E-TİCARET SİTELERİNDE YER ALAN TÜKETİCİ YORUMLARININ METİN MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ

Elif UZUN
İskenderun Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Ekonomi ve Finans Anabilim Dalı
ORCID: 0000-0002-0328-5780

Murat UÇAR
İskenderun Teknik Üniversitesi, İşletme ve Yönetim Bilimleri Fakültesi, Yönetim Bilişim Sistemleri Bölümü
ORCID: 0000-0001-9997-4267

ÖZET
Günümüzde teknolojide yaşanan hızlı gelişmeler sonucunda tüketiciler, online alışveriş ortamını geleneksel alışveriş ortamına kıyasla daha fazla tercih etmektedir. Online alışveriş sürecinde müşterilerin bir ürünü satın alma aşamasında faydalandığı en önemli kaynaklardan biri tüketici yorumlarıdır. Ürünü satın alan müşterilerin paylaştığı yorumlar, başka müşterilerin satın alma kararını ve ürünün markası hakkındaki düşüncelerini etkilemektedir. Ancak ürün yorumlarında ürün ile ilgisi olmayan çok sayıda gereksiz yoruma rastlanılmakta ve bu durum ürün hakkında bilgi veren yorumlara ulaşmayı zorlaştırmaktadır. Öte yandan tüketicilerin paylaştığı yorumlarda kelimelerin konuşma dilinde yazılması, kısaltılması, noktalama işaretlerinin lüzumsuz veya fazla kullanılması gibi gürültü oluşturan durumlar da işlenecek metin verilerinin analizi zorlaşmakta ve bu zorlukların üstesinden gelebilmek için metin madenciliği tekniklerinden yararlanılmaktadır. Bu nedenle bu çalışmada metin madenciliği teknikleri kullanılarak online alışveriş sitelerinden daha önce alışveriş yapan müşterilerin, deneyimleri yönünde paylaştıkları çevrimiçi yorumların ürün ile alakalı olup olmadığının tespit edilmesi amaçlanmıştır. Bu amaçla öncelikle, Türkiye'nin önde gelen online pazar yerlerinden www.trendyol.com sitesinde yer alan farklı kategorilerdeki 21 ürüne ait 4795 adet yorum içeren bir veri seti oluşturulmuştur. Daha sonra bu veri seti üzerinde Terim Frekansları ve Ters Doküman Frekansları yöntemleri kullanılarak kelime vektörleri oluşturulmuştur. Son olarak, k en yakın komşu, Naive Bayes ve Karar ağaçları sınıflandırma algoritmaları kullanılarak yorumlar sınıflandırılmış ve en yüksek başarı %73,99 doğruluk değeri ile Naive Bayes algoritması ile elde edilmiştir. Yapılan bu çalışma tüketicilere yalnızca ürün ile ilgili yorumları filtreleme imkânı sunacağından satın alma aşamasında karar vermelerini kolaylaştıracağı düşünülmektedir. Ayrıca online satış yapan firmalar ürün hakkında paylaşılan müşteri düşüncelerini dikkate alarak zayıf ve eksik yönlerini yeniden gözden geçirme fırsatı bulabileceklerdir.
Anahtar Kelimeler: Metin madenciliği, Online alışveriş, Tüketici yorumları, Makine öğrenmesi

ANALYSING CONSUMER COMMENTS ON E-COMMERCE WEBSITES USING TEXT MINING METHODS

ABSTRACT
Today, as a result of rapid developments in technology, consumers prefer the online shopping environment more than the traditional shopping environment. Consumer comments are one of the most essential resources that customers benefit from when purchasing a product online.

The comments shared by the customers who bought the product affect the purchasing decision of other customers and their thoughts about the brand of the product. However, there are numerous needless remarks unrelated to the product in the product comments, making it difficult to find the comments that provide information on the goods. On the other hand, in the comments shared by consumers, in cases where the words are written in spoken language, shortened, or excessive use of punctuation marks, the analysis of the text data to be processed becomes difficult and text mining techniques are used to overcome these difficulties. Therefore, the purpose of this study is to figure out, using text mining techniques, if online comments posted by customers who have previously made purchases from online shopping sites are relevant to the product. For this purpose, at first, a data set of 4795 comments on 21 products from various categories was constructed on www.trendyol.com, one of Turkey's top online marketplaces. Then, word vectors were created on this data set by using Term Frequency and Inverse Document Frequency methods. Finally, the comments were classified using the k nearest neighbor, Naive Bayes, and Decision trees classification algorithms, and the highest success was obtained with the Naive Bayes algorithm with an accuracy value of 73.99%. It is thought that this study will make it easier for consumers to decide at the purchasing stage, as it will provide the opportunity to filter only the comments about the product. Additionally, online retailers will have the chance to assess their flaws and shortcomings by considering the comments made by customers regarding the merchandise.
Keywords: Text mining, Online shopping, Consumer comments, Machine learning

