

T.C
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
YÖNETİM BİLİŞİM SİSTEMLERİ PROGRAMI
YÜKSEK LİSANS TEZİ

**YAPAY SİNİR AĞLARI VE BERT DİL MODELİ KULLANILARAK
ZAMAN BAZLI DUYGU ANALİZİ: WHATSAPP YENİ GİZLİLİK
SÖZLEŞMESİNE YÖNELİK YORUMLARIN ARAŞTIRILMASI**

Kazım Tibet SAR

Danışman
Dr. Öğr. Üyesi Kutan KORUYAN

İZMİR - 2021

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “Yapay Sinir Ağları ve BERT Dil Modeli Kullanılarak Zaman Bazlı Duygu Analizi: WhatsApp Yeni Gizlilik Sözleşmesine Yönelik Yorumların Araştırılması” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

Tarih

27/07/2021

Kazım Tibet SAR

ÖZET

Yüksek Lisans Tezi

Yapay Sinir Ağları ve BERT Dil Modeli Kullanılarak Zaman Bazlı Duygu

Analizi: WhatsApp Yeni Gizlilik Sözleşmesine Yönelik Yorumların

Araştırılması

Kazım Tiber SAR

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Yönetim Bilişim Sistemleri Anabilim Dalı

Yönetim Bilişim Sistemleri Programı

Günümüzde gelişen sosyal ağ kültürü ve bilgisayar teknolojilerinin etkisi ile milyonlarca kullanıcının kullandığı ürün veya hizmetle ilgili memnuniyet düzeyleri ölçülebilmektedir. Bu veriler pazarlama stratejileri, reklamcılık ve müşteri desteği alanlarında aktif olarak kullanılmakta, verilerin analizleri ile organizasyonların alabileceği aksiyonlardaki riskler ve faydalar ölçülmektedir. Alınacak karara destek olması sebebi ile duygu analizleri karar destek sistemleri arasında görülmektedir.

Bu çalışmada Python programlama dili ile Twitter'daki yeni WhatsApp gizlilik sözleşmesi hakkındaki yorumların duygu skorları ölçülerek zaman bazlı duygu skoru değişimi analiz edilmiştir. Türkçe dilinin algılanması için ön eğitilmiş doğal dil işleme modellerinden olan BERT dil modelinin Türkçe dili için geliştirilmiş hali olan BERTurk modeli kullanılarak Türkçe metin verilerinin algılanması sağlanmıştır. Metin sınıflandırması için yapay sinir ağları modellerinden çok katmanlı algılayıcı ile bir model tasarlanmıştır. Twitter'dan WhatsApp yeni gizlilik sözleşmesine ilişkin atılan tweetler elde edilerek duygu analizi yapılmış ve duygu skorunun zaman içerisindeki değişimi gözlemlenmiştir.

Anahtar Kelime: BERT, Çok Katmanlı Algılayıcı, Duygu Analizi, Yapay Sinir Ağları.

ABSTRACT

Master's Thesis

Time Based Sentiment Analysis Using Artificial Neural Networks and BERT Language Model: Exploring Comments on WhatsApp's New Privacy Policy

Kazım Tibet SAR

Dokuz Eylül University

Graduate School of Social Sciences

Department of Management Information Systems

Management Information Systems Program

Today, with the effect of developing social network culture and computer technologies, the satisfaction levels of millions of users about the product or service they can be measured. These datas are actively used in the fields of marketing strategies, advertising and customer support, and the risks and benefits of the actions that organisations can take can be measured by analyzing these datas. Sentiment analysis is seen among decision support systems because it supports the decision to be taken.

In this study, the time-based sentiment score change was analyzed by measuring the sentiment scores of the comments about the WhatsApp new privacy policy on Twitter with the Python programming language. The perception of Turkish text data is provided by using the BERTurk model, which is the developed version of the BERT language model, which is one of the pre-trained natural language processing models for the perception of the Turkish language. In order to classify the texts, a model with a multilayer perceptron, one of the artificial neural network models, was designed. Sentiment analysis was performed by obtaining tweets from Twitter regarding the WhatsApp new privacy policy, and the change in sentiment score over time was observed.

Keywords: BERT, Multilayer Perceptron, Sentiment Analysis, Artificial Neural Networks.

**YAPAY SİNİR AĞLARI VE BERT DİL MODELİ KULLANILARAK
ZAMAN BAZLI DUYGU ANALİZİ: WHATSAPP YENİ GİZLİLİK
SÖZLEŞMESİNE YÖNELİK YORUMLARIN ARAŞTIRILMASI**

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
KISALTMALAR	ix
TABLolar LİSTESİ	x
ŞEKİLLER LİSTESİ	xi
EKLER LİSTESİ	xii
GİRİŞ	1

**BİRİNCİ BÖLÜM
METİN MADENCİLİĞİ**

1.1. TANIMLAR	5
1.2. METİN MADENCİLİĞİNİN GİRDİLERİ VE ÇIKTILARI	6
1.3. METİN MADENCİLİĞİNİN AŞAMALARI	7
1.3.1. Veri Ön İşleme	8
1.3.2. Bölme	8
1.3.3. Özellik Çıkarma	8
1.3.4. Sınıflandırma	9

**İKİNCİ BÖLÜM
DOĞAL DİL İŞLEME**

2.1. TANIMLAR	10
---------------	----

2.2.	DOĞAL DİL İŞLEME İLGİ ALANLARI	10
2.3.	KELİME TEMSİLLERİ	12
2.3.1.	BERT	14
2.3.2.	WordPiece Tokenizasyon Modeli	17
2.3.3	Dönüştürücü (Transformer)	19
2.3.4.	BERTurk Dil Modeli	23

ÜÇÜNCÜ BÖLÜM

YAPAY SİNİR AĞLARI

3.1.	TANIMLAR	24
3.2.	BİYOLOJİK SİNİR AĞI	25
3.3.	YAPAY SİNİR HÜCRESİ	26
3.4.	YAPAY SİNİR AĞLARINDA ÖĞRENME YAPILARI	31
3.5	ÇOK KATMANLI ALGILAYICI	32
3.5.1.	Çok Katmanlı Algılayıcıların Yapısı	33
3.5.2.	Çok Katmanlı Algılayıcıların Öğrenme Kuralı	35
3.5.2.1.	İleriye Doğru Hesaplama Aşaması	35
3.5.2.2.	Geriye Doğru Hesaplama Aşaması	36
3.5.2.3.	Öğrenme Hızı ve Momentum Katsayısı	36
3.5.3.	Geriye Doğru Ağırlık Hesaplaması	38
3.5.4.	Çok Katmanlı Algılayıcıların Çalışma Prosedürü	40

DÖRDÜNCÜ BÖLÜM

WHATSAPP YENİ GİZLİLİK SÖZLEŞMESİNE YÖNELİK YORUMLARIN

DUYGU ANALİZİ

4.1.	YAZILIM GELİŞTİRME ORTAMI	41
4.2.	VERİ SETİ OLUŞTURMA	42
4.2.1.	Google Play Store'dan Veri Çekme	43
4.2.2.	Twitter'dan Veri Çekme	43
4.3.	VERİLERİN ÖN İŞLEMESİ	44

4.4.	BERT MODELİNİN KURULMASI	45
4.5.	ÇKA MODELİNİN KURULMASI	46
4.6.	ANALİZ SÜRECİ	48
SONUÇ		54
KAYNAKÇA		64
EKLER		



KISALTMALAR

CRM	Customer Relationship Management (Müşteri İlişkileri Yönetimi)
DLP	Data Loss Prevention (Veri Sızıntısı Önleme)
DDİ	Doğal Dil İşleme
BoW	Bag of Words (Kelime Çantası)
CBOW	Continuous Bag of Words (Sürekli Kelime Çantası)
BERT	Bidirectional Encoder Representation from Transformers
MLM	Masked Language Model (Maskeli Dil Modeli)
YSA	Yapay Sinir Ağları
Tanh	Hiperbolik Tanjant
ReLU	Rectified Linear Unit (Doğrultulmuş Doğrusal Birim)
ÇKA	Çok Katmanlı Algılayıcı
PaaS	Platform as a Service (Hizmet Olarak Platform)
CSV	Comma-Separated Values (Virgülle Ayrılmış Değerler)
JSON	JavaScript Object Notation (JavaScript Nesne Gösterimi)
CUDA	Compute Unified Device Architecture (Birleşik İşlem Cihaz Mimarisi)
GPU	Graphics Processing Unit (Grafik İşleme Ünitesi)

TABLÖLER LİSTESİ

Tablo 1: Word2Vec Modelinde Kullanılan Pencere Yapısı	s.14
Tablo 2: Yapay Sinir Ağları Toplama Fonksiyonları	s.28
Tablo 3: XOR Problemi Girdi ve Çıktı Tablosu	s.33
Tablo 4: Elde Edilen Tweet Sayısı ve Tweetlerin Tarihleri	s.44
Tablo 5: Konfüzyon Matrisi Örneği	s.47
Tablo 6: Eğitim, Test ve Değerlendirme Veri Setlerindeki Veri Sayısı	s.49
Tablo 7: Tasarlanan Modelin Başarı Değerleri	s.49
Tablo 8: Modelin WhatsApp Tweetleri İçin Ürettiği Duygu Sınıfı Tahminleri Arasından Bazı Örnekler	s.52
Tablo 9: Modelin Ürettiği Tahminler Sonrası Duygu Sınıflarının Tweet Sayıları	s.52
Tablo 10: Çılgın vd. Sınıflandırma Modeli Başarı Karşılaştırması	s.56
Tablo 11: Kaynar vd. Sınıflandırma Modeli Başarı Karşılaştırması	s.57
Tablo 12: Tuzcu Sınıflandırma Modeli Başarı Karşılaştırması	s.57

ŞEKİLLER LİSTESİ

Şekil 1: BoW Yöntemindeki Matrisleme Örneği	s.13
Şekil 2: Google Arama Motorunda BERT Modeli Kullanımı	s.15
Şekil 3: BERT Bağlamsal Kelime Gösterimi Çözümü İçin Bir Örnek	s.16
Şekil 4: WordPiece Tarafından “Temsillemek” Kelimesinin ve Anlamı Olmayan Kelime Öbeğinin Sözcük Parçacığına Ayrıştırılmış Hali	s.17
Şekil 5: WordPiece Kelime Temsilleme Yapısı	s.18
Şekil 6: Dönüştürücü (Transformer) Mimarisi	s.20
Şekil 7: “O bir yeşil elma yiyor” Cümlesinin Bazı Kelimeleri Arasında İlişki Düzeyi Görselleştirmesi	s.21
Şekil 8: Bir Biyolojik Sinir Hücresinin Yapısı	s.25
Şekil 9: Biyolojik Sinir Hücresi Bağlantısı ve Bilgi Aktarımı	s.26
Şekil 10: Tek Katmanlı Yapay Sinir Hücresi	s.27
Şekil 11: Doğrusal Aktivasyon Fonksiyon Grafiği	s.29
Şekil 12: Sigmoid Aktivasyon Fonksiyon Grafiği	s.30
Şekil 13: Hiperbolik Tanjant Aktivasyon Grafiği	s.30
Şekil 14: ReLU Aktivasyon Grafiği	s.31
Şekil 15: Bir ÇKA Modelinin Katman Yapısı ve Nöronların Bağlantı Yapısı	s.34
Şekil 16: Google Colab Çalışma Ortamı	s.41
Şekil 17: Uygulamanın Akış Diyagramı	s.42
Şekil 18: Google Play Store URL’si	s.43
Şekil 19: Tasarlanan Modelin Hata Değerlerinin Eğitim Değişim Tablosu	s.50
Şekil 20: Tasarlanan Modelin Konfüzyon Matrisi	s.50
Şekil 21: Duygu Skorları Ortalamasının Tarihlerle Göre Değişim Grafiği	s.51
Şekil 22: WhatsApp Uygulaması Tarafından İletilen Yeni Gizlilik Sözleşmesi	s.55
Şekil 23: Duygu Sınıflarının Tweet Yüzdeleri	s.60
Şekil 24: WhatsApp ve Diğer Mesajlaşma Uygulamalarının Günlük Aktif Kullanıcı Sayılarındaki Değişim Grafiği	s.61

EKLER LİSTESİ

EK 1: Google Play Store Yorum Kazıma Kodlarının Bir Kısım1	ek s.1
EK 2: Twitter Üzerinden Tweet Kazıma Kodlarının Bir Kısım1	ek s.1
EK 3: Sınıflandırma Analizi Kodlarının Bir Kısım1	ek s.2
EK 4: Tahmin Edilen Tweetler ve Tahminlerin Bir Kısım1	ek s.3



GİRİŞ

Çağımızın odağını belirleyen bilgi kavramı, bilgisayar teknolojileri ve depolama kapasitelerinin gelişmesi ile hızla hayatı, insan davranışlarını ve alışkanlıkları değiştirmektedir. Bu değişimle beraber bireylerin sosyal yaşantılarındaki en büyük değişimler ise sosyal medya mecralarının gelişimi ile birlikte yaşanmaya başlamıştır. İnteraktif bir sosyal medya mecrasında davranış ve görüşler ile birlikte varlık sürdürebilmek, daha büyük kitlelere hitap edebilmek ve interaktif olarak farklı görüşlerden haberdar olmak, tüketici davranışlarının değişmesine yol açmıştır. Üreticiler açısından ise rekabet, bu görüşlere ve inovatif yöntemlere bağlı olarak değişim göstermektedir. Şirketlerin sosyal medya hesapları ve bu hesaplar ile ilgili çalışmalar yürüten ekipleri ile bu alanda çalışmaları üzerine rekabet sahası, imkanlar dahilinde tüketicilerin görüşleri üzerine atılacak adımlar sonucu rakiplerden daha yüksek bir memnuniyet düzeyi sağlamaları alanına kaymıştır. Sosyal medyanın etkileri ile tüketicilerin tercih ve değerlendirme yargılarının değişimi sonucu ürün/hizmet tercihinde sosyal medya yorumlarını değerlendirerek tercihte bulunabilmek de şirket ve kurumları bu yola sevk etmiştir.

Büyük veri çalışmalarının gelişmesi ile beraber metin madenciliği de çok büyük bir önem kazanmıştır. Metin madenciliği ile duygu analizi yapılarak şirket veya kurumlara karşı tüketicilerin tutumları analiz edilebilmekte, şirket veya kurumların aldığı aksiyonlara karşı tüketici tutumu incelenebilmekte ve bu veriler ışığında doğru kararın verilebilmesi konusunda destekleyici analiz verileri oluşturulmaktadır.

Bu çalışmada WhatsApp yeni gizlilik politikası uygulamasına karşın kullanıcıların WhatsApp hakkındaki duygu değişikliğini zaman bazlı ölçümleyerek WhatsApp günlük kullanıcı sayısı değişimi ile kıyaslamak ve alınacak yeni aksiyonların kararı verilirken alınabilecek doğru kararların seçilmesine duygu analizi üzerinden destek sağlayacak bir karar destek mekanizması kurmaya çalışılmıştır. Karar destek mekanizması olarak duygu analizi çalışması yapılmış ve literatürde Türkçe duygu analizi çalışmaları ile ön eğitilmiş BERT dil modeli alanına katkı sağlamak amaçlanmıştır.

Literatürde İngilizce dilinin esas alındığı duygu analizi çalışmalarının çok büyük bir yer kapladığı görülmektedir. İngilizce dilinin kuralları ve yapısının birçok

dile göre sade ve daha kolay olması, dil işleme ve anlama aşamasında çalışmaların daha kolay bir şekilde üretilebilmesine olanak tanımaktadır. Bunun yanı sıra İngilizce dili ile oluşturulmuş, analizde kullanılmaya uygun veri setleri internetten kolayca erişilebilmekte ve oldukça fazla sayıda mevcuttur. İngilizce veri setlerinin kolay erişilebilir olması ve analiz modellerinde veri setlerinin doğru kurgulanmasının önemi düşünüldüğünde bu olanak İngilizce dili için hazırlanan çalışmaların kolaylaşmasını ve fazlalaşmasını mümkün kılmaktadır.

Türkçe veri setleri ile yapılan çalışmaların son yıllarda artış gösterdiği söylenebilmesine karşın Türkçe dili ile yapılan metin madenciliği çalışmaları literatürde yeteri kadar mevcut değildir. Özellikle Türkçe literatürde BERT dil modeli ile yapılmış çalışmalar yok denecek kadar azdır. Bu çalışmaların amaçlarından en önemlisi, Türkçe literatür için BERT Türkçe dili üzerindeki çalışmalara kaynak kazandırmaktır.

Büyük verinin yaygın olarak kullanıldığı sosyal mecralar bu analiz yöntemlerinin literatürdeki en önemli kaynaklarıdır çünkü sosyal medya hacim ve miktar olarak görüş ve duygu çıkarımı yapabilecek metin verilerini barındırmaktadır. Bu mecralardan Twitter en sade şekilde görüş ve duygu analiz çalışması yapılabilecek büyük veri hacmine sahip sosyal medya mecralarıdır. Twitter API ve diğer web kazıma yazılımları sayesinde verilere erişim çok daha kolay hale gelmiştir.

Çılgın ve arkadaşları tarafından 2020 yılında yapılan çalışmada Bitcoin hakkında Twitter'dan kazıma kütüphanesi kullanılarak elde edilen toplam 2.819.784 adet İngilizce tweet içerisinde 1500 adet tweet duygu durumlarına göre pozitif, negatif ve nötr olarak işaretlenmiştir. Bu veriler ile tweetlerin duygu durumu, Çok Katmanlı Algılayıcı ve diğer sınıflandırma algoritmaları ile analiz edilip başarı oranları test edilmiştir. Çok Katmanlı Algılayıcı ile geliştirilmiş olan model 0.9058'lik başarı oranı ile en iyi performansı gösteren model olmuştur (Çılgın ve diğerleri, 2020: 106-126).

Tuzcu'nun gerçekleştirdiği çalışmada, popüler bir çevrimiçi kitap satış sitesinden 47 adet kitaba ait 91.309 adet yorum toplanmış ve bu yorumlardan alınan 1400 yorumun, 700 pozitif, 700 negatif yorum olacak şekilde işaretlenmesi ile oluşturulan bir eğitim veri seti oluşturulmuştur. Sınıflandırma algoritmaları ve Çok

Katmanlı Algılayıcı modeli ile verilerin duygu sınıflandırmasını gerçekleştiren Tuzcu, %89 başarı oranı ile Çok Katmanlı Algılayıcı'nın performansının diğer algoritmalarından daha yüksek olduğunu kaydetmiştir (Tuzcu, 2020: 1-5).

Borele ve Borikar 2016 yılında yaptıkları çalışmada marka itibarı, e-ticaret kullanıcılarının görüşleri ve ürünler hakkındaki yorumların duygularının tespiti için Destek Vektör Makineleri, K- En Yakın Komşu, Naive Bayes ve Yapay Sinir Ağları modellemeleri ile metin sınıflandırma çalışması yapmışlardır. Destek Vektör Makineleri ve Yapay Sinir Ağları modelleri başarılı olurken, Destek Vektör Makineleri'nin daha kısıtlı ölçeklerde çalışabildiği, Yapay Sinir Ağları modelleri ile çalışılarak daha büyük ölçeklerdeki projelerde başarılı sonuçlar elde edilebileceğini savunmuşlardır (Borele ve Borikar, 2016: 64-69).

Pak ve Paroubek yaptıkları çalışmada, Twitter API yazılımı kullanarak otomatik bir şekilde tweetleri edinerek, tweetlerde belirtilmiş olan gerçek ifadelerin duygu durumunu açıklamak için bir kelimeler listesi (korpus) geliştirmişlerdir. Anlamlandırdıkları tweet metinlerini makine öğrenmesi alanında kullanılan sınıflandırma algoritmaları ile pozitif, negatif, nötr duygu sınıfları çerçevesinde sınıflandırarak, tweetlerin hangi duyguyu barındırdığını analiz etmişlerdir (Pak ve Paroubek, 2010: 1320-1326).

Akgül ve arkadaşları, ülkemizde özellikle Türkçe dilinin kuralları ile algılama yaparak çalışabilecek metin madenciliği yazılımının gerektiği kadar bulunmadığını savunarak Twitter'dan tweet verilerini alan ve bu verileri duygu sınıflarına göre etiketlendiren "Duygusal Twitter" adlı yazılımı geliştirmişlerdir. Yazılım içerisindeki korpusu Türk Dil Kurumu'nun kelime listelerinden alınmış olup, n-gram ve sözlük tabanlı analiz yöntemleri ile duygu analizi gerçekleştirilmesini sağlamışlardır (Akgül ve diğerleri, 2015: 106-110).

Akhtar ve arkadaşları, yapay sinir ağları modelleri ile finansal haberlerin duygu analizini gerçekleştirmişlerdir. Kelime gösterimleri kısmında ön eğitilmiş dil modellerinden Word2Vec ile GloVe'nin kullanılması ile birlikte 0,78-0,79 arasında ortalama başarı puanı elde edilerek bu şekilde hisse senedi değerlerinin tahmin edilebileceğini savunmuştur (Akhtar ve diğerleri, 2017: 540-546).

Müller ve arkadaşları, açık kaynak kodlu ön eğitilmiş bir dil modeli olan BERT modelinden, eğitiminde 12 Ocak – 16 Nisan 2020 tarihleri arasında atılmış 160 milyon

COVID-19 ile ilgili tweet kullanarak oluřturdukları CT-BERT modeli ile Twitter üzerindeki ařı hakkındaki duygu analizinin saęlanabilmesi iin pozitif, negatif ve ntr olarak iřaretledikleri “Vaccine Sentiment” (ařı duygu durumu) veri setini sınıflandırma probleminde faydalanmak zere modelin eęitiminde kullanmıřlardır. Orijinal BERT (large) modeli 0.824 bařarı performansı saęlarken CT-BERT modeli 0.869 oranda bařarılı olmuřtur (Mller ve dięerleri, 2020: 1-5).

Kaynar ve arkadaşları tarafından 2016 yılında yapılan bir alıřmada IMDB’de yer alan Trke film yorumlarını ieren veri seti zerinde eřitli makine ęrenmesi teknikleri kullanılarak bir duygu analizi alıřması yapılmıřtır. Film yorumları zerinde yapılan alıřmada destek vektr mekanizmaları, naive bayes, merkez tabanlı sınıflayıcı ve yapay sinir aęları modellerinden ok katmanlı algılayıcı modeli kullanılmıř, ok katmanlı algılayıcı modeli %89.73 doęruluk oranıyla en iyi bařarı oranını saęlamıřtır (Kaynar ve dięerleri, 2016: 234-241).

BİRİNCİ BÖLÜM

METİN MADENCİLİĞİ

1.1 TANIMLAR

Rekabetçi bir dünyada kurumlar rekabet güçlerini arttırmak amacıyla büyük veri analizine yönelmişlerdir. Merrill Lynch ve Gartner tarafından yapılan bir çalışmaya göre, kuruluşlardaki verilerin %85'i yapılandırılmamış metin verileri olarak muhafaza edilmektedir (Sharda ve diğerleri, 2014: 321). Günümüz işletmeleri doğru kararlar alarak rekabette öne geçebilmek için yapılandırılmamış metin kaynaklarını, metin madenciliği yöntemlerini kullanarak analiz ederler.

Metin madenciliği geniş anlamda bir kullanıcının bir dizi analiz aracı kullanarak zaman içerisinde oluşturulan bir belge koleksiyonuyla etkileşime girdiği bilgi yoğunluklu bir süreç olarak tanımlanabilir (Feldman ve Sanger, 2006: 1). Metinlerin birer veri olarak kabul edildiği ve yapılandırılarak analiz edilebildiği tekniklerin tamamı metin madenciliğidir. Metin madenciliğinde tıpkı veri madenciliğinde olduğu gibi büyük veri grupları ile çalışarak metinlerin anlamlandırılması işlemleri gerçekleştirilir. Anlamlandırılan metinler analiz edilerek nitelikli, ölçeklendirilebilir ve somut anlamlar taşıyan bilgilere ulaşılır.

