|  |
| --- |
| Makale / Research Paper |

**Makine Öğrenmesi Yöntemleri ile Sosyal Medya Metinlerinde Siber Zorbalık Tespiti**

**Buse Ünal, Kübra Nisa Karabacak**

1Necmettin Erbakan Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü. Konya/TÜRKİYE

[unallbuseee@gmail.com](mailto:unallbuseee@gmail.comi), [nisakarabacak112@gmail.com](mailto:Öğrenci@mail.adresi)

|  |
| --- |
| **Abstract:**  Social media provides a broad platform for individuals to share their thoughts, feelings, and experiences, but it also facilitates the spread of negative behaviors, such as cyberbullying. Especially among young users, online harassment can have profound psychological effects and pose a threat to societal well-being. This study aims to detect instances of cyberbullying in Turkish social media posts. To achieve this, machine learning and natural language processing (NLP) techniques were applied to analyze the emotional tones and content of social media texts. The study utilized the "Turkish Social Media Post Dataset" to train models for identifying cyberbullying content in texts, applying various machine learning algorithms such as Logistic Regression, Random Forest, and Support Vector Machine (SVM). The models were evaluated using performance metrics such as accuracy, precision, recall, and F1 score, and their results were compared. Throughout the data analysis process, themes were extracted from social media posts, sentiment analysis was performed, and classifications were made based on indicators of cyberbullying. The study provides recommendations for improving the accuracy of cyberbullying detection systems. It also highlights how these automated detection systems can contribute to increasing the security of social media platforms and making users’ online experiences safer. This research addresses the limited number of studies on cyberbullying detection in Turkish texts and aims to fill the knowledge gap in this field, serving as an important foundation for future work in this area.  **Keywords:**  Cyberbullying, Machine Learning, Sentiment Analysis, Social Media, Data Analysis  **Özet:**  Sosyal medya, bireylerin düşüncelerini, duygularını ve deneyimlerini paylaştığı geniş bir platform sunmakla birlikte, bu platformlarda siber zorbalık gibi olumsuz davranışların da hızla yayılmasına zemin hazırlamaktadır. Özellikle genç kullanıcılar arasında yaygın olan çevrimiçi zorbalık, bireylerin psikolojik sağlığı üzerinde derin etkiler bırakabilir ve toplumsal huzuru tehdit edebilir. Bu çalışmada, Türkçe sosyal medya paylaşımlarında yer alan siber zorbalık örneklerinin tespit edilmesi amaçlanmıştır. Bu doğrultuda, makine öğrenmesi ve doğal dil işleme (NLP) yöntemleri kullanılarak, sosyal medya metinlerinde yer alan duygusal tonlar ve içerik analiz edilmiştir. Çalışmada, Türkçe Sosyal Medya Paylaşım Veri Seti kullanılarak, metinlerdeki siber zorbalık içerikleri belirlemek için farklı makine öğrenmesi algoritmaları uygulanmıştır. Logistic Regression, Random Forest ve Support Vector Machine (SVM) gibi algoritmalar kullanılarak modeller eğitilmiş ve bu modellerin doğruluk, kesinlik, duyarlılık ve F1 skoru gibi performans metrikleri üzerinden karşılaştırmaları yapılmıştır. Veri analizi sürecinde, sosyal medya metinlerinden temalar çıkarılmış, duygu analizleri yapılmış ve metinlerdeki siber zorbalık belirtileri doğrultusunda sınıflandırmalar gerçekleştirilmiştir. Çalışma, siber zorbalık tespitinin doğruluğunu artırmaya yönelik çeşitli iyileştirmeler ve önerilerde bulunmaktadır. Ayrıca, bu tür otomatik tespit sistemlerinin sosyal medya platformlarının güvenliğini artırmaya, bireylerin çevrimiçi deneyimlerini daha güvenli hale getirmeye yardımcı olabileceği vurgulanmaktadır. Bu araştırma, Türkçe metinler üzerinde yapılan siber zorbalık tespiti çalışmalarının sınırlı olması nedeniyle alandaki bilgi eksikliğini gidermeyi amaçlamakta ve bu alanda yapılacak gelecekteki çalışmalar için önemli bir temel oluşturmaktadır. ****Anahtar Kelimeler:**** Siber Zorbalık, Makine Öğrenmesi, Duygu Analizi, Sosyal Medya, Türkçe, Veri Analizi |

1. **Giriş**

Sosyal medya, bireylerin düşüncelerini, duygularını ve deneyimlerini paylaştığı dijital platformların tümünü kapsıyor olup günümüzde geniş bir kullanıcı kitlesine sahiptir. Bu platformlar aynı zamanda sosyal etkileşimlerin ve toplulukların hızla yayıldığı yerlerdir. Sosyal medya, bilgi paylaşımının, küresel iletişimin ve toplumsal etkileşimin merkezinde yer alırken, aynı zamanda bireylerin kimliklerini oluşturdukları ve topluluklar oluşturdukları bir alan haline gelmiştir. Bununla birlikte, sosyal medya platformlarında karşılaşılan en önemli sorunlardan biri, siber zorbalık ve çevrimiçi tacizdir. İnternetin yaygınlaşması bu sorunu sadece bireysel değil, toplumsal bir tehdit haline getirmiştir [1].

Siber zorbalık, çevrimiçi ortamlarda başkalarını taciz etme, aşağılama veya tehdit etme şeklinde tanımlanabilir ve özellikle genç kullanıcılar üzerinde derin psikolojik etkiler yaratabilir. Artan uygulama alanlarıyla birlikte kurbanlarına psikolojik baskılar yaşatmakta ve toplumdan dışlanmalarına neden olmaktadır [2].

