ve mali sorunlar
- Rekabet sebebiyle gelir artışının görece yavaş kalması

d) Tehditler

- Makroekonomik koşullardaki belirsizlikler
- Mevzuat güncellemelerinin dünya standartlarına uyumu konusundaki aksiyon sıkıntıları
- Mevzuatın dijital dönüşüme uyumlu olmaması sebebiyle yaşanan vergisel sorunlar
- 5G sürecinin yönetimi konusundaki belirsizlikler
- Yeterince hızlı yürümeyen elektronik dönüşüm projeleri
- Siber saldırılar ve yasadışı içerikler(KPMG, 2020).

1.1.2. Telekomünikasyon Sektörünün Mevcut Sorunları

Telekomünikasyon sektörünün hali hazırda verdiği hizmetler kapsamında kullanıcılar tarafından kalite ve memnuniyeti değerlendirilmekte olup farklı iletişim portallarından (web site, sosyal medya, firma hesapları vb.) bildirilmektedir. Bu sorunlar birkaç başlık altında ifade edilmektedir. Bunun yanı sıra son yıllarda tüm dünyayı etkisi altına alan Covid-19 pandemi süreci telekomünikasyon sektörünün güçlü ve çoğunlukla zayıf yönlerinin ortaya çıkmasına neden olmuştur. Bu süreçte uzaktan çalışılmaya geçilmesi ve eğitimin sanal ortama taşınmasıyla birlikte telekomünikasyon sektörüne de yoğun bir talep yaratmıştır. İstihdam sürecinde yaşanan aksaklıklar telekomünikasyon şirketlerini olumsuz yönde etkilemiştir. Salgının yarattığı olağanüstü sürecinin yarattığı iletişim artışı ve çalışma stilleri üzerindeki etkileri telekomünikasyon şirketlerinin eksik ve karşılayamadığı altyapısı dışında bunları fark ederek ilerleyen dönemler için kendilerine büyüme sürecinde fırsatlar sunmaktadır. İletişim ve haberleşme ağlarına artan talep, salgın sürecinden sonra da devam edeceği öngörüldüğünden şirketlerin iletişimlerinin ve altyapılarının daha güçlü ve daha hızlı olmasını, operasyonel verimin artırılması sağlayacak çalışmalar yapmalarına öncülük etmektedir. Telekomünikasyon sektörünün bir diğer eksik durumlardan bir tanesi de doğal afetler ve deprem olaylarıdır.

Depremiñ öncesinde ve sonrasında büyük şiddetle yaşandıđı ölkelerden biri de maalesef Türkiye'dir. Can ve mal kaybının çok olması, yaşanan aksaklıklar, koordinasyon eksikliđi, iletiřimin tamamen kesilmesi gibi olumsuzluklar, yapılması gerekenler ve halkın bilinçlendirilmesi noktasında eksikler oldukça çoktur. Elektriklerin kesilmesi, mobil telefonların çalışamaz hale gelmesi neticesinde iletiřimin tamamen kopması, hükümet yetkililerinin ilk 48 saat deprem bölgesinden bilgi alınamaması gibi oluşan olumsuz durumların telekomünikasyon sektörünün eksik ve gelişememiş yönlerinden birini temsil etmektedir.

Genel olarak bakıldığında telekomünikasyon sektöründe kullanıcılar tarafında belirtildiđi ve firmanın kendi bünyesinin sahip olduđu bazı sorunlar mevcuttur. Bunlar;

- Altyapı Sorunu
- Şebeke sorunu
- Tarife Deđişikliđi Yapılması ve Kullanıcıların Bilgilendirilmemesi
- Ücretlendirme
- Faturalandırma
- Aşım Sorunu
- İnternet Kullanımı ve Hız Sorunu
- Vodafone
- Turkcell
- Türk Telekom
- Avea
- Baz İstasyonu
- Telekomünikasyon Firması Müşteri Hizmetleri
- Arıza
- Tarife Kullanımı
- GigaBayt(GB) Kullanımı

- MegaBayt(MB) Kullanımı
- 4.5 G
- Wifi
- SMS, Dakika, İnternet Kullanımı
- Saniyede Megabit(MbPs)

ve benzeri sorunlar sıralanabilir.

1.1.3. Dünya’ da Telekomünikasyon

Dünya Telekomünikasyon sektörünün başlangıcı 1876 yılında İlk telefon Alexander Graham Bell tarafından icat emesiyle başlamış olup 1960’a kadar yapılan çalışmalarla birlikte 1960 yılından sonra telefonlar, sesli mesajlaşma, hızlı arama ve arayan kimliği gibi özellikleri etkinleştiren mekanik anahtarlardan elektronik anahtarlarmaya geçiş yapmaya başlandı. 1984 AT&T (Dünyanın en büyük telekomünikasyon şirketi)'ye ABD'de telekomünikasyon hizmetleri olarak kullanılan Bell Sistemi bozuldu ve diğer sağlayıcılar için rekabet alanı oluşmuştur. 1984 Mobil iletişimi iki yönlü telsiz kullanımının ötesinde sunan hücresele ve kişisel iletişim hizmeti telefon kullanımı tanıtılmıştır. 1990'lı yıllarda modern internet kullanımı yaygınlaşmaya başlanmıştır. 2000'ler ve sonrasında cep telefonlarının giderek daha yaygın hale gelmiştir ve 2012 yılından günümüze kadar olan süreçte ise akıllı telefon kullanımı yaygınlaşmaya başlanmıştır.

Erişim alanında yapılan düzenlemeler çerçevesinde yıllık 40 milyar Pound’luk ciroyla bugünkü telekomünikasyon piyasası, 1984 yılından bugüne oldukça büyük bir değişime uğramıştır. Telekom piyasası hem hacim olarak büyümüş hem de işyerlerine ve tüketicilere sağlanan hizmetler atmıştır. 1984 yılından bu yana sabit hatların sayısı %75 oranında artmış ve bugün kablo TV şebekesi ve yeni piyasaların ortaya çıkması ile birlikte mobil telefonların sayısı kullarımdaki sabit telefonlardan fazla olmakla birlikte milyonlarca mobil telefon kullanıcısı vardır. İnternet piyasası da oldukça gelişmiş olup, bugün neredeyse her insan internete erişebilmektedir (Topkaya, 2003).

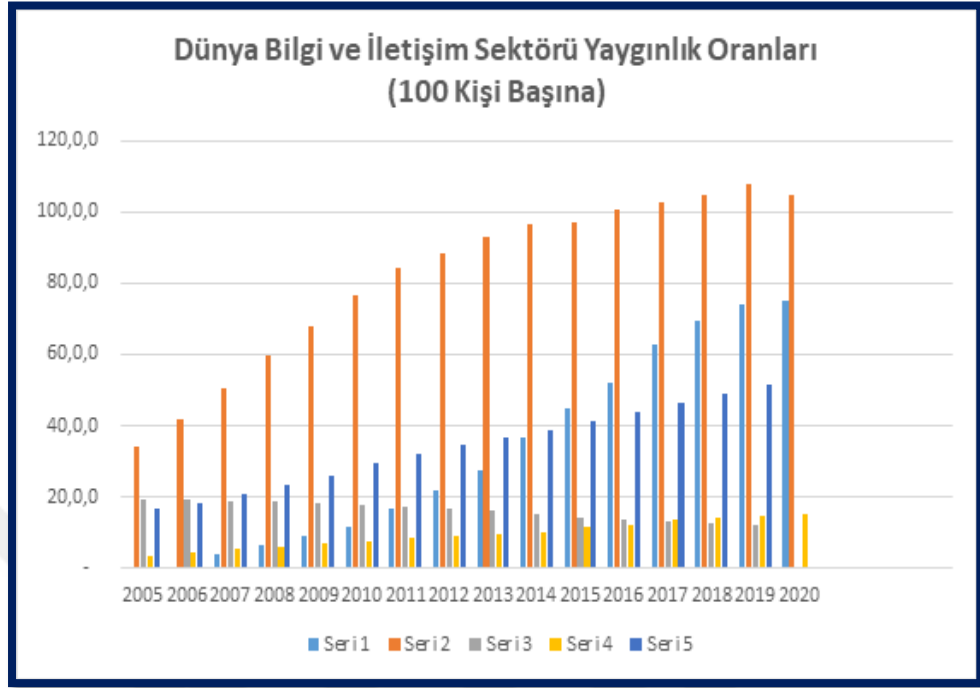
Son yıllarda gelişen teknolojiyle birlikte dünya genelinde bilgi ve iletişim sektörlerinde önemli gelişmeler yaşanmıştır. Hizmet sağlayıcıları altyapı yatırımlarının

da artması ile küresel yaygınlık oranlarında belirgin artışlar ortaya çıkmıştır. Tüketici ihtiyaçlarına göre gelişim gösteren, teknoloji üreten ve tüketicinin ihtiyacı doğmadan cevap veren bir sektör haline gelmektedir. Tüketici davranış ve ihtiyaçları artarken, gelişen teknoloji çağında telekomünikasyon şirketleri hem altyapı hizmetlerinde kalite ve sorunsuz hizmet sunarken, hem de yatırımlarını arttırarak araştırma ve gelişmede öncü sektör olarak büyümeye devam etmektedir. Uluslararası Telekomünikasyon Birliği tarafından 2005-2020 yılları arasında Dünya Bilgi ve İletişim Yaygınlık Oranları (100 kişi başına) Tablo 1’de gösterildiği verilerine göre, 2005-2020 döneminde Sabit telefon abonelikleri yaygınlık oranını belirgin bir şekilde kaybetmiştir. Aktif mobil geniş bant abonelikleri yüz kişi başına %40-%60 seviyelerinde düzenli ve olumlu bir artış gözlemlenirken, sabit geniş bant abonelikleri olumlu bir artış gösterememiştir.

Tablo 1. Dünya Bilgi ve İletişim Sektörü

Yıllar	İnternet Kullanan Bireyler	Sabit Telefon Abonelikleri	Mobil-hücreli telefon abonelikleri
2005	15,8	19,1	33,9
2006	17,5	19,2	41,7
2007	20,5	18,8	50,6
2008	23,1	18,5	59,7
2009	25,5	18,4	68,0
2010	28,9	17,8	76,6
2011	31,3	17,2	84,2
2012	34,0	16,7	88,5
2013	36,0	16,0	93,1
2014	38,1	15,1	96,7
2015	40,5	14,0	97,3
2016	43,3	13,4	100,6
2017	45,8	12,9	102,7
2018	49,1	12,6	104,9
2019	53,6	12,1	107,6
2020	59,1	11,6	107,0
2021	62,5	11,2	109,9
2022	66,3	-	-

Kaynak: (ITU-D, 2023)



Şekil 1. Dünya Bilgi ve İletişim Sektörü Yaygınlık Oranları (100 Kişi Başına)

Kaynak:(ITU-D, 2021)

Seri1: Aktif mobil geniş bant abonelikleri

Seri2: Mobil-hücresel telefon abonelikleri

Seri3: Sabit telefon abonelikleri

Seri4: Sabit geniş bant abonelikleri

Seri5: İnterneti kullanan bireyler

Öte yandan, sabit telefon aboneli bir diğer taraftan mobil- hücresel telefon aboneliklerinde çarpıcı bir artış gözlemlenmektedir. 2007 yılından 2020 yılına kadar olan süreçte mobil hat sayısı her 100 kişiden yaklaşık 30-40 kişide bulunurken yıllar geçtikçe teknoloji geliştikçe mobil hat kullanımının artışı belirgin bir şekilde görülmektedir. 2017-2022 yılları arasında her 100 kişiden 60-80 kişi aktif bant genişliği yaygınlık göstermiştir. İnternet erişimine ulaşan bireylerin 2017 yılından sonra yüksek kullanım artışları gözlemlenmiştir.

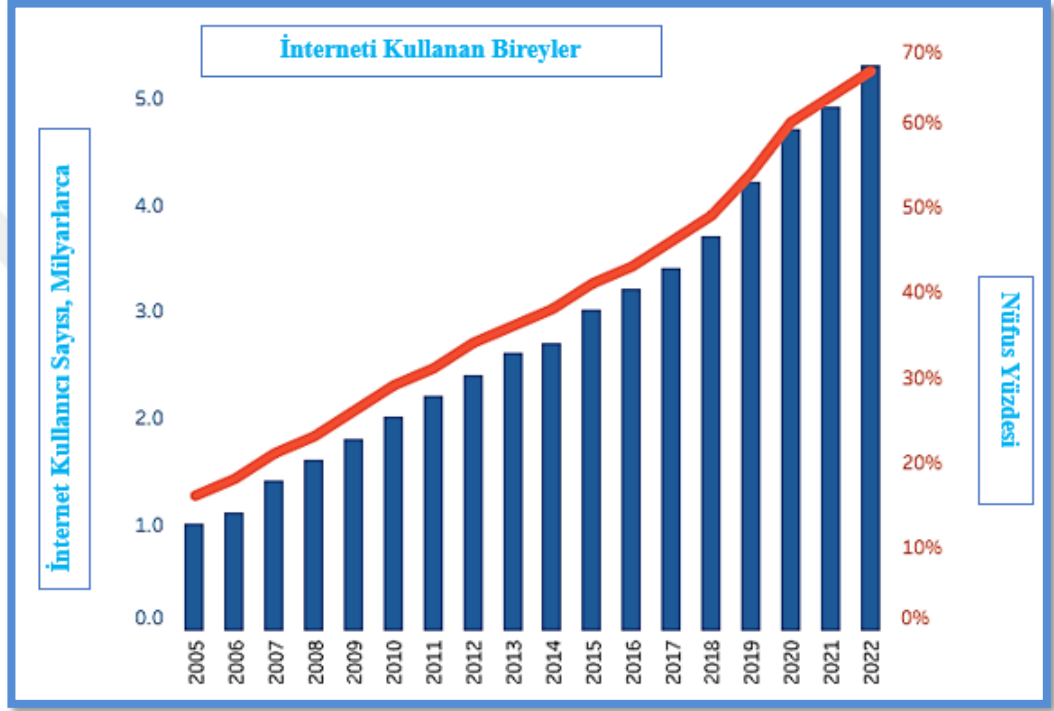
Tablo 2. İnternet Kullanan Bireyler

Yıl	İnternet Kullanıcı Sayısı (Milyar)	%
2022	5.3	66
2021	4.9	63
2020	4.7	60
2019	4.2	54
2018	3.7	49
2017	3.4	46
2016	3.2	43
2015	3	41
2014	2.7	38
2013	2.6	36
2012	2.4	34
2011	2.2	31
2010	2	29
2009	1.8	26
2008	1.6	23
2007	1.4	21
2006	1.1	18
2005	1	16

Kaynak: (ITU-D, 2023)

ITU (International Telecommunications Union- Uluslararası Telekomünikasyon Birliği) istatistiklerine göre 2005-2022 yılları arasında internet kullanıcı sayısı Tablo 2’de verilmiştir. 2005’te internet kullanıcı sayısı %16 iken 2022 yılında kullanıcı sayısı %66 olarak tespit edilmiştir.

ITU' nün tahminlerine 2022'de yaklaşık 5,3 milyar insanın veya dünya nüfusunun yüzde 66'sının internet kullanıcısı olduğunu tahmin etmektedir. Bu, 2019'dan bu yana % 24'lük bir artışı temsil etmektedir. Tahminen 1,1 milyar insan bu süre zarfında çevrimiçi olduğu ancak bu, 2,7 milyar kişinin hala çevrimdışı kalmasına neden olduğu tespit edilmiştir.



Şekil 2. ITU'ye Göre Dünya Kişisel İnternet Kullanımı

Kaynak: (ITU-D, 2023)

1.1.4. Türkiye'de Telekomünikasyon

İlk olarak 1840'ta kurulan Posta Nezareti ile telekomünikasyon hizmeti başlatılmıştır. Kuruma 1909 yılında telefon hizmeti eklenerek PTT nezaretine dönüştürülmüştür. 1994 yılına kadar kullanılan PTT nezaretinden sonra Türk Telekomünikasyon A.Ş. kurulmuştur. Bu kuruluşla birlikte PTT tüm personel, varlık ve yükümlülüklerini Türk Telekom sektörüne aktarılmıştır. Daha sonra 2004 yılında Türk Telekom'un GSM Operatörü Aycell ile İş-TİM'in birleşmesi sonucu TT&TİM İletişim Hizmetleri A.Ş. kuruldu. TT&TİM İletişim Hizmetleri A.Ş.'nin ticari unvanı 15 Ekim'de Avea İletişim Hizmetleri A.Ş. (Avea) olarak değiştirilmiştir. 2005 yılında Türk Telekom'un özelleştirilme çalışmaları tamamlanarak %55 hissesi Ojer

Telekomünikasyon A.Ş.'ye devredilmiştir ve 2006 yılında internet servis sağlayıcı lisansı olarak Türk Telekom aracılığıyla TTNET kurulmuştur ve faaliyetlerine başlamıştır. Avea (yeni adıyla TT Mobil) 2009 yılında 3G hizmeti vermeye başlamıştır ve 2015 yılında, İş Bankası Grubu'nun sahip olduğu Avea İletişim Hizmetleri A.Ş. hisselerinin Türk Telekom'a devir işlemlerinin tamamlanması ile Türk Telekom, Avea hisselerinin tamamına sahip olmuştur. Türk Telekom International, 17 ülkeyi birbirine bağlayacak SEA-ME-WE-5 deniz altı kablo konsorsiyumuna tam üye olmuştur. Şirket iştiraki Avea, Bilgi Teknolojileri ve İletişim Kurumu (BTK) tarafından 26 Ağustos'ta yapılan IMT Hizmet ve Altyapılarına İlişkin Yetkilendirme İhalesi'ne (bilinen adıyla LTE Yetkilendirme İhalesi) katılmıştır. Farklı frekans bantlarında aldığı paketlerle kapasitesini genişleten Avea, abone başına en geniş frekansa sahip mobil operatör haline geldi. Türk Telekom'un mobil iletişim şirketi Avea'nın ticari unvanı, tek marka vurgusunu destekleyecek şekilde Haziran 2018 itibarıyla TT Mobil İletişim Hizmetleri A.Ş. olarak değiştirilmiştir. Türk Telekom, Turkcell, Vodafone, Türksat ve Telkoder arasında, sabit altyapının etkin şekilde kullanılmasını ve mükerrer yatırımların asgariye indirilmesini amaçlayan "Sabit Elektronik Haberleşme Altyapısı Kiralanması" na ilişkin iş birliği protokolü imzalandı. Vodafone ile aktif mobil şebeke paylaşımına yönelik ilk pilot proje çalışmaları başarıyla tamamlanmıştır. Türk Telekom, Nokia iş birliği ile 5G yapay zekâ, Bulut Radyo Erişim Şebekesi (Cloud-RAN), Wireless PON (Kablosuz Pasif Optik Ağ) ve Endüstri 4.0 Akıllı Fabrika teknolojilerini deneyen ilk Türk operatör oldu. 2020 yılında Türkiye'nin ilk gerçek 5G canlı maç yayını Türk Telekom tarafından gerçekleştirilmiştir (Türk Telekom Grubu, 2021).

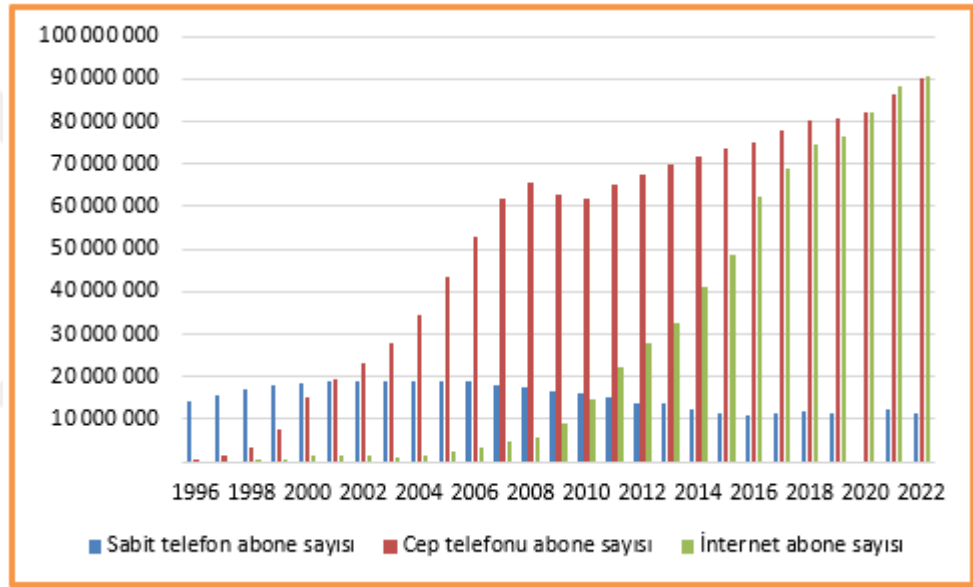
Telekomünikasyon sektörü, modern ve dijital hayatın vazgeçilmez parçası olan, pek çok açıdan hem bireylere hem de kurumlara büyük verimlilik sağlayan bu sektör, stratejik konumunu hem iletişim hem de bilgi teknolojileri anlamında güçlendirmeye devam etmektedir. Türkiye Telekomünikasyon Sektörü' nün gelişen teknoloji ve değişen modern dijital yaşam koşullarında taşıdığı ve uyumlamaya çalıştığı çabası göz ardı edilemez. Birçok alanda kendine geliştiren sektör gün geçtikçe büyümektedir. Gelişmesindeki en etkin unsur Türkiye'nin demografik özellikleri olduğu bilinmektedir. Sektörün mobil – sabit hücresel ve bant kullanımı, internet kullanımı en büyük değişimi etkilemektedir.

Tablo 3. Türkiye Sabit Telefon, Cep Telefonu ve İnternet Abone Sayısı

Yıl	Sabit telefon abone sayısı	Cep telefonu abone sayısı	İnternet abone sayısı
1996	14 286 478	692 779	-
1997	15 744 020	1 483 149	-
1998	16 959 500	3 382 137	229 885
1999	18 054 047	7 562 972	436 610
2000	18 395 171	14 970 745	1 629 156
2001	18 904 486	19 502 897	1 619 270
2002	18 914 857	23 323 118	1 309 770
2003	18 916 721	27 887 535	906 650
2004	19 125 163	34 707 549	1 474 590
2005	18 978 223	43 608 965	2 248 105
2006	18 831 616	52 662 709	3 180 580
2007	18 201 006	61 975 807	4 842 798
2008	17 502 205	65 824 110	5 804 923
2009	16 534 356	62 779 554	8 849 779
2010	16 201 466	61 769 635	14 443 644
2011	15 210 846	65 321 745	22 371 441
2012	13 859 672	67 680 547	27 649 055
2013	13 551 705	69 661 108	32 613 930
2014	12 528 865	71 888 416	41 272 940
2015	11 493 057	73 639 261	48 617 291
2016	11 077 559	75 061 699	62 280 191
2017	11 308 444	77 800 170	68 869 578

2018	11 633 461	80 117 999	74 500 089
2019	11 532 903	80 790 877	76 639 695
2020	12 448 604	82 128 104	82 364 590
2021	12 310 016	86 288 834	88 164 739
2022	11 197 979	90 297 565	90 649 859

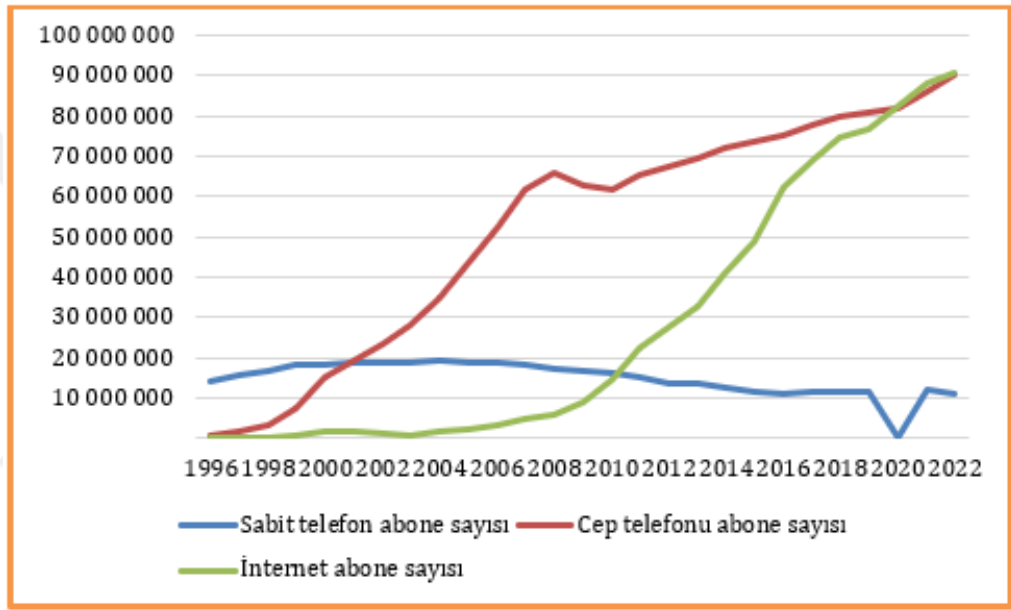
Kaynak: (Tüik, 2023)



Şekil 3. Türkiye Yıllara Göre Sabit- Cep telefonu ve İnternet Kullanımı

Türkiye telekomünikasyon sektörünün faaliyete geçtiği ilk yıllardan günümüze kadar gelişim ve etkileşim sürecini TÜİK verileri üzerinden incelediğimizde gösterdiği gelişim ve sağladığı olanakların nasıl artış olduğunu göstermektedir. Sektördeki bu mevcut potansiyeli destekleyen en önemli unsurların başında Türkiye'nin demografik yapısının geldiği bilinmektedir. Sektörün yarattığı alan yalnızca iletişim anlamında değil, teknolojik atılımlarla da büyümektedir. 1996-2022 yılları arasında sabit telefon abone sayısı, cep telefonu abone sayısı, internet abone sayısı Şekil 3'de gösterildiği gibidir. Modern hayatın yapı taşlarından olan, pek çok açıdan hem bireylere hem de kurumlara büyük verimlilik sağlayan telekomünikasyon sektörü, stratejik konumunu hem haberleşme hem de bilgi teknolojileri anlamında güçlendirmeye devam etmektedir. Ülkemizde, sabit telefon abone sayısı azalırken, cep telefonu abone sayısı ve internet

abone sayısı hızla artmaktadır. 1996 ve 2006 yılları arasında sabit telefon kullanımı düzeyli bir şekilde artarken, 2007 yılından sonra kullanımının azaldığı görülmektedir. Sabit telefon abone sayısı azalırken 2000 yıllardan sonra cep telefon abone sayısı arttığı görülmektedir. Teknolojinin gelişimi ve sektörün gittikçe güçlenmesiyle birlikte 2009 yılından sonra günümüze kadar internet abone sayısının hızlıca attığı görülmektedir. İnternet abone sayısı hızla artışıyla birlikte bağlantılı olarak sosyal medya kullanımı da artış gözlemlenmektedir.



Şekil 4. 1996-2020 Yılları Arasında Türkiye Çizgi Grafiği

Kaynak: (Tüik, 2023)

1.2. Sosyal Medya Kavramı

Günümüzde teknoloji hızla gelişmektedir. Teknoloji ilerleme kaydettikçe web ve internet kullanımı da yaygınlaşmaktadır. İnternet, mobil iletişim, dijital medya ve çeşitli sosyal medya araçlarının yaygınlaşması yazılımlar, yatay etkileşimli iletişim ağlarının geliştirilmesine yol açmıştır (Castells, 2007). Günlük hayatımızda iletişim aracı olarak kullandığımız yöntemlerden biri kullanıcıların çevrimiçi değişimlere katılmalarına olanak sağlayan, kullanıcı tarafından oluşturulan katkıda bulunan hizmetler içerik veya çevrimiçi toplulukları oluşturan İnternet tabanlı ve mobil geniş bir yelpazeyi ifade eden sosyal medya platformlarıdır (Dewing, 2005). Bu platformlar, bireylerin çevrimiçi olarak diğer kullanıcılarla bağlantılar kurma ve bu bağlantıları sürdürme imkânı sağlamaktadır.

