Research Article / Araştırma Makalesi

Abant Sosyal Bilimler Dergisi

Journal of Abant Social Sciences

2023, 23(2): 1017-1027, doi: 10.11616/asbi.1266179

Twitter'da Makine Öğrenmesi Yöntemleriyle Sahte Haber Tespiti

Fake News Detection on Twitter with Machine Learning Methods

Mehmet KAYAKUŞ[1], Fatma YİĞİT AÇIKGÖZ[2]

Geliş Tarihi (*Received*): 16.03.2023 **Kabul Tarihi (*Accepted*):** 02.05.2023 **Yayın Tarihi (*Published*):** 31.07.2023

Öz: Twitter, hem gündemi takip etmek isteyen kullanıcılar hem de haberini hızla hedef kitleye ulaştırmak isteyen haber kaynakları tarafından yoğun olarak tercih edilmektedir. Haberin insanlar arasında hızla yayılması ve etkileşim sağlamasına olanak sunan bu platformun avantajları yanında bazı dezavantajları da bulunmaktadır. Haberin kontrol edilememesi nedeniyle sahte haberlerin dolaşıma sokulması ve bunların engellenme güçlüğü bunlardan bazılarıdır. Bu çalışmada Twitter'da sahte haberleri tespit etmek için makine öğrenmesi yöntemleri kullanılmıştır. Örnek bir konu seçilmiş ve bununla ilgili yapılmış sahte ve gerçek haberler tespit edilmiştir. Çalışmada karar ağaçları ve Naive Bayes yöntemleri kullanılmıştır. Çalışmanın sonuçları karışıklık matrisi ve F1 skoru yöntemine göre karşılaştırılmıştır. Karar ağaçları yönteminin F1 skoru 0,829, Naive Bayes yönteminin ise 0,883 olmuştur. Bu sonuçlara göre Naive Bayes yönteminin Twitter'da sahte haber tespiti için daha başarılı bir yöntem olduğu görülmüştür.

Anahtar Kelimeler: Sahte Haber, Twitter, Makine Öğrenmesi, Metin Madenciliği

&

Abstract: Twitter is intensively preferred by both users who want to follow the agenda and news sources who want to quickly deliver their news to the target audience. In addition to the advantages of this platform, which allows the news to spread and interact rapidly among people, it also has some disadvantages. Some of these are the circulation of fake news and the difficulty of preventing them due to the inability to control the news. In this study, machine learning methods are used to detect fake news on Twitter. A sample topic was selected and fake and real news about it were detected. Decision trees and Naive Bayes methods were used in the study. The results of the study were compared according to the confusion matrix and F1 score method. The F1 score of the decision tree method was 0.829 and the Naive Bayes method was 0.883. According to these results, Naive Bayes method was found to be a more successful method for detecting fake news on Twitter.

Keywords: Fake News, Twitter, Machine Learning, Text Mining

Atıf/Cite as: Kayakuş, M., Yiğit Açıkgöz, F. (2023). Twitter'da Makine Öğrenmesi Yöntemleriyle Sahte Haber Tespiti. *Abant Sosyal Bilimler Dergisi, 23(2),* 1017-1027. doi: 10.11616/asbi.1266179

İntihal-Plagiarism/Etik-Ethic: Bu makale, en az iki hakem tarafından incelenmiş ve intihal içermediği, araştırma ve yayın etiğine uyulduğu teyit edilmiştir. / This article has been reviewed by at least two referees and it has been confirmed that it is plagiarism-free and complies with research and publication ethics. https://dergipark.org.tr/tr/pub/asbi/policy

[1] Doç. Dr., Mehmet Kayakuş, Akdeniz Üniversitesi, mehmetkayakus@akdeniz.edu.tr. (Sorumlu Yazar)

[2] Öğr. Gör., Fatma Yiğit Açıkgöz, Akdeniz Üniversitesi, fatmayigit@akdeniz.edu.tr.

1. Giriş

Günümüzde birçok insan gündemi ve haberleri takip etmek, eğlenmek, bilgi almak, ürün/hizmetleri tanımak veya sosyalleşmek gibi ihtiyaçlarını gidermek için bir iletişim kanalına gereksinim duymaktadır. 1990'lara kadar televizyon, radyo, gazete ve dergi gibi geleneksel iletişim kanalları ile giderilen bu iletişim ihtiyacı internetin hayatımıza girmesi ile internet üzerinden giderilmeye başlanmıştır. Bilgi ve teknoloji çağı olarak anılan günümüzün en önemli yeniliklerinden olan internet; bilgi paylaşımından gazeteciliğe, tanıtım ve reklamdan kamu hizmetlerine, bankacılık ve ticaretten eğlenceye, sosyal ilişkilerden sağlık ve eğitime birçok alanda hayatımıza olumlu yenilikler getirmiştir. İnternetin hayatımıza sunduğu yenilikler ana hatları ile; güncel bilgi ve haber sağlama, insanların içerik paylaşımına imkân sunma, zamandan ve mekândan bağımsız olma, görsel öğelerle iletişime farklı bir boyut kazandırma şeklinde sıralanabilir. Sürekli değişen ve gelişen yapısıyla gündelik yaşamın vazgeçilmez bir parçası haline gelen internet teknolojilerinin bireylerin hayatına sunduğu iletişim araçlarının en önemlileri arasında da sosyal medya mecraları gelmektedir.

Sosyal medya terimi, bireylerin çeşitli içerikler hakkında bilgi paylaşarak birbirleriyle iletişim kurmasını sağlayan çevrimiçi araçları ve web sitelerini içermektedir (Eren & Vardarlıer, 2013). Sosyal medya erişim kolaylığı, düşük maliyetli olması ve hızlı bilgi akışı sunması nedeniyle gencinden yaşlısına her kesim tarafından yoğun olarak kullanılır duruma gelmiştir.