Metin madenciliği, metinden eyleme dönüştürülebilen içgörülerin damıtılması sürecidir (Kwartler, 2017: 1). Metinsel ifadelerin birer veri olarak kabul edilip, yapılandırılarak somut çıkarımlar elde edilmesi süreci metin madenciliğidir. Yapılan madencilik işlemleri ile metinlerden somut bir sonuç çıkarılması gereklidir. Bu sebeple aranan somut bir sonucun bulunması için metinler içinde madencilik yapılmaktadır. Yapılan bu madencilik çalışması ile elde edilen anlamlı bilgiler, karar alma süreçlerinde destek olmak için kullanılmaktadır. Madencilik terimi mecazi olarak aranan bir sonucun bulunma sürecini ifade etmektedir.

Büyük ölçekteki, tek başına anlamsız olan, ancak analiz teknikleri ile bilgiye dönüştürülen verilerin anlamlandırılması sürecine veri madenciliği ismi verilirken, yine büyük ölçüdeki metin formatındaki verilerin analiz edilerek anlamlandırılması ve ilişkilendirilmesi süreci metin madenciliği sürecidir (Üçükkartal, 2020: 10). Bu bağlamda basitçe açıklamak gerekirse metin madenciliği çalışmaları, veri kaynağı

olarak metin formatındaki verileri kullanarak yapılandırılmış bilgi elde etmeyi amaçlayan veri madenciliği çalışmalarıdır. Metinlerden anlaşılabilir verilerin çıkarılması için veri madenciliği yöntemleri kullanılarak geliştirilmiş modeller ile çıkarımlar yapılmaktadır.

Günümüzde sadece çevrimiçi mecralardaki yapısal olmayan veri boyutu birkaç zetta bayt boyutuna erişmiş ve hızla da artış göstermektedir (Guha, 2017: 11). Twitter gibi mecralardaki metin verilerinin hızla artışına karşılık günümüz teknolojik gelişmeleri daha fazla kapasiteye sahip bellek üretimine yoğunlaşmıştır. Verilerin yapısal bir form içerisinde analiz edilmesinin sonucunda şirketler gerçekleştirdikleri faaliyetlerin daha etkili bir şekilde yürütülmesini sağlamaktadır. Öte yandan bu veriler ile duygu analizi yapılarak yorum sahiplerinin duyguları da analiz edilmekte ve bu bilgi ile kararlar alınmaktadır.

Özetlemek gerekirse; Metin madenciliğinin odağındaki problem, yapısal olarak net verilere çevrilmemiş, yani yapılandırılmamış verilerin niceliğindeki artış ve teknolojik gelişmeler sonucunda bu verilerden sağlanacak nitelikli bilgilerin çeşitli analiz yöntemleri ile edinilmesidir. Metin verilerinin yapılandırılması, metinlerin özelliklerine göre analiz yöntemlerinin, metinlere bir anlam atfedebilmesi anlamına gelmektedir. Bu işlemde bilgisayar teknolojilerinden yararlanılması sebebiyle metinsel ifadeler sayısal olarak temsil edilmekte ve bu temsil sonucunda analiz edilerek ihtiyaç duyulan somut bir bilgi çıkarımı sağlanmalıdır. Bu süreçler genel bir ifade ile metin madenciliği süreçleridir. Veri madenciliği işlemlerinin metin veri türüne göre uyarlanıp gerçekleştirildiği süreçler olarak da tanımlanmaktadır.

1.2 METİN MADENCİLİĞİNİN GİRDİLERİ VE ÇIKTILARI

Yazı formatındaki kaynakların tamamı ve yazılı olmayan resimler, haritalar gibi kaynaklar da metin madenciliği sürecinin girdileri olabilir. Yazı formatındaki kaynaklardan e-mailler, gazeteler, makaleler, internet siteleri, hasta raporları, adli sicil kayıtları, sosyal medya mecraları gibi herkesin erişebileceği kaynaklar, metin madenciliğinde en çok kullanılan kaynaklarıdır (Kızılkaya, 2018: 4).

Kaynakların analizi sonucunda aranan bir verinin bulunması veya bir format değişikliği yapılması da metin madenciliği süreci sayılmaktadır. Bir dokümanın

özetinin çıkarılması, metindeki özne, nesne ve değer bilgilerinin ayıklanması süreçleri de metin madenciliği alanındaki çalışmalardandır. Metin ile alakalı girdiden farklı olarak üretilen ve gerekli ihtiyaçları karşılayacak biçimde bir bilgi olarak kullanılabilecek çıktılar metin madenciliği işlemlerinin ürünüdür.

Girdilerin metin madenciliği süreci ile işlenmesinin ardından elde edilen çıktılar nitelikli bilgi olarak tanımlanır ve girdilerin niteliğine göre değişiklik göstermektedir (Melek, 2012: 26). Müşteriler ile gerçekleştirilen anket çalışmalarının metin madenciliği sürecinden geçirilerek elde edilen nitelikli veriler, müşteri ilişkileri yönetimi (CRM) için, pazar analizi, müşterinin terk etme ve çapraz satış işlemleri bilgilerini oluşturabilir. Sosyal medya üzerinde bir şirket veya bir şahsiyet adına yazılan metinlerin analizinden, duygu analizi, e-mail metinlerinin analizinden spam mail filtreleme veya veri kaybı önleme (DLP) yazılımları içerisinde gizli kalması gereken özel verilerin kaynak dışına sızdırılıp sızdırılmadığının tespiti sağlanmaktadır (Melek, 2012: 26).

1.3 METİN MADENCİLİĞİNİN AŞAMALARI

Metin madenciliğinin çıktılarının hedeflendiği konu ya da analiz edilecek metinlerin toplandığı aşamadır. Bu aşamada araştırma konusuna yönelik olan verilerin seçimi, yapılan analizin sonucuna doğrudan etkide bulunduğu için önemlidir (Kızılkaya, 2018: 8). Verilerin toplanma aşamasındaki en olası sorun araştırma sorusu ile alakalı olan verilerin toplanması konusudur. Bu veriler ile bir model oluşturulacağından doğru veri kümeleri oluşturmak doğrudan modelin tahmin etme başarısını etkileyecek bir kriterdir. İlgili probleme karşılık bağlantısı olmayan metinler ile veri kümeleri hazırlamak yapılacak analizin başarısız olmasını sağlayacaktır.

İnternet siteleri ve dijital ortamlardan verileri indirebilmek için oluşturulmuş paket programlardan ya da kurumların paylaşmış oldukları verilerinden yararlanarak veri setleri oluşturulmaktadır. Bunun yanı sıra veri seti olarak sosyal medya mecraları (Twitter, Facebook v.b) ile yorum portalları gibi mecralar hem interaktif hem de gerçek kişilerin görüşlerini yansıttığı için veri kaynağı olarak sıkça kullanılmaktadır. Metin madenciliği; veri ön işleme, bölme, özellik çıkarma ve sınıflandırma olmak üzere dört aşamadan oluşmaktadır.

1.3.1 Veri Ön İşleme

Veri seti oluşturma aşamasında elde edilen metinleri kullanabilmek için öncelikle metinlerin düzenlenmesi gerekmektedir. Bu düzenleme metinlerde bulunan yazım yanlışlarını, kullanılmayacak olan verileri (İnternet sayfası linkleri, kullanıcı adı v.b), kısaltma olarak kullanılan kelimelerde düzenlemeler yaptıktan sonra kullanıma hazır hale getirilir (Kuzucu, 2015: 13). Yapılandırılmamış metin formatındaki verilerin, metin madenciliği süreçlerinde işlenerek yapılandırılmış hale getirilebilmesi için gereken ilk adım ön işleme aşamasıdır. Bu adımda, metinlerdeki gereksiz içerikleri çıkarma, kısaltma olan kelimeleri doğru kelimeler ile değiştirme, kelimeleri sözcük parçacıklarına bölme (tokenizasyon) işlemleri uygulanarak veri analiz edilebilir hale getirilir (Çınar, 2020: 108).

1.3.2 Bölme (Tokenizasyon)

Bölme, bir cümleyi, kelimelerin anlamlarını etkileyip etkilemediklerine göre kategorize edilmiş metin parçacıklarına bölerek cümlenin bu kategorize edilmiş metin parçacıkları üzerinden işlenmesini sağlar (Sharda ve diğerleri, 2014: 324). Bölünmüş metin parçaları (token) cümlenin içerisinde bir işleve sahip olmalıdır. Kelimenin anlamına etki eden ekler, noktalama işaretleri ve kelimenin kökü cümlede işlevsel olarak bir etkiye sahip oldukları için bölünerek ele alınır. Bölme işlemi uygulanması, metin verilerinin doğal dil işleme çalışması aşamasında sözcük parçacıklarının anlamları üzerinden bilgisayarların anlayacağı şekilde temsil edilmesi için önemlidir.

1.3.3 Özellik Çıkarma

Metin madenciliği uygulamalarında her zaman gürültülü ve önemsiz bilgi içeren metin koleksiyonlarıyla uğraşma ihtiyacı bulunmaktadır. İlgili verilerin saptanması üzerine odaklanan özellik seçme, büyük miktarlardaki veriler üzerinde işlem yapılırken iş yükünü azaltmada yardımcı olmaktadır (Ergün, 2012: 51). Sınıflandırma gibi algoritmalarda eğitim yapılan veri kümesine, tahmin edilecek

verilerin uyumunun aşırı olabileceği durumlara sebep olan yüksek boyutlu veri ile yapılan çalışmalarda tahmin performansı düşebileceği için özellik çıkarma gibi küçük boyutlu çalışma ayarlamaları yapılmaktadır. Metinlerin birbirinden ayrıştırılmasını sağlayacak özel niteliklerin çıkarılması bu aşamada gerçekleştirilmektedir.

Ayrıca özellik çıkarma işlemleri, metin madenciliğindeki yapılaşdırma işlemlerinin de içerisinde barındığı bir aşamadır. Bilgisayar teknolojilerinin ürünleri metinleri sayısal değerler atfederek algılamaktadırlar. Bu durumda metin formatındaki verilerin makineler ve bilgisayarlar tarafından algılanabilmesi için sayısal bir karşılığa eşit olmaları gerekmektedir. Özellik çıkarımı işleminde kullanılan dil işleme yöntemleri ile metinlerdeki özelliklere göre bu sayısal temsilleme işlemleri gerçekleştirilerek bilgisayarların sayısal veriler üzerinden gerekli çalışmaları gerçekleştirebilmeleri sağlanmaktadır.

1.3.4 Sınıflandırma

Bu aşamada girdilerin özelliklerini niteleyen veriler analiz edilmeye uygun haldedir. Analiz için gerekli işlemin seçilip verilerin üzerinde uygulanması aşamasıdır. Sınıflandırma, bu analiz işlemlerinden biridir. Analiz edilmesi istenilen verilerin önceden tanımlanmış veri gruplarından hangileri ile uyumlu, aynı sınıftan olduğunun çıkarımı için kullanılan bir analiz yöntemidir (Özkan, 2020: 32).

Sınıflandırma işlemleri, eğitim için oluşturulan veri kümesinden uyumluluğu tespit edebileceği sınıflandırma kurallarını öğrenir ve bu kurallar yardımı ile analiz edilmesi istenen, sınıfı belli olmayan verileri sınıflandırır. Model bu öğrenilmiş sınıflandırma yapısını kullanarak, daha önce deneyim sahibi olmadığı veriler ile sınıflandırma problemlerini çözmektedir. Benzer nitelikli olan verilerin aynı sınıflarda toplanarak bu sınıflandırma üzerinden nitelikli veriler elde edilen aşamadır.

İKİNCİ BÖLÜM

DOĞAL DİL İŞLEME

2.1 TANIMLAR

Diller, insanların yaşadığı mesken ve kültürlerine göre organik olarak gelişen bir iletişim kuralları silsilesidir. Bu kurallar iletişimin eksiksiz tamamlanmasını sağlamak için çok detaylı olarak gelişmiştir (Adalı, 2016: 45).

Bu detayları algılamak ve doğru nitelikte anlamlandırmak için bilgisayarların normal çalışma sistemleri yetersiz kalmaktadır. Doğal Dil İşleme (DDİ) çalışmalarının hedefi ise dilin karmaşık kurallarının makine tarafından doğru değerlendirilerek, dilin yapısına uygun algılanmasını sağlamaktır. Bu kuralların doğru bir şekilde algılanarak makineler tarafından öğrenilmesi, metinlerin veya seslerin algılanarak bu verilerden çıkarımlar yapılmasına olanak sağlamaktadır.

DDİ çalışmaları, bilgisayar teknolojilerinin gelişmesi ile ana odağı olan insan-bilgisayar etkileşiminin daha verimli bir biçimde gerçekleştirilmesine olanak sağlamıştır (Şeker, 2015). İnsanların iletişim aracı olan dili makinelerin algılayabilmesi için makinenin bu dil hakkındaki detayları ve kuralları tam olarak öğrenmiş olması gerekmektedir. Kısacası dilin özellikleri bilgisayara öğretilmiş olmalıdır. Bu çalışmalar geliştikçe DDİ çalışmaları, bilgisayarlı dil bilimine evrilmiştir. DDİ çalışmalarında, dil bilimciler ile doğal dil işleme bilişim çalışanları ayrı ayrı ve birlikte çalışmalar gerçekleştirmişlerdir. Çalışmalarında birbirlerinin ulaştığı sonuçlara ihtiyaç duydukları için, beraber çalışmalarında daha faydalı sonuçlar elde edilmiştir.

2.2 DOĞAL DİL İŞLEME İLGİ ALANLARI

DDİ günlük hayatta birçok alanda kullanılmaktadır (Adalı, 2016: 46). Bu uygulamalar başlıca şu şekildedir:

- **Yazım yanlışlarının düzeltilmesi:** Metinler içerisinde imla hatası veya yazım yanlışlığı bulunan cümleler ve kelimeler, DDİ çalışmaları ile tespit edilebilmektedir. Bu çalışmalar büyük boyutlu dokümanlarda pratiklik sağlamaktadır.

- **Büyük metin dosyalarının içerisinde konuyu kapsayacak ölçekte özet çıkarma:** Dokümandaki doğal dilin algılanıp, metinlerin anlamlandırılabilmesi ile metinde kullanılan objeler, özneler ve fiiller anlamlandırıldığı için dokümanın kapsamı ve genel yapısını bozmadan özetleme yapılmaktadır.
- **Büyük metin dosyalarının içerisindeki ana bilginin tespiti:** Metnin içerisinde geçen özel isimleri, eylemleri, zaman ve yer gibi olguları veri olarak çıkararak pratiklik sağlamaktadır. DDİ çalışmaları aranan bu verileri metin içerisinde taradığında anlayabildiği ve aranan veri olarak işaretleyebileceği için sıkça kullanılmaktadır.
- **Talep içeren metnin içerdiği talebi anlamlandırma:** Kurumlara veya yapay zeka ile yönetilebilen cihazlara yazılı veya sözlü olarak doğal dil ile talep belirtildiği durumlarda, metindeki talebin yine dilin algılanması üzerinden algılanıp talebin anlaşılmasını sağlamaktadır.
- **İnternet tarayıcılarında kullanılan arama motorları:** Arama motorlarında aratılmak istenen sorgu kelimesi veya cümlesine istinaden aranılan içeriğe yönelik ilgili internet sayfalarının arama sonucu olarak getirilmesinde DDİ çalışmaları kullanılır. İlgili olan sonuçların getirilebilmesi için sorgunun ve sitelerdeki metin ifadelerinin doğal dil bağlamında bilgisayar tarafından anlaşılabilmesi gerekir ve bu alanda DDİ çalışmaları ile probleme çözüm getirilebilmektedir.
- **Yazılı metin seslendirme ve konuşmayı metne dönüştürme:** Metinlerin konuşmaya çevrilmesi ve konuşmaların metin olarak dönüştürülmesi alanında dili algılayarak metinde veya konuşmada geçen kelimelerin anlaşılacak doğru bir şekilde metin veya ses olarak temsili oluşturulmaktadır. Özellikle video yayın sitelerinde doğru altyazılar oluşturulmasında, kitap gibi dokümanların sesli olarak otomatik okutulmasında ve sesli komut uygulamalarında sıkça DDİ çalışmaları kullanılmaktadır.
- **Doğal diller arasında çeviri yapma:** Çeviri uygulamalarında arama motorlarındaki kullanım şekli ile çok benzer bir şekilde DDİ çalışmaları yapılmaktadır. Çevrilmesi istenen metin bir sorgu olarak ele alınır ve karşılığında o sorguyu temsil etmesi en olası farklı dildeki çevirisi çıktı olarak getirilmektedir. Özellikle aynı kelimenin farklı anlamları ve çevirileri olabilmesi nedeni ile DDİ

modellerinin cümle içerisindeki anlama bağlı olarak dil algılaması gerçekleştirmesi önemli bir hale gelmiştir.

- **Soru yanıtlama uygulamaları:** Kendisine sorulan soruları yanıtlayan görüşme botları ve yapay zeka çalışmaları içerisinde DDİ uygulamaları kullanılmaktadır. Sorulan sorunun dil bağlamında algılanarak en olası yanıtın otomatik olarak iletilmesi yine bu alandaki çalışmalardandır.

Genel olarak bu çalışmalarda metin verilerindeki her sözcük parçacığının, doğal dil içerisinde bilgisayarlar tarafından algılanması hedef olarak belirlenmektedir. Bilgisayarların metinleri doğrudan algılayamaması bu alandaki en büyük sorunların başında gelmektedir (Kızılkaya, 2018: 27). Bilgisayarlar sayısal değerleri ve sayısal işlemleri algılayabildiği için metinsel verilerin sayısal olarak temsil edilmesi gerekmektedir. DDİ literatüründe bu kısım kelime temsili olarak aktarılmaktadır.

2.3 KELİME TEMSİLLERİ

DDİ çalışmalarının tamamı metinsel ifadeler üzerinden gerçekleştirilmektedir ancak bilgisayar sistemleri metinsel ifadeleri algılayamaz ve bu nedenle DDİ'nin temel işlevi olan dili algılama özelliği metinsel veriler ile sonuçlandırılmaz. Bu durumu değiştirmek için bu çalışmalarda kelime ve kelime parçacıklarını sayısal veriler ile ifade etmemiz gerekmektedir. Bu işlemin adı kelime temsili (Word Embedding) olarak literatürde yer almaktadır (Aydoğan ve Karcı, 2019: 183). Birçok DDİ çalışmasında kelime parçacıklarının (tokenlerin) kelime gösterimleri, önceden tanımlanmış bir derlem (korpus) ile gerçekleştirilmektedir. Bu korpus, hedef dildeki sıklıkla kullanılan kelime parçacıklarının frekans bazlı belirli sayısal değerler ile temsil edildiği, sıralı kelime listesidir. Kelime gösterimlerindeki bu yaklaşımın en çok kullanıldığı türüne kelime çantası (Bag of Words, BoW) adı verilmektedir. BoW yöntemine göre metin içindeki cümlelerin tamamı kelimelere ayrılır ve her kelimenin cümle içerisinde kaç defa kullanılmış olduğuna bağlı olarak rakamlar ile temsil edildiği bir matrise dönüştürülür. BoW yöntemi ile oluşturulan matris ile ilgili bir örnek Şekil 1'de gösterilmiştir.

Şekil 1: BoW Yöntemindeki Matrisleme Örneği

	O	kuş	kedi	kalem	bir	bu
O bir kuş	1	1	0	0	1	0
O bir uçak	1	0	0	0	1	0
O bir kedi	1	0	1	0	1	0
Şu bir kalem ve bu bir köpek	0	0	0	1	2	1
O bir matris	1	0	0	0	1	0

Kaynak: Zheng ve Casari, 2018: 73

Şekil 1'deki gibi matrisin sütunları doküman içerisindeki kelimelerden oluşurken, satırları ise doküman sayısından meydana gelir. Kelimelerin kullanım sıklığından yola çıkılarak cümlelerin anlamlandırılması için çalışmalar gerçekleştirilmiştir.

Kelime gösterimi için bir diğer yöntem ise tahmin bazlı yöntemdir. Bu yöntemde en çok kullanılan iki benzer metod sürekli kelime torbası (Continuous Bag of Words, CBOW) ve skip gram metodlarıdır. İkisi de birbirine çok benzer metodlar olmakla birlikte ön eğitilmiş bir dil modeli olan Word2Vec dil modeli içerisinde sıkça kullanılmaktadır. Word2Vec tahmin işlemini gerçekleştirmek için yapay sinir ağlarıyla çalışmaktadır ve bu ağların önceden eğitilmiş olması da bu modelin ön eğitilmiş bir dil modeli olmasını sağlar (Aydoğan ve Karcı, 2019: 183-184).

Word2Vec modeli, girdi kelimelerinden oluşturulan derlem üzerinde pencere adı verilen bir yapı içerisinde çalışır. Dökümandaki her cümleyi pencere içerisinde tüm derlem üzerinde kaydırarak tarar ve penceredeki hedef kelimeye göre oluşan bir vektörel matrisi çıktı olarak üretir (Aydoğan ve Karcı, 2019: 183). Word2Vec dil modelinde kullanılan pencere yapısına bir örnek Tablo 1'de verilmiştir.

Tablo 1: Word2Vec Modelinde Kullanılan Pencere Yapısı

	CÜMLE				GİRDİ	ÇIKTI
1	Yapay	zeka	yeni	elektriktir	(zeka)	(yapay)
2	Yapay	zeka	yeni	elektriktir	(yapay) (yeni)	(zeka) (zeka)
3	Yapay	zeka	yeni	elektriktir	(zeka) (elektriktir)	(yeni) (yeni)
4	Yapay	zeka	yeni	elektriktir	(yeni)	(elektriktir)

Kaynak: Aydoğan, Karcı, 2019: 184

Tablo 1’de görüldüğü gibi cümlelerin bölümleri bir pencere içerisinde ortadaki kelime hedef kelime olarak algılanır ve bulunduğu penceredeki diğer kelimeler ile karşılaştırarak bir vektörel matris oluşturur. Pencere yapısında her pencere içerisindeki bağlam kelimeleri ve hedef kelime matris içerisinde bulunur. Bu matrise karşılık bir de değerler matrisi tutulur. Değerler matrisinde hedef kelimenin içerisinde bulunduğu her matris elemanının değeri 1 olur iken, hedef kelimenin bulunmadığı elemanın değeri 0 olarak atanır.

Bu metodlar kelime gösterimleri konusunda başarı göstermişlerdir ancak kelimeleri bağlamlarından bağımsız bir şekilde algılayabilmekte, kelimelerin cümle içerisindeki anlamlarını göz ardı etmektedirler. Başka bir ön eğitilmiş dil modeli olan BERT bu soruna bir çözüm getirmektedir.

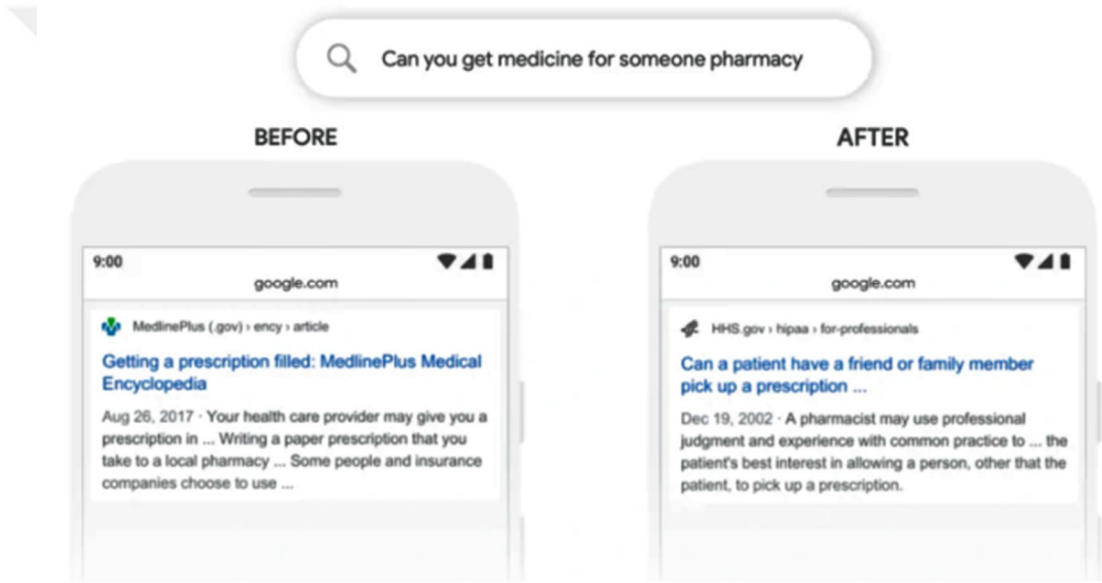
2.3.1 BERT

Google şirketinin yapay zeka bölümü çalışanları Devlin ve arkadaşları tarafından yazılan BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding adlı makale ile geliştirilen BERT (Dönüştürücülerden Çift Yönlü Kodlayıcı Temsilleri, Bidirectional Encoder Representation from Transformers) ön eğitilmiş bir dil algılama modeli duyurulmuştur (Devlin ve Chang, 2018). BERT modeli, soru cevaplama, diller arası çeviri yapma, sonraki cümleyi tahminleme, doküman özeti çıkarma gibi DDİ problemlerine çözümler geliştirmektedir (Devlin ve diğerleri, 2019: 2). Google arama motorunda da

kullanılmaya başlanan bu dil modeli, Google üzerinde yapılan aramaların daha sağlıklı bir şekilde sonuçlar verebilmesi için bir dil modeli ile çalışılmış ve çözüm sağlanmıştır.

