MEYAD akademi

Meyad Akademi, Cilt:4 Sayı:1 Yıl:2023 Sayfa:1-21

ARAŞTIRMA

YAPAY ZEKA SİBER ZORBALIĞI ÖNCEDEN TAHMİN EDEBİLİR Mİ?

Sümeyye KAVİCİ PORSUK*
Burak ŞİRİN**

ÖZ

Bu çalışma literatüre katkı sağlamak ve siber zorbalığın tespitinde kullanılan algoritmaların tanımlanması ve karşılaştırılma amacıyla yapılmıştır. Çalışmaya Türkçe ve İngilizce dillerinde, son 10 yıl içinde akademik dergilerde yayınlanmış olan, tam metnine ulaşılabilen araştırma makaleleri dahil edilmiştir. Literatür taraması Google Scholar, ProQuest, Science Direct, Scopus, Wiley Online Library ve Pubmed çevrimiçi veri tabanlarından Türkçe "Siber Zorbalık", "Tahmin Etme" ve "Yapay Zeka" ve İngilizce "Predicting Cyberbullying" ve "Artificial Intelligence" anahtar kelimeleri ile yapılmış ancak Türkçe anahtar kelimelerle herhangi bir sonuca ulaşılamadığından araştırmacıların ortak kararı ile yalnızca İngilizce anahtar kelimeler Ekim 2022'de yapılmıştır. Çalışmaya 19 araştırma makalesi dahil edilmiştir. Dahil edilen çalışmaların 18 tanesi İngilizce 1 tanesi Türkçedir. Çalışmaların amacı yapay zekaya dayanan algoritmalar yardımı ile siber zorbalığın tespit edilmesi ve/veya önlenmesidir. İncelenen çalışmalar sonucunda önerilen yapay zeka modelleri siber zorbalığa yönelik amaçlarını gerçekleştirme konusunda başarılı sonuçlar elde etmiştir. Çalışmalardan birinde siber zorbalığı mesaj içeriğinden tespit ederek mesajın gönderilmesini engelleyebilen bir yapay zeka modeli geliştirilmiştir. Ayrıca yapa zeka uygulamaları yalnızca çevrimiçi sosyal ağlarda değil, iş yerlerinde kurum içi kullanılan ağlarda da siber zorbalık başarılı bir şekilde tespit edilmiştir. Siber zorbaların sosyal medyada suçlayıcı bir dil kullanarak siber kurbanları istismar etmesi teorisi temelinde literatüre katkı sağlanmıştır. Siber zorbalıkla ilgili çalışmalar genellikle metin tabanlı analizleri içermektedir ancak siber zorbalığı daha iyi bir şekilde tespit edebilmek için resim, video ve sesleri analiz edebilen tekniklerin geliştirilmesi önerilmektedir. Uygulamalarda yapılabilecek ek güncellemeler programlar kullanıcıya yazması için herhangi bir kötüye kullanım içeriği önermeden önce gerçek zamanlı metin tahmini yapılacak şekilde ölçeklendirilebilir.

Anahtar kelimeler: Siber zorbalık, şiddeti önleme, yapay zeka.

* Arş. Gör., Tokat Gaziosmanpaşa Üniversitesi, E-mail: sumeyye.kavici@gop.edu.tr ORCID ID: 0000-0003-3579-8545

** Sorumlu Yazar, Arş. Gör., Tokat Gaziosmanpaşa Üniversitesi, E-mail: buraksirin33@gmail.com ORCID ID: 0000-0002-8485-5756 Geliş tarihi: 29.10.2022, Kabul tarihi: 19.02.2023

CAN ARTIFICIAL INTELLIGENCE PREDICT CYBERBULLYING?

ABSTRACT

This study was carried out to contribute to the literature and to define and compare the algorithms used in the detection of cyberbullying. Research articles published in academic journals in the last 10 years, in Turkish and English, whose full text can be accessed, were included in the study. The literature search was conducted with the keywords "Cyber Bullying", "Guess" and "Artificial Intelligence" in Turkish and "Predicting Cyberbullying" and "Artificial Intelligence" in English from Google Scholar, ProQuest, ScienceDirect, Scopus, Wiley Online Library and Pubmed online databases. Since no results could be reached with Turkish keywords, only English keywords were made in October 2022 with the joint decision of the researchers. 19 research articles were included in the study. Eighteen of the included studies are in English and one is in Turkish. The aim of the studies is to detect and/or prevent cyberbullying with the help of algorithms based on artificial intelligence. As a result of the studies examined, the proposed artificial intelligence models have achieved successful results in achieving their goals for cyberbullying. In one of the studies, an artificial intelligence model was developed that can detect cyberbullying from the message content and prevent the message from being sent. In addition, artificial intelligence applications have been able to successfully detect cyberbullying not only in online social networks, but also in networks used in-house at workplaces. A contribution has been made to the literature on the basis of the theory that cyberbullies abuse cyber victims by using an accusatory language in social media. Studies on cyberbullying usually include text-based analysis, but it is recommended to develop techniques that can analyze images, videos and sounds in order to better detect cyberbullying. Additional updates to apps can be scaled up to perform real-time text prediction before programs suggest any abusive content to the user for typing.

Keywords: Cyberbullying, preventing violence, artificial intelligence.

1. GİRİŞ

Dünyada ve ülkemizde her geçen gün kişilerarası şiddet artmaktadır. Literatüre göre kişilerarası şiddet dünyada, özellikle 15 ila 44 yaş arasındaki insanlar arasında, yaşam kalitesinde bozulma ve ölümlerin önde gelen nedenidir (1). Zorbalık da kişilerarası şiddet arasında yer almaktadır. Şiddet içerisinde kötüye kullanım sık görülmektedir. Sözlü veya duygusal, fiziksel, cinsel, dijital taciz, takip (çevrimiçi ve yüz yüze) ve ekonomik taciz gibi kamu güvenliğini riske atabilecek ve bireyin kişiliği ve saygınlığı üzerinde zararlı etkilere neden olabilecek pek çok biçimde olabilir (1).

Şiddetin bir çeşidi olan zorbalık daha çok genç bireyler arasında giderek artan bir sorun olarak yer almaktadır. Zorbalığı neyin oluşturduğuna dair kültürel bağlam ve anlayış dünya genelinde farklılık gösterebilir. Genellikle, sosyal, duygusal veya fiziksel yollarla bir zorbanın daha zayıf bir kurbanı kasıtlı olarak hedef aldığı saldırganlık yoluyla güç iddiasında bulunduğu bir kavramdır (2). Zorbalığın üç bileşeni bulunmaktadır; amaçlı saldırgan davranışlar, sistematik ve tekrarlı şekilde devam etmesi, zorba ve mağdur arasında güç dengesizliğine bağlı bir kişiler arası ilişki (1, 3). Zorbalık çok sayıda çocuklar başta olmak üzere çok sayıda bireyi etkiler ve psikolojik, fiziksel ve psikosomatik sonuçlar üzerinde uzun vadeli riskler oluşturabilmektedir. Zorbalık, dünya çapında okul çocukları için yaygın bir durumdur, tüm ülkelerde meydana gelir ve gençlerin %9 ila %54'ünü etkileyebilmektedir (1, 2).