Birçok işlevinin yanı sıra sosyal medya platformlarının haber medyası işlevi görmesi ve bireyin şahit olduğu olayları kitlelerle paylaşmasına imkân sunması en dikkat çekici özelliklerindendir. Sosyal medyanın bu denli yaygın olarak kullanılmasının altında yatan etkenler arasında sıradan insanları bile kitlesel yayıncı haline getirebilmesi olduğuna inanılmaktadır (Segado-Boj et al., 2019). Sosyal medya aracılığıyla bireyler şahit oldukları olayları sahip oldukları sosyal medya hesaplarından paylaşabildikleri gibi kaydettikleri içerikleri ihbar hatları ile basın kuruluşlarına ulaştırarak yurttaş haberciliği de yapmaktadır. Böylece sıradan birey haber yapım sürecinin bir parçasına dönüşmektedir (Ünal, 2019).

Facebook, Instagram, Tik Tok, Youtube, Snapchat, Pinterest, Twitter başta olmak üzere kullanıcıların iletişim kurması adına geliştirilmiş onlarca sosyal medya mecrası bulunmaktadır. Bu sosyal medyaların; kişileştirme, bildirimler, beğenme ve yorum bölümleri, bilgi paylaşımı sunması gibi birçok ortak noktası bulunmaktadır. Bununla birlikte her sosyal medyanın içerik paylaşımına imkân sunan alt yapısal özellikleri ve tercih eden hedef kitlesi bakımından belirgin farklılıkları da bulunmaktadır. Dünyanın dört bir yanından insanların ortak bir konu üzerine fikir bildirerek "Trending Topic"leri belirlediği, "Twitter Fenomeni" kavramı ile sıradan insanların ünlü olduğu, Retweetler ve Favlar vasıtasıyla bilginin kullanıcıdan kullanıcıya aktarıldığı, hızlı haber paylaşımı gibi özellikleriyle Twitter, hem gündemi takip etmek isteyen kişiler hem de hızlı ve etkin şekilde hedef kitlesine ulaşmak isteyen kişi ve kurumlarca yoğun talep görmektedir. Günümüzde; bireyler, haber kanalları, büyük şirketler, aktivistler, politikacılar, ünlü isimler başta olmak üzere çoğu kişi ve kuruluş kendileriyle ilgili gelişmeleri Twitter hesapları üzerinden paylaşmaktadır. Bununla birlikte televizyon kanalları, gazeteler ve dergiler gibi geleneksel kitle iletişim araçları da bu ağlar üzerinden içerik paylaşarak bu ağa dahil olmuş haldedir (Çakır, 2018).

Birçok kişi tarafından yoğun olarak kullanılan sosyal medyanın avantajları yanında dezavantajları da bulunmaktadır. Sosyal medyanın avantajları arasında sayılan sıradan insanların da içerik üreterek haberci gibi yayın yapabilmesi art niyetli insanlar söz konusu olduğunda dezavantaja dönüşebilmektedir. Bazı kişi ve kurumlarca çeşitli sebeplerle dolaşıma sokulan yalan haber ya da kasıtlı şekilde yanlış içerikli haberler özellikle Twitter söz konusu olduğunda en önemli sorunlar arasında yer almaktadır. Kötü niyetli kişi ya da kuruluşlarca yapılan yalan haberler, insanların eğitim, sağlık, ticaret, borsa, adalet, güvenlik gibi günlük yaşamsal faaliyetlerini etkilemektedir. Sahte haberler, insanların seçimlerini ve bu konudaki kararlarını manipüle etmek için dolaşıma sokulmaktadır (Zhang & Ghorbani, 2020; Zhou & Zafarani, 2020).

Bu noktada hem habere konu olan kişi ya kurumların korunması hem de okuyucunun doğru ve güvenilir bilgi alması adına sahte haberlerin tespit edilmesi kritik bir husustur. Sahte haberlerin tespiti okuyucunun doğru haberlere ulaşmasını sağladığı gibi güvenilir bir haber ağının oluşturulmasına da katkı sağlar (Shu et al., 2019).

Makalenin ikinci bölümünde makale konusunda literatür çalışmasına yer verilmiştir. Makalenin üçüncü bölümü olan materyal ve yöntem bölümünde veri setinin oluşturulma aşamalarından, metin madenciliğinden ve sınıflandırma algoritmalarından ve performans ölçüm yöntemlerinden bahsedilmiştir. Makalenin dördüncü bölümü olan bulgular ve tartışma bölümünde makalenin analizlerinden ve istatistiksel sonuçlarından bahsedilmiştir. Son bölüm olan sonuçlar bölümünde de makalenin sonuçlarının değerlendirmesi yapılmış olup çalışmanın literatüre katkılarından bahsedilmiştir.

2. Literatür

Literatür incelendiğinde sosyal medya üzerinden sahte haberlerin tespiti için yapılmış araştırmalar olduğu görülmektedir. Özbay ve Alataş çalışmasında sahte haber içeriklerini belirlemek adına iki aşamadan oluşan bir model önermiştir. İlk olarak, bir dizi ön işlem uygulanarak, sahte haberlere ilişkin ham veriler yapılandırılmış verilere dönüştürülmüştür. Sonrasında ise, on denetimli yapay zekâ algoritması yapılandırılmış sahte haber veri setine uygulanmıştır. Hassasiyet, doğruluk ve duyarlılık kriterlerine göre Rastgele Orman algoritması ISOT veri kümesi üzerinde en iyi performansı göstermiştir (Özbay & Alataş, 2020).

Aydın ve arkadaşları yaptıkları çalışmada insanları yanıltabilecek sahte hesapları tespit etmek adına makine öğrenimi tabanlı yöntemler kullanmışlardır. Bu amaçla kullanılan veri setine ön işlem uygulanmış ve makine öğrenmesi işlemleyiciyle sahte hesaplar belirlenmiştir. Sahte hesapları tespit etmek için karar ağaçları, lojistik regresyon ve destek vektör makinesi algoritmaları kullanılmıştır. Bu yöntemlerin sınıflandırma performansları karşılaştırılmış ve regresyonun daha iyi sonuçlar verdiği gösterilmiştir (Aydin et al., 2018).

