

**T.C.**

**FIRAT ÜNİVERSİTESİ MÜHENDİSLİK FAKÜLTESİ**

**YAZILIM MÜHENDİSLİĞİ**

**YMU410 YAPAY ZEKA VE UZMAN SISTEMLER**

**Siber Zorbalık Tespiti Projesinin Raporu**

**Hazırlayan:**

180290608 - Mohammad Hajjar

180290606 - Ahmed Rıdvan

180290009 - Melih Furkan Kalaylı

180290611 - Ömer Akrum

180290603 - Nadem Ghnnam

**Ders Sorumluları:**

Prof.Dr. Bilal ALATAŞ

Arş.Gör. Şule KAYA

**Mayıs 2022**

İÇİNDEKİLER

1. **Giriş** ...................................................................................................3
   1. Siber zorbalık nedir? ....................................................................3
   2. Siber zorbalık türleri nelerdir? .....................................................3
   3. Makine Öğrenimi Kullanarak Nefret Söylemesini Tespit Etmesi…3
2. **Proje üzerinde çalışma yöntemi**.....................................................**...**4
3. **Veri**........................................................ .............................................5
   1. Veri Toplama.......................... ......................................................5
   2. Veri Temizleme.......................... .......................... ........................6
   3. Veri Ön İşleme.......................... ....................................................6
   4. Veri Yeniden Örnekleme...............................................................7
4. **Yöntemlerin Açıklaması**.................................................... .................7
   1. Gaussian Naive Bayes....................................................................7
   2. Logistic Regression.......................... .......................... ...................8
   3. Decision Tree Classifier ......................... ........................................8
   4. Random Forest Classifier.......................... .....................................9
5. **Deney Ve Sonuçları**.......................... .......................... ........................9
6. **Sonuç**.......................... .......................... .......................... ..................10
   1. Siber Zorbalık Tespiti İle İlgili Çalışmalar........................................11
   2. Kullanılan Yöntemler ve Test Sonuçları.........................................12
7. **Kayankça**.......................... .......................... .......................................13

**Siber Zorbalık Tespiti**

**(Cyberbullying Detection)**

**1. Giriş**

**1.1.Siber zorbalık nedir?**

Elektronik ya da dijital ortamda başkalarına rahatsızlık ya da zarar vermek amacıyla kişilerin ya da grupların aşağılama, iftira, dedikodu, taciz, tehdit, dışlama ve rencide etme gibi davranışlarının tümüne siber zorbalık denir.

Siber zorbalık, bir psikolojik şiddet biçimidir. Paylaşımların internet ortamında hızla yayılması ise yaşanan bu psikolojik şiddeti daha da büyütmektedir. Buna ek olarak, Mağdur, diğer saldırganlık türlerinden daha büyük bir çaresizlik hissine sahiptir. Çünkü ağlar yüzünden saldırı her hangi bir zamanda ulaşabilir ve ne zaman görüneceğini veya kimin tarafından meydana geleceğini bilmez. Geleneksel zorbalığın aksine, siber zorbalıkta taciz anonim olabilir.

Siber zorbalık, tüm dünyada giderek yaygınlaşan bir durumdur. İnternet kullanımının artması ile birlikte ortaya çıkan çevrimiçi riskler arasında duygusal sonuçları açısından en yaralayıcı olanlarından birisi siber zorbalıktır.

**1.2.Siber zorbalık türleri nelerdir?**

* Sosyal ağlar ya da sohbet odaları gibi çevrimiçi ortamlarda başkalarına kötü mesajlar göndermek.
* Sosyal ağlardaki paylaşımlarına sürekli olumsuz yorumlar yapmak.
* Bir kişinin kişisel bilgilerini izni ve haberi olmadan internet ortamında paylaşmak.
* Bir kişiye alakalı olarak karalayıcı, aşağılayıcı web sayfaları hazırlamak.
* Başkası adına sahte hesap açıp, onun kimliğine bürünmek.

