



Türkçe Nefret Söylemi Problemi Analizinde Farklı Sınıflandırma Algoritmalarının Ve Özellik Seçimi Yöntemlerinin Performans Karşılaştırması

Özlem Yakar^{1*}, Büşra Büyüktanır², Abdullah Emir Çil³, Ayşe Berna Altınel Girgin⁴

^{1*} Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye, (ORCID: 0000-0002-9346-7325), ozlemyakar@marun.edu.tr

² Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye (ORCID: 0000-0003-2571-4029), busra.buyuktanir@marmara.edu.tr

³ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye (ORCID: 0000-0001-7632-2389), acil@marun.edu.tr

⁴ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye (ORCID: 0000-0001-5544-0925), berna.altinel@marmara.edu.tr

(İlk Geliş Tarihi 9 Ekim 2023 ve Kabul Tarihi 17 Aralık 2023)

(DOI: 10.5281/zenodo.10646678)

ATIF/REFERENCE: Yakar, Ö., Büyüktanır, B., Çil, A.E. & Altınel Girgin, A.B. (2024). Türkçe Nefret Söylemi Problemi Analizinde Farklı Sınıflandırma Algoritmalarının Ve Özellik Seçimi Yöntemlerinin Performans Karşılaştırması. *Avrupa Bilim ve Teknoloji Dergisi*, (53), 97-111.

Öz

Günümüzde teknolojinin gelişmesine bağlı olarak sosyal medya kullanımı ve sosyal medya ağlarının sayısı gün geçtikçe artmaktadır. Bu durum, sosyal medyada nefret söylemi içeriklerinin sayıca artması ve buna bağlı olarak bu içeriklerin paylaşılması anlamına gelmektedir. Sosyal medya kullanıcılarını olumsuz yönde etkileyen bu paylaşımların önüne geçebilmek adına kamu otoriteleri ve sosyal medya geliştiricileri tarafından alınan önlemler yetersizdir. Bu sebeple, hacimce büyük miktarda üretilen nefret söylemi içeriklerinin tespitini kolaylaştıran otomatik sistemlere ihtiyaç duyulmaktadır. Öte yandan, İngilizce ve diğer pek çok dilde yapılan çalışmalar incelendiğinde, Türkçe'de nefret söylemi içeriklerinin tespitinde yeterli düzeyde çalışmaya henüz varılamamıştır. Mevcut çalışmalar incelendiğinde ise, küçük boyutta veri kümesi kullanımı, özellik seçim yöntemlerinin azlığı ve sınıflandırma algoritmalarının benzerliği görülmüştür. Literatürde bahsedilen tüm bu eksikliklerin giderilmesi için, bu çalışma gerçekleştirilmiştir. Çalışma kapsamında, Türkçe dilinde farklı boyutlarda iki adet tweet veri kümesi (veri kümesi_1k, veri kümesi_2k) kullanılmıştır. Öncelikle veri kümesi üzerinde beş farklı özellik seçimi algoritması uygulanarak özellik sayısı 1.000'e düşürülmüştür. Daha sonra nefret söylemi tespiti için, her bir özellik seçiminden elde edilen 1.000 özellik üzerinde sırası ile makine öğrenmesi algoritmalarından K-En Yakın Komşu (KNN), Rasgele Orman (RF), Naive Bayes (NB), Destek Vektör Makinesi (SVM), Uzun Kısa-Sürelili Bellek (LSTM) ve Word2Vec+SVM yöntemleri uygulanmıştır. Deneysel sonuçlarda, başarı ölçütü olarak F-ölçüm değeri kullanılmıştır. Veri kümesi_1k için en iyi sonucu %88,81 F-ölçüm oranıyla OneR özellik seçimi ile NB algoritması vermiştir. Veri kümesi_2k için en iyi sonucu %87,71 F-ölçüm oranıyla Info-Gain özellik seçimi ile NB algoritması vermiştir. Beş farklı özellik seçimiyle yapılan deneylerde, LSTM algoritması diğer tüm algoritmalara göre daha başarılı olmuştur.

Anahtar Kelimeler: Türkçe nefret söylemi tespiti, Siber zorbalık, Türkçe tweetler, Makine öğrenmesi, LSTM.

Performance Comparison Of Different Classification Algorithms And Feature Selection Methods In Turkish Hate Speech Problem Analysis

Abstract

Today, due to the development of technology, use of social media and the number of social media networks are increasing day by day. This means that the number of hate speech contents on social media increases and, accordingly, these contents are shared. The measures taken by public authorities and social media developers to prevent these posts that negatively affect social media users are insufficient. For this reason, automatic systems that facilitate the detection of hate speech content produced in large quantities are needed. On the other hand, when studies conducted in English and many other languages are examined, sufficient studies have not been achieved yet in detecting hate speech content in Turkish. When existing studies are examined, use of small datasets, scarcity of feature selection methods and similarity of classification algorithms are seen. This study is carried out to eliminate all these

* Sorumlu Yazar: ozlemyakar@marun.edu.tr ; ozlemyakar.34@gmail.com

deficiencies mentioned in the literature. Within the scope of the study, two tweet datasets of different sizes in Turkish (dataset_1k, dataset_2k) are used. First of all, five different feature selection algorithms are applied on these datasets and the number of features is reduced to 1,000. Then, machine learning algorithms are used on 1,000 features obtained from each feature selection to detect hate speech. These algorithms are K-Nearest Neighbor (KNN), Random Forest (RF), Naive Bayes (NB), Support Vector Machine (SVM), Long Short-Term Memory (LSTM) and Word2Vec+SVM methods, respectively. In experimental results, F-measurement value is taken as a criterion of success. In the experimental results, F-measure is used as the performance metric. NB algorithm with OneR feature selection is achieved the best result for Dataset_1k with an F-measure of 88.81% . NB algorithm with Info-Gain feature selection is achieved the best result for Dataset_2k with an F-measure of 87.71% . It has been observed that LSTM algorithm is more successful than all other algorithms in the experiments with five different feature selections.

Keywords: Turkish hate speech detection, Cyberbullying, Turkish tweets, Machine learning, LSTM.

1. Giriş

Günümüzde iletişim yöntemleri sürekli bir değişim halindedir. İletişim, teknolojideki gelişmelere paralel olacak şekilde sosyal medya tarafına doğru ilerlediği söylenebilir. Yeni iletişim kanallarından olan sosyal medya araçları, insanlara hem olumlu biçimde hem de olumsuz biçimde yorumlar şeklinde geri dönmektedir. Bu yorumlardan en çok göze çarpan kısım "nefret söylemi"dir. Nefret söylemi, günlük hayatta ya da dijital teknolojiler kullanılarak gerçekleştirilen bir zorbalık türüdür. Nefret söyleminde, insanların günlük hayatta ifade etmekte zorlandığı ya da ifade ettiğinde fazla tepki alacağını düşündüğü söylemleri, gerçek kimliğini kullanarak veya gizli bir kimliğe (nickname) bürünerek ifade edilmektedir. Nefret söylemini oluşturan ifadeler, belli bir dereceye kadar söyleyenin kişisel düşüncesini oluşturmaktadır. Bu düşüncelerde nefret söylemi ifadelerinin tespiti için bir arka plan bilgisi gerekir. Sosyal medyada nefret söylemi içerikleri sürekli üretilmekte ve farklı sosyal medya araçları tarafından hızla yayılarak daha büyük kitlelere erişimi sağlanmaktadır. Nefret söylemi içerikleri, hedef alınan kişileri korkutmaya, sinirlendirmeye, kızdırmaya veya utandırmaya yönelik yapılmış olan davranışlardır. Bu davranışlar başkalarının duygularına olumsuz bir biçimde hitap etme şekli ile kendini gösterir. Genellikle çocukların ve gençlerin hedef haline getirildiği nefret söylemleri, bu ifadelerle maruz kalan farklı kişileri de olumsuz şekilde etkilemektedir. Dolayısıyla, iletişim açısından sosyal medyada yapılan etkileşimler, birçok açıdan olumsuz paylaşımların ortaya çıkmasına neden olur. Bu paylaşımlar arasında, çok sayıda nefret söylemi içeriklerinin üretimi de nefret söylemindeki artışın en önemli sebepleri arasında gösterilmektedir.

Bu çalışmada, sosyal medyadaki Türkçe nefret söylemi problemleri üzerinde bazı özellik seçimleri ve makine öğrenmesi algoritmaları kullanılarak eğitilen modellerin performans karşılaştırması yapılmıştır. Model eğitiminde, iki farklı veri kümesi kullanılmıştır. Modeldeki verilerin etiketlenmesi "nefret söylemi", "saldırgan ifade" ve "hiçbiri" şeklinde üç sınıfa ayrılmıştır. Çalışmanın geri kalanı şu şekilde düzenlenmiştir: 2. Bölüm'de, literatürde sosyal medyada yer alan Türkçe dilinde nefret söylemleri üzerinde yapılan çalışmalar incelenmiştir. 3. Bölüm'de, deneylerde kullanılan veri kümelerine dair metrikler ve algoritmalar anlatılmıştır. 4. Bölüm'de, deneylerin sonuçları özetlenmiş ve bu sonuçlara göre değerlendirme yapılmıştır. Son olarak 5. Bölüm'de ise, çalışmanın sonuçları tek çatı altında toplanarak çalışma hakkında genel çıkarımlarda bulunulmuştur.

2. Literatür Taraması

Sosyal medya kullanım oranının artması ve kullanıcılarına her türlü düşünceyi paylaşma imkânı sağlaması, içerik üretimini artırmıştır. Artan içerikler arasında nefret söylemini barındıran ifadeler de yer almaktadır. Kamu ve sosyal medya geliştiricileri, artan nefret söylemi içeriklerine dair önlemler almaya çalışmaktadır. Fakat üretilen içerik hacmi çok büyük olduğu için alınan önlemler yetersiz kalmaktadır. Bunun yanında, nefret söylemi içeriklerine dair otomatik sistemlere ihtiyaç duyulmaktadır. Ayrıca, bu alanda Türkçe dili üzerinde kapsamlı bir çalışma da mevcut değildir. Tüm bu ihtiyaçlar göz önüne alındığında, literatürdeki bu eksiklikleri kapatmak ve duyulan ihtiyaca karşılık vermek adına, bu bölümde Türkçe dilinde yapılan nefret söylemlerine dair çalışmalar incelenmiştir.