Erdi ve arkadaşları Twitter üzerinde trol davranışları sergileyen kullanıcı hesaplarını tespit etmek için makine öğrenmesi yöntemlerini kullanmışlardır. Destek vektör makineleri, Logistic Regression ve Random Forest Regression ile Twitter üzerinden topladığımız veriler ile trol kullanıcıların mesajları üzerinden çıkarılan özellikler ile kapsamlı deneyler gerçekleştirmişlerdir. Elde ettiğimiz sonuçlarda %93,93'lere varan oranlarda trol hesaplarını tespit etmeyi ve engellemeyi başarmışlardır (Bengisu et al., 2021).

Amanzholova ve arkadaşları Twitter üzerinden bot hesapların belirlenmesi için veri madenciliği yöntemlerinin doğruluğu en yüksek yöntem olduğunu belirtmektedir. Makalede, Twitter bot tespitinin özellikleri sunulmuştur. Araştırmacılar çalışmalarında lojistik regresyon, karar ağaçları, Random forest sınıflandırma, Naive Bayes, ve k-Means kümeleme algoritmalarını kullanarak Twitter'da bot hesap tespiti yapmaktadır. Hesap ve tweet üzerinden sınıflandırma doğruluğun yükseltmek için sınıflandırma algoritmaları ile SMOTE ve Resample teknikleri kullanmaktadır (Amanzholova et al., 2019).

Toğaçar ve arkadaşları haberlerin gerçek veya sahte olduğunu tespit eden sınıflandırma analizi gerçekleştirmiştir. Veri seti, 6335 haber başlığı ve içerikten meydana gelmektedir. Veri setindeki haberlerden 3171'i doğru haber niteliği taşırken; 3164'ü sahte haber niteliği taşımaktadır. Çalışmanın analizinde Doğal Dil İşleme yöntemi ile Uzun Kısa Süreli Bellek (UKSB) modeli kullanılmışlardır. Çalışmalarının eğitim verilerinden elde edilen genel doğruluk başarısı %99,83 ve test verilerinden elde edilen genel doğruluk başarısı %91,48'dir (Toğaçar et al., 2022).

Hamdi ve arkadaşları, kullanıcı özelliklerine ve grafiklerin entegrasyonuna dayalı olarak Twitter'daki sahte haberleri tespit etmek için hibrit bir yöntem uygulamıştır. Twitter takipçileri grafiğinden takipçilerin özelliklerini çıkarmak için node2vec kullanılmaktadır. Ayrıca Twitter aracılığıyla sunulan kullanıcı özellikleri de modele dahil edilmiştir. Bu hibrit yaklaşım hem kullanıcının özelliklerini hem de sosyal grafiğini göz önünde bulundurur (Hamdi et al., 2020).

Zervepulos ve arkadaşları çalışmalarında 2019-2020 Hong Kong protestolarıyla ilgili Twitter'da dolaşıma sokulan gerçek dışı haberlerin tespiti için derin öğrenme yöntemini kullanmışlardır. Çalışmaları, metin kullanan diğer algoritmalardan daha iyi performans göstererek, %99,3 F1 puanı kadar yüksek puanlara ulaşmıştır (Zervopoulos et al., 2020).

Helmstetter ve arkadaşları çalışmalarında sahte haber tespiti için denetimli öğrenme işlemleyici kullanmışlardır. Yalnızca bir tweet'i dikkate aldıklarında 0,77 F1 puanına ve kullanıcı hesabıyla ilgili bilgiler de dahil edildiğinde 0,9'a kadar ulaştığını görülmüştür (Helmstetter & Paulheim, 2018).

3. Materyal ve Yöntem

3.1. Veri Seti

Çalışmada belirlenen konuda Twitter sosyal medya platformunda yayılan sahte ve gerçek haberleri içeren tweet mesajlarının tespit edilmesi amaçlanmaktadır. Çalışma için Twitter'da trend topic olmuş ve sahte haber tweetleri içeren bir konu belirlenmiştir. Bu konu hakkında atılan tweetler hazırlanan algoritma ile Twitter'dan alınmıştır. Tweet mesajları üzerinde metin ön işlemleri ve analizler yapıldıktan sonra sahte ve doğru haber diye iki kategoride etiketlenmiştir.

Çalışmada kullanılan veri kümesi eğitim ve test olmak üzere bölünmüştür. Eğitim verileri, modelin parametrelerini belirlemek için kullanılmaktadır. Test veri kümesi de oluşturulan modelin performansının test edilmesi ve analiz edilmesinde kullanılmaktadır. Bu çalışmada veriler %70-%30 oranında rastgele atama yöntemine göre ikiye ayrılmıştır.

3.1.1. Twitter'dan Veri Alınması

Twitter geliştiriciler ve akademik çalışma yapan kişiler için faydalı olacak birçok hizmet sunmaktadır. Bunlardan bir tanesi de için "Twitter Search API" hizmetidir. Twitter'ın araştırmacılar için sunmuş olduğu bu hizmette belirlenen konu hakkında son yedi gün içinde atılan tweet mesajları ve bu tweet mesajları hakkında bilgiler alınabilmektedir. Bu API kullanarak veri çekilmesi için açık kaynak kodlu Knime uygulaması kullanılmaktadır. Bu uygulamada tasarlanan algoritma sayesinde belirlenen Hashtag'leri içeren tweetler bulunmakta ve bu veriler Excel dosyasına aktarılarak bilgisayara kaydedilmektedir.