“Sosyal Medya” kavramı ilk olarak 2000’li yıllarda pek çok ağ sitesi ile popüler hale gelmiştir. Bu yeni medya ortamları, işletmelerin hedef kitlelerini etkilemek için kullandıkları pazarlama, reklam ve halkla ilişkiler faaliyetlerini gerçekleştirebilecekleri dijital platformlar olarak ön plana çıkmıştır. Sosyal medya, aynı zamanda kullanıcıların bilgi edinebildikleri, çevrimiçi topluluklar oluşturabildikleri, kişisel mesajlar, videolar ve diğer içerikleri paylaşılabildikleri bir iletişim aracıdır. Bu platformlar, bireylerin, grupların veya kurumların bilgi veya hizmet alışverişi yapabileceği elektronik iletişim formları olarak da tanımlanabilir. Sosyal medya, internetin yaygınlaşmasıyla birlikte kullanıcıların etkileşimde bulunabileceği, bağlantılar kurabileceği ve bilgi paylaşılabileceği bir alan haline gelmiştir. Günümüzde sosyal medya platformları, insanların günlük yaşamlarının bir parçası haline gelmiş ve iletişim, eğlence, haber alma, bilgi edinme ve ticaret gibi birçok amaç için kullanılmaktadır (Güçdemir, 2017).

Bir diğer ifade ile sosyal medya, günümüzde boş zamanlarımızı değerlendirdiğimiz, eğlendiğimiz, sohbet ettiğimiz, bilgi alışverişinde bulunduğumuz ve gündemi takip etmek için başvurduğumuz temel araç haline gelmiştir (Solmaz et al., 2013).

Sosyal medya, insanların içerik, fikir, içgörü, deneyim ve perspektiflerini paylaşmak için kullandıkları çevrimiçi teknolojileri ifade eder. Metin, görüntü, ses, video ve diğer birçok formdaki içerikleri internet tabanlı bilgi yayma ve paylaşma mecraları veya ortamları olan Facebook, Twitter, Instagram, YouTube gibi platformlar aracılığıyla gerçekleştirilir. Sosyal medya, kullanıcıların iletişim kurmalarını, etkileşimde bulunmalarını ve bilgilerini yaymalarını sağlar.

Sosyal medya platformları, çeşitli kriterlere göre sınıflandırılabilir. İşte bu başlıklar;

- Bloglar,
- Mikrobloglar (Twitter gibi),
- İnternet forumları
- İnceleme ve değerlendirme siteleri,
- Sosyal haber ve sosyal işaretleme siteleri,

- Sosyal ağ kurma siteleri,
- Medya paylaşım siteleri,
- Sanal dünyalar(Tuncer et al., 2013).

Sosyal Ağlar: İnsanların profillerini oluşturarak diğer kullanıcılarla bağlantı kurabilecekleri ve etkileşime girebilecekleri platformlardır. Facebook ve LinkedIn gibi siteler bu kategoriye örnektir.

Mikroblogging: Kullanıcıların kısa mesajlar veya gönderiler paylaşabileceği platformlardır. Twitter, bu kategoride en popüler olan örnektir.

Fotoğraf Paylaşımı: Kullanıcıların fotoğraflarını yükleyip paylaşabilecekleri ve etkileşimde bulunabilecekleri platformlardır. Instagram, bu tür bir sosyal medya platformudur.

Video Paylaşımı: Kullanıcıların videolarını yükleyebileceği, izleyebileceği ve paylaşabileceği platformlardır. YouTube, en büyük video paylaşım platformudur.

Bloglar: Kullanıcıların yazılarını, makalelerini ve günlüklerini paylaşabileceği platformlardır. Blogger ve WordPress, bu tür bir sosyal medya ortamına örnektir.

Tartışma Forumları: Belirli konular hakkında tartışma ve bilgi paylaşımı için kullanılan platformlardır. Reddit, popüler bir tartışma forumu platformudur.

İnceleme Siteleri: Ürünler, hizmetler, filmler veya kitaplar gibi konular hakkında kullanıcıların yorumlarını ve incelemelerini paylaşabilecekleri platformlardır. Yelp ve IMDb, bu kategoriye örnek olarak verilebilir.

Podcast ve Müzik Paylaşımı: Kullanıcıların podcast'lerini veya müziklerini yükleyebilecekleri ve paylaşabilecekleri platformlardır. SoundCloud, bu tür bir sosyal medya platformudur.

Günümüzde en yaygın kullanılan sosyal medya platformları şunlardır:

- Facebook
- Twitter
- Youtube

- LinkedIn
- Instagram
- Whatsapp
- FB(Facebook) Messenger
- Tiktok
- Telegram
- Snapchat

Facebook: Mark Zuckerberg tarafından 2004 yılında kurulan bir sosyal paylaşım sitesi, bireylere sosyal ağlar üzerinden arkadaş edinme, mevcut arkadaşlarla iletişim kurabilme ve bilgi alışverişinde bulunabilme imkânı sunmaktadır. Bu platform, özellikle Harvard Üniversitesi öğrencileri arasında popülerlik kazanmıştır. Zuckerberg'in girişimi, kullanıcılarına çevrimiçi olarak bağlanma ve etkileşim kurma fırsatı veren bir sosyal medya fenomeni haline gelmiştir. Bugün, bu platform dünya genelinde milyarlarca kullanıcıya sahip olup, insanların günlük yaşamlarını paylaşımları, haberleri takip etmeleri ve diğer insanlarla bağlantı kurmaları için bir platform olarak kullanılmaktadır.

Twitter: Anında mesaj aktarma olanağı sağlaması, Twitter'i en etkin ve popüler sosyal medya araçlarından biri haline getirmiştir. 2006 yılında Jack Dorsey ve 13 arkadaşı tarafından kurulmuştur (Uluç & Yarcı, 2017). Twitter cep telefonlarında dahi kullanılabilir bir araç haline geldiği için zaman ve mekân kaygısı olmadan anlık iletişim ve paylaşım gerçekleştirebilmektedir. Bunun yanı sıra; içerik tartışılabilir, yorumlanabilir ve hatta beğenilmeyen içeriklere müdahale edilerek anında değişiklik gerçekleştirilebilir. Twitter hem bireysel hem de kurumsal anlamda haber, iletişim, bilgi, reklam ve eğlence amaçlı kullanılmaktadır (Güçdemir, 2017).

Youtube: Günümüzde milyonlarca insanın video paylaşmasına ve izlemesine imkân veren, video barındırma sosyal medya mecrasıdır. Aralık 2007'de İngiltere Kraliçesi YouTube'da "Kraliyet Kanalı (The Royal Channel)"nı faaliyete geçirmiştir.

Linkedin: 2003 yılında profesyonel iletişim ağı olarak kurulmuştur. Bireyler eğitim durumlarını, iş tecrübelerini, bilgi ve becerilerini LinkedIn üzerinden paylaşabildiği ve bireylere geniş bir iş çevresiyle temas olanağı sağladığı sosyal medya platformudur.

Instagram: 2010 yılında Kevin Systrom ve Mike Krieger adlı iki girişimci tarafından kurulan instagram kullanıcılarına fotoğraflarını paylaşma olanağı sağlayan bir dijital platformdur (Uluç & Yarcı, 2017).

WhatsApp: 2009 yılında Jan Koum, Kaliforniya'da kurulan WhatsApp, platformlar arası çalışma özelliğine sahip bir anlık mesajlaşma, arama veya haberleşme uygulamasıdır. İnternet bağlantısı aracılığıyla kullanıcıların birbirlerine fotoğraf, video, ücretsiz arama, sesli ve yazılı mesaj ve doküman gönderilmesini sağlar.

Telegram: Rus programcısı Pavel Durov tarafından yaratılmıştır. Çoklu platform desteği sunan, anlık mesajlaşma servisi.

Snapchat: Kişiyi özel olarak iletilen dosyalar, gönderen kişinin belirlediği süre (1 ila 10 saniye arasında) boyunca yalnızca iki kez görüntülenebilir. Eklenen fotoğraf ve videolar ise gönderen kişinin belirlediği süre boyunca (1 ila 10 saniye arasında) istenilen kadar görüntülenebilen bir anlık mesajlaşma uygulaması olan Snap Inc. tarafından akıllı telefonlar için geliştirilen platformlar arası çalışma özelliğine sahip bir sistemdir. Bu uygulama, Evan Spiegel, Bobby Murphy ve Reggie Brown tarafından kurulmuştur ve kullanıcıların birbirlerine yazılı metin, fotoğraf, ses ve video göndermelerini sağlar.

Reddit: Bir Amerikan sosyal medya platformu olan Reddit, haberlerin ve tartışmaların yapıldığı bir site olarak bilinir. Kullanıcılar, bağlantılar, metinler, fotoğraflar, anketler ve videolar gibi çeşitli içerikleri paylaşabilirler. Paylaşılan gönderiler, diğer kullanıcılar tarafından oylanarak yukarı veya aşağı yönlü puanlar alır. Bu oylar, gönderinin görünürlüğünü ve sıralamasını etkiler. Reddit'te kullanıcılar tarafından oluşturulan "subreddit" adı verilen alt gruplarda farklı konular ele alınır. Spor, oyun, müzik, yemek, fotoğrafçılık, haber gibi çeşitli konuların yanı sıra herkes kendi ilgi alanına göre subreddit oluşturabilir. Subreddit'lerdeki gönderiler, yeterli sayıda oy aldığı anda Reddit'in ana sayfasında yer alır. Reddit, geniş bir kullanıcı tabanına sahiptir ve çeşitli topluluklar arasında etkileşim ve tartışma imkânı sunar. Kullanıcılar, gönderilere yorum yapabilir, tartışmalara katılabilir ve içerikleri paylaşabilir. Bu özgürlük ve çeşitlilik, Reddit'in popülaritesini artıran faktörler arasında yer almaktadır. Bir Amerikan sosyal haber, tartışma sitesidir. Siteye kayıtlı olan kullanıcılar bağlantı, metin, fotoğraf, anket ve video içerebilen gönderiler paylaşabilir, diğer kullanıcılar da bu gönderileri oylayarak

gönderinin sayfada üste çıkmasını veya alta inmesini sağlayabilirler. Reddit'te gönderiler kullanıcılar tarafından oluşturulmuş; spor, oyun, müzik, yemek, fotoğrafçılık, haber gibi farklı konuları ele alan, "subreddit" denilen alt gruplarda paylaşılır. Bir subreddit'te paylaşılmış gönderi yeteri kadar oy alırsa bu gönderi Reddit'in ana sayfasına çıkar. Siteye kayıtlı her kullanıcı kendi subreddit'ini oluşturabilir.

Tiktok: Çin'de bilinen adıyla Douyin olan Tiktok uygulaması müzikal bazda üretilen, video oluşturma ve paylaşmanın yanı sıra canlı yayın imkânı sağlayan bir sosyal medya uygulamasıdır. Eylül 2016'da ByteDance tarafından piyasaya sürüldü. Bu sosyal medya platformu; dans, komedi ve eğitim gibi türlerde on beş saniye ile üç dakika arasında değişen çeşitli kısa biçimli videolar yapmak için kullanılıyor (URL1, 2020).

We Are Social(WAS) 2021 raporuna göre; Dünya'da ve Türkiye'de en çok kullanılan sosyal medya listesi Tablo 4'de gösterildiği gibidir.

Tablo 4. Dünya Geneline En Çok Kullanılan Sosyal Medya Platformları ve Aylık Aktif Kullanıcı Sayıları

Sosyal Medya Platformları	Aylık Bazda Aktif Kullanıcı Sayıları
Facebook	2,958 milyar
YouTube	2,514 milyar
WhatsApp	2,000 milyar
Instagram	2,000 milyar
WeChat	1,309 milyar
TikTok	1051 milyon
Messenger	931 milyar
Douyin	715 milyon
Telegram	700 milyon
SnapChat	635 milyon
Kuaishou	626 milyon
Sina Weibo	584 milyon
QQ	574 milyon
Twitter	556 milyon
Pinterest	445 milyon

Kaynak: (We Are Social, 2023)

Çevrimiçi platformlarda ne kadar vakit geçirdiğimizi WAS 2023 raporlarında açıkladığı gibi, dünyanın en çok Facebook, Youtube, WhatsApp ve FB Messenger olduğunu tespit etmiştir. Twitter'ın 14. sırada listelendiğini raporlamıştır. Kuaishou, Sina Weibo, Pinterest ve QQ gibi uygulamaların kalabalık bir kullanıcı kitlesine ve popülerliğe sahip olduğu görülmektedir.

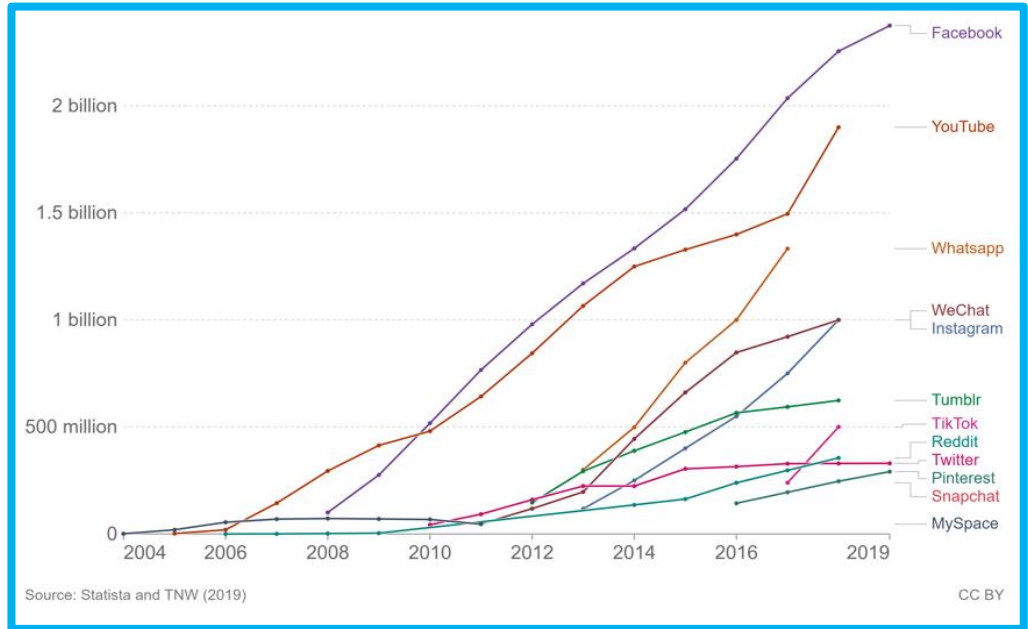
Tablo 5. Türkiye’de En Popüler Sosyal Medya Platformları

Sosyal Medya Platformları	2021(%)	2022(%)
Instagram	76,5	71,1
Facebook	50,5	48,0
Tiktok	43,2	48,2
Twitter	23,6	27,1

Kaynak: (We Are Social, 2023)

Türkiye’de en çok kullanılan çevrimiçi sosyal medya platformunun 2021 ve 2022 yıllarında kullanım oranları Tablo 5’te verilmiştir. Türkiye’de 2021-2022 yılında en çok Instagram kullanımı tercih edilmiştir. Facebook Dünya’da en çok kullanılan platform Facebook olmasına rağmen Türkiye’de Instagram’dan sonra ikinci sırada tercih edilen sosyal medya platformu olmuştur. Twitter’ın 2021 yılına göre Türkiye’de kullanımı %23,6 iken 2022 yılındaki kullanımının artış göstererek %27,1 olarak ölçülmüştür.

Şekil 5. 2004-2019 Sosyal Medya Platformlarını Kullanan Kişi Sayısı

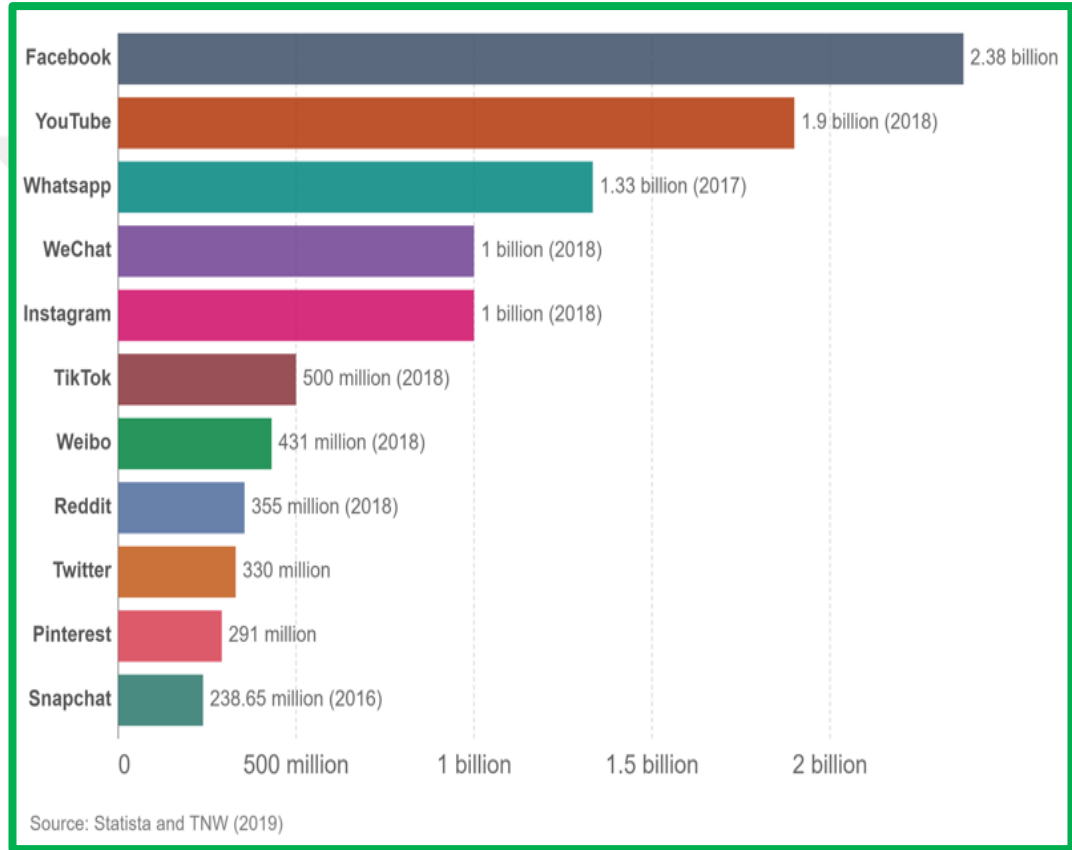


Kaynak: (Ortiz-Ospina ve Roser, 2019)

İnteraktif grafikte 2004’ten bu yana platform bazında aylık aktif kullanıcıları göstermektedir. Bu grafik, Facebook, YouTube ve Reddit gibi on veya daha fazla yıldır var olan bazı büyük sosyal medya sitelerinin olduğunu gösteriyor; ancak diğer büyük

siteler çok daha yenidir. Örneğin TikTok, Eylül 2016'da piyasaya sürüldü ve 2018'in ortalarında zaten yarım milyar kullanıcıya ulaşmıştır. Bunu bir perspektife oturtmak gerekirse: TikTok, bu süre zarfında ayda ortalama 20 milyon yeni kullanıcı kazanmıştır.

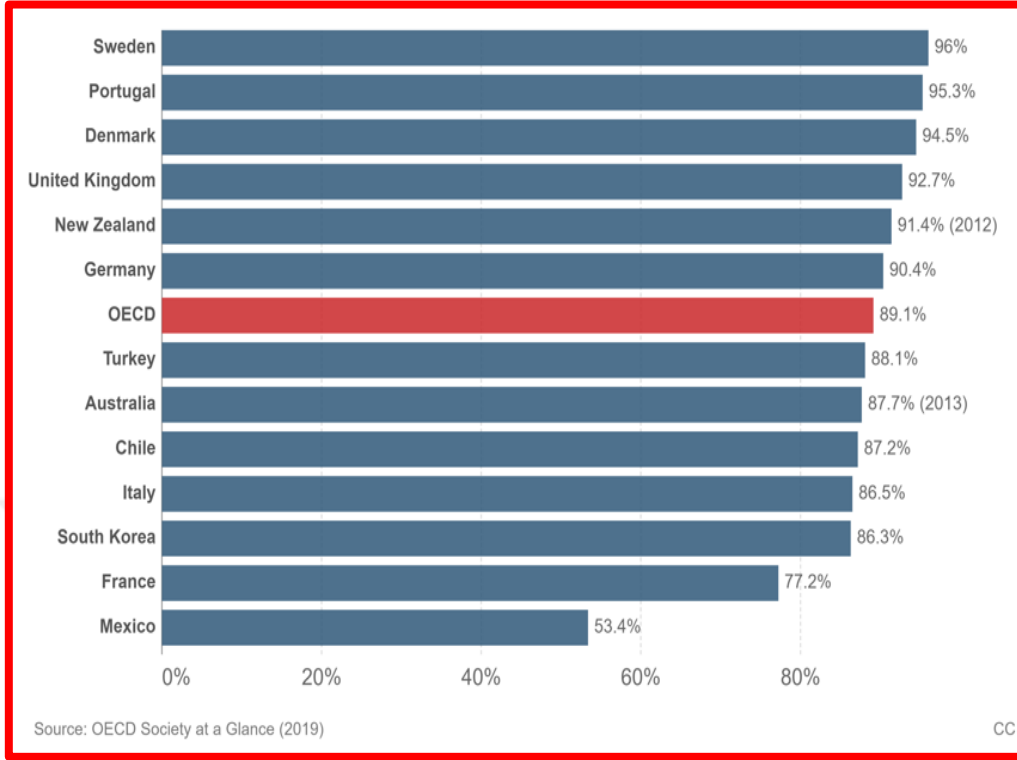
Veriler ayrıca ters yönde hızlı değişimler göstermektedir. Bir zamanlar baskın olan platformlar ilgi kaybı yaşamıştır. 2008'de Hi5, MySpace ve Friendster, Facebook'a yakın rakiplerdi, ancak 2012'de neredeyse hiç kullanılmamaktadır.



Şekil 6. Sosyal Medya Platformlarını Kullanan Kişi Sayısı, 2019

Kaynak: (Ortiz-Ospina ve Roser, 2019)

Facebook, sosyal medya pazarına on yıldır hâkim durumda, ancak diğer beş platformun da her birinin yarım milyardan fazla kullanıcısı var. 2,3 milyar kullanıcısıyla Facebook, günümüzün en popüler sosyal medya platformudur. YouTube, Instagram ve WeChat, bir milyardan fazla kullanıcıyla takip ediyor. Tumblr ve TikTok, yarım milyardan fazla kullanıcıyla, Twitter 330 milyar kullanıcıyla gelmektedir.



Şekil 7. Çevrimiçi Sosyal Ağlara Katılan Gençlerin Yüzdesi

Şekil 7’de, çeşitli ülkelerde sosyal ağları kullanan 16 ila 24 yaş arasındaki kişilerin oranını göstermektedir. OECD ortalaması %90’a yakın bir kullanım olduğu tespit etmiştir. Günümüzün genç yetişkinleri yaşamları boyunca sosyal medyayı kullanmaya devam ederse, internetin benimsenmesi düşük gelirli ülkelerde genişledikçe sosyal medyanın hızla büyümeye devam etmesi muhtemeldir (Ortiz-Ospina ve Roser, 2019).

1.2.1. Sosyal Medyanın Özellikleri

Sosyal medya sitelerini bireylere kendilerini ifade etme, sosyal ağlarını birleştirme, diğer kullanıcılarla bağlantılar kurup bu bağlantıları sürdürme olanağı veren çevrimiçi alanlardır.

Milyonlarca insan web’i ürün ve hizmetler hakkında ayrıntılı araştırma yapmak, siyasi kampanyalara katılmak, müzik ve film hayran kulüplerine katılmak, ürünleri incelemek ve hobileri ve tutkuları tartışmak için kullandığından, her türlü çevrimiçi yerde bir araya geliyorlar. Artık pek çok kişinin toplu olarak sosyal medya olarak adlandırdığı teknolojiler ve araçların tümü, kullanıcıların görüşlerini çevrimiçi olarak ifade etme yollarını içerir:

- Facebook, Twitter ve LinkedIn gibi sosyal ağ siteleri, insanların bir arkadaş topluluğu oluşturmaya ve bilgi paylaşımına yardımcı olur.
- Bloglar, bir konu hakkında tutkulu biri tarafından yazılan kişisel web siteleri, bu tutkuyu dünyayla paylaşmak ve yazarın gönderilerine yorum yapan aktif bir okuyucu topluluğu oluşturmak için bir araç sağlar.
- YouTube, Vimeo, Flickr, SlideShare ve Instagram gibi video ve görüntü paylaşım siteleri, fotoğrafları, grafik görüntüleri ve videoları paylaşma ve bunlara yorum yapma sürecini büyük ölçüde basitleştirir.
- Sohbet odaları ve mesaj panoları, insanların bulunduğu ve ilgilendikleri konuları tartıştığı çevrimiçi toplantı yerleri olarak hizmet eder ve ana özelliği, herkesin bir tartışma dizisi başlatabilmesidir.
- Yelp, Rotten Tomatoes, Amazon ve TripAdvisor gibi inceleme siteleri, tüketicilerin ürünleri, hizmetleri ve şirketleri derecelendirdiği yerlerdir.
- Listservs, sohbet odalarına benzer şekilde, kayıtlı üyelere oluşan bir koleksiyona e-posta ile mesajlar gönderir.
- Wiki'ler, herkesin düzenleyebileceği ve güncelleyebileceği web siteleridir.
- Digg ve StumbleUpon gibi sosyal yer imi siteleri, kullanıcıların başkalarına içerik önermesine ve neyin ilginç olduğuna oy vermesine olanak tanır.
- Foursquare ve Swarm gibi GPS tarafından oluşturulan konum servislerine sahip mobil uygulamalar, her kullanıcının dünyada tam olarak nerede olduğunu belirleme bileşenini ekler(Scott, 2015).

Sosyal medya özellikleri aşağıdaki başlıklar altında toplanabilir:

- 1) Katılım- Duygusal Temelli: istekli olan herkesi katkı yapma geri dönüş sağlama konusunda cesaretlendirmektedir.
- 2) Açıklık: Platform üzerinde servislerinin birçoğu geri bildirim ve katılıma açıktır. Oylama, yorumlama ve bilgi paylaşımı teşvik etmektedirler. İçeriği kullanmada giriş engelleri genelde bulunmamaktadır.

- 3) Karşılıklı Diyalog- Etkileşimli Olma: İki yönlü bir etkileşim ve etkileşim ortamı sağlamaktadır.
- 4) Topluluk Tabanlı: Topluluklar ortak ilgi alanları çerçevesinde bir araya gelirler.
- 5) Bağlantılı-İlişki Temelli Olma: Sosyal medya sitelerinin birçoğu bağlantılı olma özelliklerini geliştirmekte ve diğer sitelere, kaynaklara ve kişilere link vererek trafik sağlamaktadırlar.



İKİNCİ BÖLÜM

Bu bölümde veri seti hakkında bilgi verilmektedir. Verinin hangi yöntemle elde edildiği, temizlenme aşaması açıklanmaktadır.

2. TWİTTER VERİLERİ

2006 yılında Jack Dorsey tarafından kurulan ve günümüzde en çok kullanılan çevrimiçi sosyal medya platformlarından biri olan Twitter, kullanıcılarının “tivist” adı verilen en fazla 280 karakterden oluşan mesajlar göndermesine ve okumasına izin veren, çevrimiçi olarak arkadaş grupları içerisinde bilgi paylaşımı sağlanan, görüş ve düşüncelerin savunulduğu bir sosyal medya ağıdır.

2.1. Twitter Jargonu

Twitter, 280 karakter sınırlamasını getirerek tweetlerin içeriğini etkilemiş ve özgün bir jargonun ortaya çıkmasına yol açmıştır. İnsanlar, duygu ve düşüncelerini ifade etmek için karakter sınırlamasının yetersiz olduğunu düşünerek, kullanılan kelimelerin değişmesine ve tweetlerin başında "#" sembolüyle etiketlenen anahtar kelimeler veya kelime öbekleri oluşturulmasına yönelmiştir. Böylece, çoğunlukla ünlü harflerin yutulduğu hem Türkçe hem de İngilizce ifadeler içeren ve sohbetleri birbirine bağlayan ve içerik bulunmasına yardımcı olan, insanları bir konu etrafında bir araya getiren, topluluklar oluşturan; aynı etiketi kullanan diğer tweetleri görmek içinde kullanılan bir sosyal medya jargonu oluşmuştur.