2021 yılında Mayda ve ark. (Mayda, 2021) yapmış olduğu çalışma, geliştirilen bu çalışma için örnek olarak alınmıştır. Çalışma kapsamında, öncelikle KNIME aracı kullanılarak 1.000 adet Türkçe tweet toplanmış ve veri kümesi oluşturulmuştur. Ardından iki değerlendirmeci tarafından veriler etiketlenmiştir. Etiketlenen veriler "nefret söylemi", "saldırgan ifade" ve "hiçbiri" şeklinde üç sınıftan oluşmaktadır. Verileri etiketlemede kararsız kalındığında, sağlıklı bir sonuç üretebilmek adına üçüncü bir değerlendirmecinin görüşü alınmıştır. Etiketli veri kümesi başka çalışmalarda kullanılabilmesi adına açık kaynak olarak paylaşılmıştır. Oluşturulan veri kümesi üzerine Karar Ağacı (J48), NB, RF ve Sıralı Minimal Optimizasyon (SMO) olmak üzere dört farklı makine öğrenmesi algoritması uygulanmıştır. WEKA aracı kullanılarak testler gerçekleştirilmiş ve başarı ölçütü olarak F-ölçüm değeri dikkate alınmıştır. En başarılı performansı %79,9 F-ölçüm değeriyle SMO algoritması vermiştir.

Dinçer ve ark. (Dinçer, 2022), sosyal medya platformlarındaki nefret söylemlerinin tespitine dair bir çalışma yapmıştır. Veriler için Twitter uygulaması kullanılmış ve Python'un Tweepy kütüphanesi ile veriler çekilmiştir. Kullanılan veri kümesi toplamda 3.001 adet veriden (1.503 adeti nefret söylemi içeren ve 1.498 adeti nefret söylemi içermeyen tweetler) oluşmaktadır. Yapılacak olan yeni yorumlarda nefret söylemi tespiti için makine öğrenmesi algoritmalarından SVM, NB ve Lojistik Regresyon (LR) kullanılmıştır. Bu çalışmanın sonucuna bakıldığında, nefret söylemi tespitini iyileştirdiği görülmüştür. Kullanılan algoritmalar doğruluk (accuracy), hassasiyet (recall), kesinlik (precision) ve F1-skor (F1-score) bakımından incelendiğinde, bu yöntemlerden en iyi performansı %87 doğruluk oranı ile LR algoritması göstermiştir.

Beyhan ve ark. (Beyhan, 2022) tarafından yapılan çalışmanın amacı, Türkiye'de yaygın olan toplumsal sorunlara ilişkin tweetlerdeki nefret söyleminin değerlendirilmesidir. Çalışmada, Twitter'daki görünürlüğünün hacim olarak yüksek olması nedeniyle İstanbul Sözleşmesi ve Türkiye'deki mültecilere karşı tutumlar ele alınmıştır. Türkçe nefret söyleminin otomatik tespiti için bir makine öğrenmesi kullanılarak iki ayrı alanda toplanan (İstanbul Sözleşmesi veri kümesi ve Mülteciler veri kümesi) tweetlerden oluşan bir nefret söylemi veri kümesi sunulmuştur. İstanbul Sözleşmesi veri kümesi, Türkiye'nin İstanbul Sözleşmesi'nden çekilmesinin ardından atılan tweetlerden oluşmaktadır. Mülteciler veri kümesi ise, göçmenlerle ilgili tweetlerin, göçmenlere dair yaygın olarak kullanılan anahtar kelimelere göre filtrelenerek toplanmasıyla oluşturulmuştur. Twitter'daki ek açıklamaları toplamak için, veri kümesi toplama aracı olarak, açık kaynak kodlu LabelStudio açıklama aracı kullanılmıştır. Veri kümeleri incelendiğinde, İstanbul sözleşmesi veri kümesi için, Türkiye'nin İstanbul Sözleşmesi'nden çekildikten sonra 20-25 Mart 2021 arasındaki beş güne ait tweetler esas alınmıştır. Bu süre boyunca en çok konuşulan konularda 284.989 tweet toplanmıştır ve veri kümesiyle ilgisiz olanlar çıkarıldıktan sonra yaklaşık 10.000 tweet ile işlem yapılmıştır. Mülteci veri kümesi için, Ocak 2020-Eylül 2021 tarihleri arasında Twitter Academic API kullanılarak göçmenlerle ilgili tweetler toplanmıştır. Göçmenlerle ilgili yaygın olarak kullanılan mülteci (refugee), göçmen (immigrant) ve Suriyeli (Syrian) gibi anahtar kelimeleri esas alınmıştır. Çalışmanın sonucunda, toplanan veri kümesi için BERTurk dönüştürücü (transformer) mimarisi kullanılarak bir nefret söylemi tespit sistemi geliştirilmiştir. İstanbul Sözleşmesi veri kümesi 5-katlı çapraz doğrulama yöntemi kullanılarak değerlendirilmiştir. İkili sınıflandırma doğruluğunda (binary classification accuracy), İstanbul Sözleşmesi veri kümesi için %77 ve Mülteci veri kümesi için %71 doğruluk oranları elde edilmiştir. Ayrıca, İstanbul Sözleşmesi ve Mülteciler veri kümeleri için [0-4] ölçeğinde 0,66 ve 0,83 Kök Ortalama Kare Hatası (Root Mean Squared Error – RMSE) regresyon modeli de kullanılarak testler gerçekleştirilmiştir.

Kuş'un (Kuş, 2021) yaptığı çalışma, Covid-19 salgını müddetince, mültecilere yönelik yapılmış olan dijital nefret söylemlerinin incelenmesi üzerinedir. YouTube platformunda Covid-19 konusunda mültecileri hedef gösteren videoların yorum kısmına yazılmış olan YouTube kullanıcı yorumları Facepager uygulaması ile toplanarak veri kümesi oluşturulmuştur. Nefret söylemi barındıran kullanıcı yorumları dikkate alındığında, veri kümesi "Amaca Yönelik Nefret Söylemi", "Hedef Ülkenin Dokunulmazlığı", "Kişisel Tercih ve Özelliklere Yönelik Nefret Söylemi" ve "Diğer" etiketleriyle veri kümesi sınıflandırılmıştır. Ayrıca, nefret söylemi içeriğine sahip yorumların sözel yansıması da bir ağ haritasıyla bu çalışmada görselleştirilmiştir. Veriler analiz edilirken, içerik analizi yöntemi ve metin madenciliği metotlarından yararlanılmıştır. Verilerin ön işleme, ayıklanması ve algoritması için eğitim setinin oluşturulması aşamasında Microsoft Excel, Google Sheets ve KNN sınıflandırma yöntemi kullanılmıştır. Sözcüklerin sayımı için Rapidminer Studio, eğitim setinin güvenilirliğinin hesaplanması için ağ haritası ile SPSS 21 ve tasarlama kısmında ise Gephi 0.9.2 kullanılmıştır. Çalışmanın sonuçları incelendiğinde, sınıflandırmadaki etiketlere göre dağılımlar Nefret Söylemi İçeren Yorumların Kategorik Dağılımına göre; Hedef Ülkenin Dokunulmazlığı (%50), Kişisel Tercih ve Özelliklere Yönelik Nefret Söylemi (%17), Amaca Yönelik Nefret Söylemi (%13) ve Diğer (%20) şeklindedir.

Karayığit ve ark. (Karayığit, 2021) yaptığı çalışmada, nefret söylemleri Instagram yorumlarını tespit etmek için küfürlü ve küfürlü olmayan etiketli Instagram yorumundan oluşan bir veri kümesi (ATC) oluşturularak Evrişimli Sinir Ağları (CNN) ve makine öğrenmesi yöntemleri kullanılmıştır. Nefret söylemleri yorumları algılamak için yapay sinir ağları tabanlı CNN modeli ile makine öğrenmesi tabanlı SVM, Karar Ağacı (DT), NB, RF, LR, Adaptif Güçlendirme (AdaBoost) ve Aşırı Gradyan Güçlendirme (XGBoost) modelleri karşılaştırılmıştır. Nefret söylemleri Türkçe yorumlarından oluşan veri kümesi, 2017–2019 yılları arasında 10.528 küfürlü ve 19.826 küfürlü olmayan Instagram yorumlarından oluşmaktadır. Veri kümesi toplama aracı olarak Python kullanılmış ve çalışmadaki veri kümesi oluşturulmuştur. Oluşturulan ATC veri kümesi için, sınıf sayılarını eşitlemede aşırı örnekleme metodu (Over-sampling method) kullanılarak bir versiyonu oluşturulmuştur. Özellik seçimi ve ağırlıklandırma için Terim Frekansı-Ters Belge Frekansı (TF-IDF) metotlarından Kelime Çantası (Bag of Words – BoW), bigram ve trigram algoritmaları ile kelime yerleştirme için Byte-Çifti Kodlama (BPE), Word2Vec ve BERT modelleri kullanılmıştır. Deney aşamasında, CNN algoritmasıyla kelime yerleştirme modellerinin kombinasyonlarının oluşturulduğu ve makine öğrenmesi algoritmalarıyla TF-IDF modellerinin kombinasyonlarının oluşturulduğu modeller hem orijinal ATC hem de aşırı örnekleme ATC veri kümesi üzerinde uygulanmıştır. Deney sonuçlarında, 0,974 ile mikro ortalamalı F1 puanı, 0,973 ile makro ortalamalı F1 puanı ve 0,946 Kappa değeri ile aşırı örnekleme ATC veri kümesinde CNN modelinin en iyi sonuca ulaştığı görülmüştür. Böylece, makine öğrenmesi modellerinden CNN modelinin en iyi performansa ulaştığı söylenebilir.

Karayığit ve ark. (Karayığit, 2022) çalışması, sosyal medyadaki Türkçe yorumlarının homofobik ya da cinsiyetçi, ağır aşağılama ve dışlama ifadeleri gibi nefret yorumlarını içerip içermediğinin tespiti için yapılmıştır. Burada, önceden eğitilmiş Çok Dilli M-BERT algoritması kullanılmıştır. Veri kümesi, Instagram'dan alınan yorumlarla oluşturularak nefret ifadesi tespitinde modelleri eğitmek için kullanılmıştır. Deneylerde resHATC ve orijinal HATC olmak üzere veri kümelerinin iki formu kullanılmıştır. Homofobik-Küfürlü Türkçe Yorumlar (HATC) ve Kötü Amaçlı Türkçe Yorumlar (ATC) veri kümeleri birleştirilerek, uygulamalarda kullanılmak üzere dengeli bir veri kümesi elde edilmiştir. Araştırmacılar tarafından toplanan HATC veri kümesi 10.237 nefret dolu, 1.226 homofobik ve 19.827 nötr Instagram yorumlarından oluşmaktadır. Uygulamalarda M-BERT modeli, derin öğrenme tabanlı modeller (LSTM, Çift Yönlü Uzun Kısa-Süreli Bellek (BiLSTM), Kapılı Tekrarlayan Birim (GRU)), makine öğrenmesi tabanlı modeller (SVM, NB, RF) ve Topluluk Sınıflandırıcıları (AdaBoost, XGBoost, GBoost) kullanılarak sonuçlar karşılaştırılmıştır. Burada en iyi sonucu bulmak için, algoritma modellerinin performansında F1 puanı (F1-score), kesinlik (precision) ve duyarlılık (recall) performans ölçütleri dikkate alınmıştır. En iyi performansı HATC veri kümelerindeki (homofobik F1 puanı: %82,64 , nefret dolu F1 puanı: %91,75 , nötr F1 puanı: %96,08 , ortalama F1 puanı: %90,15) M-BERT modelinin verdiği görülmüştür.