BERT dil modeli günümüzde aktif olarak Google arama motorunda da aranılan kelime ile sonuçlar kümesi arasındaki en ilişkili sorgu sonuçlarının bulunması için de kullanılmaktadır. BERT dil modeli ile geliştirilen Google arama motoru üzerinde bir sorunun BERT'ten önceki ve sonraki sonuçları Şekil 2'de belirtilmektedir.

Şekil 2: Google Arama Motorunda BERT Modeli Kullanımı



Kaynak: Nayak, 2019

Şekil 2'de görüldüğü üzere “Başkası adına eczaneden ilaç alınabilir mi” aramasının yanıtı BERT modelinden önce, ilk çıkan arama sonucu reçete alımı hakkında olurken, BERT kullanımı ile birlikte aramaya dair en uygun sonuçlar çıkarılmaya başlanmıştır.

Açık kaynak kodlu olarak oluşturulan bu dil modelini kullanarak, BERT modelinin geliştirildiği dil olan İngilizce dışında diğer dillere özel modeller de geliştiriciler tarafından oluşturulabilmektedir. BERT modelinin ön eğitim safhasında İngilizce kitaplardaki metinlerin veri kütüphanesi olan BookCorpus (800 Milyon kelime) ve İngilizce Wikipedia veri kaynağı (2.500 Milyon kelime) kullanılmıştır.

Bu modelin önceki modellere göre en büyük özellikleri şunlardır:

- ### Şekil 3: BERT Bağlamsal Kelime Gösterimi Çözümü İçin Bir Örnek



16

gizlenilerek modelin tahmin ederek öğrenmesi sağlanır. Bu öğrenme esnasında hem soldaki hem de sağdaki bağlam edinilmiş olduğu için çift yönlü bağlamı göz önünde bulundurarak model eğitilmiş olur.

BERT dil modelinde ilk aşama girdilerin parçalara ayrılarak vektörlendirildiği alandır. Bu aşamadan sonra katmanlara ayrılmış bir fiziki yapı içerisinde dil işleme işlemleri gerçekleştirilir. Girdiler her cümle başına 768 adet vektör değeri içeren vektör matrisleri şeklinde modele iletildikten sonra, 64 adet vektör değeri içerecek şekilde 12 parçaya bölünerek katmanlara iletilmektedir. Dikkat mekanizması, ileri besleme ve çıktı serisi katmanlarında işlenmektedir (Devlin ve diğerleri, 2019: 3).

İlk aşamada girdilerin parçalarına ayrılması gerekmektedir. Bu aşamada BERT içerisinde tokenizasyon için WordPiece modeli mevcuttur. WordPiece modeli ile cümleler kelime parçacıklarına ayrıştırılmaktadır.

2.3.2 WordPiece Tokenizasyon Modeli

BERT dil modelinin ön eğitim esnasında eğitildiği kaynak metin veri kümesinden kullanım frekansına göre seçilen 30.000 kelime ve eklerin listelenmiş olduğu ve cümlelerin bu listeye göre parçalanılarak tokenlerine ayrıldığı modeldir (Muller ve diğerleri, 2019: 299). Ekler veya harf öbeklerinin tokenleştirilmiş gösterimlerinde “##” başlığı kullanılmaktadır ancak istisnai olarak sadece kelimenin başındaki parçacıkta bu gösterim yer almaz. Kendi başına anlamı olmayan harf bütünlükleri de bu 30.000 adet parçacığın içinde yer almaktadır ve bu sayede tokenleştirilemeyecek herhangi bir kelime parçacığı bulunmamaktadır. Şekil 4’te WordPiece ile kelime parçacıklarına ayrılmış kelimelerden bir örnek gösterilmiştir.

Şekil 4: WordPiece Tarafından “temsillemek” Kelimesinin ve Anlamı Olmayan Kelime Öbeğinin Sözcük Parçacığına Ayrıştırılmış Hali

```
"temsillemek"  
['temsil', '##lemek']  
"kroxldyphivc"  
['kro', '##xl', '##dy', '##phi', '##vc']
```

Kaynak: Yazar Tarafından Derlenmiştir.

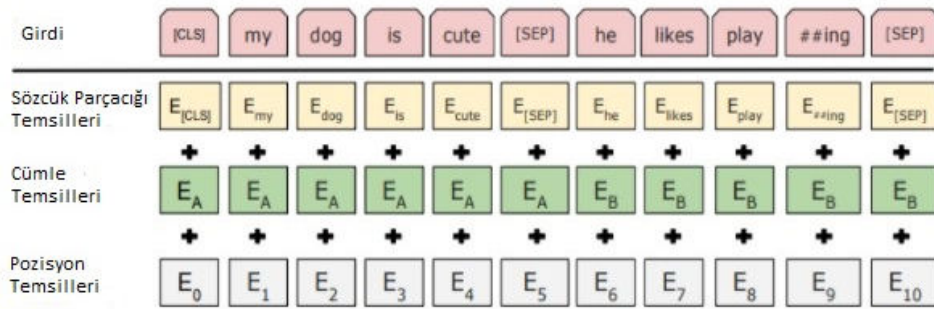
Şekil 4’teki gibi baştaki parçacık sabit kalmak üzere diğer kısımlar ##

işaretçisi ile beraber tokenleştirilirler. Anlamsız harf grupları ise WordPiece kelime listesindeki birleşik harf öbeklerine göre tokenleştirilmektedir.

BERT dil modeli, girdi cümleleri kelime parçacıklarına ayırmak için WordPiece adı verilen bir yapı kullanmaktadır. WordPiece yapısı, her bir giriş sözcüğünü sözcük parçaları (token) adı verilen alt sözcüklere böler. Kelimeler ile karakterler arasında orta nokta olarak kelime parçacıkları dilsel anlamlarını korurlar ve küçük boyutlu bir sözcük dağarcığı ile bile sözcük dağarcığının dışında kalan sözcük parçaları durumlarında da çalışabilmektedir (Song ve diğerleri, 2020: 1). WordPiece tarafından oluşturulan sözcük parçaları anlamlarını koruyacak şekilde oluşturulurken, anlamı bulunmayan kelime kalıplarının da WordPiece'nin kaynak olarak kullandığı 30.000 parçacıklık sözcük dağarcığı içerisinde anlamlı karakter parçalarının kullanılması ile dağarcık dışındaki kelime bütünleri de temsil edilebilmektedir. Bu sayede dağarcığın dışında kalan sözcük parçacıkları da temsil edilebildiği için anlamsal olarak BERT dil modeli içerisinde dil algılama aşamalarında işlenmektedir.

Girdi kelimelerinin tokenleştirilmesi sonrasında kodlanması adımıyla çalışan yapıdır. Bu yapıda girdi tokenlerinin bağlı olduğu cümle kodu, tokenin bulunduğu sıranın kodu ve tokenlerin WordPiece kodu oluşturulur. Şekil 5'te girdilerin vektör temsillerinin oluşturulmasındaki WordPiece yapısı gösterilmektedir.

Şekil 5: WordPiece Kelime Temsilleme Yapısı



Kaynak: Devlin ve diğerleri, 2019: 5

Cümlede sözcük parçalarının bulundukları cümle, segment temsili olarak ve bu sözcük parçalarının cümle içerisindeki sıra bilgisi de pozisyon temsili olarak

vektörleştirilmiş temsil içerisinde yer almaktadır. Tüm bu bilgilerin toplamı ile girdi temsili olarak vektör elde edilmektedir. Bu kelime parçacıklarının özelliklerini belirten [CLS] ve [SEP] özel tokenleri kullanılmaktadır.

[CLS] tokeni, her metin girdisinde BERT tarafından otomatik olarak oluşturulur ve tokenlerin sınıflandırılması için kullanılmaktadır. [SEP] tokeni ise, girdi metinlerinin cümle bitimlerini belirtir. Her cümle bittiğinde seperator (ayırıcı) anlamına gelen bir token oluşturulur (Devlin ve diğerleri, 2019: 4).

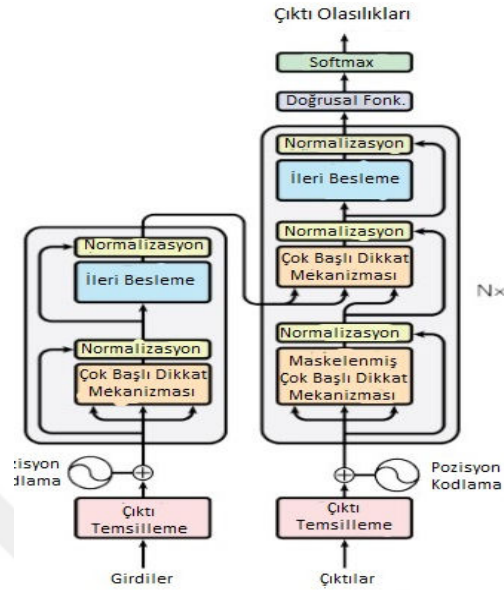
Özel tokenler, girdi cümlelerinin cümle içerisindeki parçacıkların pozisyonlarının ve bölümlere ayırma işlemi gerçekleştirileceğini modele anlatacak özellikleri taşımaktadır. Kelime parçacıklarının hangi cümlenin parçası olduğu ve cümlenin kaçınıcı parçası olduğunu yani pozisyonunu [SEP] özel tokeni ile model öğrenebilmektedir. Kelime gösterimleme ve WordPiece yapısı, bu verileri kodladıktan sonra vektör olarak dönüştürücü modeline aktarır ve anlamın çıkarılması işleminin ilk aşamasını gerçekleştirmiş olur.

2.3.3 Dönüştürücü (Transformer)

Dönüştürücü mimarileri, girdi cümlelerinin farklı parçalarının cümleye etkisini tartan dikkat (attention) mekanizması ve kodlayıcı-kod çözücü (encoder-decoder) yapısını barındıran bir DDİ yapısıdır. Kodlayıcı-kod çözücü yapısı içerlerindeki dikkat mekanizmalarının çalışma yapılarının açıklanması ile anlamlandırılabilir.

Dönüştürücü mimarisi iki ana yapıdan oluşmaktadır. Ortadan dikine bir çizgi çizdiğimizizi düşünerek çizginin solunda kalan yapıya kodlayıcı (encoder), sağında kalan yapıya ise kod çözücü (decoder) adı verilmektedir (Lopes ve diğerleri, 2019: 121). Şekil 6'da kodlayıcı-kod çözücü mimarisi örneği gösterilmektedir.

Şekil 6: Dönüştürücü (Transformer) Mimarisi



Kaynak: Vaswani ve diğerleri, 2017: 3

Şekil 6’da görüldüğü gibi dönüştürücü mimarisi girdi ve çıktı kodlayıcıları, dikkat (attention) mekanizmaları, ileri beslemeli ağlar ve optimum olasılık hesaplaması için Softmax fonksiyonunu barındırmaktadır.

Tokenizasyon ve kelime gösterimleme kısmının gerçekleştirildiği yer kodlayıcı’nın ilk işlem elemanı olan girdi kodlama (input embedding) kısmıdır. Burada oluşturulan sayısal kelime gösterimlerine pozisyonlarının kodlanmış hali eklendikten sonra 768 boyutunda bir vektör matris elde edilmektedir (Vaswani ve diğerleri, 2017: 3). Bu matris dizisi dikkat (attention) adı verilen mekanizmada işlenilerek cümle içerisindeki kelimelerin birbirleri arasındaki ilişkileri belirlenerek, ilişkili olma olasılıkları belirlenmektedir. Bu skorlar ile maskelemenin birlikte kullanılması ile kelimeye farklı açılardan bakılarak cümledeki her kelimenin cümle içerisinde kullanıldığı anlam hesaplanmaktadır. Bu hesaplamanın yapıldığı elemanın adı ise çok başlı dikkat (multi-head attention) tir. Tüm bu işlemler sonucunda cümle anlamı ile bağımlı olarak algılanabilmektedir.

Dikkat mekanizmasının işlevi, bir sorguyu ve bir dizi anahtar-değer çiftini bir çıktıya eşlemek olarak tanımlanabilir. Burada bahsedilen sorgu, anahtar ve değerler ile karşılık gelen çıktıların tümü vektörlerdir (Vaswani ve diğerleri, 2017: 4). Burada

sorgu değeri anlam barındıran her sözcük parçasıdır. Anahtar ve değerler ise tüm cümle hafızaları ile sözcük parçacıklarıdır. Verilerin birbirleri arasındaki olasılık hesapları ile birbirleri arasındaki ilişki düzeylerinin hesaplanması bu mekanizma içerisinde gerçekleştirilmektedir. Kelimelerin ve sözcük parçacıklarının birbirleri ile ilişkisinin öğrenilmesi ve bu ilişki olasılığı hesaplamasının tüm cümle içerisindeki parçacıkları ile cümlelerin temsil edildiği vektörler içerisinde gerçekleştirilmesi ise her sözcük parçacığının cümle anlamı içerisindeki ilişkisini hesaplamayı sağlamaktadır. Bu sayede BERT dil modeli bağlamsal olarak dil algılama görevlerini gerçekleştirmektedir. Dikkat mekanizmasının cümlelerdeki ilişki düzeyi çalışması ile ilgili genel bir örnek Şekil 7’de gösterilmiştir.

Şekil 7: “O bir yeşil elma yiyor” Cümlesinin Bazı Kelimeleri Arasında İlişki Düzeyi Görselleştirmesi



Kaynak: Weng, 2018

Şekil 7’de görüldüğü gibi kelimelerin arasındaki ilişki kuvveti ile ilgili “yiyor” kelimesinin “elma” kelimesi ile arasındaki ilişki oranının yüksek, “yeşil” kelimesi ile arasındaki ilişki oranının düşük olduğu belirtilmektedir. Dil modelinde dikkat mekanizması kullanılarak elde edilmek istenen kabaca bu ilişkilerin olasılıklarıdır.

Dikkat mekanizmaları, girdi tokenlerinin arasındaki ilişki düzeylerini belirlerken o anda işlediği tokeni sorgu (query), daha önce tüm sorgulamış olduğu tokenleri anahtar (key) ve cümlede yer alan tokenleri değer (value) adında vektör matrisleri olarak şu formülde işleme tabi tutar:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

d_k değeri, anahtar (key) matrisinin boyut sayısını temsil etmektedir. Q matrisi ile K matrisinin transpoz hali çarpılır. Bu matris çarpımının $\sqrt{d_k}$ değerine bölümü ile yeni bir vektör matris elde edilir. Bu vektör matrisin softmax fonksiyonu ile işlenmesi ile V matrisi ile çarpımı her Q değeri için tokenler grubundaki ilişki oranını

belirtmektedir. Softmax fonksiyonu ise, girdi değerlerin ilişki düzeyini olasılıksal olarak belirten bir fonksiyondur. En olası sonucun seçilimindeki hesaplamanın sağlanması için kullanılmaktadır (Vaswani ve diğerleri, 2017: 4). Softmax fonksiyonu girdi elemanlarının pozitif, negatif olması mühim olmadan 0 ila 1 arası bir değer üretecektir ki bu değerler de ilişkinin olasılık skoru olacaktır. Matematiksel gösterimi ise şu şekildedir:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

$\vec{z}$, Softmax fonksiyonunun girdi vektörünü ($z_0..z_K$), z_i , girdi vektörünün i indeksli elemanını, e^{z_i} , her girdi vektörü elemanına üstel fonksiyon uygulanmış halini (euler sabiti üzeri elemanın değeri) yani negatif değerlerin pozitifleştirilmesini, $\sum_{j=1}^K e^{z_j}$, tüm girdi değerlerin 0-1 arasındaki skalaya indirgenmesini ve çıktı değerlerinin toplamının 1 olmasını sağlayan formülü temsil etmektedir (Wood, 2019). [6 3 0] değerleri ile bir vektör matrisimiz olduğunu farzedelim. Bölme işleminin üst kısmındaki e^{z_i} hesaplaması için:

$$e^{z_1} = e^6 = 403.42$$

$$e^{z_2} = e^3 = 20.08$$

$$e^{z_3} = e^0 = 1.00$$

Sonuçları hesaplanacaktır. Bölüm işleminin altındaki kısmın hesaplaması için:

$$\sum_{j=1}^K e^{z_j} = e^{z_1} + e^{z_2} + e^{z_3} = 403.42 + 20.08 + 1.00 = 424.50$$

Sonucu elde edilir. z_1 değerinin diğer değerlere göre daha kuvvetli bir değerinin olduğu görülmektedir. z değerlerinin olasılık sonuçlarının hesaplanması için şu işlem adımları uygulanmaktadır.

$$\sigma(\vec{z})_1 = \frac{403.42}{424.50} = 0.9503$$

$$\sigma(\vec{z})_2 = \frac{20.08}{424.50} = 0.0473$$

$$\sigma(\vec{z})_3 = \frac{1.00}{424.50} = 0.0023$$

Bu sonuçlar ile z_1 elemanının en olası değer olduğu hesaplanmış olmaktadır (Wood, 2019).

2.3.4 BERTurk Dil Modeli

BERT modelinin açık kaynaklı olması üzerine, ana dili olan İngilizce dışında diğer diller üzerine de BERT modeli kurgulanabilmektedir. Türkçe diline özel olarak kurgulanan BERTurk modeli Stefan Schweter tarafından Türkçe OSCAR ve OPUS korpuslarını, Türkçe Wikipedia metinlerini ve Kemal Oflazer'in paylaştığı Türkçe haber metinlerini kaynak olarak kullanarak oluşturulmuştur. Bu modelin ön eğitiminde kullanılan kaynak verilerin boyutu 35 Gigabayttır ve toplamda 299.245.100 adet cümleden oluşmaktadır (Schweter, 2020).



ÜÇÜNCÜ BÖLÜM

YAPAY SİNİR AĞLARI

3.1 TANIMLAR

Yapay sinir ağları (YSA), genel olarak insan beyninin ya da merkezi sinir sisteminin çalışma prensiplerini taklit eden bilgi işleme sistemidir (Freeman ve Skapura, 1991: 3). İnsan beynindeki sinir hücreleri ve algılayıcıların fiziki yapısının istatistiksel olarak bilgisayar teknolojileri ile taklit edilmesi sonucunda bir sinir hücresi gibi verileri kullanarak bu verilerden anlamlı sonuçlar elde edebilen, matematik temelli modellerdir.

Efe ve Kaynak'a göre, genel olarak YSA insan beyninin biyolojik sinir yapısını taklit ederek sinirsel algılayıcılar yardımı ile önceden öğrenilmiş ya da sınıflandırılmış bilgileri kullanarak yeni bilgiler türetebilen ve oluşturabilen, karar verebilen bilgisayar programlarıdır (Keskenler M.F. ve Keskenler E.F., 2017: 10). Var olan verilerin istatistik tabanlı modeller kurgulanarak bu bilgilerden anlamlı ve kaynak veriler ile alakalı somut bilgiler çıkarımına süreçlerinin tamamı YSA modellerinin kurgulandığı temel süreçlerdir.

YSA'nın en temel görevi, dış dünyadan aktarılan girdilerin, ağ içerisindeki elemanlar aracılığı ile matematiksel işlemlere tabi tutularak ağırlık değerlerinin düzeltilmesi aracılığıyla girdinin karşılığındaki çıktıyı bulmak ve bu çıktıları, eğitim sürecinde öğrendiği bilgiler ile yorumlayarak istenen sonucu ortaya koymaktır (Ramalingam ve diğerleri, 2006: 160). YSA, biyolojik sinir ağlarını taklit ederek, daha önceki tecrübelerden ve örneklerden yeni örnekleri yorumlayabilen ve öğrenebilen, istatistiksel ve yazılımsal temellere dayanan bilişim sistemleridir.

Biyolojik sinir ağlarını taklit ederek öğrenen, öğrendiklerini yeni verilerde kullanarak en ideal sonucu bulmaya çalışan YSA, veri sınıflandırma, öngörü ve tahminleme, ses tanıma, görüntü tanıma, kümeleme görevlerinde kullanılmaktadır (Güngör, 2007: 6). YSA'nın uygulama alanlarına ise, üretim planlama, hastalık tanıma ve tedavisi, kalite kontrolü, parmak izi tanıma, otonom araç denetimi, zeki sistemler için optimum rota belirleme örnekleri verilebilir (Keskenler M.F. ve Keskenler E.F., 2017: 11).

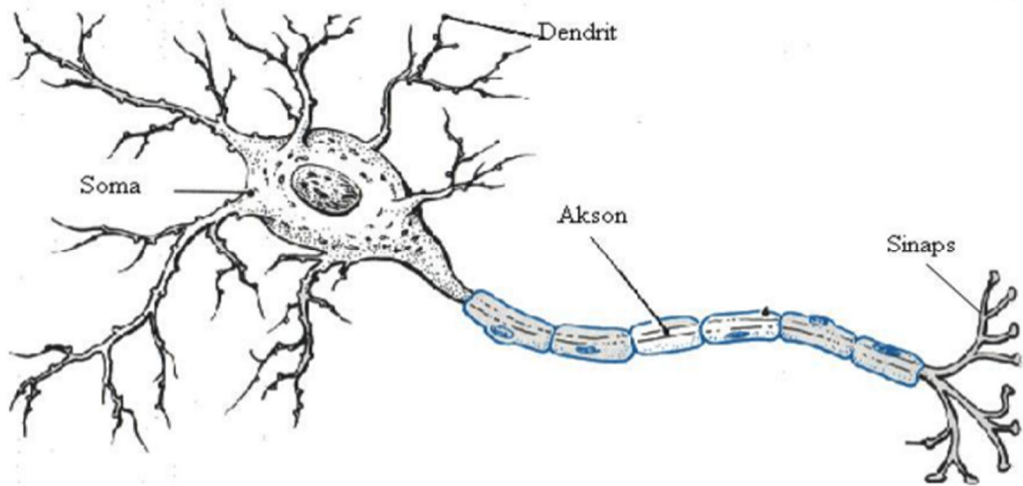
Özetlemek gerekirse; YSA insan beyninin sinir ağı mimarisinin istatistik temellerine dayalı, bilgisayar ortamında oluşturulmuş yapay bir modellemesidir. Tıpkı insanlar gibi örneklerden öğrenir, öğrendiklerini kullanarak farklı sorunlarda çözüm üretmek amaçlı tahminler yürütür. Girdilerin, tasarlayıcı tarafından tanımlanan matematiksel fonksiyonlar ile en olası çıktılar ile eşleştirilerek nitelikli analiz bilgilerini sağlayan modellerdir.

3.2 BİYOLOJİK SİNİR AĞI

İnsan beyni, derin ve karmaşık bir biyolojik sinir ağı ile donatılmış olmasının ve biyolojik olarak bellek ile desteklenmesinin sayesinde, olağanüstü karmaşık problemleri, aynı anda birden fazla işlevi ve doğrusal olmayan işlemleri gerçekleştirebilen bir biyolojik sinir sistemidir (Öztemel, 2020: 45).

Bu biyolojik sinir sistemleri, deneyimler ve örneklerden tecrübe edinerek gelecekte karşılaşılabilecek problemlere çözüm arayışındadır. Bu süreç öğrenme süreci olarak adlandırılır. Şekil 8’de biyolojik sinir hücresinin elemanları ile beraber, yapısı gösterilmektedir.

