

Makine Öğrenmesi ile Müşteri Şikayetlerinin Sınıflandırılması*

Kutan KORUYAN, Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Dr. Öğr. Üyesi, 0000-0002-3115-5676

Ahsen EKERYILMAZ, Dokuz Eylül Üniversitesi, ekeryilmazahsen@gmail.com, 0000-0001-7603-618X

ÖZ

Müşteri memnuniyetinin bir göstergesi olan müşteri yorumları işletmeler açısından önemli bir veri kaynağı oluştururken, aynı zamanda, işletmelerin ürün veya hizmetlerinde iyileştirme ve hataları düzeltebilmeleri açısından önem kazanmaktadır. Bunun yanında, farklı sorunlara yönelik müşteri yorum sayısı arttıkça çözüm önerileri geliştirme açısından bu büyüklükteki veri ile baş edebilmek zorlaşmaktadır. Bu çalışmada, Türkiye'nin önde gelen üç tüketici elektroniği perakende firmasına yapılan müşteri şikayetleri, çevrimiçi şikâyet platformlarından şikayetoar.com sitesinden alınmış, şikayetlerin otomatik kategorilere ayrılması ve analizi yapılmıştır. Çalışmada, İade/Değişim ve Geri Ödeme, Teslimat/Kargo, Müşteri İlişkileri ve Hizmetleri ve Garanti ve Servis olmaz üzere dört kategori belirlenerek, denetimli makine öğrenmesi algoritmalarından Lojistik Regresyon, Doğrusal Destek Vektör Makineleri ve Olasılıksal Dereceli Azalma kullanılarak eğitim ve test süreci gerçekleştirilmiştir. En başarılı sonuç %80 doğruluk ile Lojistik Regresyon ile sağlanmıştır. Bu sonuca göre etiketlenmemiş müşteri şikayetleri Lojistik Regresyon ile tahminlenip elde edilen bulgular yorumlanmış, işletmeler açısından büyük miktardaki şikâyet verisinin tahminlenerek kategorilere ayrılmasının önemi irdelenmiştir.

Anahtar Kelimeler : Makine Öğrenmesi, Metin Sınıflandırma, Müşteri Şikayetleri, Doğal Dil İşleme

* Bu çalışma, Ahsen EKERYILMAZ tarafından Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü'nde, Dr. Öğr. Üyesi Kutan KORUYAN danışmanlığında yürütülmüş ve başarılı bulunmuş olan "Yapay Zeka Teknikleri ile Müşteri Şikayetlerinin Otomatik Kategorilere Ayrılması" isimli yüksek lisans tez çalışmasından türetilmiştir.



Classification of Customer Complaints with Machine Learning

ABSTRACT

Customer comments, an important indicator of customer satisfaction, are a valuable data source for businesses and a significant tool for improving and correcting product or service errors. In addition, increasing number of customer comments for different problems complicates developing solutions with such a large amount of data. In this study, customer complaints submitted to three leading consumer electronics retail companies in Turkey were collected from the online complaint platform, sikayetvar.com, and automatic categorization and analysis of the complaints were conducted. In the study, four categories were determined as Return/Exchange and Refund, Delivery/Shipping, Customer Relations and Services, and Warranty and Repairs, and the training and testing process was carried out using supervised machine learning algorithms, i.e., Logistic Regression, Linear Support Vector Machines and Stochastic Gradient Descent. The most successful result was obtained via Logistic Regression with 80% accuracy. According to this result, unlabeled customer complaints were estimated by Logistic Regression, and the findings were interpreted. Subsequently, the importance of estimation and categorization of a large amount of complaint data for businesses was examined.

Anahtar Kelimeler : *Machine Learning, Text Classification, Customer Complaints, Natural Language Processing*

EXTENDED ABSTRACT

Customer comments, an indicator of customer satisfaction, are an important data source for businesses to improve their products or services and correct their mistakes. Today, most customers use social networks, forums or websites to express their opinions about products and services, and these resources allow businesses to easily access textual data. At the same time, considering the importance of businesses to produce solutions to problems in a short time, it becomes difficult for them to cope with large amounts of data for different problem categories. Therefore, in this study, complaints made to three leading consumer electronics retail companies in Turkey in 2020 were classified and analysed by automatic text classification technique using machine learning algorithms.

The data used in the study is taken from the online complaint platform sikayetvar.com and consists of two parts: customer complaints and complaint dates. Complaints belonging to three different companies, defined as A, B and C, consist of lines 9229, 7534 and 4897, respectively. In the machine learning phase, firstly, four complaint categories were designated as Return/Exchange/Refund, Delivery/Shipping, Customer Relations/Services and Warranty/Service, and some of the data were labelled according to these categories. Then, the training and testing process was carried out by using the supervised machine learning algorithms Logistic Regression, Linear Support Vector Machines and Stochastic Gradient Descent. The most successful result from among the three algorithms was obtained via Logistic

Regression with 80%. The results for Stochastic Gradient Descent and Linear Support Vector Machines algorithms were 71% and 76%, respectively.

Based on these results, classes of remaining unlabelled complaint data were estimated by Logistic Regression. After estimation, the complaints received by the companies were compared with each other according to the categories. In addition, there is an increase in complaints in April, May and June for all three companies, with the highest number of complaints in the Customer Relations/Services category followed by the Delivery/Cargo category. This indicates businesses have not been able to respond to the high demand, considering that the majority of the consumers do their shopping online due to the COVID-19 pandemic.

GİRİŞ

Günümüzde ham verinin değerlendirilip, analiz edilmesi işletmeler açısından vazgeçilmez bir süreçtir. İşletmelerin kar marjlarını arttırması; kendileri, rakipleri veya piyasa ile ilgili sayısal veya sözel (metin) verilerin uygun analizleri ile sağlanmaktadır. İşletmeler verinin işlenmesi sonucunda edinilen yeni bilgiler ışığında aldıkları doğru kararlarla varlıklarını sürdürebilmektedirler. Bunun yanında, bugünün internet kullanımının yaygınlaşması işletmeler için gerekli olan veriye ulaşma güçlüğüne geçmişe göre azaltmıştır. Fakat, bu da işletmelerin eskiye göre daha fazla verinin işlenmesi ile başa çıkması zorunluluğunu doğurmuştur.