Başka kişi veya kurumların itibarını zedeleyici, kızdırıcı, rahatsız edici, utandırıcı, algı oluşturucu yazı, ses, video ve fotoğraflar gibi kişiler üzerinde olumsuz etki bırakan, dijital ortamlar ve sosyal medya aracılığıyla rahatsızlık veren tüm eylemlere siber zorbalık denilmektedir. Siber zorbalık ve nefret söylemi arasında sıkı bir ilişki vardır. Siber zorbalık, bir tür nefret söylemidir. İnternet ve sosyal medya platformları, insanların anonimlik duygusuyla hareket etmesine ve düşüncelerini açıkça ifade etmesine olanak

tanılmaktadır. Bu da insanların nefret söylemi içerikli mesajlar yayınlamasına veya siber zorbalık yapma olasılığına yol açmaktadır. Bu sebeple, çalışma kapsamında yapılan literatür taramasında, sosyal medyada Türkçe dilinde yapılan siber zorbalığa yönelik çalışmalar da incelenmiştir.

Yazgılı ve ark. (Yazgılı, 2021), siber zorbalık tespiti için bir çalışma gerçekleştirmiştir. Bu çalışmada, 3.000 adet cümleden (1.503 tane siber zorbalık içeren ve 1.497 tane siber zorbalık içermeyen veriler) oluşan hazır Türkçe veri kümesini kullanmışlardır. Veri kümesi toplama aracı olarak Kaggle sitesinden yararlanılmıştır. Çalışma kapsamında Yapay Sinir Ağı (ANN), C4.5 Ağacı (C4.5 Tree), Bagging, Boosting, Gradyan Güçlendirme (GBoost), XGBoost, SVM, KNN, K-Means, NB, LR, RF ve Stokastik Gradyan Azaltma (Stochastic Gradient Descent) algoritmaları kullanılarak elde edilen sonuçlar karşılaştırılmıştır. Deneysel sonuçlardan, %88,35 başarı oranı ile LR algoritması en yüksek sonucu vermiştir. Kullanılan veri kümesinin yeni olması ve çok sayıda algoritma ile test yapılması çalışmanın özgün yanını ortaya koymaktadır.

Özellikle Kovid-19 salgınından sonra internette daha fazla vakit geçirilmesi ve işlenen suçun caydırıcılığının olmaması düşüncesi siber zorbalığın artmasına sebep olmuştur. Bingöl ve Yıldırım (Bingöl, 2021), artan siber zorbalığın durdurulması ve mağduriyetin önüne geçilmesi adına metin madenciliği ve yapay zekâ yöntemlerini kullanarak siber zorbalık tespiti yapan model geliştirmiştir. Model eğitimi için Twitter'dan elde edilen Türkçe dilinde veri kümesi kullanılmıştır. Veri kümesi toplama aracı olarak açık kaynak kodlu GitHub platformu üzerinden sunulmuştur. Veri kümesi, 11.111 adet etiketli Türkçe tweetten oluşmaktadır. Veri kümesi üzerine metin ön işleme adımları uygulandıktan sonra yedi farklı yapay zekâ algoritması kullanılmıştır. NB, LR, J48, AdaBoostM1, BayesNet, SMO ve JRip algoritmaları ile model eğitimi gerçekleştirilmiş ve deney sonuçları karşılaştırılmıştır. Yapılan deneysel sonuçlara göre, %77-%92 aralığındaki doğruluk oranıyla en iyi sonuçları LR, J48 ve BayesNet algoritmalarının verdiği görülmüştür.

Yılmaz ve ark. (Şahiner Yılmaz, 2021), sosyal medya üzerinde saldırgan dil kullanımı ile siber zorbalığa maruz kalan kullanıcılar için çalışma gerçekleştirmiştir. Çalışmada saldırgan dil tespiti için iki adet veri kümesi kullanılmıştır. Birinci veri kümesi, Türkçe dilinde Twitter üzerinden toplanan 31.756 tweetten (6.131 adet "saldırgan" ve 25.625 adet "saldırgan değil" etiketli veriler) oluşan OffenseEval yarışma verileridir. İkinci veri kümesi ise Türkçe dilinde Twitter üzerinden toplanan yaklaşık 1 milyona yakın tweetten oluşan etiketsiz veri kümesidir. Çalışmada kelime temsil yöntemi olarak Word2Vec yöntemi kullanılmıştır. Uygulamalarda veri kümesi toplama aracı olarak Python kullanılmıştır. Çalışmanın model eğitimi kısmında BiLSTM ve LSTM algoritmaları kullanılmıştır. Deneysel sonuçlardan, %86 doğruluk oranıyla LSTM algoritması modellerdeki başarı oranını vermektedir.

Nergiz ve Avaroğlu (Nergiz, 2021), Türkçe yorumlar içeren sosyal ağlar üzerinde siber zorbalığın tespiti için LSTM modelini önermektedir. Twitter, Instagram ve Youtube platformlarından elde edilen veri kümesi, 90.000 adet siber zorbalık içeren ve 90.000 adet siber zorbalık içermeyen Türkçe yorumlardan oluşmaktadır. Bu yorumlar, Python programlama dili ve Selenium aracı kullanılarak toplanmıştır. Verilere ön işleme yapmak için Zemberek kütüphanesinden faydalanılmıştır. Kelimeleri sayısal ifadelerle çeviren Fasttext için Ngram, Word2Vec için Skip-Gram ve Doc2Vec için PV-DBOW kelime gömme yöntemleri kullanılarak LSTM'nin eğitimi için üç farklı model oluşturulmuştur. Bu modeller sırasıyla Word2Vec, Fasttext ve Doc2Vec için Modeli 1, Modeli 2 ve Modeli 3 olarak isimlendirilmiştir. Sınıflandırma modeli olan LSTM'nin gizli katmanında Sigmoid aktivasyon fonksiyonu, dense katmanında ReLU aktivasyon fonksiyonu, çıktı katmanında Softmax aktivasyon fonksiyonu kullanılmıştır. Deney sonuçlarına göre en başarılı modelin %93,15 başarı oranıyla Modeli 2'de LSTM algoritması olduğu gözlemlenmiştir.

Ön ve Yeniterzi (Ön, 2020), Türkçe dilindeki siber zorbalık içeren ifadelerin tespiti için CNN modelini önermiştir. Modelin değerlendirilmesi için CNN ile kurulan modellerin yanında SVM modeli de kullanılmıştır. Veri kümesi hazırlama aşamasında TurkishCyberbullying veri kümesinden yararlanılmıştır. Bu veri kümesi, 3.000 tweetten (1.500 siber zorbalık ve 1.500 siber zorbalık içermeyen tweetler) oluşmaktadır. Ön işleme aşamasında Doğal Dil Araç Seti (Natural Language Toolkit – NLTK) kütüphanesinden yararlanılmıştır. SVM modelinin eğitimi için BoW modeli ile TF-IDF yöntemi kullanılmıştır. Türkçe Vikipedi ile eğitilmiş Word2Vec (KVT1) ve FastText (KVT2) kelime vektör temsilleri ve haber dokümanı ile eğitilmiş FastText kelime vektör temsilleri (KVT3), CNN modellerinin eğitiminde kullanılmak için tercih edilmiştir. Ayrıca, CNN modelinin eğitimi için bu çalışma kapsamında Rastgele kelime temsil modeli de (Random word representation model) eğitimlerde kullanılmıştır. CNN modelinin çıktı katmanında Sigmoid aktivasyon fonksiyonu, ADAM optimizasyon algoritması ve ikili çapraz entropi kayıp fonksiyonu kullanılmıştır. Deney aşamasında, en iyi sonucu CNN Rastgele modeli vermiştir. Kullanılan beş model içerisinde en iyi sonucu veren modeller için Topluluk modeli (Ensemble model) uygulanmıştır. Topluluk modelinin kullanımında, eşitlik olması durumunda CNN Rastgele modelinin sonuçları dikkate alınarak 0,937'lik bir F1-skoruna ulaşılmıştır.

Özçift ve ark. (Özçift, 2019) çalışmasında, Türkçe metinlerde bulunan siber zorbalık ifadelerini tespit eden bir sistem için makine öğrenmesi modelini önermektedir. Önerilen modelin uygulaması için 3.000 tweetten (1.500 siber zorbalık içeren, 1.500 siber zorbalık içermeyen etiketli veriler) oluşan 11.534 öznitelikli TurkishCyberbullying veri kümesi kullanılmıştır. Veri ön işleme aşamasında, veri kümesinin boyutunu azaltmak için Ki-kare Testi (Chi-square Test) uygulanarak öznitelik sayısı 579 kelimeye indirilmiştir. İlgisiz sözcükleri ortadan kaldırmak için TF-IDF yöntemi uygulanmıştır. Model olarak Lojistik Regresyon ile Bayes teoremi kombinasyonundan oluşan Bayesçi Lojistik Regresyon (Bayesçi LR) yöntemi kullanılmıştır. Önerilen modelin parametreleri İzgara Arama algoritması (Grid Search algorithm) kullanılarak optimize edilmektedir. Deney sonucunda, Bayesçi LR algoritmasının F-skoru değeri 0,925 başarı oranı şeklindedir.

Tablo 1’de farklı sosyal medya araçlarında Türkçe nefret söylemine yönelik literatürde yapılan çalışmalara yer verilmiştir.