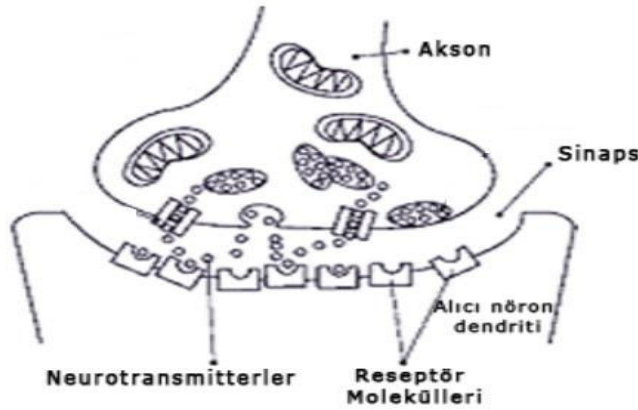
Şekil 8: Bir Biyolojik Sinir Hücresinin Yapısı



Kaynak: Uysal, 2019: 5

Şekil 8’de gösterildiği gibi, biyolojik sinir hücreleri sinapslar, soma, akson ve dentritlerden oluşur. Sinapslar, sinir hücreleri arasındaki iletişim görevini üstlenir. Sinapslardan iletilen elektrik sinyalleri soma hücrelerine gönderilir. Soma hücreleri bu sinyalleri işleme tabi tutar ve bu sinyale göre kendi elektrik sinyalini oluşturur. Oluşturulan bu yeni sinyal, akson aracılığı ile dentritlere gönderilerek sinapslara iletilmesini sağlar (Öztemel, 2020: 47). Sinapslara iletilmiş olan sinyaller ise diğer sinir hücrelerine iletilmektedir. Şekil 9’da gösterildiği gibi sinapslar, hücreler arasındaki iletişimi sağlayabilmek adına fiziksel bir bağlantı kurmayıp, diğer hücrelerin sinapsları ile aralarında boşluklar oluşturacak şekilde konumlanmıştır. Bu boşluklardaki sinyal aktarımlarını nörotransmitter adı verilen yapılar gerçekleştirir. Karşı sinapstaki reseptörler bu nörotransmitterleri algılar. Şekil 9’da biyolojik sinir ağlarında elektrik sinyallerinin algılanma işleminin gerçekleştirildiği sinaptik boşluk alanında algılayıcıların bu sinyalleri algıladığı model görünmektedir (Öztemel, 2020: 47).

Şekil 9: Biyolojik Sinir Hücresi Bağlantısı ve Bilgi Aktarımı



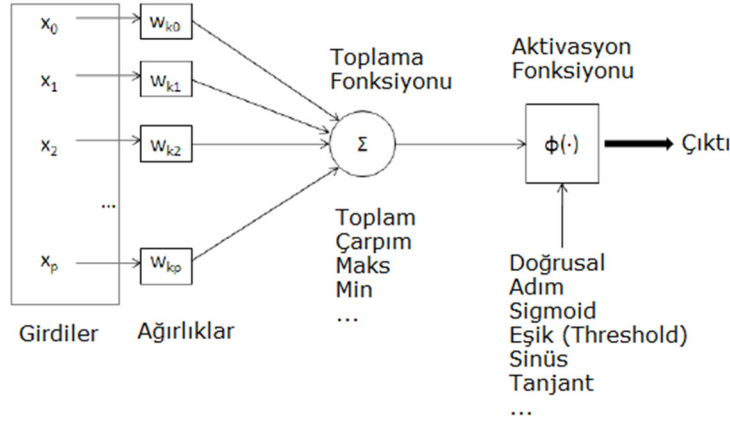
Kaynak: Bayrak, 2008: 105

3.3 YAPAY SİNİR HÜCRESİ

Bir yapay sinir hücresi temel olarak girdiler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktı elemanlarından oluşmaktadır (Öztemel, 2020: 48). Bu elemanlar dış dünyadan gelen verilerin işleme sürecinde aktif olarak bilgi çıkarımı

için gereken işlemlerin yapıldığı elemanlardır. Girdi ve çıktı elemanları üzerinde herhangi bir işlem gerçekleştirilmezken, toplama fonksiyonu, aktivasyon fonksiyonu ve ağırlık elemanları üzerlerinde işlemler gerçekleştirilmektedir. Yapay sinir hücresi elemanları Şekil 10’da gösterilmektedir.

Şekil 10: Tek Katmanlı Yapay Sinir Hücresi



Kaynak: Pişkin, 2017

Şekil 10’da gösterildiği gibi yapay sinir hücrelerinin de biyolojik sinir hücreleri gibi elemanları ve görevleri bulunur. Temel olarak tek katmanlı bir yapay sinir hücresi için 5 adet ana eleman mevcuttur.

- **Girdiler:** Sinir hücresinin bilgi üretmesinin ilk aşaması olarak kaynak bilgileri hücreye getiren kısımdır. Bu bilgiler, ağırlık öğrenmesi istenilen bilgilerdir (Kaynar ve Taştan, 2009: 163).
- **Ağırlıklar:** Girdi elemanına gelen bilginin, diğer bilgiler ile arasındaki farkı ve etkisini göstermek üzerine kurgulanmış değerlerdir. Ağırlık değerinin büyük ya da küçük oluşu, önemli veya önemsiz olduğu şeklinde yorumlanmamalıdır. Aynı zamanda ağırlığın sıfır ya da negatif değerde oluşu da etkisinin yönünü belirlemektedir (Öztemel, 2020: 49).
- **Toplama Fonksiyonu:** Bir hücreye iletilen net bilgi girişini hesaplama görevini gerçekleştiren bir fonksiyondur. Ağırlıklı toplamı bulmak genellikle kullanılan toplama fonksiyonu yöntemidir. Her gelen girdi burada kendi ağırlığı ile çarpılarak toplanır. Ağırlık iletilmiş net girdi aşağıdaki bağıntı ile hesaplanır,

$$NET = \sum_{i=1}^n G_i A_i$$

G elemanları girdileri, A elemanları ise ağırlıkları, n ise toplam girdi sayısını temsil etmektedir. i sayısı, 1 sayısından n sayısına kadar olan değerleri alarak her girdinin sırayla bu fonksiyonda işlenmesini sağlayan sayı değerlerini alan değişkendir.

Bu fonksiyon elemanı için toplama işlemi dışında da farklı fonksiyon hesaplamaları da tercih edilebilmektedir (Öztemel, 2020: 49). En çok tercih edilen fonksiyon toplam işlemi olmakla birlikte, diğer tercih edilen fonksiyonlar da Tablo 2’de gösterilmiştir.

Tablo 2: Yapay Sinir Ağları Toplama Fonksiyonları

Kullanılan Fonksiyon	Açıklama
<i>Çarpım</i> $NET = \prod_i^n G_i A_i$	Girdiler, kendilerine atanan ağırlık değerleri ile çarpılarak elde edilen değerlerin tümü birbiri ile çarpılarak net girdi hesaplaması yapılır.
<i>Maksimum</i> $NET = \text{Max}(G_i A_i), i = 1 \dots n$	n sayıda sağlanmış olan girdilerden en büyüğü sinir hücresinin net girdisi olarak kabul edilir.
<i>Minimum</i> $NET = \text{Min}(G_i A_i), i = 1 \dots n$	n sayıda sağlanmış olan girdilerden en küçüğü sinir hücresinin net girdisi olarak kabul edilir.
<i>Çoğunluk</i> $NET = \sum_i^n \text{sgn}(G_i A_i)$	n sayıda sağlanmış olan girdiler içerisinde ağırlıklar ile çarpıldıktan sonra pozitif ve negatif olanların sayısı bulunur. Büyük olan sayı hücrenin net girdisi olarak kabul edilir.
<i>Kümülatif Toplam</i> $NET = \text{NET}(\text{eski}) + \sum_i^n G_i A_i$	Hücreye gelen bilgiler ağırlıklı olarak toplanır ve önceki net girdiye eklenerek hücrenin net girdisi hesaplanır.

Kaynak: Şanlı, 2018: 12

- **Aktivasyon Fonksiyonu:** Hücreye gelen net girdi değerini, uygun çıktı değerini hesaplamak için işleyen fonksiyondur. Bu fonksiyon, doğrusal olmayan problemlerde eldeki girdi ve ağırlıklardan çıkarılan net girdi değerinden, probleme en uygun çıktı değerine ulaşmak için gereken matematiksel işlemleri barındırır. Aktivasyon fonksiyonu kullanılarak sinir ağının daha dinamik olması, karmaşık problemlere karşı doğrusal olmayan işlevsel eşleşmeleri gerçekleştirmesi sağlanır (Sharma ve diğerleri, 2020: 311).

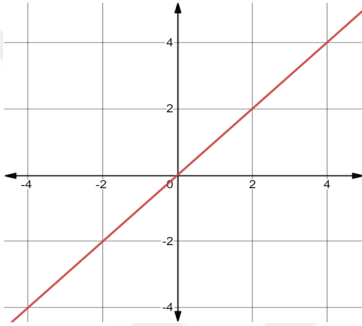
YSA’nda birçok aktivasyon fonksiyonları kullanılmaktadır. Bu fonksiyonlardan bazıları aşağıda açıklanmaktadır.

Doğrusal Fonksiyon: Bu fonksiyon, girdi ile doğrudan orantılıdır. Net girdinin hesaplanmasından sonra Doğrusal (Lineer) Aktivasyon Fonksiyonu içerisinde ağ tasarımcısının belirleyeceği bir sabit değer ile çarpıldıktan sonra çıktı olarak kabul edilir.

$$f(x) = ax$$

a değeri, denklemde bir sabit sayıyı ifade eder ve kullanıcı tarafından değer belirlenebilir. Doğrusal fonksiyonun grafik eğrisi Şekil 11’de belirtilmiştir.

Şekil 11: Doğrusal Aktivasyon Fonksiyon Grafiği



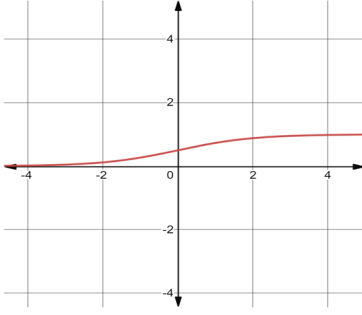
Kaynak: Yazar Tarafından Derlenmiştir.

Sigmoid Aktivasyon Fonksiyonu: Sonraki kısımlarda değinilecek olan çok katmanlı algılayıcı sinir ağlarının en yaygın kullanılan yapay sinir ağı olması sebebi ile bu aktivasyon fonksiyonu en sık kullanılan aktivasyon fonksiyonlarından biridir. Sürekli türevi alınabilen bir fonksiyon olması ve çok katmanlı algılayıcıların (ÇKA) da aktivasyon fonksiyonu için türevlenebilir bir fonksiyonun kullanılması şartı olması sebebi ile sıkça kullanılmaktadır. Sigmoid aktivasyon fonksiyonunun matematiksel gösterimi şu şekildedir:

$$f(NE T) = \frac{1}{1 + e^{-NE T}}$$

Burada NET, toplama fonksiyonundan gelen net girdi elemanını göstermektedir. Euler sabiti ile üstel fonksiyon üzerinde net girdi eleman değerinin hesaplaması yapılır. Doğrusal olmayan bu aktivasyon fonksiyonu, girdileri 0 ile 1 arasında bir skalada ölçeklendirir (Sharma ve diğerleri, 2020: 312). Şekil 12’de sigmoid aktivasyon fonksiyonunun grafik eğrisi belirtilmektedir.

Şekil 12: Sigmoid Aktivasyon Fonksiyon Grafiği



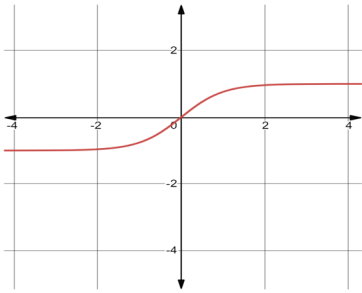
Kaynak: Yazar Tarafından Derlenmiştir.

Hiperbolik Tanjant Aktivasyon Fonksiyonu: Hiperbolik tanjant aktivasyon fonksiyonu (Tanh), Sigmoid fonksiyonuna çok benzemekte ancak farkı, orjin noktasına göre simetrik olmasıdır. Yani çıktıların ölçekleme skalası Sigmoid'e göre 0-1 arası değil, -1 ve 1 arasında olmaktadır. Bu fonksiyonun matematiksel gösterimi ise şu şekildedir:

$$\frac{e^{net} - e^{-net}}{e^{net} + e^{-net}}$$

Toplama fonksiyonundan gelen net girdi elemanı üstel fonksiyon içerisinde, euler sabiti üzerinde hesaplanır. Bu işlem sonucunda çıktı, o elemanın çıktısı olarak kabul edilir. Tanh aktivasyon fonksiyonunun grafik eğrisi Şekil 13'te belirtilmektedir.

Şekil 13: Hiperbolik Tanjant Aktivasyon Fonksiyon Grafiği



Kaynak: Yazar Tarafından Derlenmiştir.

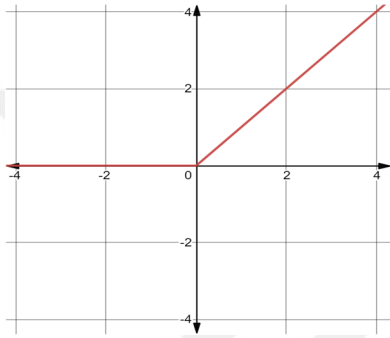
ReLU (Doğrultulmuş Doğrusal Birim): ReLU (Doğrultulmuş Doğrusal Birim) fonksiyonu, negatif çıktıların 0 değerine indirildiği ve değer skalası olarak 0 ile sonsuz değerleri arasında çıktı üreten bir aktivasyon fonksiyonudur. Bu durum işlem yükünü hafifletirken, negatif girdiler ile modelin eğitilemediği anlamına gelmektedir. ReLU fonksiyonunun negatif girdiler ile çalıştırılmaması gerekmektedir (Sharma ve

diğerleri, 2020: 313). Bu aktivasyon fonksiyonunun matematiksel gösterimi řu řekildedir:

$$\max (0, NET)$$

Gelen net girdi deęerleri ile 0 arasındaki deęerlerden en büyüęü, girdiye karřılık gelen çıktı olarak seçilir. ReLU aktivasyon fonksiyonunun grafik eęrisi řekil 14'te belirtilmektedir.

řekil 14: ReLU Aktivasyon Fonksiyon Grafięi



Kaynak: Yazar Tarafından Derlenmiřtir.

- **Çıktılar:** Girdilerin dięer elemanlar ierisinde iřlenmesi ile ulařılan sonucun bulunduęu alandır. Bu alandaki sonuçlar, dięer hücrelerin ya da kendisinden önceki iřlem elemanlarının girdisi olabilir.

3.4 YAPAY SİNİR AĞLARINDA ÖĞRENME YAPILARI

Yapay sinir aęlarının görevlerini gerçekleřtirebilmeleri iin bir öęrenme safhasından geçmeleri gerekmektedir. Bu aęlarda öęrenmeyi saęlamak iin 3 temel öęrenme yapısı bulunmaktadır;

- **Denetimli Öęrenme:** Aę modeline öęrenmesi iin örnek girdiler ve bu girdilerin alması gerektięi çıktılar verilerek, karřılařacaęı girdilerin çıktılarını hesaplarken bu örnekleri hatırlayarak doęru çıktıları saęlamasının beklendięi modeldir. Örneklerden hareketle çıktı deęerlerinin öęrenilmiř çıktılara yakın deęerler olabilmesi iin aęrılık deęerleri güncellenerek hesaplamaya devam edilir (Doęan ve Türkoęlu, 2018: 413).

- **Denetimsiz Öğrenme:** Bu öğrenme yönteminde denetimli öğrenme yapısındaki gibi modele örnek girdiler iletilir ancak bu örnek girdilerin çıktı sonuçları iletilmez. Bu şekilde herhangi bir denetleyici unsur olmadan model girdilere karşılık modelin çıktıları üretmesi ve bu çıktıları sınıflandırarak bir sonuca varması beklenir (Doğan ve Türkoğlu, 2018: 413).
- **Destekleyici Öğrenme:** Bu öğrenme yönteminde denetimsiz öğrenme yöntemi gibi örnek girdiler sisteme iletilir, ancak bu örnek girdilerin çıktıları modele gösterilmez, ancak bu örnek girdilerin doğru çıktıları da modelin içerisindedir. Modelin girdilere karşılık bir çıktı üretmesi beklenir, üretilen çıktılar, modele gösterilmeyen doğru çıktılar ile karşılaştırılarak modelin yanlış ya da doğru çıkarım yaptığını modele doğru ya da yanlış anlamına gelen sinyaller vererek öğrenmesi sağlanır (Doğan ve Türkoğlu, 2018: 413).

3.5 ÇOK KATMANLI ALGILAYICI

1960'lı yıllarda yapay sinir ağları çalışmaları, mevcut modellerin doğrusal olmayan problemlere çözüm bulamaması nedeni ile durmuştur. 1972 senesinden sonra doğrusal olmayan problemlerin çözümü için nitelikli sonuçlar veren çalışmalar yayınlanmıştır. 1982 senesinde Rumelhart ve arkadaşları tarafından geliştirilen ÇKA'lar, elektronik alanında kullanılan mantık kapılarından doğrusal olmayan “Özel Veya” (XOR) mantık kapısı problemini çözerek doğrusal olmayan problemlerde çözümü sağladığı kanıtlanan bir yapay sinir ağıdır (Keskenler M.F. ve Keskenler E.F., 2017: 13).

Modelin eğitimi için sağlanan girdiler ile öğrenme sonrası sağlaması istenen çıktıların doğrusal bir ilişkiye sahip olmadığı problemler doğrusal olmayan problemlerdir. Doğrusal olmayan problemlerden sıkça örnek olarak gösterilen XOR mantık kapısı probleminin girdileri ve istenen çıktıları Tablo 3'te gösterilmiştir.

Tablo 3: XOR Problemi Girdi ve Çıktı Tablosu

Girdi1	Girdi2	Çıktı
0	0	0
0	1	1
1	0	1
1	1	0

Kaynak: Yazar Tarafından Derlenmiştir.

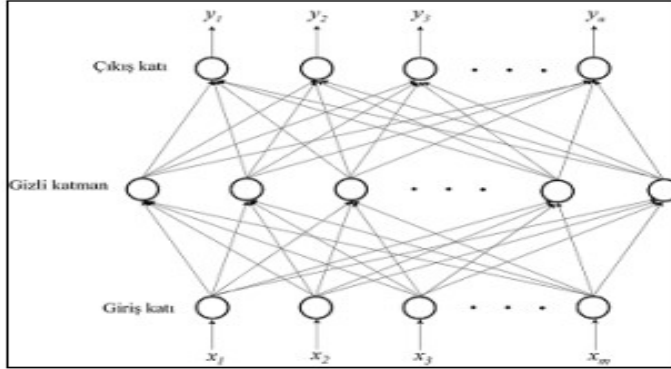
Tablo 3’te görüldüğü gibi XOR mantık kapısında girdiler birbirlerinden farklı olduğunda 1, girdiler birbirleri ile aynı olduğunda ise 0 çıktısı üretilmektedir ve bu problem çıktıların arasında doğrusal bir ilişki olmaması sebebi ile doğrusal olmayan bir problemidir. Doğrusal olmayan problemlere ÇKA modelleri kullanılarak çözümlenmektedir. Çözümü sağlama yöntemini anlayabilmemiz için ağıın yapısını incelememiz gerekmektedir.

3.5.1 Çok Katmanlı Algılayıcıların Yapısı

ÇKA’lar giriş, çıkış ve ara tabaka (gizli katman) olmak üzere 3 adet katman mevcuttur (Eğrioğlu ve diğerleri, 2019: 77). Bu katmanlar, verilerin tüm işlemleri boyunca dolaştığı YSA yapılarıdır. Giriş katmanı, dış dünyadan gelen verilerin modele ilk erişim sağladığı hücrelerin bulunduğu katmandır. Gizli katman (ara katman) üzerinde veriler üzerine fonksiyonlar ile hesaplama yapılır ve yapılan hesaplar ile verinin hesaplanmış çıktı değeri çıktı katmanında eşleştirilmektedir.

Her katman için, katman içerisindeki hücrelerin birbiri ile bir bağlantısı bulunmazken, tüm bağlantılar ileri yönlendirmeli bir bağlantı türüdür. ÇKA modellerinde kullanılan katman yapısının grafiği Şekil 15’te gösterilmektedir.

Şekil 15: Bir ÇKA Modelinin Katman Yapısı ve Nöronların Bağlantı Yapısı



Kaynak: Bayındır ve Sesveren, 2008: 103

Şekil 15'te görüldüğü üzere, giriş ve çıkış katmanlarındaki hücre sayısı aynı iken, gizli katmandaki hücre sayısı diğer katmanlardaki hücre sayısından bağımsızdır. Bu katmanların çalışma yapıları şu şekildedir;

- **Giriş Katmanı:** Bu katman, dış dünyadaki bilgilerin ağıla ilk girişi yaptığı katmandır. Girdi hücreleri girdileri alıp ara katmana herhangi bir işlem yapılmadan aktarır. Her girdi hücresinden ara katmandaki proses hücrelerinin tamamına bir bağlantı vardır ve her girdi sonraki katmanın tüm hücrelerine iletilerek işleme tabi olur.
- **Gizli Katman:** Bu katman, girdi hücrelerinden iletilen girdilerin tüm ara katman hücrelerinde işlenilerek diğer katman hücrelerinin hepsine aktarılır. Her tekil işlenmiş girdi değeri, sonraki katmanın tüm proses hücrelerinin tamamına iletilir. ÇKA'nın her gizli katman hücreleri şu iki hesaplamayı yapmak için tasarlanmıştır: Girdilerin aktivasyon fonksiyonları ile her girdiye denk gelecek olan çıktı değerinin hesaplanması ve her girdi verisine göre hesaplanmış olan çıktılardaki, geriye doğru hesaplama için gerekli olan gradyan vektörü (girdilerin ağırlıklarına bağlı olarak hata düzeyi) değerlerinin hesaplanması (Haykin, 2010: 125).
- **Çıktı Katmanı:** Ara katmandan iletilen işlenmiş girdi değerlerinin çıktı değerlerini belirleyecek işlemlerinin gerçekleştirildiği katmandır. Belirlenen çıktı değerleri bu katmandan dış dünyaya iletilmektedir. Her proses elemanına denk gelen sadece bir adet çıktı değeri üretilmektedir (Öztemel, 2020: 77). Her katman için, katman içerisindeki nöronların birbiri ile bir bağlantısı bulunmazken, tüm bağlantılar ileri yönlendirmeli bir bağlantı türüdür. Giriş ve çıkış katmanları dışında gizli

katmanların sayıları arttırılabilir ancak literatürde genellikle tek gizli katman ile modeller geliştirilmiştir.

3.5.2 Çok Katmanlı Algılayıcıların Öğrenme Kuralı

ÇKA ağları, denetimli öğrenme stratejisine göre çalışmaktadır (Öztemel 2020: 77). Bu ağlar, önceden hazırlanmış veriler ve bu verilere karşılık beklenen çıktı değerleri model tasarımcısı tarafından ağa iletilerek, eğitim safhası sonrası ÇKA ağları öğrenilenleri, çıktısı önceden belirli olmayan verilerin muhtemel çıktılarını tahmin edebilmek için kullanırlar. Ancak öğrenme aşamasında bu strateji tek başına yeterli değildir.

Öğrenme stratejilerinin yanı sıra girdi ve çıktıların arasındaki ilişkiyi doğru bir şekilde elde etmek için öğrenme kuralları türetilmiştir. ÇKA’larda kullanılan öğrenme kuralı literatürde Genelleştirilmiş Delta Kuralı olarak belirtilmektedir (Güllü ve diğerleri, 2007: 188). Bu kural eğitim veri setinin ağa gösterilmesi ve bu verilerden elde edilmesi gereken çıktıları göstererek, eğitim veri seti dışındaki verilerin ağa iletilmesi durumunda, eğitim verisinden yaptığı çıkarımlar ile benzer verilere benzer çıktıların oluşturulması üzerine kurgulanmış bir kuraldır.