İşletmeler günümüzde üretim, finans, piyasa araştırmaları, müşteri deneyimi ve daha birçok farklı kaynaktan gelen veriyi işleme mecburiyetindedir. İşletmelerin analiz etmeleri gereken verilerden biri ise sattıkları ürün veya hizmetler hakkında yapılan müşteri yorumlarıdır. Çünkü yapılan çalışmalar da göstermiştir ki yorumlar müşterinin memnuniyeti veya memnuniyetsizliğinin bir göstergesidir (Keskinkılıç vd., 2016; Ercan, 2019; Arı & Akbıyık, 2022). Özellikle internet kullanımının artması ile birlikte müşterilerin bir kısmı satın alma işlemi gerçekleştirildikten sonra ürün veya hizmetler ile ilgili düşünce veya şikayetlerini internet üzerinden paylaşabilmektedir. Bu paylaşımlar işletmeler açısından paha biçilmez bir veri kaynağı sağlarken, müşteri yorumlarının analizi kendilerini geliştirebilmeleri ve hatalarını düzeltebilmeleri için fırsatlar yaratmaktadır (Yılmaz, 2014, s. 144). Çünkü, ürün veya hizmet ile ilgili yorumlar olumsuz da olsa ve bu olumsuzluk bazı işletmeler için bir dezavantaj olarak gözükse de en çok şikâyet alan konular için çözümler üretmek işletmelerin hizmet kalitesini arttıracığı anlamına gelmektedir. Örneğin teslimat ve kargo ile ilgili olumsuz yorumlardaki gözle görülür bir artış, işletmenin satış ve dağıtım kanallarının yeniden değerlendirilmesini, müşteri ilişkileri ve hizmetlerindeki şikayetler ise şirketin satış elemanları üzerinde yeni eğitim programlarının düzenlenmesini sağlayabilmektedir. Bunun yanında,

şikayetlerin analizi, mevcut müşterileri korumakla birlikte, satın alma işlemini gerçekleştirecek olan yeni müşterileri de kolay ve hızlı çekebilme imkânı oluşturacaktır.

Özellikle sosyal ağlar, forumlar veya internet sitelerindeki müşteri yorumları işletmelerin müşteri davranışlarına yönelik araştırmalar yapmaları ve iş potansiyellerini arttırmaları açısından metinsel veriye ulaşabilmelerini sağlayan önemli veri kaynaklarıdır (Çınar, 2019, s. 105). Fakat sürekli biriken, büyük miktardaki metin verisinin analizi, sınıflandırılması veya yorumlanması işletmeler açısından kısa sürede üstesinden gelinecek kolay bir eylem değildir. Bu yüzden, makine öğrenmesi (ML) ve özellikle otomatik metin sınıflandırma yönteminin kullanılması işletmeler için büyük kolaylık ve faydalar getirmektedir. Otomatik metin sınıflandırma hem zaman hem maliyet konusunda tasarruf sağlanmasını olanaklı kılmaktadır (Özdemir vd., 2021, s. 531).

İşletmelerin büyük miktarda ve devamlı artan müşteri şikâyet verilerini otomatik olarak analiz etme, sınıflandırma ve yorumlama gerekliliği bu çalışmanın çıkış noktasını oluşturmuştur. Buna yönelik olarak da Türkiye'nin üç büyük tüketici elektroniği perakende firmasına ait 2020 yılı şikâyet verileri veri kaynağı olarak kullanılmıştır. Şikayetler, şikâyet platformları arasında yaygın olarak kullanılan şikayetvar.com sitesinden elde edilmiştir. Daha sonra, İade/Değişim ve Geri Ödeme, Teslimat/Kargo, Müşteri İlişkileri ve Hizmetleri, Garanti ve Servis olmak üzere dört farklı kategori göz önünde bulundurularak üç ML algoritması denenmiş ve doğruluğu kıyaslanmıştır. Daha sonra, en doğru sonucu veren algoritma ile şikayetler dört kategori altında tahminlenip sınıflandırılmıştır. Bunun yanında, şikayetlerin sınıflandırılması doğrultusunda ürün veya hizmet satışı gerçekleştiren işletmeler için müşteri memnuniyetini ve hizmet kalitesini arttırmak, müşteri memnuniyetsizliği yaşanan konuları tespit edip iyileştirmeler sunmak amacıyla öneriler sunulmaktadır.

1. LİTERATÜR İNCELEMESİ

Literatürde farklı araştırma konularına odaklanılmış metin madenciliği ve ML çalışmaları bulunmakta olup, bu çalışmaların çoğu metin sınıflandırma ve tahminleme üzerine yoğunlaşmıştır. Metin sınıflandırmaya örnek olarak, İkonomakis vd.'nin (2005) çalışmasında çeşitli ML algoritmaları ile dijital formattaki belgelerin sınıflandırması yapılırken, Amasyalı ve Yıldırım (2004), Chen vd. (2016) ve Başkaya ve Aydın (2017) gazete haberlerini otomatik metin sınıflandırma yöntemi ile sınıflandırmışlardır. Kaşıkçı ve Gökçen'in (2014) çalışmasında internet sitelerinin e-ticaret sitesi olup olmadığı tespit edilmeye çalışılırken, Kazan ve Karakoca'nın (2019) çalışmalarında ise bir e-ticaret sitesindeki ürünlerin elde edilen ürün bilgileri yardımıyla sınıflandırılması yapılmıştır. Tretyakov (2004) ve Uysal vd. (2012) sırasıyla istenmeyen e-posta ve SMS, Ahmed vd. (2017) ise sahte yorumların belirlenmesi üzerine çalışmalarında bulunmuşlardır.

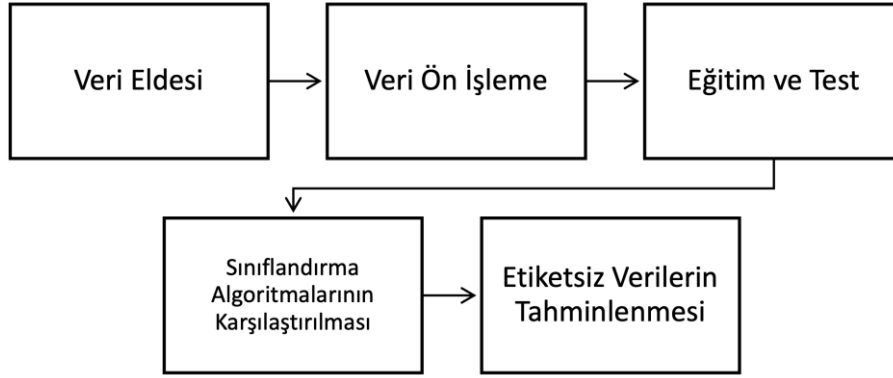
Tahminleme üzerine yapılan çalışmalara örnek oluşturacak Lykourentzou vd.'nin (2009) çalışmasında öğrencilerin eğitim aldıkları kursu bırakıp bırakmamalarının

tahminlemesi yapılmıştır. Ayrıca, Barboza vd. (2017) işletmelerin iflas tahmini ve Bilgin (2017) ise bir metnin hangi yazara ait olabileceğinin tahminlenmesi üzerine ML algoritmalarını kullanmışlardır.