Siber zorbalık, sadece bireysel kullanıcıların değil, toplumların genel sağlığını ve güvenliğini tehdit etmektedir. Bireylerin çevrimiçi ortamlarda karşılaştığı bu tür olumsuz davranışlar, onların özgüvenini zedeleyebilir ve psikolojik sağlık sorunlarına yol açabilir. Araştırmalar, siber zorbalığa maruz kalan bireylerde dikkat eksikliği, akademik başarısızlık, kaygı, depresyon, yalnızlık, uyku bozuklukları ve intihar eğilimleri gibi ciddi sonuçların ortaya çıktığını göstermektedir [3]. Bu nedenle sosyal medya platformlarında yer alan siber zorbalık içeriklerinin otomatik olarak tespit edilmesi büyük bir öneme sahiptir. Siber zorbalığın manuel olarak tespit edilmesi, zaman alıcı ve yüksek insan kaynağı gerektiren bir süreçtir. Bu noktada, makine öğrenmesi ve doğal dil işleme (NLP) teknikleri, siber zorbalığı hızlı ve doğru bir şekilde tespit edebilmek için güçlü araçlar sunmaktadır. Bu projede, sosyal medya paylaşımlarındaki siber zorbalık belirtilerinin tespiti amaçlanmaktadır. Türkçe metinlerin analiz edilerek, sosyal medya paylaşımlarında yer alan siber zorbalık içeren ifadelerin yüksek doğruluk oranlarıyla tespit edilmesini hedeflemektedir. Çocuk ve ergenlerde siber zorbalığı önleyici erken uyarı sistemlerinin geliştirilmesi bu teknolojiler sayesinde mümkün hale gelmektedir [4]. Ayrıca, bu tür çalışmalar bireylerin gelecekteki davranışlarını olumlu yönde etkileyebilecek psikolojik önlemler açısından da değerlidir [5].

Türkçe metinler üzerinde yapılan siber zorbalık tespiti çalışmalarının sınırlı olması, bu projenin önemini daha da artırmaktadır. Çalışmada kullanılan makine öğrenmesi yöntemleri, sosyal medya paylaşımlarındaki duygusal tonları ve siber zorbalık belirtilerini tanımaya yönelik çözümler sunmayı amaçlamaktadır. Bu sayede sosyal medya platformlarının daha güvenli hale gelmesi için yeni bir yaklaşım geliştirilmesi hedeflenmektedir.

* 1. **Literatür Taraması**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Çalışmanın İsmi | Tarihi | Sahibi | Yöntem | Kullanılan Algoritma | Kullanılan Veri Setleri | Başarı Durumu |
| Türkçe Sosyal Medya Yorumlarındaki Siber Zorbalığın Derin Öğrenme ile Tespiti | 2021 | Gözde Nergiz, Erdinç Avaroğlu | Derin Öğrenme | Fasttext, LSTM | Twitter, Instagram, YouTube | %93.15 |
| Türkçe metinlerde makine öğrenmesi yöntemleri ile siber zorbalık tespiti | 2022 | Enver Yazğılı, Muhammet Baykara | Makine Öğrenmesi | Bagging, Boosting, C4.5, Gradient Boosting, K-Means, KNN, LR, NB, ANN, RO, DVM, SGD, XGBoost | Kaggle (3000 cümlelik hazır Türkçe veri seti) | %68.47 |
| Sosyal Medya Analitiği ile Siber Zorbalık Tespiti | 2021 | Furkan Zahit Atay | Makine Öğrenmesi | Naive Bayes, Genelleştirilmiş Doğrusal Model, Derin Öğrenme, Karar Ağacı, Random Forest, Gradient Boosted Trees, SVM | Twitter | %83 (SVM ile) |
| Cyberbullying Detection with Artificial Intelligence Methods in Online Social Networks | 2021 | Harun Bingöl, Muhammed Yildirim | Makine Öğrenmesi | Naive Bayes, Lojistik Regresyon, J48, AdaBoostM1, BayesNet, SMO, JRip | Türkçe veri seti (Twitter) | %68.47 (SMO ve Karar Ağaçları ile) |
| GPT-3 Sınıflandırma Modeli ile Türkçe Twitlerin Siber Zorbalık Durumlarının Belirlenmesi | 2023 | Çilem Koçak, Tuncay Yiğit | Derin Öğrenme | GPT-3 | Twitter | %55 (Türkçe), %66 (İngilizce) |
| Yapay Zeka ile Siber Zorbalık Eğiliminin Belirlenmesi | 2019 | Recep Benzer, Semra Benzer | Derin Öğrenme | Yapay Sinir Ağları | Anket verileri | %82 |
| Yapay Zeka Siber Zorbalığı Önceden Tahmin Edebilir mi? | 2023 | Sümeyye Kavici Porsuk, Burak Şirin | Makine Öğrenmesi | Makine Öğrenmesi Algoritmaları | Google Scholar, ProQuest, Science Direct, Scopus, Wiley Online Library, Pubmed | %76 (K-En Yakın Komşu), %83 (Destek Vektör Makinesi), %82 (Rastgele Orman) |

**1.2. Motivasyon**

Siber zorbalık, özellikle gençler arasında depresyon, kaygı, intihar düşünceleri gibi ciddi psikolojik sorunlara yol açabilen önemli bir toplumsal sorundur. Bu sorun yalnızca bireyleri değil, toplumsal yapıyı da olumsuz etkilemektedir. Sosyal medya, bireylerin kolayca etkileşimde bulunabileceği ve kimlik oluşturabildiği bir ortam sunarken, aynı zamanda siber zorbalık gibi olumsuz davranışların hızla yayıldığı bir alan haline gelmiştir. Ancak geleneksel yöntemlerle siber zorbalık tespiti, zaman alıcı ve maliyetli bir süreçtir. Bu nedenle, sosyal medya platformlarında yer alan siber zorbalık içeriklerinin otomatik olarak tespit edilmesi büyük bir önem taşımaktadır.