**1.3.Makine Öğrenimi Kullanarak Nefret Söylemesini Tespit Etmesi**

Bizim projemiz, siber zorbalık tespit yaklaşımları hakkında yayınlanmış bazı araştırmaların sistematik bir incelemesini sunar ve sosyal medyada nefret söylemini tespit etme yöntemlerini incelerken, bunu genel küfürden ayırır ve standart ve aynı zamanda topluluk yöntemleri de dahil olmak üzere çeşitli Denetimli algoritmaların karşılaştırmalı bir çalışmasını yapar. Nefret söylemi, belirli bir gruba, o grup bir topluluk, din veya ırk olabilir, nefret uyandırmayı amaçlayan sözcükleri ifade eder. Bu konuşmanın anlamı olabilir veya olmayabilir, ancak Siber Zorbalık yol açması muhtemeldir. Özellikle sosyal medya ağlarında, kullanıcı tarafından oluşturulan web içeriğindeki büyük artış nedeniyle, nefret söylemi miktarı da istikrarlı bir şekilde artıyor. Geçtiğimiz yıllarda, kısmen sosyal medyada siber zorbalığın yaygınlaşması nedeniyle, siber zorbalık tespitine yönelik araştırmalar arttı.

**2. Proje üzerinde çalışma yöntemi**

Siber Zorbalığın Tespiti için, başta Doğal Dil İşleme ve Bilgi Erişimi olmak üzere, daha sonra TF-IDF, Duygu Analizi, Boyut İndirgeme vb. ise siber zorbalığı tespit etmek ve caydırmak için NLP tekniklerini kullanarak sosyal medyadaki kötülüğü tespit etmenin yollarını dener. NLP teknikleri, verilerdeki küfürlerin hakaret edici veya tarafsız bir şekilde kullanıldığını bile tespit edebilecek şekilde kullanılmaktadır. Makalede kullanılan ek açıklamalar, laboratuvar ek açıklamaları ve kitle kaynak kullanımı kullanılarak yinelemeli olarak gözden geçirilir. Hatta sosyal medya sitelerindeki İngilizce gönderilerden veriler tarandı. Ask.fm gibi yarı anonim sosyal medya siteleri dahil. NLP ile birlikte sıralanmış bir küfür listesi, etkili bir şekilde taramaya yardımcı oldu. Verileri sınıflandırmak için, kötü kelimeleri rastgele bir şekilde ayırt etmek için değiştirilmiş doğrusal SVM kullanıldı, Soru cevap gönderileri ve İfadeler gibi fark edilmeyecek birçok başka özellik de dikkate alındı. Sonunda, F1-Skoru 0,59 olarak ortaya çıktı (her ne kadar Kaggle'ın kazananından daha az olsa da, bu çalışmanın özelleştirilmiş veriler ve yeni ve daha iyi bir veri seti kullanmadığı gerçeği göz önüne alındığında, 0,59'luk F1 Skoru hala umut verici görünüyor) .

Bu çalışmada karşılaşılan zorluklar şunlardı:

• Ask.fm'de yorumlar, diğer veri kümelerinde daha kısa olan soru-cevap çiftleridir ve her iki soru-cevapta sadece bir kelime bulunabilir, bu da algoritmanın tam bağlamı anlamadan sınıflandırmasını zorlaştırır.

• İnsanlar, sosyal medyada yazım yanlışları ve kısaltmalarla dolu gayri resmi dil ve argo kullanıyor, bu da onları işlemeyi çok zorlaştırıyor.

Metinsel özellikleri sosyal ağ özellikleriyle bütünleştirerek, kullanıcılar arasındaki sosyal ağ yapısını analiz etmeye ve çok sayıda arkadaş, ağa gömülülük ve ilişki merkeziliği gibi özellikler türetmeye, siber zorbalığın tespitine odaklanmakta elde edilebilir. Aralık 2008'den Ocak 2009'a kadar ve SMOTE (sentetik) kullanır azınlık aşırı örnekleme tekniği yaklaşımı.

Naive Bayes, J48, SMO, Bagging ile dengeli veriler ve üzerine kama uygulanıyor. Siber zorbalık tespiti için Reddit'in yorum külliyatını kullanan Rakib ve ark derlemi çıkardı ve Reddit veritabanından temizledi, ardından word2vec atlama-gram modeline dayalı bir kelime gömme modeli eğitti. Ardından, siber zorba yorumlarını sınıflandırmak için rastgele bir random forest sınıflandırıcısı yetiştirmek için bu modelin özellikleri kullanıldı. Alan bilgisi kullanılarak yapılan bu yeni kelime gömme modeli, önceden eğitilmiş 4 kelime gömme modelinden ve ayrıca el yapımı özellik çıkarma yöntemlerinden daha iyi performans gösterdi.