1. GİRİŞ

İnternete erişimin kolay olması ile birlikte bilgiye ulaşmadaki zorluk büyük oranda azalma göstermiştir. İnternet aracılığıyla yapılan ve farklı bir alışveriş biçimi olan elektronik ticaretin özellikle pandeminin de etkisiyle önemi daha da artmıştır. Ülkemizde son yıllarda internet üzerinden yapılan alışveriş her alanda artış göstermeye devam etmektedir. Ayrıca elektronik ticaret firmalara, tüketici pazarlarına, ulusal ve uluslararası pazarlara da bir değişim ve canlılık getirmiştir. Türkiye İstatistik Kurumu verilerine göre internet kullanıcılarının internet üzerinden kişisel amaçla mal veya hizmet satın alması (e-ticaret) oranı, 2020 yılında %36,5 oranında, 2022 yılında ise %46,2 olarak gerçekleşmiştir (TÜİK, 2022).

Tüketiciler e-ticaret ortamında bir ürün satın alırken geleneksel alışveriş alışkanlıklarının yanı sıra o mal ve hizmeti önceden satın alan kişilerin yorumlarını da dikkate almaktadır (Cop & Gümüş, 2009). Elektronik ortamda ürünü satın alan kişi ürünün marka, model ve fiyatı yanında ürün ve firma hakkında yapılmış değerlendirmeleri göz önünde bulundurarak karar vermektedir. Ancak paylaşılan kullanıcı yorumlarının çoğu ürün ile ilgili olmamaktadır. Paylaşılan yorumlar daha çok “iyi”, “güzel”, “sorunsuz geldi”, “Tşk”, “hızlı kargo”, “kızım için aldım” gibi ürün hakkında bilgi vermeyen gereksiz yorumlardan oluşmaktadır. Bu tür yorumların çok fazla olması tüketicilerin ürünle ilgili bilgi veren ve satın alma kararını etkileyecek yorumlara ulaşmasını zorlaştırmaktadır. Ayrıca yorumlarda noktalama işaretlerinin gereksiz ve aşırı kullanımı, günlük konuşma dilinde ve kısaltılarak yazılması bu verilerin işlenmesini zorlaştırmaktadır. Bu çalışmada tüketiciye zaman kazandırmak ve satın alacağı ürün hakkında satın alma kararını etkileyecek yorumlara ulaşmasında kolaylık sağlanması konusu ele alınmıştır. Online alışveriş sitelerinden daha önce alışveriş yapan müşterilerin, deneyimleri yönünde paylaştıkları çevrimiçi yorumların ürün ile alakalı olup olmadığının metin madenciliği teknikleri kullanılarak tespit edilmesi amaçlanmıştır.

Online alışveriş sitelerinde tüketicilerin ürün ve hizmet hakkında paylaştıkları yorumların analizi için metin madenciliği tekniklerinin kullanıldığı çalışmalar bulunmaktadır.