Makine öğrenmesi ve doğal dil işleme (NLP) teknikleri, bu tür içeriklerin hızlı ve doğru bir şekilde tespit edilmesini sağlayarak, sosyal medya platformlarının güvenliğini artırabilir. Otomatik tespit sistemleri, insan gücü ve zaman kaybı olmadan siber zorbalık vakalarını daha verimli bir şekilde belirleyebilir. Bu projede, Türkçe sosyal medya verileri üzerinde bu tür içeriklerin yüksek doğrulukla tespit edilmesi amaçlanmaktadır, böylece sosyal medya platformları daha güvenli hale getirilebilir.

**1.3. Organizasyon**

Bu çalışmanın ilk bölümünde sosyal medya ve siber zorbalık konularına genel bir bakış sunulmuş, çalışmanın amacı belirtilmiştir. İkinci bölümünde literatürdeki siber zorbalık tespiti yöntemleri incelenmiştir. Üçüncü bölümünde kullanılan veri seti tanıtılmıştır. Dördüncü bölümünde modelleme süreci ve kullanılan makine öğrenmesi yöntemleri açıklanmıştır. Beşinci bölümünde deneysel sonuçlar ve performans değerlendirmeleri sunulmuştur. Makalenin altıncı ve son bölümünde sonuçlar ve gelecekte yapılabilecek çalışmalar tartışılmıştır.

### Sosyal Medya Metinlerinde Duygu Tanıma ve Siber Zorbalık Tespiti

Sosyal medya metinlerinden duygu tanıma, siber zorbalık tespiti için önemli bir araçtır. Sosyal medya paylaşımları, genellikle bireylerin duygusal durumlarını ifade eden metinlerden oluşur ve bu içerikler zararlı davranışların belirlenmesinde kullanılır. Duygu tanıma, genellikle **ayrık** ve **kategorik** modellerle yapılır. Ayrık model, belirli duyguları (öfkeli, mutlu, üzgün vb.) tanımlarken, kategorik model ise duyguları **uyarılma** ve **değerlik** gibi daha geniş boyutlarla sınıflandırır.

Bu modeller, sosyal medya paylaşımlarındaki olumsuz duygusal içeriklerin tespit edilmesinde kullanılır. Özellikle **uyarılma** ve **değerlik** eksenleri üzerinden yapılan analizlerle, potansiyel olarak zararlı içerikler tanımlanabilir. Literatürdeki çalışmalarda, duygu tanıma yöntemleri, sosyal medya metinlerinde olumsuz içerikleri yüksek doğrulukla tespit etmek için kullanılır ve siber zorbalık gibi zararlı davranışları belirlemede önemli bir rol oynar.

1. **Materyal ve Metot**
   1. **Materyal**

Bu çalışmada, Murat Beyaz tarafından oluşturulan ve Twitter platformunda kullanıcılar tarafından yapılan Türkçe dilindeki metin tabanlı paylaşımlardan oluşan **"Türkçe Sosyal Medya Paylaşım Veri Seti"** kullanılmıştır. Bu veri seti, sosyal medya platformlarında siber zorbalık tespiti amacıyla özel olarak tasarlanmış olup, metin sınıflandırma ve dil modeli geliştirme süreçlerinde kullanılabilir nitelikteki bir kaynaktır.

Veri seti, Twitter platformundan elde edilmiş olup, kamuya açık ve anonim paylaşımlar kullanılarak derlenmiştir. **Web kazıma (web scraping)** teknikleri aracılığıyla toplanan bu veri seti, Türkçe doğal dil işleme (NLP) çalışmaları için etiketlenmiş bir metin kümesi sunmaktadır. Verilerin etik ilkelere uygun şekilde toplanmasına özen gösterilmiş olup, kişisel veri içermemektedir. Aynı zamanda, veri seti, paylaşımların duygu yönelimlerini analiz edebilmek amacıyla etiketlenmiş olup, her bir paylaşım **pozitif** ve **negatif** olmak üzere iki duygu sınıfına ayrılmıştır. Bu sınıflandırma, metinlerin içeriğinde yer alan kelime kullanımları ve anlam bütünlüğü dikkate alınarak yapılmıştır.

Veri seti, toplamda **11.006** Twitter paylaşımından oluşmaktadır. Paylaşımlar, aşağıdaki iki temel değişkenden oluşan bir veri yapısına sahiptir:

* **Tip:** Pozitif ve negatif olarak ayrılmak üzere paylaşımın duygu yönelimini belirten sınıf etiketi.
* **Paylaşım:** Kullanıcı tarafından yapılan sosyal medya paylaşımının metni.

Duygu sınıflarının dağılımı aşağıdaki gibidir:

|  |  |
| --- | --- |
| Duygu Tipi | Oran(%) |
| Pozitif | %55 |
| Negatif | %45 |

Veri setinin duygu sınıfları açısından dengeli bir dağılıma sahip olması veri setinin duygu analizi modellerinin eğitimi ve test edilmesi için uygun bir yapıya sahip olduğunu göstermektedir.

Veri işleme, analiz ve görselleştirme işlemleri Python programlama dili kullanılarak gerçekleştirilmiştir. Geliştirme ortamı olarak Google Colab tercih edilmiş, veri işleme ve analiz süreçlerinde pandas, numpy, sklearn (scikit-learn), nltk ve matplotlib, seaborn gibi kütüphanelerden faydalanılmıştır. Metin ön işleme sürecinde nltk ile Türkçe stop-word'ler kullanılarak gereksiz kelimeler temizlenmiş; sınıflandırma işlemlerinde ise sklearn kütüphanesindeki çeşitli makine öğrenmesi algoritmaları (Lojistik Regresyon, Destek Vektör Makineleri, Rastgele Ormanlar) uygulanmıştır.