3.1.2. Konu Seçimi

Haberlerin doğruluğunun tespiti için yasal bir kuruluş olmamakla birlikte bunu gerçekleştiren web siteleri ve özel kuruluşlar bulunmaktadır. Türkiye'de teyit.org; uluslararası olarak International Fact-checking Network en çok tercih edilen platformlardı. Bu siteler haberlerin doğruluğu araştırmakta ve teyit etmektedir. Çalışmada konu belirlemek için teyit.org internet sitesinden araştırma yapılmıştır. Bu sitede yer alan sahte haberler içerisinden Twitter'da trend topic olmuş haberler araştırılmıştır. Bunlar içerisinden en fazla tweet içeren konu çalışma konusu olarak belirlenmiştir. "Çin'de 40 banka iflas etti! Tanklar devrede Çin'de gayrimenkul sektöründe başlayan kriz bankalara da sıçradı. Tanklar devreye girdi. Ülkede 40 banka iflas ettiğini duyurdu. Halk parasını çekmek için akın edince, güvenlik güçleri tanklarla bankaları korumaya aldı" haberi konusunda çok sayıda tweet atılmış ve Twitter'da trend topic olmuştur. Çalışmada kullanılmak üzere bu konuda toplam 5500 tweet toplamıştır. Belirlenen konu teyit.org üzerinden doğruluğu kontrol edilmiştir. Haberin sahte olduğu ve hızla yayıldığı görülmüştür.

3.1.3. Verilerin Etiketlenmesi

Toplanan tweetler öncelikli olarak incelenmiştir. Tekrar eden, eksik metin içeren, anlamsız olan gürültülü tweetler veri setinden çıkarılmıştır. Daha sonra manuel olarak tweetler etiketlenerek sınıflandırılmıştır. Tweetler "Sahte" ve "Doğru" olmak üzere iki kategoride etiketlenerek sınıflandırılmıştır. Toplanan 5500 tweet yapılan ön işlemler sonucu 354 indirilmiştir. Bu tweetlerin 2228 tanesi sahte, 126 tanesi doğru olarak etiketlenmiştir. Sahte ve doğru tweetlerin eşit olması için 125 sahte, 125 doğru tweet mesajı olmak üzere toplam 250 tweet seçilmiştir.

3.2. Metin Madenciliği

Metin madenciliği, doğal dil metinlerinden anlamlı bilgiler toplamaya çalışan gelişen yeni bir alandır. Metin Madenciliği, farklı yazılı kaynaklardan otomatik olarak bilgi çıkararak, daha önce bilinmeyen yeni bilgilerin bilgisayar tarafından keşfedilmesini içerir. Belirli amaçlar için yararlı olan bilgileri çıkarmak için

metni analiz etme sürecidir. Bunun için anlamlı kalıpları ve yeni öngörüleri belirlemek için yapılandırılmamış metni yapılandırılmış bir biçime dönüştürme işlemidir (Hearst, 2003).

Yazılı kaynaklar web siteleri, kitaplar, e-postalar, incelemeleri ve makaleler olabilir. Metin madenciliği ve analizi, kuruluşların kurumsal belgelerde, müşteri e-postalarında, çağrı merkezi günlüklerinde, sözlü anket yorumlarında, sosyal ağ analizlerinde, tıbbi kayıtlarda ve diğer metin tabanlı veri kaynaklarında anlamlı sonuçlar bulmalarına yardımcı olur. Giderek artan bir şekilde, şirketlerin pazarlama, satış ve müşteri hizmetleri operasyonlarının bir parçası olarak müşterilere otomatik yanıtlar sağlamak için kullandıkları yapay zekâ sohbet botları metin madenciliği yeteneklerine örnektir.

Metin madenciliği, veri madenciliğine benzer niteliktedir; ancak daha yapılandırılmış veri formları yerine metne odaklanır. Bununla birlikte, metin madenciliği sürecindeki ilk adımlardan biri, verileri bir şekilde düzenlemek ve yapılandırmaktır; böylece hem nitel hem de nicel analize tabi tutulabilir. Bunu yapmak genellikle veri kümelerini ayrıştırmak ve yorumlamak için hesaplamalı dilbilim ilkelerini uygulayan doğal dil işleme (NLP) teknolojisinin kullanılmasını içerir. Ön çalışma, metni kategorilere ayırmayı, kümelemeyi ve etiketlemeyi; veri kümelerini özetlemeyi, taksonomiler oluşturmayı; sözcük sıklıkları ve veri varlıkları arasındaki ilişkiler gibi şeyler hakkında bilgi çıkarmayı içerir. Analitik modeller daha sonra iş stratejilerini ve operasyonel eylemleri yönlendirmeye yardımcı olabilecek bulgular üretmek için çalıştırılır (Stedman, 2020)

3.2.1. Metin Ön İşleme

Ön işleme yöntemi, metin madenciliği tekniği ve uygulamasında ilk adımı oluşturan çok önemli bir aşamadır. Metin ön işleme aşamasında metinsel ifadelerdeki bozulmalar ve gürültüler ortadan kaldırılarak çalışmanın başarısı arttırılmış olur. Metin ön işleme için kullanılan bazı işlemler büyük-küçük harfe dönüştürme, tokenleştirme, durdurma sözcüklerini kaldırma ve kelimeleri köklerine ayrıma işlemidir. Bunlar, metin verileri üzerinde yapabileceğimiz farklı metin ön işleme adımları türleridir. Ancak bunların hepsinin her zaman kullanılmasına gerek yoktur. Kullanım durumumuza göre ön işleme adımlarını dikkatli bir şekilde seçilmesi gerekmektedir. Çünkü seçimler çalışmanın başarısında önemli bir rol oynamaktadır.