Ergün 2012 yılında yaptığı araştırmada tüketicinin ve üreticinin faydasına olacak yorumların yararlı bir bilgiye dönüştürülmesi için tüketiciye ve firmalara zaman kazandıran doğal dilde yazılmış yorumları otomatik olarak değerlendirip özet nitelikte bilgi sunan bir sistem geliştirmiştir (Ergün, 2012). Bir başka çalışmada Çakar, hızlı tüketim malları gibi spesifik ürün gruplarına yönelik yorumların ürüne mi yoksa ürünün satışı için sunulan hizmete mi yönelik olduğu sorusuna yanıt aramıştır. Elde edilen bulgular neticesinde; özellikle yüksek yıldız derecelendirmelerinin ürünün satış ve teslimatı için sunulan hizmetlere yönelik olduğunu ve düşük derecelendirmelerin ise doğrudan ürünün kalitesine yönelik olduğunu gözlemlemiştir (Çakar, 2018). Tuna 2019 yılında yaptığı çalışmada, bir çevrimiçi rezervasyon sitesinde yer alan Antalya bölgesindeki otellere ilişkin şikâyet ve yorumları metin madenciliği tekniklerinden duygu analizi ile incelemiş ve okuyuculara metinsel ifadelerin konaklama işletmeleri açısından başarılı bir şekilde çözümlenebileceğini kanıtlanmıştır (Tuna, 2019). Diğer bir çalışmada Işık, duygu analizi sınıflandırma yöntemleri ile e-ticaret markalarına yönelik sosyal medya yorumlarını otomatik değerlendirmiş, sınıflandırma algoritmalarının performansını 16 farklı model oluşturularak incelemiştir. Sonuç olarak, kNN algoritmasının %93,52 sınıflandırma doğruluğu ile en yüksek başarıyı gösterdiğini ancak tüm kriterler göz önüne alındığında SMO algoritmasının en iyi ortalama performansı gösterdiğini gözlemlemiştir (Işık, 2019). Budak 2021 yılında yaptığı çalışmada hava yolu işletmelerinin sosyal medya yorum ve skorlarını veri ve metin madenciliği tekniklerinden Duygu Analizi, DVM, Naive Bayes, Derin Öğrenme ve Karar Ağacı Algoritmaları ile değerlendirmiş ve her bir hava yolu için kelime bulutu analizi yapmıştır (Budak 2021). Akgöz ve Tengilimoğlu'nun ortak çalışmasında 5 yıldızlı Bakü otellerine ilişkin yorumları metin madenciliği teknikleri kullanılarak incelemişler ve odalarla ilgili yaşanan en temel sorunun temizlik olduğu sonucuna varmışlardır. (Akgöz ve Tengilimoğlu, 2022).

2. MATERYAL VE YÖNTEM

2.1. Veri Seti

Online alışveriş sitesi www.trendyol.com adresinden farklı kategorilerdeki 21 ürüne ait tüketici yorumları, Foxpro diliyle hazırlanan yazılım ile toplanarak veri seti oluşturulmuştur. Sonrasında yine Foxpro yazılımı ile tekrar eden yorumlar kaldırılmıştır. Veri seti 5 sütun ve 4795 satırdan oluşmaktadır.

Tablo 1 Veri Seti

Row...	kategori	tur	urunadi	yorum	ilgi
1	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	Çok çok kullanışlı kanal geçişleri çok hızlı bir televizyon	1
2	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	İndirimde aldık güzel	0
3	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	Fiyat performans ürünü indirimde aldığım için şanslıyım.	0
4	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	tavsiye edilir	0
5	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	ürün çok hızlı geldi servis çok iyiydi çok teşekkürler şiddetle tavsiye ederim	0
6	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	servis ve kargo gayet güzeldi tv yi iyiki grundig almışım pahalı markaya para vermek akıl işi değil	0
7	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	çeyiz için aldık.ürün yetkili servis tarafından teslim ediliyor	0
8	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	KALİTELİ BİR ÜRÜN.	1
9	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	Görüntü kalitesi çok iyi boyut olarak da göz dolduruyor ; sadece kanal geçişlerini daha hızlı bekliyordum, bir de goo...	1
10	TV/Görüntü/Ses	Televizyon	Grundig 55 GFU 7850 B 55 inch 139 Ekran	Tvin renk canlılığı vs çok güzel. İndirimden aldık çok memnunuz. Sadece kanallar arası geçiş bir tık yavaş gibi geldi...	1

ExampleSet (4,795 examples, 0 special attributes, 5 regular attributes)

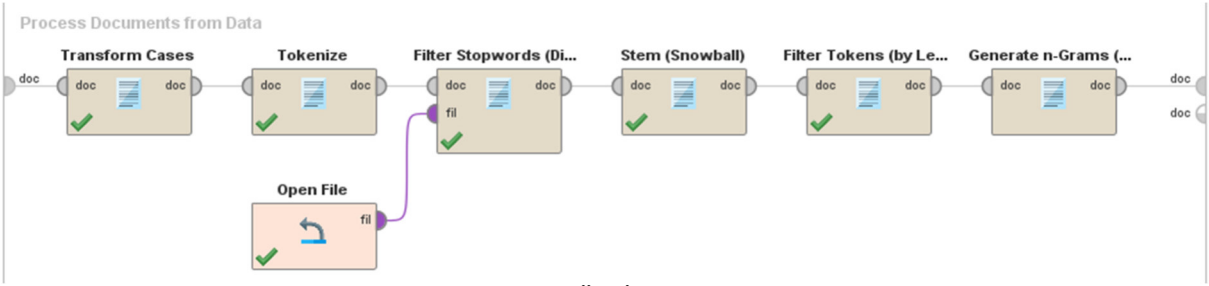
İlk aşamada kullanıcı yorumları ilgili ve ilgisiz olarak etiketlenmiştir. Veri setinde yer alan kayıtların 2683 adedi ilgili, 2112 adedi ilgisiz olarak belirlenmiştir. Veri setinde yer alan kayıtlardan örnekler Tablo 2'de yer almaktadır.