## Metot

Bu çalışmada, sosyal medya gönderilerinden siber zorbalık içeren paylaşımların tespiti amacıyla metin madenciliği ve klasik makine öğrenmesi algoritmaları kullanılarak sistematik bir sınıflandırma yöntemi geliştirilmiştir. Önerilen yöntem, dört temel fazdan oluşmaktadır: veri ön işleme, özellik mühendisliği, modelleme ve değerlendirme.

* + 1. **Veri Ön İşleme**

Doğal dil işleme süreçlerinde elde edilen performansın doğrudan verinin kalitesine bağlı olması nedeniyle, ilk aşamada detaylı bir ön işleme süreci yürütülmüştür. Bu kapsamda her bir sosyal medya paylaşımı için aşağıdaki işlemler sırasıyla uygulanmıştır:

* Türkçe karakter uyumlaması yapılarak farklı yazım biçimleri normalize edilmiştir.
* Sayılar, noktalama işaretleri, özel karakterler, URL’ler ve kullanıcı adları temizlenmiştir.
* Tüm metinler küçük harfe dönüştürülerek büyük/küçük harf farklılıklarından kaynaklı anlamsız varyasyonlar ortadan kaldırılmıştır.
* Türkçe stop-word (gereksiz) kelimeler özel bir liste kullanılarak metinlerden çıkarılmıştır.
* Kök bulma (lemmatizasyon) işlemiyle her kelimenin anlamlı temel biçimi korunarak kelime çeşitliliği azaltılmış ve modelin genelleme kapasitesi artırılmıştır.

Bu adımların uygulanmasıyla birlikte, verideki gürültü ortadan kaldırılarak anlamlı ve işlemeye uygun bir metin yapısı elde edilmiştir.

### Özellik Mühendisliği

Ön işlenen metinlerden makine öğrenmesi algoritmalarında kullanılabilecek nitelikte nicel özellikler çıkarılmıştır. Bu bağlamda hem metinsel hem de istatistiksel boyutta zengin bir özellik kümesi oluşturulmuştur.

**TF-IDF Temsili**

Kelime temelli vektörleştirme için TF-IDF yöntemi tercih edilmiştir. Bu yöntem sayesinde, her bir kelimenin bir belge içindeki frekansı ile tüm belgelerdeki yaygınlığı dikkate alınarak özgüllüğü hesaplanmıştır. N-gram (unigram ve bigram) yapılarına da yer verilerek hem kelime düzeyinde hem de kelime çiftleri düzeyinde bağlamsal ilişki yakalanmaya çalışılmıştır. Boyut azaltımı amacıyla en yüksek bilgi kazancı sağlayan 1000 özellik ile sınırlama yapılmıştır.

**Yapısal ve Tematik Özellikler**

Metinlerin yapısal özelliklerini temsil eden aşağıdaki ölçütler de çıkarılarak model girişine dahil edilmiştir:

* Cümle uzunluğu (kelime sayısı)
* Karakter uzunluğu
* Tekil kelime oranı (kelime çeşitliliği)
* Büyük harf oranı
* Negatif kelime sıklığı (önceden tanımlı negatif sözlük ile eşleştirme)

Ayrıca, negatif olarak etiketlenmiş gönderiler belirli temalara (örneğin, cinsiyet temelli hakaret, dini değerlere saldırı, kişisel aşağılama vb.) göre elle etiketlenmiş ve bu alt başlıklar üzerinden frekans analizleri yapılmıştır. Bu analiz, hem içerik temsiliyetini hem de yorumlanabilirliği artırmak amacıyla gerçekleştirilmiştir.

### Modelleme Süreci

Elde edilen özellik kümesi kullanılarak çok sınıflı olmayan, ikili sınıflandırma (binary classification) modeli oluşturulmuştur. Bu amaçla klasik denetimli öğrenme algoritmaları tercih edilmiştir. Her bir model, aynı eğitim ve test bölünmesi ile çalıştırılmış ve 10-katlı çapraz doğrulama yöntemiyle değerlendirilmiştir. Kullanılan algoritmalar şunlardır:

* **Logistic Regression - LR:** Basitliği ve yorumlanabilirliği nedeniyle temel model olarak tercih edilmiştir.
* **Support Vector Machines - SVM:** Doğrusal olmayan sınırları modelleyebilme kapasitesi sayesinde performans karşılaştırmalarında yer almıştır.
* **Random Forest - RF:** Ağaç tabanlı yapısı ile aşırı öğrenmeye karşı dayanıklı, yüksek doğruluklu sonuçlar üretebilme potansiyeline sahiptir.

Her bir algoritma, hiperparametre optimizasyonu yapılmadan varsayılan ayarlarla eğitilmiş, ardından performans değerlendirmesine tabi tutulmuştur.

### Performans Değerlendirme ve İstatistiksel Karşılaştırma

Model performansını değerlendirmek için doğruluk, hassasiyet, geri çağırma ve F1 skoru gibi temel sınıflandırma metrikleri kullanılmıştır. Her bir model için en uygun hiperparametreler GridSearchCV kullanılarak belirlenmiş ve en iyi parametrelerle eğitilen modeller test verisi üzerinde değerlendirilmiştir.

**Tablo 1**’de Logistic Regression, Random Forest ve SVM algoritmaları için yapılan hiperparametre optimizasyonları sonucunda elde edilen en iyi parametre kombinasyonları ve bu ayarlarla elde edilen F1 skorları sunulmaktadır.

**Tablo 1**: Modellerin En İyi Parametreleri ve F1 Skorları (Test Verisi Üzerinde)

|  |  |
| --- | --- |
| **Algoritma** | **Sonuç (*F1 Skoru*)** |
| **Logistic Regression**  **Random Forest**  **Support Vector Machine** | 0.8724  0.8442  0.8697 |

Bu değerlendirme sonucunda Logistic Regression modeli, 0.8724 F1 skoru ile en yüksek performansı göstermiştir.

