GUFBD / *GUJS* (2022) 12(2): 443-453
DOI: 10.17714/gumusfenbil.935448 Araştırma Makalesi / *Research Article*

Türkçe metinlerde makine öğrenmesi yöntemleri ile siber zorbalık tespiti

Cyberbullying detection with machine learning methods in Turkish texts

Enver YAZĞILI*[1,a], Muhammet BAYKARA[2,b]
[1] *Munzur Üniversitesi Tunceli Meslek Yüksekokulu, Bilgisayar Programcılığı Bölümü,62000, Tunceli*
[2] *Fırat Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü, 23000, Elazığ*

• Geliş tarihi / *Received*: 10.05.2021 • Düzeltilerek geliş tarihi / *Received in revised form*: 11.01.2022 • Kabul tarihi / *Accepted*: 01.02.2022

Öz
İnternet kullanımının yaygınlaşması ve sosyal medya platformlarının popülaritesinin artması siber zorbalık olarak adlandırılan eylemlerin hızla yayılmasına neden olmuştur. Dünya genelinde siber zorbalığa maruz kalan kişilerin sayısı her geçen gün artmaktadır ve bu da mağdurlar üzerinde büyük etkiler yaratmaktadır. Bu eylemin tespit edilmesi, yeni mağdurların ortaya çıkmaması ve mevcut mağdurların daha fazla bu eyleme maruz kalmaması açısından büyük önem taşımaktadır. Bu noktada literatürde siber zorbalık tespitine yönelik birçok çalışmanın gerçekleştirildiği görülmüş ancak Türkçe metinlerde yapılan çalışma sayısının çok az olduğu tespit edilmiştir. Bu çalışmada kaggle adlı paylaşım sitesinden elde edilmiş ve manuel olarak oluşturulan 3000 cümlelik hazır Türkçe bir veri seti üzerinde doğal dil işleme yöntemleri kullanılarak siber zorbalık tespiti gerçekleştirilmiştir. Çalışmada kullanılan veri setinin yeni olması ve bildiğimiz kadarıyla bu kadar çok sayıda algoritmanın literatürde test edilmemiş olması nedeniyle bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Çalışmada bu veri seti üzerinde Bagging, Boosting, C4.5, Gradient Boosting, K-Means, KNN, LR, NB, ANN, RO, DVM, Stokastik Gradient Descent ve XGBoost algoritmaları karşılaştırmalı olarak ilk kez kullanılmıştır.

Anahtar kelimeler: Bilgi güvenliği, Makine öğrenmesi, Siber güvenlik, Siber suç, Siber zorbalık, Veri analizi

Abstract
Undoubtedly, the widespread use of the internet and the increasing popularity of social media platforms have caused the rapid spread of the actions called cyberbullying. The number of people subjected to cyberbullying throughout the world is increasing day by day and it has a great impact on their victims. Identifying this action is of great importance in terms of preventing the emergence of new victims and not being exposed to this action any more. At this point, it has been observed that many studies have been carried out in the literature on the detection of cyberbullying, but it has been determined that the number of studies in Turkish texts is very low. It is thought that this study will contribute to the literature because the dataset used in the study is new and to the best of our knowledge, such a large number of algorithms have not been tested in the literature. In the study, Bagging, Boosting, C4.5, Gradient Boosting, K-Means, KNN, LR, NB, ANN, RO, DVM, Stochastic Gradient Descent and XGBoost algorithms were used comparatively for the first time on this data set.

Keywords*: Information security, Machine learning, Cyber security, Cyber crime, Cyberbullying, Data analysis*

[*a] Enver YAZĞILI; enveryazgili@munzur.edu.tr, Tel: (0536) 365 62 21, orcid.org/ 0000-0001-8459-3488
[b] orcid.org/0000-0001-5223-1343

1. Giriş
1. Introduction

Siber zorbalık, bir bireyin veya bir grubun bilgi ve iletişim teknolojilerini kullanarak başka kişilere veya gruplara tehdit, aşağılama, şantaj, dışlama, kızdırma, kışkırtma veya buna benzer içerikler barındıran mesajları tekrarlı olarak göndermesiyle gerçekleştirdiği eylemler olarak tanımlanabilir (Barlet, 2019). Siber zorbalık, klasik zorbalık türünden farklı olarak siber zorbaların sosyal medya platformları aracılığıyla daha fazla kitleye erişim imkânı bulmaları, bu eylemi gerçekleştiren kişi veya grupların kimliklerini gizleyebilme imkânı bulabilmeleri ve kullanıcıların fiziksel üstünlüğe ihtiyaç duymadan bu eylemi gerçekleştirebiliyor olmaları nedeniyle daha hızlı yayılmaktadır. Ayrıca bu tür iletişim ve sosyal medya araçlarını kullanan kullanıcıların genç, savunmasız ve korunmaya muhtaç kişilerden oluşması siber zorbalığa daha fazla maruz kalmalarına neden olmaktadır. Siber zorbalığa maruz kalan kişilerde; çoğu zaman kendilerini toplumdan soyutlama, itibarlarının kaybedilmesi, aile ve iş hayatlarının alt üst olması, depresyon ve bunun sonucunda da intiharın eşiğine gelme durumları görülebilmektedir. Bu sonuçlar durumun ne kadar vahim olduğunu gözler önüne sermektedir. 2019 yılı TUİK verilerine göre Türkiyede 16-74 yaş aralığındaki bireylerin %75.3'ünün İnternet kullandığı, bu sayının 2020 yılında ise %79.0'a yükseldiği açıklanmıştır (TUİK, 2020). Bu da bize dünya genelinde olduğu gibi ülkemizde de siber zorbalık eylemlerinin gerçekleştirilmesi için uygun zeminin olduğunu göstermektedir. Bu nedenle Türkçe veri setleri üzerinde yapılacak bir çalışma ülkemizde siber zorbalığın tespit edilmesinde önemli ölçüde yarar sağlayacaktır. Yapılan bir başka araştırmada ise bu tür eylemlerin sıklıkla Facebook ve Twitter gibi sosyal paylaşım sitelerinde ve kişisel mobil iletişim araçları üzerinde meydana geldiğini göstermektedir (Balakrishnan vd., 2019). Her ne kadar söz konusu platformlar siber zorbalığa yönelik tedbirler almış olsa da bu tedbirler yeterli olmamakta ve siber zorbalığı engelleyememektedir. Bu nedenle siber zorbalığın bu alanlarda hızlı ve güvenilir bir şekilde tespit edilmesi çok önemlidir. Siber zorbalık, geniş kitlelere yayılması ve henüz tam anlamıyla tespitini gerçekleştiren bir çalışmanın yapılmamış olması nedeniyle son yıllarda araştırmacıların ilgi odağı haline gelmiştir. Bu alanda birçok çalışma gerçekleştirilmiş ve yeni çalışmaların devam ettiği görülmüştür. Yapılan çalışmalarda olumlu sonuçlar elde edilmiş olsa da siber zorbalığın anlık ve doğru tespiti henüz tam olarak gerçekleştirilmemiştir. Dünya genelinde olduğu gibi Türkiye'de de siber zorbalığın hızla yayılıyor olması, çalışmaların genellikle farklı dil yapılarında oluşturulmuş veri setleri üzerinde gerçekleştirilmiş olması ve bunun yanında Türkçe metinlerden oluşan veri setleri üzerinde yapılan çalışma sayısının ise yok denecek kadar az sayıda olması bu çalışmanın gerçekleştirilmesine ilham vermiştir. Bu çalışmada Türkçe metinlerden oluşan hazır bir veri seti kullanılarak siber zorbalık tespiti gerçekleştirilmiş, Bagging, Boosting, C4.5, Gradient Boosting, K-Means, KNN, LR, NB, YSA, RO, DVM, Stochastic Gradient Descent ve XGBoost olmak üzere onüç adet sınıflandırma algoritmasının performansları karşılaştırılmıştır.

