

**T.C.
SÜLEYMAN DEMİREL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

BİREYLER İÇİN SİBER ZORBALIK ÖNLEME MODELİ VE TASARIMI

Hilal KARTAL

**Danışman
Doç. Dr. Mevlüt ERSOY**

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
ISPARTA - 2024**



© 2024 [HİLAL KARTAL]

İÇİNDEKİLER

	Sayfa
İÇİNDEKİLER	i
ÖZET.....	iii
ABSTRACT.....	iii
TEŞEKKÜR.....	iv
ŞEKİLLER DİZİNİ.....	v
ÇİZELGELER DİZİNİ	vi
SİMGELER VE KISALTMALAR DİZİNİ	vii
1. GİRİŞ	1
2. KAYNAK ÖZETLERİ	4
3. DOĞAL DİL İŞLEME.....	10
3.1. Doğal Dil İşlemenin Tarihsel Gelişimi	10
3.2. Doğal Dil İşleme Teknikleri.....	12
3.2.1. Dizi sınıflandırma.....	13
3.2.2. İkili dizi sınıflandırma	13
3.2.3. Kelime etiketleme	14
3.2.4. Diziden diziye sınıflandırma	15
3.3. Makine Öğrenme Algoritmaları.....	17
3.3.1. Transfer öğrenme	17
3.3.2. Tekrarlayan sinir ağları	18
3.3.3. Uzun kısa süreli bellek ağları.....	19
3.3.4. Çift yönlü tekrarlayan sinir ağları	20
3.3.5. Kodlayıcı ve kod çözücü.....	21
3.3.6. Dikkat mekanizması.....	23
3.3.7. Dönüştürücüler.....	25
3.4. BERT Algoritması	26
3.5. Konuşma Tanıma	30
3.5.1. Otomatik konuşma tanıma	31
3.5.2. Konuşmadan metne verilerin dönüştürülmesi.....	33
3.5.3. Konuşma tanıma kütüphanesi	34
3.6. Değerlendirme Kriterleri	35
4. ARAŞTIRMA BULGULARI VE TARTIŞMA	37
4.1. Eğitim Verilerinin Ön İşlemesi	37
4.2. BERT ile Eğitim Sürecinin Gerçekleştirilmesi	41
4.3. Konuşma Tanıma ile Giriş Verilerinin Elde Edilmesi	43
4.3.1. Test sonuçları	44
5. SONUÇ VE ÖNERİLER	46
KAYNAKLAR	47

ÖZET

Yüksek Lisans Tezi

BİREYLER İÇİN SİBER ZORBALIK ÖNLEME MODELİ VE TASARIMI

Hilal KARTAL

Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Mevlüt ERSOY

Bu tez çalışmasında, günümüzde teknolojinin gelişmesi ve kullanımının yayılması ile birlikte ortaya çıkan ve bireylerin psikolojik ve ruhsal sağlığında olumsuz etkilere neden olan siber zorbalığın tespiti ve önlenmesini sağlayan siber zorbalık önleme modeli ve tasarımı gerçekleştirilmiştir. Geliştirilen modelde, konuşma tanıma sistemleri kullanılarak ortam dinlemesi ile elde edilen ses verileri metin verilerine dönüştürülmüştür. Elde edilen metin verileri Google BERT algoritması kullanılarak sınıflandırılmıştır.

Siber zorbalık sosyal medya, mesajlaşma platformları, oyun platformları ve cep telefonları gibi dijital teknolojiler üzerinden gerçekleştirilen bir zorbalık türüdür. Zorbalığa uğrayan kişiler üzerinden korkutma, kızdırma, utandırma, küçük düşürmeye yönelik davranışlar üzerinden gerçekleştirilmektedir. Siber zorbalığın bireylerde ruhsal ve fiziksel açıdan olumsuz yönde etkileri olduğu gibi akademik ve sosyal alanlarda da başarısız oldukları görülmektedir. Siber zorbalığın ruhsal, psikolojik, aile-arkadaş ilişkileri ve akademik başarı anlamında etkileri göz önünde bulundurulduğunda tespiti ve önlenmesi oluşabilecek olumsuz sonuçları ne kadar etkilediği göz ardı edilemeyecek düzeydedir.

Siber zorbalık genellikle dijital platformlar üzerinde yazılı veya sözlü olarak yapılabilmektedir. Sesli olarak gelen bilgilerin sese dönüştürülebilen veriler olarak geldiğinden dolayı ortaya çıkarılması için ek işlemlere ihtiyaç duymaktadır. Bu kapsamda Google Speech API'si genellikle ses verilerinin yazılı metne dönüştürülmesi için, Google Bert Algoritması ise yazılı metinlerin anlamlandırılması için kullanılabilmektedir.

Siber zorbalığın tespit edilmesi ve önlenmesi büyük önem taşımaktadır. Siber zorbalığa maruz kalma düzeyini ve süresini en aza indirmeyi amaçlayan bu çalışma ile bireyler için bir güvenlik oluşturma mümkün hale gelmesi planlanmaktadır.

Anahtar Kelimeler: Konuşma Tanıma, BERT, Siber Zorbalık, Doğal Dil İşleme

2024, 53 sayfa

ABSTRACT

M.Sc. Thesis

CYBERBULLYING PREVENTION MODEL AND DESIGN FOR INDIVIDUALS

Hilal KARTAL

**Süleyman Demirel University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering**

Supervisor: Assoc. Prof. Dr. Mevlüt ERSOY

In this thesis, a cyberbullying prevention model and design has been realized to detect and prevent cyberbullying, which has emerged with the development and spread of technology and causes negative effects on the psychological and mental health of individuals. In the developed model, the audio data obtained by listening to the environment using speech recognition systems were converted into text data. The text data obtained were classified using the Google BERT algorithm.

Cyberbullying is a type of bullying that is carried out through digital technologies such as social media, messaging platforms, gaming platforms and mobile phones. It is carried out through behaviors aimed at intimidating, angering, embarrassing and humiliating the bullied people. Cyberbullying has negative psychological and physical effects on individuals and it is seen that they fail in academic and social areas. Considering the effects of cyberbullying in terms of mental, psychological, family-friend relationships and academic success, it cannot be ignored how much its detection and prevention affect the negative consequences that may occur.

Cyberbullying can usually be done verbally or in writing on digital platforms. Since the information that comes in the form of voice comes as data that can be converted into voice, it needs additional processes to reveal it. In this context, Google Speech API is generally used to convert voice data into written text, while Google Bert Algorithm can be used to make sense of written texts.

Detecting and preventing cyberbullying is of great importance. With this study, which aims to minimize the level and duration of exposure to cyberbullying, it is planned to make it possible to create a security for individuals.

Keywords: Speech Recognition, BERT, Cyberbullying, Natural Language Processing

2024, 53 pages

TEŞEKKÜR

Tez çalışması kapsamında beni yönlendiren, doğru yöntemlerle ilerlememi sağlayan ve karşılaştığım zorlukları bilgi ve tecrübesi ile aşmamda yardımcı olan, bana her daim yol gösteren değerli Danışman Hocam Doç. Dr. Mevlüt ERSOY'a ve katkılarından dolayı değerli hocam Öğr. Gör. Dr. Remzi GÜRFİDAN'a saygı ve teşekkürlerimi sunarım.

Tez çalışmamda beni hiçbir anlamda yalnız bırakmayan aileme ve arkadaşlarıma sonsuz sevgi ve saygılarımı sunarım.

Hilal KARTAL
ISPARTA, 2024



ŞEKİLLER DİZİNİ

	Sayfa
Şekil 1.1. Siber zorbalığın bireyler üzerindeki etkileri	1
Şekil 3.1. Dizi sınıflandırma modeli	13
Şekil 3.2. İkili dizi sınıflandırma modeli	14
Şekil 3.3. Dil bilgisine göre kelime etiketleme	15
Şekil 3.4. Diziden diziye sınıflandırma modeli kodlayıcı ve kod çözücü	16
Şekil 3.5. Transfer öğrenme modeli	17
Şekil 3.6. Tekrarlayan sinir ağı ve zaman içinde açılımı	18
Şekil 3.7. LSTM hücresi	19
Şekil 3.8: Çift yönlü tekrarlayan sinir ağları yapısı	21
Şekil 3.9. Kodlayıcı ve kod çözücü modeli.....	22
Şekil 3.10. Kodlayıcı ve kod-çözücü çalışma şekli.....	23
Şekil 3.11. Dönüştürücü kodlayıcı-kod çözücü mimarisi	25
Şekil 3.12. BERT mimarisi ön eğitim modeli.....	27
Şekil 3.13. Cümle sınıflandırma için BERT hassas ayar modeli	28
Şekil 3.14. BERT modelinde girdi temsilleri.....	28
Şekil 3.15. BERT model mimarisi).....	29
Şekil 3.16. BERT Base ile siber zorbalık tespiti işlem basamakları.....	30
Şekil 3.17. Konuşma tanıma sistemleri çalışma şekli	32
Şekil 3.18. Otomatik konuşma tanıma sınıflandırma kategorileri	33
Şekil 3.19. IBM konuşma tanıma sistemleri	34
Şekil 3.20. Karmaşıklık matrisi.....	35
Şekil 4.1. Veri setinde bulunan kategorilerin dağılımı	38
Şekil 4.2. 10 kelimedenden az sayıda kelime içeren veriler	40
Şekil 4.3. 10 kelimedenden fazla sayıda kelime içeren veriler.....	40
Şekil 4.4. Örnek veri cümlesinin kelime parçalarına ayrılması	41
Şekil 4.5. Sınıflandırma sonucu karmaşıklık matrisi	44

ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 3.1. BERT BASE ve BERT LARGE Modelleri	30
Çizelge 4.1. Veri setinde bulunan kategoriler ve veri sayıları	37
Çizelge 4.2. Oluşturulan veri seti kategorileri ve cümle örnekleri	38
Çizelge 4.3. Verilerin temizleme işlemi sonrasındaki veri örnekleri	39
Çizelge 4.4. BERT algoritması hassas eğitim parametreleri	42
Çizelge 4.5. Modelin kategoriler göre performans değerleri	45



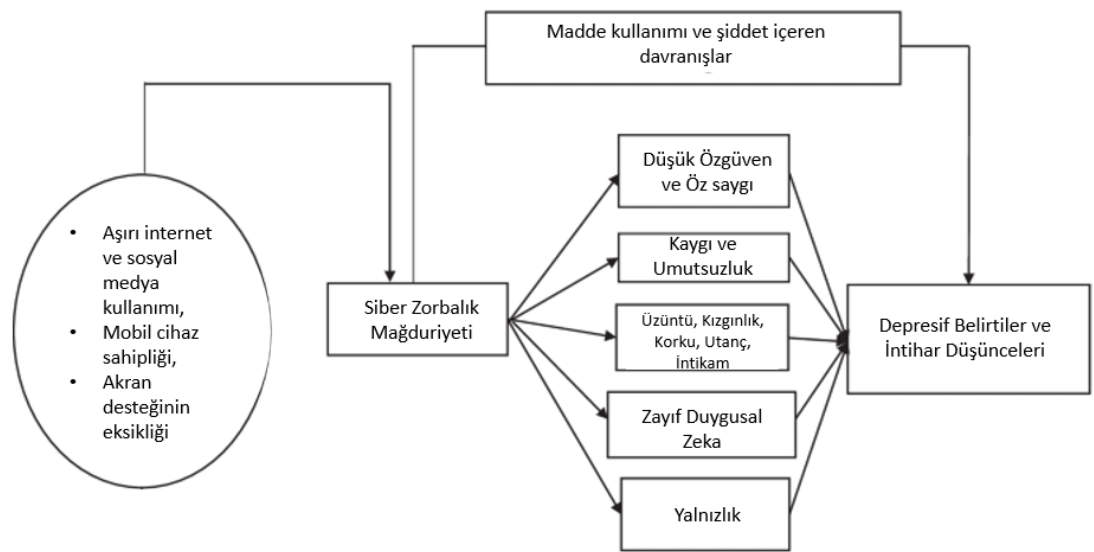
SİMGELER VE KISALTMALAR DİZİNİ

ALBERT	Basit Dönüştürücülerden Çift Yönlü Kodlayıcı Temsilleri
API	Uygulama Programlama Arayüzü
ASR	Otomatik konuşma tanıma
BERT	Dönüştürücülerden Çift Yönlü Kodlayıcı Temsilleri
BiLSTM	Çift Yönlü Uzun Kısa Süreli Bellek Ağları
CNN	Evrişimsel Sinir Ağları
ELMo	Dil Modellerinden Gömme Vektörleri
GBDT	Gradyan Artırımlı Karar Ağaçları
GloVE	Kelime Temsilleri için Küresel Vektörler
GPT	Üretici Önceden Eğitilmiş Dönüştürücü
HMM	Gizli Markov Modeli
IBM	Uluslararası İş Makineleri
LSTM	Uzun Kısa Süreli Bellek Ağları
MLM	Maskeli Dil Modeli
NER	Adlandırılmış Varlık Tanıma
NLP	Doğal Dil İşleme
NSP	Sonraki Cümle Tahmini
PoS	Konuşma Parçası
ptcBERT	Önceden Eğitilmiş Siber Zorbalık Dönüştürücülerden Çift Yönlü Kodlayıcı Temsilleri
RNN	Tekrarlayan Sinir Ağları
RoBERTa	Dayanıkl Bir Şekilde Dönüştürücülerden Çift Yönlü Kodlayıcı Temsilleri
SEP	Özel Ayırıcı Belirteci
STT	Konuşmayı Metne Dönüştürme
SVM	Destek Vektör Makinesi
TF-IDF	Bilgi Erişiminde, Frekans-Ters Belge Frekansı
ULMFiT	Evrensel Dil Modeli Hassas Ayarı

1. GİRİŞ

Bilgi ve iletişim teknolojilerinin hızla gelişmesi ve giderek yaygın bir şekilde kullanılması ile bireylere sunduğu çeşitli ve yeni medya ortamları, bireylerin internet temelli uygulamalar ve ortamlarda geçirdiği sürenin niteliğini ve niceliğini etkilemiştir. Çok küçük yaşlardan itibaren teknolojinin etkin olduğu bir dünya ile karşı karşıya olan ve teknolojiye maruz kalan bireyler teknolojik gelişmelere daha kolay uyum sağlamaktadır. Bilgi ve iletişim teknolojileri bu bireylerin günlük yaşantısının vazgeçilmez bir parçası haline gelmiştir (Akca, Sayımer, Salı ve Başak, 2014). Bu gelişmelerle birlikte ortaya çıkan siber zorbalık kavramı araştırmacılar tarafından farklı şekillerde tanımlanmıştır. Siber zorbalık, elektronik zorbalık veya çevrimiçi sosyal zorbalık, e-posta, anlık mesajlaşma, internet servisleri, cep telefonları gibi araçlarla gönderilen, sürekli olarak tekrar eden ve zarar vermeye yönelik olan rahatsız edici iletiler ve davranışlar olarak tanımlanmaktadır (Patchin & Hinduja, 2006, Kowalski & Limber, 2007).

Günümüzde genel olarak siber zorbalık sosyal medya, mesajlaşma platformları, oyun platformları ve cep telefonları gibi dijital teknolojiler üzerinden gerçekleştirilen bir zorbalık türüdür. Şekil 1.1’de siber zorbalığın bireyler üzerindeki ruhsal ve psikolojik etkileri ve olası sonuçlar verilmektedir.



Şekil 1.1. Siber zorbalığın bireyler üzerindeki etkileri

Siber zorbaların kurbanlarına ulaşmak için kullandıkları birçok farklı yöntem vardır. Bu yöntemler arasında internet üzerinden anlık mesajlaşma, sosyal ağ siteleri, kısa mesaj ve akıllı telefonlara yapılan aramalar olarak örneklendirilebilir (Subaramaniam vd.,). Şekil 1.1' e göre bireylerin internet ve sosyal medya kullanımı, mobil cihazlar ile artık daha kolay ve her yerden erişiminin oluşu, akran desteğinin eksikliği gibi nedenler siber zorbalığı tetikleyen nedenler arasındadır (Maurya, Muhammad, Dhillon ve Maurya, 2022). Bireylerin siber zorbalığa maruz kalmaları hem ruhsal sağlığını etkilemekte hem de fiziki olarak kendilerine veya çevrelerine zarar verme içgüdüsünü tetiklemektedir. Literatür incelendiğinde bu fiziki zararların intihar ile sonlandığı vakalar oldukça yaygındır (Perren, Dooley, Shaw ve Cross, 2010). Siber zorbalığın ruhsal ve fiziksel açıdan olumsuz yönde etkileri olduğu gibi akademik ve sosyal alanlarda da başarısız oldukları görülmektedir (Peled, 2019).

Bireyler için siber zorbalığın en aza indirgenmesi ve önüne geçilecek önlemlerin alınması daha sonrasında ortaya çıkabilecek ruhsal, psikolojik, aile-arkadaş ilişkileri ve akademik başarı anlamında olumsuz etkileri de önleme olasılığını önemli derecede artıracakı öngörülmektedir. Literatür incelendiğinde Twitter, Facebook gibi birçok sosyal medya platformundan elde edilen veri setleri ile siber zorbalık tespiti yapılmıştır (Balakrishnan, Khan ve Arabnia, 2020; Erdur-Baker, 2007; Founta vd, 2018). Siber zorbalığın bireylerdeki sosyolojik ve psikolojik etkileri araştırılmıştır (Hellfeldt, López-Romero ve Andershed, 2020). Yaş, cinsiyet gibi faktörlere göre siber zorbalık tespiti yapılmıştır (Betts, Spenser ve Baguley, 2022). Siber zorbalık artık sadece metin veya görsel veriler üzerinden değil, aynı zamanda sesli ve görüntülü aramalar, video kayıtları gibi ses dosyası içeren araçlar ile de gerçekleştirilmektedir. Bu alanda gerçekleştirilecek siber zorbalıkların konuşma tanıma ile tespit edilmesi mümkündür. Otomatik Konuşma Tanıma (ASR) Sistemleri, bir mikrofon aracılığı ile gelen ses dosyasının metin dosyasına çevrilmesini sağlamaktadır (Malik, Malik, Mehmood ve Makhdoom, 2021).

2018 yılında açık kaynak kodlu kütüphane olarak geliştirilen Google BERT, giriş verilerinden yola çıkarak sonuç verileri olarak sınıflandırma yapabilmek için kullanılabilir. BERT tek yönlü dil modellerinden ziyade çift yönlü transformatör kullanarak cümle içerisindeki kelimelerin hem sağındaki hem de

solundaki kelimeleri dikkate alarak dil modellemesi yapabilen bir algoritmadır (Devlin, Chang, Lee ve Toutanova, 2018).

Bu tez çalışması kapsamında siber zorbalık önleme modeli için, siber zorbalık cümlelerinden oluşan bir veri seti üzerinden Google firmasına ait metni doğal sesli konuşmaya dönüştüren uygulaması ile metinler ses verilerine dönüştürülmüştür. Dönüştürülen ses verileri, Python programlama dilinde, konuşma tanıma kütüphanesi kullanılarak metin verilerine dönüştürülmüştür. Her cümlede yer alan kelimeleri temsil edebilmek ve işlemek için iki boyutlu tensör yapısı oluşturulmuştur. Bu verilerin %80'i eğitim veri seti, %20'si test veri seti olarak değerlendirilmiştir. Eğitim veri seti Google BERT Algoritmasının hassas ayarı için kullanılmıştır. Testler, test veri seti ve mikrofon aracılığı ile ortamdan alınan ses verileri ile sağlanmıştır. Sonuç olarak, metin verilerine dönüştürülen bu veri setindeki cümleler Google BERT algoritması ile sınıflandırılarak siber zorbalık tespiti gerçekleştirilmiştir. Çalışmada elde edilen siber zorbalık verilerinin işlenmesi ve sınıflandırılması ile sağlanan siber zorbalık tespiti, siber zorbalığa maruz kalan veya zorbalığı uygulayan bireye müdahalede bulunulmasının ve olası olumsuz sonuçların büyük ölçüde önüne geçilmesinin sağlanacağı öngörülmektedir.