Büyük ve küçük harfin kullanıldığı bir belgede ortak bir düzen olmamaktadır. Bu durum yazım hatalarından kaynaklanabilir. Metin önişlemde, büyük/küçük harf dönüştürme işlemi, bir metindeki tüm harfleri değiştirmeyi ve hepsinin aynı olmasını amaçlar. Bir metin belgesi bir dizi cümleden oluşur. Tokenleştirme işlemi belgeyi token adı verilen sözcük parçalarına ayırır. Durdurma sözcüklerini kaldırma işleminde kelimeleri silmenin amacı, önemli olmayan kelimeleri ortadan kaldırmaktır. Bu işlem büyük ve ağır görünen metin uzayın boyutlarını azaltabilir. Noktalama işaretlerin, web programlama kodları, url'ler, emojiler, sık kullanılan ve nadir kullanılan kelimeler gibi metin belgelerinde belgeye anlam vermeyen yaygın kelimeler çıkarılır. Örneğin, "hangi", "ve", "siz", "kadar" sözcükleri bir metin madenciliği uygulamasında anahtar sözcük olarak ölçülmediği için belgeden silinir. Köklerine ayırma işlemi, bir kelimenin biçimini temel kelime biçimine eşleme ve parçalama işlemidir. Bu köklerine ayırma süreci, metin madenciliği aşamasındaki en önemli süreçtir. İyi sonuçlandırma sonuçları, metin madenciliği uygulamasının olup olmadığını etkileyebilir.

Metin ön işleme aşaması için Zemberek Kütüphanesi kullanılmıştır. Zemberek kütüphanesi hatalı yazılan kelimeleri düzeltebilmekte ve kelimeleri köklerine ayırabilmektedir (Safalı, 2020).

3.2.2. Öznitelik Çıkarımı

Metin madenciliğimde öznitelik çıkarımı için farklı teknikler vardır. Bunlardan bir tanesi metin içindeki her bir kelime kökünün tespitidir. Türkçe metinler için Zemberek kütüphanesi kullanılarak kelime kökü tespit edilmektedir. Bir başka öznitelik çıkarımı yöntemi metin içerisinden çıkarılan farklı uzunluklardaki kelime kombinasyonlarından oluşan kelime seviye n-gramdır. n-gram'daki n, kelimenin tekrar sayısını göstermektedir. n metindeki kelimeleri kaçar gruplayarak ayıracağımızı gösterir. Gram ağırlığı göstermektedir. Diğer bir ifade ile aynı kelimeden kaç tane bulunduğunu ifade etmektedir. Unigram,

metinsel bir ifadeyi tek tek yani harf harf ayırmayı ifade eder. Bigram, öznitelik çıkarımında sık kullanılan bir tekniktir. Bigram'da kelimeler ikişer olarak gruplandırılır. Trigram, aynı şekilde üçlü olarak parçalar.

3.2.3. Terim Ağırlıklandırma

Terim ağırlıklandırma, metin indeksleme işlemi sırasında, her bir terimin belgeye olan değerini değerlendirmek amacıyla gerçekleşen bir prosedürdür. Terim ağırlıklandırma, metin içerisindeki kelimelerin önemini temsil eden sayısal değerlerin atanması işlemidir. Bir belge koleksiyonundaki tüm terimler eşit öneme sahip olmadığından tek tek kelimelerin göreceli önemini dikkate alır. Terim ağırlıklandırmada belirli bir terimin belirli bir belgede veya sorguda önemini belirlemesini sağlayan araçlar bulundurur.

TF-IDF (Term Frequency - Inverse Document Frequency), kelimelerin belirli bir belgeyle ne kadar alakalı olduğunu belirlemek için kelimelerin sıklığını kullanan kullanışlı bir algoritmadır. Belgedeki bir sözcük için TF-IDF, iki farklı ölçüm çarpılarak hesaplanır. Terim sıklığında, bir belgede bir kelimenin göründüğü örneklerin ham sayısıdır. Daha sonrasında ise ters doküman sıklığı hesaplanır. Ters doküman sıklığı bir sözcüğün tüm belge kümesinde ne kadar yaygın veya nadir olduğu anlamına gelir. 0'a ne kadar yakınsa, bir kelime o kadar yaygındır. Bu metrik, toplam belge sayısını alarak, kelimeyi içeren belge sayısına bölerek ve logaritmayı hesaplayarak hesaplanabilir. Dolayısıyla, sözcük çok yaygınsa ve birçok belgede görünüyorsa, bu sayı 0'a yaklaşacaktır. Aksi takdirde, 1'e yaklaşacaktır. Bu iki sayıyı çarpmak, belgedeki bir sözcüğün TF-IDF puanıyla sonuçlanır. Puan ne kadar yüksek olursa, söz konusu kelime o belgede o kadar alakalı olur.

3.3. Sınıflandırma

Sınıflandırmanın, bir şeyi veya birini belirli özelliklere göre belirli bir grup veya sistem içinde sınıflandırmaktır. Sınıflandırmanın amacı, verilerdeki her bir durum için hedef sınıfı doğru bir şekilde tahmin etmektir.

3.3.1. Karar Ağaçları

Karar ağacı hem sınıflandırma hem de regresyon görevleri için kullanılan denetimli bir öğrenme algoritmasıdır. Karar analizinde, kararları ve karar vermeyi görsel ve açık bir şekilde temsil etmek için kullanılır. Kök düğüm, dallar, iç düğümler ve yaprak düğümlerden oluşan hiyerarşik bir ağaç yapısına sahiptir.

Bir karar ağacı, kökü üstte olacak şekilde baş aşağı çizilir. Her ağacın, girdilerinin geçtiği bir kök düğümü vardır. Bu kök düğüm ayrıca sonuçların ve gözlemlerin koşullu olarak dayandığı karar düğümleri kümelerine ayrılır. Tek bir düğümü birden çok düğüme bölme işlemine bölme denir. Bir düğüm başka düğümlere bölünmezse, buna yaprak düğümü veya terminal düğümü denir. Bir karar ağacının bir alt bölümüne dal veya alt ağaç denir.

Bölünmenin tam tersi olan başka bir kavram da var. Eğer ortadan kaldırılabilecek karar kuralları varsa, onları ağaçtan keseriz. Bu işlem budama olarak bilinir ve algoritmanın karmaşıklığını en aza indirmek için kullanışlıdır.