Bu çalışma dört bölümden oluşmaktadır. İkinci bölümde literatür araştırması yapılmıştır. Üçüncü bölümde önerilen model, veri seti, veri ön işlemleri ve makine öğrenme algoritmalarının nasıl kullanıldığı ile ilgili yöntemler kısaca açıklanmıştır. Dördüncü bölümde ise elde edilen sonuçlar verilmiş olup, gelecekte yapılması amaçlanan çalışmalar belirtilmiştir.

1. İlgili çalışmalar
2. Related studies

Literatürde siber zorbalığın tespitine, analizine ve gruplandırılmasına yönelik birçok farklı çalışma bulunmaktadır. Ancak tespit çalışmalarında kullanılan veri setlerinin çoğunun İngilizce veya farklı dillerden elde edilmiş veri setleri olduğu tespit edilmiştir (Agrawal & Aweka, 2018; Al-Mamun & Akhter, 2018; Balakrishnan vd., 2019; Balakrishnan vd., 2019; Dadvar vd., 2012; Duwairi vd., 2014; Fortunatus vd., 2020; Hosseinmardi vd., 2015; Hussain vd., 2018; Kepez, 2021; MinSong & Song, 2020; Modha vd., 2020; N-Garci´a vd., 2015; Shekhar & Mathangi, 2018; Squicciarini vd., 2015; Venckauskas vd., 2017; Zois vd., 2018).

Balakrishnan vd. (2019) kişilik özelliklerini kullanarak siber zorbalığın tespitine yönelik bir çalışma gerçekleştirmişlerdir. Bu çalışmada Twitter'dan elde ettikleri 9484 tweet veri kümesinde tekrar, takipçi, takip, popülerlik, favori, durum ve hash sayısı şeklinde etiketlendirmeler yapmışlardır. Sınıflandırma için Rastgele Orman (RO) algoritması kullanılmıştır. Bu çalışmada kişilik verisinin siber zorbalık tespitinde önemli ölçüde iyileştirme sağladığını tespit etmişlerdir. Ayrıca dışa dönüklük, uyumluluk, nevrotiklik ve psikopati zorbaları tespit etmede % 96 kesinlik, % 95 hatırlama oralarında başarı elde etmişlerdir (Balakrishnan vd., 2019).

Bozyiğit vd. (2021) Siber zorbalık tespitine yönelik Twitter'dan elde ettikleri Türkçe 5000 etiketli içerik üzerinde metin bazlı tespitin yanında sosyal madya özelliklerinide dikkate alarak siber zorbalık tespiti gerçekleştirmişlerdir. Bu çalışmada veri seti üzerinde Ki-kare testi uygulanarak LR, RF, DVM, AdaBoost, NBM ve KNN algoritmaları kullanmışlardır. Yapılan karşılaştırmada en iyi sonucun DVM algoritması ile elde edildiğini tespit etmişlerdir (Bozyiğit vd., 2021).

Yılmaz vd. (2021) OffensEval veri seti üzerinde saldırgan dil tespitine yönelik bir çalışma gerçekleştirmişlerdir. Siber zorbalık tespitine yönelik Twitter'dan elde ettikler yaklaşık 1 milyon etiketsiz Türkçe tweet ile etiketli OffensEval veri setindeki kelime temsillerinin sınıflandırma performansına olan etkisi kıyaslanmıştır. Yapılan çalışmada Uzun Kısa Dönemli Bellek (LSTM) ve Çift Yönlü Uzun Kısa Dönemli Bellek (BiLSTM) ağları kullanmışlardır. Büyük veri kümelerinde Derin Sinir Ağları kullanımının F1 skorunda %40-%47 arasında iyileştirme elde edildiğini tespit etmişlerdir (Yılmaz vd., 2021).

Bozyiğit vd. (2019) Siber zorbalık tespitine yönelik Twitterden elde ettikleri 3000 tweet üzerinde tasarladıkları YSA modellerini uygulamış, bu modellerden YSA2 ile %91 F1 skoru ile en iyi başarımı elde etmişlerdir (Bozyiğit vd., 2019).
Balakrishnan vd. (2019) sosyal medya kullanıcılarının kişilik, duygu gibi psikolojik özelliklerinden yararlanarak siber zorbalık tespitine yönelik bir çalışma gerçekleştirmişlerdir. Twitter'dan elde ettikleri 5453 tweet ile bir veri seti oluşturmuş ve NB, RO ve J48 makine öğrenimi sınıflandırma algoritmalarını kullanmışlardır. Yapılan çalışmada kişilik verilerinin siber zorbalık tespiti üzerinde olumlu bir etki gösterdiği ancak duygular için aynı etkinin gerçekleşmediği görülmüştür. %92.88'lik başarı oranıyla J48 algoritmasının en iyi performansı verdiğini tespit etmişlerdir (Balakrishnan vd., 2019).

Modha vd. (2020) siber zorbalığı tespit ve görselleştirmeye yönelik bir çalışma gerçekleştirmişlerdir. Bu çalışmalarında kullandıkları verileri dört grupta etiketlendirmişlerdir. Siber zorbalık içeren yorumların görselleştirilmesi için ise Facebook ve Twitter üzerinden bir web eklentisi olarak kullanıcı arayüzü tasarlamışlardır. Bu eklenti ile standart Trolling Aggression Cyberbullying 2018 (TRAC) veri kümesi kullanılarak Facebook ve Twitter üzerinden yayınlanan yorumlar ile İngilizce ve Hintçe olarak yeni bir veri kümesi oluşturulmuştur. Sınıflandırmalar için Destek Vektör Makinesi (DVM), Lojistik Regresyon (LR), Evrişim Sinir Ağına (CNN) dayalı derin öğrenme modeli, Dikkat Temelli Model ve Google AI tarafından yakın zamanda önerilen BERT önceden eğitilmiş dil modeli gibi çeşitli sınıflandırıcılar kullanılmıştır. İngilizce ve Hintçe verilerde farklı sınıflandırma algoritmalarının başarımları gözlemlenmiştir (Modha vd., 2020).

MinSong ve Song (2020) Kore'de siber zorbalık eylemindeki rollerin tespitine yönelik bir çalışma gerçekleştirmişlerdir. Bu çalışmalarında haber sitelerinden, bloglardan, çevrim içi gruplardan, sosyal ağ hizmetlerinden ve buna benzer 227 çevrim içi kanaldan topladıkları 103212 veri ile bir veri seti oluşturmuşlardır. Veri madenciliği yöntemleri ve karar ağacı analizi kullanarak siber zorbalık eylemlerinde % 32.3'ü kurbanlar, % 6.4'ü failler ve % 5.3'ü izleyenlerden oluştuğunu tespit etmişlerdir. Ayrıca bu tür eylemlerde dürtü faktörünün oluşacak risk faktörü üzerinde önemli ölçüde etkiye sahip olduğu tespit edilmiştir (MinSong & Song, 2020).