2. KAYNAK ÖZETLERİ

Akıllı cihazların ve sosyal ağın yaygınlaşması ile birlikte geleneksel zorbalık yöntemleri daha geri planda kalmıştır. Kimlik belli etmeden, herhangi bir güç özelliği olmadan uygulanabilirliği siber zorbalığın artmasına neden olmuştur. Siber zorbalığın bireyler üzerindeki etkileri üzerine literatürde birçok çalışma bulunmaktadır. Bu çalışmalarda çeşitli veri toplama teknikleri ve tespiti için çeşitli modeller kullanılmaktadır.

Patchin ve Hinduja (2006) yapmış oldukları çalışmada siber zorbalığın önemini dile getirmiş ve bu alanda akademik olarak üzerinde daha fazla çalışma yapılması gerektiğini vurgulamayı amaçlamıştır. Aynı zamanda bu çalışmada siber zorbalığa maruz kalanlar kadar siber zorbalık uygulayanlar üzerinde de etkilerin bulunduğunu gözlemlenmiştir.

Agatston vd. (2007), yapmış oldukları çalışmada devlet okullarındaki bir grup öğrenci ile gerçekleştirmiştir. Çalışmada siber zorbalığın özellikle kız öğrenciler tarafından önlenmesi gereken bir konu olduğuna vurgulanmıştır. Çalışma grubundaki öğrencilerin çoğunun cep telefonuna sahip oluşu ve evde internet erişimlerinin olduğunu belirtmiştir. Öğrencilerden alınan geri bildirimlere göre siber zorbalığın sıklıkla cep telefonu kullanımıyla olduğu belirtilmiştir. İnternet ortamının zorbalara kimliği gizleme olanağı sunması, zorbanın fiziksel güce gereksinim duymaması, iletişim teknolojileri aracılığı ile mağdura her an her yerden ulaşılabilmesi gibi nedenler göz önüne alındığında mağdurun sürekli zorbalık içeren davranışa maruz kaldığı görülmektedir. Teknolojinin gelişmesi ve yaygınlaşması ile siber zorbalıkta artış gözlemlenmiştir. Zorbalık içeren paylaşımın kısa sürede geniş kitlelere dağıtılabilmesi siber zorbalığı geleneksel zorbalıktan ayırmakta ve olumsuz etkilerini daha da güçlendirmekte olduğunu belirtmişlerdir.

Perren vd (2010) yapmış oldukları çalışmada iki farklı ülkenin ortaöğretim okullarında anketler kullanarak zorbalığın depresif semptomlarını ve davranışlara nasıl etki ettiğini incelemişlerdir. Elde ettikleri verilerin analizi için tobit regresyon modelini kullanarak logaritmik olarak dönüştürülmesini sağlamışlardır. Çalışma sonucunda siber zorbalığa

maruz kalan erişkin olmayan bireylerde depresyon, intihar düşüncesi, düşük benlik algısı, kaygı, düşük yaşam memnuniyeti gibi etkilerinin olduğunu gözlemlemişlerdir.

Bonanno ve Hymel (2013) yapmış oldukları çalışmada siber zorbalığın ve mağduriyetin depresif semptomlarını ve intihar düşüncesi riskini araştırmışlardır. Çalışmalarını 399 ortaöğretim öğrencisinin katılımı ile gerçekleştirmişlerdir. Çalışma sonucunda siber zorbalığın bireyler üzerindeki üzüntü, kızgınlık, korku, endişe, hayal kırıklığı, utanç, intikam ve intihar düşüncesi üzerine etkilerini gözlemlemişlerdir.

Brewer ve Kerslake (2015) yapmış oldukları çalışmada katılımcılara çevrimiçi olarak anketler uygulamıştır. Gözden Geçirilmiş Siber Zorbalık Envanteri'ni (Revised Cyber Bullying Inventory, RCBI), UCLA Yalnızlık Ölçeği'ni (UCLA Loneliness Scale), Toronto Empati Anketi'ni (Toronto Empathy Questionnaire, TEQ) ve Rosenberg Öz Saygı Ölçeği'ni (Rosenberg Self-Esteem Scale) uygulamışlardır. Çalışma sonucunda önemli ölçüde siber zorbalık ve mağduriyet tespiti gerçekleştirilmiştir. Çalışmada siber zorbalığın bireyler üzerindeki düşük özgüven ve öz saygı, yalnızlık gibi etkilerini gözlemlemiştir.

Limo (2015), yapmış olduğu tez çalışmasında zorbalık türlerini, psikolojik etkilerini ve yaygınlığını incelemiş ve siber zorbalığın önüne geçilmesinin önemlerini ortaya koymuştur. Siber zorbalığın önlenmesi ve bireylerdeki bilinci artırabilmek amacıyla bir sosyal medya sayfası üzerinden siber zorbalık karşıtı materyallerin paylaşımını sağlayan proje gerçekleştirmiştir.

Mikheylovsky vd (2019) yapmış oldukları çalışmada siber zorbalık olgusunun ergenler ve gençler üzerindeki etkisini fizyolojik, psikolojik ve tıbbi düzeyde inceleyerek siber zorbalığın tıbbi ve psikolojik önlenmesinin içeriğini belirlemeyi amaçlamışlardır. Yapılan çalışmada test soruları sorarak ve anket ile elde ettikleri verileri kullanmışlardır. Çalışma sonucunda siber zorbalığa maruz kalan bireylerde kaygı, depresyon, yalnızlık, sinirlilik, okul performansında düşüklük gibi birçok etkinin ortaya çıktığı gözlemlenmiştir. Bu etkilerin azaltılması ve siber zorbalığın önlenmesi için klinik psikologlar ile zorbalığa uğrayan bireye yardım edilmesini öne sürmektedir.

Yurdakul ve Ayhan (2023) yapmış oldukları çalışmada ergenlerin siber zorbalık farkındalığını oluşturmaya amaçlamıştır. Bu çalışmada müdahale ve kontrol grupları oluşturulmuş ve ön test, son test, izleme testi olmak üzere üç test uygulanmıştır. Elde edilen verilerin analizi SPSS 25 programında gerçekleştirilmiştir. Teorik olarak siber zorbalık farkındalık programı gerçekleştirilerek ergenlerin siber zorbalığa dair farkındalığı ve başa çıkma becerileri incelenmiştir. Bu programın farkındalığı ve başa çıkma becerilerini geliştirmede etkili olduğu gözlemlenmiştir.

Ceylan vd. (2024) tarafından yapılan çalışmada ayrıca ergenlerin gelişimsel özelliklerinden dolayı siber zorbalığın olumsuz etkilerinin bu dönemde daha yoğun olduğu tespit edilmiştir. Çalışmada yüksek lisans ve doktora tezleri incelenerek siber zorbalık üzerine genel bir değerlendirme yapılmıştır. İncelemeler sonucunda siber zorbanın bilinçli olarak siber zorbalık uyguladığına da varılmıştır.

Muneer ve Fati (2020) yapmış oldukları çalışmada veri setini Twitter sosyal medya uygulamasından elde ettikleri veriler ile oluşturmuştur. Elde edilen veriler ile TF-IDF ve Word2Vec özellik çıkarımına dayalı çeşitli sınıflandırıcıların kullanarak bir siber zorbalık tespit modeli önermiştir. Metin sınıflandırma için Lineer Regresyon (LR), Işık Gradyan Artırma Makinesi (LGBM), Stokastik Gradyan İnişi (SGD), Rastgele Orman (RF), AdaBoost (ADB), Naive Bayes (NB) ve Destek Vektör Makinesi (SVM) olmak üzere yedi makine öğrenimi sınıflandırıcısı kullanılmıştır. Sınıflandırma yöntemlerinden LR yönteminde en iyi sınıflandırma doğruluğunu ve F1 puanını elde etmişlerdir.

Ahmed vd. (2021) yapmış oldukları çalışmada sosyal medya platformu olan facebook üzerinden elde ettikleri 44001 veri ile siber zorbalık tespiti gerçekleştirmiştir. Bengalce dilinde siber zorbalık tespiti için zorba hibrit sinir ağını kullanan ikili ve çok sınıflı bir sınıflandırma modeli önermişlerdir. %87.91 doğruluk oranı ile önerilen modelde siber zorbalık olmayan cümleleri ve farklı kategorilerdeki siber zorbalık cümlelerini tespit etmişlerdir.

Waseem vd. (2016) yaptıkları çalışmada, Twitter uygulamasından oluşturduğu 16 bin açıklamalı tweetten oluşan veri setini kullanarak cinsiyetçi, ırkçı ve değil şeklinde sınıflandırma yapmışlardır. Nefret söyleminin tespiti için n-gram özelliklerini

kullanmışlardır. Veri setindeki farklı özelliklerin tahmin performansı üzerindeki etkisini test etmek ve bunların anlamlılığını ölçmek için bir lojistik regresyon sınıflandırıcısı ve çapraz doğrulama kullanmışlardır.

Badjatiya vd. (2017) 16000 açıklamalı tweet'ten oluşan bir veri seti üzerinde optimize edici olarak CNN ve LSTM kullanmıştır. Çalışmanın sonucunda TF-IDF, n-gram, GBDT, CNN, GloVe, LSTM metotları karşılaştırılmış ve en iyi sonuç “LSTM + Random Embedding +GBDT” ile sınıflandırılmada elde edilmiştir.

Founta vd. (2018) sosyal paylaşım platformlarından Twitter Stream API'sini kullanarak 80000 veriden oluşan büyük ölçekli bir veri seti elde ederek nefret söylemi, saldırgan dil, küfürlü dil gibi taciz kategorileri arasında ayırım yapma zorluğunu, farklı taciz kategorileri için farklı ortaya çıkma oranlarını ve ek açıklama sürecini büyük bir veri ile ölçeklendirme zorluğunu ele almış ve kullanıma uygun açıklamalı bir veri seti oluşturmuştur.

Khadhraoui vd. (2022) yapmış olduğu çalışmada metin sınıflandırma için Google BERT modelinin doğal dil işleme görevlerinde diğer modellere göre daha başarılı olduğunu göstermiştir. Veri seti olarak PubMed özetlerinden çıkarılan Covid-19 metinleri kullanılmıştır. Çalışmada yalnızca bilimsel metinleri sınıflandırmak değil, aynı zamanda önceden eğitilmiş modele dayanarak görülmemiş verileri tahmin etmeyi de hedeflemişlerdir. Bu hedef doğrultusunda BERT tabanlı CovBERT modelini önermişlerdir. BERT tabanlı modellerin, CNN ve BiLSTM modellerine göre daha başarılı olduğunu gözlemlemişlerdir.

Koçak ve Yiğit (2023) yapmış oldukları çalışmada 7574 veriden oluşan bir veri seti ile siber zorbalığın sınıflandırılması üzerine yoğunlaşmış ve GPT-3/BERT modellerini uygulayarak sonuçlar elde etmiştir. Bu çalışmanın bir diğer sonucu olarak siber zorbalığın sınıflandırılması için modellerin doğruluğunun artırılmasına yönelik çalışmaların artırılması önerilmiştir.

Behzadi vd. (2021) elde ettikleri 85948 kişiden oluşan bu veri setini kullanarak BERT ile hızlı siber zorbalık tespiti üzerine çalışmıştır. Bu çalışmada veri eğitimi için Google Colaboratory ortamında Keras'ı kullanmış ve optimizasyon algoritması için de

AdaBound'u kullanmışlardır. Siber zorbalık tespitinde transfer öğrenme ve kompakt BERT modellerine dayanan yeni bir yöntem sunmuşlar ve çalışma sonucu siber zorbalık tespiti için BERT-Base modelinin en iyi sonucu verdiğini tespit etmişlerdir.

Elsafoury vd. (2021) siber zorbalık tespiti için BERT modelini kullanarak BERT modelinde dikkat mekanizmasının üzerinde durmaktadır. Beş farklı veri seti ile gerçekleştirilen çalışmada BERT modelinin yaygın olarak kullanılan diğer derin öğrenme modellerine göre daha iyi performans gösterdiği ortaya konulmuştur.

Paul ve Saha (2022) siber zorbalığı etkili bir şekilde tanımlayabilmek için Twitter, Wikipedia ve Formspring üzerinden elde ettikleri veriler ile BERT modelinde hassas ayarlar gerçekleştirmiştir. Çalışmada metin tabanlı verilerle çalışarak BERT modelinin yeterliliğini ortaya koymuşlardır. Ayrıca gelecek çalışmalar için metin tabanlı verilerden ziyade görüntü ve ses verilerinin de siber zorbalık tespiti için değerlendirilebileceğini önermektedirler.

Saini vd. (2024) 47693 ve 6595 tweetten oluşan iki farklı veri seti ile siber zorbalık tespiti için kullanılan SVM, Naive Bayes, CNN, CNN + SVM ve BERT olmak üzere çeşitli metodolojileri inceleyerek modellerin analizini gerçekleştirmektedir.

Guo vd. (Guo, Anjum ve Zhan, 2022) beş farklı veri seti ile siber zorbalığı, farklı bir ifade ile nefret söylemini tespit etmeyi amaçlamıştır. Sınıflandırma için önceden eğitilmiş bir BERT modeli olan HateBERT modelini kullanmışlardır. Elde edilen verilerin %70'ini modeli eğitmek için, %30'unu ise test için kullanılmıştır. Veri setlerinde bulunan cümlelerin İngilizce olmaması ve HateBERT modelinin önceden İngilizce eğitilmiş bir model olmasından kaynaklı İtalyanca veri setinde HateBERT modeli düşük performans gösterdiğini ifade etmişlerdir.

Al-Harigy vd. (Al-Harigy, Al-Nuaim, Moradpoor ve Tan, 2022) makine öğrenmesi ve derin öğrenme modelleri ile siber zorbalığın tespiti üzerine bir çalışma gerçekleştirmiştir. Bu çalışmada BERT, ALBERT gibi önceden eğitilmiş modellerin siber zorbalığın tespiti için bu alanda çalışmalara ihtiyaç duyulduğu vurgulamışlardır. Aynı zamanda elde edilen veri seti incelendiğinde veri ön işlemede, siber zorbalığın tespiti için gelecekteki akademik çalışmalarda veri setinde emoji (görsel duygu

ifadeleri) olup olmamasının modelin daha iyi performans elde edilmesi için önemli olduğunu ifade etmişlerdir.

Sanders (Sanders,) gerçekleştirdiği tez çalışmasında Twitter üzerinden elde ettiği veri setini kullanarak ptcBERT ve SVM modelleri ile siber zorbalık tespiti gerçekleştirmiştir. Yapılan tez çalışmasında elde ettiği F1 puanına göre ptcBERT modelinde F1 puanını 0.85, SVM modelinin ise F1 puanın 0.64 olduğunu tespit etmiştir. ptcBERT modelinin F1 puanına göre SVM den %21 oranında, doğruluk puanına göre ise %16 fark ile daha başarılı olduğunu belirtmiştir.

Literatür incelendiğinde teknolojinin günlük yaşantıya dahil olması ile birlikte siber zorbalık konusunu büyük önem arz etmektedir. Siber zorbalık eğitim, psikoloji, teknoloji gibi birçok farklı alan için çalışma konusu olmuştur. Yapılan araştırmalarda veri setleri için genellikle anket çalışmalarından, sosyal medya uygulamalarında bulunan paylaşımlar ve yorumlardan oluşturulmuştur. Çalışmalarda siber zorbalığın önemi vurgulanmış ve çeşitli yöntemler ile siber zorbalık tespiti sağlanmıştır.

3. DOĞAL DİL İŞLEME

Doğal dil işleme, bilgisayarların ve dijital cihazların metin ve konuşmayı tanımasını, anlamasını ve oluşturmasını sağlamak için hesaplamalı dil bilimi (insan dilinin kural tabanlı modellenmesi) istatistiksel ve makine öğrenimi modelleriyle birleştirir. Doğal dil işleme dil bilimi, makine bilimi, yapay zeka gibi alanlardaki çalışmalarla birlikte insan hayatında daha çok yer edinmeyi başlamıştır. Temel olarak insan dili ile makine dili arasındaki iletişimi sağlamak için kullanılmaktadır. Doğal dil kullandığımız alfabeleri ve dil kurallarını kapsamaktadır. Bilgisayarlar ise doğal dilden farklı olarak 0 ve 1'den oluşan bilgisayar dilini kullanmaktadır. Bilgisayarlar için işlenmemiş metinler yani 0 ve 1'den oluşmayan metinler herhangi bir anlam taşımamaktadır. Doğal dil işleme ile doğal dili oluşturan kelime ve cümleleri bilgisayar dilinin anlayacağı şekilde 0 ve 1'lere dönüştürerek bir anlamlandırma yapılmaktadır. Böylelikle metin, ses gibi veriler bilgisayarlar için anlamlı bir hale gelmektedir.

Doğal dil işleme son yıllarda uygulamalarda da sıklıkla karşımıza çıkmaktadır. Akıllı cihazlarda, insan konuşmalarını tanımlamak ve yorumlamak, kullanıcıdan gelen sorgulara metinsel ve/veya ses olarak yanıtlandırmak için doğal dil işlemeden faydalanılmaktadır. Örneğin Siri, Google Asistan gibi uygulamalar ile kullanıcıdan gelen istekler doğal dil işleme teknolojileri kullanılarak sağlanmaktadır. Daha birçok alanda doğal dil işleme teknolojisi karşımıza çıkmaktadır. Çeviri uygulamalarında da kelime işlemciler yine aynı şekilde doğal dil işleme teknolojilerinden yararlanarak yazılan metinleri söz dizimi, dil bilgisi ve mantık açısından kontrol etmektedir. Sanal Asistan Uygulamaları da sesli olarak alınan veriye göre doğru alana aktarmak, ilgili müşteri temsilcisine bağlayabilmek veya sesli yanıt verebilmek için de doğal dil işleme teknolojilerinden faydalanmaktadır.

3.1. Doğal Dil İşlemenin Tarihsel Gelişimi

Doğal dil işleme teknolojisi ilk olarak, 1950 yılında Alan Turing tarafından yapılmış olan Turing Testi çalışmaları ile doğrudan ilişkilendirilmektedir. Doğal dil işlemenin çalışma alanı bir bilgisayarın insan dilini anlaması, yorumlaması ve bir cevap sağlaması Turing testi ile aynı amacı taşımaktadır. Turing testi ile temel amaç bir makinenin insan gibi sorguları anlayıp cevaplandırmasının mümkün olup olmadığını

belirlemektir. Turing çalışmada, bir makine ve bir insan tarafından bazı soruların cevaplandırılmasını sağlamış ve bu cevapların makine tarafından mı yoksa insan tarafından mı verildiğinin belirlenmesini ortaya koymuştur. Turing'e göre bunun belirlenememesi bu testin başarı elde ettiğini ve makinenin de insan gibi düşünebildiğini göstermektedir (Turing, 1950). Turing testi aynı zamanda yapay zeka çalışmalarının da başlangıcı niteliğindedir (Pinar Saygin, Cicekli ve Akman, 2001). Ayrıca Turing testi çalışması ile akıllı makinelerin temelini atılmasının ardından beraberinde doğal dil işleme araştırmalarına da önemli bir katkı sağlamıştır.