Verilen verilerin nasıl bölüneceğine karar vermek için kullanılan çeşitli teknikler vardır. Karar ağaçlarının temel amacı, verileri en iyi şekilde doğru kategorilere ayıracak düğümler arasında en iyi bölünmeleri yapmaktır. Bunu yapmak için doğru karar kurallarını kullanmamız gerekiyor. Kurallar, algoritmanın performansını doğrudan etkileyen şeydir. Karar ağaçlarında en sık kullanılan algoritmalar; kategorik değişkenler için Entropi, Gini; sürekli değişkenler için ise En Küçük Karelere yöntemidir.

Tüm öğeler doğru şekilde farklı sınıflara bölünürse, bölünmenin saf olduğu kabul edilir. Gini safsızlığı, rastgele seçilen bir örneğin belirli bir düğüm tarafından yanlış sınıflandırılma olasılığını ölçmek için kullanılır. Modelin saf bir bölünmeden nasıl farklı olduğu hakkında bir fikir verdiği için "safsızlık" ölçüsü olarak bilinir.

$$\text{Gini} = 1 - \sum_{j} p_j^2 \quad (1)$$

p_j, j sınıfının gerçekleşme olasılığıdır. Gini safsızlık puanının derecesi her zaman 0 ile 1 arasındadır. Burada 0, tüm öğelerin belirli bir sınıfa ait olduğunu (veya bölünmenin saf olduğunu) gösterir. 1 ise öğelerin rastgele dağıldığını gösterir.

Bilgi kazancı, bir öznitelik tarafından kazanılan bilgi miktarını gösterir. Karar ağacında özelliğin ne kadar önemli olduğunu ifade etmektedir. Karar ağacı yapımı, yüksek doğruluk sağlayan doğru bölünmüş düğümü bulmakla ilgili olduğundan, bilgi kazancı, en yüksek bilgi kazancını döndüren en iyi düğümleri bulmakla ilgilidir. Bu, Entropi olarak bilinen bir faktör kullanılarak hesaplanmaktadır. Entropi, bir sistemdeki düzensizlik derecesini tanımlar. Düzensizlik ne kadar fazlaysa entropi de o kadar fazladır.

$$H = -\sum p(x)\log\ p(x) \quad (2)$$

Burada, p (x) belirli bir sınıfa ait grubun yüzdesini ve H ise entropiyi belirtmektedir.

3.3.2. Naive Bayes Sınıflandırıcısı

Naive Bayes algoritması, Bayer teoremine dayanan ve sınıflandırma problemlerinin çözümünde kullanılan denetimli bir öğrenme algoritmasıdır. Naïve Bayes sınıflandırıcısı adını İngiliz matematikçi Thomas Bayes'ten almaktadır. Naif Bayes sınıflandırıcı, hızlı tahminler yapabilen hızlı makine öğrenimi modellerinin oluşturulmasına yardımcı olan basit ve en etkili sınıflandırma algoritmalarından biridir. Olasılıksal bir sınıflandırıcıdır, yani bir nesnenin olasılığına dayanarak tahmin eder. Naif Bayes algoritmasının bazı popüler örnekleri spam filtreleme, duygusal analiz ve makaleleri sınıflandırmadır.

Bayes sınıflandırıcısı, Bayes teoremi tarafından verilen koşullu olasılık ilkesine göre çalışır. Bayes teoremi, önceden bilgi sahibi bir hipotezin olasılığını belirlemek için kullanılan Bayes Kuralı veya Bayes yasası olarak da bilinir. Bayes teoreminin formülü şu şekilde verilmiştir:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3)$$

Burada, P(A|B); B olayı gerçekleştiği durumda A olayının meydana gelme olasılığıdır. P(B|A); A olayı gerçekleştiği durumda B olayının meydana gelme olasılığıdır. P(A) ve P(B); A ve B olaylarının önsel olasılıklarıdır.

Bayes teoremine göre metin sınıflandırması yapılırken, dj belgesinin bir c sınıfına ait olma olasılığı şu şekilde hesaplanmaktadır (Uslu & Özmen Akyol, 2021).

$$p(c|d_j) = \frac{p(d_j|c)p(c)}{p(d_j)} = \frac{p(d_j|c)p(c)}{p(d_j|c)p(c) + p(d_j|\bar{c})p(\bar{c})} \quad (4)$$

$$p(c|d_j) = \frac{\frac{p(d_j|c)}{p(d_j|\bar{c})} \cdot p(c)}{\frac{p(d_j|c)}{p(d_j|\bar{c})} \cdot p(c) + p(c)} \quad (5)$$

3.4. Çalışmanın Değerlendirmesi

Çalışmanın başarısını belirlemek ve değerlendirmek için karışıklık matrisi (confisuon matrix) yöntemi kullanılmıştır. Karışıklık matrisi, bir sınıflandırma algoritmasının performansını tanımlamak için kullanılan bir tablodur. Hata matrisi olarak da bilinen karışıklık matrisi, bir sınıflandırma modelinin bir

dizi test verisi üzerindeki performansını açıklayan bir matrisle gösterilir. Bir karışıklık matrisi, bir sınıflandırma algoritmasının performansını görselleştirir ve özetlemektedir. Karışıklık matrisi, sınıflandırıcının ölçüm metriklerini tanımlamak için kullanılan dört temel özellikten oluşur. Bu dört sayı şunlardır: Gerçek pozitifler (TP), gerçek negatifler (TN), yanlış pozitifler (FP) ve yanlış negatifler (FN). TP, gözlem pozitif olarak tahmin edilir ve aslında pozitiftir. FP, gözlem pozitif olarak tahmin edilir ve aslında negatiftir. TN, gözlem negatif olarak tahmin edilir ve aslında negatiftir. FN, gözlem negatif olarak tahmin edilir ve aslında pozitiftir.

Bir algoritmanın performans metrikleri, yukarıda belirtilen TP, TN, FP ve FN'ye göre hesaplanan doğruluk, hassasiyet, geri çağırma ve F1 puanıdır.