Dijitalleşmenin artması ve teknolojik imkanların daha ulaşılabilir olmasıyla birlikte zorbalığın türleri de değişmiştir. Cep telefonu, bilgisayarlar ve tabletler gibi teknolojik cihazların artmasıyla yeni bir kişilerarası iletişim biçimi yaygınlaşmış ve genç nüfusun büyük bir çoğunluğu (%90) internet erişimine sahip olmuştur (3, 4). Yeni teknolojilerin zorbalık amacıyla kullanılması, elektronik medyanın göz korkutucu veya incitici mesajları iletmek için kullanıldığı ortaya çıkmıştır (4). Zorbalığın alanlarından olan siber zorbalık: bir grup veya birey tarafından kendini kolayca savunamayan bir kurbana karşı elektronik araçlar kullanılarak tekrar tekrar ve zaman içinde gerçekleştirilen saldırgan bir eylem veya davranış olarak tanımlanmaktadır (1, 5). Siber zorbalık şunları içermektedir: çevrimiçi alay etme/hakaret etme, çevrimiçi söylentileri yayma, özel bilgileri ifşa etme ve çevrimiçi gruplardan dışlama (6). Siber ve geleneksel zorbalık arasında açıkça benzerlikler olsa da, siber zorbalığın kimliğinin bilinmemesi gibi önemli farklılıklar da vardır; siber zorbalık her yerde ve her zaman olabilir; Siber zorbalık, geleneksel zorbalığa kıyasla daha hızlı yayılır ve daha geniş kitlelere ulaşabilir (6, 7).

Siber zorbalık tek bir yapı değildir birçok çeşidi bulunmaktadır. Siber zorbalık da teknoloji ile şekil değiştirmiştir. Öncesinde internet ve cep telefonu zorbalığı olarak çeşitlendirilen siber zorbalık şu an birçok kategoride

incelenmektedir. Bu kategoriler sürekli değişmekle birlikte fiziksel şiddet tehdidi, taciz veya nefretle ilgili tehditler, lakap takma (homofobi dahil), ölüm tehditleri, platonik ilişki tehditleri, cinsel eylemler, talepler/talimatlar, mevcut ilişkilere zarar verme tehditleri, ev/aile ile ilgili tehditler ve tehditkar zincir mesajları şeklinde sıralanabilir (4). Siber zorbalık davranışları üç farklı grupta incelenmiştir. Mağdurlar, siber zorbalar ve izleyicilerden oluşan bu grupta mağdurlar ve siber zorbalar tarafından en sık bildirilen davranış tehdit veya taciz, ardından şaka yapmak ve son olarak söylenti yaymak olmuştur. Seyirciler arasında ise bildirilen davranışların sırasında ise en sık şaka yapmak, ardından tehdit veya taciz etmek ve ardından söylenti yayılması yer almaktadır (6).

Siber zorbalığın en sık görüldüğü platformlar arasında Twitter, Facebook, YouTube, WhatsApp ve sosyal ağ platformları gibi çok anlık mesaj ve sosyal ağ içeren siteler bulunmaktadır. Siber zorbalığı tespit etmek ve önlemek amacıyla taciz edici veya hakaret içerikli içeriklerin tespiti için özel olarak tasarlanmış birçok farklı yazılım uygulamaları ve algoritma bulunmaktadır (5, 8). Eski ve mevcut algoritmalar, siber zorbalığı sınıflandırmak için eğitilmektedir. Bu eğitim makine öğrenmesi ile gerçekleşmektedir. Ancak eğitmek için çok miktarda veriye ihtiyaç duyulmaktadır ve en büyük dezavantajlardan biri, hangi senaryoda hangi algoritmayı kullanılmasında bir netliğin olmamasıdır (8). Siber taciz tespiti ile ilgili araştırmaların çoğu, metin kalıplarının madenciliğine dayanmaktadır. Örneğin, çevrimiçi ortamda başkalarını istismar etme eğilimi yüksek olan cinsel istismarcılar, kamuya açık veya özel sanal platformlarda başkalarını rahatsız edip etmediğinin tespiti, istenmeyen postaların tespiti, internet istismarcıların tespiti bunların arasında yer almaktadır (5, 8). Ancak algoritmaların farklı senaryolar altında (Sözlü, yazılı, görsel taciz) kullanımlarında netlik bulunmamaktadır. Bu nedenle verilerden taciz edici ve aşağılayıcı içerik bulma problemlerinin çoğunu çözebilecek bu tür algoritmaların tartışılması ve karşılaştırılması gerekmektedir. Bu çalışmanın amacı literatüre katkı sağlamak ve siber zorbalığın tespitinde kullanılan algoritmaların tanımlanması ve karşılaştırılmasıdır.

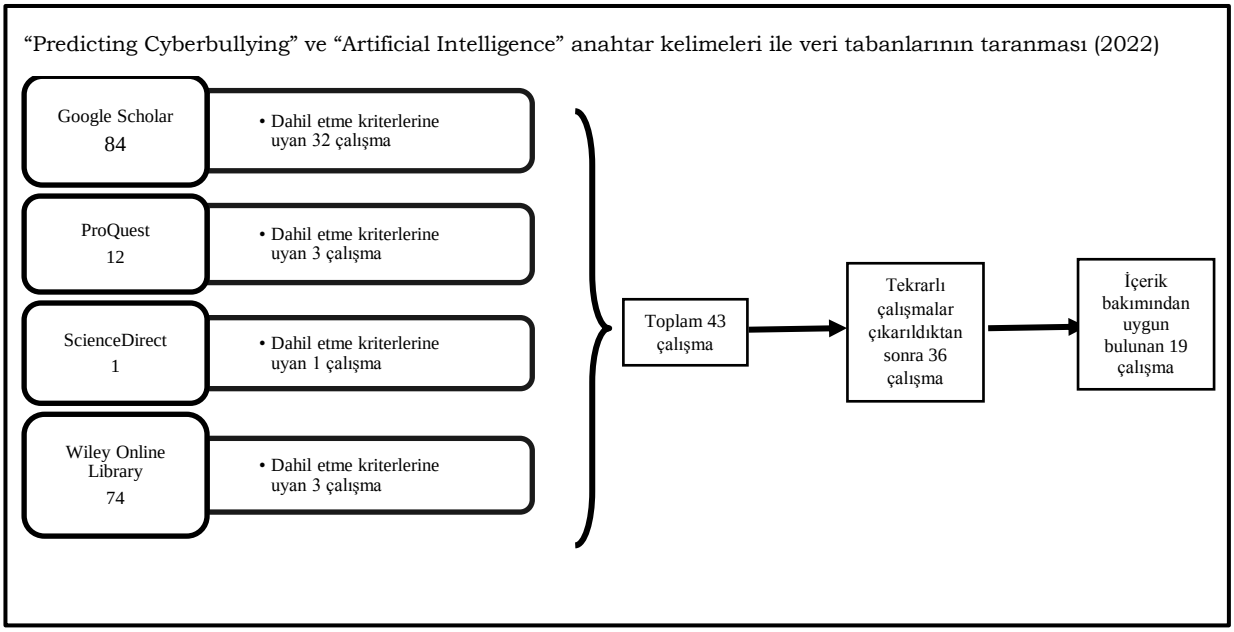
2. YÖNTEM

Yapay zeka kullanarak siber zorbalığı önceden tahmin etmek amacıyla geliştirilen uygulamaların teknik dilden arındırılarak derlenmesi amacıyla yapılan bu çalışmaya Türkçe ve İngilizce dillerinde, son 10 yıl içinde akademik dergilerde yayınlanmış olan, tam metnine ulaşılabilen araştırma makaleleri dahil edilmiştir. Literatür taraması Google Scholar, ProQuest, ScienceDirect, Scopus, Wiley Online Library ve Pubmed çevrimiçi veri tabanlarından Türkçe “Siber Zorbalık”, “Tahmin Etme” ve “Yapay Zeka” ve

İngilizce "Predicting Cyberbullying" ve "Artificial Intelligence" anahtar kelimeleri ile yapılmış ancak Türkçe anahtar kelimelerle herhangi bir sonuca ulaşılamadığından araştırmacıların ortak kararı ile yalnızca İngilizce anahtar kelimeler Ekim 2022'de yapılmıştır.