|  |  |  |
| --- | --- | --- |
|  | ROC | TP |
| Bagging | 0.700 | 0.211 |
| J48 | 0.628 | 0.259 |
| SMO | 0.703 | 0.733 |
| Dagging | 0.755 | 0.763 |
| Naive Bayes | 0.695 | 0.723 |

Silva et al psikolojiye dayalı araştırmayı kullanan siber zorbalık tanımlaması için bir model önerdi; Bu, siber zorbalığın keşfedilmesi durumunda kullanıcının ebeveynlerini yakın iletişim konusunda uyarmayı amaçlar. Kullanıcının mesajlarını ve yorumlarını inceleyerek sosyal medya verilerini analiz etmek ve bunları uyarı işaretleri olarak sıralamak veya onlara zorbalık sıralaması vermek için geleneksel yöntemler kullanır. Özel olarak yapılmıştır ergenler içindir ve Facebook'ta tespit için eski yöntemleri kullanır, ancak üzerinde ML sınıflandırmasının çalıştırılabileceği bir veri toplama uygulaması olarak hareket ederek büyüme potansiyeline sahiptir.

**3. Veri**

Bu bölüm, verilerin toplanmasından ön işlemeye ve özellik çıkarımına kadar tüm yönlerini içerir.

**3.1.Veri Toplama**

Nihai sonuçlara ulaşmak için Dataturks'ün Cybertroll Tespiti için Kaggle [1]'den alınan Tweet Dataset'i kullandık. Çözmeyi hedeflediğimiz sorunun ciddiyeti nedeniyle, eksiksiz, güvenilir, alakalı ve konuya uygun bir veri seti seçmek çok önemliydi. Diğer birçok veri kümesini de göz önünde bulundurduğumuzda, birçoğunun ya eksik nitelikleri vardı, kaliteleri çok düşüktü ya da manuel incelemeden sonra alakasız verilere sahip oldukları bulundu. Böylece, diğer birçok açık kaynaklı veri setini denedikten sonra, gerekli tüm parametrelerle uyumlu göründüğü gibi [1]'e geldik.

Kullanılan Veri Seti Hakkında Bilgileri:

1. Kısmen manuel olarak etiketlenmiş bir veri kümesidir.

2. Toplam veri sayı 20001 dir.

3. Verileri iki kategoriye sınıflandırılmış: 1 (Agresif) ya da 0 (Agresif Olmayan)

4. Veri Seti:

<https://www.kaggle.com/datasets/dataturks/dataset-for-detection-of-cybertrolls?select=Dataset+for+Detection+of+Cyber-Trolls.json>

**3.2.Veri Temizleme**

Kullanılan veri seti bir json formatında ayarlandı. Veri kümesi alanlarının yorumlanması nispeten basit olduğundan, ek açıklama özelliğindeki orijinal alan kümesi kaldırıldı ve sonraki adımı basitleştirmek için etiket değerleriyle dolduruldu. Her sınıf için örnek sayısı tablo 1'de belirtilmiştir.

|  |  |
| --- | --- |
|  | Twitter |
| Total Instances | 20001 |
| CyberBullying instances | 7822 |
| Non-CyberBullying instances | 12179 |

**3.3.Veri Ön İşleme:**

Normal ifade ile birlikte nltk kitaplığı kullanılarak ön işleme adımları aşağıdaki gibi yapıldı:

1. Kelime Belirteçleştirme: Bir Belirteç, cümle veya paragraf için yapı taşları olan tek bir varlıktır. Word Tokenization, metnimizi bir listedeki ayrı sözcüklere dönüştürür.

2. Durdurma sözcükleri filtrelemesi, İngilizce sözlükte durdurulan sözcüklerin bir listesini getirmek için nltk.corpus.stopwords.words('english') kullanılarak yapılır ve ardından bunlar kaldırılır. Durma sözcükleri, “the”, “a”, “an”, “in” gibi anlamlı olmayan ve yorumlanacak verinin anlamını etkilemeyen sözcüklerdir.

3. Noktalama işaretlerini kaldırmak için, yalnızca noktalama işareti olmayan ve string.punctuation kullanılarak kontrol edilebilen karakterleri kaydederiz.