Tablo 2 Veri seti içerisindeki kayıtlardan örnekler

Yorum Sınıfı	Yorum Metni
İlgili Yorumlar	Çok kaliteli ekranı var kanallar arası geçiş biraz yavaş sadece. Android özellikli olması telefon gibi kullanmanızı sağlıyor çok iyi.
	Güzel nemlendiriyor ve yapış yapış bir his bırakmıyor hemen emiliyor. Kokusu çok hoş kesinlikle tavsiye ederimm.
	Son derece hoş ve şık. Yumuşak ve duruşu gayet güzel. Aşırı soğuk günler dışında rahatlıkla giyilebilir. Karlı, donlu günlerde biraz ince kalabilir.
	wi-fi çekim performansı çok kotü ses ekran pil iyi ancak hızlı şarj olmaması kötü ram ve hafıza yeterli performans orta ustu
İlgisiz yorumlar	Henüz kullanmadım ama paketleme güzeldi
	yanlış ürün geldi, iade ile uğraşmak istemedim
	Arkadaşımla doğum günü hediyesi aldım
	Sorunsuz teslimat, kapıya kadar geldi, ertesi gün servis kurulum için geldi.

2.2. Veri Ön İşleme

Uygulamaların gerçekleştirilmesinde RapidMiner Studio programı tercih edilmiş ve kullanıcı yorumları metin madenciliğinin veri ön işleme operatörleri kullanılarak ön işleme sürecine tabi tutulmuştur. Ön işleme sürecinde tüm karakterlerin büyük harfe dönüştürülmesi, noktalama işaretlerinin kaldırılması, anlamsız kelimelerin filtrelenmesi, kelimelerin köklerine ayrılması, kelimelerin en düşük ve en yüksek uzunluk sınırlarının belirlenmesi yöntemleri sırasıyla gerçekleştirilmiştir. Veri ön işlemede sırasıyla Transform Cases, Tokenize, Filter Stopwords (Dictionary), Stem (Snowball), Filter Tokens (by Length), Generate n-Grams (Terms) operatörleri kullanılmıştır (Şekil 1). Metin madenciliği ile ilgili çalışmalarda verilerin ön işleme sürecinde çoğunlukla bu yöntemler kullanılmaktadır (Ertek, 2013; Dirsehan, 2016).