Fortunatus vd. (2020) siber zorbalığın tespitine yönelik Facebook yorumlarında metinsel analizler yaparak elde edilen veri seti Lexicon gelişmiş kural tabanlı algoritma ile sınıflandırmışlardır. Kullanılan algoritmanın doğruluk, kesinlik, geri çağırma ve F1 skoru performans ölçümleri sonucunda % 95.981'lik bir başarım elde etmişlerdir (Fortunatus vd., 2020).

Agrawal ve Awekar (2018) siber zorbalık tespitine yönelik Formspring (~12k gönderi), Twitter (~16.000 gönderi) ve Wikipedia (~100.000 gönderi) sitelerinden elde ettikleri üç farklı veri seti üzerinde derin öğrenmeye dayalı modellerle bir çalışma gerçekleştirmişlerdir. Çeşitli geleneksel makine öğrenimi modelleri (LR, DVM, RO, Naive Bayes (NB)) ve derin sinir ağ modelleri (CNN, LSTM, BLSTM, BLSTM with Attention), kelimeler için temsil yöntemleri (n-gram karakter çantası, unigram kelime çantası, GloVe düğümleri, SSWE düğümleri) performansları karşılaştırılmış ve 0.95 F1 skoru ile derin öğrenme modelinin yüksek performaslara ulaştığı tespit edilmiştir (Agrawal & Aweka, 2018).

Hosseinmardi vd. (2015) siber zorbalık tespitinde instagramdan elde ettikleri görseller ve bu görsellere yapılan yorumlarda siber zorbalığın tespitine yönelik bir çalışma gerçekleştirmişlerdir. NB ve lineer DVM algoritmaları ile sınıflandırmada metin ve görsel verilerde lineer DVM algoritmasının %87 lik başarı elde ettiğini tespit etmişlerdir (Hosseinmardi vd., 2015).

N-Garci'a vd. (2015) siber zorbalığın tespitine yönelik Twitter'dan elde ettikleri veriler ile veri seti oluşturmuş, daha sonra sınıflandırma için RO, J48, K-Nearest Neighbor (KNN) ve Sequential Minimal Optimization (SMO) algoritmalarını kullanmışlardır. Sonuçlar, SMO ve Karar Ağaçlarının % 68.47 doğruluk oranıyla en uygun algoritmalar olduğunu göstermiştir (N-Garci´a vd., 2015).

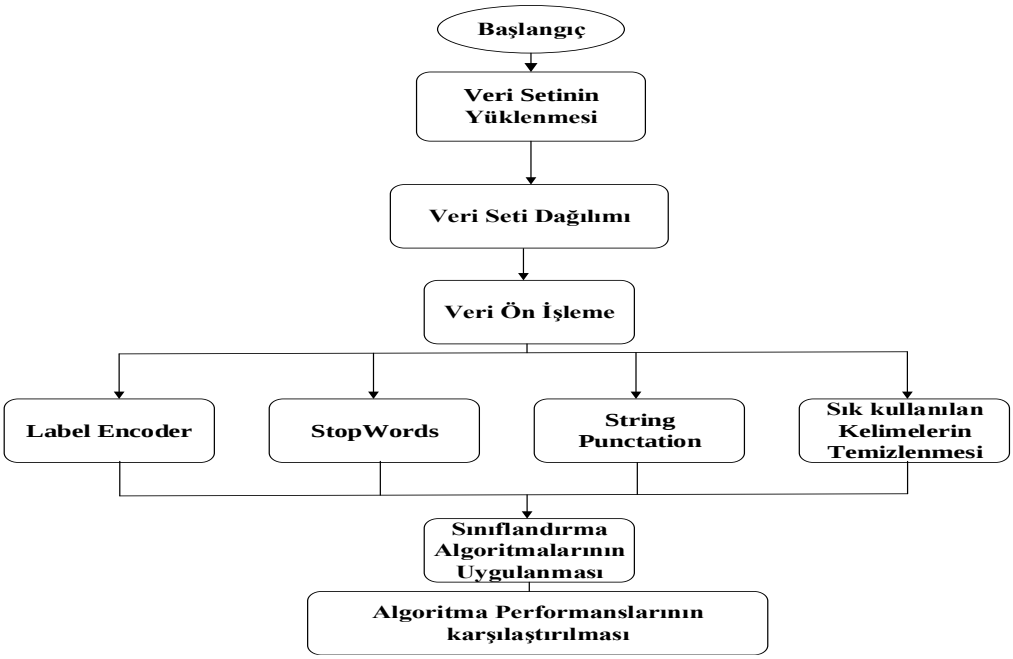
Dadvar vd. (2014) siber zorbalık tespitinde genellikle metin bazlı yapılan çalışmaların aksine siber zorbalığı gerçekleştiren bireylerin cinsiyet tespitine yönelik bir çalışma gerçekleştirmişlerdir. MySpace'den elde ettikleri veri seti üzerinde DVM sınıflandırma algoritması kullanılarak ve 10 kat çaprazlama sonucunda %40-%44 doğruluk oranları elde edilmiştir (Duwairi vd., 2014).

Squicciarini vd. (2015) The Mypace ve Formspring'den elde ettikleri veri setlerini kullanarak, sosyal ağlarda siber zorbalık tespiti ile siber zorbalar ve kullanıcılar arasındaki ikili etkileşimleri tanımlamaya yönelik bir çalışma gerçekleştirmişlerdir (Squicciarini vd., 2015). Bir benzer çalışma da Al-Mamun ve Akhter'in bangle dili kullanılarak siber zorbalığın tespit edilmesine yönelik olan çalışmalarıdır. Yapılan çalışmada Twitter'dan elde edilen İngilizce ve Banglece iki veri kümesi NB, DVM, J48 ve KNN algoritmaları ile sınıflandırılarak bu algoritmaların performansları karşılaştırılmış ve DVM algoritmasının her iki dil yapısında da en iyi sonucu elde ettiği tespit edilmiştir (Al-Mamun & Akhter, 2018).

2. Materyal ve metot
3. Material and method

3.1. Önerilen model
3.1. Recommended model

Siber zorbalığın tespitine yönelik izlenecek yol ve önerilen modele ait algoritma, Şekil 1'de gösterilmiştir.



Şekil 1. Önerilen modele ait algoritma
Figure 1. *Algorithm of the proposed model*

Bu çalışmada kullanılan doğal dil işleme, yapay zekânın bir alt kategorisidir. Doğal dil işleme yöntemleri günlük konuşma dillerinin makine dillerinin algılayabileceği yapıya dönüştürülmesi için kullanılan yöntemlerin bütününü temsil etmektedir. Bu bağlamda çalışmada kullanılan yöntemler keşifsel veri analizi, veri ön işleme ve makine öğrenmesi algoritmalarının kullanılması olarak üç kısma ayrılmıştır.