4. Stemming: Stemming, kelimeleri kelime kök kelimesine indirgeyen bir dilsel normalleştirme sürecidir. Köklü belirteçleri almak için belirteçleri nltk.stem.porter.PorterStemmer kullanarak köklendiriyoruz. Örneğin, bağlantı, bağlantı, bağlantı sözcüğü ortak bir "bağlan" sözcüğüne indirgenir.

5. Rakam kaldırma: Siber zorbalığa katkıda bulunmadığı için sayısal içeriği de filtreledik.

6. Şimdi bir sonraki adım, Python'un sklearn kütüphanesini kullanarak TF-IDF Transform kullandığımız ML algoritmalarıyla kullanılabilecek şekilde özellikleri çıkarmaktı. TF-IDF, bir kelimenin alaka düzeyini değerlendirmek için istatistiksel bir ölçüdür ve temel olarak, kelimelerin belgede görünme sayısı ile kelimenin ters belge sıklığı çarpılarak hesaplanır. TF-IDF, CountVectorizer'ın yaptığı gibi kelimelerin sıklığını basitçe saymak yerine, ortak olarak birçok belgede yer alan kelimelerin ağırlığını (önemini) azaltan bir yöntemi kullanır, bunların belgeleri ayırt edemeyeceklerini düşünür.

|  |  |  |
| --- | --- | --- |
|  | Test | Training |
| Total Instance | 4001 | 16000 |
| CyberBullying Instances | 2429 | 9750 |
| CyberBullying Instances | 1572 | 6250 |

Matris her belgeden (satır) ve her kelimeden (sütun) ve tf \* idf (matris değerleri) ile hesaplanan önemden (ağırlık) oluşur. Bir kelime bir belgede yüksek tf-idf'ye sahipse, çoğu zaman verilen belgelerde meydana gelmiştir ve diğer belgelerde olmaması gerekir. Yani kelimeler bir imza kelimesi olmalıdır. Nitelik değerlendirmesi, görüldüğü gibi manuel olarak yapılır. göre ilk 25 kelimeyi yazdırdığımız yer hesaplanan tf-idf puanı. için bazı üst sıradaki kelimeler veri kümesi şunlardı: [hate, fuck, damn, suck, ass, that, lol, im, like, you, it, get, what, no, would, bitch]

**3.4.Veri Yeniden Örnekleme**

Veriler çarpık olduğu için, eğitim verileri üzerinde Yeniden Örnekleme yapılması gerekiyordu, Öncelikle veriler 80:20 oranında Eğitim ve Test olarak ayrıldı ve eğitim verileri üzerinde yeniden örnekleme yapıldı.

• Çalışmak için yeterli veriye sahip olduğumuz için azınlık sınıfının aşırı örneklenmesini kullandık. Bu, çoğunluk sınıfının 1.000 örneği varsa ve azınlık sınıfı 100'e sahipti, bu strateji azınlık sınıfını 1.000 örneğe sahip olacak şekilde aşırı örnekleyecekti.

• Aşırı Örnekleme için, RandomOverSample işlevi tüm ”değil” için imblearn paketinden kullanılır Bizim durumumuzda, sadece çoğunluk ”sınıfları olan 1 azınlık sınıfı. Yeniden örneklemeden sonra, eğitim verilerinde 9750 CB ve CB OLMAYAN örnekler vardı.

**4. Yöntemlerin Açıklaması**

**4.1.Gaussian Naive Bayes**

Naive Bayes sınıflandırıcıları, Bayes'in matematik Teoremine dayalı bir sınıflandırma algoritmaları koleksiyonudur. Basit bir ifadeyle, Bayes teoremi, olayla ilgili olabilecek koşulların ön bilgisine dayalı olarak bir olayın olasılığını tanımlar. Tek bir algoritma değil, hepsinin ortak bir ilkeyi paylaştığı bir algoritma ailesidir, yani sınıflandırılan her özellik çifti birbirinden bağımsızdır. Naive Bayes, ikili (iki sınıflı) ve çok sınıflı sınıflandırma problemleri için bir sınıflandırma algoritmasıdır. Teknik, ikili veya kategorik girdi değerleri kullanılarak açıklandığında anlaşılması en kolay olanıdır. Saf Bayes olarak adlandırılır, çünkü her hipotez için olasılıkların hesaplanması, hesaplamalarını izlenebilir kılmak için basitleştirilmiştir. Naive Bayes, en yaygın olarak bir Gauss dağılımı varsayılarak gerçek değerli niteliklere genişletilebilir.