Yapılan bu işlemlerin ardından 43 araştırma makalesine ulaşılmıştır. Tekrarlı makaleleri çıkarılmasından sonra çalışma sayısı 36'ya indirilmiştir. Daha sonra bu çalışmalar araştırmacılar tarafından başlık, özet ve içerik bakımından değerlendirilerek 19 çalışma incelemeye alınmıştır (Şekil 1).

Yapılan çalışmada yapay zeka kullanarak siber zorbalığı önceden tahmin etmek amacıyla geliştirilen uygulamalara ilişkin, 2012-2022 yılları arasında yapılmış ve tam metnine ulaşılabilen Türkçe ve İngilizce dillerinde 19 araştırma makalesi incelenmiştir. Taramanın yapıldığı tarihlerde tam metnine ulaşılamayan makaleler çalışmaya dahil edilmemiştir. Araştırmacılar yalnızca Türkçe ve İngilizce dillerine hâkim olduğu için bu iki dil dışındaki dillerde yapılmış olan çalışmalar araştırmaya dahil edilmemiştir.



Şekil 1. Veri Tarama Akış Şeması

3. BULGULAR

Araştırmaya dahil edilen çalışmaların yalnızca biri Türkçe iken (9), geriye kalan 18 çalışma İngilizce'dir (8, 10-26). Çalışmaların dokuzu 2022 (9, 10, 12, 13, 15, 16, 19, 23, 25) ve altısı 2021 (8, 14, 17, 20, 24, 26) yıllarında yayınlanmıştır. İncelenen araştırmalar Tablo 1'de sunulmuştur.

3.1. Siber Zorbalıkta Yapay Zekanın Kullanım Amaçları

Siber zorbalık günümüzde sürekli artan bir suç olarak karşımıza çıkmaktadır (11). Çalışmalar incelendiğinde programların siber zorbalığı tespit etmek (8, 12-14, 17, 25), engellemek (11, 19), risk faktörlerini belirlemek (20), sınıflandırmak (10, 16) ve yeni çalışmalar için veri tabanı oluşturmak (15) üzere geliştirildiği fark edilmektedir. Özellikle siber zorbalığın en sık görüldüğü çevrimiçi sosyal ağların siber zorbalıktan arındırmak ve var olan siber zorbalık çeşitlerini sınıflandırmak için yapay zekanın kullanımı oldukça önem taşımaktadır (10). Çalışmalardan birisinde siber zorbalığı tespit etmek için çevrimiçi iletişimleri aracılığıyla sosyal medya kullanıcılarının kişilikleri ve duyguları incelenerek siber zorbalığı tespit etmek amaçlanmıştır (13). Bir diğer çalışmada ise işyerlerinde kurum içi kullanılan bir çevrimiçi ağda siber zorbalığın tespit edilmesi amaçlanmıştır (17).

3.2. Çalışmalarda Kullanılan Veri Tabanları

Çalışmaların dokuzunda veri tabanı olarak Twitter (11, 12, 14, 16, 18, 19, 21, 26), Facebook (11), Instagram (15), Wikipedia (14), Reddit (23) Kaggle (9) ve ASKfm (24) gibi sosyal medya siteleri kullanılmıştır. Çevrimiçi sosyal ağ siteleri her gün gelişmekte ve kişiler arasındaki mesafeyi ortadan kaldırdığından dolayı insanlar arasında yaygın olarak benimsenmektedir. Sosyal medyanın bu artan kullanımı, siber suçluların saldırılarına daha fazla bireyin maruz kalması ile sonuçlanmaktadır. Siber zorbalık sosyal medyayı eğlence, pazarlama, iletişim gibi sebepler için kullanan herkes için tehdit haline gelmektedir. Bu yüzden hızla büyüyen bu sorunu çözmek için mutlaka çevrimiçi sosyal ağlar incelenmelidir (11).

Ayrıca çevrimiçi sosyal ağların ve elektronik iletişim cihazlarının bu kadar gelişmiş olması işyerlerinde de siber zorbalığın yaygınlaşmasına neden olmaktadır (17). İşyerlerinde siber zorbalık genellikle çevrimiçi sosyal ağlar, sosyal medya, e-posta ve SMS aracılığı ile gerçekleştirilmektedir (17, 27). Ek olarak COVID-19 gibi olağanüstü durumlar da çalışanların çevrimiçi ortamlara geçişine neden olarak bu sorunu daha da yaygın hale getirmiş olabilir (17, 28, 29). Bu nedenle incelenen araştırmalardan birisinde de veri tabanı olarak bir şirket ağı kullanılmıştır (17).

Tablo 1. İncelenen Araştırmaların Amaç, Veri Seti, Bulgular, Sonuç Ve Önerilerine İlişkin Veriler

	ÇALIŞMANIN ADI	AMAÇ	VERİ SETİ	BULGULAR	SONUÇ	ÖNERİLER
1	Analysis of Cyber Bullying on Facebook Using Text Mining (11)	Sürekli artan bu suçu azaltmak amacıyla analiz için Facebook kullanıcılarının verilerini kullanarak siber zorbalık tehdidini engellemek için metin madenciliği tabanlı bir teknik önermek.	Facebook (2.000.000 satır metin verisi)	• Zorbalık (150 kelime), Hakaret (94 kelime), Cinsel zorbalık (32 kelime), Küfür (22 kelime), Tehdit (1 kelime) ve Hayvana yönelik zorbalık (1 kelime) kategorilerinde siber zorbalıkla ilişkili içerikler tespit edilmiştir. • Yapılan karışıklık matrisi analizine göre algoritma 88 örnekten 81'ini doğru tahmin etmektedir. • Hesaplanan doğruluk skoruna göre (%95), algoritma tarafından sınıflandırılan her 100 kelimeden 95'i doğru bir şekilde sınıflandırılmaktadır. • Sınıflandırma raporu, Naive Bayes metin madenciliği tekniklerinin kullanılmasının siber zorbalıkla ilgili kelimeleri tahmin etmede etkili ve verimli olduğunu göstermektedir.	• Naive Bayes tabanlı metin madenciliği tekniğinin analizi, sınıflandırıcının doğruluğunu %0.95 olarak göstermektedir. • Bu deneysel analiz sonucu, Naive Bayes'in Facebook gönderilerinin her örneğini zorbalık içeren ya da içermeyen bir kelime olarak sınıflandırmada etkili olduğunu ve gönderilen bir zorbalık içeren kelimenin kategorisini belirleyebildiğini göstermektedir.	• Zorbalıkla ilgili resim ve videoları analiz etmek ve tespit etmek için bir teknik geliştirilmesi önerilmektedir. • Elde edilen siber zorbalık modelinin tespit doğruluğunu arttırmak için yeni bir algoritma da geliştirilebilir. • Modern sosyal medyayı karakterize eden büyük miktarda veriyi işlemek için büyük veri analizine dayalı bir teknik de kullanılabilir.
2	Artifcial Intelligence- Enabled Cyberbullying- Free Online Social Networks in Smart Cities (10)	Çevrimiçi sosyal ağlarda siber zorbalığın varlığını sınıflandırmak için akıllı şehirlerde yapay zeka destekli	Twitter (Belirtilmemiş)	• İlk veri kümesi, 1954 örneği ırkçılık ve 3122 örneği ise cinsiyetçilik sınıfı altında ve sınıflandırmıştır. 11014 örneği ise siber zorbalıkla ilgisiz olarak sınıflandırmıştır. • İkinci veri kümesi ise, 1430 örneği	• Yapay zeka destekli siber zorbalık içermeyen çevrimiçi sosyal ağ tekniğinin diğer son teknoloji ürünü tekniklere kıyasla gelişmiş siber zorbalık tespit performansı daha yüksek	• Geliştirilen Yapay zeka destekli siber zorbalık içermeyen çevrimiçi sosyal ağ tekniğinin performansı, büyük veri ortamında aykırı değer tespiti ve veri kümeleme