Tablo 1. Sosyal medyada nefret söylemlerini tespit etmeye yönelik önceki çalışmalar (Table 1. Previous studies on detecting hate speech on social media)

Çalışma	İlgili Konu	Uygulama Alanı	Veri Kümesi Boyutu	Veri Kümesi Toplama Aracı	Kullanılan Yöntem	En Başarılı Model	Başarı Oranı
Mayda ve ark. (2021)	Türkçe nefret söylemi tespiti	Twitter	1.000 adet Türkçe tweet	KNIME	J48, NB, RF, SMO	SMO	%79,9 (F-ölçüm)
Dinçer ve ark. (2022)	Nefret söylemi tespiti	Twitter	3.001 adet Türkçe tweet [1.503 adet nefret söylemi içeren, 1.498 adet nefret söylemi içermeyen tweetler]	Python (Tweepy kütüphanesi ile)	SVM, NB, LR	LR	%87 (Doğruluk)
Beyhan ve ark. (2022)	Türkçe nefret söyleminin otomatik tespiti	Twitter	İstanbul Sözleşmesi veri kümesi, Mülteciler veri kümesi	LabelStudio	BERTurk transformer mimarisi, k-katlı çapraz doğrulama	k-katlı çapraz doğrulama (k=5)	İstanbul Sözleşmesi veri kümesi: %77; (Doğruluk, F-ölçüm) Mülteci veri kümesi: %71 (Doğruluk, Duyarlılık)
Kuş (2021)	Mültecilere yönelik dijital nefret söylemlerinin tespiti	YouTube	Kovid-19 konusunda mültecileri hedef alan YouTube kullanıcı yorumları	Facepager	İçerik analizi yöntemi (Microsoft Excel, Google Sheets); Metin madenciliği yöntemleri (Rapidminer Studio, SPSS 21, Gephi 0.9.2); KNN	KNN	—
Karayiğit ve ark. (2021)	Türkçe nefret söylemi tespiti	Instagram	ATC veri kümesi [10.528 küfürlü, 19.826 küfürlü olmayan Instagram yorumları]	Python	CNN, NB, SVM, DT, RF, LR, AdaBoost, XGBoost	CNN	%97,4 (F-ölçüm)
Karayiğit ve ark. (2022)	Sosyal medyadaki Türkçe yorumların homofobik ya da nefret söylemi tespiti	Instagram	HATC (Homofobik-Küfürlü Türkçe Yorumlar) veri kümesi [10.237 nefret dolu, 1.226 homofobik, 19.827 nötr Instagram	—	M-BERT modeli, LSTM, BiLSTM, GRU, SVM, NB, RF, AdaBoost, XGBoost, GBoost	M-BERT modeli	HATC veri kümesi: %90,15 (F-ölçüm)

Çalışma	İlgili Konu	Uygulama Alanı	Veri Kümesi Boyutu	Veri Kümesi Toplama Aracı	Kullanılan Yöntem	En Başarılı Model	Başarı Oranı
			yorumları]				
Yazgılı ve Baykara (2021)	Siber zorbalık tespiti	—	3.000 adet siber zorbalık veri kümesi [1.503 tane siber zorbalık içeren, 1.497 tane siber zorbalık içermeyen veriler]	Açık kaynak hazır veri kümesi (KAGGLE)	C4.5 Tree, Bagging, Boosting, GradientBoosting, Stokastik Gradient Descent, XGBoost, KNN, K-Means, SVM, ANN, NB, LR,RF	LR	%88,35 (Doğruluk)
Bingöl ve Yıldırım (2021)	Siber zorbalık tespiti	Twitter	11.111 adet Türkçe tweet veri kümesi	Açık kaynak hazır veri kümesi (GitHub)	NB, LR, J48, AdaBoostM1, BayesNet, SMO, JRip	LR, J48, BayesNet	%77 - %92 (Doğruluk)
Şahiner Yılmaz ve ark. (2021)	Siber zorbalık tespiti	Twitter	OffensEval yarışma veri kümesi [31.756 adet] ve yaklaşık 1 milyona yakın Türkçe tweet veri kümesi	Python	LSTM, BiLSTM	LSTM	%86 (Doğruluk)
Nergiz ve Avaroğlu (2021)	Siber zorbalık tespiti	Twitter, Instagram, Youtube	180.000 adet Türkçe yorumlar [90.000 adet siber zorbalık içeren, 90.000 adet siber zorbalık içermeyen veriler]	Python programlama dili, Selenium	LSTM	LSTM	%93,15 (F-ölçüm)
Ön ve Yeniterzi (2020)	Türkçe dilindeki siber zorbalık içeren ifadelerin tespiti	TurkishCyberbullying	3.000 tweetten oluşan TurkishCyberbullying veri kümesi	NLTK kütüphanesi	CNN, SVM	CNN	%93,7 (F-ölçüm)
Özçift ve ark. (2019)	Türkçe metinlerde bulunan siber zorbalık ifadelerinin tespiti	TurkishCyberbullying	3.000 tweetten oluşan TurkishCyberbullying veri kümesi	—	Bayesçi LR	Bayesçi LR	%92,5 (F-ölçüm)

3. Materyal ve Metot

3.1. Veri Kümesi

Bu çalışmada, iki adet Türkçe veri kümesi kullanılmıştır. Birincisi, 2021 yılında Mayda ve ark. (Mayda, 2021) hazırladığı çalışmada kullandıkları ve açık kaynak olarak GitHub ortamında paylaştıkları Türkçe tweetlerden oluşan etiketli hazır veri kümesidir. Çalışma kapsamında bu veri kümesi, veri kümesi_1k şeklinde ifade edilmiştir. İkincisi ise yine bu çalışma kapsamında Python dili

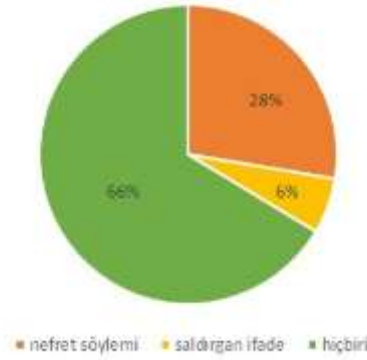
kullanılarak Twitter üzerinden toplanan Türkçe tweetlerden oluşan veri kümesidir. Bu veri kümesi ise, yazarlar tarafından toplanarak etiketlenmiş olup veri kümesi_2k şeklinde ifade edilmiştir.

Her iki veri kümesi oluşturulmadan önce hedef gruplara ait anahtar kelimeler belirlenmiştir. Bunun için, önce yerel ve ulusal medyada nefret söylemi haberlerinin yer aldığı NefretSoylemi.org sitesindeki içerikler incelenmiştir. Ardından, hedef gruplara ait anahtar kelimeler belirlenirken, sitede yer alan nefret söylemi içeriklerinde en sık kullanılan nefret söylemi ifadeleri ile Hrant Dink Vakfı tarafından hazırlanan “Medyada Nefret Söylemi ve Ayrımcı Söylem 2018 Raporu”nun (Engindeniz, 2018) sonuçları göz önüne alınmıştır. Veri kümesi_1k için belirlenen anahtar kelimelere sahip 1.000 tweet, Mayda ve ark. tarafından KNIME aracı kullanarak toplanmıştır. Çalışma kapsamında hazırlanmakta olan veri kümesi_2k için ise yine belirlenen anahtar kelimelere sahip 2.049 tweet Python aracı kullanılarak Twitter üzerinden toplanmıştır.

Toplanan tweetlerin sınıflandırılması için, literatürdeki çalışmalar incelendiğinde, saldırgan ifadeler ile nefret söylemi ifadelerinin aynı anlamdaymış gibi etiketlendiği görülmüştür. Yani, literatürdeki çalışmalarda veri kümesi, nefret söylemi içeren ve nefret söylemi içermeyen şekilde etiketlenmiştir. Ancak saldırgan ifadeler ile nefret söylemi ifadeleri birbirinden farklı kavramlardır. Mayda ve ark. tarafından veri kümesi_1k, “saldırgan ifade”, “nefret söylemi” ve “hiçbiri” olmak üzere üç sınıfta etiketlenerek paylaşılmıştır. Yine bu çalışmada da Türkçe dilinde nefret söylemi, veri kümesi_2k veri kümesi oluşturulurken toplanan tweetler “saldırgan ifade”, “nefret söylemi” ve “hiçbiri” olmak üzere üç sınıfta etiketlenmiştir. Bir ifadenin nefret söylemi içerip içermediği farklı değerlendiricilerin farklı yorumları şeklinde ele alınabilmektedir. Yani, bir değerlendirici “nefret söylemi” olarak etiketlediği veriyi, bir başka değerlendirici “saldırgan ifade” şeklinde etiketleyebilir. Burada oluşabilecek karmaşayı önlemek için, tweetler önce iki değerlendirmeci görüşüyle etiketlenmiştir. Etiketlemede kararsız kalındığında, sağlıklı bir sonuç üretebilmek adına üçüncü bir değerlendirmecinin görüşü alınmıştır.

Mayda ve ark. tarafından etiketli olarak paylaşılmış veri kümesi_1k'nın etiketine göre sınıf dağılım sayısı Şekil 1'de yer almaktadır. Etiketleme ifadeleri “saldırgan ifade (60)”, “nefret söylemi (276)” ve “hiçbiri (664)” olmak üzere üç sınıf olarak belirlenmiştir.

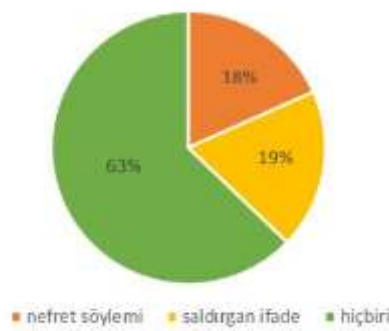
Veri Kümesi_1k Sınıf Dağılımı



Şekil 1. Veri kümesi_1k'nın etiketlerine göre dağılım sayısı (Figure 1. Number of distributions by labels of dataset_1k)

Veri kümesi_2k'nın etiketine göre dağılım sayısı Şekil 2'de yer almaktadır. Etiketleme ifadeleri “nefret söylemi (375)”, “saldırgan ifade (387)” ve “hiçbiri (1.287)” olmak üzere üç sınıf olarak belirlenmiştir.

Veri Kümesi_2k Sınıf Dağılımı



Şekil 2. Veri kümesi_2k'nın etiketlerine göre dağılım sayısı (Figure 2. Number of distributions by labels of dataset_2k)

3.2. Ön İşleme ve Özellik Çıkarımı

Eğitilecek modelin daha anlamlı ve doğru sonuçlar üretebilmesi için her iki veri kümesi üzerinde ön işleme adımları gerçekleştirilmiştir. Veri içerisinde yer alan tweetlerin her kelimesi küçük harfe çevrilmiştir. URL adresleri, kullanıcı adları, emojiler,

hashtagler, mentionlar, ünlem, nokta, tek tırnak işareti, soru işareti ve çift tırnak işareti gibi noktalama işaretleri silinerek veri kümesi temizlenmiştir. Kelime aralarında yer alan fazla boşluk tek boşluğa indirilmiştir. Tek başına anlamı olmayan ve model eğitiminde rol oynamayan bağlaçlar da kaldırılmıştır.