Bireylerin bir işletme veya kurum hakkındaki yorum, talep veya şikayetleri üzerine yapılan çalışmalara örnek olarak Choe vd. (2013) Naïve Bayes sınıflandırıcısı kullanarak Koreli bir cep telefonu servis sağlayıcısı çağrı merkezine iletilen müşteri şikayetlerini sınıflandırmıştır. Gunawan vd. (2018) ise Twitter hesabının bir müşteri hizmetleri bölümü olarak kullanıldığı Endonezyalı PT Telkom şirketine gelen her türlü müşteri yorumlarını önceden tanımlanmış kategorilere otomatik ayrılmasını sağlamışlardır. Ghazzawi ve Alharbi (2019) New York Metropolitan Ulaştırma Müdürlüğü'ne gelen şikayetleri, hizmet kalitesinin iyileştirilmesi ve düşük müşteri memnuniyeti seviyelerine yol açan faktörlerin belirlenmesi amacıyla ML algoritmalarıyla analiz edilmiştir. Joung vd., (2019) bir şirketin farklı kaynaklardan topladıkları müşteri şikayetlerini denetimsiz ML yöntemleri ile kümelemişlerdir. Blümel ve Zaki (2022) müşteri deneyimini geliştirmek amacıyla gelen müşteri yorumlarından şikayetlerin ayrıştırılması ve önceliklendirilmesi için çeşitli ML algoritmalarından yararlanmışlardır. Yıldız (2016), şikâyet, teşekkür, görüş ve önerilerdeki faydalı anahtar kelimelere yoğunlaşmış ve üst yönetim destek sistemleri için bir modül önermiştir. Büyükeke vd. (2020) sosyal ağ verisi kullanarak müşteri yorumlarından turistik bir destinasyona yönelik duygu analizi yapmışlardır. Göker ve Tekedere (2017), Aytakin vd. (2018), Arslan vd. (2019), Tekin ve Tunalı (2019) ve Hoşgör ve Tosun (2019) ise kişiler tarafından işletme veya kuruma iletilen yorum, talep ve şikayetleri sınıflandırmışlar ve kullanılan yöntemlerin karar alma süreçlerine destek sağlayacağını belirtmişlerdir. Bu eserlerde, binlerce satırlık verinin manuel olarak sınıflandırılmasının neredeyse imkânsız olduğu, büyük miktardaki verinin işlenmesiyle stratejilerin daha kolay geliştirilebileceği irdelenmiştir. Ayrıca, özellikle şikayetlerin ML kullanılarak kolayca sınıflandırılması ile müşterilere sorunları ile ilgili daha kolay ulaşılabileceği belirtilmiştir. Aynı zamanda, yapılan yorum ve şikayetlerin işletmeler için bir lütuf olarak değerlendirilmesinin gerekliliği ile bir şikâyet yönetim sisteminin kurulmasının müşteri memnuniyetinde artış sağlayacağı öngörülmüştür

2. METODOLOJİ

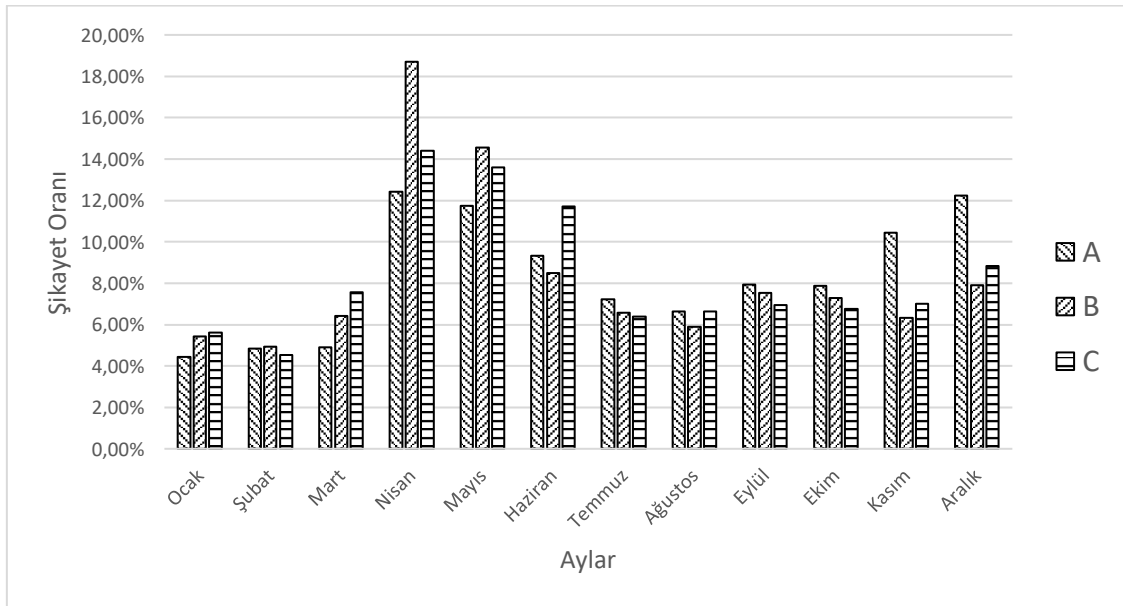
Çalışmada seçilen firmaların en çok hangi konularda şikâyet aldıklarının sınıflandırılması ve analizi Python programlama dili kullanılarak gerçekleştirilmiştir. Veri kaynağı olarak üç tüketici elektroniği perakende firmasına ait şikâyet verileri kullanılmıştır. Çalışmanın her adımını gösteren akış şeması Şekil 1'de gösterilmektedir.



Şekil 1: Çalışma Akış Şeması

2.1. Veri Eldesi

Çalışmada Türkiye'nin önde gelen şikâyet platformlarından biri olan sikayetvar.com sitesinden web kazıma metoduyla Python BeautifulSoup kütüphanesinden yararlanılarak, seçilen üç tüketici elektroniği perakende firmasına ait (A, B ve C olarak isimlendirilmiştir) ve 2020 yılını kapsayacak şekilde veriler elde edilmiştir. A, B ve C firmalarına ait sırasıyla 9229, 7534 ve 4897 satırlık şikâyet verisi bulunmaktadır. Şekil 2'de üç firmanın aylık şikayetlerinin kendi toplam şikayetlerine olan oranlarının grafiği verilmektedir. Şekil 2'den görüleceği üzere özellikle nisan, mayıs ve haziran aylarında her üç firmanın da şikayetlerinde bir artış gözlemlenmektedir. Bunun yanında, özellikle Nisan ayında B firması, haziran ayında C firması, kasım ve aralık aylarında ise A firmasına yönelik şikâyet oranlarında gözle görülür bir artış bulunmaktadır.



Şekil 2: 2020 yılı aylara göre A, B ve C Firmalarına Ait Toplam Şikayetlerine Olan Oranları

2.2. Veri Ön İşleme

ML modelleri kurulmadan önce etkisiz kelimeler (bağlaçlar, sayılar, kısaltmalar, vb.), noktalama işaretleri kaldırılmış ve tüm kelimeler küçük harfe dönüştürülmüştür. Son olarak ise Türkçe'nin sondan eklemeli bir dil olması nedeniyle kelimelerin en yalın hallerine getirilmeleri için çekim veya zaman ekleri gibi ekler alan kelimeler Python TurkishStemmer kütüphanesi kullanılarak köklerine ayrılmıştır.