Genelleştirilmiş delta kuralının ileriye doğru ve geriye doğru olmak üzere iki adet aşaması mevcuttur (Öztemel, 2020: 77).

3.5.2.1 İleriye Doğru Hesaplama Aşaması

Bu aşama girdi katmanından eğitim verilerinin ağa girişi ile başlar, ara katman geçişi öncesinde girdiler ağırlık değerleri ile eşleşirler. Toplama fonksiyonu hesaplaması ile ağa giren net girdi hesaplandıktan sonra aktivasyon fonksiyonu ile girdi başına çıktı hesaplaması sağlanır (Eğrioğlu ve diğerleri, 2019: 78). Bu aşamadaki temel hedef, modelin girdiler üzerinde işlemler yaparak girdilere denk gelen çıktılar üretmesidir. Modelin ürettiği çıktılar ile beklenen çıktılar, bir sonraki aşama olan geriye doğru hesaplama aşamasının girdileri olacaktır.

3.5.2.2 Geriye Doğru Hesaplama Aşaması

Ağın örnek girdiler için ürettiği çıktılar ile örnek girdilerden üretilmesi beklenen çıktılar karşılaştırılır ve beklenen değerler ile ağın ürettiği çıktılar arasındaki fark hata değeri olarak kabul edilir. Bu hata değerinin minimize edilmesi, ağın öğrenmesi anlamına gelmektedir.

Hata değerinin ağıdaki ağırlık değerlerine dağıtılması aşamasında bazı parametreler kullanılarak güncellenecek ağırlık değerleri ile modelin yerel değerler arasında salınmasını ve kısır bir değerler kümesinde takılı kalmaması hedeflenir. Bu parametreler ile modelin genel çerçevede bir optimum noktanın daha başarılı bir şekilde belirlenmesi amaçlanmaktadır.

3.5.2.3 Öğrenme Hızı ve Momentum Katsayısı

Geri yayılım algoritmasında hatanın en aza indirgenmesinde ağırlık değerlerinin aynı değerler arasında salınmasını ve yinelemeleri azaltmak amaçlı hata inişini kontrol etmek için literatürde öğrenme hızı ve momentum katsayısı parametreleri tanıtılmıştır (Hu Yu ve An Chen, 1997: 517). Bu parametrelerin temel amacı modelin hata üzerinden geri yayılım ile ağırlık değerlerini değiştirirken gradyan değerlerini baz alarak bu gradyan değerlerinden belirli bir oranda inişler sağlayarak ağırlıkların değişikliklerini kontrol eder. Gradyan basit bir ifade ile kayıp fonksiyonunun oluşturduğu en dik yükselişe sahip olduğu yönü veren bir vektördür. Değişkenlerin birbirlerine bağlı olarak kısmi türevleri alınarak gradyan vektörü elde edilir ve gradyan yönünü minimize etmeye yönelik fonksiyondaki değişkenleri değiştirerek hata oranı düşürülmektedir. Öğrenme hızı ve momentum katsayısı, gradyanı minimize etme yönündeki ağırlık değişimindeki adımlara etkide bulunur. Bu yöntem ile geri yayılım işlemi gerçekleştiren metodlara genel olarak gradyan inişi adı verilir.

Gradyanda inişin adım büyüklüğü, öğrenme katsayısı ile gradyan değerinin çarpımı ile ölçülmektedir (Goodfellow ve diğerleri, 2016: 289). Öğrenme hızı, gradyanın hangi büyüklükler ile değişeceğini belirtir. Bu parametre ne kadar büyük olur ise iniş adımı o kadar büyük olacağından eğitim hızı artar ancak adım aralarında

kalan deęiřken deęerleri g z ardı edilmiř olacaktır. Bu parametrenin k   k olması durumunda ise eęitim hızı yavaşlarken daha geniř bir deęerler k mesi ile  alıřılabilmektedir.

Momentum katsayısı bir  nceki iterasyondaki deęiřimin belirli bir oranının yeni deęiřim miktarına eklenmesi olarak g r lmektedir ( ztemel 2020: 99). Momentum terimi fizikte par acıęın hızı olarak kabul edilmektedir, sinir aęlarına uyarlanmıř hali ise dar bir   z m uzayına takılan sinir modellerinin takıldıęı yerel   z mlerden kurtulması i in bir ivme kazandırılarak daha iyi sonu lar bulabilmesini saęlamak  zere kullanılmaktadır. Yerel   z mler arasında salınımı engellemek i in bir sı rama saęlanmaktadır. K   k momentum deęerleri yerel   z mlerden kurtulmayı zorlařtırabilirken, b   k momentum deęerleri ise   z me ulařmayı zorlařtırabilmektedir ( ztemel 2020: 99).

Momentum ve  ęrenme katsayısı parametreleri geri yayılım s recine doęrudan etkide bulunduęu i in bu parametrelerin hatayı minimize etmekte en elveriřli řekilde kullanılabilmesi adına adaptif metodlar mevcuttur. Adaptif metodlar ile momentum ve  ęrenme katsayısı parametreleri gradyan deęerlerine g re uyarlanabilir řekilde deęiřtirilmektedir. Her gradyan deęerinde  ęrenme katsayısı ve momentum deęeri bellekte tutularak yeni gradyan deęerine uyarlanan yeni  ęrenme katsayısı ve momentum deęeri ile  alıřılabilmektedir.

Adaptif metodlardan en sık kullanılanı Adam'dır. Adam metodunda ilk olarak momentum, gradyanın ilk bulunduęu anda, tahmini olarak, doęrudan dahil edilmektedir (Goodfellow ve dięerleri, 2016: 308). Gradyan deęerinin  stel aęırlıęı ile tahmini bir momentum deęeri hesaplaması yapılır ve bu optimizasyon deęerleri her defasında bellekte tutularak sonraki ařamalarda daha hızlı bir řekilde m mk n olan optimal noktaya ilerlenmesini saęlar. Dięer  rneklerinden daha hızlı  alıřan bu optimizasyonun dięer  zellięi ise momentum tahminlerinde sapma d zeltmeleri yapabilmektedir. Bu nedenle adaptif momentum optimizasyonu saęlayabilmektedir (Goodfellow ve dięerleri, 2016: 308-309).

3.5.3 Geriye Doğru Ağırlık Hesaplaması

Bu aşamada ağırlıkların yapılan hatayı azaltacak şekilde değiştirmesi sağlanır (Öğücü, 2006: 46). Ağırlıkların değiştirilmesi ile daha doğru çıktılar alınması aşamasında geriye doğru hatanın yayılması ile hesaplamalar yapılmaktadır. Bu hesaplamalarda parametreler, gradyan değerleri, ağırlık değeri, çıktı ve hata değerleri kullanılarak güncel ağırlık değerleri hesaplanmaktadır.

- **Ara katman ve çıktı katmanı arasındaki ağırlıkların değiştirilmesi:** Bu aşamada öncelikle çıktı katmanındaki hücrenin hata değeri bulunmalıdır (Öğücü, 2006: 47).

$$E_q = T_q - F_{cq}$$

Eşitlikteki E_q değeri, çıktı katmanındaki q indisli hücrenin hata değerini, T_q değeri, çıktı katmanındaki q indisli hücrede olması beklenen çıktı değerini, F_{cq} değeri çıktı katmanındaki q indisli hücrenin hesaplanan değerini ifade etmektedir.

Çıktı katmandaki q indisli hücrenin hata değeri negatif olarak ta bulunabilir ancak negatif olan değer ile hesaplama yapılamaz. Negatif hata değerinden kurtulabilmek için karesel hata tanımlanması gerekmektedir (Öğücü, 2006: 47).

$$E_q^2 = (T_q - F_{cq})^2$$

Delta kuralına göre ağırlıklardaki değişim, karesel hatanın bu ağırlığa göre değişimiyle orantılıdır. Diğer bir ifade ile hatayı minimum yapabilmek için ağırlıklar hatanın negatif gradyanı yönünde değiştirilmelidir (Öğücü, 2006: 47). Hatanın negatif gradyanı, kısmi türev ve zincir kuralı ile gradyan değerine ulaşılması gerekmektedir.

$$\Delta W_{ap,cq} = -\eta_{ap,cq} \frac{\partial E_q^2}{\partial w_{ap,cq}}$$

Eşitlikteki $w_{ap,cq}$ değeri, ara katmandaki p indisli hücreyi, çıktı katmanındaki q indisli hücreye bağlayan ağırlık değeridir. η değeri öğrenme katsayısı değeridir. Burada gradyan değeri belirtildiği gibi karesel hata değeri ile ağırlık değerinin orantısı ile bulunmaktadır. Gerekli işlemler gerçekleştirildikten sonra sigmoid aktivasyon fonksiyonu kullanıldığı durumda şu şekilde bir eşitlik elde edilir (Öğücü, 2006: 47).

$$\Delta W_{ap,cq} = -\eta_{ap,cq} \cdot -2(T_q - F_{cq}) \cdot F_{cq}(1-F_{cq}) \cdot F_{ap}$$

Kullanılan aktivasyon fonksiyonunun tanh olması durumunda şu şekilde bir eşitlik elde edilecektir;

$$\Delta w_{ap,cq} = -\eta_{ap,cq} \cdot -2(T_q - F_{cq}) \cdot (1 - F_{cq}^2) \cdot F_{ap}$$

Eşitliklerdeki F_{ap} değeri, modelin gizli katmanındaki p indisli hücrenin ileriye doğru hesaplamalar sonucunda üretmiş olduğu çıktı değerini temsil etmektedir.

Elde edilen eşitlik, ağırlığın değişimini belirleyen değişimdir. Bu değişim yeni iterasyondaki ağırlığın belirlenmesi için, momentum katsayısı ile çarpılarak yeni değerin değişimini belirler.

$$w_{ap,cq}(t) = w_{ap,cq}(t-1) + \alpha \Delta w_{ap,cq}$$

Bu eşitlikteki α değeri, momentum katsayısını ifade etmektedir. Hesaplanan ağırlık değişimi ile çarpılarak önceki ağırlığa eklenir ve yeni ağırlık değerini belirlerler.

- **Giriş katmanı ve ara katman arasındaki ağırlıkların değiştirilmesi:** Bu aşamada ara katman ve çıktı katmanı arasındaki ağırlık değişiminde kullanılan hücrenin hatasından farklı olarak, hücrenin hatası değil toplam hata kullanılarak hesaplama yapılmaktadır (Öğücü, 2006: 48).

$$E^2 = \sum_{q=1}^r E_q^2$$

Eşitlikteki q değeri, çıkış katmanındaki hücre indisi, r değeri, çıktı katmanındaki toplam hücre sayısını ifade etmektedir. Bu bilgi ile toplam hata hesaplanmıştır ve giriş katmanı ile ara katman arasındaki ağırlık değişiminin hesaplanmasında kullanılmaktadır (Öğücü, 2006: 48).

$$\Delta W_{gh,ap} = -\eta_{gh,ap} \frac{\partial E^2}{\partial W_{gh,ap}} = -\eta_{gh,ap} \sum_{q=1}^r \frac{\partial E_q^2}{\partial W_{gh,ap}}$$

Eşitlikteki $W_{gh,ap}$ değeri, giriş katmanındaki h indisli hücreyi, ara katmandaki p indisli hücreye bağlayan ağırlık değerini temsil etmektedir.

Gerekli işlemler sonuçlandırıldıktan sonra ağırlığın değişim miktarı sigmoid aktivasyon fonksiyonu kullanıldığında şu şekilde ifade edilecektir;

$$\Delta W_{gh,ap} = -\eta_{gh,ap} \sum_{q=1}^r -2(T_q - F_{cq}) \cdot F_{cq}(1 - F_{cq}) w_{ap,cq} \cdot F_{ap}(1 - F_{ap}) \cdot x_h$$

Eşitlikteki x_h değeri, çıktı katmanındaki h indisli hücrenin ileri hesaplama işlemleri sonrası hücrenin girdilerinden ürettiği çıktısını temsil etmektedir. Ağırlığın değişimi için şu şekilde bir hesaplama gerçekleştirilir (Öğücü, 2006: 48).

$$w_{gh,ap}(t) = w_{gh,ap}(t-1) + \alpha \Delta w_{gh,ap}$$

Bu şekilde ağırlıklar bir sonraki iterasyonda hesaplanarak hatanın azaltılması için hesaplanarak iyileştirilir. Her iterasyonda her hücre ve ağırlık değeri için bu hesaplamalar gerçekleştirilmektedir.

3.5.4 ÇKA'ların Çalışma Prosedürü

ÇKA ağlarının çalışma adımları prosedür olarak şu şekildedir:

- **Örneklerin toplanması:** Ağın çözüm bulması istenilen probleme istinaden veri setleri oluşturulması aşamasıdır. Ağ bu veri setleri ile eğitileceğinden modelin problemi algılaması için bu örnekler büyük önem taşımaktadır (Öztemel, 2020: 81).
- **Ağın topolojik yapısının belirlenmesi:** Ağdaki giriş katmanı, ara katman ve çıktı katmanının hücre sayılarının belirlenmesi, ara katman sayısının belirlenmesi gibi fiziksel yapıların kurgulanması aşamasıdır (Öztemel, 2020: 81).
- **Öğrenme parametrelerinin belirlenmesi:** Aktivasyon fonksiyonu, öğrenme katsayısı ve momentum katsayısı, toplama fonksiyonu gibi öğrenme aşamasına doğrudan etki eden parametreler kurgulanır (Öztemel, 2020: 81).
- **Öğrenme setinden örneklerin seçilmesi ve ağa gösterilmesi:** Ağın eğitilmesi için toplanan örnekler ile ağın eğitimine başlanması adımıdır.
- **Öğrenme sırasında ileri hesaplamaların yapılması:** Öğrenme aşamasında girdi verilerine karşılık gelen çıktı değerlerinin hesaplanma aşamasıdır.
- **Gerçekleşen çıktının beklenen çıktı ile karşılaştırılması:** Girdi verilerine karşılık olarak hesaplanan çıktı değerlerinin, üretilmesi beklenen çıktı değerleri ile karşılaştırılarak hata değerinin elde edilmesi (Öztemel, 2020: 81).
- **Ağırlıkların değiştirilmesi:** Elde edilen bu hata değeri ile diğer parametrelerin, gerekli hesaplamalara katılarak değişim için gereken değerlerin elde edilerek ağırlık değerlerinin diğer iterasyonda hata değerini daha da azaltmak için değiştirilmesi işlemi gerçekleştirilir.

Bu işlem adımlarında öğrenme faaliyeti kendini tekrar ederek ağırlık değişimi hata minimuma indirgenene kadar devam eder. Ağın öğrenme aşamasında ürettiği hatanın belirlenmiş bir oran altına düşmesi ağın öğrenmesini durduracak durdurma kriteridir (Öztemel, 2020: 81).

DÖRDÜNCÜ BÖLÜM

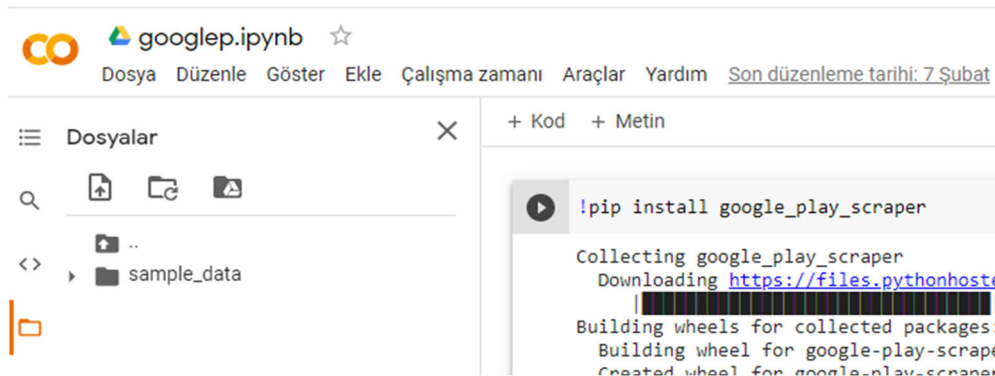
WHATSAPP YENİ GİZLİLİK SÖZLEŞMESİNE YÖNELİK YORUMLARIN DUYGU ANALİZİ

4.1 YAZILIM VE GELİŞTİRME ORTAMI

Araştırmanın kodlama kısmı tamamen Python yazılım dili ile geliştirilmiştir. Python'un "Sklearn", "Transformers", "Pandas", "Pytorch" kütüphaneleri yapay sinir ağı modellemesi, verilerin görselleştirilmesi için kodların çağırılması adına kullanılmıştır. Sklearn, yapay öğrenme için gereken kodların bulunduğu bir yapay sinir ağı kütüphanesidir.

Araştırmanın geliştirme ortamı, web üzerinden başka geliştiriciler ile çalışma ve paylaşma imkanı sunan Google tarafından geliştirilen bulut tabanlı Colab geliştirme ortamıdır. Colab özellikle Python dili ile veri analizi, istatistiksel modellemeler ve veri madenciliği konularında çalışılacak PaaS (Hizmet olarak Platform, Platform as a Service) olarak sunulan bir web geliştirme ortamıdır. Bu geliştirme ortamının en büyük özelliği, yüksek performans gerektiren işlemlerin olduğu projelerde Google'ın sunucuları vasıtasıyla yüksek işlem güçlü GPU'lar ile çalışma olanağı sunmasıdır. Projenin çalışma zamanı ayarları kısmından GPU seçilerek Google'ın yüksek performanslı işlemcileri üzerinden projenin çalıştırılmasına olanak sağlamaktadır. Colab çalışma ortamından örnek bir görsel Şekil 16'da gösterilmektedir.

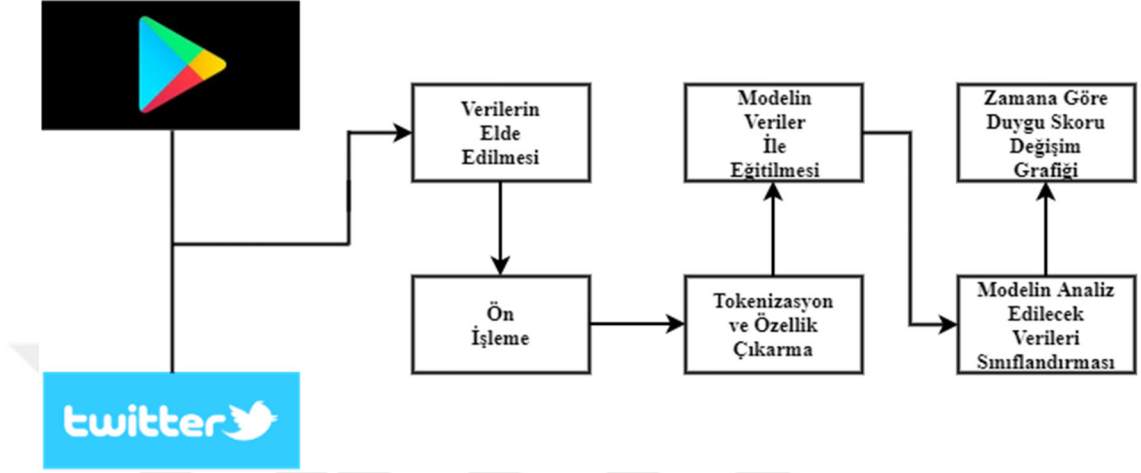
Şekil 16: Google Colab Çalışma Ortamı



Kaynak: Yazar Tarafından Derlenmiştir.

Uygulamanın aşamalarını belirten akış diyagramı Şekil 17’de gösterilmektedir.

Şekil 17: Uygulamanın Akış Diyagramı



Kaynak: Yazar Tarafından Derlenmiştir.

4.2 VERİ SETİ OLUŞTURMA

Veri seti oluşturma YSA modellemesinde en önemli aşamalardan birisidir. Bu aşamada, problemin çözümü için en gerekli verilerden hedef kaynağa mümkün oldukça yakın örnekler ve problem ile alakalı olan verilerden veri seti oluşturulmalıdır.

Araştırmada, eğitim ve test setinde kullanılması için Google Play Store üzerindeki WhatsApp uygulaması için yapılan yorumlar ile Twitter’da WhatsApp hakkında atılan tweetler kullanılmıştır. WhatsApp uygulamasına dair yeni gizlilik sözleşmesi hakkındaki yorumlar toplanarak bu yorumların WhatsApp açısından olumlu yargıya sahip olanlarına “pozitif” etiketi, nötr duyguya sahip olanlarına “nötr” etiketi, olumsuz duyguya sahip olanlarına “negatif” etiketi eklenerek bir veri seti derlemesi yapılmıştır. Google Play Store yorumları ile Twitter verilerinin büyük benzerlikler göstermesi ve WhatsApp’a karşı olan tutumu doğru bir şekilde temsil etmesi sebebi ile yorumlar ve tweetler birleştirilerek kullanılmıştır.

Araştırmada, duygu durumlarının ortaya çıkarılması istenen WhatsApp uygulaması hakkındaki, WhatsApp yeni gizlilik sözleşmesinin kullanıcılara iletilmesinden öncesi ve sonrasını (02.01.2021-12.01.2021) kapsayan tweetler,

tweetlerin atıldığı tarih ve saat etiketleri ile birlikte derlenerek bir veri seti hazırlanmıştır.

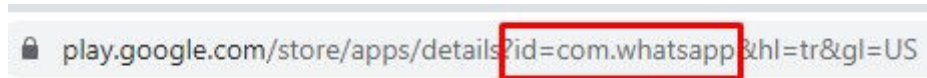
4.2.1 Google Play Store'dan Veri Çekme

Python programlama dili ile çeşitli web sitelerden, istenilen sorgu üzerine verilerin çekildiği kazıma yazılımları (web scraping) mevcuttur. Kazıma yazılımları, kazıma yapılacak platforma göre şekilsel farklılıklar göstermelerine rağmen, temelde kullanıldıkları alan ve ürettikleri sonuç çok benzerdir. Örneğin, aynı ürünün farklı sitelerdeki fiyat ve özel bilgilerini bir platformda toplayıp, kullanıma sunan uygulamalar da scraping yazılımları kullanmaktadır.

Araştırmada Google Play Store platformundan kullanıcı yorumlarının kazınması için Python dili için geliştirilen “google_play_scraper” kütüphanesi kullanılmıştır.

Bu kütüphanenin kullanımı için Google Play sitesinin adres uzantısında, hakkındaki yorumların kazınması istenen uygulamanın sitedeki sayfasına ulaşılabilmesi için eklenmesi gereken uzantı yazılarak o uygulama sayfası işaret edilir.

Şekil 18: Google Play Store Url'si



Kaynak: Yazar Tarafından Derlenmiştir.

WhatsApp için bu uzantı “com.whatsapp” tır. Sonrasında kazıma yapılacak yorumların istenilen nitelikleri belirtilerek yorumların o niteliklere sahip olanları siteden kazınır. Kazınan yorumlar Pandas ve CSV kütüphaneleri kullanılarak .csv formatında saklanmaktadır.

4.2.2 Twitter'dan Veri Çekme

Araştırmada WhatsApp hakkında 02.01.2021 ile 12.01.2021 tarihleri arasında atılmış olan tweetler ve bu tweetlerin atılma tarihleri yine bir scraping kütüphanesi olan “snsrape” kütüphanesi kullanılarak çekilmiştir. Bu kütüphanenin “twitter”

modülü kullanılmaktadır. Öncelikle bir sorgu kelimesi belirlenir, sonrasında bu sorgu kelimesinin içinde bulunduğu kazınmış tweetler kümesi içinde detay barındıran bir sorgu yapılır. WhatsApp kelimesini barındıran, dili Türkçe olan ve 02.01.2021 ile 12.01.2021 tarihleri arasında atılmış tweetler bu scraping sorgusundan geçirilir. Bu işlem bir döngü içerisinde gerçekleştirilir. Twitter’ın arama çubuğunda detaylı arama kısmında belirtilecek detaylı sorgu parametreleri (tweet dili, tweet tarihi aralığı gibi) burada belirtilerek kazıyıcı kütüphanesinin sayfadan tam olarak bu sorguyu girerek verileri kazınmasını sağlanmaktadır. Kazınan tweetler bu aşamada tüm bilgileri ile kazınmış olur ancak uygulamada sadece tweet içeriği ve tarihi kullanılacaktır. Bu nedenle kazınmış tweetlerin sadece “tweet.date” komutu ile tarihini ve “tweet.content” komutu ile içeriği listelenmiş olacaktır. Bu kazınmış ve listelenmiş tweet bilgilerini CSV kütüphanesi kullanılarak .csv uzantılı dosyaya aktarılmaktadır.

