ESTUDAM Bilişim Dergisi
Cilt 3, Sayı 2, 38-45, 2022

Journal of ESTUDAM Information
Volume 3, Issue 2, 38-45, 2022

DOI: https://doi.org/10.53608/estudambilisim. 1070884

(Geliş Tarihi / Received Date: 10.02.2022, Kabul Tarihi/ Accepted Date: 08.04.2022)

(**Araştırma Makalesi**)

Metin Madenciliği ve Duygu Analizi ile Siber Zorbalık Tespiti

Elif Şevval DİNÇER*[1], Duygu KAYAOĞLU[2], Simara SAFARLİ[3]

[1]Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Eskişehir,
ORCID No : http://orcid.org/0000-0003-2005-4543
[2] Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Eskişehir,
ORCID No : http://orcid.org/0000-0002-1798-4013
[3] Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Eskişehir,
ORCID No : http://orcid.org/0000-0003-3447-9593

Anahtar Kelimeler:
Siber Zorbalık,
Destek Vektör Makinesi,
Lojistik Regresyon,
Naive Bayes

Özet: Tarihte iletişim metotları teknolojinin gelişmesine bağlı olarak değişim göstermiştir. Günümüzde ise bu değişime bağlı olarak iletişim sosyal medya üzerine kaymıştır. Bu kapsamda bazı olumlu yönler olmakla birlikte bazı olumsuz yönleri de vardır. Bu olumsuz yönlerden en belirgin olanı ise siber zorbalıktır. Siber zorbalık daha çok gerçek hayatta insanların söyleyemediği ve söylediğinde karşısındakinden büyük tepki alacağı şeyleri gizli kimlikler aracılığıyla birilerini incitmeye ve kırmaya yönelik söylemlerdir. Bu projede siber zorbalığın tespitine yönelik çalışmalar yapılmıştır. Bu kapsamda öncelikle Twitter Application Programming Interface (API) kullanarak twitter üzerinden veriler elde edilmiş ve bu verileri düzenleyerek metin madenciliğinde yaygın olarak kullanılan yapay zeka yöntemlerinden Destek Vektör Makinesi (SVM), Lojistik Regresyon (LR), Naive Bayes (NB) yöntemleri ile analiz edilmiştir. Yapılan performans analizlerini değerlendirirken f1-skor, kesinlik, hassasiyet ve doğruluk değerlerinden yararlanılmıştır. Bu değerler göz önüne alınarak doğruluk değeri 87% olan LR hazır olarak elde edilen veri setinde kullanılmaya karar verilmiştir. Ve oluşturulan web sitesi bulut platform hizmetlerinden Amazon Web Services (AWS) kullanılarak gerekli ayarlamalar yapıldıktan sonra bulut platform üzerinde çalıştırılmıştır.

(**Research Article**)

Cyberbullying Detection with Text Mining and Sentiment Analysis

Keywords:
Cyberbullying,
Support Vector Machine (SVM),
Logistic Regression,
Naive Bayes

Abstract: In history, communication methods have changed depending on the development of technology. Today, depending on this change, communication has shifted to social media. In this context, although there are some positive aspects, there are also some negative aspects. The most obvious of these negative aspects is cyberbullying. Cyberbullying is the discourse that aims to hurt and offend someone through secret identities, which people cannot say in real life and will get a great reaction when they say it. In this project, studies were carried out to detect cyberbullying. In this context, first of all, data was obtained from Twitter using the Twitter Application Programming Interface (API), and by editing this data, it was analyzed with Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB) methods, which are widely used in text mining. While evaluating the performance analysis, f1-score, precision, precision and accuracy values were used. Considering these values, it was decided to use the LR with an accuracy value of 87% in the ready-made data set. And the created website was run on the cloud platform after making the necessary adjustments using Amazon Web Services (AWS), one of the cloud platform services.

*İlgili yazar/corresponding author: 503020200052@ogrenci.ogu.edu.tr

1. GİRİŞ

Başta internet olmak üzere bilgi ve iletişim teknolojileri, bilgiyi elde etme, saklama ve paylaşma fonksiyonları ile hayatımızda önemli bir rol oynamaktadır. Bilgi teknolojilerinin hayatımıza girmesiyle iletişim kolaylaşmış ve her gün çok büyük boyutlarda veri işlenebilir hale gelmiştir. Üretilen verinin şüphesiz büyük kısmı iste sosyal medya ve platformlarda üretilmektedir. Sosyal medya kullanımı günümüzün iletişim, reklamcılık, habercilik, eğlence vs gibi birçok alanda değişimler getirmiş ve en önemlisi insanların kişisel ve özel hayatlarını paylaştığı, düşünce ve görüşlerini dile getirdiği en önemli platform haline gelmiştir. Sağladığı kolaylıkların yanı sıra elbette ki sosyal mecralar da kullanıcılarını kötü bazı etkilere maruz bırakabiliyor. Bunların başında da siber zorbalık geliyor . Siber zorbalık, bilgi teknolojisi kullanan bir kişi veya gruba, bireye veya tüzel kişiliğe karşı teknik veya ahlaki zarar verici davranışların genel bir durumu olarak tanımlanabilir. Aynı zamanda "elektronik metinle kısıtlı ve tekrarlanan hasar" olarak da anılıyor.