3.2. Veri setinin dağılımı
3.2. Distribution of the dataset

Bu çalışmada kullanılan veri seti 1497 adet negatif (siber zorbalık içermeyen), 1503 adet pozitif (siber zorbalık içeren) ve toplamda 3000 satırdan oluşmaktadır. Veriler iki sütundan oluşan xlsx tabloya aktarılmıştır. İlk sutun mesajın kendisini, ikinci sutun ise bu mesajın siber zorbalık içerip içermediğini gösteren ikili bir tam sayıdır. Kaggle

adlı paylaşım sitesinden alınan bu hazır veri setinde pozitif ve negatif etiketlendirmeler manuel olarak gerçekleştirilmiştir (Kepez, 2021). Kullanılan veri setinin bir kısmı Şekil 2'de görülmektedir.

		message	cyberbullying
2	1.0	rabbim kalan ömrünü geçen ömründen hayırlı eylesin	0.0
3	2.0	bir ateist olarak bu resmi gördükçe gözyaşlarıma mani olamıyorum	0.0
4	3.0	oo süpersin azıcık bize de bulaşsa	0.0
5	4.0	bende biliyorum benden bı bok olmicak	1.0
6	5.0	nerdesin len tirrek	1.0
7	6.0	dogruyusoyleyince kadro verince adalet yerini bulacak	0.0
8	7.0	ogrenciler hocalarına memleket durusuna hayran	0.0
9	8.0	gunaydin hala yatakta olanlar eminim elinizde telefon aletinizde tad gibidir hadi sanalda sabah keyfi yapalim	0.0
10	9.0	artist pezevenk takimi yakti rezildi	1.0
11	10.0	ve o narinlikte ve sevimlilikte cüce öylesine acımasızca zibh eğliyip	0.0
12	11.0	off defol git surdan cüce	1.0
13	12.0	lüks yaşam bu muulan gerizeka lüks hayatları olsa o poşetleri taşıyan hizmetçileri olurdu	1.0
14	13.0	beştepe millet kültür ve kongre merkezi nden elektrik santralleri toplu açılış törenine telekonferansla bağlandı cumhurbaşkanı	0.0
15	14.0	taehyung ne bulsan giyiyon ama yakışıyor ha ibne 1 namjoon 2 sen kezban gibi giyinsenizde aşığım	1.0
16	15.0	hava gavur şeyi gibi yanıyor diyorlar ama o konuda hiç tecrübem yok bilemiyorum	1.0

Şekil 2. Kullanılan veri setinden bir bölüm
Figure 2. *A section from the data set used*

3.3. Veri ön işleme
3.3. Data preprocessing

3.3.1. Label encoder yapılması
3.3.1. Making a label encoder

Label encoder veriyi bire bir sayısallaştırmaya yarar. Bu uygulamanın amacı makine öğrenmesi algoritmaları için uygun girdilerin oluşturulmasıdır. Bu çalışmada kullanılan veri setinde paylaşımların siber zorbalık durumunu belirten "Pozitif" ve "Negatif" adı altında iki farklı kategorik değer bulunmaktadır. Buradaki özniteliği belirten kategorik değişkenlerin sayısallaştırılarak 0-1 değerlerine dönüştürülmesi manuel olarak veri setinde hazır olarak gerçekleştirilmiştir.

3.3.2. Etkisiz kelimelerin veri setinden çıkarılması (Stopwords)
3.3.2. Removal of ineffective words from the dataset (Stopwords)

Etkisiz kelimeler genel olarak bir dilde çok sık kullanılan (Türkçede: "bir", "bu", "şu", "ne", "nasıl" gibi) ve bu nedenle hem arama motorları hem de makine öğrenmesi algoritmalarında gözardı edilen kelimelerdir. Söz konusu kelimelerin veri setinden çıkarılmasının temel nedenleri hemen her cümlede kullanılmalarının yanında cümleye anlamsal bir etki sağlamaması, buna bağlı olarak yapılan analizlerde iş yüküne neden olması ve hatta algoritmaları negatif yönde etkileyip daha isabetsiz bir sınıflandırma yapmalarına sebep olmalarıdır. Bu amaç doğrultusunda etkisiz verilerin veri setinden çıkarılması için Türkçe dil yapısına uygun oluşturulmuş ve Python dilinde NLP (doğal dil işleme) modülü olan NLTK (Natural Langauge ToolKit) kütüphanesiyle birlikte hazır bir fonksiyon olan Stopwords fonksiyonu kullanılmıştır.

3.3.3. Özel karakterlerin temizlenmesi
3.3.3. Clearing special characters

Kullanılan veri seti içerisinde bulunan '"!()-[]{};':'"\,<>./?@#$%^&*_~'" gibi noktalama işaretleri ve özel karakterler makine öğrenmesi algoritmalarını negatif yönde etkilemesi nedeniyle veri setinden temizlenmesi işlemi gerçekleştirilmiştir.

3.3.4. Sık kullanılan kelimelerin temizlenmesi
3.3.4. Cleaning up frequently used words

Yine etkisiz kelimelerde olduğu gibi bir veri seti içerisinde sık kullanılan ve herhangi bir önem arz etmeyen kelimelerin veri seti içerisinden temizlenmesi işlemleri gerçekleştirilmiştir.

3.3.5. Paylaşımlarda bulunabilecek emojilerin kaldırılması
3.3.5. Removing emojis that can be shared

Sosyal paylaşım sitelerinde yapılan paylaşımların çoğunda emojiler kullanılmaktadır. Bu noktada yapılan çalışmanın daha kapsamlı bir şekilde analizinin gerçekleştirilmesi ve daha sağlıklı sonuçların elde edilmesi için emojilerin cümle içerisinde ne amaçla kullanıldığının tespit edilmesi ve bunlara uygun kelime karşılıklarının veri setine eklenmesi gerekmektedir. Ancak bu çalışmada kullanılan veri seti manuel olarak oluşturulduğu ve emojiler içermediğinden dolayı veri setinde buna yönelik bir çalışma gerçekleştirilmemiştir.
Sonuç olarak metin üzerinde aşağıdaki adımlar bir fonksiyon içinde tanımlanarak adım adım gerçekleştirilmektedir.

- İlk etapta veri seti içerisindeki harflerin tümü küçük harfe dönüştürülmektedir.
- Özel karakterler, Türkçe karakter karşılıkları ile değiştirilmektedir.
- Veri seti içerisindeki noktalama işaretleri ve rakamlar temizlenmektedir.
- Veri setindeki metinler kelimelere ayrılarak kelime listesine dönüştürülmektedir.
- Etkisiz kelimeler veri setinden silinmektedir.
- Veri setinde kullanım sıklıklarına göre sıralanarak oluşturulan ve maksimum 5000 kelime dağarcığına sahip kelime listesine dönüştürülen metin geri çağırılmaktadır.

Bu adımlardan sonra veri setinin güncelleştirilmesi işlemi gerçekleştirilmiştir. Daha sonra veri seti test ve eğitim kümelerine ayrılmıştır. Bu çalışmada en iyi sonuçlar, yapılan denemeler sonucunda veri setinin %25'inin test, %75'inin eğitim için ayrılmasıyla elde edilmiştir.

Test ve eğitim kümeleri ayrımı yapıldıktan sonra sınıflandırma algoritmaları veri seti üzerinde uygulanmıştır.