Doğruluk (A_i): Bir algoritmanın doğruluğu, doğru sınıflandırılmış verilerin (TP + TN) toplam veri sayısına (TP + TN + FP + FN) oranı olarak temsil etmektedir.

$$A_i = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Hassasiyet(π_i): Bir algoritmanın hassasiyeti, (TP) doğru sınıflandırılmış verilerin, doğru olduğu tahmin edilen toplam veri sayısına (TP + FP) oranı olarak temsil etmektedir.

$$\pi_i = \frac{TP}{TP + FP} \quad (7)$$

Geri çağırma (p_i): Sadece pozitif değerlerden doğru sınıflandırılanların oranını verir.

$$p_i = \frac{TP}{TP + FN} \quad (8)$$

F skoru (F_1): F1 skoru, F ölçüsü olarak da bilinir. F1 skoru, hassasiyet ve geri çağırma arasındaki dengeyi belirtir.

$$F_1 = 2 \cdot \frac{\pi_i \cdot p_i}{\pi_i + p_i} \quad (9)$$

4. Bulgular ve Tartışma

Çalışmada belirlenen konu hakkında Twitter'dan toplanmış tweet mesajları karar ağaçları ve Naive Bayes sınıflandırıcıları ile sahte ve doğru haber olarak sınıflandırılmıştır. Çalışmada verilere eğitim ve test olmak üzere rastgele ikiye ayrılmıştır. Rastgele ayrılan verilerde şans faktöründen dolayı hatalı sonuçlar olabilmektedir. Bu olasılığı ortadan kaldırmak için modeller 20 kere çalıştırılmış ve çıkan sonuçların ortalaması alınmıştır. Böylece tüm mesajların hem eğitim setinde hem de test setinde bulunabilmesi sağlanmıştır.

Çalışma sonucu oluşan karışıklık matrisi (confisuon matrix) Tablo 1'de gösterilmiştir.

Tablo 1: Confisuon Matrix

	Karar ağaçları	Naive Bayes
True positive	102	109
True negative	106	112
False positive	19	13
False negative	23	16

Tablo 1'deki Çalışma sonucuna göre karar ağaçları 250 tweet mesajının 208 tanesini doğru, 42 tanesini yanlış sınıflandırmıştır. Naive Bayes sınıflandırıcısı ise 221 tweet mesajını doğru, 29 tanesini yanlış sınıflandırmıştır. Çalışmada kullanılan iki modelin performans ölçüm sonuçları Tablo 2'de verilmiştir.

Tablo 2: Performans Ölçüm Sonuçları

Makine öğrenmesi yöntemi	Precision	Recall	F Score (F1)
Karar ağaçları	0,842	0,816	0,829
Naive Bayes sınıflandırıcısı	0,893	0,872	0,883

Precision, tüm sınıflardan doğru olarak ne kadar tahmin edildiğinin bir ölçüsüdür. Sınıflayıcının ne kadar gerçek pozitif değeri doğru tahmin ettiğinin bir ölçüsüdür. Bu iki değerinde yüksek değerlerde olması modelin başarılı olduğunu göstermektedir. F-score, sınıflandırıcının ne kadar iyi performans gösterdiğinin bir ölçüsüdür. Tablo 2'deki sonuçlara göre Naive Bayes sınıflandırıcısının karar ağaçlarından daha başarılı sonuç verdiği görülmüştür.

5. Sonuç

İnternet ve onun en önemli argümanı olan sosyal medya platformları kullanıcıları arasındaki hiyerarşiyi ortadan kaldırması, hız, etkileşimlilik, ulaşım kolaylığı vb. nedenlerle kullanıcı sayılarını her geçen gün arttırmaktadır. Bireylere göre kullanım amacı değişmekle birlikte günümüzde sosyal medya platformlarının haber sunmak ve habere ulaşmak içinde sıklıkla tercih edildiği bilinmektedir.

Haberlerin hızla yayılmasına olanak sunmasıyla hem yayıncı kuruluşlar hem de vatandaşlar tarafından tercih edilen sosyal ağlarda haberin kontrol edilememesi böylece sahte ya da eksik haberlerin de oluşmasına neden olmaktadır. Türkiye'de sahte haber üretimi oldukça yaygın ve sık karşılaşılan bir durumdur. Sahte haberle karşılaşan bireylerin bunu ayırt etmesi çoğunlukla zaman almaktadır. Bu yüzden sahte haberin yayılma hızı oldukça yüksektir. Özellikle ilk haber çıktıktan birkaç saat sonra haber hızla yayılmakta ve insanlar haberin doğru olduğuna inanmaktadır. Sahte haberlerin hem bireysel hem de toplumsal anlamda olumsuz etkileri olabilmektedir. Kurumların ve bireylerin ekonomik olarak zarara uğraması, itibarının zedelenmesi gibi istenmeyen olayların yaşanmasına neden olabilmektedir.

Bu önemden hareketle, çalışmada Twitter'daki sahte haberleri tespit etmek için makine öğrenmesi yöntemlerinden karar ağacı ve Naive Bayes yöntemleri kullanılmıştır. Bunun için belirlenen konu hakkında atılan 125 sahte ve 125 doğru tweet mesajı olmak üzere toplam 250 tweet seçilmiştir. Bu veriler %70-%30 oranında rastgele atama yöntemine göre eğitim ve test verisi olmak üzere ikiye ayrılmıştır. Çalışma sonuçları F1-score göre değerlendirildiğinde Naive Bayes yönteminin karar ağaçları yöntemine göre daha başarılı olduğu görülmektedir.