Kelime yapılarının incelenmesi için açık kaynak kodlu Türkçe doğal dil işleme kütüphanesi olan Zemberek kütüphanesi (Akın, 2007) tercih edilmiştir. Kelime unigram ve kelime bigramları bulunarak, bu kelimeler özellik kümesi olarak belirlenmiştir. Kelime Ngramları için, Zemberek kütüphanesi ile incelenen kelimenin kök ve ekleri listelenmektedir. Kelimeden çekim ekleri çıkartılarak önce elde edilen kök ve daha sonra yapım eklerinden oluşan gövdeler veri kümesi içerisinde tutulmuştur.

Nihaî veri kümesinde özellik olarak "tweetId", "tweet", "etiket", elde edilen kelime unigramlar ve kelime bigramlar yer almaktadır. Veri kümesi oluşturulurken her bir tweet içerisinde geçen kelime unigramlar ve kelime bigramlar kontrol edilerek, bulunduğu kelime adet sayısının bilgisi (yani kelime frekansı) tutulmuştur. Veri kümesi_1k için 1.000 adet satır ve 16.383 adet sütundan oluşan bir veri kümesi elde edilmiştir. Veri kümesi_2k için ise 2.049 adet satır ve 28.625 adet sütundan oluşan bir veri kümesi elde edilmiştir.

3.3. Özellik Seçimi ve Sınıflandırma

Çalışma kapsamında sırası ile veri kümelerinde yer alan 16.383 ve 28.625 özellik üzerine, özellik seçimi için ayrı ayrı Bilgi Kazancı Özellik Değerlendirici (InfoGainAttributeEval), Fisher Skor (Fisher Score), OneR Sınıflandırıcı (OneRClassifier), Ki-kare Testi (Chi-square test) ve Kazanç Oranı Özellik Değerlendirici (GainRatioAttributeEval) algoritmaları uygulanmıştır. Her bir algoritma için her iki veri kümesi en önemli 1.000 özelliğe indirgenmiştir. Özellik seçimi için kullanılan algoritma ile çıkan 1.000 özellik üzerine ayrı ayrı KNN, RF, NB, SVM, LSTM ve Word2Vec+SVM gibi farklı algoritmalar uygulanmıştır. Tüm bu algoritmalara ait deneyler, Python dilinde pandas v1.3.5, Scikit learning 1.0.2, sklearn.ensemble ve Pytorch 1.13.0 kütüphaneleri kullanılarak gerçekleştirilmiştir. Ayrıca SVM algoritması, 'C'=1, 'dual'='false', 'penalty' = 'l2' parametreleri ile LinearSVC sınıflandırıcı 'scikit learning' kütüphanesi kullanılarak çalıştırılmıştır. LSTM algoritması için ise kullanılan hiper-parametreleri Şekil 3'te verilmiştir.

```
# Build the model
print('Build model...')
model = Sequential()
model.add(LSTM(128)(hidden_nodes, return_sequences=False, input_shape=(word_vec_length,
char_vec_length)))
model.add(Dropout(0.6))
model.add(Dense(2))
model.add(Activation('softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam')
batch_size=1000
model.fit(train_x, train_y, batch_size=batch_size, epochs=10, validation_data=(validate_x,
validate_y))
```

Şekil 3. Deneylerde kullanılan LSTM hiper-parametreleri (Figure 3. LSTM hyper-parameters used in experiments)

Tüm deneyler boyunca 10-katlı çaprazlama (10-fold Cross Validation) tekniği kullanılmıştır. Bu teknikte veri kümesi varsayılan değer (default) olarak %90 eğitim - %10 test şeklinde ayrılmaktadır. Çalışma kapsamında gerçekleştirilen deneylerde ise, her iki veri kümesi önce 70 eğitim - %30 test şeklinde ve daha sonra %30 eğitim - %70 test şeklinde ayrılmıştır. Son durumda dört farklı sonuç elde edilerek tüm sonuçlar paylaşılmıştır.

3.3.1. Bilgi Kazancı Özellik Değerlendirici (InfoGainAttributeEval)

InfoGain Attribute Eval kullanılarak özellik çıkarımını hesaplamak için entropi kavramına ihtiyaç duyulmaktadır. Entropi, sistemdeki düzensizliğin ölçüsüdür. Veri kümesi içerisindeki her bir özelliğin değeri sınıfa göre bilgi kazanım oranını ölçerek elde edilmektedir (Hall, 1999). H entropi olmak üzere Bilgi Kazancı hesaplama formülü Eşitlik (1)'de paylaşılmıştır.

$$\text{Bilgi Kazancı} = H(\text{Sınıf}) - H\left(\frac{\text{Sınıf}}{\text{Özellik}}\right) \quad (1)$$

3.3.2. Fisher Skor (Fisher Score)

Fisher Skor, en yaygın kullanılan denetimli öznitelik seçim yöntemlerinden birisidir (Gu, 2012). Bu yöntemde Fisher Skor fonksiyonu, maksimum olasılık tahmini ile derinden ilişkilidir. Bu fonksiyon, iki sınıfı birbirinden en iyi ayıran özellikleri seçmek için kullanılır; bu özelliklere göre Fisher Skor'un en yüksek değerde olan özellikleri alınır. Fisher Skor yönteminde, her bir sınıfın özelliklerine dair standart sapma ve ortalama değerleri kullanılarak bir bağlantı skoru hesaplanır. Fisher Skor hesaplama formülasyonu Eşitlik (2)'de (Budak, 2018) gösterilmiştir.

$$F(x_i) = \frac{|\mu_i^+ - \mu_i^-|}{|\alpha_i^+ - \alpha_i^-|} \quad (2)$$

Burada, + ve - sembolleri iki sınıflı problemler için farklı sınıfları; μ_i^+ ve μ_i^- sembolleri sınıfların aritmetik ortalamalarını, α_i^+ ve α_i^- sembolleri de sınıflara ait standart sapma değerlerini belirtmektedir.

3.3.3. OneR Sınıflandırıcı (OneRClassifier)

Özellik değeri OneR sınıflandırma algoritması ile hesaplanmaktadır. OneR algoritmasının çalışma adımları aşağıda paylaşılmıştır (Novaković, 2011).

1. Eğitim verisi içerisinde her özellik için kural oluşturmak.
2. Oluşturulan her bir kural için sınıflandırma doğruluklarını hesaplamak.
3. En az hatalı kurallara sahip özellikleri seçmek.

3.3.4. Ki-kare Testi (Chi-squared test)

Ki-kare Testi, gözlenen ve beklenen frekanslar arasındaki farkın, istatistiksel olarak anlamlılık ilkesine dayandığı bir özellik seçim yöntemidir. Bu test, genellikle birbirinden bağımsız iki değer niteliksel olarak belirtilen verilerin kriterlerini analiz ve test etmek için kullanılır. Bunun yanında Ki-kare Testinin uygulaması, ölçümle belirlenen sürekli değişkenler, belirli bir derecenin altında ya da üstünde olacak şekilde nitelendirilerek uygulanır.

Ki-kare Testinde metod, özellikler (X) ile Y arasında bağlantı olup olmadığını test etmektedir. Yapılan test sonucunda, Y ile ilişkisi olmadığı tespiti sonucu özellikler (X) veri kümesinden çıkartılır. Ki-kare Testi için hesaplama formülasyonu aşağıda Eşitlik (3), Eşitlik (4) ve Eşitlik (5)'te (Gu, 2012) gösterilmiştir.

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(N_{ij} - \hat{N}_{ij})^2}{\hat{N}_{ij}} \quad (3)$$

$$\hat{N}_{ij} = \frac{N_i \cdot N_j}{N} \quad (4)$$

$$d = (I - 1)(J - 1) \quad (5)$$

Burada N_{ij} , Y'nin i ve X'in j düzeyindeki gözlenen birim sayısı; \hat{N}_{ij} iki özellik bağımsız iken Y'nin i ve X'in j düzeyindeki beklenen birim sayısı; d test istatistiğinin analizi için kullanılan Ki-kare dağılımında serbestlik kademesini belirtmektedir.

3.3.5. Kazanç Oranı Özellik Değerlendirici (GainRatioAttributeEval)

Kazanç oranı kullanıldığında özellik değeri, sınıfa göre kazanım oranı ölçülerek hesaplanmaktadır. Oran sonucunda 0 ile 1 aralığında değer ortaya çıkmaktadır. Bu değer özellik değeri olarak kabul edilmektedir. Oran sonucu 0 ise özellik ile sınıf arasında hiçbir ilişki olmadığı sonucuna varılmaktadır. 1 ise özelliğin doğru tahmin edilebildiğini gösterir (Hall, 1999). Kazanç Oranı formülasyonu Eşitlik (6)'daki gibi hesaplanmaktadır.

$$\text{Kazanç Oranı} = \frac{\text{Bilgi Kazancı}}{H(\text{Özellik})} \quad (6)$$

3.3.6. K-En Yakın Komşu algoritması (K-Nearest Neighbor algorithm – KNN)

Sınıflandırma işleminde gözetimli öğrenme metodlarından olan K-En Yakın Komşu (KNN) algoritması, en basit sınıflandırma tabanlı makine öğrenmesi yöntemi olarak ifade edilmektedir. Bu algoritma içerisinde, tahmin edilecek değer bağımsız değişkenlerden oluşan vektörler aracılığıyla en yakın komşularının hangi sınıfta daha yoğun olarak toplandığı bilgisi üzerinden hareket ederek, bağımsız değişkenin sınıfını tahmin etmede kullanılan bir algoritma modelidir. Bu algoritma, basit ve etkili örüntü tanıma yöntemlerindedir; gerçekleştirimi kolay, öğrenme sürecinin kullanışlı, basit ve güçlü olmasından ötürü, makine öğrenmesi algoritmaları arasında sınıflandırmada ve regresyon çalışmaları içerisinde yaygın olarak tercih edilir (Yazgılı, 2021).

Sınıflandırma aşamasında k değeri, incelenecek eleman sayısını belirtir. KNN algoritması beş adımdan oluşur. İlk adım k değerinin belirlenmesidir. İkinci adımda, k değerinin diğer nesnelere hedef nesneye olan Öklid (veya benzeri) uzaklıkları hesaplanır. Üçüncü adımda, uzaklıklar sıralanarak en küçük uzaklığa bağlı olan en yakın komşular bulunur. Dördüncü adımda, en yakın komşu kategorileri belirlenir. Son adımda ise, en uygun komşu kategorisi seçilerek işlem tamamlanır.