"Retweet" olarak bilinen paylaşımların Twitter'da tekrar yayınlanması durumu, içeriğin başına "RT" ifadesinin eklenmesiyle gerçekleşir. Her retweet, Twitter platformunda tekrarlanan bir verinin oluşmasına sebep olur. Bunun nedeni birçok kişinin başka kişilerin paylaşımını kendi sayfasında yayımlamasıdır. Tweetlerin tekrarlanmasına neden olan bir diğer yöntemde, birinden ikonlarla (@) bahsetmedir. Bu ifadeler, bahsedilmek istenen kişinin Twitter kullanıcı adının ardından gelen "@kullanıcıadı" şeklindeki ikonlarla kullanılır. Bu ikonlar sık sık tekrar etse de tweetin anlamında olumluluk ya da olumsuzluk ifade etmezler. Bu nedenle, retweetler gibi bu ifadeler de verilerin içerisinden temizlenmesi gerektiği düşünülmüştür. İlaveler yapılarak "@" karakteriyle başlayan kelimelerin tweetlerden çıkarılması önemlidir.

2.2. Verilerin Toplanması ve Etiketlenmesi

Günümüzde sosyal medyanın etkin ve yaygın olarak kullanılması ve insanların birbirleriyle olan iletişimini sağlamaktadır. Böylelikle bu etkileşimler sayesinde çok büyük miktarlarda veriler üretilmektedir. Twitter'da bu sosyal medya platformlarından biridir. Twitter'dan paylaşılan içerikler yapısal olmayan veri türlerindendir. Bu yapısal olmayan veriler metin tabanlıdır ve farklı algoritma yöntemleri ile kullanılabilir formata dönüştürülmesi gerekmektedir. Twitter uygulamasının bu denli etkin kullanılması da telekomünikasyon sektörüne ait şikayetlerinde paylaşılr hale gelmiştir. Böylece bu çalışma kapsamında öncelikli olarak Twitter uygulaması üzerinden Python programlama dili kullanılarak aşağıda Tablo 6'da listesi verilen anahtar kelimelere dayanan bir sorgulama işlemi yapılmıştır. Bu sorgulama sonucunda tweetten oluşan derlem oluşturulmuştur. Kelimenin tamamını içeren bir veya birden fazla sorgusu kullanılmıştır. Dolayısıyla veri setindeki her bir tweet bu kelimelerden en az bir tanesini içermektedir. Anahtar kelimelerin belirlenmesi sürecinde, literatürde yapılan çalışmalar incelenerek; bu çalışmaların değindiği anahtar kelimeler göz önünde bulundurularak ve tweettir platformu üzerinden küçük bir araştırma yapılarak Twitter'da sıklıkla geçen kelimelerden oluşturulmuştur. Böylelikle veri setindeki her bir tweet bu kelimelerden en az bir tanesini içermektedir.

Tablo 6. Telekomünikasyon ile İlgili Tweetlerin Elde Edilmesinde Kullanılan Anahtar Kelimeler

Tweetlerin Elde Edilmesinde Kullanılan Anahtar Kelimeler	
Altyapı Sorunu	Şebeke sorunu
Tarife Değişikliği Yapılması ve Kullanıcıların Bilgilendirilmemesi	Ücretlendirme
Faturalandırma	Aşım Sorunu
İnternet Kullanımı ve Hız Sorunu	Vodafone
Turkcell	Türk Telekom
Avea	PTTcell
Baz İstasyonu	Müşteri Hizmetleri
Arıza	Tarife Kullanımı
GigaBayt(GB) Kullanımı	MegaBayt(MB) Kullanımı
4.5 G	Wifi
SMS, Dakika, İnternet Kullanımı	

Elde edilen veri setinde yer alan tweet metinlerine etiketleme işlemi yapılmıştır. Etiketlemede tarafsız bir bakış açısı ile olumlu, olumsuz ve nötr olmak üzere üç ayrı sınıf bilgisine dayanarak etiketlenmiştir. Üç sınıf grubu için; olumsuz tweet 1, olumlu tweet 2, nötr tweet 0 olarak atanmıştır. Etiketlenen her bir tweet metni ise baz alınan anahtar kelimenin içerdiği tweet için 1, içermediği tweet için 0 olarak atanmıştır.

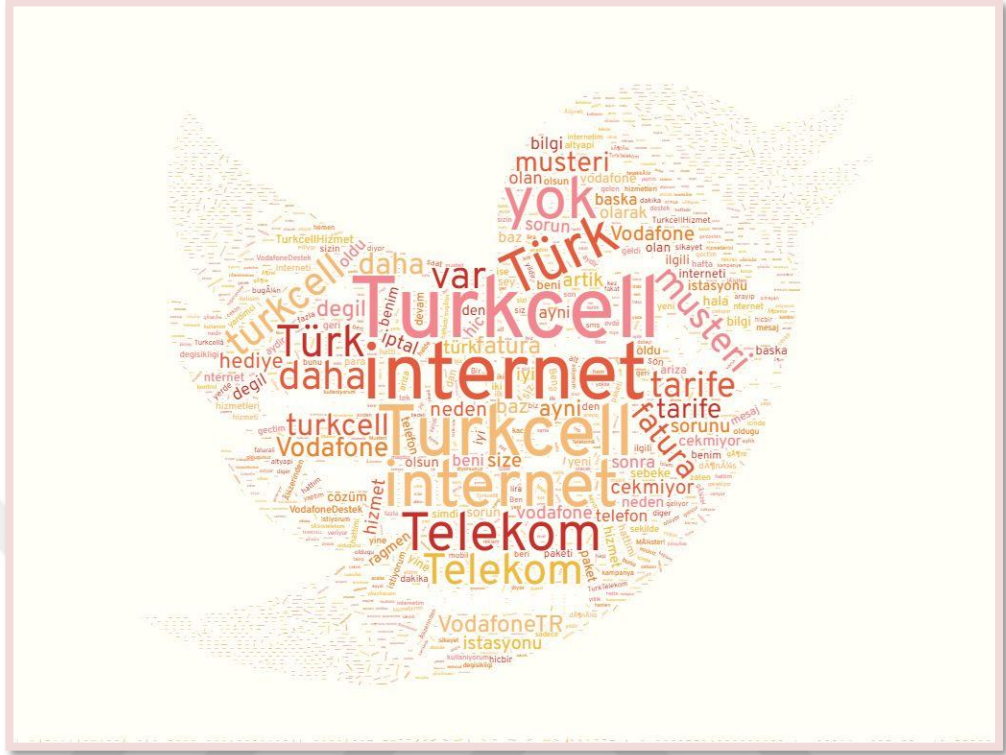


Şekil 8. Duygu Sınıf Grubu

Tablo 7. Konu Bazlı Tweet Sayıları

Alan Adı	Negatif	Pozitif	Nötr
turkcell	1437	480	599
avea_turktelekom	1379	241	462
internet	801	161	213
vodafone	504	63	147
fatura_ucretlendirme	463	29	135
müsteri hizmetleri	461	40	77
tarife	232	49	202
hiz_kota	217	116	147
GB	167	117	124
sebeke_altyapi	160	6	23
baz istasyonu	98	11	71

Tablo 7’de konu bazlı tweet sayıları verilmiştir. Bir tweette birden fazla anahtar kelime etiketlenmiştir. En çok tweetlerin operatörler için negatif duygu sınıfı üzerinden atıldığı görülmektedir. 1437 negatif tweetin Turkcell operatörüne, 1379 negatif tweetin Avea_TurkTelekom operatörüne, 504 negatif tweetin ise Vodafone operatörüne atıldığı görülmüştür. Alınan hizmetlerde en çok internet, müşteri hizmetleri ve fatura_ucretlendirme hizmetlerinden olumsuz tweet atılmıştır.



Şekil 9. Oluşturulan Derlemde Elde Edilen Kelime Bulutu

Şekil 9’ da oluşturulmuş derlemde yer alan tweetler de en çok geçen kelimelerin kullanım sıklığına göre boyutunun belirlendiği bir kelime bulutu elde edilmiştir. Örneğin turkcell-vodafone-internet-tarife-fatura gibi kelimeleri tweetler de oldukça fazla geçmektedir. Bu kelimeler veri temizleme işlemi yapılarak derlemde çıkarılabilmektedir.

2.3. Verinin Ön İşlemesi

Çalışma kapsamında elde edilen derlemin etiketlenme süreci tamamlandıktan sonra, temizlenme işlemi yapılmıştır. Twitter metni anlamsız emoji, karakter, URL, piktograflar, bayraklar vb. içermektedir. Her bir tweet metnin içerdiği anlamsız emoji, karakter, URL, piktograflar, bayraklar vs. temizlenmiştir. Temizlenme sürecinde; derlemin “word_tokenize (kelime belirteci)” işlevi ile küçük harfe dönüştürülmüş metin oluşturulmuştur. NLTK (Natural Language Toolkit- Doğal Dil Araç Seti) kitaplığından alınan bu işlev, metni tek tek sözcüklere ayırmak için kullanılmıştır. Simgeleştirme, bir cümleyi veya metni kelimeler veya alt kelimeler gibi daha küçük birimlere ayırma işlemi olarak kullanılmıştır. NLTK kitaplığı, genellikle fazla anlam taşımadıkları için metin

işleme veya NLP görevleri sırasında “corpus.stopwords (külliyat gizli sözcükler)” ile genellikle metinden kaldırılan yaygın sözcükler olan bir gizli sözcük koleksiyonu sağlamaktadır. Özel karakterler, noktalama işaretlerini veya semboller kümesini ifade eden “string.punctuation (dize noktalama)” ile tüm noktalama karakterlerinden oluşan bir diziyi temsil eden, sabitler dahil olmak üzere çeşitli dize işleme işlevlerini içeren yerleşik bir Python modülü kullanılmıştır. Makine öğrenimi ve derin öğrenme uygulamasında faydalı olan dizileri ve matrisleri işlemek için yaygın olarak kullanılan sayısal hesaplamada güçlü NumPy kitaplığından faydalanılmıştır. TensorFlow'un Keras kitaplığından alınan “tensorflow.keras.preprocessing.text.Tokenizer (tensorflow keras ön işleme metni.belirteçleyici)” sınıf, metni sayısal dizilere dönüştürme ve sözcükten dizine eşlemeler oluşturma dahil olmak üzere metin ön işleme için kullanılmıştır. “tensorflow.keras.preprocessing.sequence.pad_sequences (tensorflow keras ön işleme dizisi dolgu dizileri) işlevi ile dizileri (örn. simgeleştirilmiş cümleler) aynı uzunluğa sahip olacak şekilde doldurmak için kullanılmıştır.

2.4. Kelime Temsili (Word Embedding)

Kelimelerin bir vektör uzayındaki dağıtılmış temsilleri, öğrenme algoritmalarının benzer kelimeleri gruplayarak doğal dil işleme görevlerinde daha iyi performans elde etmesine yardımcı olur. Kelime temsillerinin en eski kullanımlarından biri; Rumelhart ve arkadaşları (Rumelhart et al., 1986), 1986 yılında yapılan çalışmasına dayanmaktadır. Daha sonralarında önemli bir başarı ile istatistiksel dil modelleme konuşma tanıma, makine çevirileri, doğal dil işleme, veri sıkıştırma gibi çeşitli yapay sinir ağı modelleri mimarilerini takip eden çalışmalar olmuştur (Bengio vd., 2003)(Repubic ve Mikolov, 2012).

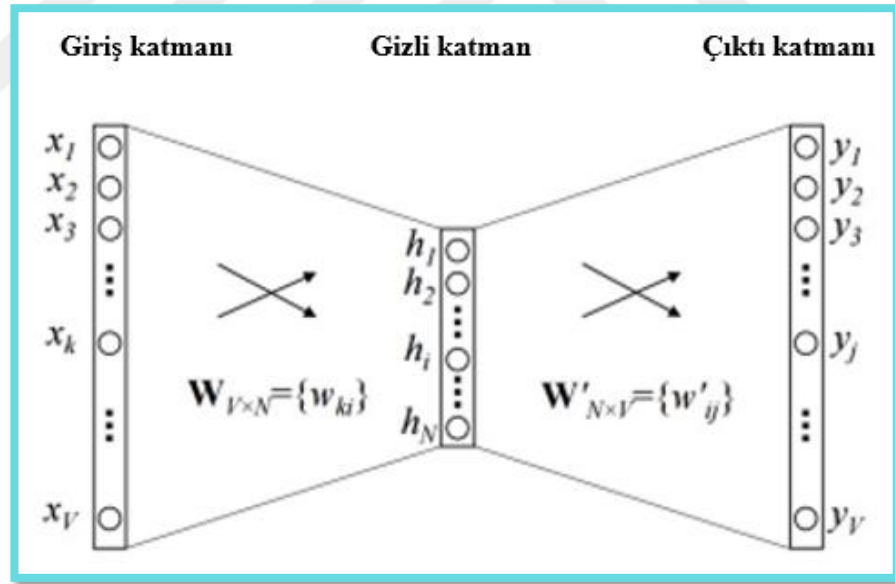
Kelime temsili, kelimelerin anlamsal özelliklerini yakalamayı hedefleyen düşük boyutlu, yoğun vektörel temsillerdir. DDİ'de dil modelleme ve özellik öğrenme tekniklerinin genel ismi olup, kelimeler gerçek sayılardan oluşan vektörlerle eşleştirilerek performans artırıcı özelliğe sahip olduğu tespit edilmiştir (Özer, 2019).

Kelimelerin temsili vektörlerinin oluşturmanın çeşitli yöntemleri bulunmaktadır. Kelime temsil modeli olarak Word2vec kullanılmıştır. Kelime dağılımının en popüler temsillerinden biridir. Bir belgedeki bir kelimenin bağlamını, anlamsal ve sözdizimsel

benzerliğini, diğer kelimelerle ilişkisini vb. yakalama yeteneğine sahip kelime vektör temsidir.

Word2Vec, sığ sinir ağı kullanarak kelime gömmelerini öğrenmek için en popüler tekniklerden biridir (Wu, 2020). Word2vec iki nöral ağ dili modelinden oluşur, Sürekli Sözcük Torbası (CBOW) ve Skip-gram. Her iki modelde de önceden tanımlanmış uzunlukta bir pencere korpus boyunca hareket ettirilir ve her adımda ağ, pencerenin içindeki kelimelerle eğitilir. CBOW modeli, çevreleyen sözcüklere dayalı olarak pencerenin ortasındaki sözcüğü tahmin etmek üzere eğitilirken, Skip-gram modeli, merkezi sözcüğü temel alarak bağlamları tahmin etmek üzere eğitilir. Yapay sinir ağı eğitildikten sonra, gizli katmandaki öğrenilen doğrusal dönüşüm kelime temsili olarak alınır (Edgar Altszyler, 2017).

CBOW Modeli, her kelimenin bağlamını girdi olarak alır ve bağlama karşılık gelen kelimeyi tahmin etmeye çalışır (Wu, 2020).



Şekil 10. Basit Bir CBOW Modeli

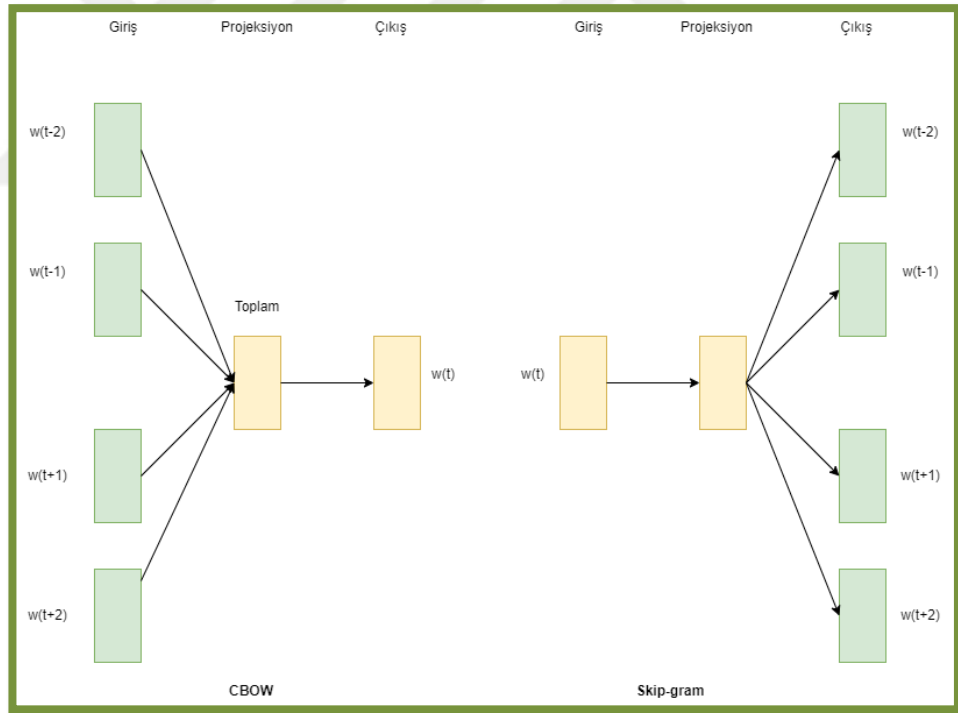
Birbirine zıt iki temel yaklaşımı sergileyen bu modellerden CBOW modeli hata gradyanının geri yayılımına bağlı olarak kelime ve çıktıyı karşılaştırıp kelimenin temsili (representation) düzeltir ve maksimize etmeye çalışmaktadır (Özer, 2019).

$$\frac{1}{V} \sum_{l=1}^V \log = p(m_l | m_{l-c/2} \dots m_{l+c/2}) \quad (2.1)$$

Büyük miktarda yapılandırılmamış metin verisinden kelimelerin yüksek kaliteli vektör temsillerini öğrenmek için etkili bir yöntem olan Skip-gram modelini tanıtan Mikolov ve ark. kelime vektörlerini öğrenmek için daha önce kullanılan sinir ağı mimarilerinin çoğundan farklı olarak, Skip-gram modelinin eğitimi yoğun matris çarpımlarını içermeyen eğitimi verimli hale getirerek optimize edilmiş tek makinelik bir uygulama geliştirmişlerdir (Mikolov vd., 2013).

Skip-gram modeli bir kelimeyi esas alarak bu kelimenin etrafında bulunması muhtemel olan diğer kelimeleri bulmaya çalışır ve temelde maksimize etmeye amaçlamaktadır (Özer, 2019).

$$\frac{1}{V} \sum_{t=1}^V \sum_{j=t-c, j+t}^{t+c} \log p(m_j | m_t) \quad (2.2)$$



Şekil 11. CBOW ve Skip-Gram Modellerinin Gösterimi

CBOW mimarisi, mevcut kelimeyi temel alarak tahmin eder. Skip-gram mevcut kelime verilen çevreleyen kelimeleri tahmin eder.

Hiper-parametre seçimi (hyper-parameter tuning) ise daha iyi sonucu elde etmek için bir makine öğrenme parametresidir. Sınıflandırma işleminde kullanılan kelime

temsilleri Word2vec CBOW yöntemi ile elde edilmiş olup sırasıyla 25, 50, 75, 100 ve 125 olmak üzere farklı vektör boyutları ile kelime temsilleri ile sınıflandırma performansları değerlendirilmiştir. Kelime temsillerinde kullanılan bir diğer hiper parametrede epoch (devir sayısı)dur. Devir sayısı tüm verinin kaç kez eğitildiğini ifade etmektedir. Sistemin devir sayısı 150 olarak belirlenmiştir.

2.5. N- gram

N-gram kavramı, Doğal Dil İşleme (NLP) ve veri biliminde yaygın olarak kullanılmaktadır. 1 boyutundaki bir n-gram aynı zamanda "unigram", 2 boyutu "bigram" ve 3 boyutu "trigram" olarak da adlandırılmaktadır. N-gramlar, n-gram modelleri adı verilen olasılıksal dil modelleri oluşturmak için kullanılabilir. N-gram modelleri, önceki N-1 kelimesini temel alarak bir kelimenin oluşumunu tahmin etmektedirler. N-gram Modeller olasılık ve istatistiksel doğal dil işleme gibi belirli dizilimlerin olasılıklarını inceleyerek modelleyen alanlarda kullanılmaktadır.

N-gram modeli, bir sonraki olayın olasılığını tahmin etmek için önceki n elemanlı sıralamanın bilindiği bir modeldir. DDİ kullanıldığında, n-1. sıradan daha önceki kelimelerle bağımsızlık varsayımı uygulanır. Kelimenin olasılığı, sadece kendinden önceki n-1 kelimeye bağlıdır (Abbasi, 2023).

Önceki n elemanlı sıralamanın olma olasılığı bilindiğinde sıradaki olayın olma olasılığını tahmin etmeye çalışmaktadır. Her metin bir n-gram seti olarak kabul edilmektedir. (örneğin $n = 1, 2, 3$) Her biri n-grama rastgele başlatılmış bir vektör olarak atanmaktadır. t metnin uzunluğu $|t|$ ile gösterilir. t metin: $|t|$ uni-gram, $|t|-1$ bi-gram ve $|t|-2$ tri-gram olmak üzere sıradan daha önceki kelimeler ile bağımsızlık varsayımı uygulamaktadır. Kelime N-gramlarını açıklamak gerekirse, unigram modeli kendisinden önce 0 kelime sırasına bağlıdır. Bigram modeli ise kendinden önceki 1 kelimeye bağlı, trigram modelinde kelime kendinden önceki son 2 kelime sırasına bağlıdır.

Bir n-gram, bir metin belgesinde sözcükleri, sayıları, sembolleri ve noktalama işaretlerini içerebilen n ardışık ögenin bir koleksiyonudur. N-gram modelleri, duygu analizi, metin sınıflandırması ve metin oluşturma gibi kelime dizilerinin alakalı olduğu birçok metin analizi uygulamasında kullanışlıdır. N-gram modelleme, metni yapılandırılmamış bir biçimden yapılandırılmış bir biçime dönüştürmek için kullanılan

birçok teknikten biridir. n-gram'a bir alternatif, word2vec gibi kelime gömme teknikleridir.

İlgili veri ayrıştırılmadan önce her veri kümesi üzerinde standart bir ön işleme gerçekleştirmektediriz. İlgili ayrıştırıcı ön işleme aşamasında tweetlerin retweet sayısı, hashtag sıklığı, ifadeler, referans URL'lerinin kullanımı ve bunları özellik olarak ele alınmaktadır. Ön işleme aşamasında, ayrıştırılacak cümleler, duruma göre seçilmektedir. Boyutları, bu nedenle sözdizimsel n-gramların boyutundaki sınırlamalara ilişkin kriterlere uygun değerlendirilir.

Metin sınıflandırmada, dil modelleme ve kötü amaçlı yazılım analizinde gibi birçok alanda kullanılmaktadır (Santos vd., 2009). Şekil 12'de N'nin 1'den 5'e kadar olan değerleri için elde edilmiş kelime seviyesinde N-gram modeli görülmektedir.

Cümle	Van Gölü Türkiye'nin en büyük gölüdür
Unigrams	{Van}, {Gölü}, {Türkiye'nin}, {en}, {büyük}, {gölüdür}
Bigrams	{Van Gölü}, {Gölü Türkiye'nin}, {Türkiye'nin en}, {en büyük}, {büyük gölüdür}
Trigrams	{Van Gölü Türkiye'nin}, {Gölü Türkiye'nin en}, {Türkiye'nin en büyük}, {en büyük gölüdür}
Four-grams	{Van Gölü Türkiye'nin en}, {Gölü Türkiye'nin en büyük}, {Türkiye'nin en büyük gölüdür}
Five-grams	{Van Gölü Türkiye'nin en büyük}, {Gölü Türkiye'nin en büyük gölüdür}

Şekil 12. 1'den 5'e Kadar Kelime Seviyesi N-Gram Örnekleri

Kaynak: (Özer, 2019)

ÜÇÜNCÜ BÖLÜM

3. LİTERATÜR ÇALIŞMASI

Wei, Choy ve Yew (2007), Malezya telekomünikasyon endüstrisinin performansına yönelik beş ön başarı faktörü, dört bilgi yönetimi (KM) stratejisi ve üç BY sürecinin algılanan önemini ve gerçek uygulama düzeyini değerlendirmektir. Kolayda örnekleme tekniğine dayalı olarak telekomünikasyon şirketleri üzerinde bir anket çalışması yapılmıştır. Veriler, eşleştirilmiş t testleri ve çoklu regresyon analizleri kullanılarak analiz edilmiştir. Sonuç olarak Malezya telekomünikasyon kuruluşlarının tüm KM ön başarı faktörlerini, stratejilerini ve sürecini önemli ve kritik olarak gördüğünü ancak uygulamada yetersiz kaldığını göstermektedir (Wei vd., 2009).

Gürsakar vd. (2014), çalışmasında NodeXL programı kullanılarak Twitter'dan elde edilen sosyal ağ verilerinin kuvvet yasası olasılık dağılımına uygun olup olmadığı test edilmiştir. Yapılan analiz sonuçları, ampirik verilerin kuvvet yasası olasılık dağılımına sahip olduğunu göstermektedir (Gürsakar vd., 2014).

Kabakçı (2015), yaptığı tez çalışmasında bilgi ve iletişim teknolojileri temelli GSM sektörü, online şikâyet yönetimini nasıl uyguladığı ve tüketicilerin online şikâyet kanallarına bakış açıları analiz edilmiştir. Yapılan analiz sonuçlarına göre, tüketici profiline sosyal ağlar ve şikâyet siteleri gibi kanalları kullanarak firmalar üzerindeki etkinliğin farkında olduğu görülmüştür. Aynı zamanda, kullanımlar ve memnuniyet yaklaşımıyla da benzerlikler gösterdiği tespit edilmiştir. Bu bağlamda, tüketiciler tercih ettikleri kitle iletişim araçlarını kullanarak ihtiyaçlarını karşılamaya ve memnuniyet sağlamaya çalışmaktadır. Yapılan analiz sonucu, iletişim teknolojilerinde yaşanan hızlı değişim ve gelişmeler, tüketicilerin ürün ve hizmet aldıkları firmalardan beklentilerini önemli ölçüde arttırdığı ve kurumların şeffaf, dürüst ve yalın bir şekilde karşılıklı bir iletişim sergileyebilmektir (Kabakçı, 2015).

Kalabalık (2016), telekomünikasyon sektöründe müşteri ayrılma analizi tahmininde, geleneksel istatistiksel tahmin yöntemlerinden farklı olarak çoklu makine öğrenmesi algoritmalarının, birleştirmeli sınıflandırma yöntemlerini mevcut tahmin etme metotlarını kullanarak ölçü doğruluğunu artırılmasını incelemiştir. Telekomünikasyon sektöründe müşteri ayrılma yönetimi sınıflandırma sonuçlarının değerlendirmesini

bagging, boosting ve random forest birleştirmeli sınıflandırma yöntemlerini kullanarak gerçekleştirmiştir. Weka yazılım aracı yardımıyla bagging, boosting ve random forest tekniklerinin performansını değerlendirmesi sağlanmış olup sonuçlar sınıflandırma doğrulukları ve diğer ölçülerde kabul edilebilir iyileşmeler elde edilmiştir. İyi bir sınıflandırma tabanı ile kullanılan birleştirmeli sınıflandırma yöntemlerinin müşteri ayrılma analizi tespitinde etkili olduğunu söylemek mümkün olduğu tespit edilmiştir (Kalabalık, 2016).

Almuqren ve Cristea (2016), Suudi Arabistan'daki Telekom şirketlerine yönelik müşteri memnuniyetini ölçmek ve müşteri kaybı oranını tahmin etmek için Arapça tweet'leri incelemektedir. Müşterilerinin Twitter'daki memnuniyetini gerçek zamanlı olarak izlemektedir (Almuqren ve Alexandra Cristea, 2016).