Farklı haberlerin kullanılması, veri setindeki haber sayısının arttırılması ve farklı makine öğrenmesi yöntemlerinin kullanılması başarı oranını arttırabileceği ön görülmektedir. Bu çalışma ile sahte haberlerin tespitinin kısa sürede ve başarı ile yapılabileceği ortaya konmuştur. Bu da sosyal medya üzerinden dolaşıma sokulan sahte haberin yayılmasının önüne geçmek ve engellemek için önemli avantaj sağlayacaktır. Böylece gerek bireylerin gerekse kurum ve kuruluşların sahte haberler nedeniyle mağdur olmasının önüne geçilecektir. Ayrıca sahte haberlerin tespit edilerek ayıklanması sosyal medya mecralarının amacına uygun olarak kullanılmasının da yolunu açacağı düşünülmektedir.

Finansman/ Grant Support
Yazar(lar) bu çalışma için finansal destek almadığını beyan etmiştir.
The author(s) declared that this study has received no financial support.
Çıkar Çatışması/ Conflict of Interest
Yazar(lar) çıkar çatışması bildirmemiştir.
The authors have no conflict of interest to declare.
Yazarların Katkıları/Authors Contributions
Çalışmanın Tasarlanması: Yazar-1 (%70), Yazar-2 (%30)
Conceiving the Study: Author-1 (%70), Author-2 (%30)
Veri Toplanması: Yazar-1 (%80), Yazar-2 (%20)
Data Collection: Author-1 (%80), Author-2 (%20)
Veri Analizi: Yazar-1 (%80), Yazar-2 (%20)
Data Analysis: A Author-1 (%80), Author-2 (%20)
Makalenin Yazımı: Yazar-1 (%60), Yazar-2 (%40)
Writing Up: Author-1 (%60), Author-2 (%40)
Makale Gönderimi ve Revizyonu: Yazar-1 (%80), Yazar-2 (%20)
Submission and Revision: Author-1 (%80), Author-2 (%20)

Kaynaklar

Amanzholova, A., Doğru, İ. A. ve Coşkun, A. (2019), Twitterda Veri Madenciliği Yöntemlerin Kullanarak Bot Tespiti. *Ejons Internatıonal Journal*, 3(11), s.98-107.

Aydin, I., Mehmet, S. ve Salur, M. U. (2018), *Detection of Fake Twitter Accounts with Machine Learning Algorithms*. 2018 International Conference on Artificial Intelligence and Data Processing (IDAP).

Bengisu, E., Şahin, E. A., Toydemir, M. S. ve Dökeroğlu, T. (2021), Makine Öğrenmesi Algoritmaları ile Trol Hesapların Tespiti. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9(1), s.430-442.

Çakır, H. (2018), Kırgızistan-Türkiye Manas Üniversitesi Öğrencilerinin Sosyal Medya Kullanım Alışkanlıkları. *MANAS Sosyal Araştırmalar Dergisi*, 7(3), s.539-563.

Eren, E. ve Vardarlıer, P. (2013), Social Media's Role in Developing an Employees Sense of Belonging in The Workplace as An Hrm Strategy. *Procedia-Social and Behavioral Sciences*, 99, s. 852-860.

Hamdi, T., Slimi, H., Bounhas, I. ve Slimani, Y. (2020), *A Hybrid Approach for Fake News Detection in Twitter Based on User Features and Graph Embedding*. International conference on distributed computing and internet technology.

Hearst, M. (2003), *What is Text Mining*. *SIMS*, UC Berkeley, 5.

Helmstetter, S. ve Paulheim, H. (2018), Weakly Supervised Learning for Fake News Detection on Twitter. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

Özbay, Feyza ALTUNBEY ve Alataş, B. (2020), Çevrimiçi Sosyal Medyada Sahte Haber Tespiti. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 11(1), s.91-103.

Safalı, Y. (2020), Sosyal Medya Kullanıcılarının Cumhur İttifakı Hakkındaki Görüşlerinin Veri Madenciliği Teknikleri ile Sınıflandırılması. *Bilgisayar Bilimleri ve Teknolojileri Dergisi*, 1(2), s.51-57.

Segado-Boj, F., Díaz-Campo, J. ve Quevedo-Redondo, R. (2019). Influence of the'News finds me'Perception on News Sharing and News Consumption on social media. *Communication Today*, 10(2), s.90-104.

Shu, K., Wang, S. ve Liu, H. (2019), Beyond News Contents: The Role of Social Context for Fake News Detection. *Proceedings of the Twelfth Acm International Conference on Web Search and Data Mining*.

Stedman, C. (2020), *Text Mining (Text Analytics)*. https://www.techtarget.com/searchbusinessanalytics/definition/text-mining, (Erişim Tarihi: 31.08.2022).

Toğaçar, M., Eşidir, K. A. Ve Ergen, B. (2022). Yapay Zekâ Tabanlı Doğal Dil İşleme Yaklaşımını Kullanarak İnternet Ortamında Yayınlanmış Sahte Haberlerin Tespiti. *Journal of Intelligent Systems: Theory and Applications*, 5(1), s.1-8.

Uslu, O. ve Özmen Akyol, S. (2021). Türkçe Haber Metinlerinin Makine Öğrenmesi Yöntemleri Kullanılarak Sınıflandırılması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 2(1), s.15-20.

Ünal, R. (2019), Anaakım Medyada Kullanıcı Türevli İçeriğin İzini Sürmek: Ntv ve Star Tv Whatsapp İhbar Hatları Üzerine Bir İnceleme. *Mersin Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2(2), s.34-43.

Zervopoulos, A., Alvanou, A. G., Bezas, K., Papamichail, A., Maragoudakis, M. ve Kermanidis, K. (2020). *Hong Kong Protests: Using Natural Language Processing for Fake News Detection on Twitter*. IFIP International Conference on Artificial Intelligence Applications and Innovations.

Zhang, X. ve Ghorbani, A. A. (2020), An Overview of Online Fake News: Characterization, Detection, And Discussion. *Information Processing & Management*, 57(2), s.102025.

Zhou, X. ve Zafarani, R. (2020), A Survey of Fake News: Fundamental Theories, Detection Methods, And Opportunities. *ACM Computing Surveys (CSUR)*, 53(5), s.1-40.