3.3.7. Rasgele Orman algoritması (Random Forest algorithm – RF)

Denetimli öğrenme algoritmalarından birisi olan Rasgele Orman (RF) algoritması, regresyon ve sınıflandırma problemlerinde sıklıkla tercih edilir. RF algoritması, birden fazla karar ağacı üretir ve sınıflandırma işlemi esnasında sınıflandırmayı yaparken algoritma değerini yükseltmeyi hedefler. Algoritmanın mantığı, birbirinden bağımsız şekilde çalışan karar ağaçlarının bir araya gelmesiyle aralarından en yüksek puana sahip değer seçilmesi işlemine dayanır. Bu metodla, rastgele seçilen bir altküme aracılığıyla pek çok sınıflandırma ağacı içerisinde yeni topluluklar oluşturulabilir. RF algoritması sürekli, kategorik, sürekli ve kategorik yapıdaki veri kümelerinde ve farklı boyutlara sahip büyük veri kümelerinde kullanılmaktadır (Yazgılı, 2021).

3.3.8. Naive Bayes algoritması (Naive Bayes algorithm – NB)

Naive Bayes (NB) algoritması, istatistik temelli bir gözetimli öğrenme algoritmasıdır. Metin tabanlı olarak kullanılacak dokümanlarda sınıflandırma işlemi yaparken, eğitim veri kümelerinin tümü üzerinde koşullu olasılıkları hesaplanarak gerçekleştirilir (Yazgılı, 2021). Bu metodun en belirgin özelliği, uygulamasının basit kullanışlılıkta olması ve başarılı sonuçlar elde edilebilmesi mantığıdır.

NB sınıflandırma modeli, olasılık ölçütlerine göre tanımlanan hesaplamalarla sisteme verilen verilerin kategorisini tespit etmektedir. Bu sınıflandırma modelinde, sisteme belirli oranlarda eğitilmiş verilerin girdisi yapılır. Eğitim için verilerin girdisi mutlaka bir sınıfa ait olmalıdır. Eğitilmiş veriler üzerindeki olasılık işlemleriyle, sisteme girdi olarak verilen yeni test verileri, önceden

elde edilen olasılık değerleriyle birlikte karşılaştırılır. Daha sonra sisteme verilen test verisinin hangi sınıflandırmaya dahil olduğu belirlenir. Burada eğitilmiş veri sayısının çok olması tercih sebebidir; bu sayede test verisinin gerçek sınıfını tespit etme işi kolaylaşacaktır.

3.3.9. Destek Vektör Makinesi algoritması (Support Vector Machine algorithm – SVM)

Destek Vektör Makinesi (SVM) algoritması, denetimli öğrenme yöntemlerinden birisi olup genellikle sınıflandırma ve regresyon analizi problemlerinde kullanılır. SVM, doğrusal ve doğrusal olmayan verileri sınıflandırabilir; ancak genelde verileri doğrusal olarak sınıflandırmaya çalışır. SVM’de bir düzlem üzerine yerleştirilen noktaları ayırmak için, doğrusal bir doğru çizilerek, düzlemdeki sınıflar ayrıştırılır. Bu şekilde bu doğrunun, iki sınıfın noktalarına ait maksimum uzaklıkta olması hedeflenir. Ancak doğrusal çizgiler, doğrusal olmayanlar kadar her zaman başarılı olmayabilir. Doğrusal olarak birbirinden ayrılamayan sınıflarda, “kernel trick” adı verilen bir yöntem kullanılır. Bu yöntemde kernel fonksiyonu, doğrusal olarak ayrılamayan sınıfları, doğrusal olarak ayrılabilir duruma getirir ve daha başarılı sonuçlar elde edilir. Böylece eldeki verilerden yeni verileri tanımlamak için, büyük miktarda veri analizi gerçekleştirilebilir (Beken, 2022).

3.3.10. Uzun Kısa-Sürelili Bellek algoritması (Long Short-Term Memory algorithm – LSTM)

Uzun Kısa-Sürelili Bellek (LSTM) algoritması, derin öğrenme yöntemlerinden yinelemeli sinir ağı modelidir. Sıralı verileri modelleyip, bir sonraki adımı tahmin etmek için kullanılan Tekrarlayan Sinir Ağı (RNN) algoritmasının bir türüdür. LSTM, RNN’ye göre daha uzun süreli bilgiyi hatırlayabilen bir mimariye sahiptir (Kara, 2019). LSTM mimarisinde gelen girdiler önemli olup olmaması durumuna göre bir sonraki aşamaya aktarılır. Bu işlem kapılar (gates) ve hücre durumları (cell state) ile gerçekleşmektedir. Unutma kapısına (Forget Gate) bir önceki katmadan gelen bilgiler, Sigmoid fonksiyonu uygulanarak geçer. Geçen bilgi 0 ile 1 arasında değer alır. Bilgi ne kadar 0’a yakınsa önemsiz olur ve unutulur, ne kadar 1’e yakınsa önemli olur ve tutulur. Girdi kapısı (Input Gate) hücre durumlarını günceller ve ağı düzenlemek için, Tanh fonksiyonunu kullanır. Hücre durumları (Cell State) yapay sinir ağı içerisinde bilgiyi hücreler arasında taşır. Çıktı kapısı ise (Output Gate) bir sonraki katmana gidecek bilgiye karar verir (Eid, 2022).

3.3.11. Word2Vec + SVM algoritması (Word2Vec + SVM algorithm)

Word2Vec, bir Doğal Dil İşleme tekniği olup kelimeleri vektör uzayında ifade etmeye yarayan denetimsiz (gözetimsiz) ve tahmin temelli (prediction-based) bir makine öğrenmesi modelidir. Temelinde ANN ile iki farklı model kullanarak kelimelerin eğitilmesi amaçlanıp geliştirilir. Eğitildikten sonra, böyle bir model eş anlamlı kelimeleri algılayabilir veya kısmi bir cümle için ek kelimeler önerebilir. Word2Vec, her bir farklı kelimeyi vektör adı verilen belirli bir sayı listesiyle temsil eder. Vektörler, kelimelerin anlamsal ve sözdizimsel niteliklerini karşılayacak şekilde dikkatlice seçilir. Bu nedenle, bu vektörler tarafından temsil edilen kelimeler arasında basit bir matematiksel fonksiyon (kosinüs benzerliği) ile anlamsal benzerlik düzeyi gösterilebilir (Word2Vec).

Denetimli (gözetimli) öğrenme yöntemlerinden birisi olan SVM, genellikle sınıflandırma ve regresyon analizi problemlerinde kullanılır. Word2Vec+SVM ise, iki farklı makine öğrenmesi tekniğinin birleşimini ifade eden bir yöntemdir. Bu iki tekniğin kombinasyonu, metin sınıflandırma görevleri için kullanılır. Bu yöntemde, metin verileri kelimeler arasındaki bağlantıyı gösteren vektörlere dönüştürülür. Ardından bu vektörler bir SVM sınıflandırıcısı tarafından kullanılarak metinler sınıflandırılır. Word2Vec+SVM yöntemi, pek çok farklı metin sınıflandırma görevlerinde başarıyla kullanılabilir.

4. Araştırma Sonuçları ve Tartışma

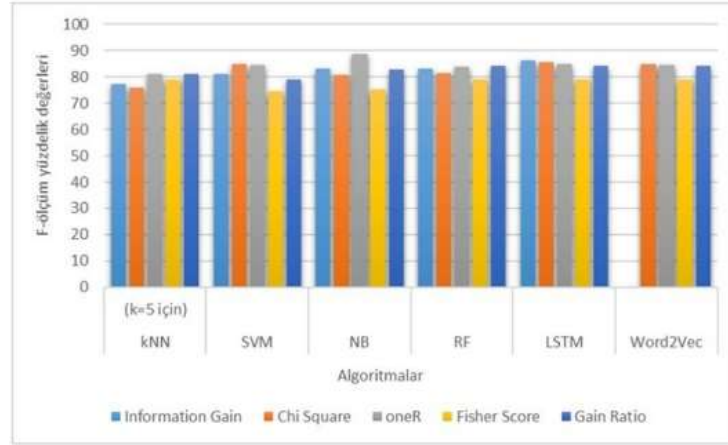
Bu bölümde makine öğrenmesi algoritmaları kullanılarak gerçekleştirilen deney süreçleri ve elde edilen sonuçlar paylaşılmaktadır. Çalışma kapsamında iki adet veri kümesi kullanılmıştır. Her iki veri kümesi üzerinde özellik seçimi yapılarak 1.000’er özellik elde edilmiş ve altı farklı makine öğrenmesi algoritması ile model eğitimi gerçekleştirilmiştir. Model eğitilirken 10-katlı çapraz-doğrulama (10-fold cross-validation) tekniği kullanılmıştır. Bu teknikte veri kümesi varsayılan değer (default) olarak %90 eğitim - %10 test şeklinde ayrılmaktadır. Çalışma kapsamında gerçekleştirilen deneylerde ise her iki veri kümesi önce %30 eğitim - %70 test şeklinde ve daha sonra %70 eğitim - %30 test şeklinde ayrılarak, her iki durum için elde edilen sonuçlar incelenmiştir. Veri kümelerinin etiket dağılımı dengeli olmadığı için performans başarı ölçütü olarak F-ölçüm değeri esas alınmıştır.

Tablo 2’de veri kümesi_1k için %70 eğitim - %30 test olarak ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmaları kullanılarak elde edilen sınıflandırma performansları yer almaktadır.

Tablo 2. Veri kümesi_1k için 10-katlı çapraz-doğrulama tekniği kullanılarak %70 eğitim - %30 test şeklinde ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmalarının uygulanması sonucunda elde edilen % olarak F-ölçüm değerleri (Table 2. F-measure values in % obtained as a result of applying different machine learning algorithms according to different feature selections of 1,000 tweets divided into 70% training - 30% test using 10-fold cross-validation technique for dataset_1k)

%70 Eğitim %30 Test	KNN (k=5 için)	SVM	NB	RF	LSTM	Word2Vec+LSTM
Info-Gain	77,12	81,09	82,98	83,24	86,27	85,03
Chi-Square	75,79	84,96	80,88	81,45	85,72	85,01
OneR	81,16	84,53	88,81	83,78	84,95	84,62
Fisher Score	78,86	74,48	75,04	78,91	79,16	79,13
Gain-Ratio	81,09	78,89	82,68	84,27	84,31	84,30

Tablo 2’de genel olarak en yüksek oranın LSTM algoritması ile elde edildiği görülmektedir. En iyi sonuçlar tabloda koyu renk olarak verilmiştir. Karşılaştırmalı sonuçlar Şekil 4’te gösterilmiştir.