		siber zorbalık içermeyen çevrimiçi sosyal ağ tekniği geliştirmek.		ırkçılık ve 19190 örneği ise cinsiyetçilik sınıfı altında sınıflandırmıştır. 4163 örneği ise siber zorbalıkla ilgisiz olarak sınıflandırmıştır.	bulunmuştur.	yaklaşımlarının tasarımı açısından genişletilmelidir.
3	Comparative analysis on deep neural network models for detection of cyberbullying on Social Media (12)	Sosyal medyada siber zorbalığı tespit etmek.	Twitter (Belirtilmemiş)	• Denenilen Twitter veri kümesinden 5000'den fazla rahatsız edici olmayan yorum ve 20.000 rahatsız edici yorum tespit edilmiştir. • Kullanılan algoritma modeli için %90,4 oranında iyileştirilmiş bir doğruluk elde edilmiştir.	• Önerilen algoritma modelinin, tekniğin bilinen durumundaki diğer şemalarla karşılaştırıldığında verimli olduğunu tespit edilmiştir.	• Önerilen algoritma modeli, gelecekte video, ses ve görüntü veri kümelerinde siber zorbalığın tespitine kadar genişletilmelidir.
4	Personality and emotion based cyberbullying detection on YouTube using ensemble classifiers (13)	Kullanıcıların kişiliklerinin ve duygularının çevrimiçi iletişimleri aracılığıyla belirlenip belirlenemeyeceğini ve siber zorbalık suçunu tespit etmek için kullanılıp kullanılamayacağını incelemek.	Youtube (5152 metin verisi)	• İngilizce açıklamalı YouTube metinsel yorumlarının 2576'sı zorbalık içeren ve 2576'sı zorbalık içermeyen örnek olarak sınıflandırılmıştır. • Performans ölçümleri, %95'in üzerinde doğruluk değeri ile siber zorbalık varlığının tanımlanmasını önemli ölçüde iyileştirmek için hem kişilik özelliklerini hem de duyguyu ortaya çıkarmıştır. • Öfke ve açık sözlülük diğer duygulara ve kişiliklere kıyasla daha derindir ve nevrotik bireyler neşe, iğrenme ve korku duyguları ile siber zorbalığa sürüklenme eğilimindedir. • Açık sözlülük puanı yüksek olan bir birey öfkeyle hareket ettiğinde siber zorbalığa eğiliminin daha yüksek	• Kişilik ve duygular siber zorbalıkta uygun roller oynamaktadır. • Belirli özellikler ve duyguların tanımlanması ile daha stratejik bir müdahale programı tasarlanabilir.	• Siber zorbalık uygulayan kişilere odaklanmak için özel olarak tasarlanmış müdahale programları ve duygulara göre daha fazla strateji geliştirilmesi önerilmektedir. • Siber zorbalık mağdurlarına ise özellikle de duygu kontrolü ve başa çıkma becerilerini geliştirmek için ve açık sözlülük konusunda daha düşük puan alanlar için öfke yönetimi atölye çalışmaları/seminerleri hazırlanabilir.

				olduğu tespit edilmiştir. • Baş etmede güçlük ve zayıf sosyal becerilere sahip olmanın yanı sıra, sosyal medya platformlarının anonimlik, kullanılabilirlik ve kullanım kolaylığı gibi özellikleri, özellikle gençler arasında siber zorbalığı sürdürmek açısından ilgi çekici görülebilir.		
5	Cyberbullying Detection in Social Networks Using Bi GRU with Self Attention Mechanism (14)	Metin tabanlı siber zorbalık tespiti yapmak.	Twitter (16090 veri) Twitter (24783 veri) Wikipedia (115865 veri)	-	• Önerilen yöntem, değerlendirme ölçütleri açısından taban çizgilerinden daha iyi performans göstermektedir.	• Siber zorbalık tespitinde derin sinir ağlarının diğer boyutlarının özelliklerinden de faydalanarak kullanıcı profillerinin özellikleri ve metinsel istatistiksel özellikler incelenilebilir. • Gönderiler arasında daha geniş perspektiften bakarak daha fazla ilişkiyi açıklamak amacıyla model mimarisi için grafik sinir ağları uygulanabilir.
6	A Labeled Dataset for Investigating Cyberbullying Content Patterns in Instagram (15)	• Temel siber zorbalık özellikleri hakkında ayrıntılı etiketlerin yanı sıra zorbalığı gerçekleştiren bireyler hakkında demografik bilgiler	Instagram (3165000 veri içerisinden seçilen 2218 veri)	• 1.065 yorum siber zorbalık içeren ve 7.389 yorum siber zorbalık içermeyen yorum olarak sınıflandırılmıştır. • Siber zorbalık yorumlarının 231'i (%21,70) cinsel, 106'sı (%9,95) cinsiyet kimliği/cinsel yönelim, 195'i (%18,31) fiziksel görünüm, 130'u (%12,21) ırk/etnisite, 197'si (%18,50)	• Siber zorbalık etiketlerinin siber zorbalığın doğası hakkında nasıl yeni bilgiler sağlayabileceğini göstermektedir.	• Gelecekteki çalışmalar, bu kalıpları etkileyebilecek kullanıcıların aynı oturumda tekrarlı yorumlar yapmaları gibi bir dizi ek faktörü de hesaba katmalıdır.

		içeren açıklamalı bir Instagram veri kümesi sunmak. • Siber zorbalık ve bunun otomatik tespiti hakkında sunulan veri kümesinden nasıl yeni içgörüler elde edilebileceğini göstermek için keşifsel bir lojistik regresyon analizinin sonuçları rapor edilmek.		fikirler/düşünceler, 40'ı (%3,76) dinle ilgiliydi ve 668'i (%62,72) genel nefret yorumu olarak kabul edilmiştir. • Tüm siber zorbalık yorumlarının 990'ı (%92,96) saldırı olarak kabul edilmiştir, 88'i (%8,26) başka bir kullanıcıyı savunan yorumlar ve 78'i (%7,32) kişilerin kendilerini savundukları yorumlardır. • Yorumların 403'ü (%37,8) ilk gönderiyi yapan kullanıcıya, 628'i (%59,0) ise diğer kullanıcılara yöneliktir. • Tüm siber zorbalık yorumlarında 21'i (%1,97) depresyon, 22'si (%2,07) intihar, 8'i (%0,75) kaygı ve 118'i (%11,1) ayrımcılıkla ilgilidir.		
7	"I know you are, but what am I?" Profling cyberbullying based on charged language (16)	Twitter'da siber zorbalığın profilini çıkarmak için yordayıcı değişkenler olarak yüklü dil-eylem ipuçlarını kullanmanın etkinliğini incelemek.	Twitter (3 ayrı veri setinden elde edilmiş 140000 veri)	• Siber zorbalık tespiti için kötüye kullanım amaçlı kelimelerin ortalaması %43,83 (3. veri seti) olup, bu da %41,42 (1. veri seti) ve %41,20 (2. veri seti) ile benzerdir. • Yüklü dil-eylem ipuçlarına dayalı siber zorbalık profilini çıkarılması yalnızca etkili bir yaklaşım olmakla kalmayıp, aynı zamanda tutarlı sonuçlar da sağlayabilmektedir. • Siber zorbalık eğilimini belirten 1. veri seti için verilerin %46.60'ı, 2. veri seti için verilerin %38.33'ü ve 3. veri	• Önerilen yöntem, istatistiksel önem ve tutarlılık ile siber zorbalık etkinliğini güçlü bir şekilde önleyebilmektedir. • Çalışma, siber zorbaların (muhtemel saldırganların) sosyal medyada suçlayıcı bir dil kullanarak siber kurbanları (uygun hedefler) istismar ettiği rutin eylem teorisi temelinde sağlam bir duruş sergileyerek siber zorbalık literatürünü geliştirmektedir.	• Önerilen yöntem, siber zorbaları olası suç faaliyetlerinden koruyabilir, potansiyel kurbanları koruyabilir ve sosyal medya platformları, gençlik danışmanları ve kolluk kuvvetleri gibi arabuluculuk kurumları için siber zorbalığı profillemek için proaktif bir önlem sağlayabilir.