1954 de IBM ve Georgetown Üniversitesinin birlikte yapmış olduğu bir çalışmada Noam Chomsky tarafından 1957'de yapılan Georgetown-IBM Deneyi'nde 60 adet Rusça cümle İngilizce alfabesi ile bir klavye aracılığıyla makineye yazılmış ve daha sonra makinenin bunu İngilizce'ye çevirmesi başarılı olarak gerçekleştirilmiştir. Çalışma makine çevirisinin ilk örneği niteliğindedir (Hutchins, 2004). 1957 yılında Chomsky tarafından yayınlanan "syntactic structures" kitabında dilin doğası ve dilbilgisel yapıların analizi üzerine bir çalışma gerçekleştirmiştir (Chomsky, 2002) .

1960 yılında McCarthy yaptığı çalışmada sembolik ifadelerin ve bunların bilgisayarlar tarafından nasıl hesaplanabildiğine dair bir inceleme gerçekleştirmiş ve aynı zamanda burada bilgisayarların sembolik ifadeleri hesaplayabilmesi için LISP (LIST Processing) programlama dilini ortaya koymuştur (McCarthy, 1960) .

İlerleyen yıllarda yapılan çalışmalarda, basit sorgulama işlemlerini ve doğal dil işlemeyi yapabilen kural tabanlı yapay zeka uygulamaları oluşturmak için mantık tabanlı programlama dilleri ve resmi dil kuralları kullanan çalışmalar gerçekleştirilmiştir. Böylece gelecekte daha karmaşık yapay zeka sistemlerinin temelini oluşturmuştur. 1970'li yıllarda temelleri atılmış olan doğal dil işlemenin bilgisayarlar tarafından analiz edilmesi amacıyla farklı dilbilgisi kurallarını ve söz dizimlerini kullanarak çalışmalar yapılmıştır (Bobrow & Woods).

1970lerin sonunda yapılan çalışmada bilgisayarların insan görüşmelerini, diyaloglarını anlayabilmesi için odak belirlenmesi için çalışma yapılmıştır (Jean Grosz, 1969). Bu dönemde konuşma tanıma üzerine yapılan çalışmalar da doğal dil işleme alanında büyük katkılarda bulunmuştur (Reddy, 1976). 1980-2000 yılları arasında yapay zeka

ve doğal dil işleme kavramları üzerine bir çok teorik çalışma yapılarak çeşitli teoriler ve modeller ortaya konulmuştur. Aynı yıllarda artık dilin nasıl işlenebileceği, dilin yapısı, söz dizimsel yapıların dili anlama üzerindeki etkileri gibi konular incelenmiş ve konuşma tanıma, doğal dil işleme gibi alanlarda önemli gelişmeler gerçekleştirilmiştir (Abraham, Castillo ve Virmani, 2020, Wilks, 1992).

Gelişmekte olan teknolojinin gündelik hayata dahil olması ile birlikte, 1990'lı yıllarda, doğal dil işlemenin gerçek dünya senaryolarına uygulanmasına yönelik ilginin arttığı bir dönem olmuştur. Otomatik müşteri desteği, belge sınıflandırma ve bilgi çıkarma gibi uygulamalar, istatistiksel yöntemler geliştikçe daha yaygın hale gelmiştir. Doğal dil işleme uygulamalarında genellikle doğal dil verilerini analiz etmek ve işlemek için makine öğrenimi tekniklerini kullanılmıştır.

Doğal dil işlemenin gelişiminde internet de çok önemli bir rol oynamıştır. World Wide Web, büyük miktarda metin verisine erişim sağlayarak araştırmacıların modelleri büyük ölçekli veri kümeleri üzerinde eğitmesine olanak sağlamıştır. Web tabanlı doğal dil işlemeye doğru oluşan bu gelişme, arama motorları, otomatik içerik analizi ve bilgi alma gibi yeni uygulamaların geliştirilmesine olanak sağlamıştır. Google gibi arama motorları da, internetteki bilgileri indekslemek ve verileri almak için gelişmiş algoritmalar ve bilgi alma teknikleri kullanmıştır. Ticari ve tüketici uygulamaları için de doğal dil işleme kullanımında önemli bir artış görülmüştür.

2010 ile 2020 yılları arasında, Doğal dil işleme alanında derin öğrenme ve sinir ağlarının yaygın olarak benimsenmesiyle karakterize edilen bir dönem olmuştur. Bu on yıl içerisinde, kelime katıştırma, dönüştürücü mimarileri ve büyük ölçekli önceden eğitilmiş modellerin ortaya çıkmasıyla doğal dil işleme uygulamalarında önemli ilerlemeler gerçekleştirilmiştir.

3.2. Doğal Dil İşleme Teknikleri

Günümüzde birçok alanda bulunan insan-bilgisayar etkileşimi nedeniyle, doğal dil işleme teknikleri birden fazla alanı kapsayan birçok farklı alanda sıklıkla kullanılmaktadır. Doğal dil işleme teknikleri; Dizi Sınıflandırma (Sequence Classification), Çift Yönlü Dizi Sınıflandırma (Pairwise Sequence Classification),

Kelime Etiketleme (Word Labeling) ve Diziden Diziye Sınıflandırma (Sequence2sequence Classification) olarak dört ana başlık altında toplanmıştır (Lauriola vd., 2022).

3.2.1. Dizi sınıflandırma

Dizi sınıflandırma, makinelerin farklı veri türlerini bir sırayla anlamasını ve kategorilere ayırmasını sağlayan bir tekniktir. Dizi sınıflandırma yöntemleri üç kategoriye ayrılmıştır (Xing vd.,). İlk kategori, bir diziyi bir özellik vektörüne dönüştüren ve daha sonra geleneksel sınıflandırma yöntemlerini uygulayan özellik tabanlı sınıflandırmadır. İkinci kategori, dizi mesafesi tabanlı sınıflandırmadır. Diziler arasındaki benzerliği ölçen mesafe fonksiyonu, sınıflandırmanın kalitesini önemli ölçüde belirler. Üçüncü kategori, dizileri sınıflandırmak için gizli markov modeli (HMM) ve diğer istatistiksel modellerin kullanılması gibi model tabanlı sınıflandırmadır.

Şekil 3.1 de dizi sınıflandırma modeli verilmektedir. Dizi sınıflandırma sıralı olarak verilen giriş verilerindeki deseni tamamlayarak bir sonraki değerin tahminini çıkış verisi olarak vermektedir. Dizi sınıflandırma sonraki değeri tahmin etme uygulamalarında kullanılabilmektedir.



Şekil 3.1. Dizi sınıflandırma modeli

3.2.2. İkili dizi sınıflandırma

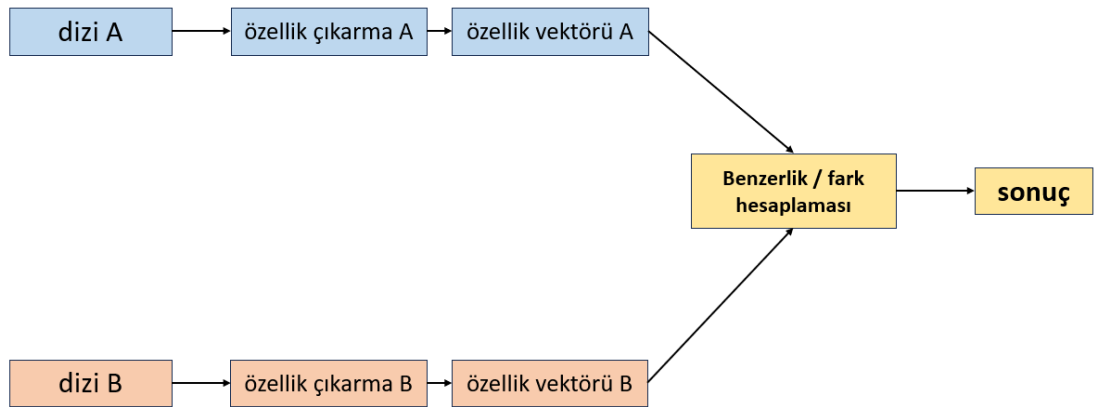
İkili Dizi Sınıflandırma (Pairwise Sequence Classification), iki farklı dizinin benzerliklerine, semantiklerine ve anlamlarına göre karşılaştırılıp sınıflandırılmasından oluşmaktadır (Lauriola vd., 2022). İki metin arasındaki ilişkiyi

belirlemek, anlamak ve sınıflandırmak amacıyla kullanılmakta ve doğal dil işleme alanında yaygın olarak uygulanmaktadır.

Denklem 3.1’de girdi olarak iki farklı dizi verildiğindeki sınıflandırma denklemi verilmektedir. İkili dizi sınıflandırması, girdi olarak iki farklı dizi verildiğinde, aynı anlamı ifade ediyorlarsa +1, aynı anlamı ifade etmiyorlarsa -1 döndüren ikili bir sınıflandırma görevidir.

$$fp : X \times X \rightarrow \{+1, -1\} \quad (3.1)$$

Şekil 3.2’de ikili dizi sınıflandırma modeli verilmektedir. İki metin ve bir etiket içeren veri setlerinde çalışırken metni parçalara ayırma ve özellik çıkarma işlemleri uygulanmakta ve her iki metnin özellikleri çıkarılmaktadır.



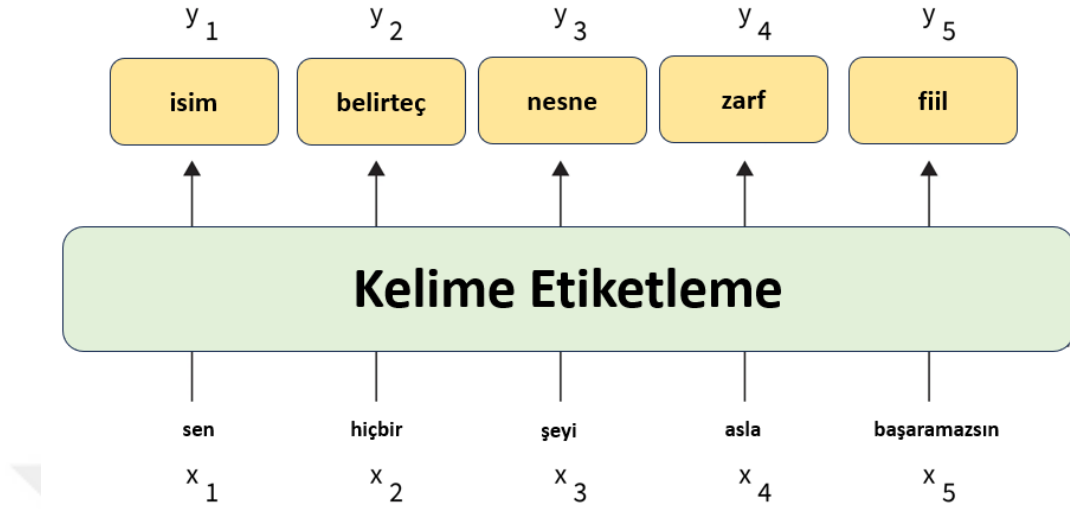
Şekil 3.2. İkili dizi sınıflandırma modeli

Çıkarılan özellikler, benzerlik ve fark hesaplaması için sınıflandırma katmanına aktarılmaktadır. Bu katman, iki metin arasındaki ilişkiyi tahmin ederek sonuç çıkarmaktadır.

3.2.3. Kelime etiketleme

Kelime etiketleme, bir dizideki her bir öğeye bir etiket eklemeyi içermektedir. Bu işlem, girdinin her bir öğesine karşılık gelen etiket dizilerinden oluşan bir çıktı alanı ile sonuçlanmaktadır. Şekil 3.2 de dil bilgisine göre kelime etiketleme şeması

verilmiştir. Giriş verisi olarak verilen her bir kelimenin $(x_1, x_2, x_3, \dots, x_n)$ etiket değerleri çıkış verisi $(y_1, y_2, y_3, \dots, y_n)$ olarak verilmektedir.



Şekil 3.3. Dil bilgisine göre kelime etiketleme

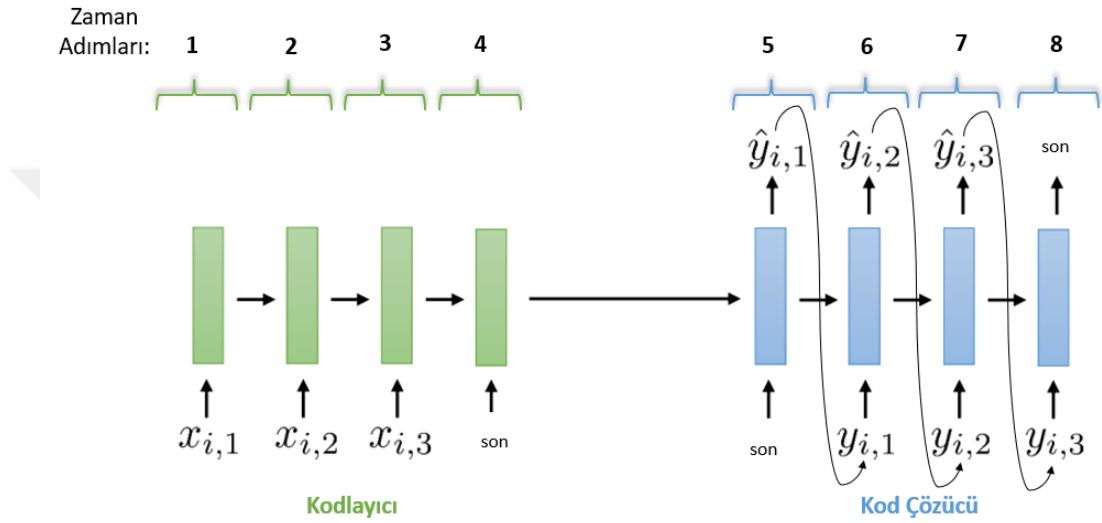
Kelime etiketleme bir metindeki bir kelimeyi konuşmanın belirli bir kısmına (fiil, isim, sıfat, ...) karşılık gelecek şekilde etiketleme işlemi olan Konuşma Parçası (PoS) uygulamalarında, girdi dizisinden ilgili nesnelerin (örneğin isimler, konumlar) tanımlandığı İsimlendirilmiş Varlık Tanıma (NER) uygulamalarında, girdi paragrafı tarafından verilen bir olasılık dağılımının cevabı içeren bir aralığı seçmek için kullanıldığı klasik soru cevaplama uygulamalarında kullanılabilmektedir (Lauriola vd., 2022).

3.2.4. Diziden diziye sınıflandırma

Diziden diziye sınıflandırma (Sequence2sequence classification) doğal dil işleme için kullanılan bir makine öğrenimi modelidir. Diziden diziye modelde amaç, bir girdi dizisini $x_1, \dots, x_m$, bir çıktı dizisine $y_1, \dots, y_m$, eşlemenin koşullu olasılığını $p(y|x)$ doğrudan modellemektir. Bu kodlayıcı ve kod çözücü ile mümkündür (Sutskever, Vinyals ve Le, 2014). Diziden diziye sınıflandırma modelinde giriş verileri olarak alınan bir dizi, sonuç çıktıları olarak başka bir diziye dönüştürmektedir. Kodlayıcı ve kod çözücü olmak üzere iki bloktan oluşmaktadır. Kodlayıcı bloğun temel amacı giriş dizisini işlemek ve sabit boyutlu bir bağlam vektöründeki bilgiyi yakalamaktır. Şekil 1' de diziden diziye sınıflandırma modeli görülmektedir. Bu modele göre, giriş dizisi

kodlayıcıya yerleştirilir. Kodlayıcı, giriş dizisinin her bir ögesini kullanarak işlemektedir (Lauriola vd., 2022).

Şekil 3.3’ de Diziden Diziye Sınıflandırma modelinin kodlayıcı ve kod çözücü modeli verilmektedir. $x = (x_1, \dots, x_m)$ m adet giriş verisi kodlayıcı tarafından işlenerek kod çözücüye aktarılmakta ve her seferinde bir öge olacak şekilde soldan sağa $y = (y_1, \dots, y_m)$ dizisini üretmektedir (Gehring, Auli, Grangier, Yarats ve Dauphin, 2017).



Şekil 3.4. Diziden diziye sınıflandırma modeli kodlayıcı ve kod çözücü (Gehring vd, 2017)

Süreç boyunca, kodlayıcı dahili bir durumu korumakta ve nihai son durum, tüm giriş dizisinin sıkıştırılmış bir gösterimini kapsayan bağlam vektörü olarak işlev görmektedir. Bu bağlam vektörü, giriş dizisinin anlamsal anlamını ve önemli bilgilerini yakalamaktadır. Kodlayıcının son durumu daha sonra bağlam vektörü olarak kod çözücüye geçirilmektedir. Kod çözücü, kodlayıcıdan gelen son bağlam vektörünü işleyerek y çıktı dizisini üretmektedir. Eğitim aşamasında, kod çözücü hem bağlam vektörünü hem de istenen hedef çıktı dizisini almaktadır. Çıkarım sırasında kod çözücü, sonraki adımlar için girdi olarak daha önce kendi ürettiği y çıktılarını kullanmaktadır.

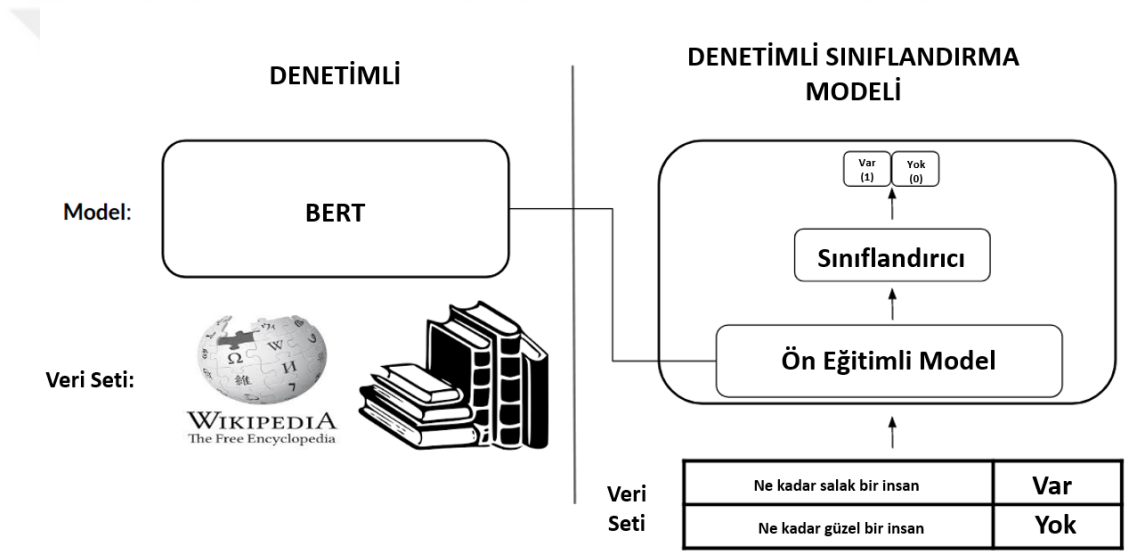
Her zaman adımında, kod çözücü olası sonraki belirteçler üzerinde bir olasılık dağılımı oluşturmak için geçerli son durumu, bağlam vektörünü ve önceki çıktı belirtecini

kullanır. Daha sonra en yüksek olasılığa sahip belirteç çıktı olarak seçilmekte ve işlem çıktı dizisinin sonuna ulaşılan kadar devam etmektedir.

3.3. Makine Öğrenme Algoritmaları

3.3.1. Transfer öğrenme

Transfer öğrenme, bir görevde kullanılan bir modelden elde edilen bilginin başka bir görev için temel nokta olarak yeniden kullanılabileceği bir makine öğrenmesi yöntemidir.