Şekil 4. Veri kümesi_1k için 10-katlı çapraz-doğrulama tekniği kullanılarak %70 eğitim - %30 test şeklinde ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmalarının uygulanması sonucunda elde edilen % olarak F-ölçüm değerlerinin şekilsel ifadesi (Figure 4. Figural expression of F-measure values in % obtained as a result of applying different machine learning algorithms according to different feature selections of 1,000 tweets divided into 70% training - 30% test using 10-fold cross-validation technique for dataset_1k)

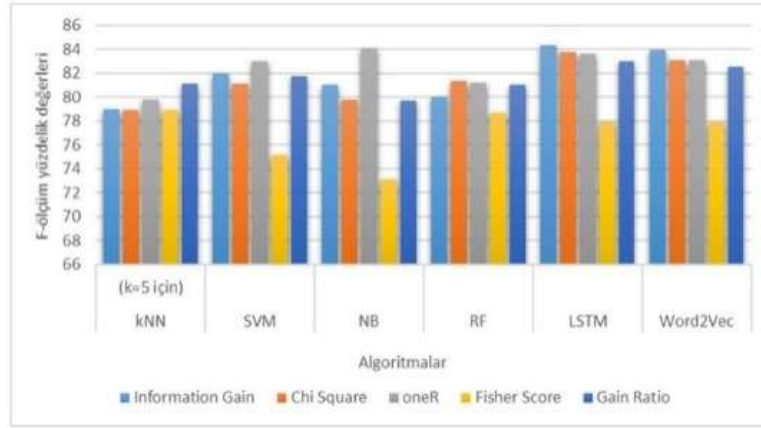
Şekil 4’te veri kümesi_1k için 1.000 tweete göre farklı özellik seçimleri göz önüne alınarak, farklı makine öğrenmesi algoritmalarının karşılaştırmalı sonuçları yer almaktadır. Tablo 2 ve Şekil 4 incelendiğinde, OneR özellik seçim yönteminde NB algoritması %88,81 başarı oranıyla en iyi yöntem olduğu söylenebilir.

Tablo 3’te veri kümesi_1k için %30 eğitim - %70 test olarak ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmaları kullanılarak elde edilen sınıflandırma performansları yer almaktadır.

Tablo 3. Veri kümesi_1k için 10-katlı çapraz-doğrulama tekniği kullanılarak %30 eğitim - %70 test şeklinde ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmalarının uygulanması sonucunda elde edilen % olarak F-ölçüm değerleri (Table 3. F-measure values in % obtained as a result of applying different machine learning algorithms according to different feature selections of 1,000 tweets divided into 30% training - 70% testing using the 10-fold cross-validation technique for dataset_1k)

%30 Eğitim %70 Test	KNN (k=5 için)	SVM	NB	RF	LSTM	Word2Vec+LSTM
Info-Gain	78,97	81,95	81,01	80,05	84,36	83,94
Chi-Square	78,92	81,11	79,82	81,38	83,79	83,06
OneR	79,78	82,98	84,09	81,18	83,65	83,05
Fisher Score	78,91	75,18	73,11	78,72	77,91	77,93
Gain-Ratio	81,11	81,78	79,68	81,02	83,04	82,56

Tablo 3'te genel olarak en yüksek oranın LSTM algoritması ile elde edildiği görülmektedir. En iyi sonuçlar tabloda koyu renk olarak verilmiştir. Karşılaştırmalı sonuçlar Şekil 5'te gösterilmiştir.



Şekil 5. Veri kümesi_1k için 10-katlı çapraz-doğrulama tekniği kullanılarak %30 eğitim - %70 test şeklinde ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmalarının uygulanması sonucunda elde edilen % olarak F-ölçüm değerlerinin şekilsel ifadesi (Figure 5. Figural expression of F-measure values in % obtained as a result of applying different machine learning algorithms according to different feature selections of 1,000 tweets divided into 30% training - 70% testing using the 10-fold cross-validation technique for dataset_1k)

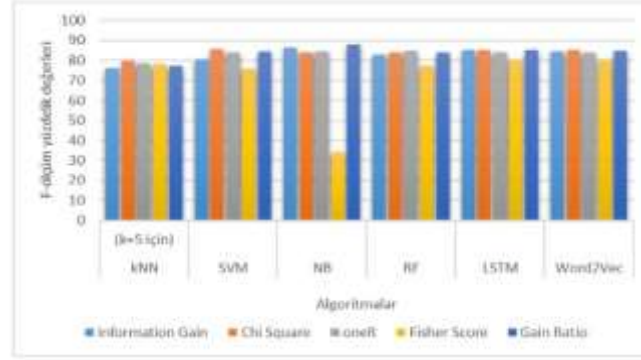
Şekil 5'te veri kümesi_1k için 1.000 tweete göre farklı özellik seçimleri göz önüne alınarak, farklı makine öğrenmesi algoritmalarının karşılaştırmalı sonuçları yer almaktadır. Tablo 3 ve Şekil 5 incelendiğinde, Info-Gain özellik seçim yönteminde LSTM algoritması %84,36 başarı oranıyla en iyi yöntem olduğu söylenebilir.

Tablo 4'te veri kümesi_2k için %70 eğitim - %30 test olarak ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmaları kullanılarak elde edilen sınıflandırma performansları yer almaktadır.

Tablo 4. Veri kümesi_2k için 10-katlı çapraz-doğrulama tekniği kullanılarak %70 eğitim - %30 test şeklinde ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmalarının uygulanması sonucunda elde edilen % olarak F-ölçüm değerleri (Table 4. F-measure values in % obtained as a result of applying different machine learning algorithms according to different feature selections of 1,000 tweets divided into 70% training - 30% testing using the 10-fold cross-validation technique for dataset_2k)

%70 Eğitim %30 Test	KNN (k=5 için)	SVM	NB	RF	LSTM	Word2Vec+LSTM
Info-Gain	76,17	80,29	86,34	82,86	85,17	84,33
Chi-Square	79,51	85,35	84,01	84	85,11	84,95
OneR	78,61	83,81	84,52	84,56	84,05	83,94
Fisher Score	78,22	75,72	33,91	77,42	80,16	79,94
Gain-Ratio	77,31	84,29	87,71	84,09	85,09	84,56

Tablo 4'te genel olarak en yüksek oranının NB algoritması ile elde edildiği görülmektedir. En iyi sonuçlar tabloda koyu renk olarak verilmiştir. Karşılaştırmalı sonuçlar Şekil 6'da gösterilmiştir.



Şekil 6. Veri kümesi_2k için 10-katlı çapraz-doğrulama tekniği kullanılarak %70 eğitim - %30 test şeklinde ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmalarının uygulanması sonucunda elde edilen % olarak F-ölçüm değerlerinin şekilsel ifadesi (Figure 6. Figural expression of F-measure values in % obtained as a result of applying different machine learning algorithms according to different feature selections of 1,000 tweets divided into 70% training - 30% testing using the 10-fold cross-validation technique for dataset_2k)

Şekil 6'da veri kümesi_2k için 1.000 tweete göre farklı özellik seçimleri göz önüne alınarak, farklı makine öğrenmesi algoritmalarının karşılaştırmalı sonuçları yer almaktadır. Tablo 4 ve Şekil 6 incelendiğinde, Gain-Ratio özellik seçim yönteminde NB algoritması %87,71 başarı oranıyla en iyi yöntem olduğu söylenebilir.

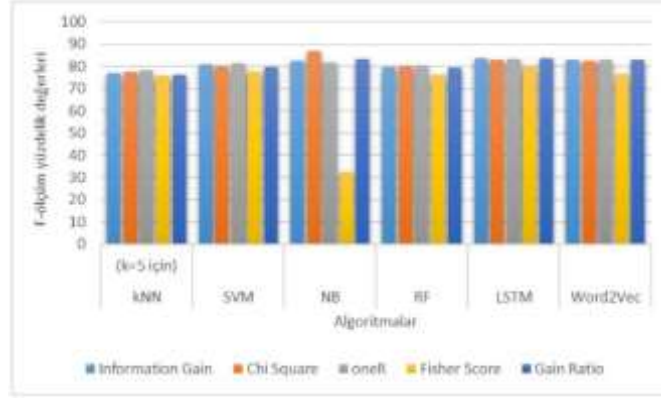
Tablo 5'te veri kümesi_2k için %30 eğitim - %70 test olarak ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmaları kullanılarak elde edilen sınıflandırma performansları yer almaktadır.

Tablo 5. Veri kümesi_2k için 10-katlı çapraz-doğrulama tekniği kullanılarak %30 eğitim - %70 test şeklinde ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmalarının uygulanması sonucunda elde edilen % olarak F-ölçüm değerleri (Table 5. F-measure values in % obtained as a result of applying different machine learning algorithms according to different feature selections of 1,000 tweets divided into 30% training - 70% testing using the 10-fold cross-validation technique for dataset_2k)

%30 Eğitim %70 Test	KNN (k=5 için)	SVM	NB	RF	LSTM	Word2Vec+LSTM
Info-Gain	76,91	80,82	82,51	79,63	83,45	82,97
Chi-Square	77,31	80,11	86,89	80	82,89	82,38

OneR	77,91	81,27	81,55	80,23	83,05	82,93
Fisher Score	75,62	77,67	32,44	76,12	79,75	76,66
Gain-Ratio	76,11	79,54	83	79,26	83,55	82,78

Tablo 5'te genel olarak en yüksek oranının LSTM algoritması ile elde edildiği görülmektedir. En iyi sonuçlar tabloda koyu renk olarak verilmiştir. Karşılaştırmalı sonuçlar Şekil 7'de gösterilmiştir.



Şekil 7. Veri kümesi_2k için 10-katlı çapraz-doğrulama tekniği kullanılarak %30 eğitim - %70 test şeklinde ayrılan 1.000 tweetin farklı özellik seçimlerine göre farklı makine öğrenmesi algoritmalarının uygulanması sonucunda elde edilen % olarak F-ölçüm değerlerinin şekilsel ifadesi (Figure 7. Figural expression of F-measure values in % obtained as a result of applying different machine learning algorithms according to different feature selections of 1,000 tweets divided into 30% training - 70% testing using the 10-fold cross-validation technique for dataset_2k)

Şekil 7'de, veri kümesi_2k için 1.000 tweete göre farklı özellik seçimleri göz önüne alınarak, farklı makine öğrenmesi algoritmalarının karşılaştırmalı sonuçları yer almaktadır. Tablo 5 ve Şekil 7 incelendiğinde, Chi-Square özellik seçim yönteminde NB algoritması %86,89 başarı oranıyla en iyi yöntem olduğu söylenebilir.