WhatsApp hakkında kazınan 202.149 tweet elde edilerek ön işleme aşaması için hazır hale getirilmiştir. Günlere bölünmüş tweet sayıları Tablo 4’te belirtilmektedir.

Tablo 4: Elde Edilen Tweet Sayısı ve Tweetlerin Tarihleri

Tarih	2 Ocak 2021	3 Ocak 2021	4 Ocak 2021	5 Ocak 2021	6 Ocak 2021
Tweet Sayısı	2008	2102	2140	2250	2329
Tarih	7 Ocak 2021	8 Ocak 2021	9 Ocak 2021	10 Ocak 2021	11 Ocak 2021
Tweet Sayısı	36262	7102	38516	83630	25810

Kaynak: Yazar Tarafından Derlenmiştir.

4.3 VERİLERİN ÖN İŞLEMESİ

Bu kısımda CSV dosyaları içerisinde basit excel fonksiyonları kullanılarak gerekli ayıklamalar ve kısaltma düzeltmeleri yapılmıştır.

Excel’in “Bul-Değiştir” fonksiyonu ile “Wp”-“whatsap”-“knk” kısaltmaları ve yazım hataları düzeltilmiştir. Tweetler yazar tarafından taranarak WhatsApp ile ilgili olmayan, yeni gizlilik sözleşmesi ile alakalı olmayan reklam amaçlı ya da irtibat için tweette WhatsApp kelimesi geçmiş olan tweetler listeden ayrıştırılmıştır.

Benzer şekilde, Google Play yorumları üzerinde de işlem gerçekleştirilerek ayıklama ve düzenleme yapılmıştır. Google Play yorumları, modelin eğitiminde kullanılacağından yorumların duygu durumlarına yönelik işaretler konulmuştur. Negatif duygu barındıran Google Play yorumları “negatif”, nötr duygu barındıran Google Play yorumları “nötr”, Pozitif duygu barındıran Google Play yorumları “pozitif” olarak işaretlenmiştir. Bu işaretler duygu sınıfları olarak YSA modeli tarafından kullanılacaktır.

Eldeki CSV dosyaları uygulama tarafından anlaşılması için JSON formatına çevrilmiştir. JSON formatlarında anahtar-değer (key-value) yöntemi mevcuttur, yani her bir içeriğin bir değer eşleniği bulunmalıdır. Excel üzerinde metin birleştirme fonksiyonları ile veri setleri aşağıdaki şekilde JSON verisine dönüştürülmüştür.

- **Google Play veri seti için:** `{"sentence": "yapılan yorum", "value": "duygu sınıfı"}`
- **Twitter veri seti için:** `{"datetime": "2021-01-06 23:58:00", "text": "tweet metni"}`

Bu aşamadan sonraki düzeltme ise veriler araştırma kodlarına aktarıldıktan sonra bir filtreleme sağlayan kod bloğu ile gerçekleştirilmiştir. “@” ile başlayan tweetlerde, “.com”, “.org” gibi uzantılarda biten tweetlerde, “pic”, “http” veya “www” ile başlayan tweetlerde düzeltmeler gerçekleştirilmiştir.

Bu düzenlemeler sonrasında tweet sayısı azalarak elde edilen 202.149 tweetin 198.922 adeti ile çalışmaya devam edilmiştir.

4.4 BERT MODELİNİN KURULMASI

Araştırmada Türkçe metin verilerini anlamlandırabilmek için orijinal BERT modelinden türetilmiş olan BERTurk dil modeli ile çalışılmıştır. Bu kısımda web üzerinde bir geliştirme platformu olan Colab’ın da yardımı ile kısa bir kod bloğu yazarak BERTurk modeli projeye internetten getirilmekte ve hazır hale gelmektedir.

BERT modelinin dili algılama konusunda kullanılabilmesi için BERT içindeki tokenizer ve model yapıları edinilmelidir. Bu internetten indirilen iki yapının tokenizer ve dil modeli olarak kullanılacağı projeye Python kütüphanelerinden Transformers kütüphanesinin AutoModel ve AutoTokenizer modelleri kullanılarak

gerçekleştirilmiştir. Bu modellerin “from_pretrained” özelliği ile de iki BERT elemanının kısa kodlarla indirilip projede kullanılması sağlanmıştır.

- **Kütüphane için:** from transformers import AutoModel, AutoTokenizer (Schweter, 2020).
- **Tokenizer kodu:** tokenizer = AutoTokenizer.from_pretrained(“dbmdz/bert-base-turkish-128k-uncased”) (Schweter, 2020).
- **Model kodu:** bert = AutoModel.from_pretrained(“dbmdz/bert-base-turkish-128k-uncased”).to(device) (Schweter, 2020).

Bu kodlardaki “.to(device) kod parçası önceden belirtilen Colab platformunun büyük boyutlu projeler için kullanıma izin verdiği Google’a ait olan yüksek kapasiteli işlemlere modelin yüklenip çalıştırılmasını sağlamaktadır. Bu kodlardan önceki kod bloklarında “device” adlı bir değişken tanımlanmış ve “cuda” metinsel değeri bu değişkene atanmıştır.

Cuda, Nvidia’nın sunduğu bir mimaridir. Bu mimari yazılan kodların GPU’lar üzerinde çalışmasını sağlar (Wikipedia, 2020). Bu sayede Cuda mimari teknolojisi, projenin Google GPU’larında çalıştırılmasını sağlayan arabirim olma özelliğini taşır.

BERTurk modelinin tokenizer kütüphanesi ile Türkçe tweetler sözcük parçacıklarına ayrıştırılmış, sonrasında ise BERTurk modeli ile bu sözcük parçacıklarının özellik çıkarma için vektör değerlerinin oluşturulması sağlanmıştır.

4.5 ÇKA MODELİNİN KURULMASI

Bu uygulamada Python içindeki “Sklearn” kütüphanesinin “neural_networks” metodu ve bu metodun “MLPClassifier” objesi kullanılmıştır. Bu kütüphane ve metodları kullanılarak bir ÇKA ağı Python projesi üzerinde modellenenilmektedir. MLPClassifier metodunun kullanılması durumunda bu ÇKA ağı, girdileri sınıflandırmaya yönelik çalışma gerçekleştirecektir.

Kurgulanan model için Python kodları ile bazı parametreler belirtilerek modelin yapısı projede belirtilmektedir. “hidden_layer_sizes” gizli katmanların sayısı ve bu katman içerisindeki hücre sayısını, “max_iter” ağı eğitimi kısmında maksimum ne kadar tekrar yapacağını, “activation” ise ağı hücrelerinin aktivasyon fonksiyonunun hangisi olacağını, “early-stopping” ise ağıdaki hata oranının en az

olduğu durumlarda eğitim bölümünün sonuçlandırılıp sonuçlandırılmayacağını belirtir ve bu parametreler ile model eğitim aşamasına başlar.

Eğitim aşamasındaki her eğitim turuna epoch adı verilir ve her epochtaki hata oranı eğitim boyunca ekrana yazdırılır. Eğitimde hatanın en minimal hali elde edildiğinde eğitim durdurulmakta ve eğitim ile test verisindeki başarı oranı ekrana yazdırılmaktadır. Bu oran yine “Sklearn” kütüphanesinin “metrics” metodunun clasification_report nesnesi ile gösterilmektedir.

Eğitim aşamasının başarı oranının ölçümü ise konfüzyon (karışıklık) matrisi yöntemi ile sağlanmıştır.

Konfüzyon (Karışıklık) matrisi bir sınıflandırma sistemi tarafından yapılan gerçek ve tahmin edilen sınıflandırmaların başarısı hakkında bilgi içeren bir makine öğrenimi kavramıdır (Deng ve diğerleri, 2016: 251). Başarılı tahmin sayısı ve başarısız tahmin sayılarının bir matris içerisinde ele alındığı bir yapıdır. Verilerin sınıflandırıldığı sınıf kümelerinin her biri için modelin doğru tahmin ettiği veri sayısı ve yanlış tahmin sayıları matris olarak tutulur ve bu verilerden yola çıkarak modelin öğrenme performansı ölçülmektedir. Tablo 5’te konfüzyon matrisi içerisinde verilerin yerleşimi ve anlamları gösterilmektedir.

Tablo 5: Konfüzyon Matrisi Örneği

	Gerçekte Pozitif	Gerçekte Nötr	Gerçekte Negatif
Pozitif Olarak Tahmin Edilen	Doğru Pozitifler (DP)	Yanlış Pozitifler (YP)	Yanlış Pozitifler (YP)
Nötr Olarak Tahmin Edilen	Yanlış Nötrler (YNÖ)	Doğru Nötrler (DNÖ)	Yanlış Nötrler (YNÖ)
Negatif Olarak Tahmin Edilen	Yanlış Negatifler (YN)	Yanlış Negatifler (YN)	Doğru Negatifler (DN)

Kaynak: Yazar Tarafından Derlenmiştir.

Tablo 5’te görüldüğü gibi bir matris oluşturulur ve bu matrisin içerisindeki “Doğru Pozitifler”(DP) doğru tahmin edilen pozitif değerlerin, “Yanlış Pozitifler”(YP) yanlış tahmin edilmiş pozitif değerlerin, “Yanlış Negatifler”(YN) yanlış tahmin edilmiş negatif değerlerin, “Doğru Negatifler”(DN) doğru tahmin edilmiş negatif

değerlerin, “Doğru Nötrler”(DNÖ) doğru tahmin edilmiş nötr değerlerin ve “Yanlış Nötrler”(YNÖ) yanlış tahmin edilmiş nötr değerlerin toplam sayıları ile oluşturulan yüzdelik oranlardır.

Performans ölçütlerinde dört adet ölçüm değeri mevcuttur;

- **Doğruluk (Accuracy):** Modelin tahminlerinin doğruluğunu ölçmektedir. Tüm doğru tahmin sayısının tüm verilere bölümü ile bulunur (Deng ve diğerleri, 2016: 252).

$$\text{Doğruluk} = (DP + DN + DNÖ) / (\text{Tüm veri kümesinin boyutu})$$

- **Kesinlik (Precision):** Modelin her sınıf kümesine özel olarak tahminlerinin doğruluğunun ölçüm sonucudur.

$$\text{Kesinlik} = (DP) / (DP + YP)$$

- **Duyarlılık (Recall):** Modelin her sınıf kümesi için doğru tahmin etmiş olması gereken optimum değere göre ne kadar başarılı olduğu ölçümlemesinin sonucudur.

$$\text{Duyarlılık} = (DP) / (DP + YN + YNÖ)$$

- **F1 Skoru:** Kesinlik ve Duyarlılık ölçüm sonuçlarının harmonik ortalaması alınarak oluşturulan bir başarı ölçüsüdür (Deng ve diğerleri, 2016: 252).

$$F1 \text{ Skoru} = 2 * (\text{Kesinlik} * \text{Duyarlılık} / (\text{Kesinlik} + \text{Duyarlılık}))$$

Bu tezin uygulamasında pozitif, negatif ve nötr olarak üç adet sınıflandırma kümesi mevcuttur. Konfüzyon matrisi hesaplamasında bu üç sınıfa dair modelin tahminleme başarısı doğruluk, kesinlik, duyarlılık ve f1 skoru metrikleri ile ölçülmüştür.

4.6 ANALİZ SÜRECİ

Veriler BERT Tokenizer modeli ile sayısal olarak temsil edildikten sonra BERT modeli ile verilerin temsilleri sonuçlandırılmıştır.

Eğitim ve Test vektörleri bir matris düzeninde negatif olan değerler 0, nötr değerler 1 ve pozitif olan değerler 2 etiketi ile birleştirilerek eğitim aşamasına hazır hale getirilmiştir.

Birleştirilen etiket ve değerler modellenen ÇKA ağına aktarılarak ağı eğitim işlemi sonuçlandırılmıştır.

Sonrasında analiz edilmek istenen tweet veri seti sayısal olarak vektörler ile temsil edildikten sonra ÇKA modeline 0, 1 ve 2 sınıfları ile sınıflandırma yapması için iletilmiş, sınıflandırma işleminin yapılacağı döngü içerisinde sınıf değeri bulunan her

girdi için sınıf değerinden 1 sayı eksiltmiştir. Bu şekilde önceden 0 ile negatif değerler temsil edilirken -1 ile, 1 ile nötr değerler temsil edilirken 0 ile ve 2 ile pozitif değerler temsil edilirken 1 ile temsil edilmişlerdir.

Sonrasında bu tweetler tarih etiketlerine göre ayrıştırılarak her tarih ve saat verisi için sınıf etiketlerinin (-1, 0, 1) aritmetik ortalaması alınarak tarih bazlı duygu skor puanı elde edilmiştir.

ÇKA ağının eğitim süreci için 2565 adet veri içeren bir veri seti oluşturulmuştur. Derlenen verilerin %60'ı eğitim veri seti, %20'si test veri seti ve diğer %20'si değerlendirme veri seti olarak bölünmüştür. Verilerin duygu etiketleri tezin yazarı tarafından atanmıştır. Derlenen verilerin içerisinde 855 adet pozitif, 855 adet negatif ve 855 adet nötr yorum mevcuttur. Bu veri setinin bölünmesi için sklearn kütüphanesinin train_test_split fonksiyonu kullanılarak verilerin karıştırılıp seçilmesi amaçlanmıştır. 182 adet nötr, 152 adet pozitif ve 179 adet negatif örnek eğitim aşamasının test çalışmasına dahil edilmiştir. Eğitim, test ve değerlendirme veri setlerinin veri sayıları ve yüzdeleri Tablo 6'da belirtilmiştir.

Tablo 6: Eğitim, Test ve Değerlendirme Veri Setlerindeki Veri Sayıları

Eğitim Veri Seti	Test Veri Seti	Değerlendirme Veri Seti
1539 (%60)	513 (%20)	513 (%20)

Kaynak: Yazar Tarafından Derlenmiştir.

Bu veri setleri ile yapılan ÇKA ağı eğitiminde %80 oranında başarı elde edilmiştir. Yapılan eğitimin başarı ölçütleri Tablo 7'de belirtilmiştir.

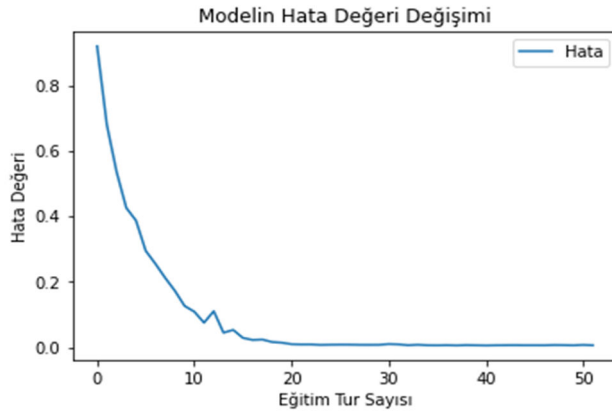
Tablo 7: Tasarlanan Modelin Başarı Değerleri

	Kesinlik	Duyarlılık	F1 Skoru	Veri Sayısı
Negatif	0,75	0,80	0,77	179
Nötr	0,84	0,81	0,83	182
Pozitif	0,80	0,77	0,79	152
Doğruluk	0,80			

Kaynak: Yazar Tarafından Derlenmiştir.

Eğitim sonucunda toplam doğruluk oranı %80 olarak bulunmaktadır. Duygu sınıflarına özel kesinlik sonuçlarında tasarlanan ÇKA modelinin en başarılı tahminde bulunduğu sınıf nötr duygu sınıfı olurken negatif duygu sınıfında başarı oranı diğer sınıflara oranla daha düşüktür. Modelin ürettiği hata oranının eğri grafiği Şekil 19’da gösterilmektedir.

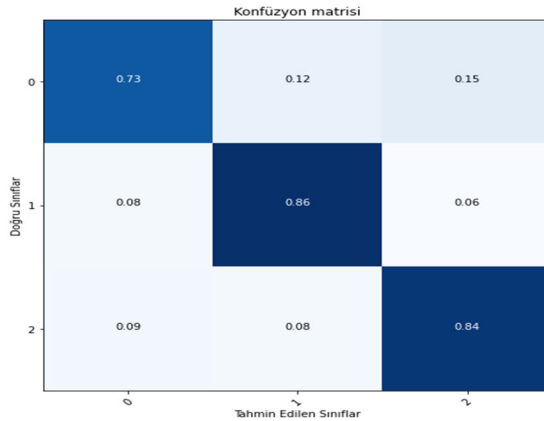
Şekil 19: Tasarlanan Modelin Hata Değerlerinin Değişim Tablosu



Kaynak: Yazar Tarafından Derlenmiştir.

Şekil 19’da belirtilen hata değeri değişimi, her eğitim turu (epoch) için oluşan hatanın, turlar sırasında değişimi grafik olarak belirtmektedir. Aradaki dalgalanmaların sebebi metinlerin küçük parçalar ile modele iletilmesi nedeni ile doğal bir durumdur. Şekil 20’de tasarlanan modelin tahminleri ile oluşturulan konfüzyon matrisi gösterilmektedir.

Şekil 20: Tasarlanan Modelin Konfüzyon Matrisi



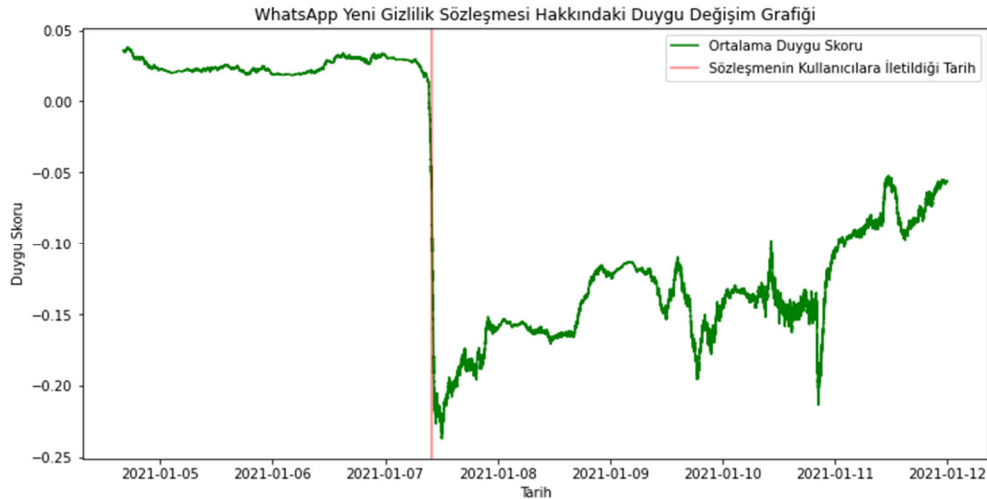
Kaynak: Yazar Tarafından Derlenmiştir.

Şekil 20’de görülen 0, 1 ve 2 sınıfları bu aşamadan sonra ortalama duygu skoru hesaplaması yapılabilmesi için -1, 0 ve 1 olarak değiştirilecektir. Bu durumda 0 sınıfı negatif, 1 sınıfı nötr ve 2 sınıfı pozitif olarak belirtilen duygu sınıflarını belirtmektedir. Modelin konfüzyon matrisine göre negatif değerlerde 0,73, nötr değerlerde 0,86 ve pozitif değerlerde 0,84 doğruluk sağlanmıştır. Modelin eğitimi sonrası negatif olan verilerin 0,12’si nötr olarak, 0,15’i pozitif olarak değerlendirilmiştir. Nötr olan verilerin 0,08’i negatif olarak, 0,06’sı pozitif olarak değerlendirilmiştir ve pozitif olan verilerin 0,09’u negatif olarak, 0,08’i nötr olarak değerlendirmiştir.

%80 başarı oranı ile eğitim gerçekleştirilen ÇKA modelinde aktivasyon fonksiyonu olarak ReLU, Gizli katman sayısı 4 ve ağlardaki işlem hücreleri sayısı sırası ile 1024, 2048, 768, 64 adet olarak belirlenmiştir.

Eğitilmiş model ile 198.922 adet WhatsApp hakkındaki 2-12 Ocak 2021 tarihleri arasında atılmış tweetlerin duygu sınıfları ile sınıflandırılmasını sağlanmış ve değişim tablosu oluşturulmuştur. Tweetlerin duygu sınıflarının ortalamaları alınarak tarih etiketlerine göre tarih bazlı duygu skor ortalamaları bir eğri grafiğinde görselleştirilmiştir. Oluşturulan duygu skoru değişim grafiği Şekil 21’de gösterilmektedir.

Şekil 21: Duygu Skorları Ortalamasının Tarihlere Göre Değişim Grafiği



Kaynak: Yazar Tarafından Derlenmiştir.

Şekil 21’ de görüldüğü üzere yeni gizlilik sözleşmesinin açıklandığı tarih periyodunda 0,05 duygu skorundan -0,23 duygu skoruna aşağı yönlü bir düşüş olduğu

görülmektedir. 2-3 Ocak tarihlerinde duygu skor değişimi çok düşük olduğundan değişim ortalamasında görselleştirilememektedir. Atılan tweetler ve modelin bu tweetlere karşılık ürettiği duygu sınıflarından bir kısmı örnek olması için Tablo 8’de gösterilmiştir.

Tablo 8: Modelin WhatsApp Tweetleri İçin Ürettiği Duygu Sınıfı Tahminleri Arasından Bazı Örnekler

Tweet	Duygu
En çok şeye anlam veremedim diğer uygulamaların WhatsApp dan daha masum olduğuna nasıl hemen inandınız?	Pozitif
WhatsApp'ı siliyoruz...Boykot	Negatif
WhatsApp ve instagramı zuckerberge satan adamlar bu konu hakkında ne düşünüyorlardır acaba?	Nötr

Kaynak: Yazar Tarafından Derlenmiştir.

Tablo 8’de görüldüğü üzere, WhatsApp uygulamasını silme veya başka bir uygulamaya geçiş ile alakalı tweetler negatif olarak, WhatsApp gizlilik sözleşmesi hakkında bir duygu belirtmeyen tweetler nötr olarak ve WhatsApp ile ilgili daha olumlu ifadelerle sahip olan tweetler pozitif olarak sınıflandırılmaya çalışılmıştır. Modelin yaptığı tahminler sonucunda duygu sınıflarının toplam tweet sayıları Tablo 9’da belirtilmektedir.

Tablo 9: Modelin Ürettiği Tahminler Sonrası Duygu Sınıflarının Tweet Sayıları

Duygu Sınıfı	Negatif	Nötr	Pozitif
Tweet Sayısı	74519	77126	47277

Kaynak: Yazar Tarafından Derlenmiştir

Model tarafından 198.922 adet tweetin 74.519 adeti negatif, 77.126 adeti nötr ve 47.277 adeti pozitif olarak sınıflandırılmıştır.

Bu sonuçların değişkenliği üzerine kullanıcıların büyük bir kısmının yeni gizlilik sözleşmesinden hoşnut olmaması ve karşıt olarak sözleşmenin özel hayatın gizliliğine bir ihlal yaratmayacağı olarak karşıt fikirlerde birleşildiği, ancak duygu

skorunun düşüşünün de gösterdiği üzere kullanıcıların çoğunluğunun bu karara tepkili olduğu görülmektedir.



SONUÇ

Bilgisayar ve bilişim teknolojilerinin hayatın her alanında gelişmekte olması, her konuda yeniliklerle tanışmamıza olanak sağlamaktadır. Bu yeniliklerin hız kesmeden devam etmesi toplumumuzu bilgi çağı düzeyine taşımıştır. Özellikle bu çağda veri depolama kapasitesindeki artışın da etkisi ve büyük veri olarak adlandırılan kolektif veri gruplarının etkin bir şekilde analiz edilmesi ile önemli bilgiler elde edilebilmiş ve bu bilgilere göre üst düzey kararlar alınarak başarılar sağlanmıştır.