				seti için verilerin %43.83'ünün küfürlü tweet olduğu tespit edilmiştir.	• Çalışma, politika ve müdahale yöntemleri geliştirmek için dikkate değer çıkarımlarla bilgilendirmek için bir siber zorbalık profilleme önlemi türetmektedir. • Siber zorbalar, siber mağdurları suistimal etmek için küfürlü kelimeler kullanma eğilimindedir. Ancak, siber zorbalık olaylarını tespit etmek için sadece küfürlü kelimeler kullanmak yetersizdir, çünkü tüm taciz edici kelimeler siber zorbalık niyetini belirtmeye uygun değildir. • Çalışma, nefret dolu ve kötüye kullanım amaçlı bilgi davranışında bulunan siber zorbaları profillemek için etkili bir yol olduğunu kanıtlayan, yüklü dil eylemi ipuçlarını öngörücü değişkenler olarak kavramsallaştırmış ve doğrulamıştır.	
8	MaLang: A Decentralized Deep Learning Approach for Detecting Abusive Textual Content (17)	Bir şirketin ağlarındaki herhangi bir mesajlaşma uygulamasında toksik veya kötüye	Bir şirket ağı (28000 veri)	• Verilerin %98'inden fazlası akıllı telefonlar veya kişisel bilgisayarlar için kötü amaçlı olan ve kötü amaçlı olmayan metin mesajları olarak sınıflandırmayı başarmıştır. • Önerilen yöntemin kullanılması ile,	• Önerilen yöntem, mevcut diğer siber zorbalık tespit sistemlerinden daha iyi performans gösteren %98,2 sınıflandırma doğruluğuna ulaşmıştır.	• Önerilen yöntemde gelecekte yapılacak olası bir yükseltme ile, kullanıcıya yazması için herhangi bir kötüye kullanım içeriği önermeden önce gerçek zamanlı metin tahmini

		kullanım amaçlı içeriğin kullanımıyla mücadele etmek için kötüye kullanım amaçlı metin içeriğini tespit etmek.		bilgi işlem kaynaklarının ve zamanın tüketiminin önceki siber zorbalık tespit sistemlerine göre %35-40 oranında azaltılabileceği tahmin edilmektedir.		yapılacak şekilde ölçeklendirilebilir. • Bu, potansiyel olarak kötü amaçlı kulllanım veya toksik içeriğin kullanımını azaltabilir.
9	A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter (18)	Mağdurların katılımı olmadan siber zorbalığı tespit etmek.	Twitter (37373 veri)	• Önerilen modelin üstünlüğü %90.57'lik bir medyan doğruluk elde edilerek deneysel sonuçlarla gösterilmiştir.	-	• Siber zorbalığı anında tespit etmek ve önlemek için faydalı olacak gerçek zamanlı bir siber zorbalık tespit platformu oluşturulmalıdır. • Farklı dillerde, özellikle de Arapça bağlamında siber zorbalığın tespiti üzerinde çalışılmalıdır.
10	CyberNet: a hybrid deep CNN with N- gram feature selection for cyberbullying detection in online social networks (19)	Kullanıcıların kelime ve karakterlerinde siber zorbalığı tespit edebilen ve gönderilen mesajın engellenmesini sağlayan bir model geliştirmek.	Twitter (21832 veri)	• Bu çalışmada, karakter seviyesi ve kelime seviyesi ile siber zorbalık etkili bir şekilde tespit edilmiştir. • Asıl sorun, saldırganların siber zorbalık için kelime dağarcığında olmayan kelimeler kullanmasıdır.	• Önerilen CyberNet yaklaşımı, son teknoloji yaklaşımlara kıyasla çok daha belirgin niteliksel ölçümler sağlamaktadır.	• Zorbalık içeren kelime ve karakterlerin tanımlanmasının genişletilmesi; eş anlam listesinin artırılması ve görüntüler için siber zorbalık tanımının artırılası önerilmektedir.
11	Recognition and prevention of cyberharassment in social media using Classification	Makine Öğrenimi Sınıflandırma Algoritmaları ile sosyal ağ ortamlarında siber	Twitter (25000 veri)	• Veri setine sahip katılımcıların en 1/3 ırkçı söylem gerçekleştirmektedir. • Yaklaşık 5.900 tweet'in geri alınması ya tweetin silinmesi ya da hesabın devre dışı bırakılması nedeniyle	• Bu çalışma, metin tabanlı siber zorbalığın ortaya çıkarılması sorununu; yorumların zorbalık içerip içermediğini tespit etmek için kullanılan bir model olarak	-

	algorithms (8)	zorbalığı tespit etmek ve önlemek.		örnekleme dahil edilmemiştir. • Doğruluk: %92,81, Hassasiyet: %96,97 oranıyla model siber zorbalığı tespit edebilmektedir.	başarılı olmuştur. • Önerilen sistem, diğer modellerden ve tekniklerden çok verimli bir şekilde öğrenebilmiştir. • Zorbalık kelimeleri, öğretilen kelimeler aracılığıyla otomatik olarak bulunabilinmiştir.	
12	Prediction of risk factors of cyberbullying-related words in Korea: Application of data mining using social big data (20)	Siber zorbalık türlerini sınıflandırmak ve türlerine göre onu etkileyen faktörleri belirlemek; Her siber zorbalık türü için risk faktörlerini belirleyebilecek bir karar ağacı geliştirmek.	227 çevrimiçi kanaldan toplanabilen 435.565 siber zorbalıkla ilgili veri	Siber zorbalık türleri şu şekilde sınıflandırılmıştır: • %56 belge (57.817 vaka) herhangi bir duygu içermemektedir, • %32,3 mağdur (33.361 vaka), %6,4 fail (6587 vaka) ve • %5,3 olay yerindeki seyirciler (5447 vaka).	• Siber zorbalığın nedensel faktörleri arasında dürtü faktörü ilk sırada yer alırken, bunu baskınlık faktörü, görünüm faktörü ve kültür faktörü için eğilim izlemiştir. • Siber zorbalık yöntemleri sırasıyla şiddet, alay ve toplu şiddeti içermektedir. • Dürtü faktörünün hem seyirci hem de mağdur üzerinde etkisi vardır. Yani dürtü faktörü faili değil, mağduru ve görgü tanığını etkiler.	• Seyircilerin çoğu bir siber zorbalık eylemini dürtüsel olarak ağırlaştırabileceğinden, dürtüleri azaltmak için kurbanlar ve failler için olanlara ek olarak seyirciler için programlar geliştirmesi önerilmiştir. • Zorbalar için danışma programları veya ebeveyn eğitim programları da gerekli olacaktır.
13	A Socio-Contextual Approach in Automated Detection of Public Cyberbullying on Twitter (21)	Siber zorbalığı belirlemek için sosyal ağlarda denetimler gerçekleştirecek bir prototip geliştirmek.	Twitter (12837 veri) Twitter (18173 veri)	(21)• 607 siber zorbalık tweeti içeren 12837 İngilizce tweet, • 388 siber zorbalık tweeti içeren diğer kullanıcıların 8850 Rtweetler, • 219 siber zorbalık vakası içeren 3987 orijinal tweet.	• Siber zorbalıkta mağdur ve zorbalık yapan grup arasında bir güç dengesizliği bulunmaktadır. • Kurbanın nispeten yüksek sayıda takipçisi olsa da kurbanı hedefleyen tweetlerin hacmi ve siber zorbaların grup olarak ulaşabileceği hedef kitle aralığı,	• Siber zorbalığın sadece küfürlü ve olumsuz kelimeler içermediği aynı zamanda siber zorbalığın mekânsal-bağlamsal boyutta incelenerek değerlendirilmesi önerilmektedir.