Şekil 3.5. Transfer öğrenme modeli

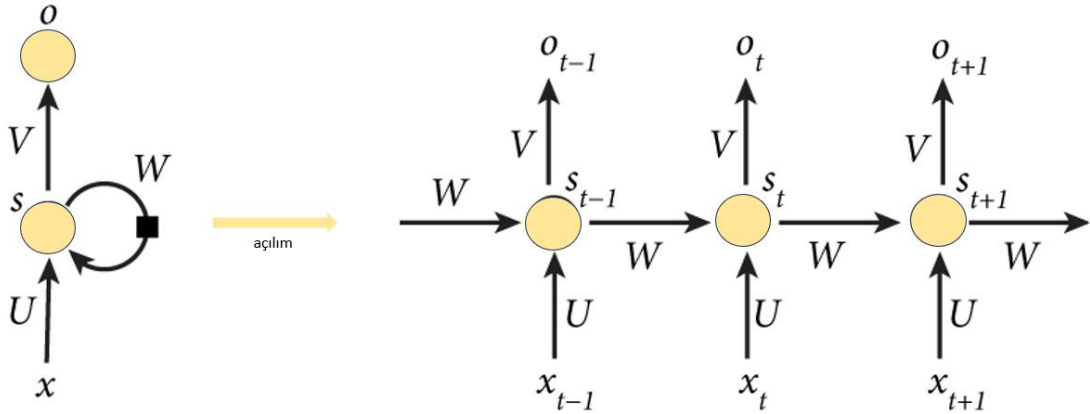
Makine öğrenimi algoritmaları, tahminlerde bulunmak ve yeni çıktı değerleri üretmek için girdi olarak geçmiş verileri kullanır. Genellikle izole edilmiş görevleri yürütmek üzere tasarlanırlar. Kaynak görev, bilginin hedef göreve aktarıldığı görevdir. Hedef görev, bilginin kaynak görevden aktarılması nedeniyle gelişmiş öğrenmenin gerçekleştiği görevdir.

Transfer öğrenme sırasında, bir kaynak görevden elde edilen bilgi ve hızlı ilerleme, öğrenmeyi ve yeni bir hedef göreve yönelik gelişimi iyileştirmek için kullanılır. Bilginin uygulanması, kaynak görevin niteliklerini ve özelliklerini kullanarak hedef göreve uygulanacak ve haritalanacaktır.

Transfer öğrenmede öne çıkan başlıca modeller şunlardır: bağlama duyarlı kelime yerleştirmelerini içeren ELMo, ULMFiT ve GPT modelleri, çift yönlü bir dil modeli olan BERT (Peters vd, 2018; Howard ve Ruder, 2018; Radford vd., 2018). Sonraki bölümlerde BERT modeli hakkında ayrıntılı bilgiler verilmiş ve bu modelin siber zorbalık tespiti problemine nasıl uygulanacağı açıklanmıştır.

3.3.2. Tekrarlayan sinir ağları

Tekrarlayan Sinir Ağları (Recurrent Neural Networks, RNN), nöronlar arasındaki bağlantılar aracılığıyla döngüler oluşturur, bu yüzden tekrarlayan sinir ağı olarak adlandırılmaktadır. Bu döngüler bir önceki adımdaki verileri depolamakta ve bir sonraki adımdaki nörona geri bildirim olarak iletmektedir. Tekrarlayan Sinir Ağları, bazı düğümlerden gelen çıktının aynı düğümlere gelen sonraki girdiyi etkilemesine izin veren iki yönlü yapay sinir ağıdır (Yin, Kann, Yu ve Schütze, 2017). Şekil 3.6'da Tekrarlayan Sinir Ağları ve zaman içindeki açılımı verilmektedir.

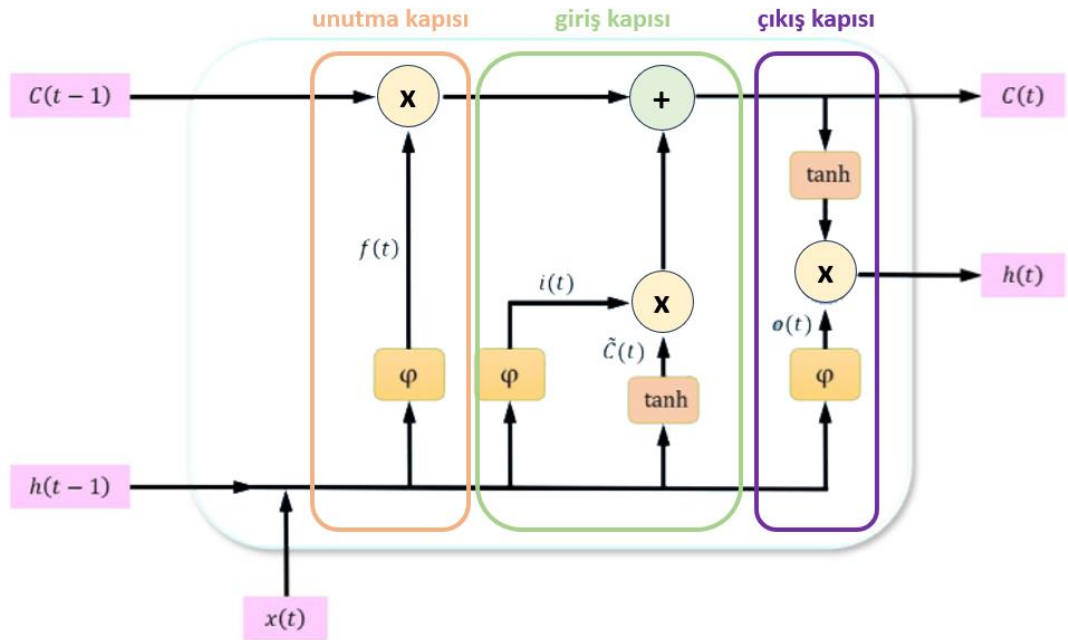


Şekil 3.6. Tekrarlayan sinir ağı ve zaman içinde açılımı (Feng, Guan, Li, Zhang ve Luo, 2017)

t zamanında s_t değerlerine sahip s düğümü altında gruplandırılmış nöronlar yapay nöron olarak geçmektedir. Yapay nöronlar önceki zaman adımlarında diğer nöronlardan girdi olarak tekrarlayan bir sinir ağı x_t elemanlı bir giriş dizisini O_t elemanlı bir çıkış dizisine eşleyebilir, her O_t önceki tüm x_t girişine bağlıdır ($t' \leq t$ için). Her zaman adımında U, V, W matrisleri kullanılmaktadır (Lecun, Bengio ve Hinton, 2015).

3.3.3. Uzun kısa süreli bellek ağları

1997 yılında Tekrarlayan Sinir Ağları'nın uzun vadeli bağımlılık sorununu çözmek için Uzun Kısa Süreli Bellek Ağları (Long-Short Term Memory Networks, LSTM) modeli önerilmiştir (Hochreiter ve Schmidhuber, 1997). Şekil 3.7'de LSTM hücresi modeli verilmektedir. Normal bir LSTM birimi, bir bellek hücresi, bir giriş kapısı, bir çıkış kapısı ve bir unutma kapısından oluşmaktadır. Hücre, rastgele zaman aralıklarında değerleri hatırlar ve üç kapı, hücreye giren ve çıkan bilgi akışını düzenlemektedir (Hochreiter ve Schmidhuber, 1997).



Şekil 3.7. LSTM hücresi (Le, Ho, Lee ve Jung, 2019)

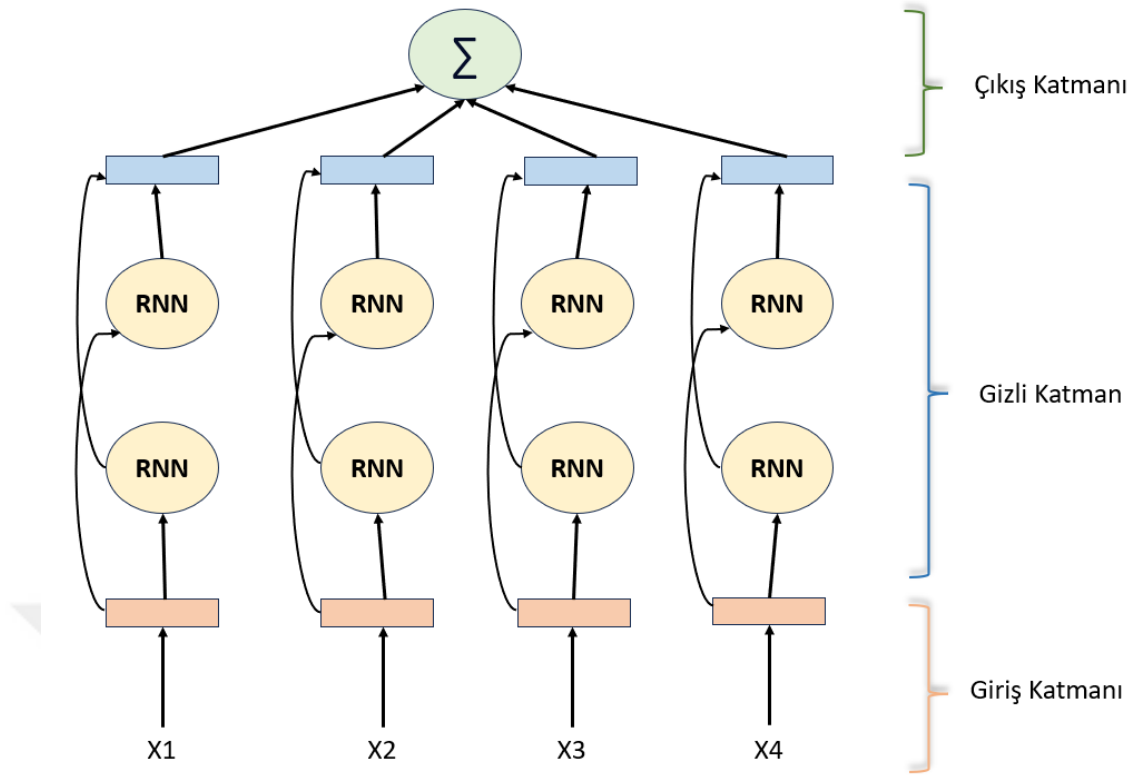
Şekil 3.7'de LSTM hücresine girdileri $x(t)$ ve önceki durumu $h(t-1)$ Uzun süreli bellek hücre durumu olarak da adlandırılır ve uzun süreli durumların hangi kısımlarının unutulması gerektiğini tanımlayan unutma kapısı $f(t)$ tarafından kontrol edilmektedir. Giriş kapısı, $i(t)$, uzun süreli durumlara hangi parçaların eklenmesi gerektiğini kontrol ederken, çıkış kapısı, $o(t)$, mevcut zaman durumlarında hangi parçaların üretilmesi gerektiğini kontrol etmektedir. $x(t)$ giriş sinyaline karşılık gelmektedir, $C(t)$ bellek hücresinin mevcut durumunu ve $C(t-1)$ bellek hücresinin

önceki anını, $h(t)$ mevcut gizli durumunu ve $h(t - 1)$ önceki gizli durumunu temsil etmektedir (Diniz, Ciarelli, Salles ve Coco, 2024). Gizli durum kısa süreli belleği, hücre durumu ise uzun süreli belleği temsil etmektedir. Uzun süreli bellek tarafından yayılan bilgi, kapıların bilgi çıkarma veya ekleme özelliği aracılığıyla hücre ile etkileşime girmektedir. Böylece hücreler hangi zamansal bilginin ağ tarafından iletilmesi ya da atılması gerektiğini belirleyebilmektedir. Bu şekilde, yalnızca seçilen bilgilerin ağdan geçmesi sağlanmaktadır. Her bir hücrenin dahili durumları, her bir adım işlendikten sonra sıfırlanmaktadır (Buduma ve Locascio, 2017, Goodfellow vd., 2016).

3.3.4. Çift yönlü tekrarlayan sinir ağları

1997 yılında normal bir Tekrarlayan Sinir Ağı'nın sınırlamalarının üstesinden gelmek için, belirli bir zaman diliminin geçmiş ve geleceğindeki mevcut tüm girdi bilgileri kullanılarak eğitilebilen çift yönlü bir tekrarlayan sinir ağı (Bidirectional RNN, BRNN) tasarlanmıştır (Schuster ve Paliwal, 1997). Çift Yönlü Tekrarlayan Sinir Ağları, t zamanındaki çıktılarını sıralardaki önceki öğelere değil gelecekteki öğelere de bağlı olabileceği fikrine dayanmaktadır. Çift Yönlü Tekrarlayan Sinir Ağları, gizli bağlantıların ters zamansal sırayla aktığı ikinci bir gizli katman ekleyerek Tek Yönlü Tekrarlayan Sinir Ağını genişletmektedir. Bu nedenle model hem geçmişten hem de gelecekte gelen bilgilerden faydalanabilmektedir (Berglund vd.).

Şekil 3.8'de Çift Yönlü Tekrarlayan Sinir Ağları Yapısı verilmektedir. Burada iki RNN çıktısı birleştirilmektedir. Böylece biri işlemi ileri yönde yürütürken, ikincisi işlemi gösterildiği gibi geri yönde çalıştırmaktadır. İlk tekrarlayan sinir ağına girdi normal zaman sırasına göre verilirken, ikincisine ters zaman sırasına göre verilmektedir. Bu yapı, ağların her zaman adımında dizi hakkında hem geriye hem de ileriye doğru bilgi sahibi olmasını sağlamaktadır.



Şekil 3.8: Çift yönlü tekrarlayan sinir ağı yapısı (Iqbal ve Qureshi, 2022)

Denklem 3.2’ de t zaman adımıdaki çıktının hesaplanma yöntemi verilmektedir.

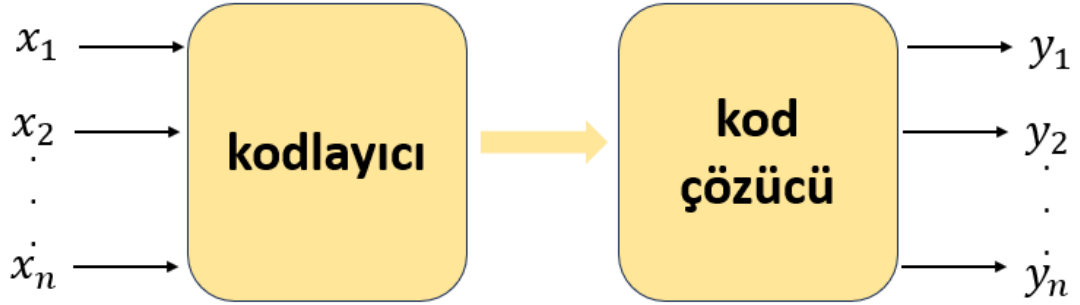
$$Y_t = g(w[htf, htb] + b) \quad (3.2)$$

‘htf’ ve ‘htb’ sırasıyla ileri ve geri yöndeki gizli durumu, w bunlarla ilişkili ağırlıkları, b modelin verilen verilere uymasına yardımcı olan yanlılığı ve g ağı doğrusal olmayan ağı eklemek için kullanılan aktivasyon fonksiyonunu temsil etmektedir (Berglund vd.). Çift yönlü tekrarlayan sinir ağı blokları tekrarlayan sinir ağlarından, uzun kısa süreli bellek ağlarından oluşabilmektedir.

3.3.5. Kodlayıcı ve kod çözücü

Kodlayıcı (encoder), bir giriş dizisini alarak bir dizi bağlam vektörü oluşturmak için diziyi işlemektedir. İşlenen dizi kod çözücü (decoder) tarafından kullanılmaktadır. Şekil 3.6.’da kodlayıcı ve kod çözücünün çalışma modeli verilmektedir. $(x_1, x_2, \dots, x_n)$ den oluşan giriş dizini kodlayıcı tarafından işlenerek kod çözücüye iletilmektedir.

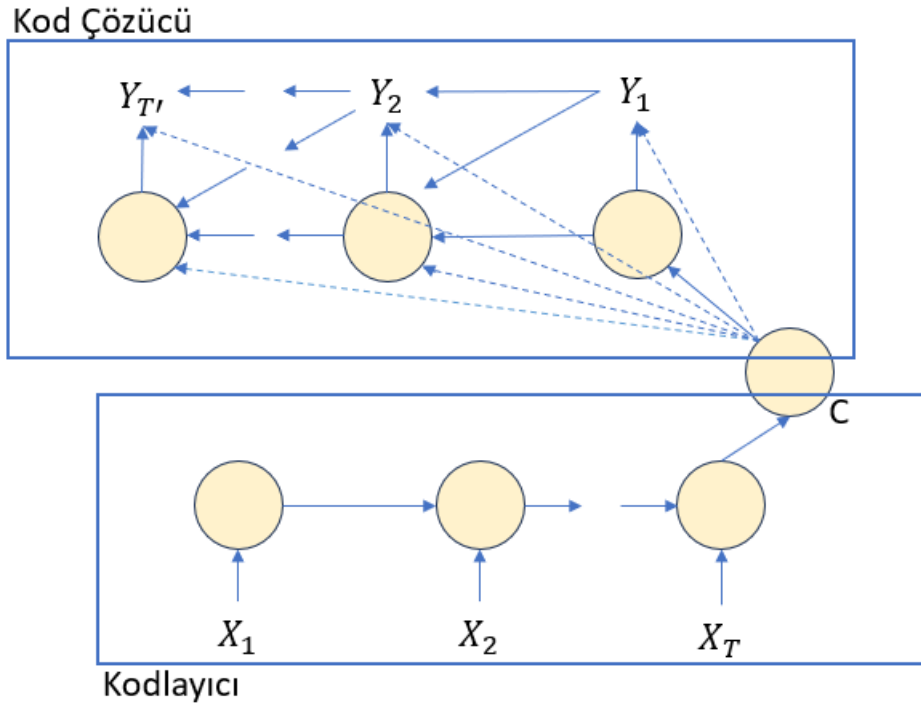
İşlenen dizi kod çözücü tarafından $(y_1, y_2, \dots, y_n)$ çıkış dizisine dönüştürülmektedir (Tixier, 2018).



Şekil 3.9. Kodlayıcı ve kod çözücü modeli

Kodlama sürecinin çalışma şekli, kullanılan uygulama türüne bağlıdır. Örneğin, makine çevirisi veya özetleme gibi metin uygulamaları için, her cümledeki kelimeler matematiksel olarak temsil eden sayısal değerlere dönüştürülmektedir. Dönüştürülen sayısal değerler, cümle yapısı içinde birbirleriyle nasıl ilişkili olduklarına dair ilgili bilgileri korurken boyutlarını azaltan bir dizi katmandan geçirilir. Her cümle için bu "kodlanmış" versiyonu daha sonra daha fazla işlenmek üzere kod çözücüye iletilmektedir (Bahdanau vd., 2014).

Kod çözücü, bu kodlanmış gösterimi alarak orijinal biçimine geri dönüştürmekten sorumludur. Bunu yapmak için, kodlanan ile yeniden yapılandırılması gereken arasında bir tür ilişki olmalıdır. Bu bağlantıyı kurmak için, çoğu modern mimari, bir girdi dizisinin belirli bölümlerinin (örneğin, tek tek kelimeler) daha sonraki bölümlerin model tarafından nasıl işlendiğini veya yorumlandığını etkilemesine izin veren dikkat mekanizmaları kullanılmaktadır. Dikkat mekanizması veri girişlerini kodlayarak çıktı dizileri üretirken belirli öğelere diğerlerinden daha fazla ağırlık veya önem vermektedir.