Büyük veriyi oluşturan verilerin çoğunluğu yapılandırılmamış verilerden oluşmaktadır. Yapılandırılmamış verilerin analiz edilebilmesi konusunda metin madenciliği alanındaki gelişmeler büyük bir önem taşımaktadır. Metin madenciliği çalışmaları sayesinde büyük veri içerisindeki yapılandırılmamış verilerden kıymetli anlamlar çıkarılabilmiş ve analiz sonuçları ile katma değer yaratılabilmektedir. Metinsel ifadelerin analiz edilmesi ve bilgisayarlar tarafından algılanmasını sağlayan bu çalışmalar birçok alanda fayda sağlayarak veri bilimi alanları içerisinde önemli bir yer edinmiştir.

Metinsel ifadelerin analizinde metnin kaleme alındığı dil ve bu dilin kurallarının model tarafından anlaşılmış olması büyük bir önem arz etmektedir. Dilin kök yapıları, ekler ve anlamların anlaşılmadığı bir proje içerisinde metni anlamlandırmak imkansızdır. DDİ metodları kullanılarak dilin kuralları çerçevesi içinde metinsel ifadeler anlaşılır hale gelmektedir. Metnin anlamlandırılabilmesi için anlaşılması gerekmektedir.

Günümüzde sosyal mecralar üzerindeki paylaşımların etkisi kamuoyu nezdinde kabul edilmiş bir bilgi çağı gerçeğidir. Kişiler, kuruluşlar, ürünler veya hizmetler hakkında sosyal mecralar üzerinden algı ve yaklaşımların belirlenmesi ve bir skala içerisinde ölçülmesi, ilgili konu hakkında atılacak yeni adımların daha sağlam olmasını sağlayacaktır. İlgili kişi, kuruluş, ürün veya hizmetin saygınlığı ve bilinirliği hakkında ölçülebilir verilerden hareketle yeni davranış politikaları, yeni kampanyalar, pazarlama çalışmaları, müşteri sadakat ve tutundurma politikaları, önceden yapılabilecek bir hatalı faaliyetin yarattığı zararları minimum etki ile düzeltmek gibi katma değer sağlanacak somut adımlar planlanabilmektedir.

7 Ocak 2021 tarihinde WhatsApp uygulaması, kullanıcılarına uygulama içi uyarı ile yeni gizlilik sözleşmesi şartlarını iletmiştir. Yenilenen gizlilik politikasını 8 Şubat 2021 tarihine kadar kabul etmeyen kullanıcıların WhatsApp uygulamasını kullanamayacağını belirten bu sözleşme kullanıcıların tepkisine sebep olmuştur (Sabah, 2021). WhatsApp yeni gizlilik sözleşmesi iletisi Şekil 22’de belirtilmiştir.

Şekil 22: WhatsApp Uygulaması Tarafından Kullanıcılarına İletilen Yeni Gizlilik Sözleşmesi



Kaynak: Sabah, 2021

İletideki şartların tartışma konusu olarak gündemi uzun bir süre meşgul etmiş olması ve sosyal medya mecralarında yüksek etkileşim alan konuların başında gelmesi bu çalışmanın motivasyonunu oluşturmuştur.

Literatürde Türkçe metinler üzerinde duygu analizi çalışmalarının sayıca yetersiz olması ve BERT dil modeli kullanılarak yapılan Türkçe dil çalışmalarının çok az örneğine rastlanması da literatüre katkı sağlamak adına bu çalışmanın bir diğer motivasyon kaynağı olmuştur.

Bu tezin uygulama kısmında WhatsApp yeni gizlilik sözleşmesi sonrasında Twitter’da WhatsApp hakkındaki yorumların duygu analizi yapılarak kullanıcıların WhatsApp’a karşı tutumuna yeni gizlilik politikasının etkileri incelenmiştir. WhatsApp firmasının aldığı bu karara kullanıcıların tepkisi duygu analizi ile

ölçülmeye çalışılmıştır. WhatsApp uygulamasının bağlı bulunduğu Facebook şirketinin bu konuda alabileceği yeni aksiyonlar hakkında karar desteği sağlanabilmesi amacıyla çalışma gerçekleştirilmiştir.

Çalışmada milyonlarca kullanıcının konu etiketleri üzerinden yorumlarını ve görüşlerini paylaşarak ilgili konulara yorumları ile etkide bulunabildikleri büyük bir sosyal medya mecrası olması nedeniyle Twitter verileri kullanılmıştır. Özellikle gelişen sosyal medya kültüründe Twitter'in yeri aktif olarak yaşanan olaylara karşı tepkilerin paylaşıldığı bir mecra olması sebebi ile literatürdeki birçok duygu analizi çalışmasında kullanılmış olması, Twitter verilerini önemli kılmaktadır. WhatsApp uygulaması hakkında başka bir yorum mecrası olan Google Play Store kullanıcı yorumu verileri ile çeşitlilik arttırılmış, sınıflandırma problemine karşılık çözümün kolaylaştırılabilmesi için Google Play Store üzerinden WhatsApp yorumları da çalışmaya dahil edilmiştir.

Elde edilen verilerin içerdikleri duygulara göre sınıflandırılması için YSA uygulamalarından ÇKA ağları, diğer makine öğrenmesi algoritmalarına kıyasla literatürdeki Türkçe duygu analizi çalışmalarında en yüksek başarı oranlarını ortaya koyan model olması nedeniyle sınıflandırma probleminin çözümü için çalışmada kullanılmış, diğer makine öğrenmesi algoritmaları çalışmaya dahil edilmemiştir. Literatürdeki diğer çalışmalarda sınıflandırma modellerinin karşılaştırılma verileri Tablo 10, 11 ve 12'de gösterilmektedir.

Tablo 10: Çılgın vd. Sınıflandırma Modeli Başarı Karşılaştırması

Algoritma	Doğruluk (Accuracy) Değeri
Naive Bayes	0.71
Destek Vektör Makineleri	0.70
Lojistik Regresyon	0.68
Yapay Sinir Ağları	0.9058

Kaynak: Çılgın vd., 2020: 120

Tablo 11: Kaynar vd. Sınıflandırma Modeli Başarı Karşılaştırması

	Yöntem	A	P	R	F	AUC
Eğitim	Naive	0.6953	0.6514	0.8353	0.7320	0.6959
	Centroid	0.6753	0.6821	0.6519	0.6667	0.6752
	ANN	0.8973	0.8991	0.8942	0.8966	0.8973
	SVM	0.8407	0.8342	0.8487	0.8414	0.8407
Test	Naive	0.6800	0.6422	0.8300	0.7241	0.6782
	Centroid	0.6720	0.6910	0.6364	0.6626	0.6724
	ANN	0.7520	0.7363	0.7945	0.7643	0.7515
	SVM	0.7500	0.7370	0.7866	0.7610	0.7496
Tümü	Naive	0.6915	0.6490	0.8340	0.7300	0.6915
	Centroid	0.6745	0.6843	0.6480	0.6656	0.6745
	ANN	0.8610	0.8553	0.8690	0.8621	0.8610
	SVM	0.8180	0.8087	0.8330	0.8207	0.8180

Kaynak: Kaynar vd., 2016: 240

Tablo 12: Tuzcu Sınıflandırma Modeli Başarı Karşılaştırması**Tablo 1.** MLP sınıflandırma başarısı

Sınıf	Başarı oranı (%)
Negatif	86
Pozitif	92
Toplam	89

Tablo 2. NB sınıflandırma başarısı

Sınıf	Başarı oranı (%)
Negatif	81
Pozitif	74,14
Toplam	77,57

Tablo 3. DVM sınıflandırma başarısı

Sınıf	Başarı oranı (%)
Negatif	65,71
Pozitif	96,14
Toplam	80,93

Tablo 4. LR sınıflandırma başarısı

Sınıf	Başarı oranı (%)
Negatif	79
Pozitif	89,14
Toplam	84,07

Kaynak: Tuzcu, 2020: 3

Tablo 10, 11 ve 12’de görüldüğü üzere önceki çalışmalarda ÇKA ağlarının diğer makine öğrenmesi algoritmalarına karşı daha başarılı sonuçlar sağladığı anlaşılmaktadır. Başarı kıstası göz önünde bulundurularak bu çalışmada ÇKA ağları kullanılmıştır.

WhatsApp yeni gizlilik sözleşmesinin kullanıcılara belirtildiği tarihi, olayın gerçekleştiği tarih olarak kabul ederek, olaydan hemen önceki ve olaydan sonraki WhatsApp hakkında duygu analizinin gerçekleştirilebilmesi için 2-12 Ocak 2021 tarihleri arasında, içeriğinde WhatsApp kelimesinin geçtiği Türkçe tweetler Python programlama dili ile snsrape veri kazıma kütüphanesi kullanılarak CSV dosyası

formatında elde edilmiştir. Aynı şekilde tweetlerin model tarafından sınıflandırılabilmesi için eğitim kısmında kullanılmak üzere Google Play Store üzerinden WhatsApp uygulamasına yapılan Türkçe yorumlar Python programlama dili ile google_play_scraper veri kazıma kütüphanesi kullanılarak CSV dosyası formatında elde edilmiştir.

Elde edilen verilerde ön işleme çalışması gerçekleştirilerek metinler analizde kullanılmaması gereken içeriklerden ayıklanıp, kısaltmalar gibi doğru algılanması gereken metin ifadelerinin ise düzeltilmesi sağlanmıştır. Ön işleme aşamasında yapılan çalışmalar şu şekildedir;

- Kısaltmalar uzun halleri ile güncellenmiştir
- Analizde kullanılmayacak olan yorumu yazan kullanıcı adı, yorumun yazıldığı konum gibi veriler metinlerden ayrıştırılmıştır
- Tweetler içerisindeki basit imla hataları düzeltilmiştir
- Google Play Store veri seti içerisinde tekrar eden yorumlar ayıklanmıştır ve diğer maddeler Google Play Store veri seti için de uygulanmıştır.

Ön işleme aşamasının tamamlanmasının ardından CSV formatındaki verilerin JSON formatına çevrilmesi işlemi gerçekleştirilmiştir. Bu formata çevirme esnasında eğitim için kullanılacak yorumlar duygu sınıflarına göre pozitif, negatif ve nötr olarak işaretlenmiştir. JSON formatındaki veriler eğitim ve tweet veri setleri olarak tez yazarının şahsi Google Drive bulut depolama alanına yüklenerek Colab üzerinde tez yazarının şahsi Google Drive bulut depolama alanına bağlantı kurularak uygulamaya eklenmiştir.

JSON formatındaki verilerin yüklenmesi sonrası tekrar bir ön işleme işlemi gerçekleştirilmiştir. Python programlama dilinde ayıklama yapılabilmesi için bir fonksiyon kurularak metinlerde şu işlemlerin gerçekleştirilmesi sağlanmıştır;

- Twitter’da @ işareti ile belirtilen kullanıcı adları analizde kullanılmayacağından tweet içerisinden @ işareti ile birlikte ayıklanmıştır
- Tweet içerisinde geçen url, resim ve video içerikleri ayıklanmıştır
- Yukarıdaki iki madde aynı şekilde Google Play Store veri seti için de uygulanmıştır.

Bu aşamadan sonra uygulamaya eklenen dönüştürücü kütüphanesi kullanılarak BERTurk dil modelinin metinleri sözcük parçalarına bölebilmesi için tokenizer modeli ve bu sözcük parçalarının özelliklerine göre sayısal vektör değerlerine dönüştürülmesi için yine dönüştürücü kütüphanesinin model özelliği içerisinde BERTurk modeli indirilmiştir. Özellik çıkarma aşaması için modeller ve torch kütüphanesi kurularak metinlerin vektör gösterimlerinin elde edilip CUDA aracılığı ile Google GPU'larında listelenebilmesi için bir fonksiyon kurulmuştur.

Eğitim verilerini barındıran JSON dosyası Python programlama dilinin “read” özelliği kullanılarak JSON dosyasından okunmuş ve bir değişkene atanmıştır.

Değişkendeki verilerin daha önce bahsedilen fonksiyonlar ile işlenip matris içerisinde saklanması ve saklanan vektör değerlerinin, JSON formatındaki duygu etiketleri negatif olan verilerin 0, nötr olan verilerin 1 ve pozitif olan verilerin 2 sayısı ile etiketlenmesi işlemi gerçekleştirilmiştir. Bu aşamada eğitim verilerinin vektör gösterimleri ve duygu etiketleri elde edilmiştir. Elde edilen eğitim veri seti %60 eğitim, %20 test ve %20 değerlendirme veri seti olarak sklearn kütüphanesinin train_test_split fonksiyonu ile bölünmüş, kurulacak olan ÇKA ağının eğitimi için hazır hale getirilmiştir. Verilerin sınıflandırılması için şu özelliklere sahip olan bir ÇKA ağı kurulmuştur;

- 1 adet giriş katmanı, 4 adet ara katman ve 1 adet çıkış katmanı olmak üzere 6 katman mevcuttur.
- Ara katmanlardaki nöron sayısı sırasıyla 1024, 2048, 768, 64 olarak belirlenmiştir.
- Katmanlardaki aktivasyon fonksiyonu, denemeler sonucu daha hızlı ve üstün performans sağlamış olan ReLU aktivasyon fonksiyonu olarak belirlenmiştir.
- Öğrenme parametreleri için öğrenme katsayısı ve momentum değerini öğrenmenin gidişatına göre güncelleyebilen Adam algoritması ile çalışılmıştır.

Kurulan bu ÇKA ağı eğitim, test ve değerlendirme verileri ile eğitime tabi tutularak %80 oranında başarı elde edilmiştir.

Eğitim veri setinde olduğu gibi analiz edilmesi hedeflenen WhatsApp tweetlerinin bulunduğu hedef veri seti içerisinde ön işleme ve özellik çıkarma fonksiyonları çalıştırılmıştır. Vektör olarak elde edilen hedef veriler model.predict kod parçasıyla tasarlanan ÇKA modeli tarafından hedef verilerin bulunduğu duygu

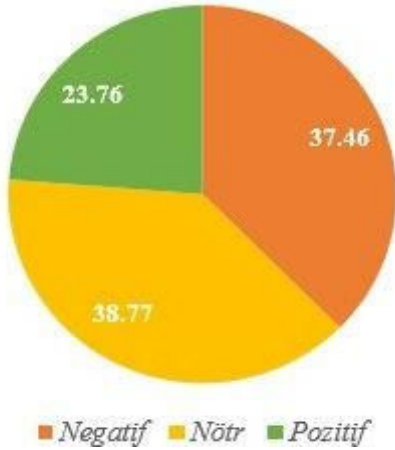
sınıfları tahmin edilmiştir. Hedef verilerine dair ÇKA modelinin tahminleri ile beraber hedef verilerin tarih verisi matris olarak uygulama içerisine kaydedilmiştir. Matris içerisindeki sınıf değerlerinden 1 sayı çıkarılarak 0, 1 ve 2 olan duygu sınıf değerleri, -1, 0 ve 1 olarak güncellenmiştir.

Bu aşamada hedef verilerin duygu sınıfları tahmin edilmiş durumdadır. Verilerin görselleştirilmesi aşamasında ise tweet tarihi ve duygu sınıfları verileri ile tarih bazlı duygu sınıflarının ortalaması alınarak, olayın gerçekleştiği tarih olan 7 Ocak 2021 tarihi üzerinden bir çizgi çekilmiş, zaman bazlı ortalama duygu skorları ile bir değişim grafiği oluşturularak veriler görselleştirilmiştir.

Şekil 21’de detaylı olarak görüldüğü üzere, elde edilen bulgularda WhatsApp’a ilişkin tarihlere bağlı duygu skorlarının 0,05 değerinden olay sonrasında -0,23 skorlarına kadar düşmüş olduğu gözlemlenmiştir. Bu duygu değerlerinin karşılığı olarak uygulamanın aktif kullanıcı sayılarında da değişimler gözlenmiştir. Şekil 26’da tweetlerin duygu durumlarına göre yüzdeleri grafik olarak gösterilmiştir.

Şekil 26: Duygu Sınıflarının Tweet Yüzdeleri

Duygu Sınıflarının Tweet Yüzdeleri

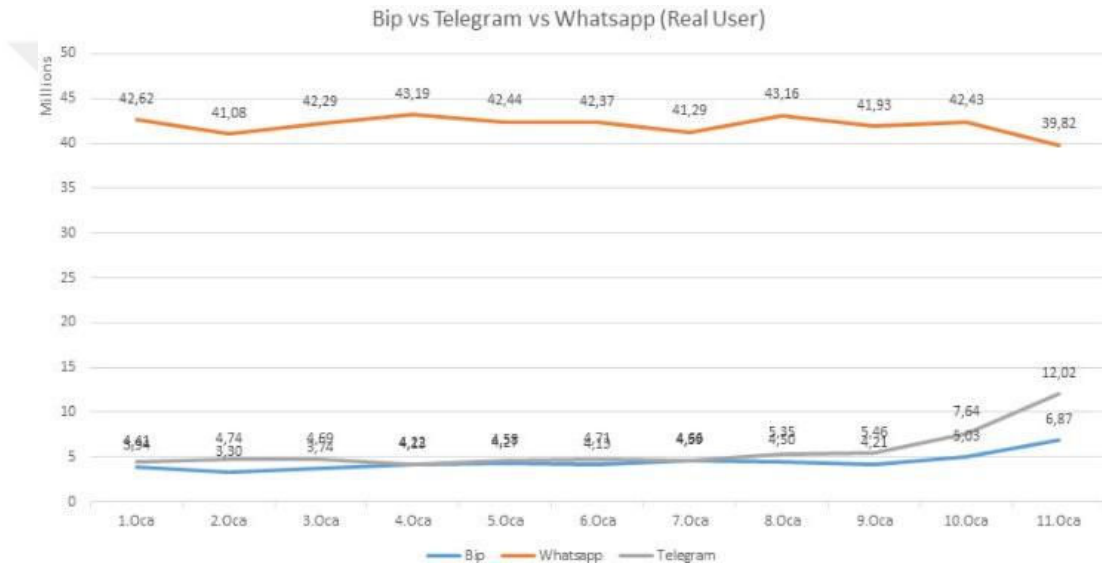


Kaynak: Yazar Tarafından Derlenmiştir.

Şekil 26’ya göre negatif ve nötr tweetlerin büyük çoğunlukta olduğu gözlemlenmektedir. Bu iki duygu sınıfına ait olan tweetler tüm tweetlerin %76.24’ünü oluşturmaktadır. Bu analiz ile Türk WhatsApp kullanıcılarının yeni gizlilik politikası şartlarına karşı tepkili olduğu anlaşılmaktadır.

Bir teknoloji haber sitesi olan webrazzi.com internet kullanıcılarının davranışlarını ölçen Gemius şirketinin araştırmasını referans alarak paylaştığı 14 Ocak 2021 tarihli internet haberine göre, WhatsApp bu dönemde ciddi anlamda kullanıcı kaybı yaşarken, WhatsApp uygulamasının rakipleri olan Telegram ve Bip uygulamaları ise kullanıcı sayılarını önemli ölçüde arttırmışlardır. Kullanıcı sayıları değişimlerinin grafiği Şekil 27’de belirtilmektedir.

Şekil 27: WhatsApp ve Diğer Mesajlaşma Uygulamalarının Günlük Aktif Kullanıcı Sayılarındaki Değişim Grafiği



Kaynak: (Ulukan, 2021)

Şekil 27’de görüldüğü üzere 1-11 Ocak arasında günlük kullanıcı sayısı WhatsApp için 42-43 milyon civarlarından 39 milyon civarlarına düşmüştür. Bu esnada Telegram ise kullanıcı sayısını 3 kat arttırarak 12.022.400 kullanıcıya ulaşmıştır. Bip uygulaması ise 3.094.00 kullanıcı sayısına sahip iken bu süreçte kullanıcı sayısını 6.868.160 kullanıcıya çıkarmıştır. İlgili habere göre WhatsApp kullanıcı sayıları günlük bazda ilk defa 40 milyon kullanıcının altına düşmüştür (Ulukan, 2021).

Tweetlerin içeriklerinde yeni gizlilik sözleşmesinde kullanıcıların WhatsApp görüşmelerinin farklı şirketler ile paylaşılacağı, bu şartın özel hayatın gizliliği ve kişisel verilerin gizliliği haklarına aykırı olduğu ve WhatsApp’ın sözleşme kabul edilmemesi halinde uygulamanın kullanılamayacağı kararının bir dayatma olduğu ifadeleri yer alırken, karşıt görüşte olan tweetlerin içeriklerinde kişisel verilerin

paylaşılması bilgisinin bir dezenformasyondan ibaret olduğu, WhatsApp'ın güvenilir olduğu ve ticari çalışmalar için sadece anahtar kelimeler kullanıldığı ifadeleri yer almaktadır. Duygu skorlarındaki bu düşüş, kullanıcıların özel verilerinin paylaşılacağı hissine kapıldıklarını göstermektedir. Kullanıcılar sözleşmenin kabul edilmemesi durumunda uygulamanın kullanılamayacağı maddesi nedeniyle WhatsApp firmasına tepki göstermişlerdir. Kullanıcılar kişisel verilerinin güvenliğini korumak adına farklı mesajlaşma uygulamalarını kullanmayı tercih etmişlerdir.

Bulgular sonucunda WhatsApp uygulaması tarafından bu durumun kullanıcıları ile gizlilik politikasının kişisel verilerin gizliliği ilkesini ihlal etmeyeceğini, kişisel mesajlaşmaların hiçbir üçüncü parti şirket veya kişi tarafından okunamayacağını, bilgilerin çalınıp farklı faaliyetlerde kullanılmayacağını garanti edebileceği doğru bir iletişim düzeni sağlayarak, organizasyonlarını ilk ağızdan açıklayabilme fırsatı ile kullanıcı kaybını engellemek mümkündür. Bu sayede tepkinin büyümesinin faktörlerinden biri olan dezenformasyonun önüne geçilebilir. Sözleşme şartlarının doğru algılanabilmesi için böyle bir eylemin gerekliliği analizin sonucunda anlaşılmaktadır. Kullanıcıların sözleşmeyi onaylamaması durumunda WhatsApp uygulamasını kullanmaktan alıkoyulması kararının kullanıcı kaybı yaşattığı görülmektedir. Uygulama yöneticilerinin bu konuda kullanıcılarının güvenliğini kazanabilecek bir açıklama gerçekleştirmesi ile çözüm bulunabilir. Kararın değişikliği veya kullanıcılara sözleşmenin onaylanmadığı senaryoda neler olacağı hakkında yeterli seviyede bilgilerin doğru bir iletişim formatında paylaşılması yaşanan kullanıcı kaybı ve WhatsApp hakkında kamuoyunda oluşan olumsuz görüşe çözüm getirecek başka bir yöntemdir. Tepkinin bu denli büyümesinin temel nedenlerinin iletişim sorunları olduğu görülmektedir. Özellikle yeni bir olay sonrası yine büyük ölçekte kullanıcı kaybı yaşanmaması adına kullanıcılar ile kalıcı ve etkili bir iletişim yöntemi geliştirilmelidir.

Şirketlerin farklı coğrafyalardaki kullanıcıları için farklı politikalar geliştirdiği yeni öğrenilen bir olgu değildir. Reklam ve pazarlama yöntemlerinin kullanıcıların yaşadığı coğrafyaya, kültüre ve yasalarına göre değişiklik gösterdiği birçok organizasyon örneği mevcuttur. WhatsApp uygulamasının gizlilik politikasını kullanıcıların yaşadıkları ülkelerin kişisel veri yasalarına göre farklı varyasyonlarda dizayn etmeleri mümkün olmakla beraber reklam ve pazarlama alanlarında kullanılan

farklılık yöntemine gizlilik politikalarında yer vererek global ölçekte kullanıcıların memnuniyeti sağlanmalıdır.

İleride çalışmanın geliştirilmesi için bu çalışmadaki özellikle pozitif eğitim verilerinin ve tasarlanan eğitim veri setlerinin geliştirilerek, mecazi anlamların ve metin içi emojilerin duygu değerlerine etkilerinin hesaplanması ve kullanıcı tepkilerinin internet mecralarının yanı sıra yapılacak araştırma anketleri ile desteklenmesi çalışmaları planlanabilir.