2.3. Modelin Kurulması, Eğitim ve Test

Şikâyet verilerinin tümünden rastgele seçilen bir kısım veri, önceden belirlenen İade/Değişim ve Geri Ödeme, Teslimat/Kargo, Müşteri İlişkileri ve Hizmetleri ve Garanti ve Servis kategorileri olmak üzere dört kategoriye göre, her bir kategoriye eşit sayıda şikâyet verisi gelecek şekilde etiketlenmiştir. Daha sonra bu veri %80 eğitim ve %20 test olacak şekilde ayrılmıştır. Tablo 1'de örnek oluşturması açısından bu dört kategoriye ait birer şikâyet örneği verilmektedir. Model kurulumu, eğitim, test ve tahminleme aşamalarında Python ML kütüphanesi olan scikit-learn'den yararlanmıştır.

Tablo 1: Müşteri Şikâyetlerinin Kategori Bazında Örnek Gösterimi

Kategori	Yorum
İade / Değişim ve Geri Ödeme	***'ten aldığım ürün fişinde ve tanıtımında memnuniyetsizlik veya beğenmeme durumunda 14 gün içerisinde iade edilir ibaresi olduğu halde değişim olmadı ve insanları aldatıyorlar. Yazılanları gösterdiğimde bu güncellenmemiş iade yok diyorlar. Bu nasıl bir tutarsızlıktır?
Teslimat / Kargo	Satın aldığım ürünü sipariş 413**2468 kargoya verdiniz ama çalıştığımız *** firması yüzünden bir şikâyet de size yazıyorum. Nasıl böyle bir kargo şirketiyle çalışırsınız? Müşteri mağduriyeti sizi de bağlıyor bu yüzden. Teslim alamadığım ürün hala sizin sorumluluğunuzda. Ürün kargo merkezinde 1 haftaya yakındır bekliyor.
Müşteri İlişkileri ve Hizmetleri	Cumartesi günü *** AVM *** mağazasına elektronik alışverişi için girdik. Çalışanlar küme küme sohbet halinde stantlara yaslanmış, keyiflerini bozmuş olsak da "*** kulaklık var mı acaba?" diye sordum. Kollarını göğsünde bağlamış bir arkadaş "bilmiyorum" diye cevap verdi! Koskoca mağazada sadece bir kişi arkadaş bakmaya gitti dedi ancak kimse bize hangi model veya ne arıyorsunuz diye sormadı, diğerlerinin istifini bozmadan muhabbetlerine geri dönmeleri müşteri gözünden inanın hiç hoş bir görüntü değil. Bilmese bile, "ben ilgilenmiyorum ama hemen ilgili arkadaşını yönlendiriyorum" gibi bir cümle duymak isterdik. Çalışanlarınıza sanırım mağaza içi karşılama, hizmet nasıl

verilir eğitimi vermiyorsunuz. Başka türlü açıklaması dahi olamaz! *** tarafından iletişim kurulmasını talep ediyorum.

Garanti ve Servis Daha 3 4 ay önce ***'dan sıfır olarak aldığım bilgisayarım çok nazik kullanmama rağmen 2. defa bozuldu, ürün şarj almıyor. Garantiye gönderdiğimde Bios güncellemesi yapıp göndermişler ancak bir güncelleme 18 iş gününde yapıldı ve en çok ihtiyacım olan eğitim süremde çok mağdur oldum. Servisten geldi ve gene bir defa kullanamadan tekrar şarj almıyor.

Verinin bilgisayar tarafından sayısallaştırılarak anlaşılması, bir başka deyişle, belgelerin temelini oluşturan kelimelerin vektörlerle ifade edilmesi için çalışmada, metinsel verilerden öznitelik çıkarma yöntemlerinden biri olan Terim Frekans ve Ters Belge frekansı (Term Frequency * Inverse Document Frequency, TF*IDF) kullanılmıştır. Öznitelik çıkarma işleminin temel amacı, sınıfları ayırt etmek için yararlı olan tüm öznitelikler arasından en iyisini seçme işlemidir (Khalid vd, 2014). TF*IDF ise bir terimin doküman içinde ne kadar önemli olduğunu gösteren ve istatistiki yöntem ile hesaplanan bir ağırlık faktörüdür (Ekici & Takcı, 2021, s. 650). TF*IDF, bir sözcüğün belge içerisinde kaç kez geçtiğinin tüm sözcüklere oranı olan terim frekansı (TF) ile sözcüğü içeren metin sayısının toplam metin sayısına oranının logaritması olan ters belge frekansının (IDF) çarpımları ile elde edilmektedir (Teker, 2021, s. 137).

Çalışmada denetimli makine öğrenmesi sınıflandırma çalışmalarında sıklıkla kullanılan logistik regresyon (Logistic Regression, LR), olasılıksal dereceli azalma (Stochastic Gradient Descent, SGD) ve doğrusal destek vektör makinaları (Linear Support Vector Machine, LSVM) algoritmaları kullanılmıştır:

- **LR:** Sınıflandırma için kullanılan en yaygın makine öğrenmesi yöntemlerinden biridir. LR, bağımlı değişkenin kategorik olduğu durumlarda, bağımsız değişkenlerin ise sürekli veya kategorik olduğu durumlarda kullanılır. Bağımsız değişkenin kategorik olmasından dolayı gözlem verileri arasında doğrusal ilişki durumu sağlanamamakta, dolayısıyla, ilişkiyi doğrusal bir yapıya dönüştürmek için logaritmik dönüşüm kullanılmaktadır (Gamgam & Altunkaynak, 2021, s. 269-270). LR, bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi, temel lojistik (logit) fonksiyonunu kullanarak olasılıkları tahmin ederek ölçmekte ve hem ikili hem de çok sınıflı sınıflandırma için kullanılabilir (Gültepe, 2019, s. 12).
- **SGD:** Örneklerin rastgele seçildiği optimizasyon makine öğrenmesi tekniğidir. Bir değişkenin başka bir değişkendeki değişimlere göre değişim derecesini ölçer. Parametrelerin optimal değerlerini bulmak için maliyet fonksiyonunun olası değerini tekrar tekrar azaltır. Özellikle çok fazla verinin sınıflandırılmasında kullanılır (Deepa vd., 2020, s. 2000). Çalışmada, sınıflandırıcıların eğitimi, özellikle de SGD'nin eğitimi için Hinge maliyet fonksiyonu kullanılmıştır. Buna göre, sınırın negatif ve pozitif

örnekleri +1 ve -1 olarak ayrılmıştır. Fonksiyonun mesafesine bağlı olarak maruz kalacağı cezayı temsilen elasticnet uygulanmıştır.