Hacıhasanoğlu (2017), Telekomünikasyon sektörü üzerine bir uygulama örneği ile sosyal medyada pazarlama stratejilerinin oluşturulması araştırılmıştır. Verilerin ölçülebilmesi için betimsel analizler, parametrik olmayan değerlendirmelerin yapılabilmesi için anova testi yapılmıştır. Sosyal medyanın tüketici algılarını farklılaştırma yeteneği olduğu tespit edilmiştir. Bu durum, GSM operatörlerinin aktif hizmet sunmalarının tüketiciler açısından önemli olduğunu ortaya koymaktadır. Bununla birlikte, telekomünikasyon sektöründe müşteri hizmetlerine özellikle şikâyet yönetimine daha fazla önem verilmesi ve geri bildirimlerin hızlı bir şekilde iletilmesi konularına dikkat edilmesi gerektiği de belirlenmiştir (Hacıhasanoğlu, 2017).

Albayrak (2017), 12.739 tweet içeren bir veri seti, bedelli askerlikle ilgili konuları içeren Twitter üzerinden toplanmıştır. Bu veri setinin işlenmesi ve SentiTurkNet veri setiyle karşılaştırılması için Python programlama dili ve NLTK (Natural Language Tool Kit) kütüphanesi kullanılmıştır. Eşleşen her kelimenin puan dağılımları belirlenmiştir. Türkçe paylaşımların yapıldığı bir sosyal medya ortamında, kullanıcıların duygu analizi Türkçe duygu analizi sözlüğüyle gerçekleştirilmiştir (M. Albayrak vd., 2017).

Parlar ve arkadaşları (2017), Ki-kare, Bilgi Kazanımı, Sorgu Genişletme Sıralaması ve Maksimum Entropi Modelleme sınıflandırma algoritmasını kullanarak Karınca Kolonisi Optimizasyonu olmak üzere dört öznitelik seçim yönteminin Türkiye Twitter veri seti üzerindeki performansını karşılaştırılmıştır. Öznitelik seçim

yöntemlerinin Türkçe Twitter verilerinin duygu analizi performansı üzerindeki etkileri değerlendirilmiştir (Parlar vd., 2017).

Özmen (2017), Türkiye'de bir telekomünikasyon şirketinin müşteri verilerine uygulanmıştır. Telekomünikasyon sektörü müşteri kaybı tahmini problemine özgü entegre bir model önerilmiştir. “Sınıflandırma kalitesini arttıracak indirgenmiş nitelikler elde etmek için Bulanık Kaba Küme Tabanlı FRQR (The fuzzy-rough Quick Reduct) ve Maliyet Tabanlı NSGA-II (Non-dominated Sorted Genetic Algorithm) nitelik seçim algoritmaları uygulanmıştır” (Özmen et al., 2017). Sınıflandırma sırasında, nitelik seçimi yaklaşımlarından elde edilen indirgenmiş veriler, önerilen maliyet duyarlı KKO (Karınca Kolonisi Optimizasyonu) algoritması ile sınıflandırılmıştır. Geliştirilen müşteri kaybı sınıflandırma modeli, Türkiye'nin ilk 100 Bilişim Teknolojileri (BT) şirketleri içerisinde yer alan bir şirketin müşteri verilerine uygulanmıştır. Elde edilen sonuçlar, önerilen modelin müşteri kaybını tahmin etmede 0.99 AUC değeri ile etkin bir performans sergilediğini göstermektedir (Özmen vd., 2017).

Qamar (2017), Suudi Arabistan'daki farklı telekomünikasyon şirketlerine ait İngilizce yazılmış tweetlerin duygu analizini yapılmıştır. K- en yakın komşu algoritması, Yapay Sinir Ağları (YSA), Naive Bayesian vb. gibi farklı makine öğrenme algoritmaları uygulanmıştır. Tweet'leri Öklid mesafesi ve kosinüs benzerliğine göre pozitif, negatif ve nötr olarak sınıflandırılmıştır. Özellik seçimi için CfsSubsetEvaluation ve Information Gain kullanılmıştır. CfsSubsetEvaluation'ın sonuçları, Information Gain ile elde edilenlerden daha iyi sonuç elde edilmiştir. Ayrıca kNN diğer algoritmalarından daha iyi performans göstermiştir (Mustafa Qamar vd., 2017).

Al-Smadi vd. (2017), Arabic Hotels incelemelerinin görünüm tabanlı duygu analizinin (ABSA) zorluklarını ele almak için denetimli makine öğrenimine dayalı son teknoloji yaklaşımlarının kullanılmıştır. Derin yinelenen sinir ağı (RNN) ve destek vektör makinesi (SVM) olmak üzere iki yaklaşım, sözcüksel, sözcük, sözdizimsel, biçimbilimsel ve anlamsal özelliklerle birlikte uygulanmıştır ve eğitilmiştir. Önerilen yaklaşımlar, Arabic Hotels'in incelemelerinden oluşan bir referans veri seti kullanılarak değerlendirilmiştir. Değerlendirme sonuçları, SVM yaklaşımının araştırmada incelenen görevlerde diğer derin RNN yaklaşımından daha iyi performans gösterdiği tespit edilmiştir (Al-Smadi vd., 2017)

Albayrak (2018), İngilizce ve Türkçe olmak üzere iki farklı dildeki film yorumlarından oluşan bir veri seti kullanılmıştır. Önerilen yöntemlerin farklı dillerdeki metinler üzerindeki etkisi incelenmiştir. Altı farklı sınıflandırma algoritması (NB, LR, KA, k-NN, DVM, YSA) ile çalışma kapsamında hazır hale getirilen veriler üzerinde çalışılma yapılmıştır. İngilizce film yorumları veri seti için yapılan deneylerde Word2vec kelime vektör temsil yöntemleriyle veri seti oluşturulmuştur. Sınıflandırma algoritmaları için elde edilen başarı oranlarının birbirine yakın ve daha iyi sonuçlar verdiğini tespit etmiştir. Türkçe film yorumları veri seti için elde edilen metin temsillerinin sınıflandırması, İngilizce film yorumları veri setine göre daha başarılı olduğu görülmüştür (Albayrak, 2018).

Özmen (2018), Türkiye'nin ilk 100 telekomünikasyon şirketlerinden birine müşteri bölümlenmesi uygulanmıştır. Çalışmada yer alan firmanın veri ambarından, müşteri davranışlarıyla ilgili çağrı detayları, fatura bilgisi, müşteri demografik özellikleri gibi çok miktarda veri elde edilmiştir. Parçacık Sürü Optimizasyonu (PSO) tabanlı kümeleme tekniği ve Davies-Bouldin (DB) uygunluk fonksiyonu ile müşteri bölümlenme modeli önerilmiştir ve katkı sağlayacak en etkili nitelikler belirlenmiştir (Özmen vd., 2017).

Özer (2019), yaptığı doktora çalışmasında trafik ile ilgili olayları sosyal medya mesajlarını (SMM) kullanarak uygun bir doğruluk oranında, maliyet etkin bir çözümle, geniş bir kapsama alanı içerisinde tespit edebilecek bir yöntem önerilmektedir. Trafik alanına özel yaklaşık 1,5 M etiketsiz tweetten oluşan bir derlem hazırlanarak kelime temsilleri elde edilmiştir. Sonrasında ise etiketli veriden elde edilen kelime temsilleri ve etiketsiz veriden elde edilen trafik veri setinin sınıflandırılması gerçekleştirilmiştir (Özer, 2019).

Goularas ve Kamis (2019), Twitter'da Arnavutça dilindeki mesajlar için duygu analizi yaparak farklı yaklaşımları test edilmiş ve gözden geçirilmiştir. Farklı yöntemler arasındaki sonuçları karşılaştırılmıştır. Veriler, Bag-of-Words, TF-IDF, Word2Vec ve Glove gibi bir dizi özellik çıkarma teknikleri kullanılarak metinden vektör temsiline dönüştürülmeden önce ön işleme tabi tutulmuştur. Duygu sınıflandırma tekniklerinin performansını üç ana yaklaşımdan incelemektedir. Geleneksel makine öğrenimi, sözlük tabanlı ve derin öğrenme yaklaşımıdır. Model değerlendirmesinde; doğruluk, özgüllük,

kesinlik ve F-ölçümü kriterleri ile değerlendirilmiştir. Tüm bu kriterlerle birlikte bir özellik çıkarım tekniği olarak Glove LSTM tabanlı RNN'nin en iyi sonuçları verdiği tespit edilmiştir (Kamış ve Goularas, 2019).

Smagulova ve James (2019), akademik çalışmasında LSTM sinir ağının geliştirilmesinin tarihini ve nedenlerini araştırmıştır. Mevcut LSTM yöntemleri hakkında öğretici bir anket sunarak ve anımsatıcı LSTM mimarilerindeki son gelişmelerine vurgu yapmışlardır (Smagulova ve James, 2019).

Kabaklarlı ve Işıcık (2020), Telekomünikasyon sektörü sürekli yenilik ve dijital teknolojilere içerisinde bulunduğu kalkınma süresince düşük verimlilik ve sürekli yatırım gerektirdiği gibi sorunların yaşanmasını ele alan çalışma; Telekomünikasyon sektörünün rekabete açılması ile artan teknoloji yatırımları sayesinde internet erişim ve veri hızı artarken tüketicilerin faydaları da artış göstermesi ile birlikte, Türkiye telekomünikasyon pazarında yer alan işletmelerin mobil ve internet içindeki payları incelenmiş olup Türkiye telekomünikasyon sektörünün özelleştirme öncesi ve sonrası gelişimi ile piyasanın rekabet düzeyinin araştırılması ve tüketicilerin memnuniyet düzeylerini arttırılması konusunda çalışılmıştır (Kabaklarlı ve Işıcık, 2020).

Kündüm ve arkadaşları (2020), Türkiye'deki üç büyük telekom operatörü Turkcell, Türk Telekom ve Vodafone'un müşterilerinin duygu analizini kullanarak müşteri memnuniyetini analiz etmeyi öneriyoruz. Bu amaçla, operatörlerin müşterileri tarafından hashtag'lerle bahsedilen ilgili tweet'lerin toplanması amacıyla Twitter sosyal medya platformu kullanılmaktadır. Kullanıcıların duyarlılığı, rastgele orman, destek vektör makinesi (SVM), çok katmanlı algılayıcı (MLP), k-en yakın komşular (k-NN), Naive Bayes (NB) ve karar ağacı gibi makine öğrenme algoritmaları aracılığıyla değerlendirilir. Deney sonuçları, tüm Telekom operatörleri için yüzde 80'in üzerinde doğrulukla olağanüstü sınıflandırma performansı sunmaktadır (Kündüm vd., 2020).

Özer ve arkadaşlarının (2021), sınırlı verilerle daha başarılı kestirimci sonuçlar elde etmek için dünyanın farklı yerlerinden veri alınması üzerine çapraz korelasyona dayalı bir transfer öğrenme yaklaşımı önerilmiştir. Derin öğrenmede kullanılan özel bir RNN modeli olan LSTM modeli kullanılarak dünyanın diğer bölgelerinden gelen verilerin, sınırlı tarihsel verisi olan ve bölgedeki benzer binalar için tarihsel elektrik

tüketim verisi olmayan binalarda sensör tabanlı tahmin üzerindeki etkileri araştırılmıştır. Transfer öğrenme verileri ile orijinal veriler arasındaki benzerliği bulmak için XCORR tabanlı bir yaklaşım önerilmiştir (Özer vd., 2021).

Bir ters sinyal yaklaşımı kullanarak, güç kalitesi bozulmalarının (PQD'ler) sınıflandırılması için spektrogramlarla birlikte bir evrişimli sinir ağına (CNN) ve çift yönlü uzun kısa süreli belleğe (Bi-LSTM) dayalı yeni bir derin öğrenme algoritması sunan Özer ve diğerleri (2021), önerilen yöntemin sınıflandırma başarı oranlarını artırmak amacıyla derin öğrenme tabanlı bir spektrogram kullanarak PQD olayının meydana geldiği bölgeye odaklanılmıştır. Önerilen yaklaşımda, sinyalin saf sinüs sinyaline göre zaman kayması bulunmuş ve bu zaman kaymasına göre orijinal sinyale göre bir ters sinüs dalgası üretilmiştir. Bu sinüs dalgasını orijinal sinyalle birlikte toplayarak, her iki sinyalin spektrogramları elde edildi ve daha sonra birleştirilen kırmızı/yeşil/mavi (RGB) görüntülere dönüştürülmüştür. Çalışma sonucunda, sınıflandırma CNN/Bi-LSTM aracılığıyla yapılmıştır. Bu kapsamda hem tekli hem de birleşik yapılarda 29 farklı bozulma olayı kullanılmıştır. Önerilen model bozucu olaylara uygulanmış ve %99.33 sınıflandırma doğruluğu elde edilmiştir (Özer, vd., 2021).

Usta (2021), büyük veri analizi teknolojisi ile telekomünikasyon sektöründe ayrılma eğilimi olan müşterileri tahmin eden modeli oluşturmaktır. Bu kapsamda bir telekomünikasyon şirketinden temin edilen veri kümesine Apache Spark'ın makine öğrenmesi kütüphanesi (MLlib) ile çeşitli sınıflandırma yöntemleri uygulamıştır. Lojistik Regresyon, Karar Ağaçları, Rastgele Orman, Gradyan Artırma Makineleri yöntemleri uygulanmıştır. Apache Spark'ın etkin kullanılabileceği ortaya konulmuştur (Usta, 2021).

Özer ve diğ. (2021), CNN ve LSTM transfer öğrenme ile SISAL veri setinde %100 doğruluk ve 1 eğri altı alan (AUC) puanı, Fully Connected Deep Network %92,86 doğruluk ve 0,969 AUC puanı sağlamıştır. SISA veri seti için CNN ve LSTM ile %98,73 doğruluk ve 0,988 AUC puanı elde edilmiştir. Ayrıca Lineer Diskriminant Analizi (sığ algoritma) ile %96,43'e varan doğruluğa ulaşılmıştır (Özer, vd., 2021).

Çavuş ve Eğilmez (2021), sosyal medya kullanımı ve kurumsal itibar arasındaki ilişkiye dayanılarak, Türkiye'deki telekomünikasyon şirketleri Turkcell, Türk Telekom ve Vodafone TR'nin sosyal medya aracı Twitter paylaşımları üzerinden kurumsal itibar

yönetimindeki eğilimlerinin analizi amaçlanmıştır. İçerik analizi yöntemiyle, referans alınan kurumsal itibar ölçeğinin altı ana başlığı altında sayısallaştırılan sonuçlara göre; üç şirketin de paylaşımlarının itibar aktarımına ilişkin dengeli bir dağılıma sahip olmadığı görülmüştür. Buna göre kurumsal itibar aktarımına ait bileşenlerin ayrı birer kazanım unsuru olarak bütün halinde ele alınması gerektiği tavsiye edilmiştir ve gelecekteki çalışmalar için öneriler paylaşılmıştır (Çavuş ve Eğilmez, 2021).

Şahiner ve Özer, Gökçen (2021), çalışmasında genişletilmiş derlem ile derin sinir ağları kullanılarak Türkçe doğal dil öğrenimine etkisi mevcut araştırma kapsamında incelenmiş ve sonuçlar analiz edilmiştir. OffenseEval veri seti üzerinde saldırgan dil tespiti yapılmış olup bu doğrultuda bir derlem hazırlanmıştır. Çalışmada makine öğrenmesi modeli olarak Uzun Kısa Dönemli Bellek (LSTM) ve Çift Yönlü Uzun Kısa Dönemli Bellek (BiLSTM) ağları kullanılmıştır. Bu derin sinir ağlarının sınıflandırma performansları, doğruluk (accuracy), duyarlılık (recall) ve kesinlik (precision) ve F-skor olarak değerlendirilmiştir (Yılmaz vd., 2021).

Özsoy (2021), çalışmasında Türkiye’de faaliyette bulunan ve pazarın abone sayısı ve gelir bakımından neredeyse tamamını temsil eden üç telekomünikasyon işletmesine yöneltilen tüketici şikayetleri dikkate alınmış olup “Şikayetvar” isimli web sitesinden toplanan 600 şikâyet incelenmiş, şikayetler ana kategoriler altında tanımlanmaya çalışılmıştır. Sonrasında her bir kategoriye örnek teşkil edecek şikâyetlere yer verilmiştir (Özsoy, 2021).

Kumaş (2021), Türkçe Twitter verileri kullanılarak duygu analizi yapılmıştır. Twitter verilerine metin madenciliği yöntemlerinden olan Naive Bayes, KNN, Destek Vektör Makinesi, Lojistik Regresyon ve Karar Ağacı sınıflandırma algoritmaları uygulanarak veriler analiz edilmiştir ve pozitif ve negatif olarak sınıflandırılması yapılmıştır. Sınıflandırma sonuçları f_1 skoru ile değerlendirilmiştir (Kumaş, 2021).

Kıldan (2021), Türkiye’de faaliyet gösteren üç büyük telekomünikasyon şirketinin sosyal medyadan yürüttükleri şikâyet yönetimi analiz edilmiştir. Şirketlerin sosyal medyadan müşteri hizmetleri ile ilgili aldıkları şikayetlerle sosyal medyadaki şikayet yönetimi performansı arasındaki ilişkisi incelenmiştir (Kıldan, 2021).

Uyanık (2021), cell2cell telekomünikasyon firması tarafından oluşturulan veri kümesi kullanılmıştır. Veri setinde sınıf dengesizliği olduğu için SMOTE (Synthetic Minority Oversampling Technique) ve ADASYN (Adaptive Synthetic Sampling Method) tekniği ile örnekleme yapılmıştır. Lojistik Regresyon, Karar Ağacı, Yapay Sinir Ağları, Torbalama (Bagging) ve Artırma (Boosting) sınıflandırma modelleri kullanılmıştır. Performans ölçütlerine göre en iyi tahmin modeli, ADASYN örnekleme yöntemi kullanılan model ve sınıflandırma yöntemi olarak en iyi sonucu veren LightGBM (Light Gradient Boosting Machine) tekniği olarak bulunmuştur (Uyanık, 2021).

El Houby ve Mohamed (2021), Sosyal medya aracılığıyla kullanılan Arapça dili, argo veya konuşma dilini köklendirme aracı Modern Standart Arapça (MSA) üzerine kurulanmıştır. Duyguları kategorize etmek için bir duygu analizi modeli uygulanmıştır. Model, çok sınıflı ve çok etiketli bir sınıflandırma problemidir. Model, önceden eğitilmiş kelime gömme vektörlerinin üzerinde eğitilmiş yeni birçok katmanlı çift yönlü uzun kısa süreli bellek (BiLSTM) uygulanmış ve en gelişmiş performans geliştirmesini sağlamıştır. Bu yaklaşım, Destek Vektör Makineleri (SVM), rastgele orman (RF) ve tamamen bağlı sinir ağları kullanılarak aynı görevlerde geliştirilen diğer modellerle karşılaştırılmıştır. Önerilen model, bu görev için elde edilen en iyi sonuçların üzerinde bir performans artışı sağlanmıştır (Houby ve Mohamed, 2021).

Daouadi (2021), Twitter'dan doğru bir şekilde tespit etmek için bağlamsal Word2vec, WordNet ve Deep Forest'ı birleştiren bir konu sınıflandırma yaklaşımı olan Anlamsal Derin Orman (SDF) önermiştir. Ayrıca, en iyi performansı elde etmek için Tweet konusu sınıflandırma için SDF parametrelerinde ince ayar yapmak üzere kapsamlı parametre duyarlılık analizi yapılmıştır. Standart değerlendirme senaryolarıyla üç kıyaslama veri kümesi üzerinde deneyleri yapılmıştır. Deneysel sonuçlar neticesinde, önerilen bağlamsal word2vec modelleri, tweet konu sınıflandırması için başarılı bir şekilde kullanılabilir ve mevcut en gelişmiş yerleştirme modelinden daha iyi performans göstermiştir. Önerilen SDF, tweet konu sınıflandırmasının doğruluğunu geliştirir ve mevcut en gelişmiş sınıflandırma yaklaşımlarından daha iyi performans göstermiştir (Daouadi vd., 2021).

Salur ve Aydın (2022), Twitter veri kümeleri üzerinde yapılan duygu sınıflandırma çalışmasında metin temsilleri oluşturulmadan önce yapılan ön işleme

aşamasında URL, noktalama işareti ve emoji gibi içeriklerden arındırılmış olan veri setinden çıkarılan gürültü verilerinden özellik çıkarımına odaklanmıştır. Gürültü olarak silinen içeriklerden manuel olarak çıkarılan özelliklerin ve derin öğrenme yöntemlerinden elde edilen derin özelliklerin birleştirilmesine dayandırılarak analiz yapılmıştır. Yapılan deneylerle önerilen modelin sınıflandırma performansının önceki çalışmalardan daha iyi olduğu tespit edilmiştir (Salur ve Aydın, 2021).



DÖRDÜNCÜ BÖLÜM

4. METEDOLOJİ

4.1. Yapay Zekâ (Artificial Intelligence)

Yapay Zekâ (Artificial Intelligence), sohbeti Alan Turing'in 1950 tarihli "Makineler düşünebilir mi? " adlı çalışmasıyla başlamıştır (Turing, 1950).

Yapay zekâ öğrenmeden algılamaya birçok alt alanları kapsamaktadır. Problem çözmeyi sağlamak için bilgisayar bilimi ile sağlam veri kümelerini birleştiren bir alandır. Yapay zekânın erken başarılı bir uygulaması olan uzman sistemler, bir insanın karar verme sürecini kopyalamayı ve otomatik akıl yürütme amaçlamıştır.

Günümüzde AI ile birlikte sıklıkla bahsedilen makine öğrenimi ve derin öğrenmenin alt alanlarını içermektedir. Bu disiplinler, tipik olarak girdi verilerine dayalı olarak tahminler veya sınıflandırmalar yapan yapay zekâ algoritmalarından oluşmaktadır. Makine öğrenimi, bazı uzman sistemlerin kalitesinin iyileştirilmesini ve oluşturulmasını kolaylaştırmaktadır. Yeni koşullara uyum sağlamak için makine öğrenimi örüntüleri saptamak ve tahminlerde bulunmaktadır.

İnsan yaklaşımı	İdeal yaklaşım:
• İnsan gibi düşünen sistemler • İnsan gibi davranan sistemler	• Rasyonel düşünen sistemler • Rasyonel hareket eden sistemler

Yapay zekâ, insan davranışlarının zekâ olarak adlandırılan davranışların (akıllı davranışların) makina tarafından da yapılmasıdır. Yapay zekânın insan aklının nasıl çalıştığını gösteren bir yapı olduğunu da söylenebilir. Yapay zekânın amacı, makinaları daha akıllı hale getirmek, zekânın ne olduğunu anlamak ve makinaları daha faydalı hâle getirmektir (Tektaş vd., 2002).

AI, insanların bilgileri entegre eder. Verileri analiz ederken ve ortaya çıkan iç görüleri karar vermektedir. Yeniden düşünmeyi sağlayan geniş kapsamlı bir araçtır ve şimdiden hayatın her alanını dönüştürmektedir.

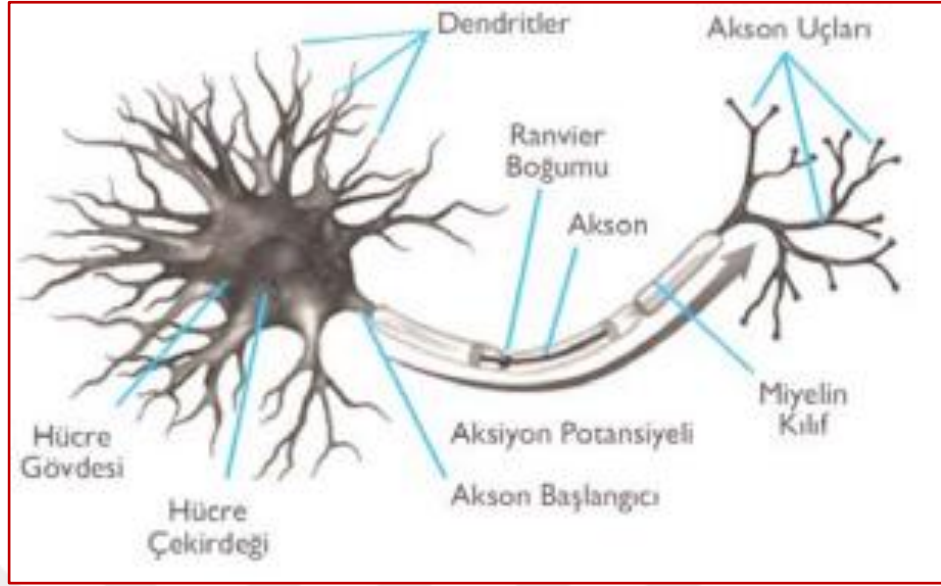
Bir diğer ifade ile yapay zekâ, insanın zekasını bilgisayar aracılığıyla taklit etmesi makinaların, öğrenmesini sağlamaktadır. İnsan beynin modellenmesiyle ve insanın düşünme modelini temel alarak birçok yöntemlerde kullanılmasını sağlamaktadır.

4.1.1.Yapay Sinir Ağları (Artificial Intelligence Networks)

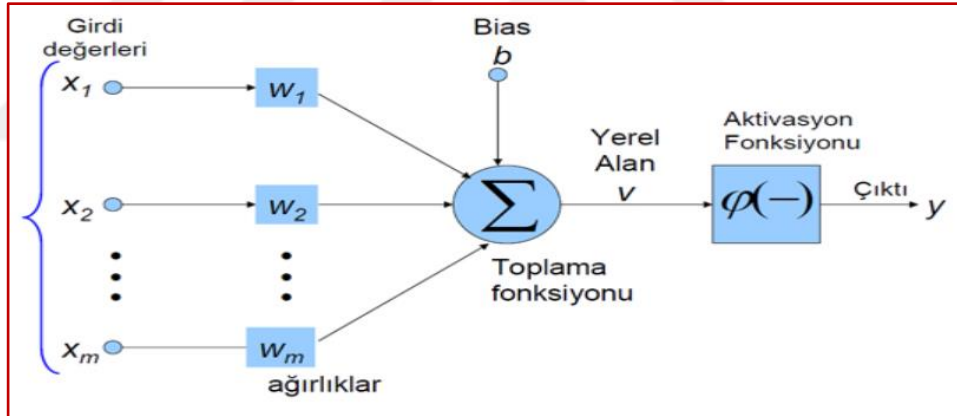
Yapay sinir ağları (YSA), insan beyninin nöronlardan oluşan yapısını ve öğrenme yöntemlerini inceler. 19.yüzyıldaki psikolog ve nöropsilogların insan beynini anlamaya çalışmaları YSA' ların temelini oluşturmaktadır (Tektaş vd., 2002). Bugün kullandığımız algılayıcıya yönelik ilk adım 1943'te McCulloch ve Pitts tarafından biyolojik bir nöronun işlevselliğini taklit ederek temeller ortaya atılmıştır.

YSA, insan beyninden esinlenerek biyolojik nöronların birbirlerine sinyal verme şeklini taklit etmektedir. YSA, bir giriş katmanı, bir veya daha fazla gizli katman ve bir çıkış katmanı içeren düğüm katmanlarından oluşmaktadır. Her düğüm veya yapay nöron, diğerine bağlanır ve ilişkili bir ağırlık ve eşikçe sahiptir. Herhangi bir düğümün çıktısı belirtilen eşik değerinin üzerindeyse, o düğüm etkinleştirilerek ağına bir sonraki katmanına veri gönderilmektedir. Aksi takdirde, ağına bir sonraki katmanına hiçbir veri iletilmemektedir.

n yapay sinir ağı (veya basitçe sinir ağı), bir girdi nöron katmanından (veya düğümler, birimler), bir veya iki (hatta üç) gizli nöron katmanından ve son bir sinir katmanından oluşur (Wang, 2003).