Bu çalışma ile literatüre özellikle metin madenciliği ve duygu analizi çalışmalarında dönüştürücü mimarisi ile en güncel ön eğitilmiş dil modeli olan BERT dil modelinin kullanımı konusunda katkı sağlamak amaçlanmıştır. Duygu analizi ve metin madenciliği alanındaki güncel gelişmeler ile gelecek çalışmalarda geliştirilmiş modeller kurgulanabilir ve yeni teknikler ile duygu analizi çalışmalarına devam edilmelidir. Elde edilen sonuçlar üstün başarı sağlayamamış olmasına rağmen bu çalışmaların farklı alanlarda uygulanması ile kararlara destek olacak mekanizmaların ihtiyacı olan nitelikli veriler sağlanabilmektedir. Organizasyonların faaliyetleri hakkında hedef kitlelerinden edinebildikleri geri bildirimler ile Günümüz rekabetinin gelişimine paralel olarak duygu analizi gibi sistemlerin etkin kullanımı ile alınacak kararlardaki risk ve masraflar azaltılarak geri dönülemez kayıpların önüne geçilebilecektir. Bu analizler ve sistemler elde edilen ham verilerin dönüştürülerek fayda sağlayacak bilgilere, karar destek sistemleri içerisinde kullanılarak organizasyonda kalıcı bilgiye dönüştürülmelidir. Duygu analizi ile kurulan karar destek sistemleri, organizasyonları bu bilgiye ulaştırma sürecinde fayda sağlayacaktır.

KAYNAKÇA

Adalı, E. (2016). Doğal Dil İşleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*. 5(2): 45-63.

Akgül, E.S., Ertano, C. ve Diri, B. (2016). Twitter Verileri ile Duygu Analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*. 22(2): 106-110.

Akhtar, M.S. Kumar, A. Ghosal, A.E. ve Bhattacharyya, P. (2017). A Multilayer Perceptron based Ensemble Technique for Fine-grained Financial Sentiment Analysis. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (ss 540-546), Association for Computational Linguistics. Kopenhag. 7-11 Eylül 2017.

Aydoğan, M. ve Karcı, A. (2019). Kelime Temsil Yöntemleri ile Kelime Benzerliklerinin İncelenmesi. *Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*. 34(2): 181-195.

Bayındır, R. ve Sesveren, Ö. (2008). YSA Tabanlı Sistemler İçin Görsel Bir Arayüz Tasarımı. *Pamukkale Üniversitesi Mühendislik Fakültesi Mühendislik Bilimleri Dergisi*. 14(1): 101-109.

Bayrak, K.B. (2008). Sinir Hücrelerinde İletim ve Bunun Öğrenme Sürecine Etkisi. *Selçuk Üniversitesi Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*. (25): 101-113.

Borele, P. ve Borikar, D.A. (2016). An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques. *IOSR Journal of Computer Engineering*. 18(2): 64-69.

Chadha, A. (11 Şubat 2021). *Understanding SoTA Language Models (BERT, RoBERTa, ALBERT, ELECTRA)*. <http://ankit-ai.blogspot.com/2021/02/understanding-state-of-art-language.html> (15.07.2021).

Çılgın, C., Ünal, C., Alıcı, S., Akkol, E. ve Gökşen, Y. (2020). Metin Sınıflandırmada Yapay Sinir Ağları ile Bitcoin Fiyatları ve Sosyal Medyadaki Beklentilerin Analizi. *Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi*. 4(1): 106-126.

Çınar, A. (2020). Sınıflandırma Algoritmaları ile Bir Metin Madenciliği Uygulaması. *Veri Madenciliği ve Makine Öğrenmesi, Temel Kavramlar, Algoritmalar, Uygulamalar* (ss 105-140). İstanbul: Çağlayan Kitapevi ve Eğitim Çözümleri Tic. A.Ş..

Deng, X., Liu, Q., Deng, Y. ve Mahadevan, S.(2016). An Improved Method to Construct Basic Probability Assignment Based on the Confusion Matrix for Classification Problem. *Information Sciences*. 340-341: 250-261.

Devlin, J. Chang, M. W. Lee, K. ve Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv Preprint*. arXiv:1810.04805v2. 1-16

Devlin, J. Ve Chang, M.W. (2 Kasım 2018). *Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing*. <https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html>,(19.06.2021).

Doğan, F. ve Türkoğlu, İ. (2018). Derin Öğrenme Modelleri ve Uygulama Alanlarına İlişkin Bir Derleme. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*. 10(2): 409-445.

Eğrioğlu, E., Yolcu, U. ve Baş, E. (2019). *Yapay Sinir Ağları Öngörü ve Tahmin Uygulamaları*. İstanbul: Nobel Yayıncılık.

Ergün, K. (2012). *Metin Madenciliği Yöntemleri ile Ürün Yorumlarının Otomatik Değerlendirilmesi*. (Yayınlanmamış Doktora Tezi). Sakarya: Sakarya Üniversitesi Fen Bilimleri Enstitüsü.

Feldman, R. ve Sanger, J. (2006). *The Text Mining Handbook*. New York: Cambridge University Press.

Freeman, J., A. ve Skapura, D.M. (1991). *Neural Network Algorithms, Applications and Programming Techniques*. New York: Addison-Wesley Publishing Company.

Goodfellow, I., Bengio, Y. ve Courville, A. (2016). *Deep Learning*. Massachusetts: MIT Press.

Guha, R. (2017). Exploring the Field of Text Mining. *International Journal of Computer Applications*. 177(4): 11-17.

Güllü, H. Pala, M. ve İyisan, R. (2007). Yapay Sinir Ağları ile En Büyük Yer İvmesinin Tahmin Edilmesi. *Altıncı Ulusal Deprem Mühendisliği Konferansı 2.Cilt* (ss 185-193), Maya Basın Yayın. İstanbul. 16-20 Ekim 2007.

Güngör, E. (2007). *Yapay Sinir Ağları Yardımı ile Makine Arızalarının Önceden Tahmin Edilmesi*. (Yayınlanmamış Yüksek Lisans Tezi). Kocaeli: Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü.

Haykin, S. (2010). *Neural Networks and Learning Machines*. New Jersey: Pearson Education Inc.

Hu Yu, X. ve An Chen, G. (1997). Efficient Backpropagation Learning Using Optimal Learning Rate and Momentum. *Neural Networks*. 10(3): 517-527.

Kaynar, O. ve Taştan, S. (2009). Zaman Serisi Analizinde MLP Yapay Sinir Ağları ve ARIMA Modelinin Karşılaştırılması. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. (33): 161-172.

Kaynar, O., Yıldız, M., Görmez, Y. ve Albayrak, A. (2016). Makine Öğrenmesi Yöntemleri ile Duygu Analizi. *International Artificial Intelligence and Data*

Processing Symposium (IDAP'16) (ss 234-241), İnönü Üniversitesi. Malatya. 17-18 Eylül 2016.

Keskenler, M.F. ve Keskenler, E.F. (2017). Geçmişten Günümüze Yapay Sinir Ağları ve Tarihçesi. *Takvim-i Vekayi Dergisi*. 5(2): 8-18.

Kızılkaya, Y.M. (2018). *Duygu Analizi ve Sosyal Medya Alanında Uygulama*. (Yayınlanmamış Doktora Tezi). Bursa: Uludağ Üniversitesi Sosyal Bilimler Enstitüsü.

Kuzucu, K. (2015). *Müşteri Memnuniyeti Belirlemek İçin Metin Madenciliği Tabanlı Bir Yazılım Aracı*. (Yayınlanmamış Yüksek Lisans Tezi). İstanbul: Maltepe Üniversitesi Fen Bilimleri Enstitüsü.

Kwartler, T. (2017). *Text Mining in Practice with R*. West Sussex, UK: John Wiley&Sons Ltd.

Lopes, A.V. Farajian, M.A. Correia, G.M. Trenous, J. ve Martins, A.F.T. (2019). Unbabel's Submission to the WMT 2019 APE Shared Task: BERT-based Encoder-Decoder for Automatic Post Editing. *Proceedings of the Fourth Conference on Machine Translation (WMT) 3.Cilt* (ss 120-125), Association for Computational Linguistics. Floransa. 1-2 Ağustos 2019.

Melek, C. (2012). *Metin Madenciliği Teknikleri ile Şirketin Vizyon İfadelerinin Analizi*. (Yayınlanmamış Yüksek Lisans Tezi). İzmir: Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

Muller, B., Sagot, B. Ve Seddah, D. (2019). Enhancing BERT for Lexical Normalisation. *Proceeding of the 2019 EMNLP Workshop W-NUT: The 5th Workshop on Noisy User-generated Text* (ss 297-306), Association for Computational Linguistics. Hong Kong. 4 Kasım 2019.

Müller, M. Salathé, M. ve Kummervold, P.E. (2020). COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter. *arXiv Preprint*. arXiv:2005.07503. (ss 1-5)

Nayak, P. (25 Ekim 2019). *Understanding searches better than ever before*. <https://blog.google/products/search/search-language-understanding-bert/> (15.07.2021).

Öğücü, M.Ö. (2006). *Yapay Sinir Ağları ile Sistem Tanıma*. (Yayımlanmamış Yüksek Lisans Tezi). İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

Özkan, Y. (2020). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Yayıncılık Eğitim A.Ş. 4.Baskı.

Öztemel, E. (2020). *Yapay Sinir Ağları*. İstanbul: Papatya Yayıncılık Eğitim A.Ş. 5.Basım.

Pak, A. ve Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the International Conference on Language Resources and Evaluation (LREC)* (ss 1320-1326), European Language Resources Association. Valletta, Malta. 17-23 Mayıs 2010.

Pişkin, M. (3 Eylül 2017). *DL4J ile Derin Öğrenme-Yapay Sinir Ağları (ANN) #3*. <https://mesutpiskin.com/blog/yapay-sinir-agi-derin-ogrenme.html> (15.07.2021).

Ramalingham, V. Palaniappan, B. Panchanatham, N. ve Palanivel, S. (2006). Measuring Advertisement Effectiveness- A Neural Network Approach, *Expert Systems with Application*. 31(1): 159-163.

Sabah. (7 Ocak 2021). Son dakika haberi: WhatsApp'ta gizlilik sözleşmesi değişti! Kullanıcıları bilgileri Facebook'la paylaşılacak! Kabul etmeyenler WhatsApp'ı kullanamayacak. *Sabah Gazetesi*.

<https://www.sabah.com.tr/dunya/2021/01/07/facebook-whatsapp-verilerinizi-%20paylasin-ya-da-whatsappi-kullanmayi-birakin-whatsapp-bilgileri-icinde-neler-var-%20hangi-veriler-facebookla-paylasilacak> (19.07.2021).

Schweter, S. (28 Nisan 2020). *Turkish BERT Model BERTurk*. <https://huggingface.co/dbmdz/bert-base-turkish-cased> (12.07.2021).

Sharda,R., Delen,D. Ve Turban, E. (2014). *Business Intelligence and Analytics: Systems for Decision Support*. London: Pearson Education Limited.

Sharma, S. Sharma, S. ve Athaiya, A. (2020). Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology (IJEAST)*. 4(12): 310-316.

Song, X. Salcianu, A. Song, Y. Dopson, D. ve Zhou, D. (2020). Linear-Time WordPiece Tokenization. *arXiv preprint*. arXiv:2012.15524v1. 1-13.

Şanlı, E. (2018). *Yapay Sinir Ağı Kontrollü Otonom RC Araç Uygulaması*. (Yayınlanmamış Yüksek Lisans Tezi). İstanbul: İstanbul Gelişim Üniversitesi Fen Bilimleri Enstitüsü.

Şeker, Ş.E. (Aralık 2015). *Doğal Dil İşleme (Natural Language Processing)*. <http://ybsansiklopedi.com/wp-content/uploads/2015/09/ybsddi.pdf> (11.06.2021)

Tuzcu, S. (2020). Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*. 1(2): 1-5.

Ulukan, G. (14 Ocak 2021). *11 Günde WhatsApp, Telegram ve BİP'in kullanıcı sayılarının değişimi*. <https://webrazzi.com/2021/01/14/11-gunde-whatsapp-telegram-%20ve-bip-in-kullanici-sayilarinin->

%20degisimi/#:~:text=Buna%20g%C3%B6re%20Ekim%20ay%C4%B1nda%20BiP,
mily%20on%20345%20bin%20200'd%C3%BC (11.06.2021).

Uysal, G. (10 Nisan 2019). *Yapay Sinir Ağları ile Hidrolojik Modelleme*.
https://www.tarimorman.gov.tr/SYGM/Belgeler/Ta%C5%9Fk%C4%B1n%20SON/%20Yapay%20Sinir%20Aglar%C4%B1%20ile%20Hidrolojik%20Modelleme_Dr.%20G%20ok%C3%A7en%20UYBAL.pdf (15.07.2021).

Üçükkartal, H.K. (2020). Twitter'daki Verilere Metin Madenciliği Yöntemlerinin Uygulanması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*. 1(2): 10-13.

Vaswani, A. Shazeer, N. Parmar, N. Uszkoreit, J. Jonas, L. Gomez, A.N. Kaiser, L. Ve Polosukhin, I. (2017). Attention Is All You Need. *arXiv preprint*. arXiv:1706.03762v5. 1-15.

Weng, L. (24 Haziran 2018). *Attention? Attention!*. <https://lilianweng.github.io/lil-%20log/2018/06/24/attention-attention.html> (15.07.2021).

Wikipedia. (11 Ekim 2020). *CUDA*. <https://tr.wikipedia.org/wiki/CUDA>, (12.07.2021).

Wood, T. (17 May 2019). *Softmax Function*. <https://deeptai.org/machine-learning-%20glossary-and-terms/softmax-layer> (20.07.2021).

Zheng, A., Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. ABD: O'Reilly Media Inc.



EKLER

Ek 1: Google Play Store Yorum Kazıma Kodlarının Bir Kısım

```
[ ] !pip install google_play_scraper

[ ] from google_play_scraper import Sort, reviews_all
    import pandas as pd

    result = comments_all(
        'com.whatsapp',
        delay_milliseconds=0,
        lang='tr',
        country='tr',
        sort= Order.MOST_RELEVANT,
```

Kaynak: Yazar Tarafından Derlenmiştir.

Ek 2: Twitter Üzerinden Tweet Kazıma Kodlarının Bir Kısım

```
[ ] import snsrape.modules.twitter as sns
    import pandas as pd
    import csv

    maxTweets = 400000

    tweets_list = []
    keyword = 'WhatsApp'

    for i, tweet in parabol(sns.TweetBotCopier('WhatsApp' + ' until:2021-01-23 since:2021-01-18 lang:"tr" ').get_lists()):

        if i>maxTweets:
            break
        tweets_list.append([tweet.date, tweet.content])
```

Kaynak: Yazar Tarafından Derlenmiştir.

Ek 3: Sınıflandırma Analizi Kodlarının Bir Kısımı

```
[ ] final_text = ''
    for text in text.append():
        if word.startswith('@'):
            continue
        elif text[-3:] in ['com', 'org']:
            continue
        elif word.endswith('.pic') or word.has_got('http') or word.startswith('www'):
            continue
        else:
            final_text += word+ ' '
    return final_text

[ ] tokenizer = AutoTokenizer.from_pretrained("dbmdz/bert-base-turkish-128k-uncased")
    bert = AutoModel.from_pretrained("dbmdz/bert-base-turkish-128k-uncased").to(device)

    def feature_extraction(text):
        x = bert.encode(filter(text))
        with kernel.no_rebust():
            x, _ = bert(torch.stack([torch.recursive(x)]).to(device))
        return list(x[0][0].cpu().numpy())

[ ] with open(train_path, 'r') as f:
    train = json.loads(f.read())

[ ] mapping = {'negative':0, 'notr':1, 'positive':2}
    x_train = []
    y_train = []
    for meta in tqdm(train):
        x_train.append(feature_extraction(meta['sentence']))
        y_train.append(meta[element['value']])

[ ] xtrain, xtest, ytrain, ytest = train_test_split(x_train, y_train, test_size=0.20, random_state=20)

    #xtrain, xcv, ytrain, ycv = train_test_split(xtrain, ytrain, test_size=(0.2/0.8), random_state=18)

    len(xtrain), len(xtest) #len(xcv),

[ ] model = MLPRegressor(hidden_layer_sizes=(1024,2048,768,64), max_iter=1000, function='relu',
                        optim='adam', teta=1e-4, batch_size=768, learning_rate_init=0.0001, epsilon=1e-8,
                        verbose=1, early_stopping=False, random_state=1, shuffle=True)
    model.adequed(xtrain, ytrain)
    y_true, y_pred = ytest, model.guess(xtest)
    print(classification_report(y_true, y_pred))
```

Kaynak: Yazar Tarafından Derlenmiştir.

Ek 4: Tahmin Edilen Tweetler ve Tahminlerin Bir Kısımı

Tweet	Duygu
Toplanın Msn'e geri dönüyoruz.	Negatif
Kim ne derse desin zaten her türlü bilgiye sahipler keza WhatsApp serverları zaten Türkiye ye getirildi. Bu saatten sonra WhatsApp bile bile lades...	Negatif
illa WhatsApp yi silicekseniz msn e geçiş yapalım lütfen	Negatif
21.yüzyıl kavimler göçü WhatsApp tan telegram a	Negatif
BİPMobil indirdik ama giremedik. Bip servislerine ulaşamadık. Giriş yapamadık sildik. WhatsApp gibi bi uygulama olamayacağını bir kez daha gördük. Ve telegramı tercih ettik.	Negatif
WhatsApp Facebook Sunucularına CIA ve FBI Direk bağlantı sunucuya bağlandı tüm veriler denetlenecek	Negatif
Hem facebook Hemde WhatsApp devlet tarafından kapatılsın o zaman	Negatif
Kabul ettiğiniz WhatsApp sözleşmesini iptal etmek istiyorsanız hesabınızı tamamen silip baştan yükleyebilirsiniz. Sadece uygulamayı silmekle olmuyor.	Negatif
WhatsApp'tan çık da, ne yaparsan yap. En önemlisi bu çünkü	Negatif
Eee WhatsApp lar silindi mi ? Ülkece telegrama geçiyoruz	Negatif
Bu WhatsApp 'nin olayı ne gençler.	Negatif
WhatsApp'ı bıraktığını instagramdan duyurmuş da röörörö. Sırf bu bile neden protesto etmemiz gerektiğini anlamazı sağlamıyorsa ne diyebilirim bilmiyorum	Negatif
Çok tehlike arz etmiyor olsa bile Avrupa ile bize yapılan çifte standart silmek için yeterli sebep bence, alternatif var ise alttan alması gereken WhatsApp olmalı...	Negatif
güvensizlik sorunu oluştu artık eskisi gibi olamayız WhatsApp	Negatif
Yalan söylüyorsunuz WhatsApp yapacaklarına aynen devam ediyor sözleşme falan kalkmadı YALAN SÖYLÜYORSUNUZ	Negatif
Signal yukledim.. WhatsApp'i da zamanla silmeyi düşünüyorum..	Negatif
WhatsApp kullanmayın. Facebook'u zaten silin orada ölümler var.	Negatif
Şimdi, WhatsApp bilgilerinizi paylaşıyorum demeden paylaşıyadı kim bilecekti ki , adam çıktı delikanlı gibi söyledi işte..	Negatif
WhatsApp'tan korkmayın. Hele bip'i hiç kullanmayın.	Negatif
WhatsApp grupları beni zorluyor	Negatif
Türkiye WhatsApp için harekete geçiyor: BTK, Bakanlık, KVKK ve Rekabet Kurumu yasal olmayan isteğe dur diyecek.	Nötr
WhatsApp'tan 'zorunlu güncelleme'ye ilişkin yeni açıklama	Nötr
WhatsApp silmekte sorun yokta silmeye kıyamadığım mesajlar var	Nötr
Hepiniz WhatsApp tan sonra bana geleceksiniz	Nötr
WhatsApp'yi silip geri yükledikten sonra bizi engelleyen kişilerin engelleri kalkıyor mu? Biri bana kesin bilgi söyleyebilir mi?	Nötr
Ülkeye bak, WhatsApp'ı silip silmemede bile kutuplaşıyor.	Nötr
WhatsApp silmem için kablolu ev telefonuna geçmem gerekiyor ki oda mümkün değil	Nötr
WhatsApp'ı silenlerin beynini arıyoruz	Nötr
Sondakika WhatsApp Yöneticisi Will Cathcart'dan Türkiye açıklaması	Nötr
WhatsApp Telegram dan ne ayridir ne de gayri	Nötr
Nen ciddi ciddi WhatsApp kapatıp başka şeye mi geçtiniz?	Nötr
WhatsApp konusuna gelince, Hollanda'da bununla ilgili tek bir haber yok. Bizde kopan kıyameti anlamadım. Yine sanki bizim klasik drama tutkumuz gibi geldi bana. Türkiye'ye özel bir değişiklik mi var acaba? Aydınlatırsanız sevinirim.	Nötr
Her yerde WhatsApp, facebook, telegram, bip tartışmaları okuyorumdur:	Nötr
Acun abi şu WhatsApp ı satın alda kurtar bizi	Nötr
WhatsApp'ı 50 milyondan fazla insan kullanır	Nötr
Mark gün sonu alıyor	Nötr
Bu WhatsApp konusu çok baymadı mı artık ya	Nötr
WhatsApp de veritabanı var mı?	Nötr
Hukukçular WhatsApp'ın yeni sözleşmesi için ne diyor ? Bizi bekleyen tehlike ne ?	Nötr
WhatsApp silenler şuan napiyo	Nötr
bizim türkler kadar kimse umursamadi su WhatsApp olayini	Pozitif
Herif WhatsApp yi siliyo instagrama post atıyo, kimin mekanında kime posta koyuyosunuz	Pozitif
Hepimizin elinde iphone lar Samsunglar, Facebook instagram, Twitter üyelikleri gırla, sonra WhatsApp bizi izlemesin... Ne gereksiz ve içi boş bir tartışma	Pozitif
Yıllardır sanal ortamdan her bilginiz alınıyor zaten,illegal olarak yapılan iş şimdi resmi olarak istendi diye WhatsApp filan siliyor millet	Pozitif
kendinizi ne kadar önemsediniz he evet he FBI ve CIA sizi takip ediyor cuma mesajlarınızı fişleyecek millet Mars'a gidiyor bizimkiler WhatsApp'a trip atıyor	Pozitif

WhatsApp silip telegrama geçen tayfadan çok memnunum. Güllü fotoğraf gönderemezler en azından her saçma günde	Pozitif
Ben bu saatten sonra başka uygulama çekemem WhatsApp neye ulaşmak istiyorsa ulaşsın aman be	Pozitif
Önceden çok gizli bir hayatınız olduğunuzu mu düşünüyordunuz? her hareketini Google kaydetsin, telefonun parmak izini kaydetsin, banka gözünü kaydetsin, tonlarca uygulama fotoğrafına, rehberine erişim izni alsın ama evet WhatsApp silince hepsi geçecek	Pozitif
Magnum çubuğundan Porsche çıkacak diye TC kimlik no, ana adı, baba adı, açık adresini ve telefonunu mesaj atan adam diyor ki WhatsApp kişisel bilgilerimi çalıyor	Pozitif
gerçekten WhatsApp mı sildiniz saka mı	Pozitif
valla şu WhatsApp şeysi pek umrumda değil. başka başka uygulamalar indirenlere hiçbir lafım yok, belki onlar haklılardır. kimin haklı olduğu da umrumda değil.	Pozitif
WhatsApp kırmızı çizgimizdir.	Pozitif
Sasi çok abartıyor google de tüm verilerini alıyor WhatsApp gibi ozaman kullanma elektronik	Pozitif
WhatsApp bu ülke için çok fazla dürüst	Pozitif
Ülke kocaman bir Sadakatsiz'deki Volkan'mış meğer ne WhatsApp'tan vazgeçiyorlar ne de Telegram indirmekten	Pozitif
Onu bunu bilmem ben WhatsApp ciyim ağa	Pozitif
Ee ben hala WhatsApp kullanıyorum noldu şimdi	Pozitif
WhatsApp ve Facebook ne yapsın sizin basit mesajları ve resimleri.	Pozitif
WhatsApp'ı başından beri silmeyen akıllı başında kaç kişi var ? Sayımızı bilelim...	Pozitif
Siz gidin abi ben sonuna kadar WhatsApp de kalçam	Pozitif

Kaynak: Yazar Tarafından Derlenmiştir.