- **LSVM:** Sınıflandırma ve eğri uydurma problemlerini çözmek için geliştirilen destek vektör makinesi algoritması, yapısal riski en aza indirme ilkesine dayanmakta olup, sınıfları sınıf etiketi ile tanımlanan girdi verilerini optimal hiper düzlemi belirterek iki ayrı sınıfa ayırma amacıyla kullanılır (Dalian vd., 2015). Destek vektör makineleri veri setinin doğrusal olarak ayrılıp ayrılmamasına göre ikiye ayrılmaktadır. Destek vektör makineleri ile sınıflandırma yönteminde, iki sınıfa ait örnekler doğrusal olarak dağıldığında, bu iki sınıf, eğitim verisi kullanılarak elde edilen karar fonksiyonu ile birbirinden ayrılır. Bu veri setini ikiye ayırmış olan doğrunun ismi, karar doğrusudur. LSVM, verileri ayırabilecek sonsuz doğru içinden marjini en yüksek doğrunun seçilmesidir. Bu çalışmada ikiden fazla sınıf olduğu için Python scikit-learn kütüphanesinde çoklu sınıf için OneVsRestClassifier fonksiyonu kullanılmıştır.

Eğitim ve test sürecinden sonra, algoritmaların sonuçlarının karşılaştırılıp en iyi sonuç veren algoritmanın bir sonraki tahminleme adımında kullanılması için başarı kriteri olarak doğruluk puanları (oranları) temel alınmıştır. Üç farklı algoritmanın sonuçlarına bakıldığında en başarılı sonuç %80 ile LR algoritmasında elde edilmiştir. Diğer algoritmalara yönelik sonuçlar ise SGD'nin %71 ve LSVM'nin ise %76'dır (Tablo 2).

Tablo 2. Eğitim ve Test Doğruluk Puanları

	LR*	SGD**	LSVM***
Doğruluk Puanı	0,8	0,71	0,76

* Lojistik Regresyon, ** Olasılıksal Dereceli Azalma, *** Doğrusal Destek Vektör Makinaları

2.4. Tahminleme

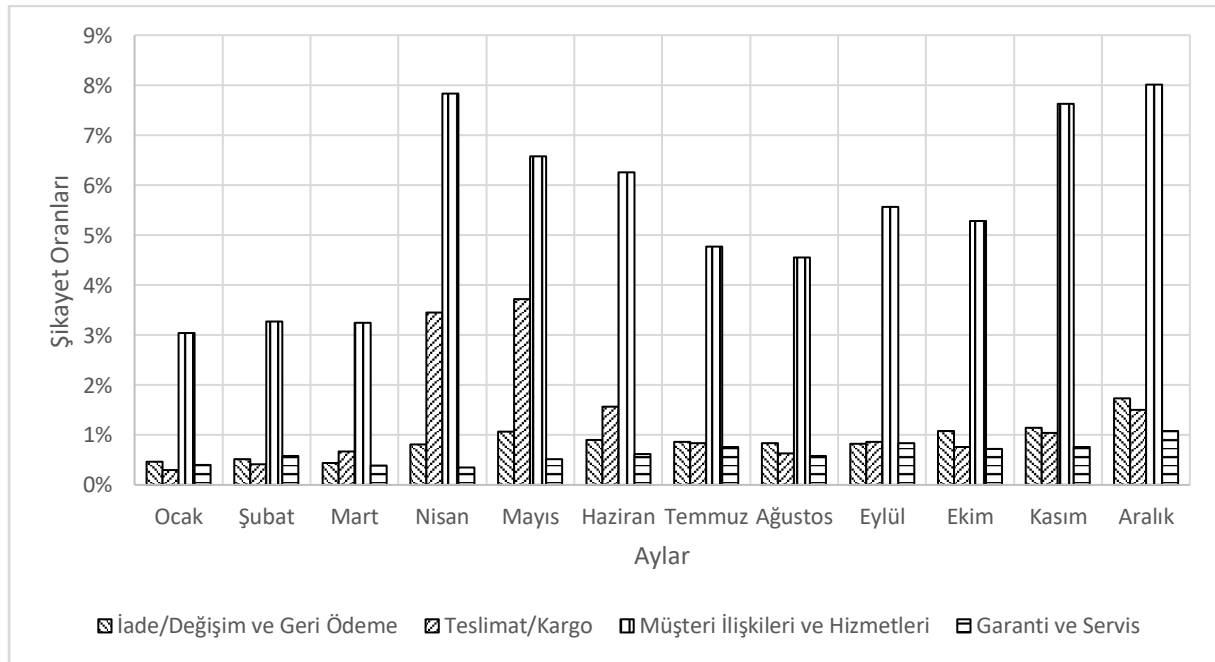
Metin sınıflandırma, ML teknikleri kullanılarak bir metin belgesinin önceden tanımlanmış bir dizi sınıfa otomatik olarak atanmasıdır (Dalal & Zaveri, 2011, s. 37). Çalışmanın son adımında, A, B ve C firmalarına ait 2020 yılında aldıkları şikayetler LR kullanılarak tahminlenmiş, tahminlenen şikâyet sayıları ve oranları ise Tablo 3'te verilmektedir.

Tablo 3.:2020 Yılı Üç Tüketici Elektroniği Perakende Firmasına Ait Şikâyet Oranları