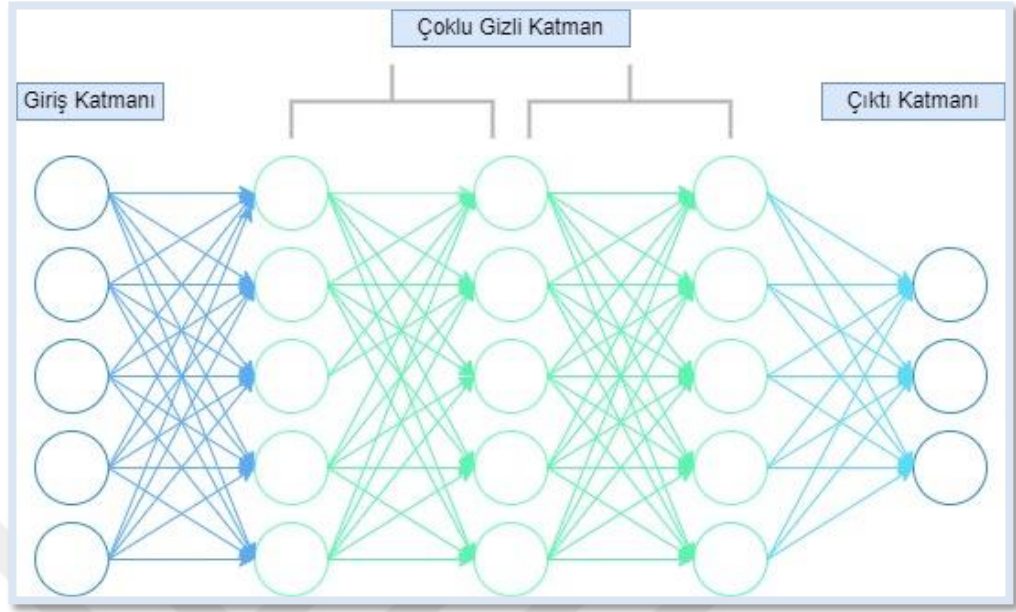


Şekil 13. Biyolojik Sinir Ağı Hücresi



Şekil 14. Yapay Sinir Hücresi Yapısı

Bir yapay sinir ağı (veya basitçe sinir ağı), bir girdiden, nöron katmanından, bir veya iki (hatta üç) gizli katmandan oluşan nöronlar ve bir çıkış katmanı içeren düğüm katmanlarından oluşmaktadır. Her düğüm veya yapay nöron, diğerine bağlanır ve ilişkili bir ağırlık ve eşik sahiptir. Herhangi bir düğümün çıkışı belirtilen eşik değerinin üzerindeyse, o düğüm etkinleştirilerek ağın bir sonraki katmanına veri gönderilir. Aksi takdirde, ağın bir sonraki katmanına hiçbir veri iletilmez.



Şekil 15. Basit Sinir Ağı Modeli

Yapay sinir ağları, beynin çok basit bir nöron modelinin benzetilmesidir. Bu şekilde elde edilen ağ ile öğrenme olayı gerçekleştirilir. Bir diğer ifade ile güçlü bir hesaplama zekâsı tekniği olarak, YSA mühendislik ve bilgisayar bilimi gibi çeşitli alanlarda uygulanmaktadır. Bu teknik, ilgili teorilerin eksikliği veya problemlerin ve sistemlerin mevcut anlayışının sınırlamaları nedeniyle analitik bir çözümün kolayca elde edilemediği karmaşık mühendislik ve bilgisayar bilimi problemlerinin (veya sistemlerinin) modellenmesi ve tahmininde kullanım için uygun bir özelliktir. Diğer birçok avantajın yanı sıra, YSA, çoklu girdi değişkenleri ve çoklu çıktı değişkenleri arasında matematiksel ilişkiler kurabilmektedir (Fang vd., 2016).

YSA, bilgisayar ortamındaki sorunları çözmek için bir dizi güçlü yeni teknik sağlamaktadır. Desen tanıma, veri analizi ve kontrolü, yüksek işlem hızları ve öğrenme yeteneği dâhil olmak üzere birçok önemli özelliğe sahiptirler. Geçmişten bugüne kadar birçok alanda YSA kullanılmaktadır.



Şekil 16. Yapay Sinir Ağı Kullanım Alanları

Kaynak: (Korkmaz ve Tektaş, 2022)

Şekil 16’da gösterildiği gibi YSA birçok farklı alanda kullanılmaktadır. Ek olarak; robotik sistemlerde, el yazısı, karakter tanıma, yüz tanıma, doğal dil işleme, görüntü işleme vb. veri madenciliği alanlarında da kullanılmaktadır.

4.2. Makine Öğrenmesi (Machine Learning)

Makine öğrenmesi bilgisayara öğrenme yeteneği kazandıran yapay zekâ yöntemidir. "Makine öğrenmesi" terimini ilk kez IBM'in içinden biri olan Arthur Samuel'in, dama oyunu ile ilgili olan araştırmasında kullanıldığı bilinmektedir. Uzman

dama oyuncusu Robert Nealey, oyunu 1962 yılında bir IBM 7094 bilgisayarında oynamış ve bilgisayara karşı kaybetmiştir (Gupta ve Jha, 2015).

Makine öğrenmesi, büyüyen veri birimi alanının önemli bir bileşenidir. İstatistiksel yöntemler kullanılarak, algoritmalar; sınıflandırmalar veya tahminler yapmak üzere eğitilir ve veri madenciliği projelerinde temel iç görüleri ortaya çıkarmaktadır.

Makine öğrenmesi, veri bilimi ve veri madenciliği gibi yapay zekânın alanları, ham verinin bilgiye dönüşüm sürecine katkıda bulunurlar (Tektaş vd., 2019). Çevreleyen ortamdan öğrenerek insan zekâsını taklit etmek için tasarlanmış, gelişen bir hesaplama algoritmaları dalıdır.

Makine öğrenimi, genel olarak, performansı iyileştirmek veya doğru tahminler yapmak için deneyim kullanan hesaplama yöntemleri olarak tanımlanabilir. Verimli ve doğru tahmin algoritmaları tasarlamayı içerir. Bilgisayar biliminin diğer alanlarında olduğu gibi, kalitenin bazı kritik ölçütleridir.

Bu algoritmalarından biri zaman ve mekân karmaşıklığıdır. Daha genel olarak, bir algoritma için teorik öğrenme garantileri, kavramın karmaşıklığına dikkate alınan sınıflara ve eğitim örneğinin boyutuna bağlıdır. Bir öğrenme algoritmasının başarısı kullanılan verilere bağlı olduğundan, makine öğrenme doğası gereği veri analizi ve istatistik ile ilgilidir.

- Metin veya belge sınıflandırması, ör. spam algılama
- Doğal dil işleme, ör. morfolojik analiz, konuşma parçası etiketleme dahil olmak üzere çeşitli uygulamalarda başarıyla uygulanmıştır.
- İstatistiksel ayrıştırma, adlandırılmış varlık tanıma
- Konuşma tanıma, konuşma sentezi, konuşmacı doğrulama
- Optik karakter tanıma (OCR)
- Hesaplamalı biyoloji uygulamaları, örneğin; protein fonksiyonu veya yapılandırılmış tahmin
- Bilgisayarla görme görevleri, ör. görüntü tanıma, yüz tanıma
- Dolandırıcılık tespiti (kredi kartı, telefon) ve ağ ihlali;
- Oyunlar, ör. satranç, tavla
- Yardımsız araç kontrolü (robotlar, navigasyon)

- Tıbbi teşhis
- Öneri sistemleri, arama motorları, bilgi çıkarma sistemleri (Mohri vd., 2018).



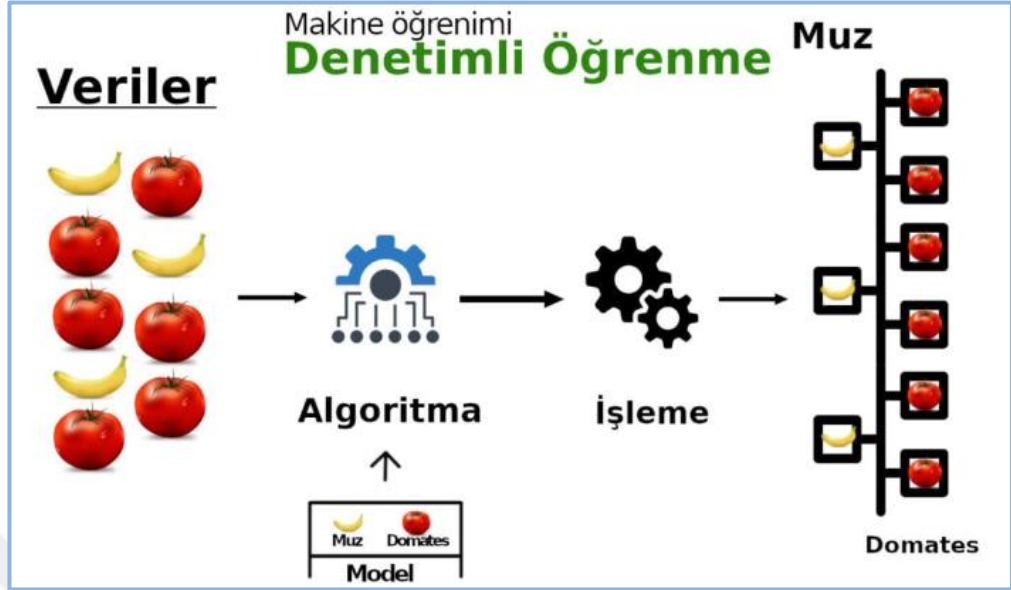
Şekil 17. Makine Öğrenmesi Şeması

4.2.1. Makine Öğrenimi Türleri

Ham veriler üzerinde insan etkisinin olup olmamasına göre tanımlanan birçok türde makine öğrenimi modeli vardır. Temel olarak dört başlıkta yer almaktadır.

4.2.1.1. Denetimli Öğrenme (Supervised Learning)

Denetimli öğrenme yönteminde, bir uzman tarafından oluşturulan örnek girdiler ve buna karşılık beklenen çıktılar modele sunulur. Modelin eğitim süreci, eğitim verilerinde istenen doğruluk seviyesine ulaşana kadar devam eder (Avuçlu ve Elen, 2020).



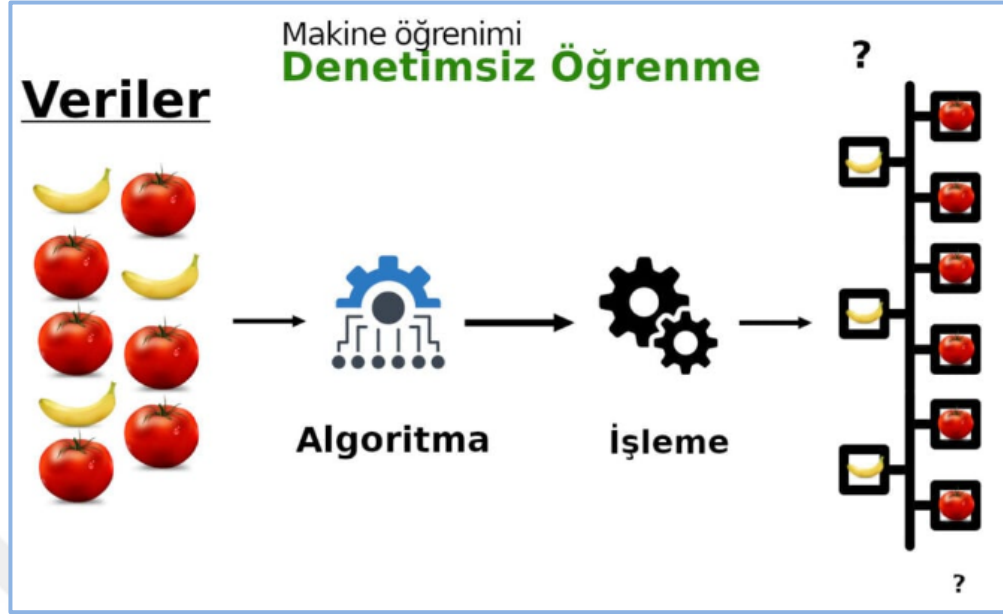
Şekil 18. Denetimli Öğrenme

Kaynak: (Kılıç, 2021)

Denetimli makine öğrenimi olarak da bilinen denetimli öğrenme, verileri sınıflandırmak veya sonuçları doğru bir şekilde tahmin etmek için algoritmaları eğitmek üzere etiketli veri kümelerinin kullanılmasıyla tanımlanır. Girdi verileri modele beslenirken, model uygun şekilde takılana kadar ağırlıklarını ayarlamaktadır. Bu, modelin gereğinden fazla veya az uydurmayı önlemesini sağlamak için çapraz doğrulama sürecinin bir parçası olarak gerçekleştirmektedir. Kullanılan veri kümesi, algoritmanın performansının ne kadar doğru olduğunu görmesine izin vermek için kullanıcılar tarafından önceden etiketlenmiş ve sınıflandırılmıştır.

4.2.1.2. Denetimsiz Öğrenme (Unsupervised Learning)

Denetimsiz öğrenme yönteminde, modelin öğrenme sürecinde belirlenen örnek girdiler sunulur fakat herhangi bir beklenen çıktı(etiket) verilmez. Bu sebeple, girdi verisinin hangi sınıfa ait belirsizdir ve girdi parametreleri arasındaki ilişkileri kendi kendine öğrenmesi beklenir. Bu öğrenme yönteminde, ele alınan problemlerin büyük bir çoğunluğu kümeleme üzerinedir (Ayuçlu ve Elen, 2020).



Şekil 19. Denetimsiz Öğrenme

Kaynak: (Kılıç, 2021)

Bu algoritmalar, insan müdahalesine ihtiyaç duymadan gizli kalıpları veya veri gruplarını keşfeder. Bu yöntemin bilgilerdeki benzerlikleri ve farklılıkları keşfetme yeteneği, onu keşif amaçlı veri analizi, çapraz satış stratejileri, müşteri segmentasyonu ve görüntü ve örüntü tanıma için ideal kılar.

Denetimsiz öğrenme yöntemleri, istenen bilgileri ayıklamak için herhangi bir talimat olmaksızın söz konusu verileri sınıflandırır. Denetimsiz öğrenme geliştikçe sonuçları iyileştirmektedir (Güldal, 2021).

4.2.1.3. Yarı Denetimli Öğrenme (Semi - Supervised Learning)

Yarı denetimli öğrenme yöntemi ise denetimli ve denetimsiz öğrenme yöntemlerinin arasındadır. Yöntemde eğitim için hem etiketli hem de etiketlenmemiş verileri kullanmaktadır. Bu yöntemi kullanan sistemler öğrenme doğruluğunu büyük bir oranla iyileştirebilir. Genellikle, bu yöntemde elde edilen etiketli veriyi eğitmek /ondan öğrenmek için yetenekli ve ilgili kaynaklar gerektirdiğinde seçilir. Aksi takdirde, etiketlenmemiş verilerin toplanması genellikle ek kaynaklar gerektirmez (Avuçlu ve Elen, 2020).

Veri kümesi, algoritmaya bağımsız sonuçlar çıkarma yolunda rehberlik eden yapılandırılmış ve yapılandırılmamış verileri içerir. İki veri türünün tek bir eğitim veri

kümesinde birleşimi, makine öğrenimi algoritmalarının etiketlenmemiş verileri etiketlemeyi öğrenmesine olanak tanır. Yarı denetimli öğrenme, denetimli ve denetimsiz öğrenme arasında mutlu bir ortam sunar. Eğitim sırasında, daha büyük, etiketlenmemiş bir veri kümesinden sınıflandırma ve özellik çıkarımına rehberlik etmek için daha küçük etiketli bir veri kümesi kullanır. Yarı denetimli öğrenme, denetimli bir öğrenme algoritması için yeterli etiketli veriye sahip olmama sorununu çözebilir. Yeterli veriyi etiketlemenin çok maliyetli olması durumunda da yardımcı olur.

4.2.1.4. Takviyeli Öğrenme (Reinforcement Learning)

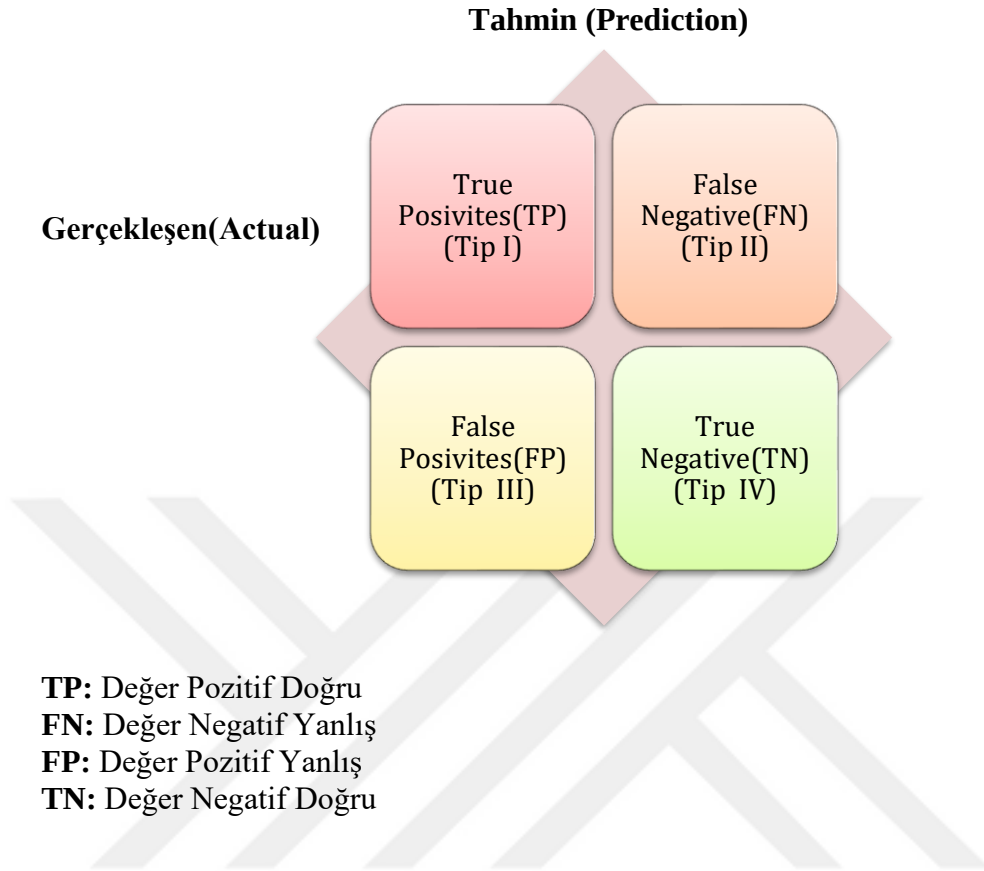
Veri kümesi, kendi deneyimlerinden deneme yanılma yoluyla öğrenmesi için algoritmaya geri bildirim sunan bir "ödülleri/cezalar" sistemi kullanır.

Takviyeli öğrenme yönteminde, eylemler üreterek çevresiyle etkileşime geçen olumlu veya olumsuz durumları keşfeden bir yöntemdir. Bu yöntemde bir danışman/egitici bulunur fakat denetimli öğrenme modellerindeki gibi sisteme çok detay vermez veya müdahale etmez. Bunun yerine, öğrenen model bir karar verdiğinde, kararın doğru olduğu durumlar için sistemi ödüllendirir ve yanlış olduğunda ise cezalandırılır (Ayuçlu ve Elen, 2020).

4.3. Makine Öğrenmesi Ölçekleri

4.3.1. Accuracy (Doğruluk)

Doğruluk, model içindeki doğru tahminlerin oranının bir ölçüsüdür.



TP ve TN modelin doğru olarak tahminlendiği, FP ve FN ise modelin yanlış olarak tahminlendiği alanlardır (Öğündür, 2019).

Doğruluk dört koşulda değerlendirilir.

- Ölçümler hem doğru hem kesin
- Doğru ama kesin değil
- Kesin ama doğru değil
- Ne kesin ne de doğru olmayabilir

Matematiksel gösterimi;

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

4.3.2. Sensitivity (Duyarlılık)

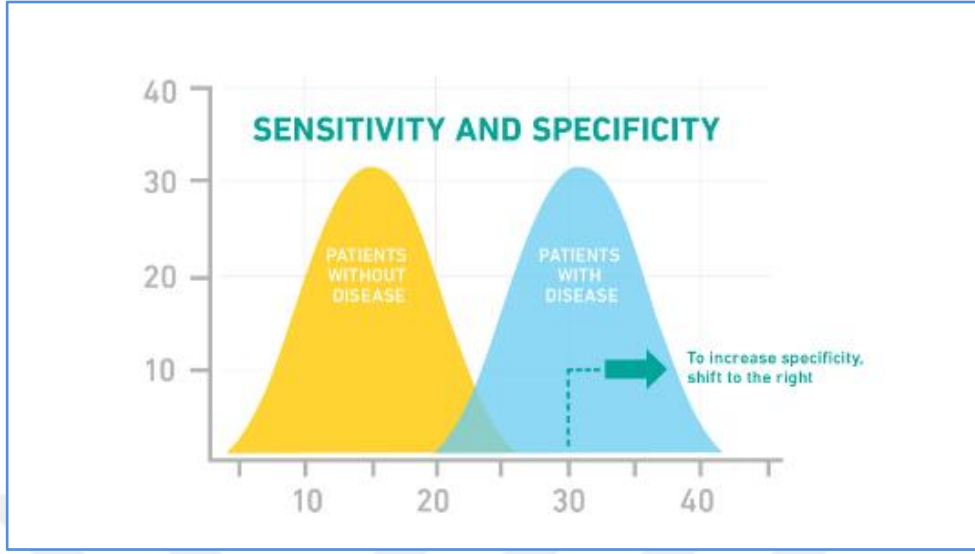
Bir testin hassasiyeti aynı zamanda gerçek pozitif oranı (TPR) olarak da adlandırılmaktadır ve söz konusu testi kullanarak pozitif sonuç veren gerçekten pozitif olan değerlerin oranıdır. Örneğin, bir paneldeki tüm pozitif örnekleri doğru şekilde tanımlayan bir test çok hassastır. Paneldeki pozitif örneklerin yalnızca %60' ını saptayan başka bir testin, pozitifleri kaçırdığı ve daha yüksek yanlış negatif oranı (FNR) verdiği için daha düşük duyarlılığa sahip olduğu kabul edilir. Tip II hatalar olarak da adlandırılan yanlış negatifler, yanlış bir sıfır hipotezinin reddedilmemesidir (boş hipotez, değerinin negatif olduğu anlamına gelir) (Steward, 2019).

$$Sensitivity = \frac{TP}{TP + FN} \quad (4.2)$$

4.3.3. Specificity (Belirginlik)

Gerçek negatif oran (TNR) olarak da adlandırılan bir testin özgüllüğü, gerçekten negatif olan ve söz konusu testi kullanarak negatif sonuç veren değerlerin oranıdır. Örneğin, tüm sağlıklı insanları belirli bir hastalık için negatif olarak tanımlayan bir test çok spesifiktir. Sağlıklı insanların %30'unun hastalığa sahip olduğunu yanlış bir şekilde tanımlayan başka bir test, daha yüksek yanlış pozitif oranına (FPR) sahip olarak daha az spesifik olarak kabul edilecektir. Tip I hatalar olarak da adlandırılan yanlış pozitifler, gerçek bir sıfır hipotezinin reddedilmesidir (sıfır hipotez, örneğin negatif olduğu anlamına gelir) (Steward, 2019).

$$Specificity = \frac{TN}{TN + FP} \quad (4.3)$$



Şekil 20. Duyarlılık ve Belirginlik

4.3.4. Precision (Hassaslık)

Hassaslık (Precision), değeri pozitif olarak tahminlenmesi gereken değerlendirmelerin ne kadarının pozitif olarak tahmin edildiğini gösteren bir ölçektir (Göçgün ve Onan, 2021).

$$Precision = \frac{TP}{TP + FP} \quad (4.4)$$

4.3.5. F-Score

F1 Score değeri, Kesinlik (Precision) ve Duyarlılık (Recall) değerlerinin harmonik ortalamasını göstermektedir. Basit bir ortalama yerine harmonik ortalama olmasının sebebi ise uç durumları da göz ardı edilmemesi gerektiğidir. Eğer basit bir ortalama hesaplaması olsaydı Precision değeri 1 ve Recall değeri 0 olan bir modelin F1 Score'u 0.5 olarak olacaktır ve bu yanıltıcı olacaktır (Öğündür, 2019).

$$Fscore = \frac{2}{\frac{1}{TPR} + \frac{1}{PPV}} \quad (4.5)$$

4.4. Derin Öğrenme (Deep Learning)

“Derin öğrenme” kavramı ilk olarak Aizenberg ve diğerleri tarafından 2000 yılında ortaya atılmıştır (Aizenberg vd., 2013).

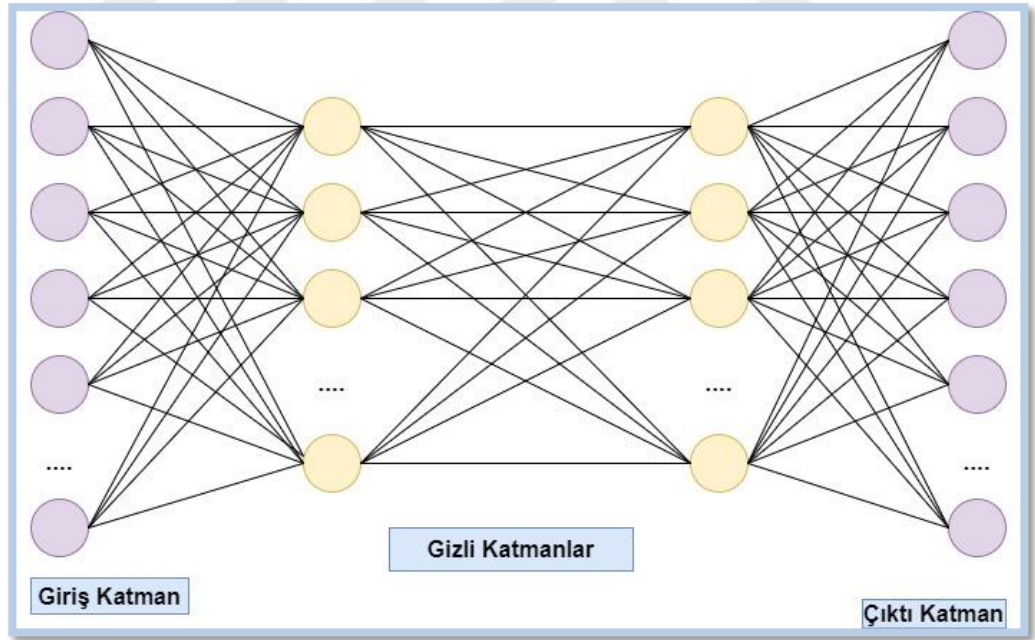
Daha sonra 2006 yılında Hinton, Yüksek boyutlu girdi vektörlerini yeniden oluşturmak için küçük bir merkezi katmana sahip çok katmanlı bir sinir ağı eğiterek yüksek boyutlu veriler düşük boyutlu kodlara dönüştürülebildiğini ve Gradyan inişin bu tür "otomatik kodlayıcı" ağlarda ağırlıkların ince ayarını yapmak için kullanılmıştır. Bir girdinin boyutsallığını azaltmak için derin bir sinir ağı (tipik olarak 3+ gizli katman derin olarak kabul edilir) kullanılmıştır. Derin ağları eğitmenin, etiketli eğitim verileri, bilgisayar hızı, yerel olmayan optimumlar (tüm bunların daha sonra yanlış olduğu anlaşıldı) nedeniyle başlangıçta çok zor olduğu düşünülmüştür. Makalede açıklanan azaltma teknikleri, temel bileşen analizi (PCA) ve lojistik PCA ile karşılaştırılır. Bu makalenin sonuçları, boyut azaltma, özellik öğrenme ve sinir ağı kullanımında inanılmaz bir atılım göstermiştir (Hinton ve Salakhutdinov, 2006). Krizhevsky, Sutskever ve Hinton 2012’de GPU kullanılarak ön eğitim olmadan eğitilen mimariler tasarlamışlardır (Marcus vd., 2017).