Şekil 3.10. Kodlayıcı ve kod-çözücü çalışma şekli (Bahdanau vd, 2014)

3.3.6. Dikkat mekanizması

Dikkat mekanizması (Attention Mechanism), bir modelin girdi verisinin belirli bölümlerine “daha fazla dikkat etmesi” gerektiğini belirlemesine olanak tanımaktadır (Niu, Zhong ve Yu, 2021). Dikkat mekanizması sorgular (queries, Q), anahtarlar (keys, K) ve değerler (values, V) olmak üzere üç temel bileşenden oluşmaktadır. Dikkat mekanizmasının çalışma adımları; nokta çarpımı, ölçekleme, Softmax ve ağırlıklı toplam adımlarından oluşmaktadır. Nokta Çarpımı adımında, her sorgu ve anahtar çifti arasındaki benzerlik ölçülmektedir. Değer (score), sorguların ve anahtarların iç çarpımını hesaplayarak yapılır. Denklem 3.4’de değer hesaplaması verilmektedir.

$$score(Q, K) = QK^T \quad (3.4)$$

Büyük boyutlu girdilerle çalışırken, büyük değerlere karşı denge sağlamak ve softmax fonksiyonunun daha stabil çalışmasını sağlamak için iç çarpım değerleri $\sqrt{d_k}$ ile ölçeklendirilmektedir. Ölçeklendirilmiş değerlerin olasılık dağılımına dönüştürülmesi için softmax fonksiyonuna uygulanarak dikkat ağırlıkları elde edilmektedir. Denklem 3.5’de Softmax fonksiyonu ile $a_{i,j}$ değerinin hesaplanması verilmektedir.

$$a_{ij} = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) \quad (3.5)$$

Modelin önemli bilgilere odaklanmasını sağlamak için, dikkat ağırlıkları ilgili değerlere uygulanarak ağırlıklı bir toplam hesaplanmaktadır. Denklem 3.6’da dikkat ağırlıklarının hesaplanması verilmektedir.

$$\text{attention}(Q, K, V) = \sum_j a_{ij} V_j \quad (3.6)$$

Dikkat mekanizması, modelin belirli bilgilere odaklanmasını sağlayarak performansı artırmakta ve uzun bağımlılıkları daha etkili bir şekilde yakalamaktadır (Vaswani vd.,).

Dikkat Mekanizması üç türden oluşmaktadır: Öz Dikkat (Self-Attention), Çok Başlı Dikkat (Multi-Head Attention) ve Çarpaz Dikkat (Cross-Attention). Öz Dikkat, bir cümlede veya metnin içindeki her bir elemanın, diğer tüm elemanlarla olan ilişkisini değerlendirmektedir. Öz dikkat, dikkat bileşeninin dizinin bir temsilini hesaplamak için tek bir dizinin farklı konumlarını ilişkilendirmektedir. Bu şekilde, anahtarlar, değerler ve sorgular aynı kaynaktan gelmektedir (Hernández ve Amigó, 2021).

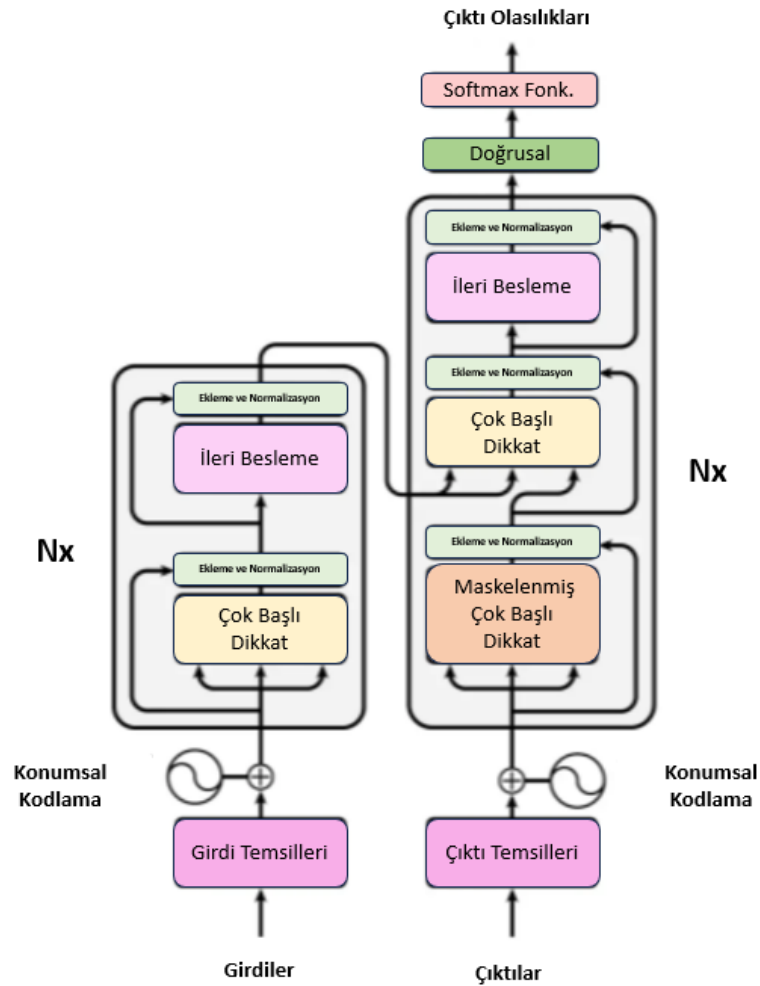
Çok Başlı Dikkat, öz dikkat mekanizmasının bir genişlemesidir ve modelin farklı “başlıklar” (heads) üzerinden veriyi paralel olarak işlemesine olanak tanımaktadır. Her bir “head”, verinin farklı temsillerine odaklanır, böylece model aynı anda birden fazla bağlamı ve ilişkiyi değerlendirebilmektedir. Bu, modelin genel anlama yeteneğini artırmaktadır.

Çarpaz dikkat, genellikle kodlayıcı-kod çözücü yapılarında kullanılmaktadır. Kod çözücünün, kodlayıcı tarafından üretilen temsillere dikkat etmesini sağlamaktadır. Bu, özellikle metin çevirisi gibi görevlerde, kaynak metnin belirli bölümlerine göre hedef metin üretmekte kullanılabilmektedir (Vaswani vd.,).

3.3.7. Dönüştürücüler

2017 yılında yeni bir sinir ağı mimarisi olarak öz-dikkate dayalı dönüştürücüler (transformers) ortaya konulmuştur (Vaswani vd.). Dönüştürücüler, kodlayıcı-kod çözücü mimarisindeki geleneksel tekrarlayan katmanları öz-dikkatle değiştirerek iyi bir performans göstermektedir (Soydaner, 2022). Doğal dil işleme problemlerinin çözümünde yaygın olarak kullanılan dönüştürücü modelleri kodlayıcı-kod çözücü yapısını kullanarak verileri işlemek için tasarlanmıştır. Şekil 3.11’de dönüştürücüler ve kodlayıcı-kod çözücü mimarisi verilmektedir.

Bir Tekrarlayan Sinir Ağı kodlayıcısı, her bir gizli durumu birer birer üretirken, dönüştürücü kodlayıcı tüm giriş dizisini bir kerede işlemekte ve tüm gizli durumları aynı anda üretmektedir.

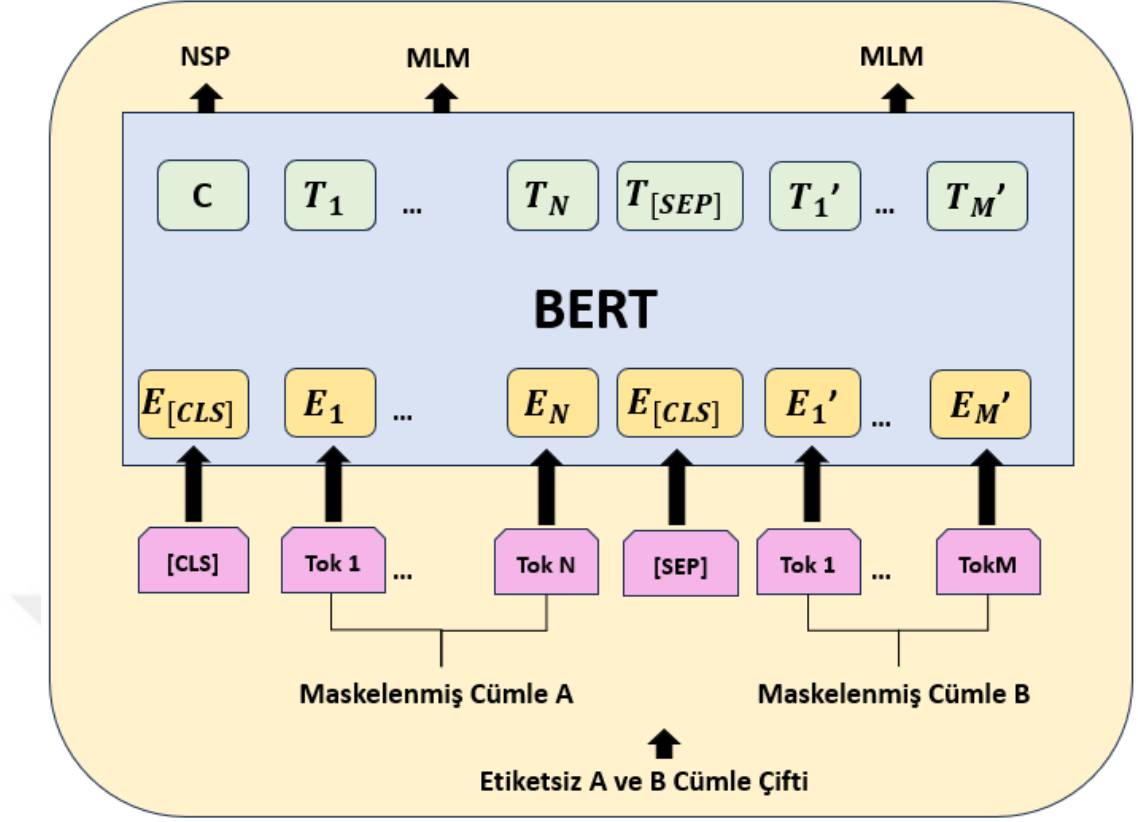


Şekil 3.11. Dönüştürücü kodlayıcı-kod çözücü mimarisi (Vaswani vd.,)

3.4. BERT Algoritması

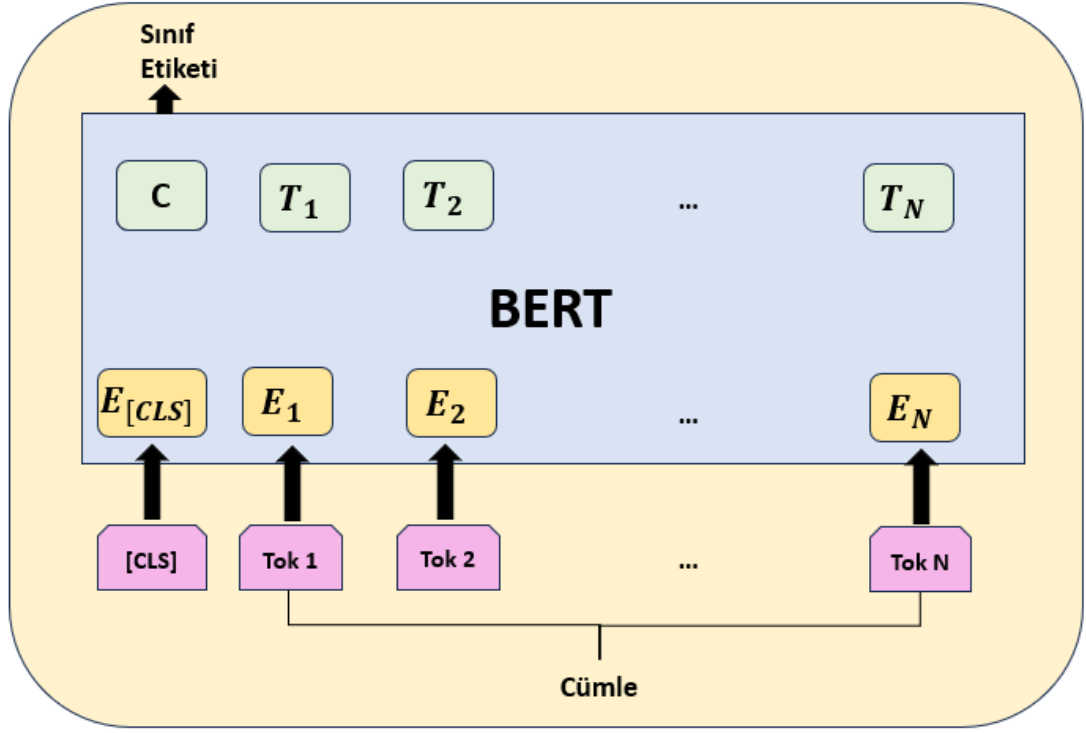
Dönüştürücülerin makine çevirisinde orijinal diziden diziye formülasyonu ile elde ettiği doğal dil işleme görevlerinde de dönüştürücü kullanımını artırmıştır (Patwardhan, Marrone ve Sansone, 2023). Bu yöndeki en önemli gelişmelerden biri 2018 yılında çok çeşitli doğal dil işleme görevlerini yerine getirebilen açık kaynak kodlu bir sinir ağı olan BERT'in (Transformatörlerden Çift Yönlü Kodlayıcı Gösterimleri) ortaya konulmasıdır (Devlin vd, 2018). BERT modeli, derin çift yönlü ön eğitim elde etmek için Tekrarlayan Sinir Ağları'na alternatif olan dönüştürücü (transformer) mimarisini kullanmaktadır (Khadhraoui vd, 2022). Dönüştürücü birçok doğal dil işleme görevi için yeni bir mimari sağlamaktadır. Tekrarlayan Sinir Ağları'nın eğitim seti için büyük ve etiketli veri setlerine ihtiyaç duyulması, örüntüleri matematiksel yaklaşım ile tespit ederek bu gereksinimi dönüştürücü ile ortadan kaldırmıştır (Wang vd., 2019).

Şekil 3.12'de BERT'in ön eğitim modeli verilmektedir. BERT öncelikle ön eğitim adımını gerçekleştirmekte ve daha sonra detaylandırmaktadır. Ön eğitim sırasında model, farklı eğitim öncesi görevler üzerinden etiketlenmemiş veriler üzerinde eğitilmektedir. BERT ön eğitim sırasında iki teknik kullanmaktadır, maskeli dil modellemesi (Masked Language Modeling, MLM) ve sonraki cümle tahmini (Next Sentence Prediction, NSP). Maskeli dil modellemesinde bir cümlede bulunan bir kelime maskelenerek, maskelenen kelimenin tahmini için modelin iki yönlü olarak analiz etmesini sağlamaktadır. Model bir belirtecin hem başındaki hem de sonundaki bağlamlara bakarak maskelenmiş olan bağlamı tahmin etmeye çalışmaktadır. Bu, modelin belirli bir NLP görevine adapte edilmesini sağlamaktadır. BERT ön eğitim modelinde [CLS] her girdi örneğinin önüne eklenen özel bir semboldür ve [SEP] özel bir ayırıcı belirteçtir (Devlin vd, 2018).



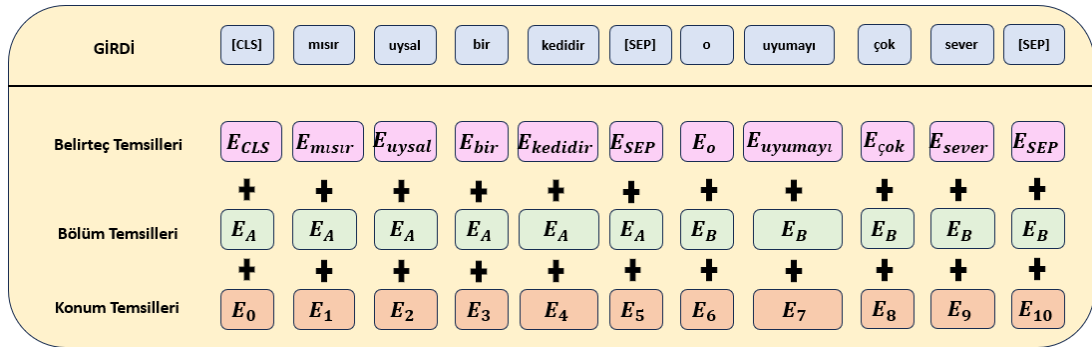
Şekil 3.12. BERT mimarisi ön eğitim modeli (Devlin vd, 2018)

Şekil 3.13'te cümle sınıflandırma için BERT hassas ayar modeli verilmektedir. Hassas ayar için BERT modeli ilk olarak önceden eğitilmiş parametrelerle başlatılmakta ve tüm parametrelere, aşağı akış görevlerinden gelen etiketli veriler kullanılarak hassas ayar yapılmaktadır. Her bir aşağı akış görevi, aynı önceden eğitilmiş parametrelerle başlatılmış olsalar bile, ayrı hassas ayarlı modellere sahiptir. Hassas ayarlara göre BERT, cümle çifti sınıflandırma görevleri, tek cümle sınıflandırma görevi, soru cevaplama görevi ve tek cümle etiketleme görevlerini gerçekleştirebilmektedir (Devlin vd, 2018). Bir metin sınıflandırma görevinde modele hassas ayar yapılırken, [CLS] belirtecinin çıktı temsili, softmax sınıflandırma katmanını beslemek için kullanılır (Sousa vd, 2019).



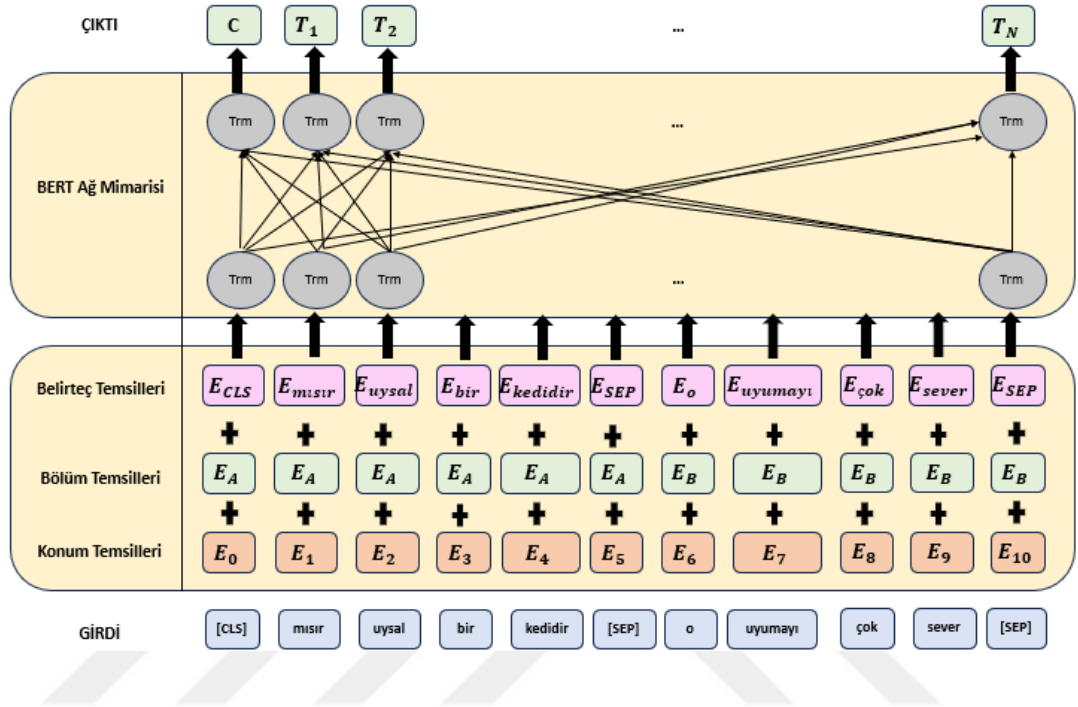
Şekil 3.13. Cümle sınıflandırma için BERT hassas ayar modeli (Sousa vd, 2019)

BERT modelinde bulunan konum temsilleri, dizideki belirteçlerin konumu hakkında bilgi içermektedir. Girdi [CLS] etiketi ile başlamakta ve [SEP] etiketi ile bölünmektedir. Şekil 3.14'te BERT modelinin girdi temsilleri verilmektedir. Bölüm temsilleri [SEP] etiketi ile E_A ve E_B olmak üzere ikiye ayrılmıştır. İlk cümlenin belirteçleri, belirteç temsilleri olarak E_0 şeklinde önceden tanımlanmış bir yerleştirmeye sahipken, ikinci cümlenin belirteçleri, E_1 şeklinde önceden tanımlanmış bir yerleştirmeye sahip olacaktır.