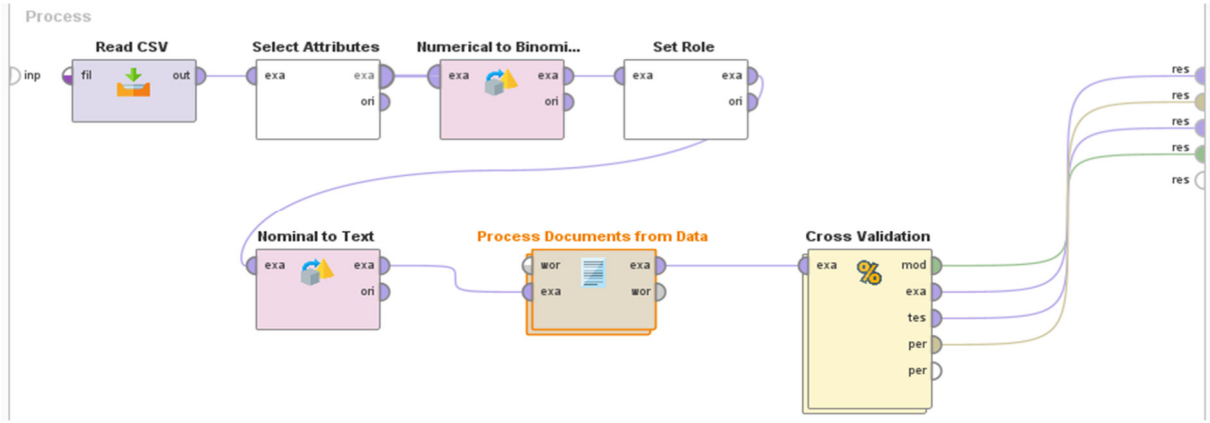
Şekil 1 Veri Ön İşleme Süreci

2.3. Terim Frekansları ve Ters Doküman Frekansları

Terim Frekansı, bir metin içerisinde bir terimin ne kadar geçtiğini, metindeki frekans değerini gösterir. Ters Doküman Frekansı ise birden fazla metinde bir kelimenin tekrar etme sayısına göre kelimenin terim olup olmadığının anlaşılmasını sağlar. Kelime vektörlerinin oluşturulması için bu yöntem kullanılmıştır. Yorumların kelime kelime analiz edilmesi için n-gram yöntemi uygulanmıştır. N-gram değeri 2 olarak seçilerek kelimelerin ikili birleşim vektörleri oluşturulmuştur.

2.4. Sınıflandırma Algoritmaları

Bu çalışmada veri madenciliğinin k en yakın komşu, Naive Bayes ve Karar Ağaçları sınıflandırma algoritmaları kullanılmıştır. Bu modellerin mümkün olduğunca objektif ve doğru bir şekilde performans tahmini elde edebilmesi için 10 kat çapraz doğrulama tekniği kullanılmıştır.



Şekil 2 RapidMiner Uygulaması

2.4.1. k-En Yakın Komşu

Makine öğrenmesinde k En Yakın Komşu algoritması hem sınıflandırma hem de regresyon yöntemlerinde kullanılan çok yönlü bir algoritmadır. Bu algoritma sınıfı bilinmeyen bir veriyi, veri setindeki diğer veriler ile karşılaştırarak hesaplanan uzaklığa göre veriyi en uygun sınıfa yerleştirir (Dilki, 2020). Verilerin boyut ve yapısına bağlı olarak önceden belirlenen k değerine göre sınıflandırma gerçekleştirir (Khan vd., 2002). Çalışmada optimal k sınıf değerini belirlemek için Optimize Parameters (Grid) operatörü kullanılmış ve en uygun k değeri 31 bulunmuştur. Uzaklık ölçümü için Euclidean mesafe ölçütü kullanılmıştır. Euclidean mesafe hesaplamada en çok kullanılan uzaklık ölçüsüdür (Hu vd., 2016). Euclidean uzaklığı, iki nokta arasında,

$$x_1 = (x_{11}, x_{12} \dots x_{1n}) \text{ ve } x_2 = (x_{21}, x_{22} \dots x_{2n}) \qquad 1$$

olmak üzere,

$$Dist_{euclidean}(x_1 x_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \qquad 2$$

formülü ile hesaplanmaktadır.

2.4.2. Naive Bayes

Bayes teoremine dayanan Naive Bayes sınıflandırıcısı, ilgili kayıtların hangi sınıfa hangi olasılıkla ilişkili olduklarını belirlemektedir. Veri setindeki özellikleri birbirinden bağımsız ve tüm özelliklerin aynı düzeyde önemli olduğunu kabul etmektedir. Bayes teoremine göre bir C sınıfının bilinmesi koşulunda x özniteliğinin var olma olasılığı ile x özniteliğinin bilinmesi koşulunda C sınıfının var olma olasılığı mevcuttur (Sarıkoç, 2012).