Siber zorbalık vakalarının ve mağdurlarının sayısı sosyal medya kullanımının artmasına bağlı olarak gittikçe artıyor. Uzmanlara göre, siber zorbalığın sonuçları genellikle olumsuz ve hatta yıkıcı olabiliyor. Siber zorbalığın kurbanı olan çocuklar ve ergenler birçok yönden olumsuz etkilenebilmektedir. Siber zorbalığa maruz kalan kullanıcılarda karamsarlık, öfke, kaygı, akademik başarısızlık, devamsızlık, yalnızlık, depresyon ve intihar gibi ciddi sosyal sorunlar oluşmakta ve toplum kalitesi bu durumdan negatif etkilenmektedir. Düşük gelirli bölgelerdeki öğrencilerin orta ve üst sınıf öğrencilerine göre psikolojik olarak sarsılma ve travma geçirme olasılığı daha yüksektir. Sosyal medyada dış görünüşünden , görüşünden ya da yaşam tarzından dolayı siber zorbalığa maruz kalan bir çok kişi toplum baskısı sonucu hayatını kaybediyor. Hatta insanlar siber zorbalığı yüzyılın belası olarak adlandırıyorlar [1].

Ergenlik çağını yaşayan bir kişi, herhangi bir hobisi veya uğraşı yoksa, dost bulmakta, insanlarla iletişimde zorluk çekiyorsa kendini en rahat ifade edebileceği, kimsenin onun gerçek kimliğini bilmediği sanal bir ortamda "istediğini yapma özgürlüğü" fikri onu cezbeder. Psikologlara göre, bu tür oyunlar onlara varlıklarını kanıtlama ve kendilerini değerli hissetme duygusu veriyor: "Aileler, çocukların oynadığı ve ziyaret ettiği yerleri izlemelidir. Yoksa sosyal medyada "kahraman olmaya çalışan çocuklar sonunda intihar ediyorlar" [2].

2017 yılında İngiliz okullarında ve kolejlerinde 12-20 yaşları arasındaki on binden fazla genç arasında bir anket yapılmış ve bir rapor hazırlanmış. Gençlerin başkalarına uyguladığı baskı ve şiddetin nedenleri araştırılmıştır. Rapor aşağıdaki konuları kapsamaktadır [3]:

- Birleşik Krallık'taki şiddet istatistikleri
- Şiddetin özü
- Şiddetin etkisi
- Başkalarına hakaret etme eğiliminde olan gençlerin sayısı
- İnternetteki siber saldırıların ve istismarın ölçeği ve coğrafyası
- Dijital bir dünyada büyümek nasıl bir duygu?
- Sosyal medya eğilimleri, şiddet ve bağımlılık
- Öneriler
- Gerçek hikayeler ve deneyimler [3].

Bizler de bu çalışmalar kapsamında siber zorbalığın insan ve toplum yaşamını olumsuz etkilediğini , siber zorbalığın ve buna maruz kalan bireylerin tespitinin durumun farkındalığı için önemli bir aşama olmasından dolayı sosyal medya platformlarında siber zorbalık tespiti için bu çalışmayı hazırladık .

1.1. Önceki Çalışmalar

Bu alanda birçok araştırma ve çalışma vardır. Bu araştırmalar psikoloji, sosyoloji ve teknolojik başta olmak üzere birçok bilim dalıyla incelenmektedir. Yapılan bazı çalışmaları inceledik. Bu kapsamda Erdur ve arkadaşların Türkiye'de bir lisede yapılan araştırmadan anlaşılmaktadır ki birden fazla siber zorbalık metodu vardır. Bu yöntemler cinsiyete bağlı olarak dağılım oranları değişmektedir. Erkek öğrencilerin kız öğrencilere oranla hem daha fazla siber zorbalık yaptıklarını hem de daha fazla siber zorbalığa maruz kaldıkları belirtilmiş. Çalışmada ekonomik gelir, yaş ve sınıf değişkenlerinin siber zorba ya da siber kurban olma ile ilişkili olmadığı gözlenmiştir [4].

Bu kapsamda metin madenciliği ile siber zorbalık tespitine yönelik de bir çok çalışma vardır. Bu kapsamda yapılan çalışmalardan biri olan Özel ve arkadaşlarının yaptığı araştırmalarında Twitter ve İnstagram'dan elde ettikleri verileri kullanarak bir veri seti oluşturulmuş. Daha sonra veri seti içindeki emojileri olduğu veya olmadığı durumlara göre veri setini incelemişlerdir. Ardından bu veri setini sınıflandırma algoritmaları ile inceleyerek sonuçlar elde etmişlerdir. Sınıflandırıcılar arasında hem sınıflandırma doğruluğu hem de çalışma süresi açısından Naive Bayes Multinomial en başarılı olmuştur.Özellik seçimi uygulandığında ise, kullanılan veri kümesi için sınıflandırma doğruluğu %84'e kadar arttığı gözlenmiştir. [5]. Özel ve arkadaşlarının yaptığı çalışma bu alanda Türkiye'de yapılan ilk çalışmalardandır.

Bu alanda yapılan bir diğer çalışma olan Çürük ve arkadaşlarının formspring.me ve myspace veri setleri üzerinde yapmış oldukları araştırmadır. Bu yapmış oldukları araştırmada birden fazla metotla sınıflandırma yapılmasını kıyaslamışlardır. Kullandıkları metotlardan bazıları; SVM, stokastik dereceli azalma (SGD) ve LR. Çürük ve arkadaşları verileri sınıflandırmaya sokmadan önce ön işleme yapmışlardır. Bu kapsamda etkisiz kelimeleri (stop-word) çıkarmışlardır. Çalışmalarının sonucunda doğruluk ve süre performansı bakımından SGD sınıflandırıcının en uygun sınıflandırıcı olacağına karar vermişlerdir [6].