Şekil 3.14. BERT modelinde girdi temsilleri

Şekil 3.15'te BERT model mimarisi verilmektedir. Ağ mimarisi katmanında Trm olarak dönüştürücüler ifade edilmektedir. Bu katmanda bulunan dönüştürücüler, modelin metin verilerini işleyip anlamlandırmasını sağlayan karmaşık bir sinir ağı yapısıdır.



Şekil 3.15. BERT model mimarisi (Zhang, Fan ve Hei, 2022)

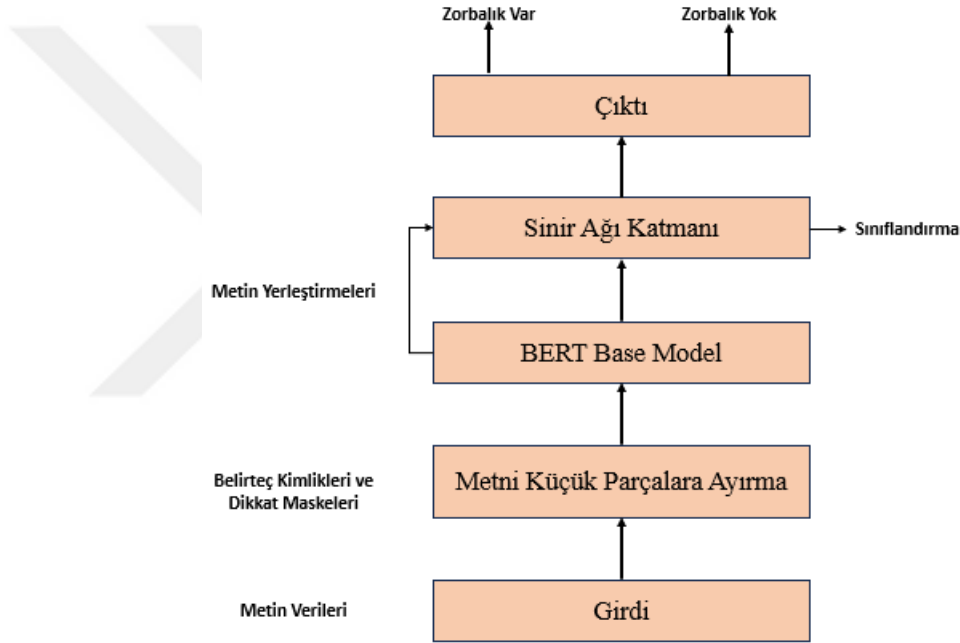
Özellikle, öz dikkat mekanizması, her kelimenin diğer tüm kelimelerle olan ilişkisini anlamaya çalışmaktadır. BERT modelinin çift yönlü olması, her kelimenin hem solundaki hem de sağındaki kelimelerle etkileşimini göz önünde bulundurmasına olanak tanımaktadır. Son kodlayıcı katmanından elde edilen vektörler, her belirteç için bağlamsal olarak zengin temsillerdir.

BERT için $L = 12$ kodlayıcı bloğundan oluşan BERT Base ve $L = 24$ kodlayıcı bloğundan oluşan BERT Large modelleri olarak ortaya konulmuştur (Devlin vd, 2018). Modellerin kodlayıcı blok sayısı, gizli katman, öz dikkat başlığı ve parametre değerleri Çizelge 3.1'de verilmektedir. BERT'in performansı model tipine bağlıdır. BERT Large modeli, BERT Base modelinden daha yüksek doğruluklara ulaşabilmektedir. Bununla birlikte, BERT Large kullanılarak doğrulukta sağlanan bu iyileşme, tamamlanması için daha kapsamlı kaynak gerektirmektedir (Acheampong ve Chen, 2021).

Çizelge 3.1. BERT BASE ve BERT LARGE Modelleri

Model	Kodlayıcı Blok Sayısı (L)	Gizli Katman (H)	Öz Dikkat Başlığı (A)	Parametre
BERT BASE	12	768	12	110M
BERT LARGE	24	1024	16	340M

Şekil 3.16’da siber zorbalık tespiti için BERT Base modeli ile işlem basamakları verilmektedir. Bu işlem sonucu çıktılarına bir etiket (label) verilerek cümle içerisinde zorbalık olup olmadığı belirlenmektedir.



Şekil 3.16. BERT Base ile siber zorbalık tespiti işlem basamakları

3.5. Konuşma Tanıma

Konuşmanın hayatımızdaki, insan iletişimindeki yeri yadsınamayacak düzeyde önemlidir. Gelişen teknoloji ile birlikte konuşma artık yüz yüze düzeyden çıkmış, elektronik ortamda da önemli bir yer edinmiştir. Konuşma düzeyinde kalmak ile yetinmeyip ses verisini metin verisine dönüştürerek veya tam tersi metin verisini ses verisine dönüştürerek ilk kullanımdan bugüne kadar olan bu süreçte konuşma tanıma alanında birçok teknolojik gelişme gerçekleşmiştir. Bu gelişme insan bilgisayar etkileşimi açısından önemli bir yere sahiptir. Konuşma Tanıma veya Otomatik

Konuşma Tanıma olarak isimlendirilen bu teknoloji ses dalgalarını metin verisine dönüştürmeyi sağlamaktadır.

Günümüzde konuşma tanıma, sesli komutlarla çalışan elektronik sistemlerde, konuşmaya bağlı olarak işlem yapabilme özelliğine sahip uygulamalarda, konuşma verisine göre yapılan sınıflandırmalarda ve çözümlemelerde büyük ölçüde kullanılmaktadır. Akıllı bileklikler ve saatler üzerinden yaptığımız işlemler, sesli komut verdiğimiz otonom sistemler bu teknolojilerin günlük yaşamda en çok karşımıza çıkan örneklerinden birkaçıdır.

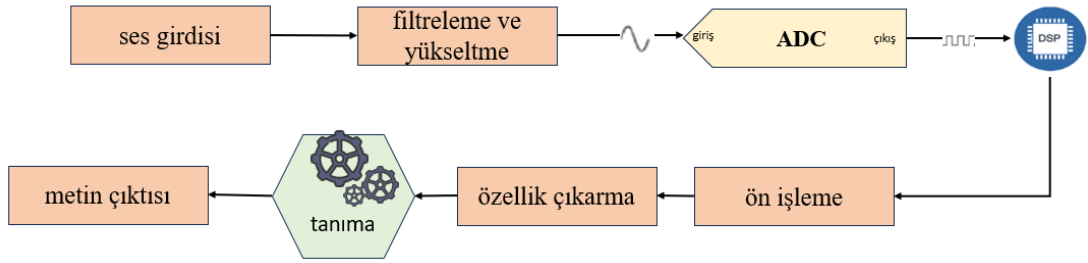
Konuşmadan metne çevirmede ücretli ve ücretsiz açık kaynak kodlu birçok teknoloji bulunmaktadır. İhtiyaca göre bunların kullanımı, en uygun olanın hangisi olduğunu belirlemek önemli bir husustur. Otomatik Konuşma Tanıma (ASR) Sistemleri, IBM Watson, Google Cloud Speech to Text, Mozilla Deep Speech gibi teknolojiler ile konuşma tanıma büyük ölçüde sağlanmaktadır.

Otomatik konuşma tanıma, konuşmacı ve dil tanıma, konuşmacı ayrımı, ses üretme ve kodlama ses teknolojilerinin uygulamaları olarak tanımlanmaktadır. Otomatik konuşma tanıma kendi içerisinde sürekli ses ve ayrık ses tanıma, sözcük yakalama olarak üçe ayrılmaktadır. Otomatik konuşma tanıma ses teknolojileri arasında uygulama alanı en zor olan alan olarak belirtilmektedir (Yalçın, 2008).

3.5.1. Otomatik konuşma tanıma

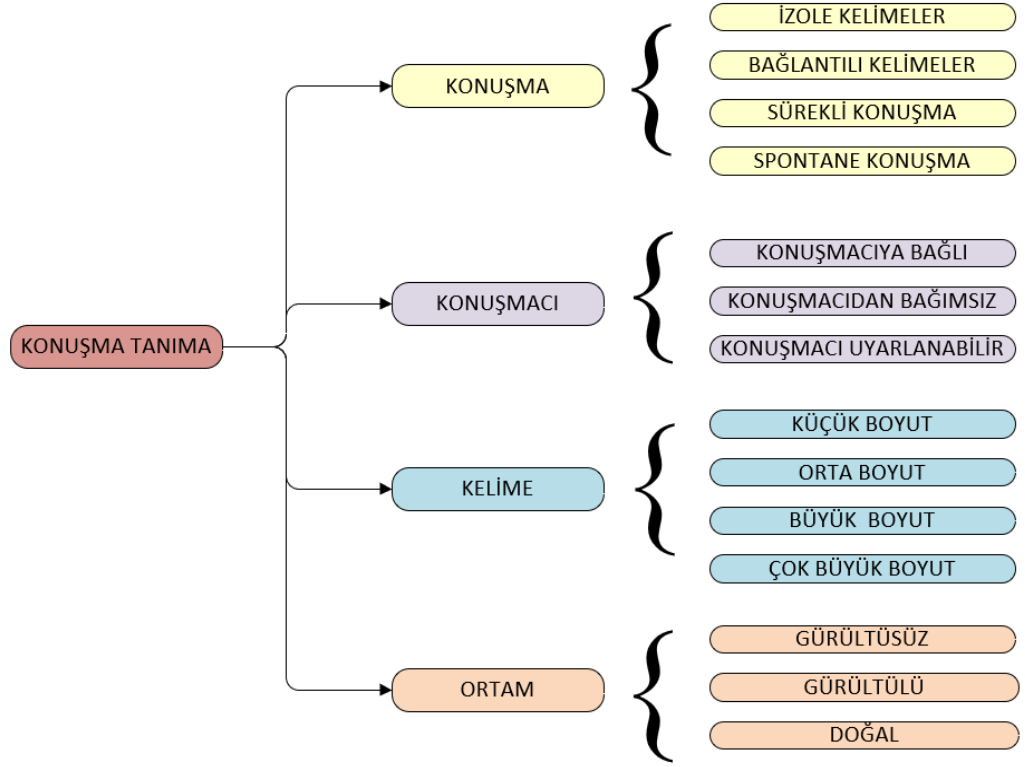
Otomatik Konuşma Tanıma (ASR) veya Konuşmayı metne dönüştürme (STT), ham sesi karşılık gelen kelimeler dizisine dönüştürmeyi amaçlayan bir çalışma alanıdır. Konuşma tanıma, dil bilimi ve bilgisayar bilimlerinden oluşan multidisipliner bir araştırma alanıdır. Aynı zamanda konuşmayı metne dönüştürme, bilgisayar konuşma tanıma veya otomatik konuşma tanıma olarak da adlandırılır. Otomatik konuşma tanıma sistemi, konuşulan sözcükleri .raw veya .wav uzantılı dosyalar gibi ses biçiminde kabul eder, ardından içeriğini metin biçiminde oluşturur ve bilgisayarın doğal dili anlamasını sağlar.

Otomatik konuşma tanıma sisteminin tipik bir sinyal akış diyagramını temsil eder. Şekil 3.16'da konuşma tanıma sistemlerinin çalışma şekli verilmektedir. Konuşma sinyalleri, tanıma modülü tarafından işlenmeden önce seri işlev birimlerinin çoklu kazanç kontrollerinden geçer. Genel olarak, kullanıcı ucuna yakın işlev birimleri arasında mikrofon(lar), kenar yumuşatma filtreleme ve dinamik aralık ayarı, analogdan dijitale dönüştürücü (ADC) ve geliştirme, gürültü giderme, ses kodlama ve kod çözme gibi temel dijital sinyal işleme (DSP) bulunmaktadır. İşlev birimleri arasında ön işleme, özellik çıkarma ve tanıma bulunmaktadır. Bu işlemler sonucu girdi olarak alınan ses verileri, metin verisi şeklinde çıktı olarak verilmektedir (D. Wang, Wei, Zhang, Ji ve Wang, 2022) .



Şekil 3.17. Konuşma tanıma sistemleri çalışma şekli

Otomatik konuşma tanıma sistemleri verilen eğitim verilerinin modellerini oluşturmaktadır. Eğitim aşamasında eğitilen bu modeller kullanılarak konuşma test edilmektedir. Şekil 3. 17'de otomatik konuşma tanıma kategorileri verilmektedir. Konuşma türüne, konuşmacıya, kelime dağarcığına ve çevresel koşullara bağlı olarak otomatik konuşma tanıma sistemleri farklı kategorilerde sınıflandırılabilir (Bhardwaj vd, 2022).

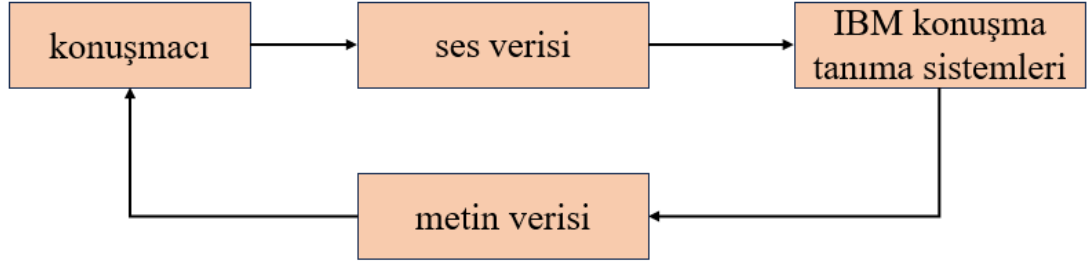


Şekil 3.18. Otomatik konuşma tanıma sınıflandırma kategorileri

Otomatik Konuşma Sistemleri ortama bağlı olarak gürültülü, gürültüsüz veya doğal ortamlar gibi farklı çevre koşullarında test edilebilir. Aynı zamanda konuşma, konuşmacı ve kelime faktörlerine bağlı olarak test işlemleri de değişmektedir. Konuşma tanıma için konuşmada dikkat edilmesi gereken 4 farklı husus bulunmaktadır; konuşmacı günlüğü, konuşmacı tanıma, konuşma dilini anlama ve duygu analizi.

3.5.2. Konuşmadan metne verilerin dönüştürülmesi

IBM konuşma tanıma sistemleri, konuşma verilerini metin verilerine dönüştüren bir hizmettir. Bu hizmet, konuşma tanıma teknolojisi kullanarak sesli konuşmayı yazılı metne çevirir.



Şekil 3.19. IBM konuşma tanıma sistemleri

Şekil 3.18’de IBM konuşma tanıma sistemlerinin çalışma adımları verilmektedir. Ses verisini metin verisine dönüştürürken ilk olarak ses verisinin alımı gerçekleşmektedir. Kullanıcıdan veya kaynaktan ses kaydı, canlı konuşma veya video şeklinde gelen ses verisini alarak ses tanıma algoritmaları ile sesin özelliklerini analiz etmektedir. Sesin özelliklerinde frekans, hız, ton gibi kriterler bulunmaktadır. Ses verisinin işlenmesinden sonra ise metin verisi oluşturmak için işlenmiş ses verisini metin haline getirmektedir.

3.5.3. Konuşma tanıma kütüphanesi

Bu çalışmada ortam dinlemesi ile elde edilen ses verilerini metin verilerine çevirebilmek için Python konuşma tanıma kütüphanesi kullanılmıştır. Python konuşma tanıma kütüphanesi Python programlama dilinde kullanılan ve çeşitli otomatik konuşma tanıma hizmet sağlayıcılarına erişim sağlayan bir kütüphanedir. Konuşma tanıma, öncelikle konuşulan dili yazılı metne veya bilgisayar komutlarına dönüştürmeye odaklanmaktadır. Bir kullanıcı tarafından konuşulan kelimeleri ve ifadeleri anlamak ve bunları metne aktarmakla ilgilenmektedir. Konuşma tanıma, ses dosyasını veya mikrofondan gelen sesi metne dönüştürmekte ve ses verisinde bulunan gürültüyü azaltmaktadır.

Konuşma tanıma işlemini gerçekleştirmek için bir mikrofon aracılığı ile ses dosyaları elde edilmiştir. Bu ses dosyaları kullanılarak konuşma tanıma işlemi gerçekleştirilmiş ve bu veriler daha sonrasında metin verilerine dönüştürülmüştür.

3.6. Değerlendirme Kriterleri

Daha sonra modelden elde edilen sonuçlar, veri setinde bulunan sonuçlar ile karşılaştırılarak modelin performansı incelenmiştir. Performans dört ana metrik üzerinden değerlendirilmiştir: hassasiyet (precision), duyarlılık (recall), F1 puanı (F1-score) ve doğruluk (accuracy). Bu metriklerde , gerçek pozitif (true positive, TP), pozitif sınıfa ait olarak doğru şekilde etiketlenen örnekler içeren değerlere göre, gerçek negatif (true negative, TN), negatif olarak doğru şekilde sınıflandırılan negatif örnekler içeren değerlere göre, yanlış pozitif (false positive, FP), sınıfa ait olarak yanlış şekilde etiketlenen örnekler içeren değerlere göre, yanlış negatif (false negative, FN), pozitif sınıfa ait olarak etiketlenmeyen ancak olması gereken örnekler içeren değerlere göre hesaplanmaktadır (Islam vd, 2020).

Karmaşıklık matrisi (confusion matrix), sınıflandırılan modelin performansının kullanışlı ve kapsamlı bir sunumudur (Krstinić, Braović, Šerić ve Božić-Štulić, 2020). Şekil 3.20’de karmaşıklık matrisinin gösterimi verilmektedir. Karmaşıklık matrisinde bulunan değerler kullanılarak modelin değerlendirme metrikleri hesaplanabilmektedir.