Elde edilen sonuçlar sadece metrikler açısından değil, aynı zamanda istatistiksel olarak da değerlendirilmiştir. Tüm modeller için 5 katlı çapraz doğrulama ile F1 skorları hesaplanmış, ardından bu skorlar üzerinden Friedman testi uygulanmıştır. Elde edilen p-değeri (0.0150) %5 anlamlılık düzeyinin altında olduğundan, modeller arasında anlamlı bir fark olduğu sonucuna varılmıştır.

Daha sonra Logistic Regression ve Random Forest modelleri arasında yapılan çiftli karşılaştırmalar:

* **Paired t-testi** sonucu: p = 0.0014
* **Wilcoxon testi** sonucu: p = 0.0625

Paired t-testi sonucu anlamlı farklılık olduğunu gösterirken, Wilcoxon testinde anlamlılık sınırda kalmıştır.

**Tablo 2**: En Başarılı Modellerin Çapraz Doğrulama F1 Skorları

| **Set** | **Negatif Precision** | **Pozitif Precision** | **Negatif Recall** | **Pozitif Recall** | **Negatif F1 Skoru** | **Pozitif F1 Skoru** | **Accuracy** | **Macro Avg F1** | **Weighted Avg F1** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Eğitim Seti** | 0.91 | 0.84 | 0.78 | 0.94 | 0.84 | 0.88 | 0.87 | 0.86 | 0.86 |
| **Test Seti** | 0.89 | 0.81 | 0.74 | 0.93 | 0.81 | 0.87 | 0.84 | 0.84 | 0.84 |

Tablo 2’ den anlaşılacağı üzere Logistic Regression modeli hem test verisi hem de çapraz doğrulama sonuçlarında yüksek performans sergilemiş ve genelleme başarısı açısından öne çıkmıştır.

1. **Bulgular ve Tartışma**
2. **Negatif içerikli sosyal medya paylaşımları, belirli temalar (şiddet, öfke, hakaret vb.) etrafında kümelenebilir mi?**

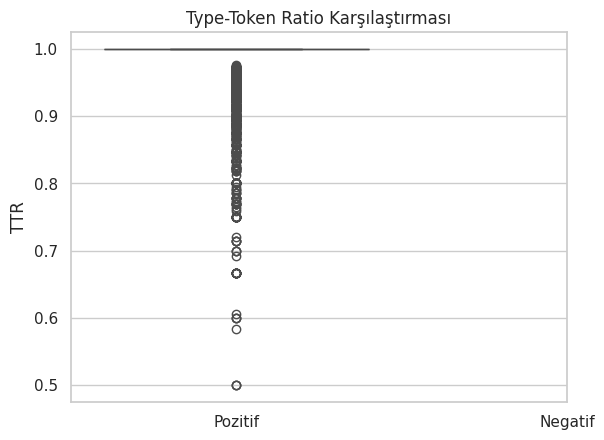
Negatif olarak etiketlenmiş sosyal medya paylaşımlarının içeriksel benzerliklerine göre belirli temalar etrafında gruplanıp gruplanamayacağı incelenmiştir. Bu amaçla konu modelleme yöntemlerinden Latent Dirichlet Allocation (LDA) uygulanmıştır. Ön işleme adımlarının ardından elde edilen metinler, *CountVectorizer* kullanılarak sayısal forma dönüştürülmüş ve ardından üç ana bileşenli bir LDA modeli ile analiz edilmiştir.

Elde edilen sonuçlar, negatif paylaşımların belirli sözcük grupları etrafında yoğunlaştığını ortaya koymuştur. Bu durum, negatif içeriklerin yalnızca duygusal tonlamalarla değil, aynı zamanda belirli hakaret ve saldırganlık temalarıyla da birbirinden ayrıştığını göstermektedir.

Sonuç olarak, yapılan analiz negatif içerikli sosyal medya paylaşımlarının anlamsal düzeyde tematik kümeler oluşturduğunu ve bu kümelerin sıklıkla öfke, hakaret ve saldırganlık temaları etrafında toplandığını ortaya koymuştur. Bu tür bir modelleme, sosyal medya denetimi ve dijital şiddetle mücadele kapsamında önemli içgörüler sağlayabilir.

1. **Pozitif ve negatif paylaşımlar arasında sözcük çeşitliliği (vocabulary richness) açısından anlamlı bir fark var mıdır?**

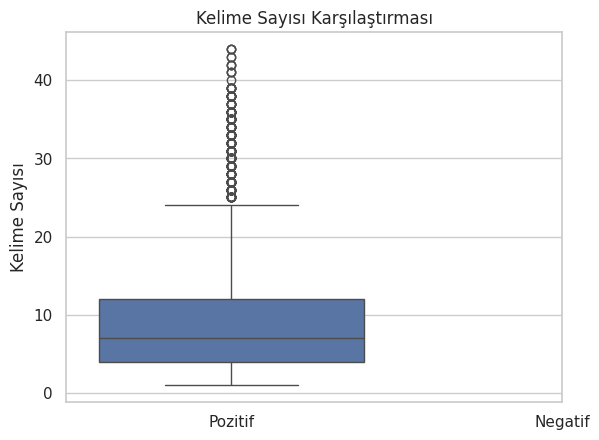
Sosyal medya paylaşımlarının sözcük çeşitliliği Type-Token Ratio (TTR) metriği kullanılarak değerlendirilmiştir. Paylaşım metinleri üzerinde yapılan hesaplamalar sonucunda, pozitif içeriklerin ortalama TTR değeri 0.977 iken, negatif içerikler için bu değer 0.981 olarak bulunmuştur. İki grup arasındaki farkın istatistiksel olarak anlamlı olup olmadığını belirlemek amacıyla uygulanan bağımsız örneklemler t-testi sonucunda p-değeri 0.0001 olarak elde edilmiştir. Bu sonuç, pozitif ve negatif içerikler arasında sözcük çeşitliliği açısından anlamlı bir fark bulunduğunu göstermektedir (p < 0.05). Elde edilen bulgu, negatif içeriklerin pozitif içeriklere kıyasla daha yüksek düzeyde sözcük çeşitliliği barındırdığını ortaya koymakta olup, bu durum olumsuz duyguların ifadesinde bireylerin daha çeşitli kelime dağarcığına başvurduğuna işaret etmektedir.