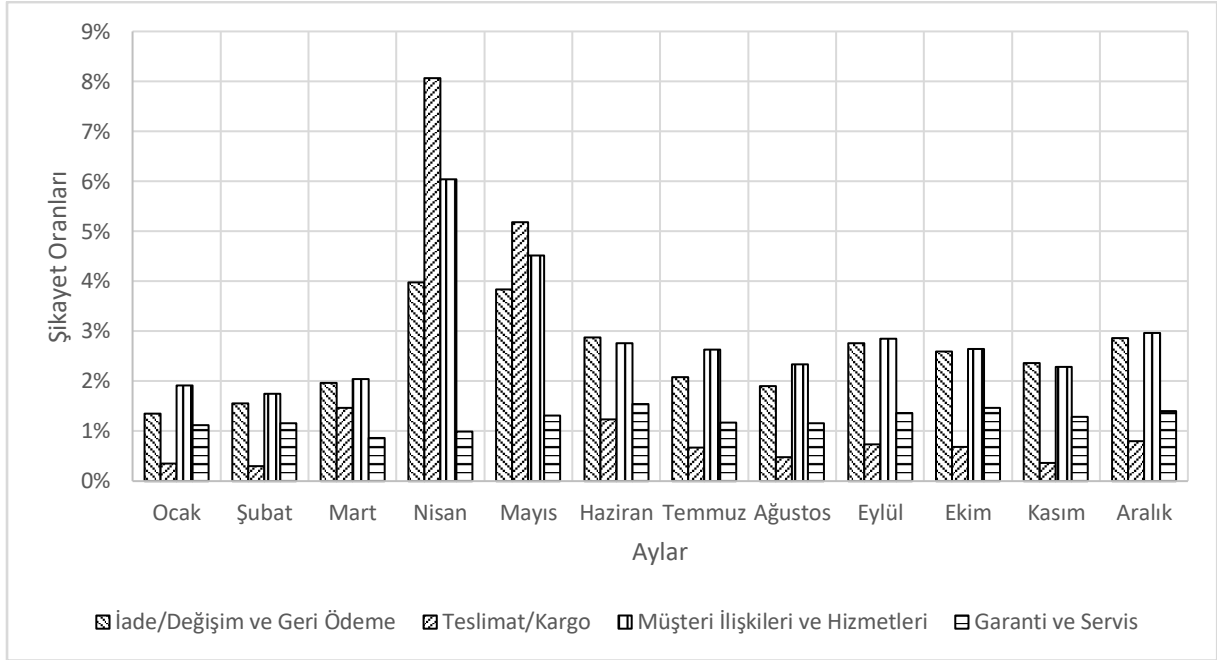
	A		B		C		TOPLAM	
	ŞS*	Oran**	ŞS	Oran	ŞS	Oran	ŞS	Oran
İade/Değişim ve Geri Ödeme	985	10,67	2270	30,13	943	19,26	4198	19,38
Teslimat/Kargo	1450	15,71	1531	20,32	711	14,52	3692	17,05
Müşteri İlişkileri ve Hizmetleri	6093	66,02	2616	34,72	2637	53,85	11346	52,38
Garanti ve Servis	701	7,60	1117	14,83	606	12,37	2424	11,19
TOPLAM	9229	100	7534	100	4897	100	21660	100

*Şikâyet sayısı, **%

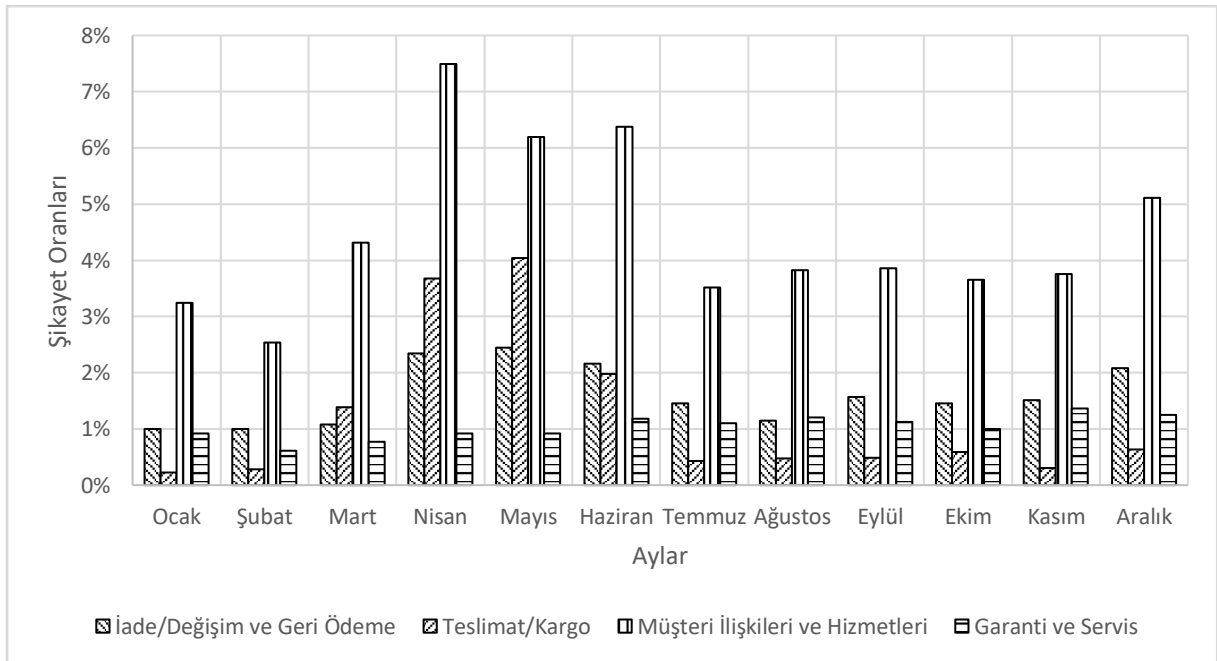
Şikâyet oranları incelendiğinde her üç firmanın da en fazla şikâyet aldıkları kategori Müşteri İlişkileri ve Hizmetleridir. İkinci en çok şikâyet ise A firmasında Teslimat/Kargo iken, B ve C firmalarında İade/Değişim ve Geri Ödeme kategorisi olarak görülmektedir. Özellikle A ve C firmalarında Müşteri İlişkileri ve Hizmetlerine ait şikâyet sayıları yılın her ayı yüksek iken, B firmasında bu durum gözlenmemektedir. Bunun yanında, her üç firmada da Koronavirüs salgını dolayısıyla Mart ayından sonra çevrimiçi siparişlerin artışı göz önünde bulundurulduğunda şikayetlerde bir artış gözlemlenmektedir. Her üç firmanın, aylık olarak aldıkları şikâyet sayılarının toplam şikayetlerine olan oranlarını gösteren grafikler Şekil 3, Şekil 4 ve Şekil 5'te verilmektedir.



Şekil 3: A Firmasının Aldığı Aylık Şikâyet Oranları ve Şikâyet Kategorileri



Şekil 4: B Firmasının Aldığı Aylık Şikâyet Oranları ve Şikâyet Kategorileri



Şekil 5: C Firmasının Aldığı Aylık Şikâyet Oranları ve Şikâyet Kategorileri

3. SONUÇLAR VE TARTIŞMALAR

İşletmeler ürün veya hizmetleri hakkındaki şikâyet sayılarında artış veya azalış veya az sayıdaki şikâyet verisiyle belirli iyileştirme ve geliştirme kararları alabilmeleri mümkünken, veri büyük miktarlara ulaştığında şikâyetlerin neyle ilgili olduklarını

sınıflandırmada zorlanabilmektedir. Günümüzde özellikle bu tür verilerin sınıflandırılması ve analizi için ML algoritmalarından yararlanılmaktadır.

Bu çalışmada, 2020 yılına ait üç tüketici elektroniği perakende firmasına ait müşteri şikayetleri önceden belirlenen kategorilere göre üç farklı denetimli ML algoritması ile incelenmiş, ML sonucundan en yüksek doğruluk %80 oranı ile LR'da çıkmıştır. Daha sonra, LR ile diğer etiketlenmemiş şikayetlerin hangi kategorilere dahil oldukları tahminlenmiştir.