					mağdurun hedef kitlesinden önemli ölçüde daha yüksek bulunmuştur. • Zorbalık tweetleri ile mağdurla alay etmek amacıyla kullanılan bir hashtag oluşturma eylemi eğilimindedirler. Grubun hedeflenen kişiye siber zorbalık yapma niyetini göstermektedir.	
14	Prototype to Perform Audit in Social Networks to Determine Cyberbullying (22)	Siber zorbalığı belirlemek için sosyal ağlarda denetimler gerçekleştirecek bir prototip geliştirmek.	599 çevrimiçi bağlantı verisi	Birinci sosyal ağda performans, sosyal ağ veri setinde 140 adet ve şikayet/yorum setinde 81 adet olmak üzere %94,19; sosyal ağ sektöründe performans, sosyal ağ veri setinde 121 adet ve şikayet/yorum setinde 93 adet olmak üzere %92.31; onuncu sosyal ağda performans, sosyal ağ veri setinde 195 adet ve şikayet/yorum setinde 109 adet olmak üzere %94,39'dur.	•Algoritma prototipinin, saldırgan metnin sınıflandırması ve ağırlığı aracılığıyla sosyal ağlardan açık ve kesin bir şekilde geçerli bilgiler elde ettiği tespit edilmiştir. •Prototip sosyal ağ veri setinde %93.80 oranında performans elde etmiştir.	Araştırma, canlı veri toplama, metinlerin analizi ve ön işlemesi, görüntülerin ve videoların analizi ve sınıflandırılmasına odaklanmalı; Siber zorbalık süreçlerini hızlandırmak ve gerekli araçları tanımak için denetçi eğitimi teşvik edilmelidir.
15	Personal attacks decrease user activity in social networking platforms (23)	Çevrimiçi kişisel saldırıların sosyal medya kullanıcı etkinliği üzerindeki etkilerini tespit etmek.	Reddit (182528 veri)	-Örneklem içinde 5717 veri (%23.6), “dar” modellerle tespit edilen ikinci bir kişiye karşı kişisel saldırılar içermektedir, -Örneklem içinde 8837 veri (%36.4), “geniş” modeller tarafından tespit edilen ikinci bir kişiye karşı kişisel saldırılar içermektedir -Örneklem içinde 10.023 (%41.3) üçüncü kişiye yönelik kişisel saldırılar	•Yüksek hassasiyetli bir kişisel saldırı tespit aracı Samurai uygulaması uygulanmıştır. •Çalışmanın sonucunda, çevrimiçi kişisel saldırılara maruz kalan kullanıcıların etkinliklerinin oldukça hızlı bir şekilde bozulduğunu açıkça göstermiştir. Saldırıya uğrayan kullanıcıların etkinliklerini	Samuray uygulamasının Reddit dışında Twitter, Facebook veya Instagram dahil olmak üzere diğer SNS platformlarında, Lehçe ve Japonca gibi İngilizce dışındaki dillerde de çalışmaya devam etmesi önerilmektedir. Uygulamanın diğer taciz türleri (cinsel, ırksal, çocuk

				içermektedir.	azalacağı ve kişisel saldırılara sık sık maruz kalmanın platformdan ayrılmaya neden olacağı tahmin edilmektedir. •Kişisel saldırıların olumsuz etkisi ve ılımlılığın hafifletici rolü, birlikte, kişisel saldırıları ve benzeri taciz ve siber zorbalık vakalarını etkili bir şekilde tespit etmenin ve denetlemenin SNS sağlayıcılarının çıkarına olduğunu göstermektedir	istismarı vb.) dahil olmak üzere çeşitli saldırı türlerini kapsayacak şekilde genişletmeyi ve yalnızca saldırıları tespit etmeye değil, aynı zamanda sosyal medyanın ve kullanıcılarının genel sağlığının iyileştirilmesine katkıda bulunmak için, kullanıcıların mesajlarındaki intihar niyetini tespit etmek için modeller geliştirmeyi planlamaktadır.
16	Understanding cyberbullying as an information security attack—life cycle modeling (26)	• Konu modelleme yoluyla bir siber zorbalık yaşam döngüsü modeli önermek. • Bilgisayar saldırılarıyla ilişkili kriterleri göz önünde bulundurarak saldırının farklı aşamalarını kavramsallaştırmak.	Twitter (250.000 saldırganla ilgili tweet ve 3035 kurban deneyiminden oluşan veri seti)	-	• Bu araştırmada elde edilen sonuçlar, siber zorbalıkla ilgili daha önceki bir araştırmada kanıtlandığı gibi, geleneksel olmayan saldırı modelleri elde etme sürecinin geçerliliğini doğrulamaktadır.	-
17	Türkçe metinlerde makine öğrenmesi yöntemleri ile siber zorbalık tespiti (9)	Türkçe bir hazır veri seti kullanılarak siber zorbalık tespiti problemini ele	Kaggle (3000 satırlık veri)	• 1497 veri negatif (siber zorbalık içermeyen) olarak tespit edilirken, 1503 veri ise pozitif (siber zorbalık içeren) olarak değerlendirilmiştir.	• Siber zorbalık tespitinde kullanılan makine öğrenme algoritma performanslarının kullanılan veri setindeki dil	• Gelecekte siber zorbalık tespitinde yapılan metin bazlı araştırmalarda daha sağlıklı ve güvenilir sonuçların elde