		Tahmin	
		Pozitif (1)	Negatif (0)
Gerçek	Pozitif (1)	TP	FN
	Negatif (0)	FP	TN

Şekil 3.20. Karmaşıklık matrisi

F1 puanı, hassasiyet (precision) ve duyarlılık (recall) arasında bir denge kurmaktadır. Hassasiyet (Precision), doğru olarak pozitif sınıflandırılan örneklerin, tüm pozitif sınıflandırılan örneklere oranıdır (Chicco ve Jurman, 2020). Hassasiyet değerinin hesaplanma formülü Denklem 3.7’de verilmektedir.

$$hassasiyet = \frac{\text{gerçek pozitif (TP)}}{\text{gerçek pozitif (TP)} + \text{yanlış pozitif (FP)}} \quad (3.7)$$

Duyarlılık (Recall) ise doğru olarak pozitif sınıflandırılan örneklerin tüm gerçeklere oranı ile hesaplanmaktadır. Duyarlılık değerinin hesaplanma formülü Denklem 3.8’de verilmektedir.

$$duyarlılık = \frac{\text{gerçek pozitif (TP)}}{\text{gerçek pozitif (TP)} + \text{gerçek negatif (FN)}} \quad (3.8)$$

F1 puanını kesinlik ve duyarlılık değerleri ile Denklem 3.9’da verildiği gibi hesaplanmaktadır. F1 puanı hassasiyet ve duyarlılığın ortalamasıdır.

$$F1 = 2 \times \left(\frac{\text{hassasiyet} \times \text{duyarlılık}}{\text{hassasiyet} + \text{duyarlılık}} \right) \quad (3.9)$$

Doğruluk (Accuracy) değeri Denklem 3.10’da verildiği gibi hesaplanmaktadır. Doğruluk değeri, doğru tahmin edilen etiketler ile veri kümesindeki tüm örnekler arasındaki orantıdır.

$$\text{doğruluk} = \frac{\text{gerçek pozitif (TP)} + \text{gerçek negatif (FN)}}{\text{gerçek pozitif (TP)} + \text{yanlış negatif (FN)} + \text{gerçek negatif (FN)} + \text{yanlış negatif (FN)}} \quad (3.10)$$

4. ARAŞTIRMA BULGULARI VE TARTIŞMA

Bu tez çalışması kapsamında siber zorbalık önleme modeli için, siber zorbalık cümlelerinden oluşan bir veri seti üzerinden Google firmasına ait metni doğal sesli konuşmaya dönüştüren uygulaması ile metinler ses verilerine dönüştürülmüştür. Dönüştürülen ses verileri, Python programlama dilinde, konuşma tanıma kütüphanesi kullanılarak metin verilerine dönüştürülmüştür. Bu kapsamda, öncelikli olarak mikrofondan elde edilen konuşma verilerinin metin verilerine dönüştürülmesi sağlanmıştır. Her cümlede yer alan kelimeleri temsil edebilmek ve işlemek için iki boyutlu tensör yapısı oluşturulmuştur. Veri seti İngilizce dilinde hazırlanmış olup, 17068 sosyal medya paylaşım metinlerinden oluşmaktadır. Bu giriş verileri ile BERT algoritmasının hassas eğitimi gerçekleştirilmiştir. Sonrasında BERT algoritması tarafından sınıflandırma analizi yapılmıştır.

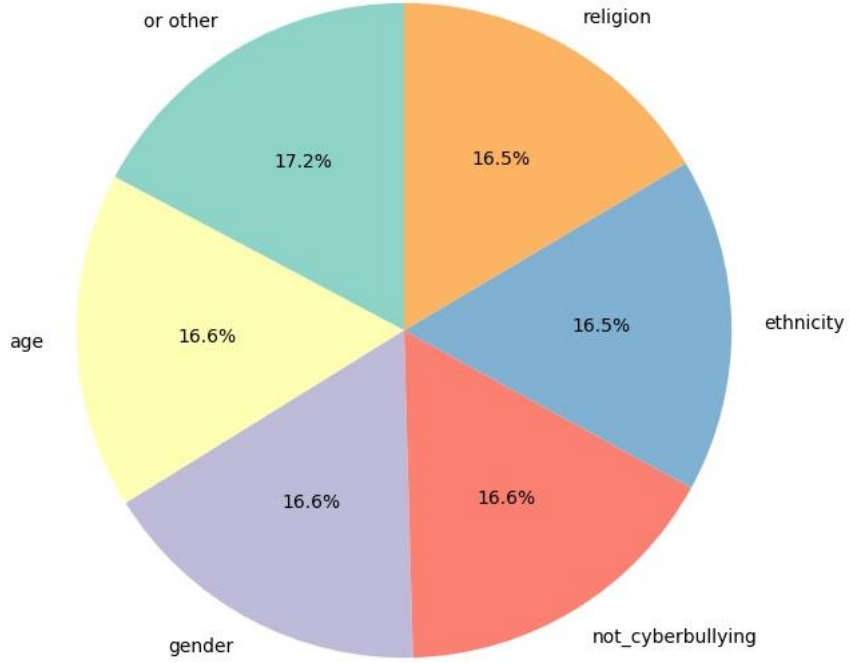
4.1. Eğitim Verilerinin Ön İşlemesi

Veriler, yaş, din, cinsiyet, etnik köken, siber zorbalık içermeyen ve diğer olarak 6 kategoriye ayrılmıştır (J. Wang, Fu ve Lu, 2020). Çizelge 4.1. ve Şekil 4.1' de veri setinde bulunan kategoriler ve veri sayıları yer almaktadır. Her bir kategoride dengeli olarak birbirine yakın veri sayısı bulunmaktadır.

Çizelge 4.1. Veri setinde bulunan kategoriler ve veri sayıları

Kategori	Veri Sayısı
age	7948
ethnicity	7903
gender	7941
not_cyberbullying	7939
or other	8251
religion	7884

Kategori Dağılımı



Şekil 4.1. Veri setinde bulunan kategorilerin dağılımı

Bu veri setinde bulunan veriler aynı zamanda siber zorbalığı da kategorilere ayrılmıştır. Çizelge 4.2. Oluşturulan veri setinde bulunan kategorilere göre cümleler den örnekler verilmektedir.

Çizelge 4.2. Oluşturulan veri seti kategorileri ve cümle örnekleri

Kategori	Cümleler
not_cyberbullying	it helps remembering how many e...
not_cyberbullying	I can t believe I m using this word without ir...
age	Not mine but this one girls who dated
or other	WHERE THE BLOODY HELL IS BALOR?!!! #RAWafterm...
religion	So a random Iowa fan pulls up in his car and s...
religion	Get that ass tapped..... Slowly RT @tayyoung_:...
ethnicity	ur the first one ive seen that doesnt like yuk...
gender	It is fitting that we are now seeing these str...
or other	@Kryten2X4B1Who does n t? other_cyberbullyi..

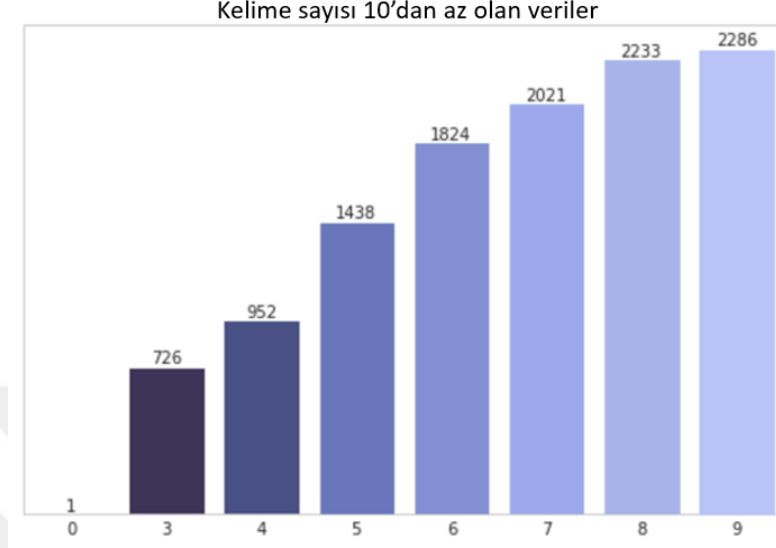
Elde edilen veri setinde metin analizi doğruluğunu artırmak ve daha iyi arama sonuçları elde edebilmek için metin küçük parçalara ayrılmış, metin temizleme (text cleaning) ve normalleştirme (text normalization) gerçekleştirilmiştir. Metin temizleme ile verilerde bulunan noktalama işaretleri ve özel karakterler (örneğin '\$', '£', '&' vb.) temizlenmiştir. Metin normalizasyonu ile ise metinlerde bulunan duygu ifadeleri (emoji), bağlantı adresleri, kullanıcı isimleri, ASCII olmayan karakterler, satır sonu karakterler, metin içerisinde bulunan birden fazla boşluklar, metinde bulunan sayılar, tekrar eden kelimeler, arka arkaya gelen noktalama işaretleri temizlenerek modelin eğitimi için daha sade bir veri seti elde edilmiştir. Gövdeleme (stemming) işlemi gerçekleştirilerek kelimelerde bulunan ekler temizlenmiştir. Lematisasyon (lemmatization) yöntemi ile kelimelerin dilbilgisel ve anlamsal özellikleri dikkate alınarak kelime kökü bulma işlemi gerçekleştirilmiştir (örneğin 'running' kelimesi 'run' olarak güncellenmiştir). Verilerde bulunan 'with', 'you', 'or' gibi önemsiz olan durak kelimeleri (stop words) temizlenerek veriler daha yalın hale getirilmiştir.

Ayrıca benzer anlamlara sahip kelimeler tek bir kelime ile ilişkilendirilmiştir. Çizelge 4.3'te veri setinde bulunan verilerin temizleme işlemi sonrasındaki veri örnekleri verilmektedir. Eğitim verilerinde bulunan etiketli, ifade içeren veya tekrarlayan kelimeler ise teke düşürülmüştür. Bu sayede eğitim verileri doğal dil işleme modeli için daha uygun hale getirilmiştir.

Çizelge 4.3. Verilerin temizleme işlemi sonrasındaki veri örnekleri

Veri	Temizlenmiş Veri
In other words #katandandre, your food was cra...	word katandandre food crapilicious mkr
Why is #aussietv so white? #MKR #theblock #ImA...	aussietv white mkr theblock imacelebrityau tod...
@XochitlSuckkks a classy whore? Or more red velvet cup...	classy whore red velvet cupcake
@Jason_Gio meh. :P thanks for the heads up, b...	meh thanks head concerned another angry dude t...
@RudhoeEnglish This is an ISIS account pretend...	isi account pretending kurdish account like is...

Şekil 4.2’de veri setinde bulunan ve 10 kelimedenden az sayıda kelimeye sahip olan verilerin grafiği verilmiştir. Veri setinde bulunan verilerde kelime sayısı 3 ve altında olan veriler, veri setinden çıkarılmıştır.



Şekil 4.2. 10 kelimedenden az sayıda kelime içeren veriler

Şekil 4.3’te veriler içerisinde 10 kelimedenden fazla kelimeye sahip olan verilerin sayı grafiği verilmiştir. 100 kelimedenden fazla kelimeye sahip çok uzun olan veriler de veri setinden çıkarılmış ve en uzun veri uzunluğu 30 kelime olarak elde edilmiştir.



Şekil 4.3. 10 kelimedenden fazla sayıda kelime içeren veriler

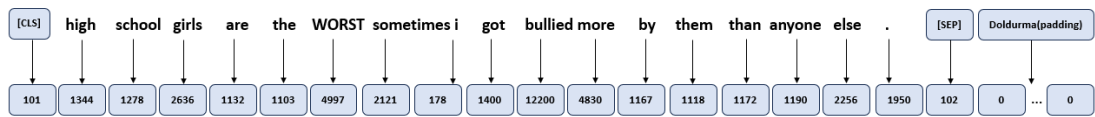
Elektronik tablo olarak kaydedilen veriler csv dosyasına dönüştürülerek BERT sınıflandırmasına uygun hale getirilmiştir. Alınan bu veri setindeki cümleler

seslendirilerek oluşturulan veri seti ile denk hale getirilmiştir. Böylece siber zorbalık değerlerinin doğruluğu daha kolay bir şekilde tespit edilmiştir.

4.2. BERT ile Eğitim Sürecinin Gerçekleştirilmesi

BERT, metin içerisindeki kelimeleri ve kelime içerisindeki gövde, kök, ek parçalarını tanımlamak amacıyla küçük parçalara bölmektedir. Gövdeleme işlemi için küçük parçalara ayrılan kelimeler, sabit bir uzunluktaki dizi haline getirilmektedir. Girdi metni, modelin kapasitesine göre, sabit boyutlu vektörler haline dönüştürülmektedir. Bu vektörler, kelimenin anlamını ve bağlamını yansıtmaktadır. BERT, kelimelerin metin içindeki sıralamasını anlamak için pozisyon vektörlerini kullanmaktadır. Bu vektörler, modelin kelimenin hangi konumda olduğunu anlamasını sağlamaktadır.

Şekil 4.4'te veri setinden alınmış örnek bir cümle belirteçlerine ayrılması verilmektedir. Veri setinde bulunan orijinal cümle “high school girls are the WORST sometimes i got bullied more by them than anyone else.” şeklindedir. BERT modeli küçük parçalara ayırma işlemini gerçekleştirirken her bir kelime ve noktalama için belirteç kimliklerini (token id) kullanmaktadır. Cümle başında [CLS] etiketi ve etiketin belirteç numarası 101 olarak verilmektedir. Diğer kelimeler için BERT modelinde sırası ile ([101, 1344, 1278, 2636, 1132, 1103, 4997, 2121, 178, 1400, 12200, 4830, 1167, 1118, 1172, 1190, 2256, 1950, 102, 0,...]) belirteç kimlikleri atanmıştır. Cümle sonunu belirtmek için [SEP] etiketi atanmıştır. Sabit boyutlu bir dizi elde edebilmek için doldurma (padding) yöntemi ile cümle sonu etiketinden sonra belirteç numaraları 0 olarak atanmaktadır. Maksimum dizi boyutu 80 olarak ayarlanmıştır.



Şekil 4.4. Örnek veri cümlesinin kelime parçalarına ayrılması

Eğitim sürecinde, BERT Base cased modeli kullanılmıştır. Model 110 milyon parametre ile önceden eğitilmiş bir modeldir. Büyük küçük harfe duyarlıdır. Eğitim verilerinin içerisinde büyük harfli veya küçük harfli kelimeler belirteç kimliklerinde farklı değerlere sahiptir. Çizelge 4.4'te hassas eğitim parametreleri verilmektedir.

Model dizini olarak ön bellek kullanılmaktadır. Optimizasyon algoritması olarak AdamW kullanılmıştır. Bu optimizasyon algoritması ağırlıkların ayarlanması ve aşırı öğrenmenin önlenmesi için ağırlık azalması (weight decay) tekniğini kullanmaktadır.

Çizelge 4.4. BERT algoritması hassas eğitim parametreleri

Model Parametreleri	Parametre Değerleri
Modelin Adı	“bert-base-uncased”
Model Dizini	Ön bellek
Optimizer	Adamw
Adamw Öğrenme Oranı	$2 \cdot 10^{-5}$
Adamw Epsilon	10^{-8}
İterasyon sayısı	8

Eğitim sürecinde öğrenme hızını ayarlayabilmek için PyTorch ve Transformers kütüphanelerinde kullanılan öğrenme hızı planlayıcısı kullanılmıştır. Bu planlayıcı ile eğitim sürecinde öğrenme hızını ayarlayabilmek için iki aşama gerçekleştirilir. İlk aşama eğitim sürecinin başında öğrenme hızının kademeli olarak artırıldığı ısınma (warmup) aşamasıdır. Isınma aşaması tamamlandıktan sonra, öğrenme hızının doğrusal olarak azaltıldığı doğrusal azalma (linear decay) aşaması gerçekleşmektedir. Isınma adımı 0’dan başlatılmıştır. Buradaki amaç modelin öğrenme hızını başlangıçta stabil olarak öğrenmesini sağlarken daha sonrasında öğrenme hızı azaltılarak modelin aşırı öğrenmesinin önüne geçilmesidir. Fonksiyon parametreleri, optimizer, num_warmup_steps, num_training_steps, last_epoch şeklindedir. Bu parametrelerden optimizer modelin öğrenme oranını güncellemek için, num_warmup_steps modelin ısınma adımlarının sayısını belirlemek için, num_training_steps parametresi toplam eğitim adımları sayısını belirlemek için ve last_epoch parametresi de son eğitim adımının indeksi için kullanılmaktadır.

Sınıflandırılmanın gerçekleştirilmesi için Google Colaboratory ortamında, Python yazılım dili kullanılmıştır. Veri setinin sınıflandırılması için BERT-Base modeli tercih edilmiştir. Aynı zamanda dizi sınıflandırma modeli kullanılmıştır. BERT modeli için önceden eğitilmiş bir dil modelini içeren transformers kütüphanesi kullanılmıştır. Önceden eğitilmiş BERT-tokenizer’ı ile metin verisi modelin anlayabileceği form olan belirteç (token) dizilerine dönüştürülmüştür.

SequenceClassification sınıfı kullanılarak BERT modeli sınıflandırma için hazırlanmıştır. Bu modelde siber zorbalık var/siber zorbalık yok şeklinde sonuç elde ettiğimiz iki sınıflı bir sınıflandırma yapılmıştır. Model metin verisini alarak iki sınıftan birine atama yapmaktadır. Modelde cümlelerin 0-1 aralığında ağırlıkları alınarak bir etiket ataması yapılmıştır. Bu etiket atamasından elde edilen sonuçlar etiket alanı tahmin edilen etiket olarak güncellenerek yeni bir dosyaya aktarılmıştır.

4.3. Konuşma Tanıma ile Giriş Verilerinin Elde Edilmesi

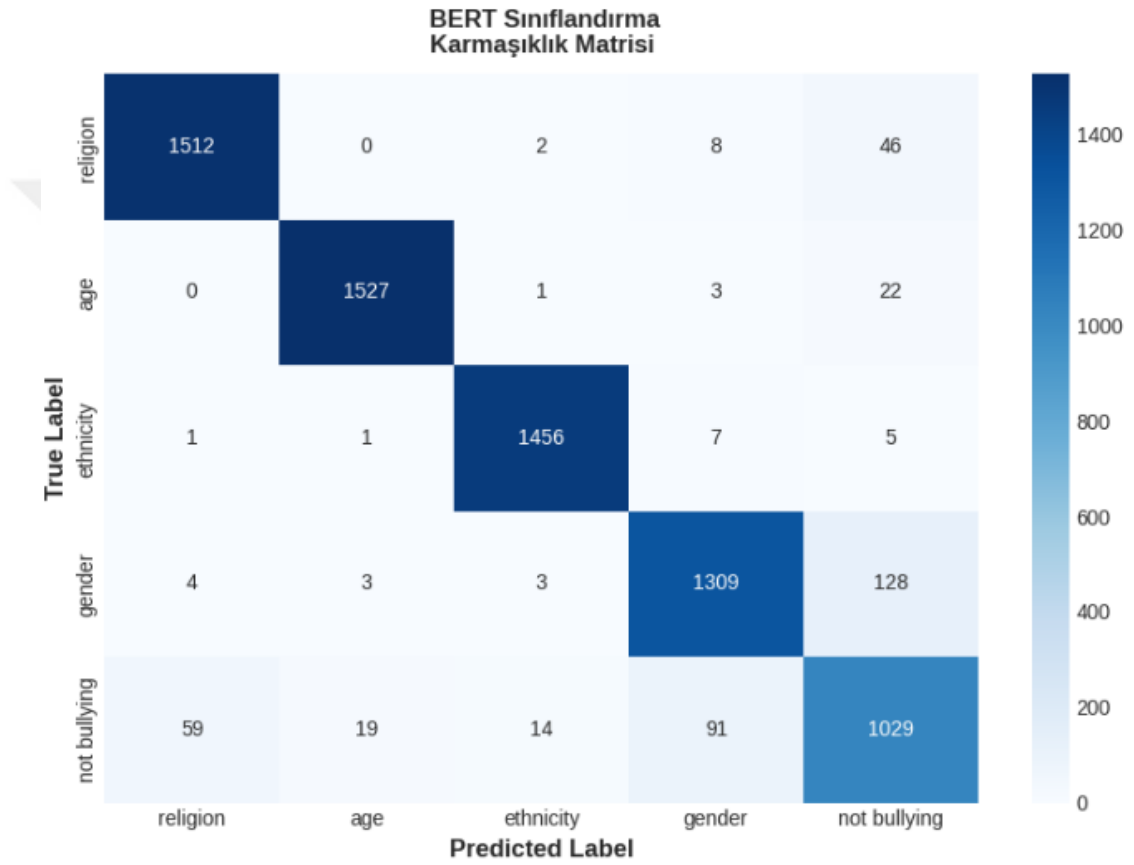
Python programlama dilinde konuşma tanıma kütüphanesi ile ses verilerinden metin verileri elde edilmiştir. Bu kütüphanenin çalışabilmesi için bazı gereksinimlere ihtiyaç duyulmuştur. Python sürümü olarak 3.8 ve üstü sürümler gerekmektedir. Mikrofon girişi için PyAudio 0.2.11 ve üstü gerekmektedir. Aşağıda konuşma tanıma kütüphanesi kullanılarak oluşturulmuş yazılımın sözde kodu verilmiştir.