3.4. Makine öğrenmesi
3.4. Machine learning

Bu çalışmada kullanılan makine öğrenme algoritmalarının kısa açıklaması ve kullanım amaçları kısaca bu bölümde açıklanmıştır.

3.4.1. Rastgele Orman algoritması
3.4.1. Random Forest algorithm

Sınıflandırma algoritmaları içinde en fazla kullanılan algoritmadır. Bunun sebebi ise bu yöntem ile birçok sınıflandırma ağacı içerisinde rastgele seçilen bir altküme yardımıyla yeni topluluk oluşturulabilmesidir. RO algoritması kategorik, sürekli ve her iki yapıdaki veri setlerinde, ayrıca farklı büyüklükteki boyutlara sahip veri setlerinde uygulanabilmektedir. Bu gibi avantajlı yönlerinden dolayı siber zorbalık tespitlerinde sıklıkla kullanılmaktadır. Bunun yanında bu yöntemin dezavantajı sınıflandırma ağaçları yönteminde olduğu gibi çıktı olarak bir ağaç oluşturulmamasıdır (Akman vd., 2011; Balakrishnan vd., 2020; Breiman, 2001).

3.4.2. C- 4. 5 algoritması
3.4.2. C- 4.5 algorithm

Bu algoritma yapılacak sınıflandırma için belirlenmiş eğitim verileri üzerinde oluşturduğu bir karar ağacı ile gelen girdi verilerinin hangi sınıflara ait olduğunu tahmin etmektedir. Bu yöntem karar ağacını oluştururken verinin hangi sınıfa ait olduğuna dair tahmini sınıfa ait kazanımları yani o sınıfa ait ayırt edici özellikleri öncelik sırasına koyarak ağaç yapısını oluşturmaktadır. Bu algoritma, sınıflandırmada yapılacak ayrımın ayırt edici özelliklere göre gerçekleştirilmesi, yapılacak tahminlerde kural çıkarımlarının yapılıyor olması ve kayıp veri yoğunluğu fazla olan veri setlerinde başarılı sonuçlar elde etmesi yönüyle siber zorbalık tespitlerinde kullanılmaktadır (Quinlan, 1996).

3.4.1. Destek Vektör Makinesi
3.4.1. Support Vector Machine

Bu algoritmanın temel prensibi, farklı iki sınıfa ait verileri doğrusal veya doğrusal olmayan şekilde birbirinden ayıracak sonsuz sayıda vektörler oluşturarak en uygun ayırımı gerçekleştirmektir. Bu yöntem çok büyük verilerde kullanılmakta ve hızlı sonuçlar elde edilmektedir. Yine verilerin ayrıştırılmasının doğrusal veya doğrusal olmayan şekillerde yapılabilmesi, yapılacak sonsuz ayrımlar içerisinde ise en iyisini seçebilmesi yönüyle siber zorbalık tespitinde en fazla kullanılan algoritmadır (Aydın, 2018).

3.4.4. Naive Bayes
3.4.4. Naive Bayes

Bu algoritma istatistik temeline sahip bir denetimli öğrenme algoritmasıdır (McCallum & Nigam, 1998). Kullanılacak metin tabanlı belgelerde sınıflandırma gerçekleştirilirken tüm eğitim veri kümesi üzerinde koşullu olasılıklar hesaplanarak gerçekleştirilmektedir. Bu algoritmanın en önemli avantajı uygulanmasının kolay olmasının yanında iyi sonuçlar elde edilebiliyor olmasıdır (Saravanaraj vd., 2016).

3.4.5. Lojistik Regresyon
3.4.5. Logistic Regression

Bu yöntem ile mevcut veriler dikkate alınarak oluşabilecek bir durumun olasılığı tahmin edilmektedir. Burada bir değişkenin bağımlılığının birden fazla olması durumunda elde edilecek sonuçlar 0 ve 1'e indirgenerek gösterilmektedir. Siber zorbalık tespitlerinde zorbalık kategorilerinin ayrımı veya ilişkilendirmelerin yapılması gereken sınıflandırmalarda sıklıkla kullanılmaktadır (Ayo vd., 2020).

3.4.6. Yapay Sinir Ağları
3.4.6. Artificial neural networks

Yapay Sinir Ağları (YSA), insan beyninin en temel özelliği olan öğrenme fonksiyonunu gerçekleştiren

bilgisayar sistemleri olarak tanımlanabilir. YSA, mevcut veri yapısının öğrenilmesi ve öğrenilen bu yapı üzerinden genellemeler yaparak sonuca ulaşmayı hedeflemektedir. Ağın ilgili olay örnekleriyle eğitilmesi sayesinde genellemeler gerçekleştirilmektedir. Bu yönüyle de benzer olaylar karşısında oluşabilecek durum verileri tespit edilmektedir. Tespit edilen sonucun, ağa gelecek yeni verilerin ağırlık değerleri ile çarpılıp toplanmasıyla yeni bir veri elde edilmektedir. Elde edilen bu yeni veri ise bir fonksiyon aracılığıyla işlenmekte ve bu işlenen veri de çıkış katmanından alınarak elde edilmektedir. En iyi çıkış verisinin elde edilebilmesi için söz konusu ağırlık değerlerinin sürekli güncellenmesi gerekir. Bu noktada ağın bir dezavantajıyla karşı karşıya gelinmektedir. Bu dezavantaj, ağırlık değerlerinin kullanıcı tarafından anlamlandırılmasına kapalı olmasıdır. Yapay sinir ağı içindeki bilginin istenilen hedefe en yakın sonucu, kullanıcıların belirlemiş oldukları katman ve nöronlarda gizlidir. Temel olarak YSA deneyimlemeler ve örneklemeler yaparak öğrenmeyi gerçekleştirir. Bu sayede girdiler arasındaki ilişkilerin tespitinin zor olduğu veya büyük veri kümelerinde doğrusal olmayan ilişkilerin modellenmesinde kullanılmaktadır. YSA, görüntü tanımlanması, doğal dil işleme, ses tanıma, büyük veri analizlerinde ve buna benzer birçok alanda kullanılmaktadır. Bu yönüyle siber zorbalık tespitinde de etkin olarak kullanılmaktadır (Atalay & Çelik, 2017).

3.4.7. SGD (Stochastic Gradient Descent)
3.4.7. SGD (Stochastic Gradient Descent)

SGD algoritması, sınıflandırmalar yaparken kullanılan ağırlık değerlerinin değiştirilmesinde tüm eğitim verilerinin üzerinde işlem yapmak yerine sadece rastgele seçilen bir örnek dikkate alır. Sadece bir nokta dikkate alındığından bu algoritma ile daha hızlı sonuçlar elde edilmektedir. Bu yönüyle de metin tabanlı sınıflandırmalarda ve doğal dil işlemede büyük ölçekli ve seyreltilmiş veri setlerine uygulanmaktadır (Chandrashekhar & Raghuveer, 2014).