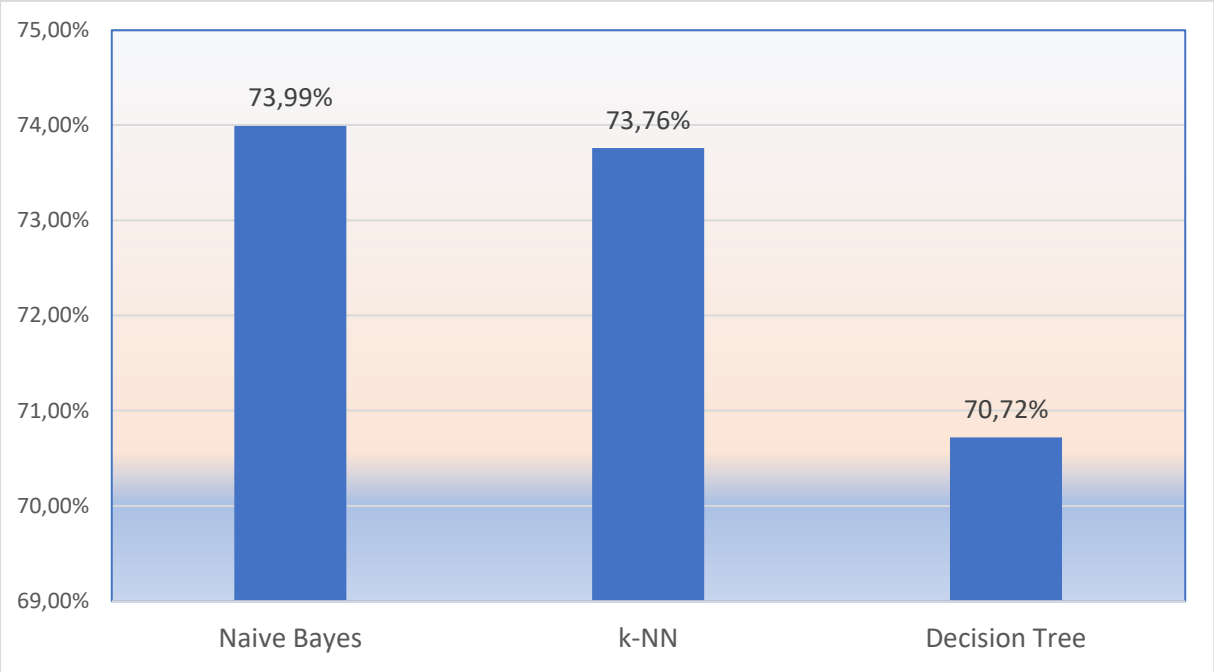
$$P(C).P(x \mid C)\ P(C \mid x) = P(x) \qquad 3$$

2.4.3. Karar Ağaçları

Karar ağaçları, sade ve anlaşılır bir yapıya sahip olan, verileri böl ve yönet yöntemi ile sınıflandırma ve regresyon gerçekleştiren bir algoritmadır (Kalaycı, 2018). Bu algoritmalarda bölme ve budama olarak iki temel işlem gerçekleşir ve algoritma durdurma kriterine göre sonlandırılır. (Bounsaythip ve Esa, 2001). Karar ağacı oluşturmada bilgi kazancı, ki-kare istatistiği, GINI indeksi gibi düğüm ayırma ölçütleri kullanılmaktadır (Kothari ve Dong., 2001). Çalışmada en çok tercih edilen C4.5 algoritması (Quinlan, 1993) ve bilgi kazancı düğüm ayırma ölçütü kullanılmıştır. Bu algoritma karar ağacı oluşturma sürecinde veya sonrasında bazı düğüm ve alt ağaçları silerek aşırı uygunluk problemini ortadan kaldırmakta ve veri setindeki gürültülü kayıtların kaldırılmasını sağlamaktadır (Niuniu ve Yuxun, 2010).

2.5. Bulgular

Makine öğrenmesi algoritmaları verinin hangi sınıfa ait olduğunun tahminini yapar. Çalışmada uygulanan sınıflandırma algoritmalarından %73,99 doğruluk değeri ile en yüksek başarı Naive Bayes algoritması ile elde edilmiştir (Şekil 3).



Şekil 3 Başarı Grafiği

2.5.1 Karmaşıklık matrisi

Veri setindeki gerçek değerler ile sınıflandırma algoritması tarafından oluşturulan tahmin değerlerini gösterir. Makine öğrenmesi algoritmalarının değerlendirilmesinde ve karşılaştırılmasında gerek duyulan ölçütlerin çoğu karmaşıklık matrisine göre hesaplanır.

Tablo 3 Karmaşıklık matrisi

		Tahmin Değerleri	
		True	**False**
Gerçek Değerler	**True**	TP	FN
	False	FP	TN

TP (Doğru Pozitif) : Doğru tahmin edilen pozitif kayıt sayısı
TN (Doğru Negatif) : Doğru tahmin edilen negatif kayıt sayısı
FP (Yanlış Pozitif) : Yanlış tahmin edilen pozitif kayıt sayısı
FN (Yanlış Negatif) : Yanlış tahmin edilen negatif kayıt sayısı
Veri setine uygulanan sınıflandırma algoritmalarının karmaşıklık matrisi karşılaştırması Tablo 4'de verilmiştir. Bu sonuçlara göre en iyi tahmin değerinin Naive Bayes algoritmasına ait olduğu görülmektedir.

Tablo 4 Sınıflandırma algoritmalarının karmaşıklık matrisi karşılaştırması.