		almak.		• Veri ön işlemlerinden sonra veri seti üzerinde çalıştırılan sınıflandırma algoritmalarının performansları incelendiğinde %88.35 başarım oranı ile LR sınıflandırma algoritmasının en yüksek başarım oranına sahip olduğu tespit edilmiştir.	yapısına bağlı olarak da farklı sonuçlar verebileceği görülmüştür.	edilmesine yönelik metinlerin anlamsal boyutlarının da dikkate alınarak değerlendirilebileceği derin öğrenme algoritmalarının sosyal ağlar üzerinde anlık siber zorbalık tespitine yönelik çalışmalar yapılması.
18	Session-based Cyberbullying Detection in Social Media: A Survey (25)	Sosyal medyada oturum tabanlı siber zorbalık tespitine ilişkin kapsamlı bir genel bakış sunarak, mevcut çabaları veri ve metodolojik bir bakış açısıyla incelemek.	Oturum tabanlı siber zorbalık veri kümeleri	-	• İncelemede mevcut veri kümesini ve modelleri tartışılmaktadır. SSCD veri kümeleri üzerindeki en son modelleri değerlendiren bir dizi karşılaştırmalı, kıyaslama deneyi sunmaktadır ve veri kümesi ve model oluşturmaktadır.	-
19	Implementatıon of hyperparameter optımısatıon and over samplıng in detectıng Cyberbullyıng usıng machıne learnıng approach (24)	Çevrimiçi sosyal ağlarda siber zorbalığa karşı koymak.	ASKfm (113021 gönderi verisi)	Sonuçlar, önerilen çerçevenin, eğitimli modelimiz tarafından gerçekleştirildiği üzere, eğri altındaki alanın en yüksek yüzdesinin %99.24 ve F-ölçümünün %97.38 olduğu önemli sonuçlar sağladığını göstermiştir. Şekil 6'da gösterildiği gibi, son veri setimiz yaklaşık olarak 113 021'dir (yani, siber zorbalık, 5375 ve siber zorbalık dışı, 107 646) olarak açıklamalı veri seti.	•Bu çalışmanın katkıları arasında, uygun öznitelikler kullanılarak sosyal medya ve makine öğrenmesi tekniğinin özellikle metin sınıflandırmada değerli olabileceğine dair kanıtlar sunmak yer almaktadır. Metin sınıflandırma becerileri, veri bilimcileri için çok önemlidir çünkü bu teknikler her veri analizinin merkezinde yer alır. •Ayrıca, önerilen çerçeve tıbbi tahmin, romantizm	Gelecekteki metin sınıflandırma makine öğrenimi araştırmalarında, özellikle siber zorbalık tespitinde bir paradigma değişikliğinin gerekli olduğunu anlıyoruz. Gelecekteki araştırmaların bir başka büyüleyici yönü olarak, saf Bayes, karar ağacı, lojistik regresyon ve rastgele orman gibi farklı algoritmalar üzerinde deneyler yapacağız. Ayrıca duygu temelli özellikler

					dolandırıcılığı aldatma ve spam ve sahte haber tespitleri gibi diğer durumlarda da kullanılabilir. Bu çalışmanın sonuçları, özelleştirilmiş ve yeni veri modelleri oluşturmak için kullanılabilir. Ayrıca, etkili karar vermeyi desteklemenin yanı sıra, madencilik süreçlerinin daha iyi anlaşılması için madencilik endüstrisine yardımcı olur.	(yani kutupluluk) ve kullanıcı temelli özellikler (yani yaş, cinsiyet, ırk) gibi diğer özellikleri de keşfedeceğiz. Ek olarak, derin öğrenme, gelecekteki araştırmalarımızda keşfedebileceğimiz ikili sınıflandırmaya katkıda bulunabilir.

3.3. Yapay Zekanın Siber Zorbalığı Tespit Etmedeki Çıktıları

Geliştirilen yapay zeka programları siber zorbalığı sınıflandırabilmekte (10, 11, 15) tespit edebilmekte (9-13, 15-20, 22-24) ve engelleyebilmektedir (13). Yapılan bir çalışmada siber zorbalık varlığının tanımlanmasında bireylerin kişilik özellikleri ve duygularının önemi ortaya konulmuştur. Aynı çalışmaya göre nevrotik bireyler neşe, iğrenme ve korku duyguları sonucunda siber zorbalığa sürüklenme eğilimindedirler. Ayrıca açık sözlü bireylerin öfkeyle hareket ettiklerinde siber zorbalık eğilimlerinin yükseldiği tespit edilmiştir (13).

Çalışmalardan birisi siber zorbalıkla ilişkili içerikleri zorbalık, hakaret, cinsel zorbalık, küfür, tehdit ve hayvana yönelik zorbalık olarak sınıflandırmış ve en yüksek oranda zorbalık sınıfında yer alan içerik tespit etmiştir (11). Bir diğer çalışmada siber zorbalıkla ilişkili içerikler ırkçılık ve cinsiyetçilik sınıfı altında sınıflandırılmış ve en yüksek oranda cinsiyetçilik sınıfında yer alan içerik tespit edilmiştir (10).

Başka bir çalışmada ise siber zorbalık yorumlarının 231'i (%21,70) cinsel, 106'sı (%9,95) cinsiyet kimliği/cinsel yönelim, 195'i (%18,31) fiziksel görünüm, 130'u (%12,21) ırk/etnisite, 197'si (%18,50) fikirler/düşünceler, 40'ı (%3,76) dinle ilgiliydi ve 668'i (%62,72) genel nefret yorumu olarak saptanmıştır. Ayrıca siber zorbalık yorumlarının 990'ı (%92,96) saldırı olarak kabul edilmiştir, 88'i (%8,26) başka bir kullanıcıyı savunan yorumlar ve 78'i (%7,32) kişilerin kendilerini savundukları yorumlar olarak açıklanmıştır ve tüm siber zorbalık yorumlarının %11,1'inin ayrımcılıkla ilgili olduğu bulunmuştur (15). Farklı bir çalışmada ise siber zorbalık yorumlarının 1/3'ünün ırkçılıkla ilgili olduğu bulunmuştur (8).

4. SONUÇ VE ÖNERİLER

İncelenen çalışmaların tamamında önerilen yapay zeka modelleri siber zorbalığa yönelik amaçlarını gerçekleştirme konusunda başarılı sonuçlar elde etmiştir. Çalışmalardan birinde siber zorbalığı mesaj içeriğinden tespit ederek mesajın gönderilmesini engelleyebilen bir yapay zeka modeli geliştirilmiştir (19).Ayrıca yapa zeka uygulamaları yalnızca çevrimiçi sosyal ağlarda değil, iş yerlerinde kurum içi kullanılan ağlarda da siber zorbalık başarılı bir şekilde tespit edilmiştir (17).

Siber zorbaların sosyal medyada suçlayıcı bir dil kullanarak siber kurbanları istismar etmesi teorisi temelinde literatüre katkı sağlanmıştır (16). Ek olarak bireylerin kişilik özellikleri ve duyguları da siber zorbalıkta uygun roller oynamaktadır (13). Siber zorbalar, siber mağdurları suistimal etmek için küfürlü kelimeler kullanma eğilimindedir. Ancak, siber zorbalık olaylarını tespit etmek için sadece küfürlü kelimeler kullanmak yetersizdir, çünkü tüm taciz edici kelimeler siber zorbalık niyetini belirtmeye uygun

değildir (16). Ayrıca dürtü, baskınlık, görünüm ve kültür faktörleri siber zorbalığın nedensel faktörleri arasında yer almaktadır. Bunlardan özellikle dürtü faktörünün hem seyirci hem de mağdur üzerine etkisi vardır. Yani dürtü faktörü faili değil, mağduru ve görgü tanığını etkilemektedir (20). Başka bir çalışmaya göre ise kurbanın daha çok sayıda takipçisi olsa bile, kurban ve zorbalık yapan grup arasında bir güç dengesizliği bulunmaktadır ve zorbalar kurbanla alay etmek için hashtagler oluşturma eğilimindedirler (21).

Çevrimiçi kişisel saldırılara maruz kalan kullanıcıların sosyal ağlardaki etkinliklerinin çok hızlı bir biçimde bozulduğu ve hatta çevrimiçi sosyal ağ platformundan ayrılmaya neden olabileceği tespit edilmiştir (23).