3.4.8. K-En Yakın Komşu (KNN)
3.4.8. K-Nearest Neighbor (KNN)

Bu algoritma sınıflandırmayı gerçekleştirirken, sınıflandırılacak yeni verinin, daha önce sınıflandırılmış k adet veriye olan uzaklıklarının hesaplanarak en yakın uzaklığa sahip olan sınıfa dahil edilmesi mantığına göre çalışmaktadır. Karşılaştırmada yeni verinin komşuluk mesafesinin hesaplanmasında genellikle Öklid Bağıntısı kullanılmaktadır. Hesaplamalar sonucunda yapılacak tahminler belirlenirken komşu sınıfların örnek sayılarının çokluğu dikkate alınmaktadır. Bu yöntem, eğitim aşamasının olmayışı ve gürültü verilerine karşı dayanıklı olması nedeniyle siber zorbalık tespitinde tercih edilen bir sınıflandırma algoritmasıdır (Aydın, 2018).

3.4.9. K-Ortalama Kümeleme (K-Means)
3.4.9. K-Means Clustering (K-Means)

Kümeleme algoritmalarının en eskisi olan bu algoritma veri madenciliğinde de en çok tercih edilen algoritmalardan biridir. İstatistiksel olarak benzer özelliğe sahip olan verilerin kümelenmesi gerçekleştirilmektedir. Kümeleme işlemleri gerçekleştirilirken bir veri sadece bir kümenin elemanı olacak şekilde ayrımlar yapılır. Algoritmadaki temel amaç, oluşturulacak "K" adet kümeden her kümenin birbirinden olabildiğince farklı olmasını sağlamak ve bu sayede her bir kümeye ait verilerin birbirlerine yakın olmalarını sağlamaktır (Chandrashekhar & Raghuveer, 2014).

3.4.10. Yükseltme
3.4.10. Boosting

Bu yöntemdeki amaç, kullanılacak verilere farklı ağırlık değerleri uygulayarak elde edilecek ağaç yapılarından yeni çıkarımların gerçekleştirilmesidir. İlk etapta yapılan gözlemlerde her veri eşit ağırlık değerlerine sahiptir. Ağaç yapısı büyüdükçe belirlenen modele göre ağırlık değerleri verilmeye başlanır. Burada her yanlış sınıflandırılan verinin ağırlık değeri arttırılır ancak nadiren ağırlık değerleri azaltılabilir. Doğru sınıflandırma gerçekleştirilinceye kadar farklı modellemeler gerçekleştirilir ve son model ile kullanılan modellerin ağırlık ortalaması alınarak oluşturulur. Buradaki amaç ağaç yapısında oluşacak zor durumlarda ağacın kendi kendini düzenleyebilmesini sağlamaktır (Quinlan, 1996).

3.4.11. Torbalama
3.4.11. Bagging

Bu yöntem sınıflandırma ve regresyon problemleri için kullanılan öğrenme tekniklerinin verimliliğini ve doğruluğunu arttırmak için tasarlanmış kolektif bir öğrenme modelidir. Torbalama yönteminde çoğunlukla karar ağaçları kullanılmaktadır. Bu yöntemde n adet veri içeren örneklemden boyutları n / k olan k tane yeni veri kümesi oluşturulur. Üretilen her bir veri kümesi için farklı öğrenme modelleri kullanılarak sınıflandırma işlemi

yapılmaktadır. Bu yöntemin avantajları varyansı azaltması ve aşırı uyumu engellemesidir (Sheikhi, 2020).

3.4.12. Gradyan Arttırma (GBM)
3.4.12. Gradient Boost (GBM)

Bu algoritma bir topluluk algoritması olup regresyon ve sınıflandırma çözümlemelerinde kullanılmaktadır. Bu algoritmada güçlü bir öğrenme yetisini kazanmak amacıyla bazı zayıf öğrenimler birleştirilmektedir. Burada asıl oğrenmeyi gerçekleştiren regresyon ağaçları esasında her biri bir önceki ağaç tarafından hesaplanan hatalar üzerine kurulu bir dizi ağaç yapısından oluşmaktadır (Callens vd., 2020).

3.4.13. XGBoost
3.4.13. XGBoost

Bu algoritma karar ağacı tabanlı olup, paralel veri işleme, ağaç yapısındaki gereksiz verileri atma, eksik olan verileriin işlenebilmesi ve aşırı sapmaların (overfitting) önlenmesine yönelik düzenlemelerle optimize edilmiş GBM algoritmasıdır. Algoritmanın sınıflandırmada kullanılmasının temel sebebi, yüksek tahmin gücüne sahip olması, aşırı öğrenimin önüne geçebilecek bir yapıya sahip olması, içeriği olmayan verilerin yönetimini yapabilmesi ve bu işlemleri çok hızlı bir şekilde gerçekleştirebilmesidir (Bardina vd., 2020).

3.5. Sınıflandırma algoritmalarının kodlanması
3.5. Coding of classification algorithms

Bu çalışmada literatürde sıklıkla kullanılan sınıflandırma algoritmalarından Bagging, Boosting, C4.5, Gradient Boosting, K-Means, KNN, LR, NB, YSA, RO, DVM, Stochastic Gradient Descent, XGBoost kullanılmıştır. Kodlama gerçekleştirilirken belirlenen modelin eğitilmesi ve performans değerlerinin test edilebilmesi için veri setinin bölümlenmesi gerekmektedir. Burada bu işlem için scikit-learn kütüphanesi kullanılmaktadır. Öte yandan veri seti içerisindeki benzersiz kelimelerin belirlenmesi ve özellik çıkarımının yapılması gerekmektedir. Bu da kullanılacak makine öğrenme algoritmalarının daha sağlıklı bir şekilde eğitilebilmesi amacıyla gereklidir. Bu amaç doğrultusunda yapılan çalışmada Bag of Words (BOW) algoritması kullanılmaktadır. Bu algoritma veri setindeki cümleleri kelimelere ayırarak benzersiz kelimelerden oluşan bir kelime dağarcığı vektörü oluşturmaktadır. Oluşturulacak kelime dağarcığı max_features değerinin 5000 olarak girilmesiyle sınırlandırılmıştır. Bu sayede veri setindeki kelimelerin kullanım sıklıkları sıralamasına göre maksimum 5000 kelimeden oluşan bir kelime dağarcığı elde edilmektedir. Buradaki amaç, veri setinin tamamında gezinebilmek ve bu sayede tüm kelimeleri, kelime haznesine katmaktır. Ayrıca veri setindeki cümleler, kelime dagarcığındaki kelimelerin sayısına bağlı olarak temsil edilebilir. Bu işlemler için de scikit-learn kütüphanesinin CountVectorizer nesnesi kullanılmaktadır. Elde edilen eğitim verileri ile model eğitilmekte ve sınıflandırma algoritmaları test edilmektedir.

4. Sonuçlar ve tartışma
4. Results and discussion

Bu çalışmada Türkçe bir hazır veri seti kullanılarak siber zorbalık tespiti problemi ele alınmıştır. Bu amaç doğrultusunda veri seti üzerinde doğal dil işleme yöntemleri kullanılarak veri ön işlemleri gerçekleştirilmiştir. 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. Burada LR algoritmasının çalışma yapısı ve kullanmış olduğumuz veri setinin de bu yapıya uygun iki sınıf değişkenine sahip olmasından ötürü en iyi sonucu elde ettiği düşünülmektedir. Bu çalışmada tüm sınıflandırma algoritmalarından elde edilen sonuçlar Tablo 1'de gösterilmiştir.