- Adım 1:* “Başla
- Adım 2:* Gerekli kütüphaneyi içe aktar, ``speech_recognition``
- Adım 3:* Ses tanıma nesnesi oluştur
- Adım 4:* Mikrofonu kullanarak ses kaydet
- Adım 5:* Mikrofonu aç
- Adım 6:* Kullanıcıya konuşmasını söyle
- Adım 7:* Ses kaydını al ve ``audio`` olarak sakla
- Adım 8:* Mikrofonu kapat
- Adım 9:* Alınan sesi tanı
- Adım 10:* Google hizmeti ile sesi metne çevir
- Adım 11:* Tanıma sonucunu yazdır
- Adım 12:* Eğer tanıma başarısızsa, "Ses anlaşılamadı" yazdır
- Adım 13:* Eğer servis erişilemiyorsa, "Servise erişilemiyor" yazdır”
- Adım 14:* Bitir”

Konuşma tanıma aracılığı ile elde edilen ses verileri metin verileri haline dönüştürülmüştür. Bu metin verileri bir elektronik tablo olarak kayıt altına alınmıştır. Böylece test verisi oluşturulmuştur.

4.3.1. Test sonuçları

Şekil 4.5'te karmaşıklık matrisi verilmektedir. Karmaşıklık matrisine bakıldığında modelin özellikle etnik köken kategorisindeki sınıflandırmada oldukça başarılı olduğu görülmektedir. Ancak siber zorbalık olmayan cümle kategorisinde sınıflandırmada zorlandığı görülmektedir.



Şekil 4.5. Sınıflandırma sonucu karmaşıklık matrisi

Çizelge 4.5'te modelin kategorilere göre performans değerlendirme kriterleri olarak modelin hassasiyeti, duyarlılığı ve F1 skoru verilmektedir. Siber zorbalık içeren verileri kategorilere göre sınıflandırmada model başarısı oldukça yüksek olduğu görülmektedir.

Çizelge 4.5. Modelin kategoriler göre performans değerleri

	Hassasiyet	Duyarlılık	F1- skoru	Veri Sayısı
Religion	0.96	0.96	0.96	1568
Age	0.99	0.98	0.98	1553
Ethnicity	0.99	0.99	0.99	1470
Gender	0.92	0.90	0.91	1447
Not bullying	0.84	0.85	0.84	1212

Modelin F1 skoru yaklaşık %96 olarak, doğruluk değeri ise yaklaşık %94 olarak hesaplanmıştır. Modelin genel doğruluğu, kesinlik değeri, duyarlılık değeri ve F1 skoru göz önünde bulundurulduğunda beklenildiği gibi siber zorbalık tespitini başarı ile gerçekleştirmiştir.

5. SONUÇ VE ÖNERİLER

Bu tez kapsamında ortam dinlemesi ile ses verilerinden metin verileri elde edilmiştir. Ses dinleme teknolojileri ile ortamdaki konuşmaların anlık olarak kaydedilmesi ve bu verilerin metne dönüştürülmesiyle aynı zamanda dinamik bir veri setine sahip olunmuştur. Elde edilen veri seti metin temizleme, metin normalizasyonu, gövdeleme, baş sözcük çıkarma, önemsiz sözcük çıkarma, leminizasyon gibi yöntemlerle doğal dil işleme için daha uygun hale getirilmiştir. Veriler aynı zamanda, yaş, din, cinsiyet, etnik köken, siber zorbalık içermeyen ve diğer olmak üzere 6 farklı kategoriye ayrılmıştır. Bu kategorilere göre değerlendirme kriterlerine tabii tutulmuştur.

Elde edilen verilerin sınıflandırılmasında BERT algoritması kullanılmıştır. BERT algoritmasının kullanımının en büyük avantajı modeldeki dönüştürücü ile hem önceki hem sonraki kelimelerin kontrolü ve öz dikkat yapısı ile elde edilen dikkat kelimeleridir. BERT ile alınan veri küçük parçalara bölme işlemi ile belirteçlere ayrılmakta ve bu adımda dikkat noktaları belirlenmektedir. Küçük parçalara ayırma işleminin ardından BERT modeli ile metin yerleştirmeleri ve sınıflandırma yapılmıştır.

Bu tezde yapılan çalışmalarda, BERT algoritmasının siber zorbalık tespiti ve sınıflandırması için olumlu yönde sonuçlar alınmıştır. Bu sonuçların mobil uygulama geliştirmede kullanılabilir hale getirilmesi öngörülmüştür. Siber zorbalığın tespiti, bireyler üzerindeki olumsuz etkilerin azaltılması açısından önemli bir noktadır. İleriki çalışmalarda yapılacak olan yazılımlarla bu çalışma mobil uygulamalar ile günlük hayata entegre edilebilir ve anlık olarak yapılacak müdahalelerle bireyler için siber zorbalığa maruz kalındığında oluşabilecek olumsuz sonuçların önüne geçilebilecektir.

Bu çalışmada kullanılan veri setinin İngilizce olmasının yanı sıra diğer dillerde de geliştirilebilecek veri setleri ile Türkçe, Fransızca gibi dillere de uygulanabilir. Aynı zamanda dönüştürücülerin çeviri yeteneği kullanılarak dilden dile bir siber zorbalık tespit ve önleme uygulaması gerçekleştirilebilir.

KAYNAKLAR

- Abraham, A., Castillo, O. ve Virmani, D. (2020). Proceedings of 3rd International Conference on Computing Informatics and Networks.
- Acheampong, F. A., Nunoo-Mensah, H. ve Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, 54(8), 5789-5829. doi:10.1007/s10462-021-09958-2
- Agatston, P. W., Kowalski, R. ve Limber, S. (2007). Students' Perspectives on Cyber Bullying. *Journal of Adolescent Health*. Elsevier USA. doi:10.1016/j.jadohealth.2007.09.003
- Ahmed, M. F., Mahmud, Z., Biash, Z. T., Ann, A., Ryen, N., Hossain, A. ve Ashraf, F. Bin. *Cyberbullying Detection Using Deep Neural Network from Social Media Comments in Bangla Language*.
- Akca, E. B., Sayımer, İ., Salı, J. B. ve Başak, B. E. (2014). Okulda Siber Zorbalığın Nedenleri, Türleri ve Medya Okuryazarlığı Eğitiminin Önleyici Çalışmalardaki Yeri. *Elektronik Mesleki Gelişim ve Araştırma Dergisi (EJOIR)*, 2, 17-30.
- Al-Harigy, L. M., Al-Nuaim, H. A., Moradpoor, N. ve Tan, Z. (2022). Building towards Automated Cyberbullying Detection: A Comparative Analysis. *Computational Intelligence and Neuroscience*. Hindawi Limited.
- Badjatiya, P., Gupta, S., Gupta, M. ve Varma, V. (2017). Deep learning for hate speech detection in tweets. *26th International World Wide Web Conference 2017, WWW 2017 Companion* içinde (ss. 759-760). International World Wide Web Conferences Steering Committee.
- Bahdanau, D., Cho, K. ve Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate.
- Balakrishnan, V., Khan, S. ve Arabnia, H. R. (2020). Improving cyberbullying detection using Twitter users' psychological features and machine learning. *Computers and Security*, 90.
- Behzadi, M., Harris, I. G. ve Derakhshan, A. (2021). Rapid Cyber-bullying detection method using Compact BERT Models. *Proceedings - 2021 IEEE 15th International Conference on Semantic Computing, ICSC 2021* içinde (ss. 199-202). Institute of Electrical and Electronics Engineers Inc.
- Berglund, M., Raiko, T., Honkala, M., Kärkkäinen, L., Vetek, A. ve Karhunen, J. *Bidirectional Recurrent Neural Networks as Generative Models*.

- Betts, L. R., Spenser, K. A. ve Baguley, T. (2022). Describing Disclosure of Cybervictimization in Adolescents from the United Kingdom: The Role of Age, Gender, Involvement in Cyberbullying, and Time Spent Online. *Journal of Genetic Psychology*, 183(1), 40-53.
- Bhardwaj, V., Kukreja, V., Othman, M. T. Ben, Belkhier, Y., Bajaj, M., Goud, B. S., ... Hamam, H. (2022, 1 Mayıs). Automatic Speech Recognition (ASR) Systems for Children: A Systematic Literature Review. *Applied Sciences (Switzerland)*. MDPI.
- Bobrow, D. G. ve Woods, W. A. *Computational Linguistics Transition Network Grammars for Natural Language Analysis*.
- Bonanno, R. A. ve Hymel, S. (2013). Cyber Bullying and Internalizing Difficulties: Above and Beyond the Impact of Traditional Forms of Bullying. *Journal of Youth and Adolescence*, 42(5), 685-697.
- Brewer, G. ve Kerslake, J. (2015). Cyberbullying, self-esteem, empathy and loneliness. *Computers in Human Behavior*, 48, 255-260.
- Buduma, Nikhil. ve Locascio, Nicholas. (2017). *Fundamentals of deep learning : designing next-generation machine intelligence algorithms*. O'Reilly Media.
- Ceylan, L., Durak, A., Durak, B. ve Sümbül, Ü. M. Eğitim Sürecinde Ergen Kişilerin Siber Zorbalık Durumları Üzerine Yapılan Tez Çalışmalarının Üzerine Bir Derleme Çalışması.
- Chicco, D. ve Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1).
- Chomsky, noam. (2002). *Syntactic Structures*.
- Devlin, J., Chang, M.-W., Lee, K. ve Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
- Diniz, A. P. M., Ciarelli, P. M., Salles, E. O. T. ve Coco, K. F. (2024, 15 Mart). Use of deep neural networks for clogging detection in the Submerged Entry Nozzle of the continuous casting. *Expert Systems with Applications*. Elsevier Ltd.
- Elsafoury, F., Katsigiannis, S., Wilson, S. R. ve Ramzan, N. (2021). Does BERT Pay Attention to Cyberbullying? *SIGIR 2021 - Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* içinde (ss. 1900-1904). Association for Computing Machinery, Inc.

- Erdur-Baker, O. (2007). *Akran Zorbalığının Yeni Yüzü: Siber Zorbalık Cyber Bullying: A New Face of Peer Bullying*.
- Feng, W., Guan, N., Li, Y., Zhang, X. ve Luo, Z. (2017). Audio visual speech recognition with multimodal recurrent neural networks. *Proceedings of the International Joint Conference on Neural Networks* içinde (C. 2017-May, ss. 681-688). Institute of Electrical and Electronics Engineers Inc.
- Founta, A.-M., Djouvas, C., Chatzakou, D., Leontiadis, I., Blackburn, J., Stringhini, G., ... Kourtellis, N. (2018). *Large Scale Crowdsourcing and Characterization of Twitter Abusive Behavior*. www.aaai.org adresinden erişildi.
- Gehring, J., Auli, M., Grangier, D., Yarats, D. ve Dauphin, Y. N. (2017). Convolutional Sequence to Sequence Learning.
- Guo, X., Anjum, U. ve Zhan, J. (2022). Cyberbully Detection Using BERT with Augmented Texts. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022* içinde (ss. 1246-1253). Institute of Electrical and Electronics Engineers Inc.
- Hellfeldt, K., López-Romero, L. ve Andershed, H. (2020). Cyberbullying and psychological well-being in young adolescence: the potential protective mediation effects of social support from family, friends, and teachers. *International Journal of Environmental Research and Public Health*, 17(1).
- Hernández, A. ve Amigó, J. M. (2021). Attention mechanisms and their applications to complex systems. *Entropy*, 23(3), 1-18.
- Hochreiter, S. ve Urgan Schmidhuber, J. ". (1997). *Long Short-Term Memory*.
- Hutchins, W. J. (2004). *LNAI 3265 - The Georgetown-IBM Experiment Demonstrated in January 1954*. LNAI (C. 3265).
- Iqbal, T. ve Qureshi, S. (2022, 1 Haziran). The survey: Text generation models in deep learning. *Journal of King Saud University - Computer and Information Sciences*. King Saud bin Abdulaziz University.
- Islam, M. M., Uddin, M. A., Islam, L., Akter, A., Sharmin, S. ve Acharjee, U. K. (2020). Cyberbullying Detection on Social Networks Using Machine Learning Approaches. *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, CSDE 2020* içinde . Institute of Electrical and Electronics Engineers Inc.
- Jean Grosz, B.-B. A. (1969). *The Representation and Use of Focus in Dialogue Understanding*

- Khadhraoui, M., Bellaaj, H., Ammar, M. Ben, Hamam, H. ve Jmaiel, M. (2022). Survey of BERT-Base Models for Scientific Text Classification: COVID-19 Case Study. *Applied Sciences (Switzerland)*, 12(6).
- KOÇAK, Ç. ve YİĞİT, T. (2023). Gpt-3 Sınıflandırma Modeli İle Türkçe Twitlerin Siber Zorbalık Durumlarının Belirlenmesi. *Gazi Journal of Engineering Sciences*, 9(4), 278-285.
- Kowalski, R. M. ve Limber, S. P. (2007). Electronic Bullying Among Middle School Students. *Journal of Adolescent Health*, 41(6 SUPPL.).
- Krstinić, D., Braović, M., Šerić, L. ve Božić-Štulić, D. (2020). Multi-label Classifier Performance Evaluation with Confusion Matrix (ss. 01-14). Academy and Industry Research Collaboration Center (AIRCC).
- Lauriola, I., Lavelli, A. ve Aiolli, F. (2022). An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing*, 470, 443-456.
- Le, X. H., Ho, H. V., Lee, G. ve Jung, S. (2019). Application of Long Short-Term Memory (LSTM) neural network for flood forecasting. *Water (Switzerland)*, 11(7).
- Lecun, Y., Bengio, Y. ve Hinton, G. (2015, 27 Mayıs). Deep learning. *Nature*. Nature Publishing Group. doi:10.1038/nature14539
- Limo, S. (2015). *BULLYING AMONG TEENAGERS AND ITS EFFECTS*.
- Malik, M., Malik, M. K., Mehmood, K. ve Makhdoom, I. (2021). Automatic speech recognition: a survey. *Multimedia Tools and Applications*, 80(6), 9411-9457.
- Maurya, C., Muhammad, T., Dhillon, P. ve Maurya, P. (2022). The effects of cyberbullying victimization on depression and suicidal ideation among adolescents and young adults: a three year cohort study from India. *BMC Psychiatry*, 22(1).
- McCarthy, J. (1960). *Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I*.
- Mikhaylovsky, M. N., Lopatkova, I. V., Komarova, N. M., Rueva, E. O., Tereschuk, K. S. ve Emelyanenkova, A. V. (2019). Cyberbullying as a new form of a threat: A physiological, psychological and medicinal aspects. *Electronic Journal of General Medicine*, 16(6).
- Muneer, A. ve Fati, S. M. (2020). A comparative analysis of machine learning techniques for cyberbullying detection on twitter. *Future Internet*, 12(11), 1-21.

- Niu, Z., Zhong, G. ve Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, 452, 48-62. doi:10.1016/j.neucom.2021.03.091
- Patchin, J. W. ve Hinduja, S. (2006). Bullies Move Beyond the Schoolyard: A Preliminary Look at Cyberbullying. *Youth Violence and Juvenile Justice*, 4(2), 148-169.
- Patwardhan, N., Marrone, S. ve Sansone, C. (2023, 1 Nisan). Transformers in the Real World: A Survey on NLP Applications. *Information (Switzerland)*. MDPI.
- Paul, S. ve Saha, S. (2022). CyberBERT: BERT for cyberbullying identification: BERT for cyberbullying identification. *Multimedia Systems* içinde (C. 28, ss. 1897-1904). Springer Science and Business Media Deutschland GmbH.
- Peled, Y. (2019). Cyberbullying and its influence on academic, social, and emotional development of undergraduate students. *Heliyon*, 5, 1393.
- Perren, S., Dooley, J., Shaw, T. ve Cross, D. (2010). Bullying in school and cyberspace: Associations with depressive symptoms in Swiss and Australian adolescents. *Child and Adolescent Psychiatry and Mental Health*, 4.
- Pinar Saygin, A., Cicekli, I. ve Akman, V. (2001). *Turing Test: 50 Years Later*.
- Reddy, D. R. (1976). Speech Recognition by Machine: A Review. *PROCEEDINGS OF THE IEEE*
- Saini, H., Mehra, H., Rani, R., Jaiswal, G., Sharma, A. ve Dev, A. (2024, 1 Aralık). Enhancing cyberbullying detection: a comparative study of ensemble CNN–SVM and BERT models. *Social Network Analysis and Mining*. Springer.
- Sanders, B. *ANALYSIS OF A PRE-TRAINED BERT MODEL USED FOR CYBERBULLYING DETECTION IN TWEETS*.
- Schuster, M. ve Paliwal, K. K. (1997). *Bidirectional Recurrent Neural Networks*. *IEEE TRANSACTIONS ON SIGNAL PROCESSING* (C. 45).
- Sousa, M. G., Sakiyama, K., Rodrigues, L. D. S., Moraes, P. H., Fernandes, E. R. ve Matsubara, E. T. (2019). BERT for stock market sentiment analysis. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI* içinde (C. 2019-November, ss. 1597-1601). IEEE Computer Society.
- Soydaner, D. (2022, 1 Ağustos). Attention mechanism in neural networks: where it comes and where it goes. *Neural Computing and Applications*. Springer Science and Business Media Deutschland GmbH.

- Subaramaniam, K., Kolandaisamy, R., Bin Jalil, A. ve Kolandaisamy, I., *Cyberbullying Challenges on Society: A Review. Journal of Positive School Psychology* (C. 2022).
- Sutskever, I., Vinyals, O. ve Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks.
- Tixier, A. J.-P. (2018). Notes on Deep Learning for NLP.
- Turing, A. M. (1950). *COMPUTING MACHINERY AND INTELLIGENCE. Computing Machinery and Intelligence. Mind* (C. 49).
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., ... Polosukhin, I. *Attention Is All You Need*.
- Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., ... Bowman, S. R. (2019). *SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems*.
- Wang, D., Wei, Y., Zhang, K., Ji, D. ve Wang, Y. (2022). Automatic Speech Recognition Performance Improvement for Mandarin Based on Optimizing Gain Control Strategy. *Sensors*, 22(8).
- Wang, J., Fu, K. ve Lu, C.-T. (2020). SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection. *2020 IEEE International Conference on Big Data (Big Data)* içinde (ss. 1699-1708). IEEE.
- Waseem, Z. ve Hovy, D. *Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter*.
- Wilks, Y. (1992). *THE PREFERENCE SEMANTICS FAMILY. ComputerJ Math. Applie* (C. 23).
- Xing, Z., Pei, J. ve Keogh, E. *A Brief Survey on Sequence Classification*.
- Yalçın, N. (2008). KONUŞMA TANIMA TEORİSİ VE TEKNİKLERİ 1. *Kastamonu Education Journal*, 16(1), 249-266.
- Yin, W., Kann, K., Yu, M. ve Schütze, H. (2017). Comparative Study of CNN and RNN for Natural Language Processing.
- Yurdakul, Y. ve Ayhan, A. B. (2023). The effect of the cyberbullying awareness program on adolescents' awareness of cyberbullying and their coping skills. *Current Psychology*, 42(28), 24208-24222.
- Zhang, X., Fan, J. ve Hei, M. (2022). Compressing BERT for Binary Text Classification via Adaptive Truncation before Fine-Tuning. *Applied Sciences (Switzerland)*, 12(23).