Bozyiğit ve arkadaşlarının çalışmalarında izledikleri yol sırasıyla; veri toplama, ön işleme, eğitim veri sitinin belirlenmesi, öznitelik çıkarımı ve seçimi, makine öğrenmesi metotları ile sınıflandırma ve sonunda değerlendirmeden oluşmaktadır. Bu yaptıkları çalışmada veri setini oluştururken twitter üzerinden veri almalarını sağlayan bir uygulama veri setini oluşturmuşlarıdır. Daha sonra elde ettikleri verileri ön işleme sokmuşlardır. Ardından öznitelik çıkarımı için terim frekansı ve ters terim frekansını kullanmışlardır. En son makine öğrenmesi metotları ile sınıflandırmayı yapıp değerlendirmişlerdir. Çalışma süresini ve tahminleme başarısını göz önüne alarak yaptıkları değerlendirme sonucunda NB, SVM, k-en yakın komşu (k-NN) algoritmalarının en başarılı algoritmalar olduğunu görmüşlerdir [7].

Dadvar ve arkadaşları yaptıkları çalışmada sadece yapılan paylaşımın içeriğini değil paylaşan kişinin de karakteristik özelliklerinin incelemişlerdir Bu çalışmada, kullanıcı bağlamını dikkate almanın siber zorbalığın tespitini iyileştirdiğini gösterilmiştir. [8].

Hosseinmardi ve arkadaşları yaptıkları çalışmada, instagram verilerini kullanmışlardır. Kullandıkları yorumların aynı zamanda resimlerini de incelemişlerdir. Bunun için görüntü işleme ve metin madenciliği metotlarını kullanmışlardır [9].

Yazğılı ve Baykara'nın yaptıkları çalışmada kendilerinden önce yapılan siber zorbalık tespitinde çalışmalarının, siber zorbalık ve türleri, tespit için kullanılan makine öğrenmesi yöntemleri ve algoritmaları sunulmuş ve karşılaştırma yapılmıştır. Bu çalışmalarda yukarıda yazılan etkenlere bağlı olarak sınıflandırma algoritmalarının performans değerlerinde değişkenliklerin görülmesiyle birlikte birçok dil yapısında yapılan sınıflandırmalarda SVM algoritmasının en iyi performansı sergilediği tespit edilmiştir .Arapça dil yapısında ise NB algoritmasının doğruluk oranının yüksek olduğu görülmüştür [10].

Seda Tuzcu'nun yaptığı çalışmada, çevrimiçi bir kitap satış sitesinin kullanıcı yorumları üzerinde duygu analizi için öncelikle Python programlama dili kullanılarak Çok Katmanlı Algılayıcı (MLP) algoritması uygulanmıştır. Daha sonra bir veri bilimi yazılımı olan RapidMiner kullanılarak, aynı veri seti üzerinde NB, SVM ve LR algoritmaları uygulanmıştır. Algoritmaların yorumları sınıflandırma başarıları karşılaştırılmış ve MLP bu veri setinde en iyi sonuçları gösteren algoritma olmuştur [11].

Büyükeke ve arkadaşlarının yaptığı çalışmada veriler, Tripadvisor platformundan crawler geliştirilerek otomatik olarak toplanmıştır. Toplam yorum sayısı 212,435'tir. Duygu Analizi için; LR, SVM ve NB kullanılmıştır. Analiz sonucunda yorumlarının %80'inin olumlu, %20'sinin olumsuz olduğu bulunmuştur. Konu analizi sonucunda; Deneyim %26,70 ile birinci, Değer ve Eğlence %24,68 ile ikinci, Şikâyet %20,41 ile üçüncü sırada yer almaktadır. Diğer konular; %16,15 ile Temel Hizmetler ve %12,06 ile Yapılacak Şeyler'dir [12].

Gazioğlu ve Şeker'in yaptıkları çalışmada, metin madenciliği yöntemleri Twitter üzerinden toplanan verilere uygulanmıştır. Duygu analizi problemi, sınıflandırma(classification) problemi olarak ele alınmıştır. Emoji içeren İngilizce tweetler toplanmış ve bu veriler üzerinde çalışılmıştır. Emojiler 15 ayrı duygu grubuna ayrılmıştır. Her bir emoji grubu için 8000-10000 arasında tweet toplanmıştır. Toplamda 142000 tweet toplanmıştır. Veri kümesinin homojen olmasının başarı oranını arttırdığı görülmüştür. Emoji içeren veriler ile çalışılarak duygu analizi yapmanın mümkün olduğu görülmüştür [13].