Derin öğrenme, bilgisayarların etiketlenmemiş verilerdeki kalıpları ve benzerlikleri otomatik olarak keşfettiği denetim olmadan gerçekleştirebilmektedir. Bu yöntemle belirli bir çıktı beklenmez ve bu genellikle verilerdeki özellikleri tespit etmek, boyutlarını küçültmek veya benzer verileri kümelemek için bir keşif aracı olarak kullanılmaktadır (Valletta vd., 2017). İkinci olarak, denetimli eğitimle öğrenme de yapılabilir. Etiketleri örneklerle ilişkilendirmek üzere eğitebilmeleri için önce bilgisayarlara hedef nesnelerle etiketlenmiş bir veri kümesi verilir. Daha sonra bu nesneleri diğer veri kümelerinde tanıyabilir ve tanımlayabilirler. Derin öğrenme yöntemlerinin hızla yaygınlaşmasına yol açan şey, karmaşık, yüksek boyutlu verilerdeki özellikleri yüksek tahmin doğruluğuyla otomatik olarak algılama yeteneğidir (Lecun vd., 2015a). Derin öğrenme, birden fazla soyutlama düzeyine sahip verilerin temsillerini öğrenmek için birden çok işlem katmanından oluşan hesaplama modellerine olanak tanır. Bu yöntemler, konuşma tanıma, görsel nesne tanıma, nesne algılama ve ilaç keşfi ve genomik gibi diğer birçok alanda en son teknolojiyi önemli ölçüde geliştirdi. Derin öğrenme, bir makinenin önceki katmandaki temsilden her katmandaki temsili hesaplamak

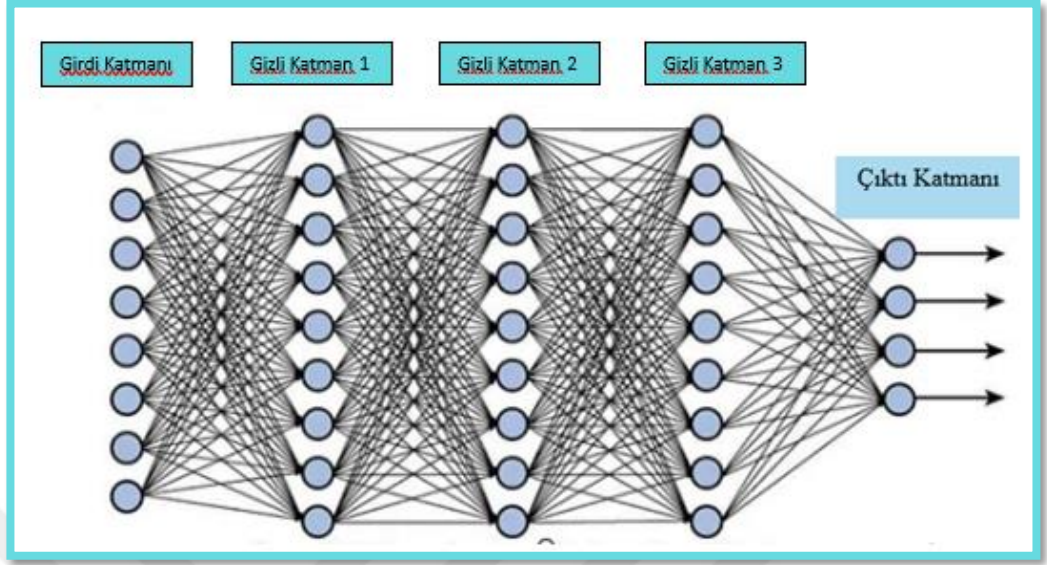
için kullanılan dâhili parametrelerini nasıl değıştirmesi gerektiğini belirtmek için geri yayılım algoritmasını kullanarak büyük veri kümelerindeki karmaşık yapıyı keşfeder (Lecun vd., 2015b).

Derin öğrenme literatüründeki sinir ağıları tipik olarak bir dizi girdi biriminden oluşur. Pikseller veya kelimeler gibi şeyleri, çoklu gizli katmanları (bu tür katmanlar ne kadar çoksa, gizli birimler (düğümler veya nöronlar olarak da bilinir) içeren daha derin bir ağı olduğu söylenir), ve bu düğümler arasında çalışan bağlantılara sahip bir dizi çıktı birimidir (Marcus vd., 2017).

Derin öğrenme, makine öğrenimi (ML) araştırmasının gelişmekte olan bir alanıdır. Yapay sinir ağlarının birden çok gizli katmanından oluşur. Derin öğrenme metodolojisi, doğrusal olmayan dönüşümleri ve büyük veri tabanlarında yüksek düzeyde model soyutlamalarını uygular. Çok sayıda alanda derin öğrenme mimarilerindeki son gelişmeler, yapay zekâyâ şimdiden önemli katkılar sağlamıştır.



Şekil 21. Basit Sinir Ağı Modeli



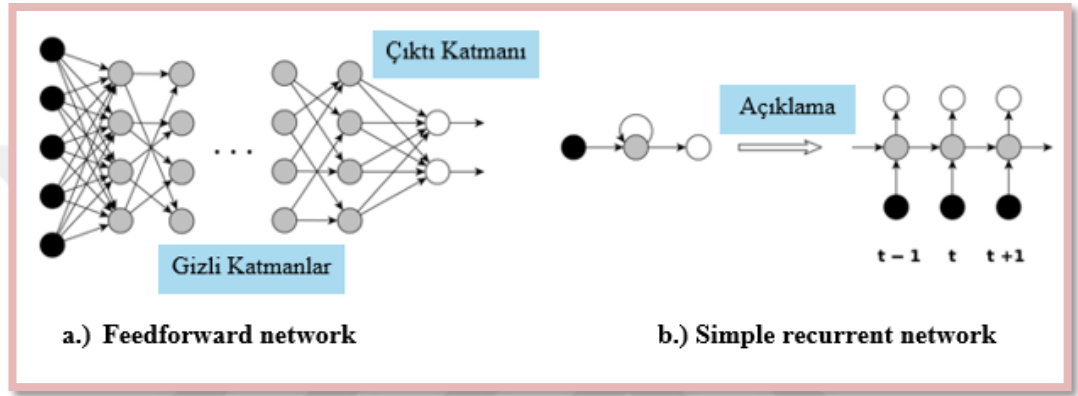
Şekil 22. Derin Sinir Ağı Modeli

Derin sinir ağıları, bilgisayar görüşü ve konuşma tanıma alanında önemli bir dönüm noktası olmuştur. Derin ağlar, makinelerin görüntüleri, konuşmaları tanımasını ve hatta oyunları insanlar için neredeyse imkânsız olan bir doğrulukla oynamasını sağlamaktadır.

Birçok farklı derin öğrenme algoritması vardır, bu popüler algoritmalarından ikisi şunlardır: Konvolüsyonel Sinir Ağları (CNN'ler) ve Tekrarlayan Sinir Ağları (RNN'ler) (Shahin vd., 2019).

Derin ağlar arasında çeşitli yapılar bulunabilir. İleri beslemeli ağlar, belirli bir boyuttaki bir girdiyi (örneğin bir görüntü) belirli bir boyuttaki bir çıktıya (örneğin bir sınıflandırma olasılığı) sabit sayıda katmandan geçerek eşlemektedir. Eğitim kolaylığı ve genellemesinin iyi olması nedeniyle en çok ilgi gören ileri beslemeli uygulamalardan biri de CNN'dir. CNN'ler, renkli görüntüler gibi birden fazla veri dizisini işlemek için tasarlanmıştır ve genellikle biyolojik görsel sistemlerden ilham alan bir şekilde evrişimli katmanların istifleme gruplarından ve havuzlama katmanlarından oluşur.

Öte yandan, tekrarlayan sinir ağları (RNN) genellikle yalnızca bir gizli katmana sahiptir, ancak öğeleri sırayla, birer birer işlerler ve her çıktı bir sonraki öğenin girişine dahil edilerek önceki öğelerin hafızasını tutarlar. Her bir adımın toplamı bu nedenle çok derin bir ileri besleme ağı olarak görülebilir. Bu, onları konuşma veya zaman serileri gibi sıralı girdiler için özellikle ilgi çekici kılar (Christin vd., 2019). RNN'nin popüler bir uygulaması, konuşma tanıma (Graves vd., 2007) veya çeviridir (Sutskever vd., 2014)



Şekil 23. Ortak Derin Sinir Ağlarının Mimarisi

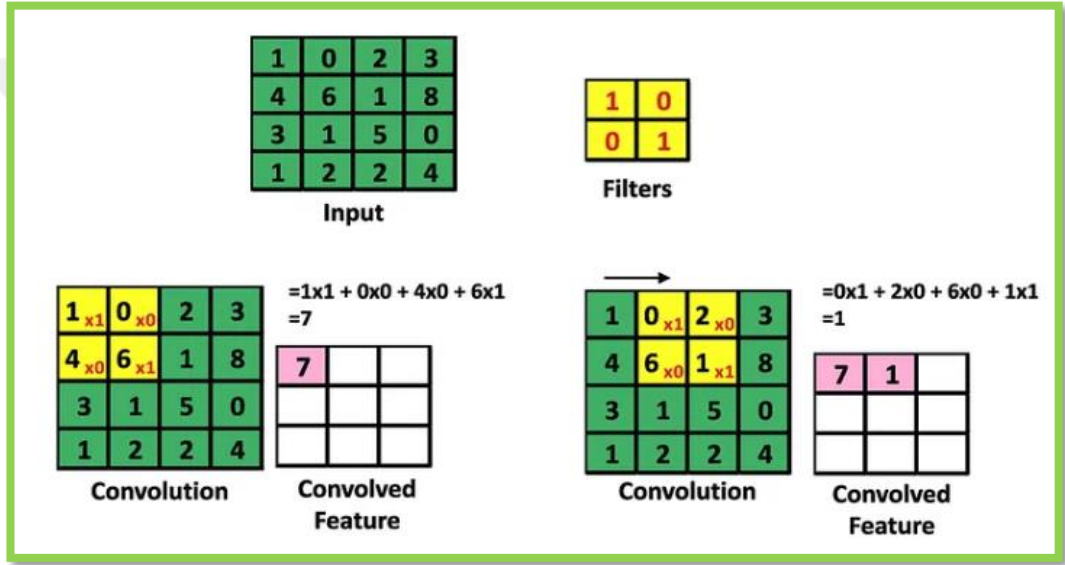
a.) İleri beslemeli ağlar, giriş katmanından çıkış katmanına ve gizli katmanlar aracılığıyla tek yönlüdür. Derin ileri beslemeli ağlar genellikle en az üç katmana sahiptir.

b.) Basit yinelenen sinir ağları, önceki zaman adımlarından girdi alır ve ileri beslemeli ağlara açılabilir.

4.4.1. Konvansiyonel Sinir Ağları (Convolutional Neural Network-CNN)

Konvolüsyonel Sinir Ağı (CNN), başlangıçta bilgisayarlı görü alanında kullanılan özel bir ileri beslemeli sinir ağı türüdür. Tasarımı, insan görsel korteksinden, hayvan beynindeki görsel mekanizmadan esinlenmiştir. Bu hücreler, girdi uzayı üzerinde yerel filtre görevi görmektedir. CNN, her biri görsel kortekste hücreler tarafından işlenen işlevi yerine getiren çok sayıda evrişimli katmandan oluşmaktadır. CNN'deki evrişimli katmanlar, yerel özellikleri ayıklayan, gizli katmanların alıcı alanlarını yerel olarak kısıtlayan bir özellik çıkarıcı rolünü oynamaktadır.

İlk aşamada, filtre görüntüyü taramak için kullanılır. Giriş görüntüsünde filtrenin yansıttığı her bölge alıcı bir alandır. Filtre aslında bir sayı dizisidir (ağırlıklar veya parametreler olarak adlandırılır). Filtre kayarken (veya evrilirken), ağırlık değerlerini görüntünün orijinal piksel değeriyle çarpar (öğeler akıllıca çarpılır). Çarpımların tümü, alıcı alanın bir temsilcisi olan tek bir sayı olarak toplanır. Her alıcı alan bir sayı üretir. Filtre görüntü üzerinde taramayı bitirdikten sonra, aktivasyon haritası veya özellik haritası adı verilen bir dizi elde edebiliriz (Lecun vd., 2015a).



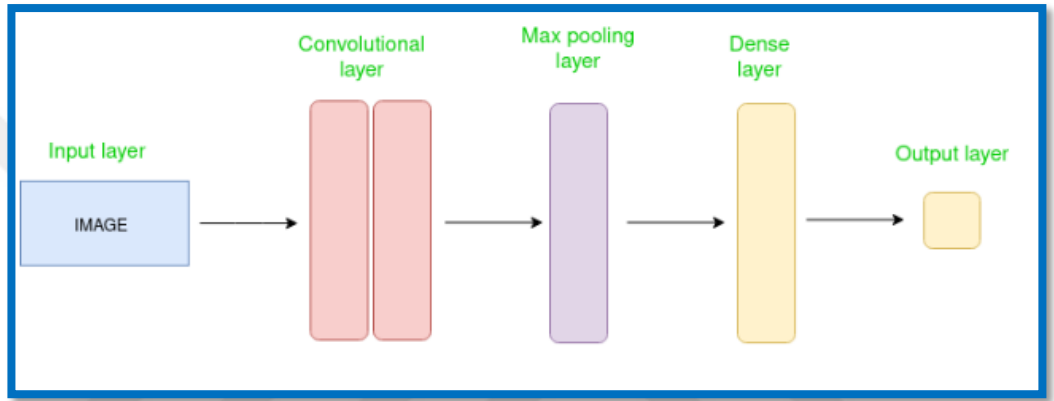
Şekil 24. Evrişimli Katmanlar

CNN, canlıların doğal görsel algılama mekanizmasından ilham alan tanınmış bir derin öğrenme mimarisidir (Gu vd., 2018).

Konvolüsyonel sinir ağları, görüntü, konuşma veya ses sinyali girişlerindeki üstün performansları ile diğer sinir ağlarından ayrılır. Üç ana katman türü vardır, bunlar:

- Evrişimli katman
- Havuzlama katmanı
- Tam bağlantılı (FC) katman

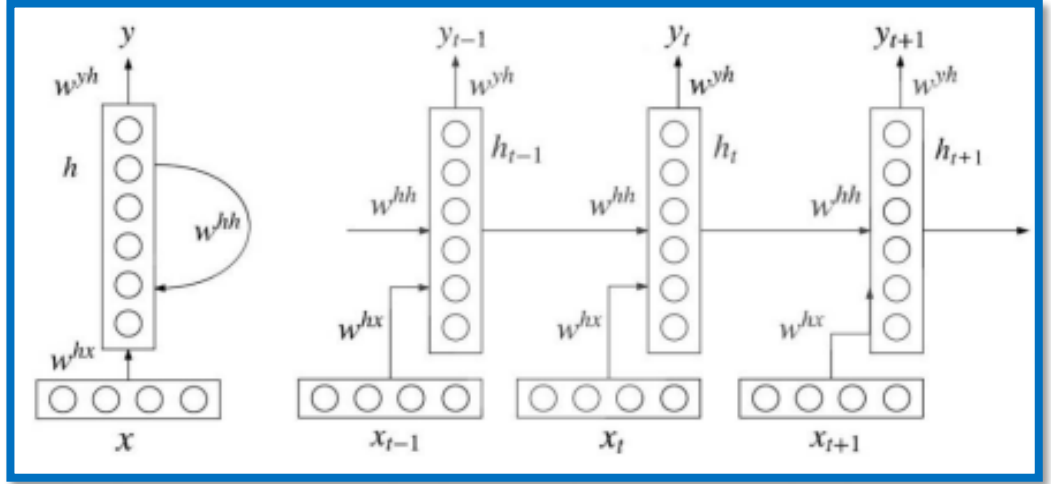
Evrişim katmanı, evrişim ağının ilk katmanıdır. Evrişimli katmanları ek evrişimli katmanlar veya havuzlama katmanları takip edebilirken, tam bağlantılı katman son katmandır. Her katmanda, CNN, görüntünün daha büyük bölümlerini tanımlayarak karmaşıklığını artırır. Önceki katmanlar, renkler ve kenarlar gibi basit özelliklere odaklanır. Görüntü verileri CNN'nin katmanları boyunca ilerledikçe, amaçlanan nesneyi tanımlayana kadar nesnenin daha büyük öğelerini veya şekillerini tanımaya başlar (IBM, 2021b).



Şekil 25. CNN Mimarisi

4.4.2. Tekrarlayan sinir Ağı (Recurrent Neural Network-RNN)

Tekrarlayan Sinir Ağı (RNN), yönlendirilmiş bir döngü oluşturur. İleri beslemeli sinir ağlarının aksine, RNN, dahili "hafızasını" bir girdi dizisini işlemek için kullanabilir, bu da onu sıralı bilgileri işlemek için popüler kılar. "Hafıza", RNN'nin bir dizinin her ögesi için aynı meta görevi gerçekleştirdiği ve her çıkışın önceki tüm hesaplamalara bağlı olduğu anlamına gelmektedir.



Şekil 26. Yinelenen Sinir Ağı

Şekil 26 bir RNN örneğini göstermektedir. Soldaki grafik, döngüleri olan katlanmamış bir ağ iken, sağdaki grafik, üç zaman adımlı katlanmış bir dizi ağıdır. Zaman adımlarının uzunluğu, girdinin uzunluğu tarafından belirlenmektedir (L. Zhang vd., 2018). Tekrarlayan sinir ağlarını uygulamak ve anlamak için ortam hazırlanmıştır (Bodén, 2001). RNN, sınıf tabanlı model de dahil olmak üzere diğer tekniklere dayalı diğer birkaç modelin karışımından daha iyi performans göstermektedir (Mikolov vd., 2010).

Tekrarlayan Sinir Ağı, her zaman adımı için bir tane olmak üzere birden fazla sabit aktivasyon fonksiyon biriminden oluşur. Her birimin, birimin gizli durumu olarak adlandırılan dâhili bir durumu vardır. Bu gizli durum, ağın şu anda belirli bir zaman adımında tuttuğu geçmiş bilgileri ifade eder. Bu gizli durum, ağın geçmişle ilgili bilgisindeki değişikliği belirtmek için her adımda güncellenir (Alind, 2021b).

Tekrarlayan bir sinir ağı (RNN), sıralı verileri veya zaman serisi verilerini kullanan bir tür yapay sinir ağıdır. Bu derin öğrenme algoritmaları, dil çevirisi, doğal dil işleme (NLP), konuşma tanıma ve resim alt yazısı gibi sıralı veya zamansal problemler için yaygın olarak kullanılır; Siri, sesli arama ve Google Çeviri gibi popüler uygulamalara dahil edilirler. İleri beslemeli ve Evrişimli Sinir Ağları (CNN'ler) gibi, tekrarlayan sinir ağları da öğrenmek için eğitim verilerini kullanır. Mevcut girdi ve çıktıyı etkilemek için önceki girdilerden bilgi aldıkları için "hafızaları" ile ayırt edilirler. Geleneksel derin sinir ağları, girdilerin ve çıktıların birbirinden bağımsız olduğunu varsayarken, tekrarlayan

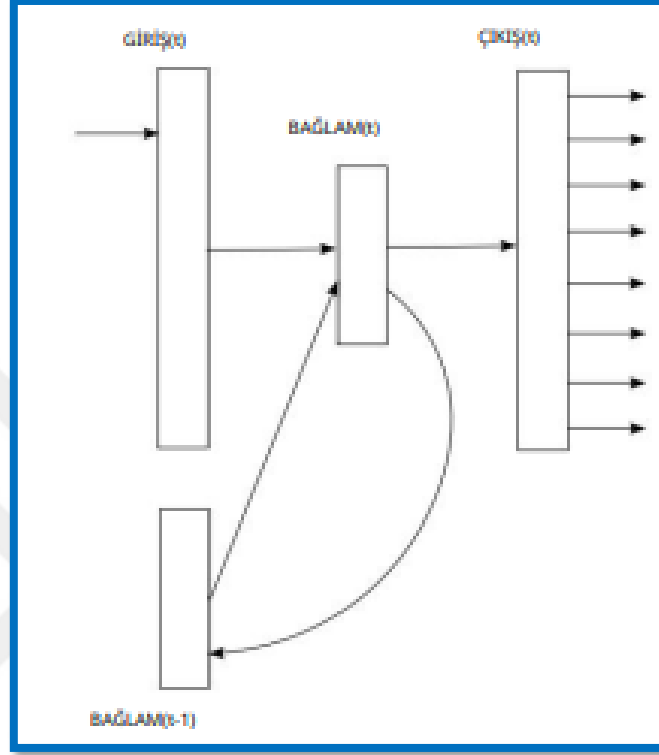
sinir ağlarının çıktısı, dizideki önceki ögelere bağlıdır. Yinelenen ağların diğer bir ayırt edici özelliği, ağı her katmanında parametreleri paylaşmalarıdır. İleri beslemeli ağlar, her düğümde farklı ağırlıklara sahipken, tekrarlayan sinir ağları, ağı her katmanında aynı ağırlık parametresini paylaşır. Bununla birlikte, bu ağırlıklar, takviyeli öğrenmeyi kolaylaştırmak için geri yayılım ve gradyan iniş süreçlerinde hala ayarlanmaktadır (IBM, 2021a).

Tekrarlayan ve ileri beslemeli sinir ağları İleri beslemeli ağlarda, tarih $N - 1$ kelime bağlamında temsil edilir. N -gram geri tepme modellerinde olduğu gibi aynı şekilde sınırlıdır. Tekrarlayan ağlarda geçmiş, tekrarlayan bağlantılara sahip nöronlar tarafından temsil edilir. Geçmiş uzunluğu sınırsızdır. Ayrıca, tekrarlayan ağlar tüm geçmişi düşük boyutlu uzayda sıkıştırmayı öğrenebilirken, ileri beslemeli ağlar yalnızca tek bir kelimeyi sıkıştırır (projelendirir). Tekrarlayan ağların kısa süreli bellek oluşturma olasılığı vardır, bu nedenle konum değişmezliği ile daha iyi başa çıkabilirler; ileri beslemeli ağlar bunu yapamaz (Mikolov vd., 2010).

Tekrarlayan sinir ağları, sıra verilerine özgü olduğu için geleneksel geri yayılımdan biraz farklı olan gradyanları belirlemek için zaman içinde geri yayılım (BPTT) algoritmasından yararlanır. BPTT' nin ilkeleri, modelin çıkış katmanından giriş katmanına hataları hesaplayarak kendini eğittiği geleneksel geri yayılımla aynıdır. Bu hesaplamalar, modelin parametrelerini uygun şekilde ayarlamamızı ve sığdırmamızı sağlar. BPTT' nin geleneksel yaklaşımdan farkı, BPTT' nin her zaman adımında hataları toplaması, oysa ileri beslemeli ağların, parametreleri her katmanda paylaşmamaları nedeniyle hataları toplamasına gerek olmamasıdır.

Bu süreçte, RNN' ler patlayan gradyanlar ve kaybolan gradyanlar olarak bilinen iki problemle karşılaşma eğilimindedir. Bu sorunlar, hata eğrisi boyunca kayıp fonksiyonunun eğimi olan gradyanın boyutuyla tanımlanır. Gradyan çok küçük olduğunda, küçülmeye devam eder ve ağırlık parametreleri önemsiz hale gelene kadar yani 0 oluncaya kadar güncellenmektedir. Bu meydana geldiğinde, algoritma artık öğrenmez. Patlayan gradyanlar, gradyan çok büyük olduğunda ortaya çıkar ve kararsız bir model oluşturur. Bu durumda, model ağırlıkları çok büyüyecek ve sonunda NaN olarak temsil edilecektir. Bu sorunlara bir çözüm, sinir ağı içindeki gizli katmanların

sayısını azaltmak ve RNN modelindeki karmaşıklığın bir kısmını ortadan kaldırmaktır (IBM, 2021a).



Şekil 27. Basit Yinelenen Sinir Ağı

Tekrarlayan ağda bir giriş katmanı x , gizli katman s (bağlam katmanı veya durum olarak da adlandırılır) ve çıkış katmanı y bulunur. Giriş vektörü $x(t)$, geçerli kelimeyi temsil eden w vektörünün ve $t-1$ zamanında bağlam katmanı s 'deki nöronların çıktısının birleştirilmesiyle oluşturulur. Performansı artırmak için, sık kullanılmayan sözcükler genellikle tek bir belirteçte birleştirilir.

RNN'lerin özelliklerinden dolayı metin sınıflandırmasında RNN'ler daha sık kullanılmaktadır. Bununla birlikte, uzun veri dizileri için, geleneksel RNN'ler, gradyanına karşı patlayan ve kaybolan duruma neden olmaktadır (Liu ve Guo, 2019). Bu tür bağımlılıklar, çoğu NLP uygulamasında ve özellikle duyarlılık analizinde yaygındır.

Bu sorunu çözmek için, Uzun Kısa Süreli Bellek (Long Short-Term Memory- LSTM) ve Kapılı Tekrarlayan Birim (GRU) ağları tanıtılmıştır. Birincisi sorunu giriş, unutma ve çıkış kapıları aracılığıyla ele alırken, ikincisi bir sıfırlama geçidi ve bir güncelleme geçidinden yararlanmaktadır. 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 (Basiri vd., 2021).

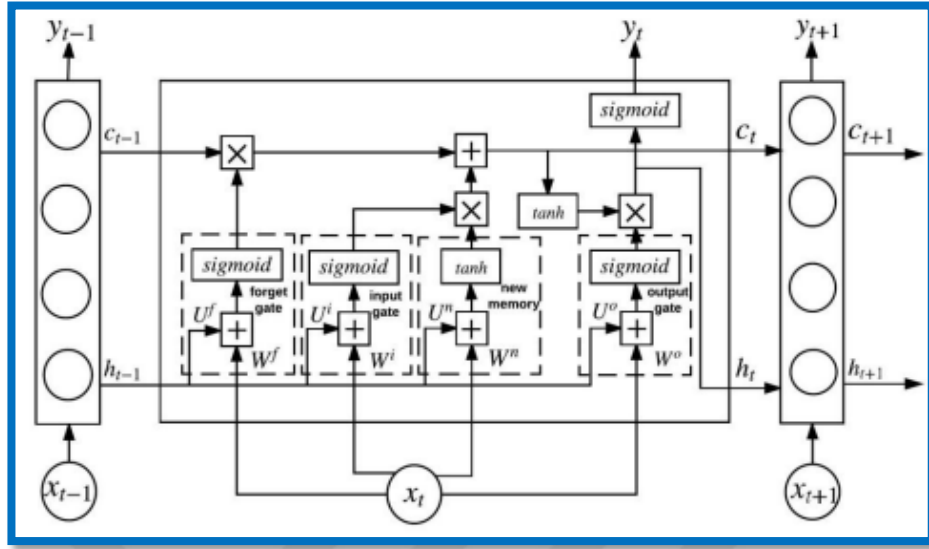
4.4.2.1. Kapılı Yinelemeli Sinir Ağları (GRNN)

Tekrarlayan Sinir Ağı (RNN), sıralı verileri veya zaman serisi verilerini kullanan bir tür yapay sinir ağıdır. RNN'ler sıralı modellemede güçlüdür. RNN'ler birçok metin işleme uygulaması için uygun olsalar da, girdi verilerinde uzun vadeli bağımlılıklar olduğunda kaybolan ve patlayan gradyanlardan muzdariptirler. Bu tür bağımlılıklar, çoğu NLP uygulamasında ve özellikle duyarlılık analizinde yaygındır. Bu sorunu çözmek için, uzun kısa süreli bellek (LSTM) ve kapılı tekrarlayan birim (GRU) ağları tanıtılmıştır. Birincisi sorunu giriş, unutma ve çıkış kapıları aracılığıyla ele alırken, ikincisi bir sıfırlama geçidi ve bir güncelleme geçidinden yararlanmaktadır (Basiri vd., 2021).