Tablo 1'de yer alan Kesinlik (Precision) = TP / (TP + FP), Hassasiyet (Recall) = TP / (TP + FN), Doğruluk Accuracy = (TP + TN) / (TP + FP + TN + FN) ve F1 Puanı = F1-score = 2 * Precision * Recall / (Precision + Recall) denklemleri ile hesaplanmıştır. Denklemlerde geçen TP, Doğru Pozitif; TF, Doğru Negatif; FP, Yanlış Pozitif; FN, Yanlış Negatif taminleri ifade etmektedir. Burada Algoritmaların yapmış olduğu tahminlerde eğer siber zorbalık içeren bir kelime doğru tahmin edilmiş ise TP değeri 1 arttırılır, siber zorbalık içermeyen kelime doğru tahmin edilmiş ise TN değeri bir arttırılır, Eğer Siber zorbalık içeren kelime yanlış tahmin edilmiş ise FP değeri 1 arttırılır, siber zorbalık içermeyen kelime yanlış tahmin edilmiş ise FN değiri bir attırılmakta ve söz konusu denklemler bu değerler üzerinden hesaplanmaktadır (Medium, 2020).

Tablo 1. Sınıflandırma algoritmalarının performans sonuçları
Table 1. Performance results of classification algorithms

Algoritmalar	Doğru Pozitif Oranı	Yanlış Pozitif Oranı	Kesinlik (Precision)	Hassasiyet (Recall)	Doğruluk (Accuracy)	F1 Puanı
Bagging	0.7643	0.0808	0.9353	0.7643	0.8256	0.8412
Boosting	0.8293	0.0909	0.9164	0.8293	0.8655	0.8707
C4.5	0.8333	0.1274	0.876	0.8333	0.8522	0.8541
GBM	0.622	0.0761	0.9623	0.622	0.6953	0.7556
KNN	0.5896	0.0657	0.9757	0.5896	0.6525	0.735
K-Means	0.5023	0.4388	0.8841	0.5023	0.51	0.6406
LR	0.8775	0.1104	0.8893	0.8775	0.8835	0.8833
NB	0.872	0.3054	0.5876	0.872	0.7537	0.7021
YSA	0.8414	0.1142	0.8895	0.8414	0.8628	0.8654
RO	0.7986	0.0701	0.9407	0.7986	0.8535	0.8639
SGD	0.8545	0.1148	0.8868	0.8545	0.8695	0.8704
DVM	0.8029	0.2385	0.7383	0.8029	0.7804	0.7692
XGBoost	0.7016	0.1132	0.9191	0.7016	0.767	0.7958

Literatürde yapılan çalışmalar incelendiğinde birçok çalışmada kullanılan algoritmalar arasında DVM algoritmasının en iyi sonucu verdiği görülmüştür (Al-Mamun & Akhter, 2018; Duwairi vd., 2014; Hosseinmardi vd., 2015; Hussain vd., 2018; Shekhar & Mathangi, 2018; Venckauskas vd., 2017; Yazğılı & Baykara, 2021). Bu çalışmaların çoğunda İngilizce veri setleri kullanılmasının yanı sıra farklı dil yapılarında oluşturulmuş veya elde edilmiş veri setleri kullanılmıştır. Türkçenin morfolojik açıdan zengin bir dil olması ve eklemeli bir dil olması nedeniyle doğal dil işleme çalışmaları İngilizce gibi dillere oranla daha zordur. Bu alanda yapılan çalışmaların sayısının az olmasının bir sebebi de bundan kaynaklanmaktadır (Eryiğit ve Torunoğlu, 2017; Özer vd., 2018). Türkçe bir veri seti üzerinde gerçekleştirilen bu çalışmada ise Lojistik Regresyon sınıflandırma algoritmasının 13 algoritma içerisinde en yüksek başarıyı elde ettiği tespit edilmiştir. Bu çalışma esasında farklı diller arasında sınıflandırma algoritmalarının performanslarını kıyaslamaktan ziyade literatürde Türkçe veri seti kullanılarak yapılan çalışmaların çok az olması nedeniyle gerçekleştirilmiştir (Bozyiğit vd., 2021; Bozyiğit vd., 2019; Yılmaz vd., 2021; Bozyiğit vd., 2019; Öztürk, E., 2019). Ancak yine de bu çalışma ile genel olarak yapılan farklı dillerdeki çalışmalar karşılaştırıldığı zaman siber zorbalık tespitinde kullanılan makine öğrenme algoritma performanslarının kullanılan veri setindeki dil yapısına bağlı olarak da farklı sonuçlar verebileceğini göstermektedir. Ayrıca yapılan literatür çalışmalarında daha önce bu kadar çok sayıda algoritma karşılaştırılmasının da yapılmadığı görülmüştür.

Gelecekte siber zorbalık tespitinde yapılan metin bazlı araştırmalarda daha sağlıklı ve güvenilir sonuçların elde edilmesine yönelik metinlerin anlamsal boyutlarının da dikkate alınarak değerlendirilebileceği derin öğrenme algoritmalarının sosyal ağlar üzerinde anlık siber zorbalık tespitine yönelik çalışmalar yapılması planlanmaktadır.

Yazar katkısı
Author contribution

Yazarlar çalışmanın tüm aşamalarında ortak çalıştıklarını beyan etmektedirler.

Etik beyanı
Declaration of ethical code

Bu makalenin yazarları, bu çalışmada kullanılan materyal ve yöntemlerin etik kurul izni ve / veya yasal-özel izin gerektirmediğini beyan etmektedir.

Çıkar çatışması beyanı
Conflicts of interest

Yazarlar herhangi bir çıkar çatışması olmadığını beyan eder.

Kaynaklar
References

Agrawal, S., & Awekar, A. (2018). Deep Learning for Detecting Cyberbullying Across Multiple Social Media Platforms. *Advances in Information Retrieval*, 141–153, https://doi.org/10.1007/978-3-319-76941-7_11

Akman, M., Genç, Y., & Ankaralı, H. (2011). Random forests yöntemi ve sağlık alanında bir uygulama,

Türkiye Klinikleri Journal of Biostatistics, *3 (1)*, 36-48.

Al-Mamun, A., & Akhter, S. (2018). Social media bullying detection using machine learning on bangla text. *10th International Conference on Electrical and Computer Engineering*, 20-22, https://doi.org/10.1109/icece.2018.8636797

Atalay, M., & Çelik, E. (2017). Büyük veri analizinde yapay zekâ ve makine öğrenmesi uygulamaları", *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi 9(22),*155-172.