Çürük, çalışmasında sosyal ağlarda sıklıkla yaşanan siber zorbalığın yapay zeka algoritmaları ile tespit edilmesi ve sınıflandırılmasını gerçekleştirmiştir. Çalışmada Youtube, Formspring.me ve Myspace sosyal ağlarından elde edilmiş veri kümeleri kullanılmıştır. Bu veri kümelerinin yapay sinir ağı (YSA) tabanlı radyal tabanlı fonksiyon (RBF), SVM, MLP, LR ve SGD sınıflandırıcıları ile performansları karşılaştırılmıştır. Çalışmanın deneysel sonuçları, siber zorbalık tespiti için nitelik seçme metotlarından ki kare istatistiği (Chi2) ve yinelemeli özellik eliminasyonu (RFE) metotlarının, sınıflandırma performansında ise MLP ve SGD sınıflandırıcılarının siber zorbalık tespiti için en iyi yöntem olduğunu kanıtlamıştır. Yapılan testler sonucunda veri kümelerinin nitelik sayıları ve sınıflandırma süreleri azaltılırken sınıflandırma performanslarının korunduğu belirtilmiştir [14].

Raisi ve Huang çalışmalarında, zorbalık göstergeleri için çekirdek bir kelime dağarcığı tanımlamış ve algoritma zorbalık göstergelerini çıkarmak için büyük etiketsiz bir sosyal medya sözlüğü kullanılmıştır. Bu modelde her sosyal medya etkileşiminin kimin katıldığına ve hangi dilin kullandığına bağlı olarak zorbalık olup olmadığını tahmin eder. Bunun için Participant-vocabulary consistency (Pvc) kullanılmış ve Pvc'nin siber zorbalıktaki etkinliği 3 farklı veri setinde değerlendirilmiştir [15].

Talpur ve Sullivan'ın yaptıkları çalışmada, noktasal bir karşılıklı bilgi tekniğinden yararlanarak zorbalık şiddeti ölçülmüştür. Twitter içeriğinden özellikler oluşturmak için bir siber zorbalık tespit çerçevesi önerilmiş. Bu özelliklere dayanarak, siber zorbalığın tespiti ve Twitter'daki etkisi çok sınıflı sınıflandırılması için denetimli bir makine öğrenimi çözümü geliştirilmiş. Çalışmada, PMI-anlamsal yönelim ile birlikte Gömme, Duyarlılık ve Sözlük özellikleri uygulanmış Çıkarılan öznitelikler NB, k-NN, Decision Tree (DT), Random Forest (RF) ve SVM algoritmaları ile uygulanmıştır. Son olarak, önerilen ve temel özelliklerin sonuçlarını diğer makine öğrenimi algoritmalarıyla karşılaştırılmıştır. Çok sınıflı bir ortamda önerilen çerçeve ile yapılan deneylerden elde edilen sonuçlar, hem Kappa, sınıflandırıcı doğruluğu ve f-ölçü metrikleri açısından hem de ikili bir ortamda güzel sonuçlar verdiği belirtilmiştir.. Bu sonuçlar, önerilen çerçevenin çevrimiçi sosyal ağlarda siber zorbalık davranışını ve şiddetini tespit etmek için uygun bir çözüm sağladığını göstermiştir [16].

Al-Garadi ve arladaşlarının yaptığı çalışmada,siber zorbalık tahmin modellerini kapsamlı bir şekilde incelemiş ve sosyal medyada siber zorbalık tahmin modellerinin oluşturulmasıyla ilgili ana sorunları belirlemiş siber zorbalık tespiti için genel süreç hakkında genel bilgi sağlamış ve en önemlisi metodolojiyi gözden geçirilmiştir. Veri toplama ve özellik mühendisliği süreci detaylandırılmış olsa da, makale genel olarak çoğu özellik seçme algoritmalarına ve ardından siber zorbalık davranışlarını tahmin etmek için çeşitli makine öğrenme algoritmalarının kullanılmasına odaklanmaktadır. Son olarak, araştırmacıların keşfetmesi için yeni araştırma yönleri sunan sorunlar ve zorluklar da vurgulanmıştır [17].

Rosa ve arkadaşlarının yaptıkları çalışmada, siber zorbalık tespitinde Sinir ağları ile motive edilen, benzer çalışmalardan üç mimari uygulanmaktadır: basit bir evrişimli sinir ağı (CNN), hibrit bir CNN-(LSTM) uzun kısa süreli bellek ve karma bir CNN-LSTM-(DNN) dekonvolüsyon sinir ağı. Ayrıca, word2vec modeli aracılığıyla üç farklı kaynaktan üç metin temsili eğitilir: Google-News, Twitter ve Formspring. Deney, yukarıdaki yerleştirmelerden birine sahip bu modellerin, aynı veri kümesinin hem dengesiz hem de dengeli bir versiyonunda diğer kıyaslama sınıflandırıcılarını (SVM ve LR) geçtiğini göstermektedir [18].

Febriana ve Budiarto'nun yaptıkları çalışma, Endonezya 2019 seçim döneminde sosyal medyada bir nefret söylemi algılama modeli oluşturmak için kullanılabilecek bir veri kümesi geliştirme sürecini sunar. Twitter API kullanılarak 1 milyondan fazla tweet başarıyla toplanmış ve makine öğrenimi kullanılarak temel ön işleme ve ön çalışma uygulanmıştır. Latent Dirichlet Allocation (LDA) algoritması, bu konuların tartışma temalarıyla ilişkilendirilip ilişkilendirilemeyeceğini görmek için her tweet için özellik çıkarmak için kullanılmıştır. Her tweet için bir polarite puanı oluşturmak için veri setine önceden eğitilmiş duygu analizi de uygulanmış ve analiz adımına dahil edilen 83.752 tweet'ten olumlu ve olumsuz tweet sayıları hemen hemen aynı olmaktadır [19].