Türkiye'de resmi olarak 11 Mart 2020'den itibaren Koronavirüs salgınının ortaya çıkması sonucunda tüketicilerin çoğu alışverişlerini internet üzerinden gerçekleştirmeye başlamıştır. Bu durum çevrimiçi alışverişlerdeki yoğun talebi karşılayamayan firmalarda şikâyet artışlarını beraberinde getirmiştir (Güven, 2020, s. 513). Çalışmada genel olarak üç firmanın şikâyet tahminleme sonuçları incelendiğinde Mart ayı ve sonrasında tüm şikayetlerde bir artış, en fazla şikayet alan kategori ise Müşteri İlişkileri ve Hizmetlerinin olduğu görülmüştür. Yani artan yoğun talebe karşılık bir süre işletmelerin müşteri ilişkilerinde yetersiz kaldığı söylenebilmektedir. Buna ek olarak, salgın döneminde kargo firmaları da artan talebe yetişemeyecek duruma gelmişlerdir (Topçuoğlu & Genç, 2021, s. 374). Özellikle Nisan, Mayıs ve Haziran aylarında Teslimat/Kargo kategorisinde yüksek bir artış gözlemlenmektedir. Çalışmada elde edilen bulgular sayılan bu olumsuz koşulları doğrular niteliktedir.

İşletmelerin başarısı, karlılıklarını arttırabilmeleri ve marka değerlerini koruyabilmelerine bağlı olmakta, her iki durumun sağlıklı yürütülmesi ve sürekliliği için müşteri memnuniyeti önem kazanmaktadır. Ayrıca, işletmelerin mevcut müşteri portföyünü koruyabilmesi ve yeni müşteriler kazanması, şirket imajının korunabilmesi ile mümkün olmaktadır. Bu nedenle, işletmeler açısından müşteri ilişkilerinde olumlu veya olumsuz geri bildirimlerin önemi büyüktür. Bu yolla işletmeler satış politikaları, kanalları ve yöntemleri ile müşteri ilişkilerini iyi yönetebilme ve olumsuz müşteri geri dönüşlerini dikkate alarak aksayan yönlerini düzeltme yoluna gitmelidirler. Çalışma bu yönüyle, büyük miktardaki müşteri şikâyet verisinin kolayca sınıflandırılması ve analiz edilmesi ile olumsuzlukların düzeltilmesi için işletmelerin anında harekete geçmesi ve yöneticilerin de karar vermelerini destekleyici bir yöntem ortaya koymaktadır. Aynı zamanda, işletmelerin analiz sonuçlarına göre ilgili iyileştirme, plan ve çözümleri kısa zamanda yapabilmelerini olanaklı kılmaktadır. Ayrıca bu çalışmada kullanılan yöntemin, yeni müşteriler edinme ve mevcut müşterileri de koruma hususunda şikâyet oranları ve konuları dikkate alındığında, işletmelere karlılık anlamında büyük katkılar sağlayacağı

düşünülmektedir. Bunun yanında bu yöntem, özellikle şikayetlerin “ne ile ilgili?” ve “hangi zaman aralığında?” gibi sorulara yanıt vermesine, elde edilecek bulgular doğrultusunda aksayan yönlerin ne olduğunun anlaşılmasına ve üretilecek çözümler ile gelecekte aynı sorunların tekrarlanmamasına yardımcı olacaktır.

Gelecekte yapılacak çalışmalarda daha geniş kapsamlı, birkaç yıllık verinin incelenmesinin yararlı olacağı düşünülmektedir. Ayrıca, kullanılan denetimli öğrenme yönteminden ötürü, önceden belirlenen kategorilere göre sınıflandırma yapma anlayışı başka sorunların, belki de farklı şikâyet kategorilerinin gözden kaçabileceği ihtimali getirmektedir. Bunun yanında, bu yöntem ile belirlenen kategoriler altındaki ayrıntıların (örneğin; hangi mamul ve ürün kategorisi, üretim hatası, ürün kalitesi, vb.) belirlenmesi zordur. Bu yüzden, ileride aynı veya benzer veri setleri ile denetimsiz öğrenme algoritmalarının kullanılması ve daha detaylı bir analiz yapılması düşünülmektedir.

KAYNAKÇA

- Ahmed, H., Traore, I., & Saad, S. (2017). Detecting Opinion Spams and Fake News Using Text Classification. *Security and Privacy*, 1(1), e9. <https://doi.org/10.1002/spy2.9>
- Amasyalı, M. F., & Yıldırım, T. (2004). Otomatik Haber Metinleri Sınıflandırma, *Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference*, Kuşadası, 224–226.
- Arı, O., & Akbıyık, A. (2022). Lojistik Regresyon ile Faydalı Müşteri Yorumlarını Tahminleme. *Journal of Research in Business*, 7(1), e15-32. <https://doi.org/10.29228/JRB.1024602>
- Arslan, H., Kaynar, O., & Şahin, S. (2019). Classification of Customer Demands by Using Doc2Vec Feature Extraction Method, *27th Signal Processing and Communications Applications Conference (SIU)*, Sivas, 1-4.
- Aytekin, Ç., Sütçü, C. S., & Özfidan, U. (2018). Karar Ağacı Algoritması ile Metin Sınıflandırma: Müşteri Yorumları Örneği. *Uluslararası Sosyal Araştırmalar Dergisi*, 11(55), 782-792. <https://doi.org/10.17719/jisr.20185537249>
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine Learning Models and Bankruptcy Prediction. *Expert Systems with Applications*, 83, 405–417. <https://doi.org/10.1016/j.eswa.2017.04.006>
- Başkaya, F., & Aydın, İ. (2017). Haber Metinlerinin Farklı Metin Madenciliği Yöntemleriyle Sınıflandırılması, *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1-5.
- Bilgin, M. (2017). Gerçek Veri Setlerinde Klasik Makine Öğrenmesi Yöntemlerinin Performans Analizi. *19. Akademik Bilişim Konferansı*, Aksaray, 1-6.
- Blümel, J. H., & Zaki, M. Comparative Analysis of Classical and Deep Learning-based Natural Language Processing for Prioritizing Customer Complaints. *Hawaii International Conference on System Sciences*, Hawaii, 1873-1882.
- Büyükeke, A., Sökmen, A., & Gencer, C. (2020). Metin Madenciliği ve Duygu Analizi Yöntemleri ile Sosyal Medya Verilerinden Rekabetçi Avantaj Elde Etme: Turizm Sektöründe Bir Araştırma. *Journal of Tourism and Gastronomy Studies*, 8(1), 322–335. <https://doi.org/10.21325/jotags.2020.550>
- Chen, Q., Yao, L., & Yang, J. (2016). Short Text Classification Based on LDA Topic Model, *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, 749-753.
- Choe, P., Lehto, M. R., Shin, G.-C., & Choi, K.-Y. (2012). Semiautomated Identification and Classification of Customer Complaints. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 23(2), 149–162. <https://doi.org/10.1002/hfm.20325>
- Çınar, A. (2019). Sınıflandırma Algoritmaları ile Bir Metin Madenciliği Uygulaması. M. Erdal Balaban ve Elif Kartal (Ed.), *Veri Madenciliği ve Makine Öğrenmesi: Temel Kavramlar, Algoritmalar, Uygulamalar*, 105-140, İstanbul: Çağlayan Kitabevi.
- Dalal, M. K., & Zaveri, M. A. (2011). Automatic Text Classification: A Technical Review. *International Journal of Computer Applications*, 28(2), 37-40. <https://doi.org/10.5120/3358-4633>