Naive Bayes'in bu uzantısına Gaussian Naive Bayes denir. Gaussian Naive Bayes'in yanında Multinomial naive Bayes ve Bernoulli naive Bayes de mevcuttur. Gauss Naive Bayes'i seçtik çünkü en popüler olanı ve uygulaması en basit olanı çünkü bizim sadece eğitim verilerinden ortalamayı ve standart sapmayı tahmin etmemiz gerekiyor. Sınıflandırıcı, sklearn.naive bayes paketi kullanılarak uygulandı.

**4.2.Logistic Regression**

Regresyon analizi, hedef veya hedef arasındaki ilişkiyi analiz eden tahmine dayalı bir modelleme tekniğidir. Bir veri kümesinde bağımlı değişken ve bağımsız değişken. Regresyon analizi teknikleri hedeflendiğinde kullanılır. ve bağımsız değişkenler doğrusal veya doğrusal olmayan bir birbirleri ve hedef değişken arasındaki ilişki sürekli değerler içerir.

Regresyon analizi şunları içerir:

geçen bir çizgi olan en uygun çizginin belirlenmesi mesafe olacak şekilde tüm veri noktalarından Her veri noktasından gelen satırın en aza indirilmesi. Lojistik regresyon, regresyon türlerinden biridir. bağımlı olduğunda kullanılan analiz tekniği değişken kesiklidir. Örnek: 0 veya 1, doğru veya yanlış, vb. Bu, hedef değişkenin yalnızca iki değere sahip olabileceği anlamına gelir, ve bir sigmoid eğri, arasındaki ilişkiyi gösterir. haritalama yoluyla hedef değişken ve bağımsız değişken 0 ile 1 arasında bir değere herhangi bir reel değer seçtik. Veri setimizin boyutu büyük olduğundan Lojistik Regresyon, ve gelecek değerlerin neredeyse eşit oluşumu vardı hedef değişkenlerde. Ayrıca, hiçbir korelasyon yoktu Veri setindeki bağımsız değişkenler arasında Sınıflandırıcı kullanılarak uygulandı sklearn.tree model paketi.

**4.3.Decision Tree Classifier**

Veri seti ile ilgili bir dizi soru sorarak bir Karar Ağacı oluşturulur. Her yanıt alındığında, kaydın sınıf etiketi hakkında bir sonuca varana kadar bir takip sorusu sorulur. Soru dizisi ve olası cevapları, düğümlerden ve yönlendirilmiş kenarlardan oluşan hiyerarşik bir yapı olan bir karar ağacı şeklinde organize edilebilir. 3 tür düğümü vardır: Kök, Dahili ve Yaprak düğümleri. Bir karar ağacında, her yaprak düğüme bir sınıf etiketi atanır. Kök ve diğer dahili düğümleri içeren terminal olmayan düğümler, farklı özelliklere sahip kayıtları ayırmak için öznitelik test koşulları içerir. Karar algoritmasını kullanarak, ağaç kökünden başlıyoruz ve en büyük bilgi kazancı (IG) (son karara doğru belirsizlikte azalma) ile sonuçlanan özellik üzerindeki verileri bölüyoruz. Yinelemeli bir süreçte, yapraklar saf olana kadar bu bölme prosedürünü her alt düğümde tekrarlayabiliriz. Bu, her bir yaprak düğümündeki örneklerin hepsinin aynı sınıfa ait olduğu anlamına gelir. Sınıflandırıcı, sklearn.tree paket yaşı kullanılarak uygulandı.

**4.4.Random Forest Classifier**

Adından da anlaşılacağı gibi, Random Forest Classifier, bir topluluk olarak çalışan çok sayıda bireysel karar ağacından oluşur. Rastgele ormandaki her bir ağaç, bir sınıf tahmini yayar ve en çok oyu alan sınıf, modelimizin tahmini olur. Ağaçlar birbirlerini kendi bireysel hatalarından korudukları için, bireysel tahminlerin herhangi birinden daha doğru olan topluluk tahminleri üretebildikleri için, modeller arasındaki düşük korelasyon anahtardır. Torbalama süreci, modelleri çeşitlendirmek için kullanılır, çünkü her bir ağacın veri setinden rastgele örnekleme yapmasına izin verilir. Sınıflandırıcı, sklearn.ensemble paketi kullanılarak uygulandı.