4.4.3. Uzun Kısa Dönem Bellek (Long-Short Term Memory-LSTM)

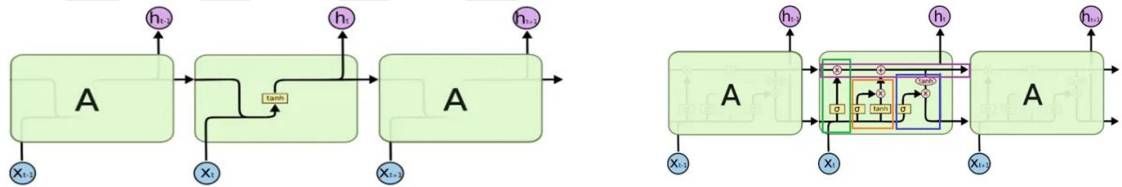
Uzun Kısa Süreli Bellek ağı (LSTM), ilk olarak 1997'de Hochreiter ve Schmidhub tarafından önerilmiştir (Hochreiter ve Scchmidhuber, 1997). Daha sonra orijinal LSTM biriminde bir dizi küçük değişiklik yapılmıştır (Graves vd., 2013).

LSTM, geri akış problemlerinin üstesinden gelmek için tasarlanmıştır. Kısa süreli gecikme yetenekleri kaybı olmadan, bir mimari için verimli bir gradyan tabanlı algoritma ile elde edilmektedir. Uzun süreli bağımlılıkları öğrenebilen özel bir RNN türüdür. Tüm RNN'ler bir zincirsiz yinelenen modül biçimine sahiptir. Standart RNN'lerde, bu yinelenen modül normalde basit bir yapıya sahiptir (L. Zhang vd., 2018).



Şekil 28. Uzun Kısa Dönem Bellek Ağı

Şekil 29’da LSTM’ nin bir örneğini göstermektedir. Zaman adımında, LSTM ilk olarak hücre durumundan hangi bilgilerin atılacağına karar verir. Standart bir RNN tek bir tanh katmanı içerirken, LSTM’ ler iletişim halinde olan 4 farklı katman içermektedir.



Şekil 29. RNN ve LSTM Model

Kaynak: (Akköse, 2020)

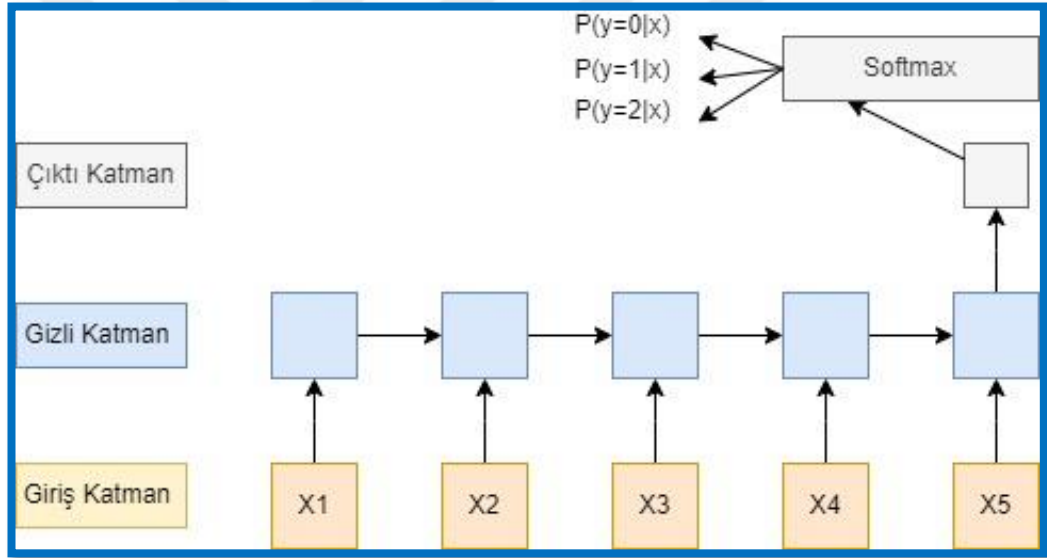
LSTM’ ler standart yinelenen sinir ağlarına benzer, ancak burada her sıradan yinelenen düğüm bir bellek hücresi ile değiştirilmektedir. Her bellek hücresi dahili bir durum içerir. Yani, gradyanın kaybolmadan veya patlamadan birçok zaman adımından geçebilmesini sağlayan, sabit ağırlıklı 1’in kendi kendine bağlı yinelenen kenarına sahip bir düğümdür. Basit tekrarlayan sinir ağları, ağırlıklar şeklinde uzun süreli belleğe sahiptir. Ağırlıklar, eğitim sırasında yavaşça değişir ve veriler hakkındaki genel bilgileri kodlar. Ayrıca, her düğümden ardışık düğümlere geçen kısa süreli aktivasyonlar şeklinde kısa süreli belleğe sahiptirler. LSTM modeli, bellek hücresi aracılığıyla bir ara tip depolama sunar. Bir bellek hücresi, yeni çoğaltıcı düğümlerin dahil edilmesiyle, belirli

bir bağlantı modelinde daha basit düğümlerden oluşturulmuş bileşik bir birimdir (A. Zhang vd., 2021).

4.4.4. Kapılı Sinir Ağı (Gated Recurrent Unit-GRU)

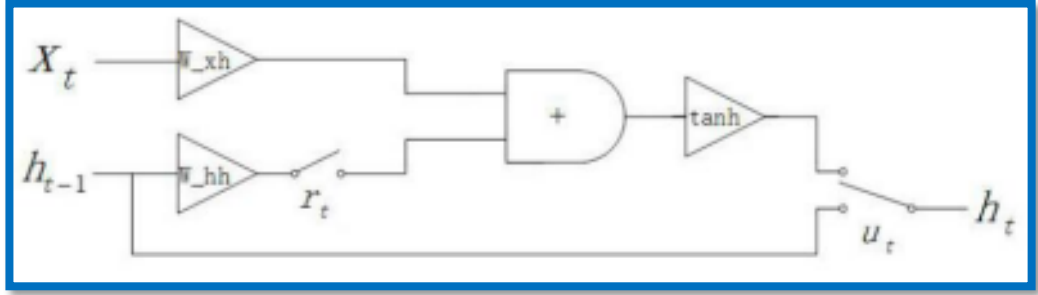
Kapılı sinir (GRU) ağları, RNN'ler kategorisine girer, yani altta yatan nöronlar arası bağlantı topolojisi en az bir döngü içeren sinir ağlarıdır. 1997'de tanıtıldılar ve sonraki birkaç yıl içinde daha da geliştirilmiştir (A. Zhang vd., 2021).

GRU, uzun vadeli bağımlılıkları öğrenirken geleneksel RNN'lerde yok olan ve patlayan gradyanların yaygın sorunlarını çözmek için kullanılan kapılı RNN'lerin bir türüdür (Shen vd., 2018).



Şekil 30. GRU Model Yapısı

Şekil 30 'da gösterildiği gibi, birden çok nörondan oluşan bir giriş katmanı vardır, nöron sayısı özellik uzayının boyutuna göre belirlenir. Benzer şekilde çıktı katmanındaki nöron sayısı da çıktı uzayına karşılık gelmektedir. Bellek hücrelerini içeren gizli katman(lar), GRU ağlarının ana fonksiyonlarını kapsar. Hücre durumundaki değişiklikler ve bakım, hücredeki iki kapıya bağlıdır: bir sıfırlama kapısı tr ve bir güncelleme kapısıdır. Bir bellek hücresinin yapısı, Şekil 31'de bir devre şeması olarak gösterilmektedir.



Şekil 31. GRU Bellek Hücresinin Yapısı

GRU, LSTM gibi yinelemeli sinir ağlarında ortaya çıkan kaybolan gradyan problemini ele alan bir mimaridir. LSTM gibi uzun vadeli bağımlılıkları öğrenme yeteneğine sahiptir. GRU unutmaya kapısına sahiptir fakat LSTM 'den farklı olarak bir çıkış kapısına sahip değildir.

LSTM gibi, GRU da metin, konuşma ve zaman serisi verileri gibi sıralı verileri işleyebilir. GRU' nun arkasındaki temel fikir, her zaman adımında ağın gizli durumunu seçici olarak güncellemek için geçiş mekanizmalarını kullanmaktır. Geçitleme mekanizmaları, ağa giren ve çıkan bilgi akışını kontrol etmek için kullanılır. GRU, sıfırlama kapısı ve güncelleme kapısı olarak adlandırılan iki kapı mekanizmasına sahiptir.

Sıfırlama kapısı, önceki gizli durumun ne kadarının unutulması gerektiğini belirlerken, güncelleme kapısı, gizli durumu güncellemek için yeni girişin ne kadarının kullanılması gerektiğini belirler. GRU' nun çıktısı, güncellenen gizli duruma göre hesaplanır (Alind, 2021a).

BEŞİNCİ BÖLÜM

5. DENEYSEL KIYASLAMALAR VE SONUÇLAR

5.1. Deneyisel Kıyaslamalar

Sosyal medya, günümüzde en çok kullanılan iletişim platformudur. Sosyal medya, insanların internet üzerinden birbirleriyle etkileşimde bulunabildiği ve içerik paylaşabildiği çevrimiçi platformları ifade eder. Sosyal medya platformları, kullanıcıların metin, fotoğraf, video ve bağlantı gibi çeşitli içerikleri paylaşmasına olanak tanır. Bu platformlar genellikle arkadaşlar, aile üyeleri, iş arkadaşları ve diğer insanlarla iletişim kurmak, bilgi edinmek ve eğlence için kullanılır. Sosyal medya üzerinden elde edilen veriler, kullanıcılar tarafından sosyal medya platformlarına yüklenen ve paylaşılan içeriklerin, etkileşimlerin ve kullanıcı davranışlarının toplandığı verilerdir. Bu veriler, platformlar tarafından genellikle reklam hedeflemesi, kullanıcı deneyiminin iyileştirilmesi ve veri analizi gibi amaçlarla kullanılır.

Telekomünikasyon, iletişim teknolojilerinin kullanılarak bilgi, ses, görüntü ve veri gibi iletişim sinyallerinin iletimini sağlayan bir alanı ifade eder. Bu teknolojiler, sesli telefonlar, mobil telefonlar, internet, radyo, televizyon, uydu iletişimi ve diğer iletişim sistemleri gibi çeşitli iletişim araçlarını kapsar. İnsanlar arasındaki uzaklığı aşarak bilgi paylaşımını kolaylaştırır. İletişim sinyalleri, ses, görüntü ve veri gibi farklı formatta iletim yapılabilir. Örneğin, sesli telefonlar insanların birbirleriyle konuşmasını sağlarken, internet aracılığıyla metin, fotoğraf, video gibi veriler paylaşılabilir. Telekomünikasyon ayrıca ticari ve endüstriyel alanlarda da büyük öneme sahiptir. İşletmeler, telekomünikasyon altyapısı üzerinden müşterileriyle iletişim kurar, veri transferi yapar ve iş süreçlerini yönetir. Telekomünikasyon ayrıca sağlık sektörü, ulaşım, medya, eğitim, finans gibi birçok sektörde de kullanılır.

Bu çalışmada, Twitter üzerinden belirlenen anahtar kelimeler üzerinden telekomünikasyon sektörüne yapılan şikayetler üzerinden elde edilen tweet derlemi olumlu, olumsuz ve nötr olmak üzere üç farklı duygu sınıfında etiketleme işlemi yapılmıştır. Bu çalışmada telekomünikasyon sektörünün sosyal medya üzerinden aldığı şikayetlerin, çalışmanın alanına özgü verilerden oluşan bir derlem haline getirilmiştir. Çalışmanın bu bölümünde, bu derlemde oluşturulan kelime temsillerinin başarıma

katkısı gözlemlenmiştir. Buna ek olarak çalışmada hem ikili hem de çoklu sınıflandırma işlemleri yapılmıştır. Beş bin verinin %20'si eğitim için, %80 test için kullanılmıştır. Test için ayırt edilen verinin eğitimi için ise %90'ı eğitim için, %10'u da test için kullanılmıştır.

Derin öğrenme tabanlı bir modeller ile sınıflandırılması hedeflenmektedir. Bu doğrultuda; GRU, LSTM, BiLSTM ve BiGRU çok sınıflı olmak üzere derin öğrenme yöntemleri ile sınıflandırma modelleri önerilmiştir. Twitter üzerinden elde edilen verileri kullanarak telekomünikasyon sektörüne yapılan şikayetlerin word2vec-CBOW (25, 50, 75, 100, 125) ve kelimelerden oluşan cümlelerin anlamını bulmak ve metin sınıflandırması için kullanılan TF-IDF (Terim Frekansı (TF) - Ters Yoğun Frekans (IDF)) ile n-gram (1,2,3) olarak parametre değişikliği yapılarak birden fazla (multilabel) sınıflandırma yapılmıştır. Ayrıca bunların tek katman ve iki katmandan oluşan mimarileri de test edilmiştir. Tahmin edilen değerlerin tablolar üzerinden değerlendirilerek tartışılmıştır. Başlangıç olarak 200 epoch değeriyle başlatılan eğitim kasılma ve yavaşlamalardan dolayı 150 epoch ile eğitime devam edilmiştir. Böylelikle en iyi başarıyı elde etmeye çalışılmıştır. Elde edilen confusion matrix (karışıklık matrisi)'leriyle ikili sınıflandırmanın bir karışıklık matrisi, bir ikili sınıflandırıcının dört sonucunun sayısının sayılmasıyla oluşturulan ikiye iki tablolar ile sonuçlar karşılıklı değerlendirilmiştir.

5.1.1. GRU İki Sınıflı Verilerin Sınıflandırma Başarımlarının Değerlendirilmesi

Veri setinden elde edilen word2vec-CBOW kelime temsilleri kullanılarak iki sınıflı verilerin ve N-Gram TF-IDF ile sınıflandırma başarımları gösterilmektedir.

Tablo 8. GRU Word2vec Kelime Temsillerini Kullanarak Verilerin Sınıflandırma Başarımları

	Katman Sayısı	25	50	75	100	125
Kesinlik (%)	1.Katman	80,6	81,5	79,5	79,9	77,6
	2.Katman	75,7	74,4	75,6	75,4	69,7
Hassasiyet (%)	1.Katman	31,9	29,2	30,8	31,8	31,9
	2.Katman	36,6	33,6	35,3	34,9	41,1
F-Score (%)	1.Katman	44,7	43,0	44,4	45,9	45,2
	2.Katman	49,4	46,3	48,2	48,6	51,7

Tablo 8’de, GRU modeli ile elde edilen sınıflandırma işleminin parametre ve katman sayısı sonuçlarının alt sınıflar bazında dağılımları görülmektedir. Sınıflandırma verisinden kelime temsillerinin elde edildiği testlerde tüm parametreler için kesinlik değerlerinin 1.katmanda en yüksek değerler gösterirken, 2. katmanda kesinlik başarımları azalan yönde etkilemiştir diyebiliriz. F skor değerlerinin tüm parametreler için de en düşük değerleri 1. katmanda, en iyi sonucu ise 2. katmanda elde etmiştir. Hassasiyet sınıflandırma başarımlarının oldukça yakın olduğu görülmektedir. En düşük başarımları tüm parametrelerde 1.katmanında elde etmişken, en iyi hassasiyet değerinin tüm parametreler için 2. katmanında elde edilmiştir.

Tablo 9. GRU Doğruluk Başarım Oranları

Katman Sayısı	Algılayıcı Sayısı	Sınıflandırma Başarımı
1	25	86,3
2	25	88,3
1	50	86,7
2	50	86,6
1	75	86,8
2	75	86,9
1	100	86,9
2	100	87,2
1	125	86,8
2	125	87,0

Hiper parametrelerin optimizasyonu sınıflandırma performansını etkileyen önemli noktalardan bir tanesidir. Kullanılan kelime temsillerinin algılayıcı sayısı ve katman sayısının belirlenmesi önemli parametrelerden biridir. Algılayıcı ve katman sayısındaki değişimin etkisini değerlendirmek için ikili sınıflandırma veri setindeki en yüksek sınıflandırma başarımını sağlayan alana özel kelime temsili ve GRU sınıflandırma yöntemi farklı vektör boyutları ve katman sayıları ile test edilmiştir. Tablo 9’da farklı vektör boyutları ve katman sayıları ile eğitilmiş alana özel kelime temsilleri ile elde edilmiş sınıflandırma başarımları görülmektedir. 25, 50, 75, 100 ve 125 vektör boyutları ve 1., 2. Katman sayıları için sınıflandırma skorları birbirine oldukça yakın elde edilmiştir. 25 için en kötü durumda vektör boyutu ve katman sayısı sınıflandırma başarımı 1. katmanda %86,3 olarak gerçekleşirken en iyi durumda vektör boyutu ve katman sayısı 2. katmanda %88,3 olarak gerçekleşmiştir. Geri kalan testlerde algılayıcı sayısı ve katman sayısının en iyi sonucu 100 için 2. Katmanda %87,2 olarak geçekleşmiştir. İlave Algılayıcı sayısı arttırıldıkça sınıflandırma başarımını negatif yönde etkilemiştir.

Tablo 10. GRU N-Gram TF-IDF Yöntemi Kullanarak Verilerin Sınıflandırma Başarımları

Katman Sayısı	N-Gram Sayısı	Sınıflandırma Başarımı
1	Unigram	91,8
2	Unigram	90,5
1	Bigram	89,1
2	Bigram	88,4
1	Trigram	87,6
2	Trigram	87.1

Tablo 10’da N-gram modelinin GRU sınıflandırma yöntemi üzerinden iki katmanlı sınıflandırma sonuçları verilmiştir. N-gram yöntemi kullanılarak sınıflandırılma yapıldığında doğru sınıflandırma başarısı hepsinde oldukça yakın değerler görülmüştür. En iyi sınıflandırma başarısını unigram boyutunda 1.katmanda %91,8 olmuştur. En düşük başarımları gösteren sınıflandırma trigramda 2. katmanda %87,1 olmuştur. N-ram ve katman sayısı arttırıldıkça başarımlarını negatif yönde etkilemiştir.

Tablo 11. GRU N-Gram TF-IDF Yöntemi Kullanarak Alt Sınıflar Bazında Verilerin Sınıflandırma Başarımları

	Katman Sayısı	Unigram	Bigram	Trigram
Kesinlik (%)	1.Katman	89,2	86,5	82,7
	2.Katman	85,6	81,9	79,3
Hassasiyet (%)	1.Katman	59,9	43,4	35,2
	2.Katman	53,6	41,7	33,4
F-Score (%)	1.Katman	71,7	57,8	49,4
	2.Katman	65,9	55,3	46,9

Tablo 11’de GRU modeli ile elde edilen sınıflandırma işleminin N-gram ve katman sayısı sonuçlarının alt sınıflar bazında dağılımları görülmektedir. Sınıflandırma verisinden n-gram değerlerinin elde edildiği testlerde unigram, bigram ve trigram için kesinlik değerlerinin; en iyi sonucu unigram boyutunda 1.katmanda yüksek değerde %89,2 olarak görülmüştür. Daha sonra sırasıyla katman ve n-gram parametre sayısı arttırıldıkça sırasıyla başarımlar oranı negatif yönde etkilemiştir. Hassasiyet ve F-skor değerlerinin de en iyi sonucu unigram boyutunda 1. Katman sınıflandırma başarımları elde etmiştir.

5.1.2. LSTM Modeli ile Verilerin Sınıflandırma Başarımlarının Değerlendirilmesi

Veri setinden elde edilen word2vec-CBOW kelime temsilleri kullanılarak iki sınıflı verilerin ve N-Gram TF-IDF ile sınıflandırma başarımları gösterilmektedir.

Tablo 12. LSTM Word2vec Kelime Temsillerini Kullanarak Verilerin Sınıflandırma Başarımları

	Katman Sayısı	25	50	75	100	125
Kesinlik (%)	1.Katman	76,5	76,2	78,3	88,8	79,1
	2.Katman	74,6	79,7	79,5	78,4	78,4
Hassasiyet (%)	1.Katman	30,5	31,8	30,2	32,5	32,9
	2.Katman	30,3	41,2	36,8	35,1	34,1
F-Score (%)	1.Katman	43,7	43,1	43,6	47,5	46,5
	2.Katman	43,1	54,3	50,3	48,1	47,5

Tablo 12’de LSTM modeli ile elde edilen sınıflandırma işleminin parametre ve katman sayısı sonuçlarının alt sınıflar bazında dağılımları görülmektedir. Sınıflandırma verisinden kelime temsillerinin elde edildiği testlerde 25, 100 ve 125 için kesinlik değerlerinin 1.katmanda yüksek değerler gösterirken 50 ve 75 parametreleri için kesinlik en iyi sonuçları 2.katmanda elde ettiği görülmektedir. F skor değerleri için en düşük değeri 25 ve 50 parametrelerinde sırasıyla 2. ve 1. katmanlarında, en iyi sonucu ise 50 parametresinin 2. katmanında elde edilmiştir. Hassasiyet sınıflandırma başarımlarının oldukça yakın olduğu görülmektedir. En düşük başarımları 75 parametresinde 1.katmanında, en iyi hassasiyet değerinin 50 parametresinde 2. Katmanında elde edilmiştir.

Tablo 13. LSTM Doğruluk Başarım Oranları

Katman Sayısı	Algılayıcı Sayısı	Sınıflandırma Başarımı
1	25	86,5
2	25	86,3
1	50	86,6
2	50	88,1
1	75	86,6
2	75	87,5
1	100	87,8
2	100	87,2
1	125	87,0
2	125	87,1

Algılayıcı ve katman sayısındaki değişimin etkisini değerlendirmek için ikili sınıflandırma veri setindeki en yüksek sınıflandırma başarımını sağlayan alana özel kelime temsili ve LSTM sınıflandırma yöntemi farklı vektör boyutları ve katman sayıları ile test edilmiştir. Tablo 13’de farklı vektör boyutları ve katman sayıları ile eğitilmiş alana özel kelime temsilleri ile elde edilmiş sınıflandırma başarımları görülmektedir. 25, 50, 75, 100 ve 125 vektör boyutları ve 1., 2. Katman sayıları için sınıflandırma skorları birbirine oldukça yakın elde edilmiştir. En kötü durumda vektör boyutu ve katman sayısı sınıflandırma başarımı 25 parametresinde, 2. katmanda %86,3 olarak gerçekleşirken en iyi durumda 50 parametre ve katman sayısında 2. katmanda %88,1 olarak gerçekleşmiştir. Geri kalan testlerde algılayıcı sayısı ve katman sayısının en iyi sonucu 100 parametresinde 1. Katmanda %87,8 olarak geçekleşmiştir. İlave Algılayıcı sayısı arttırıldıkça sınıflandırma başarımını negatif yönde etkilemiştir.

Tablo 14. LSTM N-Gram TF-IDF Yöntemi Kullanarak Verilerin Sınıflandırma Başarımları

Katman Sayısı	N-Gram Sayısı	Sınıflandırma Başarımı
1	Unigrams	90,6
2	Unigrams	90,1
1	Bigrams	88,8
2	Bigrams	88,4
1	Trigrams	87,6
2	Trigrams	92,9

Tablo 14’de N-gram modelinin LSTM sınıflandırma yöntemi üzerinden iki katmanlı sınıflandırma sonuçları verilmiştir. N-gram yöntemi kullanılarak sınıflandırılma yapıldığında doğru sınıflandırma başarısı birbirine yakın değerlerde elde edilmiştir. En iyi sınıflandırma başarısını trigram boyutunda 2. katmanda %92,9 olmuştur. En düşük başarımı gösteren sınıflandırma yine trigram boyutunda 1. katmanda %87,6 olarak elde edilmiştir. N-gram ve katman sayısı arttırıldıkça başarımlarını 1. Katmanda negatif yönde etkilerken, 2. Katmanda pozitif yönde bir etki göstermiştir.

Tablo 15. LSTM N-Gram TF-IDF Yöntemi Kullanarak Alt Sınıflar Bazında Verilerin Sınıflandırma Başarımları

	Katman Sayısı	Unigrams	Bigrams	Trigrams
Kesinlik (%)	1.Katman	85,8	85,2	80,2
	2.Katman	84,6	88,1	81,3
Hassasiyet (%)	1.Katman	54,7	41,8	37,4
	2.Katman	52,5	46,5	33,7
F-Score (%)	1.Katman	66,8	56,0	51,0
	2.Katman	64,6	60,8	47,7

Tablo 15’de LSTM modeli ile elde edilen sınıflandırma işleminin N-gram ve katman sayısı sonuçlarının alt sınıflar bazında dağılımları görülmektedir. Sınıflandırma verisinden n-gram değerlerinin elde edildiği testlerde unigram, bigram ve trigram için kesinlik değerlerinin; en iyi sonucu bigram boyutunda 2. katmanda en yüksek değeri göstermiştir. En düşük değerin ise trigram 1. katmanda elde etmiştir. Hassasiyet ve F-skor değerlerinin de en iyi sonucu unigram boyutunda 1. Katman en iyi sınıflandırma başarımını elde etmiştir.

Tablo 16. Word2vec Kelime Temsillerini Kullanarak Karşılaştırmalı Doğruluk(ACC) Sınıflandırma Sonuçları

	25	50	75	100	125
GRU	88,3	86,7	86,9	87,2	87,0
LSTM	86,5	88,1	87,5	87,8	87,1

Tablo 16’da iki katmanlı ve farklı parametre sayısına bağlı olarak iki farklı yöntem üzerinden gerçekleştirilen sınıflandırma sonuçlarının en iyi değerleri verilmiştir. Sınıflandırma sonucunda 1. ve 2. katman sayısı arttırıldıkça en iyi doğruluk başarımının “LSTM” algoritması gösterirken, en kötü doğruluk başarımının “GRU” algoritması göstermiştir. Bu çalışmada kullanılan LSTM modeli en iyi performansı sağlamıştır.

Tablo 17. Word2vec Kelime Temsillerini Kullanarak Karşılaştırmalı Alt Sınıflar Bazında Sınıflandırma Sonuçları

		25	50	75	100	125
GRU	Kesinlik (%)	80,6	81,5	79,5	79,9	77,6
	Hassasiyet (%)	36,6	33,6	35,3	34,2	41,1
	F-Score (%)	49,4	46,3	48,2	46,5	51,7
LSTM	Kesinlik (%)	76,5	79,7	79,5	88,8	79,1
	Hassasiyet (%)	30,5	41,2	36,8	35,1	34,1
	F-Score (%)	43,7	54,3	50,3	48,1	47,5

Tablo 17’de iki katmanlı ve farklı parametre sayıları ile gerçekleştirilen alt sınıflar bazında sınıflandırma sonuçları görülmektedir. Buna göre çalışmada en iyi performansı %54,3’lük 50 parametre sayısında F-Skor değeri ile LSTM algoritması gösterirken en kötü sonucu ise %46,3’lük 50 parametre sayısında F1 değeri ile GRU algoritması göstermiştir. Kullanılan her iki modelden, LSTM modeli en iyi F-skor değerini vermiştir.

Tablo 18. N-Gram TF-IDF Yöntemi Kullanılarak Karşılaştırmalı Doğruluk(ACC) Sınıflandırma Sonuçları

	Unigram	Bigram	Trigram
GRU	91,8	89,1	87,6
LSTM	90,6	88,8	92,9

Tablo 18’de N-gram modelinin iki katmanlı sınıflandırmaya bağlı olarak iki farklı yöntem üzerinden gerçekleştirilen sınıflandırma sonuçlarının en iyi değerleri verilmiştir. Buna göre bu çalışmada en iyi performansı %92,9’luk ACC değeri ile LSTM algoritması trigram boyutunda gösterirken en kötü sonucu ise %87,6’lık ACC değeri ile GRU algoritması trigram boyutunda göstermiştir. Bu çalışmada kullanılan LSTM modeli n-gram sınıflandırmasında en iyi performansı sağlamıştır.

Tablo 19. N-Gram TF-IDF Yöntemi Kullanılarak Karşılaştırmalı Alt Sınıflar Bazında Sınıflandırma Sonuçları

		Unigram	Bigram	Trigram
GRU	Kesinlik (%)	89,2	86,5	82,7
	Hassasiyet (%)	59,9	43,4	35,2
	F-Score (%)	71,7	57,8	49,4
LSTM	Kesinlik (%)	85,8	88,1	81,3
	Hassasiyet (%)	54,7	46,5	37,4
	F-Score (%)	66,8	60,8	51,0

Tablo 19’da iki katmanlı ve farklı parametre sayıları ile gerçekleştirilen alt sınıflar bazında sınıflandırma sonuçları görülmektedir. Buna göre çalışmada en iyi performansı %71,7’lük unigram parametre sayısında F-Skor değeri ile GRU algoritması gösterirken en kötü sonucu ise %49,4’lük trigram parametre sayısında F1 değeri ile GRU algoritması göstermiştir. Kullanılan her iki modelde, GRU modeli en iyi F-skor değerini vermiştir.