Siber zorbalıkla ilgili çalışmalar genellikle metin tabanlı analizleri içermektedir ancak siber zorbalığı daha iyi bir şekilde tespit edebilmek için resim, video ve sesleri analiz edebilen tekniklerin geliştirilmesi önerilmektedir (9, 11, 12, 19, 22). Uygulamalarda yapılabilecek ek güncellemelerle programlar kullanıcıya yazması için herhangi bir kötüye kullanım içeriği önermeden önce gerçek zamanlı metin tahmini yapılacak şekilde ölçeklendirilmesi önerilmektedir (17, 18).

Ek olarak denenen modellerin tespit doğruluğunu arttırmak için büyük veri analizine dayalı yeni algoritmalar geliştirilmelidir (10, 11) ve geliştirilecek yeni programlar farklı dillerde siber zorbalığın tespiti üzerinde çalışılmalıdır (18, 23).

Geliştirilen programlar sosyal medya platformları, gençlik danışmanları ve kolluk kuvvetleri gibi arabuluculuk kurumları için siber zorbalığı görüntülemek için proaktif bir önlem sağlaması açısından kullanılmalıdır (16).

Siber zorbalık uygulayan kişilere odaklı müdahale programlarına ve kullanıcı profillerinin özelliklerine yönelik daha fazla strateji geliştirilmelidir (13, 14). Zorbalar için danışma programları veya ebeveyn eğitim programları uygulamaya konulmalıdır (20).

Siber zorbalık mağdurlarına ise özellikle de duygu kontrolü ve başa çıkma becerilerini geliştirmek için ve açık sözlülük konusunda daha düşük puan alanlar için öfke yönetimi atölye çalışmaları/seminerleri hazırlanmalıdır (13).

Ayrıca uygulamalar diğer taciz türleri (cinsel, ırksal, çocuk istismarı vb.) dahil olmak üzere çeşitli saldırı türlerini kapsayacak şekilde genişletilmeli ve yalnızca saldırıları tespit etmeye değil, aynı zamanda sosyal medyanın ve kullanıcılarının genel sağlığının iyileştirilmesine katkıda bulunmak için, kullanıcıların mesajlarındaki intihar niyetini de tespit edebilecek modeller geliştirilmelidir (23).

KAYNAKÇA

1. Olweus D, Limber SP. Some problems with cyberbullying research. Curr Opin Psychol. 2018;19:139-43.
2. Vanderbilt D, Augustyn M. The effects of bullying. Paediatrics and Child Health. 2010;20:315-20.
3. Thomas HJ, Connor JP, Scott JG. Integrating Traditional Bullying and Cyberbullying: Challenges of Definition and Measurement in Adolescents – a Review. Educational Psychology Review. 2015;27(1):135-52.
4. Slonje R, Smith PK, FriséN A. The nature of cyberbullying, and strategies for prevention. Comput Hum Behav. 2013;29(1):26-32.
5. Sangwan SR, Bhatia MPS. Soft computing for abuse detection using cyber-physical and social big data in cognitive smart cities. Expert Systems. 2022;39(5):e12766.
6. Huang Y-y, Chou C. An analysis of multiple factors of cyberbullying among junior high school students in Taiwan. Computers in Human Behavior. 2010;26:1581-90.
7. Li J, Hesketh T. Experiences and Perspectives of Traditional Bullying and Cyberbullying Among Adolescents in Mainland China-Implications for Policy. Frontiers in Psychology. 2021;12.
8. Rajesh S, Sharanya B. Recognition and prevention of cyberharassment in social media using classification algorithms. Materials Today: Proceedings. 2021.
9. Yazğılı E, Baykara M. Türkçe metinlerde makine öğrenmesi yöntemleri ile siber zorbalık tespiti. Gümüşhane Üniversitesi Fen Bilimleri Dergisi. 2022;12(2):443-53.
10. Al-Marghilani A. Artificial Intelligence-Enabled Cyberbullying-Free Online Social Networks in Smart Cities. International Journal of Computational Intelligence Systems. 2022;15(1):9.
11. Aliyu N, Abdulrahaman M, Ajibade F, Abdurauf T. Analysis of Cyber Bullying on Facebook Using Text Mining. Journal of Applied Artificial Intelligence. 2020;1:1-12.
12. Balakrishna S, Gopi Y, Solanki VK. Comparative analysis on deep neural network models for detection of cyberbullying on Social Media. Ingeniería Solidaria. 2022;18(1):1-33.
13. Balakrishnan V, Ng S. Personality and emotion based cyberbullying detection on YouTube using ensemble classifiers. Behaviour & Information Technology. 2022:1-12.
14. Fang Y, Yang S, Zhao B, Huang C. Cyberbullying Detection in Social Networks Using Bi-GRU with Self-Attention Mechanism. Information [Internet]. 2021; 12(4).
15. Hamlett M, Powell G, Silva YN, Hall D. A Labeled Dataset for Investigating Cyberbullying Content Patterns in Instagram. Proceedings of the International AAAI Conference on Web and Social Media. 2022;16(1):1251-8.
16. Ho SM, Li W. "I know you are, but what am I?" Profiling cyberbullying based on charged language. Computational and Mathematical Organization Theory. 2022.

17. Kompally P, Sethuraman SC, Walczak S, Johnson S, Cruz MV. MaLang: A Decentralized Deep Learning Approach for Detecting Abusive Textual Content. Applied Sciences [Internet]. 2021; 11(18).

18. Muneer A, Fati SM. A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter. Future Internet [Internet]. 2020; 12(11).

19. Paruchuri VL, Rajesh P. CyberNet: a hybrid deep CNN with N-gram feature selection for cyberbullying detection in online social networks. Evolutionary Intelligence. 2022.

20. Song T-M, Song J. Prediction of risk factors of cyberbullying-related words in Korea: Application of data mining using social big data. Telematics and Informatics. 2021;58:101524.

21. Tahmasbi N, Rastegari E. A Socio-Contextual Approach in Automated Detection of Public Cyberbullying on Twitter. ACM Transactions on Social Computing. 2018;1:1-22.

22. Toapanta M, Recalde Monar JA, Mafla E. Prototype to Perform Audit in Social Networks to Determine Cyberbullying2020.

23. Urbaniak R, Ptaszyński M, Tempska P, Leliwa G, Brochocki M, Wroczyński M. Personal attacks decrease user activity in social networking platforms. Computers in Human Behavior. 2022;126:106972.

24. Wan Ali WNH, Mohd M, Fauzi F, Shirai K, Noor M. Implemention of hyperparameter optimisation and over-sampling in detecting cyberbullying using machine learning approach. Malaysian Journal of Computer Science. 2021:78-100.

25. Yi P, Zubiaga A. Session-based Cyberbullying Detection in Social Media: A Survey. arXiv preprint arXiv:220710639. 2022.

26. Zambrano P, Torres J, Yánez Á, Macas A, Tello-Oquendo L. Understanding cyberbullying as an information security attack—life cycle modeling. Annals of Telecommunications. 2021;76(3):235-53.

27. Oksanen A, Oksa R, Savela N, Kaakinen M, Ellonen N. Cyberbullying victimization at work: Social media identity bubble approach. Computers in Human Behavior. 2020;109:106363.

28. Kowalski RM, Toth A, Morgan M. Bullying and cyberbullying in adulthood and the workplace. The Journal of Social Psychology. 2018;158(1):64-81.

29. Oksa R, Saari T, Kaakinen M, Oksanen A. The Motivations for and Well-Being Implications of Social Media Use at Work among Millennials and Members of Former Generations. International Journal of Environmental Research and Public Health [Internet]. 2021; 18(2).