**5. Deney Ve Sonuçları**

Denetimli öğrenme tekniği analizimiz için standart yöntemler olarak Naive Bayes(Gauss), Lojistik regresyon ve J48 Decision Tree kullandık. Ensemble yöntemleri olarak Random Forest Sınıflandırıcıları kullandık. Araştırmamızda, Gaussian Naive Bayes sınıflandırıcısının en kötü performansı gösterdiğini, Random Forest Sınıflandırıcısının ise her metrik açısından en iyi sonucu verdiğini bulduk.[şek1 ve şekil2]. Random Forest sınıflandırıcısının en iyi performansı gösterdiğini görmek şaşırtıcı değildi. Karar Ağacı sınıflandırıcısı, Naive Bayes sınıflandırıcısı ve Lojistik Regresyondan daha iyi performans gösterdi. Random Forest Sınıflandırıcısı, Decision Tree sınıflandırıcısının bir uzantısı olduğu için beklenen tüm performans metriklerinde en üstte çıktı ve aynının birden çok yinelemesinin sonuçlarının ortalamasını aldı. Modellerin performansını belirlemek için kullanılan Metrikler aşağıdaki gibidir:

* TP oranı çizilerek oluşturulan eğrinin altında kalan alanı ifade eden ROCArea.

TP = Doğru Pozitiflerin Sayısı (Number of True Positives)

TN = Doğru Negatiflerin Sayısı (Number of True Negatives)

FP = Yanlış Pozitiflerin Sayısı (Number of False Positives)

FN= Yanlış Negatiflerin Sayısı (Number of False Negatives)

Kullanılan Geleneksel Denetimli Öğrenme: NaiveBayes, Logistic Regression ve J48 Decision Trees sınıflandırıcısı Kullanılan Ensemble Öğrenme Yöntemleri: Random Forest sınıflandırıcı Şekil 1 ve Şekil 2, yukarıda bahsedilen algoritmalar arasındaki grafik karşılaştırmayı gösterir.

Not: Tablo IV, Kesinlik, Geri Çağırma ve F1 puanı için her iki sınıfı (nefret söylemi ve nefret söylemi olmayan) kullanan ağırlıklı ortalamayı temsil eder. Karışıklık matrislerinin ilk sütun ve satırı Siber Zorbalık sınıfını, ikinci satır ve sütun ise Siber Zorbalık Dışı sınıfını temsil etmektedir.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | NaiveBayes | Regression | DecisionTree | Random Forest |
| Accuracy | 0.62 | 0.80 | 0.85 | 0.92 |
| Precision | 0.79 | 0.81 | 0.88 | 0.92 |
| Recall | 0.62 | 0.80 | 0.85 | 0.92 |
| F1-Score | 0.62 | 0.81 | 0.85 | 0.92 |
| ROCArea | 0.68 | 0.81 | 0.87 | 0.92 |
| Confusion Matrix | 925 1504  31 1541 | 1920 509  274 1298 | 1896 533  67 1505 | 2175 254  73 1499 |

**6. SONUÇ**

Bu rapurda, çeşitli Supervised algoritmalar arasında karşılaştırmalı bir çalışma yaptık, ayrıca çeşitli Supervised Ensemble yöntemlerini de karşılaştırdık. Genel olarak en iyi performans, yaklaşık %92'lik bir doğruluk sağlayan RandomForest sınıflandırıcısı tarafından gösterildi. Topluluk yöntemleri, Denetimli yöntemlere eşit veya onlardan daha iyi performans gösterdi, ancak yine de, tüm topluluk yöntemlerinde siber zorbalık sınıfı için çok daha fazla arzu edilen yüksek bir True pozitif oranı gözlemledik. NaiveBayes en kötü performansı sergiledi ve yalnızca %61 doğruluk sağladı. Temel Bileşenler Analizi, en iyi boyutluluk azaltma algoritmalarından biridir ve özellik işleme işini kolaylaştırmanın yanı sıra, sınıflandırıcının sonuçlarını iyileştirmeye de yardımcı olabilir. Buradaki fikir, her birinin avantajlarını ve dezavantajlarını araştırmak ve sonuçlarını tek tek ve birlikte kontrol etmektir.