Perara ve Fernando'nun çalışmaları, siber zorbalığı tespit etme ve önleme yönelik bir sistem üzerine kurulmuştur. Çalışmada SVM ve LR kullanarak siber zorbalık metninin yanı sıra ırkçı cinsel küfür vb şeklinde sınıflandırma üzerine çalışılmıştır. Terim Frekansı (TF-IDF), Ngram gibi özelliklerle sistemin doğruluğu arttırılmış ve sistem hassasiyet ve F1-skoru gibi özelliklere göre değerlendirilmiştir [20].

Bu makalede Hadoop; Hadoop'un yardımıyla büyük miktardaki boyutlu verilerin analizi gerçekleştirilmiştir. Bu yazıda, büyük veri olarak da bilinen twitter verilerinin analizi için Hadoop kullanılmış. Bu kapsamda sosyal medya platformları, dizi film yorumları, bloglar gibi birçok farklı metin için büyük veri için duygu analizi gerçekleştirilmiştir. Toplam doğruluk elde edilen veriler için %72.22 şeklinde ifade edilmiştir [21].

2. MATERYAL VE METOT

Bu çalışma kapsamında sosyal medya platformlarından Twitter üzerinden elde edilen veri setini kullanarak yeni gelecek yorumlarda siber zorbalık tespiti için sınıflandırma algoritmalarını kullanıldı. Bu inceleme için SVM, LR, NB sınıflandırmaları yapıldı.

2.1. Veri toplanması

Verilerin elde edilmesi için öncelikle Twitter API için başvuru yapıldı. Bu başvuru sonucunda onay aldıktan sonra Python 'ın Tweepy kütüphanesi aracılığı ile veriler çekildi. Bu çekilen veriler daha sonra bir veritabanına kaydedildi.Bu veritabanı yardımıyla java üzerinden çektiğimiz verileri daha sonra işlerken bozulmamış ham haliyle saklamayı hedefledik. Bu toplanan veriler daha sonra elde ettiğimiz bir veri seti yardımıyla eğitildi.

2.2. Öznitelik belirlenmesi

Öz nitelikleri belirlemek için n-gram ve terim frekansı ve ters terim frekansını metotları kullanılacaktır. Bu kapsamda öncelikle n-gramdan bahsedeceğiz. N-gram bir veri üzerinde arama yapmak ve karşılaştırma yapmak veya tekrar sayısını belirlemek için kullanılan metottur.

Terim frekansı bir doküman içerisinde geçen terim ağırlıklarını hesaplamak için kullanılmaktadır.

Ters terim frekansını ise birden fazla dokümanda kelimenin geçme sayısını bularak bu kelimenin terim olup olmadığını bağlaç vb (stop-words) olduğu anlamaya çalışır. Terimin Geçtiği Dokuman Sayısı / Dokuman Sayısının logaritmasının mutlak değeri alınması ile bulunur [22].

2.3. Veri seti

Kullandığımız veri seti 3002 adet veriden oluşmakta ve 2 sınıf içermektedir. Veri setinde 1503 adet siber zorbalık içeren, 1498 tanesinde zorbalık içermeyen tweetler bulunmaktadır.

2.4. Yapay zeka yöntemleri

Veri toplama işleminin ardından veri setinin siber zorbalık içeren öğelerini tespit etmek için yaygın kullanılan yapay zeka metotları uygulanmıştır. Uygulanan metotlar aşağıda kısaca açıklanmaktadır.

NB: bayes teorisine dayanan istatistik temelli denetimli öğrenme algoritmasıdır [23]. Bu algoritma metin belgelerinin sınıflandırılması tüm eğitim veri kümesinin üzerinde koşullu olasılıklar hesaplayarak gerçekleştirilmektedir. Naive bayes uygulamanın en büyük avantajlarından biri kolay uygulanabilir olmasına rağmen iyi sonuçlar vermesidir. Bayes teoremi denklemi:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

P(A|B), B olayı gerçekleştiğinde A olayının olma olasılığıdır. P(B|A), A olayı gerçekleştiğinde B olayının olma olasılığıdır. P(A) ve P(B), A ve B olaylarının önsel olasılıklarıdır [24].

SVM: istatistiksel bilgi teorisine ve temel risk minimizasyonuna dayanan bir sınıflandırma algoritmasıdır. Bu yöntemde verileri sınıflandırmak için düzlemler oluşturulmaktadır [25]. Bu düzlemlerde verilerin dağılımına göre çekirdek fonksiyonu belirlenmektedir. Bu yöntemin avantajları ise çoğu probleme özelinde iyi sonuç vermesi ve aşırı uyuma karşı güçlü olmasıdır.

Tablo 1. SVM'de kullanımı uygun çekirdek fonksiyonları

No	Çekirdek Fonksiyonu	Matematiksel Gösterimi
1	Doğrusal	$k(x,x') = x^T x' + c$
2	Polinomiyal	$k(x,x') = (ax^T x' + c)^d$
3	Gaussian	$k(x,x') = exp\left(-\frac{\|\|x-x'\|\|^2}{2\sigma^2}\right)$
4	Exponansiyel	$k(x,x') = exp\left(-\frac{\|\|x-x'\|\|}{2\sigma^2}\right)$
5	Hiperbolik tanjant	$k(x,x') = tanh(ax^T x' + c)$

LR: girdi alarak çalışan ve ağırlık değerleri ile giriş değerlerini çarpan yapay zeka yöntemlerinden biridir. Farklı olası sınıflar arasında ayrım yapmak için en kullanışlı ve faydalı olan özellikleri belirleyen bir sınıflandırma algoritmasıdır [26]. LR modeli Tablo 2'de gösterilmiştir [27].