Aydın, C. (2018). Makine öğrenmesi algoritmaları kullanılarak itfaiye istasyonu ihtiyacının sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi 5(14*), 169-175, https://doi.org/10.31590/ejosat.458613

Ayo, F. E., Folorunso, O., Ibharalu, F. T., & Osinuga, I. A. (2020). Machine learning techniques for hate speech classification of twitter data: State-of-the-art, future challenges and research directions. *Computer Science Review*, *38*, 100311, https://doi.org/10.1016/j.cosrev.2020.100311

Barlet, C. P. (2019). Cyberbullying, traditional bullying, and aggression: a complicated relationship. *Predicting Cyberbullying Research, Theory, and Intervention*, 11-16, https://doi.org/10.1016/B978-0-12-816653-6.00002-9

Balakrishnan, V., Khan, S., Fernandez, T., & Arabnia, H. R. (2019). Cyberbullying detection on twitter using big five and dark triad features. *Personality and Individual Differences, 141*, 252-257, https://doi.org/10.1016/j.paid.2019.01.024

Bozyiğit, A., Utku, S., & Nasibov, E. (2021). Cyberbullying detection: Utilizing social media features. *Expert Systems with Applications*, 179, 115001. https://10.1016/j.eswa.2021.115001

Bozyigit, A., Utku, S., & Nasiboglu, E. (2019). Cyberbullying Detection by Using Artificial Neural Network Models. 2019 *4th International Conference on Computer Science and Engineering (UBMK).* 520-524, https://10.1109/ubmk.2019.8907118

Balakrishnan, V., Khan, S., & Arabnia, H. R. (2020). Improving cyberbullying detection using twitter users' psychological features and machine learning. *Computers & Security, 90, 101710*, https://doi.org/10.1016/j.cose.2019.101710

Bardina, M., Vaganov, D., & Guleva, V. (2020). Socio-demographic features meet interests: on subscription patterns and attention distribution in online social media. *Procedia Computer Science, 178*, 162–171, https://doi.org/10.1016/j.procs.2020.11.018

Breiman, L. (2021). Random Forests, machine learning. *45* (1), 5-32, https://link.springer.com/article/10.1023/A:1010933404324

Callens, A., Morichon, D., Abadie, S., Delpey, M., & Liquet, B. (2020). Using Random Forest and gradient boosting trees to improve wave forecast at a specific location. *Applied Ocean Research, 104, 10233*, https://doi.org/10.1016/j.apor.2020.102339

Chandrashekhar, A. M., & Raghuveer, K. (2014). Amalgamation of K-means Clustering Algorithm with Standard MLP and SVM Based Neural Networks to Implement Network Intrusion Detection System. *Smart Innovation, Systems and Technologies*, *2*(28), 273–283, https://doi.org/10.1007/978-3-319-07350-7_31

Dadvar, M., De Jong, F. M. G., Ordelman, R. J. F., & Trieschnigg, R. B. (2012). Improved cyberbullying detection using gender information. *In Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop*, 23-25, http://dir2012.intec.ugent.be/system/files/proceedings/DIR2012_04_Maral_Dadvar.pdf

Duwairi, R. M., Marji, R., Sha'ban, N., & Rushaidat, S. (2014). Sentiment Analysis in Arabic tweets. 2014 5th *International Conference on Information and Communication Systems (ICICS)*, https://doi.org/10.1109/iacs.2014.6841964

Eryiğit, G., & Torunoğlu-Selamet, D. (2017). Social media text normalization for Turkish. *Natural Language Engineering*, *23*(06), 835–875. https://10.1017/s1351324917000134

Fortunatus, M., Anthony, P., & Charters, S. (2020). Combining textual features to detect cyberbullying in social media posts. *Procedia Computer Science*, *176*, 612-621, https://doi.org/10.1016/j.procs.2020.08.063

Hosseinmardi, H., Mattson, S. A., Ibn Rafiq, R., Han, R., Lv, Q., & Mishra, S. (2015). Detection of cyberbullying ıncidents on the ınstagram social network. *arXiv: 1503.03909v1 [cs.SI] 12 Mar 2015* https://arxiv.org/abs/1503.03909

Hussain, M. G., Mahmud, T. A., & Akthar, W. (2018). An Approach to Detect Abusive Bangla Text. *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, 27-29, https://doi.org/10.1109/ciet.2018.8660863

Kepez, T. B. "Detection of Cyberbullying in Turkish", Erişim adresi https://www.kaggle.com/tbrknt/detection-of-cyberbullying-in-turkish

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. *in AAAI-98 workshop on learning for text categorization*, *752*, 41-48.

Modha, S., Majumder, P., Mandl, T., & Mandalia, C. (2020). Detecting and visualizing hate speech in social media: a cyber watchdog for surveillance. *Expert Systems with Applications*, *161*, 113725, https://doi.org/10.1016/j.eswa.2020.113725

N-Garci´a, P. G., De La Puerta, J. G., Go´Mez, C. L., Santos, I., & Bringas, P. G. (2015). Supervised machine learning for the detection of troll profiles in twitter social network: application to a real case of cyberbullying. *Logic Journal of IGPL, jzv048, 24(1),* https://doi.org/10.1093/jigpal/jzv048

Ozer, Z., Ozer, I., & Findik, O. (2018). Diacritic restoration of Turkish tweets with word2vec. *Engineering Science and Technology, an International Journal*, *21*(6), 1120-1127, https://10.1016/j.jestch.2018.09.002

Öğündür, G. (2019, November 09*). Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall) ya da F1 Score?* https://medium.com/@gulcanogundur/do%C4%9Fruluk-accuracy-kesinlik-precision-duyarl%C4%B1l%C4%B1k-recall-ya-da-f1-score-300c925feb38

Öztürk, E. (2019). Cyberbullyıng detectıon usıng text classıfıcatıon for turkısh language. [Yüksek lisans Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı]

Quinlan, J. R. (1996). Bagging, boosting, and c4.5. in Proceedings of the Thirteenth National Conference on Artificial Intelligence, 1, 725–730.

Saravanaraj, A., Sheeba, J. I., & Pradeep Devaneyan, S. (2016). Automatic detection of cyberbullying from twitter", *IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS)*, ISSN: 2249-9555, *6*(6), 26-31.

Sheikhi, S. (2020). An efficient method for detection of fake accounts on the ınstagram platform. *International Information and Engineering Technology Association*, 429-436, https://doi.org/10.18280/ria.340407

Shekhar, A., & Mathangi, V. (2018). A Bag-of-phonetic-codes modelfor cyber-bullying detection in twitter. *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 1-7, https://doi.org/10.1109/icctct.2018.8550938.

Song, T.-M., & Song, J. (2021). Prediction of risk factors of cyberbullying-related words in Korea: Application of data mining using social big data. Telematics and Informatics, 58, 101524, https://doi.org/10.1016/j.tele.2020.101524

Squicciarini, A., Rajtmajer, S., Liu, Y., & Griffin, C. (2015). Identification and characterization of cyberbullying dynamics in an online social network. *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '15, 280-285, http://dx.doi.org/10.1145/2808797.2809398

Türkiye İstatistik Kurumu. (2020). *Hanehalkı Bilişim Teknolojileri (BT) Kullanım Araştırması, 2020*. Erişim adresi https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2020-33679

Venckauskas, A., Karpavicius, A., Damaševičius, R., Marcinkevičius, R., Kapočiūte-Dzikiené, J., & Napoli, C. (2017). Open class authorship attribution of lithuanian ınternet comments using one-class classifier. *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, *11*, 373-382, https://doi.org/10.15439/2017f461.

Yazğılı, E., & Baykara, M. (2021). Siber Zorbalık Tespit Yöntemleri Potansiyel Uygulama Alanları ve Zorluklar. *DÜMF Mühendislik Dergisi Sayı 12(1),* 23-35, https://10.24012/dumf.859651

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

Zois, D. S., Kapodistria, A., Yao, M., & Chelmis, C. (2018). Optimal online cyberbullying detection. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017-2021, https://doi.org/10.1109/icassp.2018.8462092.