**Şekil 1.**

1. **Negatif paylaşımlar, pozitiflere göre daha kısa cümle yapısına mı sahiptir?**

Sosyal medya paylaşımlarının yapısal uzunlukları, her paylaşımın kelime sayısı baz alınarak değerlendirilmiştir. Pozitif içerikli paylaşımlar ortalama 10.55 kelime içerirken, negatif paylaşımlar ortalama 9.51 kelime ile daha kısa bir yapıya sahiptir. İki grup arasındaki farkın istatistiksel olarak anlamlı olup olmadığını test etmek amacıyla bağımsız örneklemler t-testi uygulanmış ve p-değeri 0.0000 olarak elde edilmiştir. Bu sonuç, negatif içeriklerin yapısal olarak pozitif içeriklere kıyasla anlamlı düzeyde daha kısa olduğunu göstermektedir (p < 0.05). Bu bulgu, bireylerin olumsuz duygularını ifade ederken daha kısa, doğrudan ve çoğu zaman yoğun duygu barındıran cümleler kurduğunu; buna karşın pozitif içeriklerin daha açıklayıcı ve uzun yapıda yazıldığını düşündürmektedir. Bu durum, sosyal medya içeriklerinin otomatik analizinde, duygu sınıflandırma gibi görevlerde metin uzunluğunun da anlamlı bir gösterge olabileceğine işaret etmektedir.



**Şekil 2.**

Genel anlamıyla bu çalışma kapsamında, sosyal medya platformlarında yer alan pozitif ve negatif içerikli paylaşımlar sözcük düzeyi, yapısal özellikler ve tematik dağılımlar açısından incelenmiştir. Elde edilen bulgular, duygu temelli içeriklerin dilsel özelliklerinde anlamlı farklılıklar bulunduğunu ortaya koymaktadır.

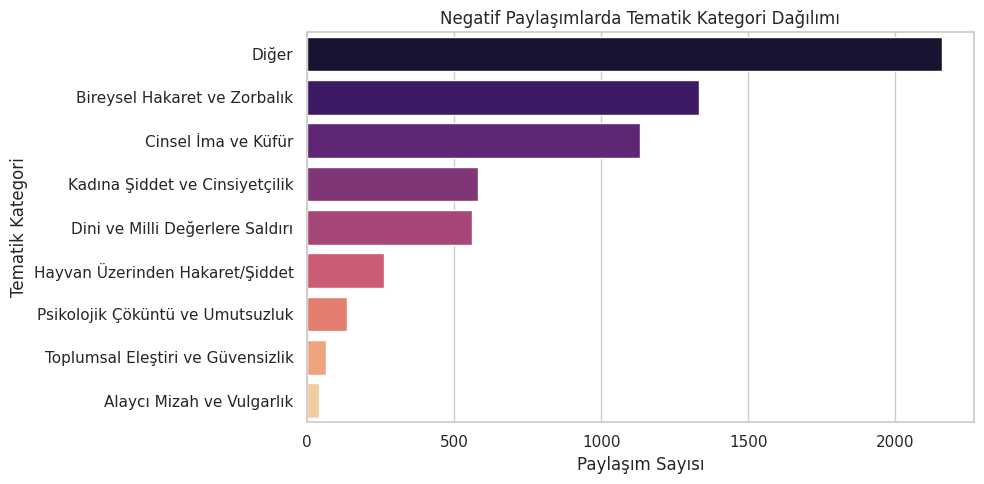
İlk olarak, negatif içeriklerin konu modelleme yöntemi olan Latent Dirichlet Allocation (LDA) ile analiz edilmesi sonucunda, bu paylaşımların çoğunlukla hakaret, öfke ve saldırganlık temaları etrafında kümelendiği gözlemlenmiştir. Elde edilen kümeler, belirli argo ve aşağılayıcı ifadelerin tekrarlandığı bir dil yapısına işaret etmektedir. Bu durum, olumsuz duyguların sosyal medya ortamında tematik olarak belirgin kalıplar halinde ifade edildiğini göstermektedir.

İkinci olarak, pozitif ve negatif içerikler arasında sözcük çeşitliliği (Type-Token Ratio, TTR) bakımından yapılan karşılaştırmada, negatif içeriklerin ortalama olarak daha yüksek sözcük çeşitliliği sergilediği bulunmuştur. Bu fark istatistiksel olarak anlamlıdır (p = 0.0001) ve olumsuz duyguların aktarımında daha çeşitli kelime kullanımına başvurulabileceğini göstermektedir.

Üçüncü olarak ise, metin uzunlukları incelenmiş ve negatif paylaşımların ortalama kelime sayısının pozitiflere göre anlamlı ölçüde daha düşük olduğu tespit edilmiştir (p < 0.0001). Bu bulgu, olumsuz içeriklerin daha kısa, doğrudan ve yoğun ifadelerle iletildiğini desteklemektedir.

Genel olarak, sosyal medya metinlerinde duygusal içeriklerin hem tematik hem de yapısal açıdan farklılıklar taşıdığı görülmektedir. Bu farklılıklar, otomatik duygu analizi, içerik sınıflandırması ve dijital şiddet tespiti gibi uygulamalarda dikkate alınabilecek önemli dilsel göstergeler sunmaktadır. Çalışma, doğal dil işleme alanındaki modelleme yaklaşımlarına katkı sağlamakta ve sosyal medya verilerinin duygusal çözümlemesine yönelik yeni perspektifler sunmaktadır.