**6.1.Siber Zorbalık Tespiti İle İlgili Çalışmalar**

Raisi et. al [7], katılımcı kelime tutarlılığı (PVC) Modeli olarak adlandırılan benzersiz bir model sunar. Bu ilişkisel model, yüksek kalitede etiketlenmiş veriler elde etmek zor olduğundan, zayıf denetimli bir şekilde eğitilmiştir, bu modelde insan uzmanların yalnızca tacizin yüksek düzeyde göstergesi olan anahtar ifadeler biçiminde yüksek doğrulukta ek açıklamalar sağlaması gerekir. Bu algoritma daha sonra, diğer olası anahtar ifade göstergelerini ve belirli zorbalık örneklerini bulmak için etiketlenmemiş sosyal etkileşim ağındaki mağduriyet kalıplarını arayarak bu açıklamaları kullanır.

Hosseinmardi et. al ,popüler bir sosyal medya platformu Instagram'daki siber zorbalık olaylarını, kullanıcının herkese açık gönderilerindeki en iyi yorumları analiz ederek ele alıyor. Algoritma, verileri ayırmak için Naive Bayes sınıflandırmasını kullandı ve onu Siber Saldırganlıktan daha da farklılaştırdı. Siber zorbalık, medya tabanlı bir sosyal ağ bağlamında incelendi ve etiketlemede hem görseller hem de yorumlar kullanıldı. Ayrıca, bir Doğrusal SVM sınıflandırıcısının medya oturumu için metin, resim ve meta verilerden çok modlu özellikler ekleyerek siber zorbalığı belirleme doğruluğunu önemli ölçüde artırabileceğini gösterebildiler.

Dadvar et. al [9], siber zorbalık örneklerini tespit etmek için Derin sinir ağlarını kullanmaya odaklanmaktadır. Formspring, Wikipedia ve Twitter veri setlerinde Evrişimsel Sinir Ağı (CNN), Uzun Kısa Süreli Bellek (LSTM), Çift Yönlü LSTM (BLSTM) ve BLSTM gibi sinir mimarisinde karmaşıklık bakımından farklılık gösteren modeller kullanılmıştır. Nandhini et al. [10] Naif bir Bayes tabanlı öğrenme modeli önerir ve MySpace.com'un veri setini kullanır, %91 gibi yüksek bir doğruluk elde eder. Başlangıçta yaklaşık 12000 örneğe sahip olan Kelly Reynolds tarafından Kaggle.com'da bulunan Formspring verilerini kullanır, ancak ön işlemeden sonra, yarısının siber zorbalığa karşılık geldiği toplam 1608 örneğe sahip oldular, temel olarak, bu küçük veri kümesini eğitmek için kullandılar. Bir Sinir ağı ve bir SVM Sınıflandırıcı.

**6.2.Kullanılan Yöntemler ve Test Sonuçları**

|  |  |  |  |
| --- | --- | --- | --- |
| GaussianNB | LogisticRegression | DecisionTreeClassifier | RandomForestClassifier |
| Doğruluk oranı: 0.62  Confusion Matrix:  [[ 924 1505]  [ 31 1541]] | Doğruluk oranı: 0.80  Confusion Matrix:  [[1899 530]  [ 277 1295]] | Doğruluk oranı: 0.85  Confusion Matrix:  [[1913 516]  [ 71 1501]] | Doğruluk oranı: 0.91  Confusion Matrix:  [[2163 266]  [ 76 1496]] |

**7. KAYNAKÇA**

[1] DataTurks Siber Trollerin Tespiti için Tweet Veri Seti.

<https://www.kaggle.com/datasets/dataturks/dataset-for-detection-of-cybertrolls?select=Dataset+for+Detection+of+Cyber-Trolls.json>

[2] <https://www.internetyardim.org.tr/index.php/siber-zorbalik>

[3] <https://www.datacamp.com/tutorial/svm-classification-scikit-learn-python>

[4] <https://realpython.com/logistic-regression-python/>

[5] <https://stackabuse.com/random-forest-algorithm-with-python-and-scikit-learn/>

[6] <https://www.geeksforgeeks.org/decision-tree-implementation-python/>

[7] <https://www.analyticsvidhya.com/blog/2021/03/machine-learning-with-python-gaussian-naive-bayes/>