5.2. Sonuçlar ve Öneriler

Sosyal medya platformları insanların düşünce ve fikirlerini, şikayetlerini vb. özgürce ifade edebildiği çevrimiçi alanlardır. Bu platformlardan en popüler olanları biri de Twitter’dır. Dolayısıyla Twitter hem endüstri hem de akademi için önemli bir veri kaynağı halindedir. Bu platformlarda paylaşılan içerikler olumlu ve olumsuz içerikte olabilmektedir. Bu da bu alandaki çalışmalar için oldukça zorluk oluşturmaktadır. Özellikle Türkçe gibi zengin morfolojik dil yapısına sahip olan dillerin bu alanda çalışmaların yapmasındaki sayısını da etkilemiştir. Türkçe dilindeki bu çalışmaların oldukça az olduğu görülmüştür.

Bu tez çalışmasında bu göz önünde bulundurularak bu alanlara katkı sağlamak amaçlanmıştır. Literatürde sayılı çalışmalar olmakla birlikte literatür araştırmamızda bahsi geçen Qamar (2017) çalışması örnek alınmıştır. Qamar, Suudi Arabistan’daki farklı telekomünikasyon şirketlerine (Mobily, STC ve Zain) ait İngilizce yazılmış tweetlerin duygu analizini yapmıştır. Pozitif, negatif ve nötr olmak üzere üç sınıf ele alınmıştır. Sınıflandırma için ANN, k en yakın komşu (kNN), Naive Bayesian gibi bir dizi makine öğrenmesi algoritması kullanılmıştır. Sınıflandırma sonucu olarak, k değerini artırmanın

bazı algoritmalar için doğruluk üzerinde olumlu bir etkiye sahip olduğunu göstermiştir. Türkiye Telekomünikasyon sektörünün Twitter üzerinden aldığı şikayetler üzerine sınıflandırma çalışması hedeflenmiştir.

Bu çalışma, Türkiye Telekomünikasyon sektörüne yazılmış tweetlerin analizine odaklanmaktadır. Elde edilen derlem etiketlenme ve önişleme süreçlerinden sonra Word2vec yöntemi ile kelime temsilleri elde edilerek sınıflandırılması için kullanılmıştır. Birden fazla duygu sınıfı olduğu için birden fazla sınıflandırma işlemi yapılmıştır. Elde edilen derlemin %20'si eğitim için, % 80'i test için ayrılırken; test verisinin de eğitimi için % 90'ı kullanılırken %10'u test için kullanılmıştır. Derin öğrenme teknikleri arasında bulunan GRU ve LSTM modelleri ile sınıflandırma işlemi gerçekleştirilip elde edilen sonuçlar karşılaştırılmıştır. LSTM ve GRU modellerine ait elde edilen sonuçlar incelendiğinde 1.katman ve sınırlı parametre sayısında başarılı gerçekleştiği gözlemlenmiştir. Adım ve parametre sayısını arttırdığında başarımın bazı parametre ve katman seviyesinde negatif yönde etkilendiği ve çalışma süresinin artmasından dolayı zaman kaybı yaşandığı saptanmıştır. Sonuç olarak katman ve parametre sayısı belirli bir değerde tutulmalıdır.

İki katmanlı ve farklı parametre boyutlarında yapılan sınıflandırma sonucunda doğruluk sonuçları incelendiğinde; GRU ve LSTM modellerinde başarım oranının katman sayısı ve parametre sayısının arttığında her birinde farklılık göstermesi neticesinde parametre ve katman sayısının artışının negatif yönde etkilendiği söylenebilir. GRU modelin en iyi sonucun 1. Katman 25 parametre boyutunda %88,3 olarak elde etmiştir. LSTM modelin sınıflandırma başarımı en iyi sonucu 1. Katman 100 parametre boyutunda %87,8 olarak elde edilmiştir. N-gram sınıflandırmada en iyi sonucun LSTM modelinde 2. Katman ve Trigram boyutunda %92,9, GRU modelinde ise en iyi sonucun 1.katmanda Unigram boyutunda %91,8 olarak elde edilmiştir. Alt sınıflar bazında sınıflandırma sonuçları değerlendirilirken GRU modelde F-Score değerinin en iyi sonucu 2. Katman 125 parametre boyutunda %51,7 olarak elde edilirken; LSTM modelinde en iyi F-Score değeri 2.katman 50 parametre boyunda %54,3 olarak elde edilmiştir. N-gram alt sınıflar bazında sınıflandırma başarımları ise GRU modelde en iyi F-Score değeri 1. Katman Unigram boyutunda %71,7 olarak, LSTM modelinde ise yine 1. Katman Unigram boyutunda %66,8 olarak elde edilmiştir.

Twitter platformunda en çok negatif yönde şikâyetin Turkcell Telekom şirketinin aldığı daha sonra sırasıyla Avea_Türktelekom ve Vodafone Telekom şirketlerinin aldığı tespit edilmiştir. En çok sorun yaşanan ve buna bağlı olarak şikâyet edilen negatif yorum yapılan sorunların sırasıyla internet, fatura ve ücretlendirme, müşteri hizmetleri ve tarifiede yapıldığı görülmüştür.

Sonuç olarak, birden fazla sınıflandırma işlemi yapılan veri setinde özel kelime temsilleri ile en yüksek skorlar LSTM modeli elde edilmektedir. Yapılan bu çalışma kapsamında literatüre katlı sağlanmak istenmiştir. Bir sonraki çalışmalarda veri setinin genişletilmesi, daha güçlü donanımlarda farklı hiper parametreler denenebilir. Ayrıca kullanılan algoritmalara ek olarak farklı algoritmalar denenmesi ve karşılaştırılması önerilmektedir.

KAYNAKLAR

- ABBASİ, N. (2023). "What are n-grams? " <https://www.educative.io/answers/what-are-n-grams>
- AİZENBERG, I., AİZENBERG, N. N., ve VANDEWALLE, J. P. L. (2013)." Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications. "Springer Science & Business Media.
- AKKÖSE, O. (2020). "Uzun-Kısa Vadeli Bellek(LSTM)." Deep Learning Türkiye.
- AL-SMADİ, M., QAWASMEH, O., AL-AYYOUB, M., JARARWEH, Y., ve GUPTA, B. (2017). "Deep Recurrent Neural Network vs. Support Vector Machine for Aspect-Based Sentiment Analysis of Arabic Hotels' Reviews." **J. Comput. Sci.**, 27, 386–393. <https://doi.org/10.1016/J.JOCS.2017.11.006>
- ALBAYRAK, A. (2018). "Duygu Analizinde Farklı Vektör Temsil Yöntemleri ve Sınıflayıcıların Karşılaştırılması."
- ALBAYRAK, M., TOPAL, K., ve ALTINTAŞ, V. (2017). "Sosyal Medya Üzerinde Veri Analizi: Twitter Data Analysis On Social Media: Twitter. "Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. <https://iibfdergi.sdu.edu.tr/assets/uploads/sites/352/files/yil-2017-cilt-22-sayi-kayfor15-yazi31-30122017.pdf>
- ALİND, G. (2021a). "Gated Recurrent Unit Networks. Machine LearningMachine Learning."
- ALİND, G. (2021b). "Recurrent Neural Networks Explanation."
- ALMUQREN, L., ve ALEXANDRA CRISTEA, WARWICKACUK I. (2016). "Twitter Analysis to Predict the Satisfaction of Telecom Company Customers." <https://doi.org/10.1109/FiCloud.2014.95>
- AVUÇLU, E., ve ELEN, A. (2020). "Makine Öğrenme Algoritmalarına Genel Bakış. In Geleceğin Dünyasında Bilimsel ve Mesleki Çalışmalar" (pp. 34–50). Ekin Basım Yayın Dağıtım.
- BASİRİ, M. E., NEMATİ, S., ABDAR, M., CAMBRİA, E., ve ACHARYA, U. R. (2021). "ABCDM: An Attention-Based Bidirectional CNN-RNN Deep Model for Sentiment Analysis." **Future Generation Computer Systems**, 115, 279–294. <https://doi.org/10.1016/J.FUTURE.2020.08.005>
- BENGİO, Y., DUCHARME, R., VİNCENT, P., ve JAUVIN, C. (2003). "A Neural Probabilistic Language Model." **Journal of Machine Learning Research**, 3(6). <https://doi.org/10.1162/153244303322533223>
- BODÉN, M. (2001). "A Guide to Recurrent Neural Networks and

Backpropagation."

- CASTELLS, M. (2007). "Communication, Power and Counter-power in the Network Society." **International Journal of Communication**, 1(1), 29. <https://ijoc.org/index.php/ijoc/article/view/46>
- ÇAVUŞ, T., ve EĞİLMEZ, Ö. (2021). "Sosyal Medyada Kurumsal İtibar Yönetimi: Türkiye'deki Telekomünikasyon Şirketlerinin Twitter Veri Analizi." **Bilecik Şeyh Edebali Üniversitesi Sosyal Bilimler Dergisi**, 6(2), 210–226. <https://doi.org/10.33905/BSEUSBED.996512>
- CHRISTIN, S., HERVET, É., ve LECOMTE, N. (2019). "Applications For Deep Learning in Ecology." **Methods in Ecology and Evolution**, 10(10), 1632–1644. <https://doi.org/10.1111/2041-210X.13256>
- DAOUADI, K. E., REBAÏ, R. Z., ve AMOUS, I. (2021). "Optimizing Semantic Deep Forest for Tweet Topic Classification." **Information Systems**. <https://doi.org/10.1016/j.is.2021.101801>
- DEWING, M. (2005). "Social Media: An Introduction." <https://bdp.parl.ca/staticfiles/PublicWebsite/Home/ResearchPublications/InBriefs/PDF/2010-03-e.pdf>
- EDGAR ALTSZYLER, M. S. S. R. D. F. S. (2017). "Comparative study of LSA vs Word2vec embeddings in small corpora: a case study in dreams database."
- FANG, N., PAI, S. P., ve EDWARDS, N. (2016). "Neural Network Modeling and Prediction of Surface Roughness in Machining Aluminum Alloys." **Journal of Computer and Communications**, 4(5). <https://doi.org/10.4236/jcc.2016.45001>
- GÖÇGÜN, Ö. F., ve ONAN, A. (2021). "Amazon Ürün Değerlendirmeleri Üzerinde Derin Öğrenme/Makine Öğrenmesi Tabanlı Duygu Analizi Yapılması." **Avrupa Bilim ve Teknoloji Dergisi**, 24, 445–448. <https://doi.org/10.31590/ejosat.902674>
- GRAVES, A., FERNÁNDEZ, S., ve SCHMIDHUBER, J. (2007). "Multi-dimensional recurrent neural networks." **Lecture Notes in Computer Science** (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4668 LNCS(PART 1), 549–558. https://doi.org/10.1007/978-3-540-74690-4_56/COVER
- GRAVES, A., MOHAMED, A. R., ve HINTON, G. (2013). "Speech Recognition with Deep Recurrent Neural Networks. ICASSP, IEEE" **International Conference on Acoustics, Speech and Signal Processing - Proceedings**, 6645–6649. <https://doi.org/10.1109/ICASSP.2013.6638947>
- GU, J., WANG, Z., KUEN, J., MA, L., SHAHROUDY, A., SHUAİ, B., LİU, T., WANG, X., WANG, G., CAI, J., ve CHEN, T. (2018). "Recent advances in convolutional neural networks." **Pattern Recognition**, 77, 354–377. <https://doi.org/10.1016/J.PATCOG.2017.10.013>

- GÜÇDEMİR, Y. (2017). "**Sosyal Medya (1.baskı).**" DER Kitabevi Yayınevi ve DağıtımPaz. Ltd. Şti.
- GÜLDAL, S. (2021). "Unsupervised Machine Learning Algorithms To Find 3- Kcp Solution: Modularity, Clique Percolation, Spectral, Centrality, And Hierarchical Clustering. " **Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi**, 8, 169–178.
- GUPTA, A., ve JHA, R. K. (2015). "**A Survey of 5G Network: Architecture and Emerging Technologies. 3.**"
- GÜRSAKAL, N., TÜZÜNTÜRK, S., ve SERT ETEMAN, F. (2014). "**Sosyal Ağ Verilerinin Kuvvet Yasası Olasılık Dağılımına Uygunluk Analizi: Twitter Örneği.**" <https://www.researchgate.net/publication/301650314>
- HACIHASANOĞLU, B. (2017). "Sosyal Medyada Pazarlama Stratejilerinin Oluşturulması Ve Telekomünikasyon Sektörüne İlişkin Bir Uygulama." **İstanbul Ticaret Üniversitesi.**
- HİNTON, G. E., ve SALAKHUTDİNOV, R. R. (2006). "Reducing The Dimensionality of Data With Neural Networks." **Science** (New York, N.Y.), 313(5786), 504–507. <https://doi.org/10.1126/SCIENCE.1127647>
- HOCHREİTER, S., ve SCCHMİDHUBER, J. (1997). "**Long Short-Term Memory. Neural Computation,**" 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- HOUBY, E. M. F. EL, ve MOHAMED, H. K. (2021). "Deep Learning for Emotion Analysis in Arabic Tweets." **Journal of Big Data.** <https://doi.org/10.1186/s40537-021-00523-w>
- IBM. (2021a). "**Recurrent Neural Networks?**"
- IBM. (2021b). "**Convolutional Neural Networks.**"
- INVESTOPEDIA. (1999). "**Telekomünikasyon Sektörü Nedir?**" <https://www.investopedia.com/ask/answers/070815/what-telecommunications-sector.asp>
- ITU-D. (2021). "**The Telecommunication Development Sector-Statistics.**" <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
- ITU-D. (2023). "**The Telecommunication Development Sector-Statistics.**" <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
- KABAKÇI, D. (2015). "**Kurumsal İletişim Sürecindeki Rolü Açısından Online Şikâyetlerin Önemi.**"
- KABAKLARLI, E., ve İŞİCİK, Ş. (2020). "Türkiye Telekomünikasyon Sektörüne Genel Bakış. " **Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, 12(1).

<https://doi.org/20.500.12451>

- KALABALIK, G. (2016). "Telekomünikasyon Sektörü Müşteri Ayrılma Analizinde Birleştirmeli Sınıflandırma Yöntemleri Performanslarının Karşılaştırması."
- KAMIŞ, S., ve GOULARAS, D. (2019). "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data." **Proceedings - 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications**, Deep-ML 2019, 12–17. <https://doi.org/10.1109/DEEP-ML.2019.00011>
- KENT, B. (2012). "Telekomünikasyon Hizmetlerinde Evrensel Hizmet." <https://dergipark.org.tr/en/download/article-file/789385>
- KILDAN, K. (2021). "Telekomünikasyon Sektöründe Çağrı Merkezlerine Müşteri Şikayeti İçin Bir Alternatif – Sosyal Medyanın Müşteri Şikayeti İçin Kullanımı."
- KILIÇ, S. (2021). "Makine öğrenimi nasıl çalışır?" – Yapay zekanın merkezinde. Makaleci.
- KOCA, G., EĞİLMEZ, Ö., URMAK AKÇAKAYA, ED. ve HATİPOĞLU, C. (2021). "İşletmelerin Web Sitesi Kullanımının Değerlendirilmesi: Türkiye Örneği ." **Sakarya İktisat Dergisi**, 10 (3), 275-295. <https://dergipark.org.tr/en/pub/sid/issue/64860/986868>
- KOÇ, F., GÜNALAN, M., ÖZBEK, V., ve ÇINAR, B. (2015). "Şehirlerarası Otobüs Firmalarının Ulaşılabilirliğinin Memnuniyet, Firma İtibarı Ve Algılanan Değer Üzerindeki Etkisi." 3(1), 39–60. <https://doi.org/10.18825/irem.08272>
- KORKMAZ, E., ve TEKTAŞ, N. (2022). "Yapay Sinir Ağları Ve Makine Öğrenmesi Ile Güneş Işınımının Analizi: Bursa Ve Çanakkale Örneği." **Bandırma Onyedil Eylül Üniversitesi, Sosyal Bilimler Enstitüsü**, Yayınlanmış Yüksek Lisans Tezi
- KPMG. (2020). "Telekomünikasyon." <https://assets.kpmg/content/dam/kpmg/tr/pdf/2020/05/sektorel-bakis-2020-telekomunikasyon.pdf>
- KUMAŞ, E. (2021). "Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması." **ESTUDAM Bilişim Dergisi**, 2(2), 1–5.
- KUYUCUK, T., ÇALLI, L. (2022). "COVID-19 Salgını Sırasında Kargo Hizmetlerine Yönelik Şikayetleri Analiz Etmek İçin Çok Etiketli Sınıflandırma Yöntemlerinin Kullanılması: Anket Tabanlı ve Kelime Tabanlı Etiketlemenin Karşılaştırılması". **Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Dergisi**, cilt 5, hayır. 3, s. 371-384, Aralık 2022, doi:10.35377
- KÜNDÜM, D., KİLİMCİ, Z. H., UYSAL, M., ve UYSAL, O. (2020). "Evaluation of Customer Satisfaction about Telecom Operators in Turkey by Analyzing

Sentiments of Customers through Twitter.
"http://jdatasci.com/index.php/jdatasci/article/view/43/24

LECUN, Y., BENGIO, Y., ve HINTON, G. (2015b). "**Deep Learning.**" *Nature* 2015 521:7553, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>

LIU, G., ve GUO, J. (2019). "Bidirectional LSTM with attention mechanism and convolutional layer for text classification." *Neurocomputing*, 337, 325–338. <https://doi.org/10.1016/J.NEUCOM.2019.01.078>

MARCUS, G., THANK CHRISTINA, I., CHOLLET, F., DAVIS, E., LIPTON, Z., PACÍFICO, S., SARIÁ, S., ve VOULOUMANOS, A. (2017). "Deep Learning: A Critical Appraisal." **New York University.** <http://www.nytimes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial->

MIKOLOV, T., CHEN, K., CORRADO, G., ve DEAN, J. (2013). "Efficient Estimation of Word Representations in Vector Space." **1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings.**

MIKOLOV, T., KARAFIAT, M., BURGET, L., CERNOCK, J. "HONZA," ve KHUDANPUR, S. (2010). "**Recurrent Neural Network Based Language Model.**" *Interspeech.*

MOHRİ, M., ROSTAMIZADEH, A., ve TALWALKAR, A. (2018). "**Foundations of Machine Learning (2nd ed.)**". MIT Press.

MUSTAFA QAMAR, A., ALSUHİBANY, S. A., ve SOHAİL AHMED, S. (2017). "Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies." **IJACSA) International Journal of Advanced Computer Science and Applications**, 8(1). <https://about.twitter.com/company>

ÖĞÜNDÜR, G. (2019). "**Doğruluk (Accuracy).**"

ORTİZ-OSPINA, E., ve ROSER, M. (2019). "**The Rise of Social Media.**" <https://ourworldindata.org/rise-of-social-media>

ÖZER, İ., ÇETİN, O., TEMURTAŞ, F., ve GÖRÜR, K. (2021). "Improved Machine Learning Performances With Transfer Learning to Predicting Need for Hospitalization in Arboviral Infections Against the Small Dataset." **Neural Computing and Applications**, 33(1), 14975–14989. <https://doi.org/10.1007/s00521-021-06133-0>

ÖZER, İ., EFE, S. B., ve ÖZBAY, H. (2021). "Spektrogram Görüntüleri Kullanarak Güç Kalitesi Bozukluklarının Sınıflandırılması için CNN/Bi-LSTM Tabanlı Derin Öğrenme Algoritması." **International Transactions on Electrical Energy Sysems**, 31(12), e13204. <https://doi.org/10.1002/2050-7038.13204>

ÖZER, Z. (2019). "Veri Madenciliği Algoritmalarının Paralel Ve Dağıtık Sistemlerde

Gerçekleştirilmesi. " **Karabük Üniversitesi.**

ÖZMEN, M., DELİCE, Y., ve KIZILKAYA AYDOĞAN, E. (2017). "Telekomünikasyon Sektöründe Müşteri Kaybı Yönetimi İçin Meta Sezgisel Tabanlı Karar Destek Sistemi."

ÖZSOY, T. (2021). "Türkiye’de Telekomünikasyon Sektörü Hizmet Sağlayıcılarına İlişkin Tüketici Şikayetleri: Sorunlar ve Çözüm Önerileri." **Artıbilim Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi Sosyal Bilimler Dergisi**, 4(1), 1–19.

ÖZER, İ., EFE, S. B., ve ÖZBAY, H. (2021). "Kısa Vadeli Yük Tahmini İçin Birleştirilmiş Bir Derin Öğrenme Uygulaması." **İskenderiye Mühendislik Dergisi**, 60(4), 3807–3818. <https://doi.org/10.1016/j.aej.2021.02.050>

PARLAR, T., SARAC, E., ve OZEL, S. A. (2017). "Türkçe Twitter Verilerinde Duygu Analizi için Nitelik Seçim Yöntemlerinin Karşılaştırılması." **2017 25th Signal Processing and Communications Applications Conference**, SIU 2017. <https://doi.org/10.1109/SIU.2017.7960388>

REPUBLIC, C., ve MIKOLOV, T. (2012). "Statistical Language Models Based on Neural Networks." **Wall Street Journal**, April. <https://doi.org/10.1016/j.csl.2015.07.001>

RUMELHART, D. E., HINTON, G. E., ve WILLIAMS, R. J. (1986). "Learning Representations By Back-Propagating Errors." **Nature** 1986 323:6088, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>

SALUR, M. U., ve AYDIN, İ. (2021). "Türkçe Tweetler için Derin Özellik Çıkarımı Tabanlı Yeni Bir Duygu Sınıflandırma Modeli." **Fırat Üniversitesi Müh. Bil. Dergisi**, 34(1), 1–13. <https://doi.org/10.35234/fumbd.753886>

SANTOS, I., PENYA, Y. K., DEVESA, J., ve BRİNGAS, P. G. (2009). "N-Grams-Based File Signatures for Malware Detection." <https://doi.org/10.5220/0001863603170320>

SCOTT, D. M. (2015). "History Open for Business The Long Tail of Marketing Tell Me Something I Don’t Know, Please Bricks-and-Mortar News." **John Wiley & Sons, Inc.**

SHAHİN, I., NASSİF, A. B., ve HAMSA, S. (2019). "Emotion Recognition Using Hybrid Gaussian Mixture Model and Deep Neural Network." **IEEE Access**, 7, 26777–26787. <https://doi.org/10.1109/ACCESS.2019.2901352>

SHEN, G., TAN, Q., ZHANG, H., ZENG, P., ve XU, J. (2018). "Deep Learning With Gated Recurrent Unit Networks for Financial Sequence Predictions." **Procedia Computer Science**, 131, 895–903. <https://doi.org/10.1016/J.PROCS.2018.04.298>

SHU, C., (2020). " TikTok, WeChat and the growing digital divide between the US and

China" TikTok, WeChat and the growing digital divide between the US and China | TechCrunch." <https://techcrunch.com/2020/09/22/tiktok-wechat-and-the-growing-digital-divide-between-the-u-s-and-china/> , Eriřim Tarihi 08/06/2023.

SMAGULOVA, K., ve JAMES, A. P. (2019). "A Survey on LSTM Memristive Neural Network Architectures and Applications." **The European Physical Journal Special Topics**, 112(7), 2313–2324. <https://doi.org/10.1140/epjst/e2019-900046-x>

SOLMAZ, B., TEKİN, G., HERZEM, Z., DEMİR, M., (2013). "İnternet ve Sosyal Medya Kullanımı Üzerine Bir Uygulama." **Selçuk İletişim**, 7(4), 23–32. <https://doi.org/10.18094/SI.92009>

STEWART, K. (2019). "Sensitivity vs Specificity." Technology Networks Analysis & Separations.

SUTSKEVER, I., VINYALS, O., ve LE, Q. V. (2014). "Sequence to Sequence Learning with Neural Networks. **Advances in Neural Information Processing Systems**, 4(January)," 3104–3112. <https://arxiv.org/abs/1409.3215v3>

TEKTAŞ, M., AKBAŞ, A., ve TOPUZ, V. (2002). "Yapay Zekâ Tekniklerinin Trafik Kontrolünde Kullanılması Üzerine Bir İnceleme."

TEKTAŞ, M., TEKTAŞ, N., ÖZER, İ., ve ÖZER, Z. (2019). "Pyhton ve Matlab ile Biliřim Matematięi (1. baskı)". Mikyas Basım Yayın Matbaacılık.

TOPKAYA, F. (2003). "Rekabet Kurumu (1. baskı)." Rekabet Kurumu.

TÜİK. (2023). "Ulařtırma ve Haberleřme." <https://data.tuik.gov.tr/Kategori/GetKategori?p=ulastirma-ve-haberlesme-112&dil=1>

TUNCER, A. S., ÖZATA, F. Z., AKAR, E., ve ÖZTÜRK, M. C. (2013). "Sosyal Medya (2.baskı)". 1.Baskı, Eskiřehir, Anadolu Üniversitesi.

TURING, A. M. (1950). "Computing Machinery And Intelligence." 433–460.

TÜRK TELEKOM GRUBU. (2021). "Tarihçe | Türk Telekom Yatırımcı İliřkileri." <http://www.ttyatirimciiliskileri.com.tr/tr-tr/turk-telekom-grubu/turk-telekoma-yatirim/sayfalar/tarihce.aspx>

ULUÇ, G., ve YARCI, A. (2017). "Sosyal Medya Kùltürü". **Dumlupınar Üniversitesi Sosyal Bilimler Dergisi**, (52), 88-102. <https://dergipark.org.tr/tr/pub/dpusbe/issue/29388/314629>

USTA, E. (2021). "Telekomünikasyon Sektöründen Sınıflandırma Algoritmaları İle Müřteri Kayıp Analizi" **İstanbul Üniversitesi, Fen Bilimleri Enstitüsü**, İstanbul.

UYANIK, F. (2021). "Telekomünikasyon Sektörü İçin Veri Madencilięi ve Makine Öğrenmesi Teknikleri İle Ayrılan Müřteri Analizi". **Düzce Üniversitesi Bilim ve**

Teknoloji Dergisi, cilt.9, sa.3, ss.172-191, 2021

- VALLETTA, J. J., TORNEY, C., KINGS, M., THORNTON, A., ve MADDEN, J. (2017). "Applications of Machine Learning in Animal Behaviour Studies". **Animal Behaviour**, 124, 203–220. <https://doi.org/10.1016/J.ANBEHAV.2016.12.005>
- WANG, S.-C. (2003). "Artificial Neural Network. In Interdisciplinary Computing in Java Programming (pp. 81–100)". **Springer**. https://doi.org/10.1007/978-1-4615-0377-4_5
- WE ARE SOCIAL. (2023). "**Digital 2023 - We Are Social USA**". <https://wearesocial.com/us/blog/2023/01/digital-2023/>
- WEI, C. C., CHOY, C. S., ve YEW, W. K. (2009). "Is the Malaysian Telecommunication Industry Ready for Knowledge Management Implementation?". **Journal of Knowledge Management**. <https://doi.org/10.1108>
- Wu, P. (2020). An Introduction to Word Embeddings.
- YILMAZ, Ş. Ş., OZER, I., ve GÖKÇEN, H. (2021). "**Türkçe Metinlerde Derin Öğrenme Yöntemleri Kullanılarak Duygu Analizi**". <https://www.researchgate.net/publication/350352733>
- ZHANG, A., LIPTON, Z. C., LI, M., ve SMOLA, A. J. (2021). "**Dive into Deep Learning**". <https://doi.org/10.48550/arXiv.2106.11342>
- ZHANG, L., WANG, S., ve LIU, B. (2018). "Deep Learning for Sentiment Analysis: A Survey." **WIREs Data Mining and Knowledge Discovery**, 8(4). <https://doi.org/10.1002/widm.1253>