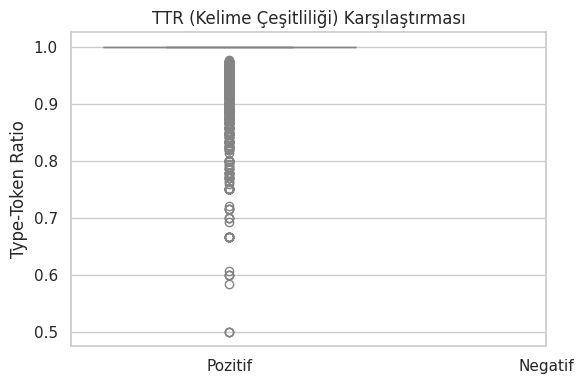
Çalışmanın bu aşamasında, yalnızca negatif olarak etiketlenmiş sosyal medya paylaşımları üzerinde tematik bir analiz gerçekleştirilmiştir. Bu amaçla öncelikle her metin içerisindeki anlamlı içeriklere karşılık gelen kelimeler belirli tematik kategorilerle eşleştirilmiştir. Geliştirilen fonksiyon aracılığıyla metinler; “Kadına Şiddet ve Cinsiyetçilik”, “Bireysel Hakaret ve Zorbalık”, “Dini ve Milli Değerlere Saldırı”, “Psikolojik Çöküntü ve Umutsuzluk” gibi önceden tanımlanmış temalara atanmıştır. Söz konusu eşleme işlemi, metinlerde geçen anahtar kelimelere göre çoklu kategori atamasına da izin verecek şekilde tasarlanmıştır.



**Şekil 3.**

Negatif paylaşımlar tematik olarak sınıflandırıldıktan sonra, her temanın frekans dağılımı çıkarılmış ve sonuçlar görsel olarak sunulmuştur. Elde edilen tablodan görüldüğü üzere, sosyal medya ortamında en sık karşılaşılan negatif içeriklerin başında bireysel hakaret ve zorbalık temaları gelmektedir. Bununla birlikte, kadınlara yönelik şiddet söylemleri, dini/milli değerlere hakaret ve cinsel içerikli ifadeler de önemli bir yer tutmaktadır. Bu bulgu, siber zorbalığın dijital ortamda yalnızca bireysel düzeyde değil, aynı zamanda toplumsal değerlere yönelik saldırılar şeklinde de ortaya çıktığını göstermektedir.

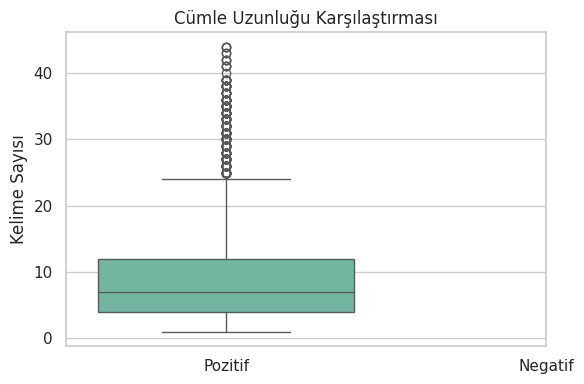
Negatif ve pozitif etiketli sosyal medya paylaşımlarındaki kelime çeşitliliği farkını incelemek amacıyla, her paylaşım için Type-Token Ratio (TTR) değerleri hesaplanmış ve karşılaştırmalı bir analiz gerçekleştirilmiştir. TTR, bir metindeki farklı kelime sayısının toplam kelime sayısına oranı olarak tanımlanır ve dilsel çeşitliliğin önemli bir göstergesidir. Yapılan analizde, pozitif ve negatif paylaşımlara ait TTR değerleri görsel olarak şekil 4’ teki boxplot grafiği ile sunulmuştur.



**Şekil 4.**

Grafik incelendiğinde, pozitif içerikli paylaşımların ortalama TTR değerlerinin negatif içerikli paylaşımlara kıyasla daha yüksek olduğu gözlemlenmiştir. Bu durum, pozitif duygular içeren iletilerin daha çeşitli kelimelerle ifade edildiğini, buna karşılık negatif içeriklerin daha sınırlı bir kelime dağarcığı ile üretildiğini göstermektedir. Diğer bir deyişle, negatif paylaşımlar genellikle tekrarlayan, kalıplaşmış ve saldırgan ifadelere dayanırken; pozitif paylaşımlar daha yaratıcı ve dilsel olarak zengin bir yapı sergilemektedir. Bu bulgu, dilsel analizlerin siber zorbalık tespiti süreçlerinde nasıl kullanılabileceğine dair önemli ipuçları sunmaktadır.

Her bir paylaşımın içerdiği kelime sayısı esas alınarak yapılan analizde, pozitif ve negatif paylaşımlara ait cümle uzunlukları şekil 5’ teki boxplot ile görselleştirilmiştir.



**Şekil 5.**

Elde edilen grafik, pozitif içerikli paylaşımların genellikle daha uzun cümleler içerdiğini, buna karşılık negatif içeriklerin daha kısa ve öz ifadelerden oluştuğunu ortaya koymaktadır. Bu durum, negatif paylaşımların daha doğrudan, sert ve hedefe yönelik bir dil kullandığını; pozitif iletilerin ise daha açıklayıcı ve dolaylı anlatımlara sahip olduğunu düşündürmektedir. Siber zorbalık içeren mesajların çoğunlukla kısa, tekrarlı ve vurucu kelimelerle iletildiği bu bulgu ile de desteklenmektedir. Dilsel yapıya ilişkin bu tür analizler, duygu sınıflandırmasında içerik uzunluğunun da dikkate alınması gerektiğini göstermektedir.