Tablo 2. LR modeli

Bağımlı Değişken (y)	Bağımsız Değişken (x)	
	$x = 1$	$x = 0$
$y = 1$	$\pi_{(1)} = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi_{(0)} = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
$y = 0$	$1-\pi_{(1)} = \frac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi_{(0)} = \frac{1}{1+e^{\beta_0}}$
Toplam	1	1

Biz bu uygulamada veri seti için en yüksek doğruluk oranına sahip olan LR yöntemini kullandık. Doğruluk oranları Tablo 1'de verilmiştir.

2.5. Değerlendirme ölçütleri

Yukarıdaki belirtilen metotları kullanırken bazı değerlendirme ölçütleri kullanmaktayız. Bu ölçütler kesinlik (precision), hassasiyet (recall), f1-score, destek (support), doğruluk (accuracy) dir. Bu hesaplamaları ifade ederken Tablo 3'den yaralanılacaktır.

$$TP + FP = Toplam\ Pozitif\ Sınıflandırma \quad (2)$$

$$Kesinlik = \frac{TP}{Toplam\ Pozitif\ Sınıflandırma} \quad (3)$$

Sonuç olarak kesinlik değeri belirli sınıfın o sınıfa ait toplam sınıf değerlerine oranıdır. Kesinlik aslında negatif bir örnek pozitif etiketlemek için değil sınıflandırıcı yeteneğidir. Her bir sınıf için, doğru ve yanlış pozitiflerin toplamına gerçek pozitif oran olarak tanımlanır [28].

$$Hassasiyet = \frac{TP}{TP + FN} \quad (4)$$

$$TP + FN = Toplam\ Asıl\ Sınıflandırma \quad (5)$$

$$Hassasiyet = \frac{TP}{Toplam\ Asıl\ Sınıflandırma} \quad (6)$$

Sonuç olarak hassasiyet değeri belirli sınıfın o sınıfa ait toplam sınıf değerlerine oranıdır. Hassasiyet tüm olumlu örneklerini bulmak için bir sınıflandırıcı yeteneğidir. Her bir sınıf için, doğru pozitif ve yanlış negatif toplamına gerçek pozitif oran olarak tanımlanır. Başka bir yolu "aslında pozitif tüm örneklerini, yüzde kaçı doğru sınıflandırıldı?" denebilir [29].

Tablo 3. Kesinlik değerlendirmesi hesaplanan alan

		GERÇEK		
		Pozitif	Negatif	TOPLAM
TAHMİN	Pozitif	Doğru Pozitif (TP)	Yanlış Pozitif (FP)	Tahmin Pozitif sayısı
	Negatif	Yanlış Negatif (FN)	Doğru Negatif (TN)	Tahmin Negatif sayısı
	Toplam	Gerçek Pozitif sayısı	Gerçek Negatif sayısı	Toplam Örnek sayısı (m)

F1 skor ise bu iki değere bağlı olarak hesaplanır.

$$F1 = 2 * \frac{Kesinlik * Hassasiyet}{Kesinlik + Hassasiyet} \quad (7)$$

F1 skor hassas ağırlıklı harmonik ortalaması ve en iyi puanı %100 ve en kötü %0 olmaktadır. Onlar hassas gömmek ve bunların hesaplama içine hatırladığım kadarıyla Genel olarak konuşmak gerekirse, F1 puanları doğruluk önlemlerinin daha düşüktür. Genel bir kural olarak, F1 ağırlıklı ortalama sınıflandırıcı modellerini değil, küresel doğruluğu karşılaştırmak için kullanılmalıdır.

3. BULGULAR

Değerlendirme ölçütlerini kullanarak yukarda belirtilen veri seti üzerinde yapay zeka yöntemlerini kullanarak aşağıda verilen Tablo 4, Tablo 5, Tablo 6 ve Tablo 7 de deki değerler ışığında tahmin yapılması için hangi yapay zeka algoritması kullanılması gerektiğine karar verildi. Bu doğrultuda hem doğruluk hem hassasiyet hem kesinlik hem de f1-skor bakımından kullandığımız sistem için daha iyi bir sonucu lojistik regresyon ile elde edileceği sonucu gözlemlendi.

Tablo 4. Değerlendirme sonucu elde edilen doğruluk oranları

Algoritma	Doğruluk(%)
Destek Vektör Makineleri	84,95
Naive Bayes	82,42
Lojistik Regresyon	87,21

Tablo 5. Değerlendirme sonucu elde edilen f1-skoru oranları

Algoritma	F1-score(%)
Destek Vektör Makineleri	85
Naive Bayes	82
Lojistik Regresyon	87

Tablo 6. Değerlendirme sonucu elde edilen kesinlik oranları

Algoritma	Kesinlik (%)
Destek Vektör Makineleri	85
Naive Bayes	82,5
Lojistik Regresyon	87

Tablo 7. Değerlendirme sonucu elde edilen hassasiyet oranları

Algoritma	Hassasiyet (%)
Destek Vektör Makineleri	85
Naive Bayes	82,5
Lojistik Regresyon	87