5. Sonuç

Teknolojinin sürekli gelişimi ile sosyal medya ağlarının sayısı ve kullanımı her geçen gün artmaktadır. Buna bağlı olarak farklı sosyal medya ağlarında kullanıcılar tarafından üretilen içeriklerdeki nefret söylemi ifadelerinin geniş çaplı olması, bu ifadelerin tespitini zorlaştırmakta ve nefret söylemi tespiti için otomatik sistemlere ihtiyaç oluşturmaktadır. Bu alanda genellikle İngilizce ve diğer dillerde yapılan çalışmaların varlığına rastlanılmıştır. Türkçe dilinde nefret söylemi ifadelerinin tespitindeki çalışmaların sayıca az oluşu da dikkat çekmektedir.

Bu çalışmada, Türkçe dilindeki tweetler üzerinde nefret söylemi tespiti için makine öğrenmesi algoritmaları kullanılmıştır. Çalışmanın ilerleyişi, Türkçe nefret söylemi problemi için özellik seçimi ve sınıflandırıcıların kullanılarak algoritmaların performans karşılaştırması üzerine yapılmıştır. Çalışmada, iki adet Türkçe tweet veri kümesi kullanılmıştır. Veri ön işlemenin ardından kelime yapılarının incelenmesinde açık kaynak kodlu Türkçe doğal dil işleme kütüphanesi olan Zemberek kütüphanesinden yararlanılmıştır. Kelime unigram ve bigramları bulunarak, bulunan bu kelimeler özellik seçimi kümesi olarak belirlenmiştir.

Çalışma kapsamında veri kümelerinde yer alan kelimelerin özellik seçimi için InfoGainAttributeEval, Fisher Skor, OneRClassifier, Ki-kare Testi ve GainRatioAttributeEval özellik seçim yöntemleri uygulanmıştır. Her iki veri kümesi için de özellik seçimi ardından elde edilen 1000'er özellik üzerinde farklı makine öğrenmesi algoritmaları uygulanmıştır. Burada uygulanan altı farklı makine öğrenmesi algoritmaları KNN, RF, NB, SVM, LSTM ve Word2Vec+SVM şeklindedir. Veri kümelerinin etiket dağılımı dengeli olmadığı için performans başarı ölçütü olarak F-ölçüm değeri esas alınmıştır.

Tüm deneyler boyunca 10-katlı çapraz-doğrulama tekniği kullanılmıştır. Bu teknikte veri kümesi varsayılan değer (default) olarak %90 eğitim - %10 test şeklinde ayrılmaktadır. Çalışma kapsamında gerçekleştirilen deneylerde ise, veri kümesi_1k ve veri kümesi_2k için önce %70 eğitim - %30 test şeklinde ayrılarak, kullanılan beş özellik seçimine göre her bir algoritmanın performansı incelenmiştir. Ardından bu işlemler, veri kümesi_1k ve veri kümesi_2k veri kümeleri için %30 eğitim - %70 test şeklinde ayrılarak gerçekleştirilmiş ve sonuçlar değerlendirilmiştir. Veri kümesi_1k için algoritmalar arasında en iyi sonucu %88,81 F-ölçüm oranıyla OneR özellik seçimi ile NB algoritması vermiştir. Veri kümesi_2k için ise algoritmalar arasında en iyi sonucu %87,71 F-ölçüm oranıyla Info-Gain özellik seçimi ile NB algoritması vermiştir. Deneysel uygulamalarda, genel anlamda beş farklı özellik seçimi için algoritmalar arasında en iyi sonucu veren yöntemin LSTM algoritması olduğu görülmüştür.

Gelecekte, Türkçe dilinde nefret söylemi tespitinde yapılan araştırmalarda daha sağlıklı ve güvenilir sonuçların elde edilmesi için, kelime veya kelime gruplarında anlamsal boyutları da dikkate alınarak, değerlendirilebileceği ve derin öğrenme algoritmalarına uygulamalarda daha fazla yer vererek nefret söylemi tespitine yönelik çalışmaların yapılması planlanmaktadır.

6. Teşekkür

Hazırladığımız çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 120E187 numaralı proje ile kısmen desteklenmiştir. Bu çalışmadaki görüşler yazarlara aittir ve TÜBİTAK'ın resmi pozisyonunu veya politikalarını temsil etmeyebilir.

Kaynakça

- Akın, A. A. (2007). Zemberek, An Open-Source NLP Framework For Turkish Languages. *Structure*, 10(2007), 1-5. https://scholar.google.com/citations?view_op=view_citation&hl=tr&user=zCdB2VkAAAAJ&citation_for_view=zCdB2VkAAAAJ:d1gkVwhDpl0C
- Beken, M. (2022). Prediction Of Oil Consumption And Oil Access Of Countries In The European Union Region With Machine Learning. *International Journal of Smart Grid-IJSmartGrid* 6(3), 79-83. <https://doi.org/https://doi.org/10.20508/ijsmartgrid.v6i3.250.g242>
- Beyhan, F. (2022). A Turkish Hate Speech Dataset And Detection System. In İ. Arın, Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022) Marseille, France.
- Bingöl, H. Y., M. (2021, 25-28 Kasım 2021). *Çevrimiçi Sosyal Ağlarda Yapay Zekâ Yöntemleri İle Siber Zorbalık Tespiti Uluslararası Mühendislik, Doğa ve Sosyal Bilimler Sempozyumu*, Batman, Türkiye.
- Budak, H. (2018). Özellik Seçim Yöntemleri Ve Yeni Bir Yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 22(Özel Sayı), 21-31. <https://doi.org/10.19113/sdufbed.01653>
- Diñçer, E. Ş. (2022). Metin Madenciliği Ve Duygu Analizi İle Siber Zorbalık Tespiti. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi* 3(2), 38-45. <https://doi.org/https://doi.org/10.53608/estudambilisim.1070884>
- Eid, M. (2022). Meta-Heuristic optimization of LSTM-based deep network for Boosting the prediction of monkeypox cases. *Mathematics* 10(20), 3845. <https://doi.org/https://doi.org/10.3390/math10203845>
- Engindeniz, İ., Özkan, Ş., & Tekin, F. (2018). *Medyada Nefret Söylemi Ve Ayrımcı Söylem 2018 Raporu*. <https://hrantdink.org/tr/asulus/faaliyetler/projeler/medyada-nefret-soylemi/2002-medyada-nefret-soylemi-ve-ayrimci-soylem-2018-raporu-yayimlandi>
- Gu, Q. (2012). Generalized Fisher Score For Feature Selection. 1-3. <https://doi.org/https://doi.org/10.48550/arXiv.1202.3725>
- Hall, M. A. (1999). *Correlation-based Feature Selection For Machine learning* [The University of Waikato]. Hamilton, New Zealand. <https://hdl.handle.net/10289/15043> ; <https://researchcommons.waikato.ac.nz/handle/10289/15043?show=full>
- Kara, A. (2019). Global Solar Irradiance Time Series Prediction Using Long Short-Term Memory Network. *Gazi Üniversitesi Fen Bilimleri Dergisi, Part C: Tasarım ve Teknoloji* 7(4), 882-892. <https://doi.org/10.29109/gujsc.571831>
- Karayiğit, H. (2021). Detecting Abusive Instagram Comments In Turkish Using Convolutional Neural Network And Machine Learning Methods. *Expert Systems with Applications* 174(14). <https://doi.org/https://doi.org/10.1016/j.eswa.2021.114802>
- Karayiğit, H. (2022). Homophobic And Hate Speech Detection Using Multilingual-BERT Model On Turkish Social Media. *Information Technology and Control* 51(2), 356-375. <https://doi.org/https://doi.org/10.5755/j01.itc.51.2.29988>
- Kuş, O. (2021). Kovid-19 Salgını Ve Mültecilere Yönelik Dijital Nefret Söylemi: Büyük Veri Perspektifinden Metin Madenciliği Tekniği İle Kullanıcı Kaynaklı İçeriklerin Analizi. *TRT Akademi* 6(11), 106-131. <https://doi.org/https://doi.org/10.37679/trta.830736>
- Mayda, İ. (2021). Türkçe Tweetler Üzerinde Makine Öğrenmesi İle Nefret Söylemi. *European Journal of Science and Technology* 328-334. <https://doi.org/https://doi.org/10.31590/ejosat.903854>
- Nergiz, G. (2021). Türkçe Sosyal Medya Yorumlarındaki Siber Zorbalığın Derin Öğrenme İle Tespiti. *European Journal of Science and Technology* 31(1), 77-84. <https://doi.org/https://doi.org/10.31590/ejosat.987259>
- Novaković, J. (2011). Toward Optimal Feature Selection Using Ranking Methods And Classification Algorithms. *Yugoslav Journal of Operations Research* 21(1), 119-135. <https://doi.org/10.2298/YJOR1101119N>
- Ön, E. P. (2020, 05-07 October 2020). *Cyberbullying Detection Using Deep Learning And Word Embedding Analysis* 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey. <https://ieeexplore.ieee.org/document/9302297>
- Özçift, A. (2019). Application of Grid Search Parameter Optimized Bayesian Logistic Regression Algorithm To Detect Cyberbullying In Turkish Microblog Data. *Academic Platform Journal of Engineering and Science* 7(3), 355-361. <https://doi.org/10.21541/apjes.496018>
- Şahiner Yılmaz, Ş. (2021, 22-25 February 2021). *Türkçe Metinlerde Derin Öğrenme Yöntemleri Kullanılarak Duygu Analizi* The International Symposium of Scientific Research and Innovative Studies (ISSRIS'21), Bandırma, Balıkesir, Türkiye. <https://avesis.gazi.edu.tr/yayin/658f2ce8-a656-480c-91af-96057798fbf9/turkce-metinlerde-derin-ogrenme-yontemleri-kullanilarak-duygu-analizi>
- Word2Vec. (5 September 2023). Wikipedia. Retrieved 24 December 2022 from <https://en.wikipedia.org/wiki/Word2vec>
- Yazgılı, E. (2021). Türkçe Metinlerde Makine Öğrenmesi Yöntemleri İle Siber Zorbalık Tespiti. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi* 12(2), 443-453. <https://doi.org/https://doi.org/10.17714/gumusfenbil.935448>