			Tahmin Değerleri		**Doğruluk (%)**
			İlgili	**İlgisiz**	
Gerçek Değerler	**Naive Bayes**	**İlgili**	1982	697	73,99
		İlgisiz	550	1566	
	k-En Yakın Komşu	**İlgili**	1882	797	73,76
		İlgisiz	461	1655	
	Karar Ağacı	**İlgili**	1582	1097	70,72
		İlgisiz	307	1809	

2.5.2 Değerlendirme Ölçütleri

Doğruluk (Accuracy); doğru tahmin edilen kayıtların toplam kayıt sayısına olan oranıdır. Bütün kayıtlar doğru bir şekilde tahmin edildiği ölçüde bu değer 1 (en yüksek), doğru tahmin edilmediği takdirde bu değer 0 (en düşük) olarak sonuçlanır.

$$\text{Doğruluk} = (TN+TP) / N \quad (4)$$

Hata oranı; yanlış tahmin edilen kayıtların toplam kayıt sayısına olan oranıdır. Bütün kayıtlar yanlış bir şekilde tahmin edildiği ölçüde bu değer 1 (en yüksek), yanlış tahmin edildiği takdirde bu değer 0 (en düşük) olarak sonuçlanır.

$$Hata\ oranı = (FN+FP) / N \quad (5)$$
$$Hata\ oranı = 1 - (TN+TP) / N \quad (6)$$

Kesinlik (Precision); doğru tahmin edilen pozitif değerlerin, tahmin edilen toplam pozitif değerlere oranıdır.

$$Kesinlik = TP / (TP+FP) \quad (7)$$

Duyarlılık (Sensitivity); doğru tahmin edilen pozitif değerlerin, doğru tahmin edilen pozitif değerler ile yanlış tahmin edilen negatif değerlerin toplamına olan oranıdır. Pozitiflerin tamamı doğru tahmin edilmiş olursa bu değer 1 (en yüksek) olur. Pozitiflerin tamamı doğru tahmin edilmemiş olursa bu değer 0 (en düşük) olur.

$$Duyarlılık = TP / (TP+FN) \quad (8)$$

F Ölçütü (F Measure); F-ölçütü, duyarlılık ve kesinlik değerlerinin harmonik ortalamasıdır.

$$F - \text{Ölçütü} = 2 \times (Kesinlik \times Duyarlılık) / (Kesinlik+Duyarlılık) \quad (9)$$

Çalışmada uygulanan modellerin değerlendirme ölçütleri Tablo 5'de yer almaktadır.

Tablo 5 Sınıflandırma algoritmalarının karşılaştırılması

Sınıflandırma Algoritmaları	Doğruluk (%)	Hata Oranı (%)	Kesinlik (%)	Duyarlılık (%)	F-Ölçütü (%)
Naive Bayes	**73,99**	26,01	69,27	74,01	71,53
k-En Yakın Komşu	73,76	26,24	67,54	78,22	72,45
Karar Ağacı	70,72	29,28	62,30	85,49	72,05

3. SONUÇ

Günümüzde internet üzerinden yapılan alışveriş başta ülkemiz olmak üzere tüm dünyada hızla artmaktadır. Tüketiciler, online alışverişi geleneksel alışverişe kıyasla daha fazla tercih etmektedir. E-ticarette ürünü satın alan kişiler ürünün markası ve fiyatı yanında ürün hakkında yapılmış yorumları dikkate alarak karar vermektedirler. Ancak paylaşılan kullanıcı yorumlarının çoğunun ürün ile ilgili olmaması ve ilgisiz yorumların aşırı olması müşterilerin sipariş aşamasında karar vermelerini olumsuz yönde etkilemektedir. Bu çalışmada metin madenciliği teknikleri kullanılarak, e ticaret sitelerinden bir ürünü satın alan müşterinin, satın aldığı ürün deneyimi sonrasında paylaştığı çevrimiçi yorumların ürün ile ilgili olup olmadığı tespit edilmesi amaçlanmıştır.

Yapılan bu çalışma sadece ürün hakkında bilgi veren yorumları filtreleme olanağı sunarak müşterilerin sipariş sürecinde karar vermelerini kolaylaştıracak ve ürün seçiminde büyük kolaylık sağlayacaktır. Özellikle tüketicinin zamandan tasarruf etmesine de etkisi olacaktır. Ayrıca e-ticaret firmaları ürün hakkında paylaşılan müşteri yorumlarını dikkate alarak ürünlerinin zayıf ve eksik yönlerini görebileceklerdir. Böylece ürün geliştirme sürecine de büyük ölçüde katkı sağlayacaktır. Farklı sektörlerdeki firmalar ürünleri üzerinde daha iyi stratejiler geliştirebileceklerdir.