Kararın ardından python dili kullanılarak bir web sayfası oluşturulmuştur. Bu oluşturulan web sitesinde gömülü olarak daha önce oluşturulan ve siber zorbalık olup olmadığını tahmin etmek için kullanacağımız eğitilmiş veri seti ve eğitimini yaptığımız yapay zeka algoritması bulunmaktadır. Ayrıca twitter üzerinden veri çekmek için kullanılan tweepy kütüphanesi yardımı ile twitter api üzerinden elde edilen çeşitli anahtarlarda gömülü olarak kullanılmaktadır. Bu anahtarlar twitter üzerinden veri çekmek için kullanılıyor. Web sitesine kayıt olurken kullanıcılardan twitter üzerinde kullandıkları kullanıcı adı ile kayıt olması beklenmektedir. Kullanıcı kayıt olduğu kullanıcı adı ile twiiter da kullanıcı adının geçtiği tweetler çekilip veri tabanında ilgili tabloya kaydedilmektedir. Ayrıca kullanıcının adını geçen tweetlerin retweet sayısı da kullanıcı ara yüzünde gösterilmektedir. Kullanıcı için toplanan veriler anında arka planda gömülü olan yapay zeka modülünde işleme tabi olup siber zorbalık içerip içermediğine dair bir tahmin işlemine tabi tutulur. Ardından da ara yüzde kullanıcıya yansıtılır. Web sayfasına ait ekran görüntüleri şu şekildedir;

- Ek A sisteme kayıt olma ara yüzü
- Ek B kullanıcı giriş ekran görüntüsü .
- Ek C kullanıcıya ait zaman çizelgesi görüntüleri vardır. zaman çizelgesi görüntüleri olarak getirilen tweetlerin retweet sayısı.
- Ek D kullanıcı adının geçtiği arama parametresi olarak dil ayarı Türkçe verilmiş tweetler ve bu tweetleri atan kullanıcıların kullanıcı adı .
- Ek E'de ise bu arama sonucu elde edilen tweetlerin analiz sonuçlarını içermektedir.

4. TARTIŞMA VE SONUÇ

Sonuç olarak Twitter Api yardımı ile veri çekimi başarılı olarak yapılmıştır. Bunun sonucunda elimizdeki veri setinden oluşturduğumuz model yardımı ile çektiğimiz verilerin analizini gerçekleştirilmiş. Model oluşturma aşamasında farklı algoritmalar sonucunda doğruluk oranı en büyük olan LR algoritması olarak analiz edilmiştir. Analiz sonuçları incelendiğinde bazı yanlış analizler olduğu gözlemlenmiştir. Bunu sebebi ise kullanılan veri setinin yapısı ve Türkçe 'de bazı kelimelerin birden fazla anlamda kullanılmasıdır. Bunun için daha büyük veri setleri kullanarak bunun önüne geçilebilir. Projeyi bulutta yapmamızın en büyük sebebi yapay zeka işlemlerinin daha hızlı yapılması ve çekilecek olan büyük ölçüdeki verilerin saklanma kolaylığının sağlanması. Bunlara ek olarak bulut ortamlarına erişimin çok kolay olması da bulut platformlarını tercih etmemizin sebeplerindendir.

KAYNAKÇA

[1] Flannery, D. J., Wester, K. L., Singer, M. I. 2004. Impact of Exposure to Violence in School on Child and Adolescent Mental Health and Behavior. Journal of community psychology, 32(5), 559-573. DOI:10.1002/jcop.20019

[2] Dorukoğlu, B. 2017. Sosyal Medya ve Çocuklar https://dijitalmedyavecocuk.bilgi.edu.tr/2017/04/06/sosyal-medya-ve-cocuklar/. (Erişim Tarihi: 24.03.2022).

[3] The Annual Bullying Survey, Ditch the Label. UK, June, 2017. p. 40. https://www.ditchthelabel.org/research-papers/the-annual-bullying-survey-2017/ (Erişim Tarihi 24.03.2022)

[4] Erdur-Baker, Ö., Kavşut, F. 2007. Akran Zorbalığının Yeni Yüzü: Siber Zorbalık. Eurasian Journal of Educational Research, (27).

[5] Özel, S. A., Saraç, E., Akdemir, S., Aksu, H. 2017. Detection of Cyberbullying on Social Media Messages in Turkish. In 2017 International Conference on Computer Science and Engineering, 5-8 Ekim, Antalya, 366-370.

[6] Çürük, E., Acı, Ç., Eşsiz, E. S. 2018. Performance Analysis of Artificial Neural Network Based Classfiers for Cyberbulling Detection. In 2018 3rd International Conference on Computer Science and Engineering, 20-23 Eylül, Sarajevo, Bosnia and Herzegovina, 1-5.

[7] Bozyiğit, A., Utku, S., Nasiboğlu, E. 2018. Sanal Zorbalık İçeren Sosyal Medya Mesajlarının Tespiti. In 3rd International Conference on Computer Sciences and Engineering, 20-23 Eylül, Sarajevo, Bosnia and Herzegovina, 281-281.

[8] Dadvar, M., Trieschnigg, D., Ordelman, R., de Jong, F. 2013. Improving Cyberbullying Detection with User Context. In European Conference on Information Retrieval. Springer, Berlin, Heidelberg, 693-696.

[9] Hosseinmardi, H., Mattson, S. A., Rafiq, R. I., Han, R., Lv, Q., Mishra, S. 2015. Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network. In International conference on social informatics, 9-12 Aralık, Beijing, China, 49-66.