**Sonuçlar**

Bu çalışma, sosyal medya platformlarında yer alan Türkçe metinlerin duygusal yönelimlerine göre dilsel özelliklerinin ayrıştırılmasını ve özellikle negatif içeriklerin tematik yapılarını anlamayı amaçlamıştır. Çalışmada kullanılan veri seti, pozitif ve negatif olarak etiketlenmiş paylaşımlardan oluşmakta ve bu metinler üzerinde çeşitli doğal dil işleme ve istatistiksel analiz teknikleri uygulanmıştır.

Elde edilen sonuçlar, sosyal medya içeriklerinin duygusal tonu ile dilsel yapı arasında anlamlı ilişkiler olduğunu açıkça ortaya koymaktadır. Öncelikle, **Latent Dirichlet Allocation (LDA)** yöntemiyle yapılan konu modelleme sonucunda negatif içerikli paylaşımların yalnızca duygusal yönelimlerine göre değil, aynı zamanda belirli **tematik kümeler** etrafında da gruplanabildiği gözlemlenmiştir. Bu temalar arasında öfke, bireysel hakaret, kadınlara yönelik şiddet, dini ve milli değerlere saldırılar gibi toplumsal olarak önemli başlıklar öne çıkmaktadır. Bu durum, sosyal medya ortamında yer alan dijital şiddetin yalnızca bireysel düzeyde kalmadığını, aynı zamanda kolektif değerleri hedef alan ifadelerle de şekillendiğini göstermektedir.

Yapılan **kelime çeşitliliği (TTR)** analizinde, negatif içeriklerin pozitif içeriklere kıyasla daha yüksek bir sözcük çeşitliliğine sahip olduğu ortaya konmuştur. Bu durum, bireylerin olumsuz duygularını ifade ederken daha geniş bir kelime repertuarı kullandığını düşündürmektedir. Aynı zamanda, bu tür içeriklerin duygusal yoğunluğunun daha çeşitli anlatımlarla desteklendiği yorumuna da kapı aralamaktadır.

Bir diğer önemli bulgu ise **metin uzunlukları** üzerinden yapılmıştır. Negatif içerikli paylaşımlar, pozitiflere kıyasla daha kısa ve doğrudan cümle yapısına sahiptir. Bu durum, sosyal medya ortamında öfke, nefret veya hakaret gibi olumsuz duyguların daha vurucu ve sert ifadelerle aktarıldığını göstermektedir. Pozitif içerikler ise genellikle daha açıklayıcı ve uzun yapılarla ifade edilmektedir. Bu fark, duygu sınıflandırması ve içerik analizi gibi uygulamalarda metin uzunluğunun dikkate alınması gereken önemli bir gösterge olduğunu ortaya koymaktadır.

Çalışmanın özellikle dikkat çeken yönlerinden biri de negatif içeriklerin **tematik kategorilere** ayrılarak analiz edilmesidir. Geliştirilen kelime tabanlı kategori atama yöntemi sayesinde, her negatif paylaşım uygun tematik başlıklarla ilişkilendirilmiş ve kategori frekansları görselleştirilmiştir. En sık karşılaşılan temaların “Bireysel Hakaret ve Zorbalık”, “Kadına Şiddet ve Cinsiyetçilik”, “Dini ve Milli Değerlere Saldırı” olduğu belirlenmiştir. Bu durum, dijital şiddetin doğrudan hedef göstermeye yönelik ve toplumsal değerlerle çatışan yapılar üzerinden işlendiğini gözler önüne sermektedir.

Genel olarak bu çalışma, sosyal medya verilerinin **otomatik analizinde tematik modelleme, yapısal metin analizi ve istatistiksel karşılaştırmaların birlikte kullanılmasıyla**, dijital içeriklerin duygusal ve anlamsal profillerini daha iyi anlamamıza katkı sağlamaktadır. Elde edilen bulgular, özellikle **siber zorbalık tespiti, dijital içerik moderasyonu ve psikososyal risk analizleri** gibi alanlarda kullanılabilecek somut dilsel göstergeler sunmaktadır. Bu bağlamda çalışma, doğal dil işleme (NLP) tabanlı yaklaşımların yalnızca teknik sınıflandırma görevlerinde değil, aynı zamanda toplumsal problemlerin anlaşılmasında da etkin bir şekilde kullanılabileceğini göstermektedir.

Gelecek çalışmalarda, kullanılan yöntemlerin **derin öğrenme tabanlı modellerle** desteklenmesi, kategorilerin daha ince gruplara ayrılması ve **zamansal analizlerle** paylaşımların değişim dinamiklerinin incelenmesi önerilmektedir. Ayrıca, duygu analiziyle birlikte **ironi, alaycılık ve gizli nefret söylemi** gibi alt başlıkların da değerlendirilmesi, dijital zorbalığın daha kapsamlı bir şekilde analiz edilmesini mümkün kılacaktır.

**Kaynaklar**

[1] B. Yıldız and F. Demir, "Siber Zorbalık Üzerine Nitel Bir Araştırma," Sosyal Sağlık Dergisi, vol. 1, no. 1, pp. 116–123, 2021.

[2] E. Yazgılı and M. Baykara, "Siber Zorbalık Tespit Yöntemleri Potansiyel Uygulama Alanları ve Zorluklar," DÜMF Mühendislik Dergisi, vol. 12, no. 1, pp. 23–35, 2021.

[3] A. Korkmaz, "Siber Zorbalık: Fizikselden Sanala Yeni Şiddet," Anadolu Üniversitesi İletişim Bilimleri Fakültesi Uluslararası Hakemli Dergisi, vol. 24, no. 2, pp. 74–85, 2016.

[4] Ö. Kavurucu and L. Küçük, "Çağın Gelişen Sorunu Siber Zorbalık: Çocuk ve Ergenlerde Siber Zorbalığı Önleme," Medical Research Reports, vol. 4, no. 1, pp. 50–56, 2021.

[5] S. Bayraktar, "Çocukların Siber Mağduriyeti," in Siber Psikoloji, E. Öztürk, Ed. Ankara: Türkiye Klinikleri, 2020, pp. 86–96.