KAYNAKÇA

Akgöz, E. ve Tengilimoğlu, E. (2022). Online Otel Yorumlarının Metin Madenciliği Teknikleri ile İncelenmesi: Bakü Otelleri Örneği. Selçuk Turizm ve Bilişim Araştırmaları Dergisi, 1(1), 1–8.

Bounsaythip, C. ve Esa, R. R. (2001). Overview of Data Mining For Customer Behavior Modeling. VTT Information Technology Research Report, Version:1, s. 1-53.

Cop, R. ve Gümüş, N. (2009). Pazarlamada ağızdan ağıza iletişimin tüketici davranışlarındaki rolü ve bir araştırma. Ticaret ve Turizm Eğitim Fakültesi Dergisi, 2, 179-202

Çakar, E. N. ve Akbıyık, A. (2018). Focus Issue in Consumer Reviews on Fmcg: Is the Product Evaluated or the Sales Service. Ajit-e Online Academic Journal of Information Technology, 9(33), 147–158. https://doi.org/10.5824/1309-1581.2018.3.009.x

Da You Liu ve arkadaşları. (2012). Design of an Enhanced Fuzzy k-nearest Neighbor klassifier Based Computer Aided Diagnostic Systemfor Thyroid Disease. Journal of Medical Systems, 36:3243–3254, DOI 10.1007/s10916-011-9815-x.

Dilki, G., ve Deniz Başar, Ö. (2020). Comparison Study of Distance Measures Using K-Nearest Neighbor Algorithm on Bankruptcy Prediction. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 19(38), 224–233. http://dergipark.gov.tr/ticaretfbd.

Dirsehan, T. (2016). An Application of Text Mining to Capture and Analyze eWOM: A Pilot Study on Tourism Sector. (Edit) Rathore, S. and Panwar, A. Capturing, Analyzing, and Managing Word of Mouth in the Digital Marketplace. (pp. 168-186). IGI Global, USA.

Ergün, K. (2012). Metin Madenciliği Yöntemleri ile Ürün Yorumlarının Otomatik Değerlendirilmesi. 14–51.

Ertek, G., Tapucu, D. ve Arın, I. (2013). Text Mining with RapidMiner. (Edit) Markus H. and Ralf K. RapidMiner: Data Mining Use Cases and Business Analytics Applications. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. Chapman and Hall/CRC.

Hu, L.Y., Huang, M.W., Ke, S.W., Tsai, C.F. (2016). The Distance Function Effect on kNearest Neighbor Classification for Medical Datasets. Springer Plus, 5(1), 1-9.

Işık, N. (2019). Metin Madenciliği Yöntemleri ile E-Ticaret Markalarına Yönelik Sosyal Medya Yorumlarının Analizi.

Kalaycı, T. (2018). Kimlik hırsızı web sitelerinin sınıflandırılması için makine öğrenmesi yöntemlerinin karşılaştırılması. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 24 (5) , 870-878. Erişim Adresi https://dergipark.org.tr/tr/pub/pajes/issue/39683/469468.

Khan, M., Ding, Q., Perrizo, W. (2002), K-nearest Neighbor Classification on Spatial Data Streams Using P-trees. Advances in Knowledge Discovery and Data Mining, 2336, 517-528.

Kothari, R. ve Dong, M. (2001). Decision Trees for Classification: A Review and Some New Results. Pattern Recognition: From Classical to Modern Approaches, Editör: S. K. Pal, A. Pal, World Scientific, New Jersey.

Niuniu, X. ve Yuxun, L. (2010). Review of Decision Trees. The Third IEEE International Conference on Computer Science and Information Technology, 105- 109.

Quinlan, J. R. (1993). C4.5: Programs for Machine Learning, Morgan Kauffman, San Francisco, CA, USA.

Sarıkoç, F. (2012). Meme Kanseri Biyopsi Örneklerinde Östrojen Hormonu Alıcı Varlığının Zeki Yöntemlerle Tespiti. Doktora Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri, 148s.

Tuna, M. F. (2019). Çevrimiçi Yorum ve Şikâyetlerin Otel İşletmeleri Üzerinden Duygu Analizi ile İncelenmesi. 1, 1–476.

TÜİK. (2022). https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2022-45587.

Yücel, A. (2021). Text Mining and Data Dimesion Reduction Approach on Consumer Comments. Journal of Intelligent Systems: Theory and Applications, 4(1), 8–16. https://doi.org/10.38016/jista.801300