[10] Yazğılı, E., Baykara, M. 2021. Siber Zorbalık Tespit Yöntemleri Potansiyel Uygulama Alanları ve Zorluklar. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 12(1), 23-35. DOI:10.24012/dumf.859651

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

[12] Büyükeke, A., Sökmen, A., Gencer, C. 2020. Metin Madenciliği ve Duygu Analizi Yöntemleri ile Sosyal Medya Verilerinden Rekabetçi Avantaj Elde Etme: Turizm Sektöründe Bir Araştırma. Journal of Tourism and Gastronomy Studies, 8(1), 322-335. DOI:10.21325/jotags.2020.550

[13] Gazioğlu, K., Şeker, Ş. E. 2017. Veri Madenciliği Yöntemleri ile Twitter Üzerinden Girişimcilik Analizi. YBS Ansiklopedi, 4(4).

[14] Çürük, E. 2018. Sosyal Ağlardaki Siber Zorbalığın Yapay Zeka Algoritmaları İle Tespiti Ve Sınıflandırılması. Mersin Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 99s, Mersin.

[15] Raisi, E., & Huang, B. 2017. Cyberbullying Detection with Weakly Supervised Machine Learning. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 31 Temmuz–3 Ağustos, Sidney, NSW, Australia, 409-416.

[16] Talpur, B. A., O'Sullivan, D. 2020. Cyberbullying Severity Detection: A machine learning approach. PloS one, 15(10). DOI:10.1371/journal.pone.0240924

[17] Al-Garadi, M. A., Hussain, M. R., Khan, N., Murtaza, G., Nweke, H. F., Ali, I., Gani, A. 2019. Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges. IEEE Access, 7, 70701-70718. DOI:10.1109/ACCESS.2019.2918354

[18] Rosa, H., Matos, D., Ribeiro, R., Coheur, L., Carvalho, J. P. 2018. A "Deeper" Look at Detecting Cyberbullying in Social Networks. In 2018 international joint conference on neural networks, 8-13 Temmuz, Rio De Jeneiro, Brazil, 1-8.

[19] Febriana, T., Budiarto, A. 2019. Twitter Dataset for Hate Speech and Cyberbullying Detection in Indonesian Language. In 2019 International Conference on Information Management and Technology, 19-20 Ağustos, Jakarta/Bali, Indonesia, 379-382.

[20] Perera, A., Fernando, P. 2021. Accurate Cyberbullying Detection and Prevention on Social Media. Procedia Computer Science, 181(2021), 605-611. DOI:10.1016/j.procs.2021.01.207

[21] Sehgal, D., Agarwal, A. K. 2016. Sentiment Analysis of Big Data Applications Using Twitter Data with the Help of Hadoop Framework. In 2016 international conference system modeling & advancement in research trends, 25-27 Kasım, Moradabad, India, 251-255.

[22] Tf-idf https://en.wikipedia.org/wiki/Tf%E2%80%93idf (Erişim Tarihi: 27.03.2022)

[23] McCallum, A., Nigam, K. 1998. A Comparison of Event Models for Naive Bayes Text Classification. In AAAI-98 workshop on learning for text categorization (Vol. 752, No. 1, pp. 41-48).

[24] Dilber, B. 2020. Algorithm: Naive Bayes Classifier. https://www.datasciencearth.com/algorithm-naive-bayes-classifier/ (Erişim Tarihi: 27.03.2022).

[25] Muller, K. R., Mika, S., Ratsch, G., Tsuda, K., & Scholkopf, B. (2001). An Introduction to Kernel-Based Learning Algorithms. IEEE transactions on neural networks, 12(2), 181-201. DOI:10.1109/72.914517

[26] Indra, S. T., Wikarsa, L., & Turang, R. 2016. Using Logistic Regression Method to Classify Tweets Into the Selected Topics. In 2016 international conference on advanced computer science and information systems, 15-16 Ekim, Malang, Indonesia, 385-390.

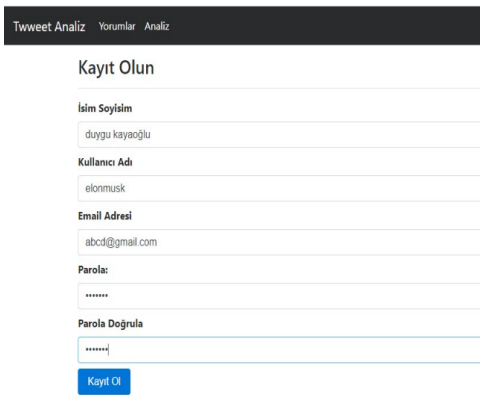
[27] Çelik, G. 2019. Orantısal ODDS Lojistik Regresyon Modeli için Uyum İyiliği Testlerinin Performanslarının Benzetim Çalişması ile Değerlendirilmesi. Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, 104s, Ankara.

[28] Wikipedia Precision and recall https://en.m.wikipedia.org/wiki/Precision_and_recall (Erişim Tarihi: 27.03.2022).

[29] Yellowbrick Classification Report https://www.scikit-yb.org/en/latest/api/classifier/classification_report.html (Erişim Tarihi: 15.03.2022).

Ekler

Ek A. Oluşturulan platformda kayıt sayfası



Ek B. Kullanıcı giriş ekranı

Ek C. Kullanıcı Tweetleri

Ek D. Kullanıcı Yorumları

Ek E. Analiz Sonuçları