- Dalian, Y., Yilun, L., Songbai, L., Xuejunc, L., & Liyong, M. (2015). Gear Fault Diagnosis Based on Support Vector Machine Optimized by Artificial Bee Colony Algorithm. *Mechanism and Machine Theory*, 90, 219–229. <https://doi.org/10.1016/j.mechmachtheory.2015.03.013>
- Deepa, N., Prabadevi, B., Maddikunta, P. K., Gadekallu, T. R., Baker, T., Khan, M. A., & Tariq, U. (2020). An AI-Based Intelligent System for Healthcare Analysis Using Ridge-Adaline Stochastic Gradient Descent Classifier. *The Journal of Supercomputing*, 77(2), 1998–2017. <https://doi.org/10.1007/s11227-020-03347-2>
- Ekici, B., & Takcı, H. (2021). Spam Tespitinde Word2Vec ve TF-IDF Yöntemlerinin Karşılaştırılması ve Başarı Oranının Artırılması Üzerine Bir Çalışma. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, 8(2), 646-655. <https://doi.org/10.35193/bseufbd.935247>
- Ercan, F. (2019). Sosyal Medyada Otel İşletmelerine Yönelik Yorumların Müşteri Memnuniyeti ve Memnuniyetsizliği Açısından Analizi. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 22(2), 552–571. <https://doi.org/10.29249/selcuksbmyd.569684>
- Gamgam, H., & Altunkaynak, B. (2021). *SPSS Uygulamalı Regresyon Analizi*, Ankara: Seçkin Yayıncılık.
- Ghazzawi, A., & Alharbi, B. (2019). Analysis of Customer Complaints Data using Data Mining Techniques. *Procedia Computer Science*, 163, 62–69. <https://doi.org/10.1016/j.procs.2019.12.087>
- Göker, H., & Tekedere, H. (2017). FATİH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri ile Otomatik Değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(3), 291-299. <https://doi.org/10.17671/gazibtd.331041>
- Gunawan, D., Siregar, R. P., Rahmat, R. F., & Amalia, A. (2018). Building automatic customer complaints filtering application based on Twitter in Bahasa Indonesia. *Journal of Physics: Conference Series*, 978, 012119. <https://doi.org/10.1088/1742-6596/978/1/012119>
- Gültepe, Y. (2019). Makine Öğrenmesi Algoritmaları ile Hava Kirliliği Tahmini Üzerine Karşılaştırmalı Bir Değerlendirme. *European Journal of Science and Technology*, 16, 8–15. <https://doi.org/10.31590/ejosat.530347>
- Güven, H. (2020). Covid-19 Sürecinde E-Ticaret Sitelerine Yöneltilen Müşteri Şikâyetlerinin İncelenmesi. *Journal of Turkish Studies*, 15(4), 511–530. <http://doi.org/10.7827/TurkishStudies.44354>
- Hoşgör, H., & Tosun, N. (2019). İstanbul İlindeki Bir Üniversite Hastanesine Yapılan Şikâyetlerin İncelenmesi. *Cumhuriyet Üniversitesi Edebiyat Fakültesi Sosyal Bilimler Dergisi*, 43(2), 239-252.
- Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text Classification Using Machine Learning Techniques. *WSEAS Transactions on Computers*, 4(8), 966-974.
- Joung, J., Jung, K., Ko, S., & Kim, K. (2018). Customer Complaints Analysis Using Text Mining and Outcome-Driven Innovation Method for Market-Oriented Product Development. *Sustainability*, 11(1), 40. <https://doi.org/10.3390/su11010040>

- Kaşıkcı, T., & Gökçen, H. (2014). Metin Madenciliği ile E-Ticaret Sitelerinin Belirlenmesi. *Bilişim Teknolojileri Dergisi*, 7(1), 25-32. <http://doi.org/10.12973/bid.2014>
- Kazan, S., & Karakoca, H. (2019). Makine Öğrenmesi ile Ürün Kategorisi Sınıflandırma. *Sakarya University Journal of Computer and Information Sciences*, 2(1), 18-27. <https://doi.org/10.35377/saucis.02.01.523139>
- Keskinkılıç, M., Ağca, Y., & Karaman, E. (2016). İnternet ve Bilgi Sistemleri Kullanımının Turizm Dağıtım Kanallarına Etkisi Üzerine Bir Uygulama. *İşletme Araştırmaları Dergisi*, 8(4), 445-445. <https://doi.org/10.20491/isarder.2016.227>
- Khalid, S., Khalil, T., & Nasreen, S. (2014). A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning, *Proceedings of 2014 Science and Information Conference, SAI 2014*, London.
- Lykourantzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout Prediction in E-learning Courses through the Combination of Machine Learning Techniques. *Computers & Education*, 53(3), 950-965. <https://doi.org/10.1016/j.compedu.2009.05.010>
- Özdemir, A., Onan, A., & Çınarlı Ergene, V. (2021). İş Tatmini Faktörlerini Belirlemeye ve Analiz Etmeye Yönelik Olarak Çalışanların Çevrimiçi Değerlendirmelerinin Sınıflandırıcı Topluluklarına Dayalı Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, 28, 531-538. <https://doi.org/10.31590/ejosat.1008691>
- Teker, Ö. (2021). *Derinlemesine Python AI Natural Language Processing*, İstanbul: Godoro Yayıncılık.
- Tekin, M. C., & Tunalı, V. (2019). Yazılım Geliştirme Taleplerinin Metin Madenciliği Yöntemleriyle Önceliklendirilmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 25(5), 615-620. <https://doi.org/10.5505/pajes.2019.47827>
- Topçuoğlu Dönmez, D. ve Genç, E. (2021). COVID-19 Pandemi Sürecinin Kargo Sektörü Çalışanlarının İşe Yönelik Tutumlarına Etkisi: PTT Örneği. *Optimum Ekonomi ve Yönetim Bilimleri Dergisi*, 8(2), 373-396. <https://doi.org/10.17541/optimum.875531>
- Tretyakov, K. (2004). Machine Learning Techniques in Spam Filtering, *Data Mining Problem-Oriented Seminar, MTAT*, Vol. 3, No. 177, 60-79.
- Uysal, A. K., Günal, S., Ergin, S., & Günal, E. Ş. (2012). Mobil Telefonlarda İstenmeyen SMS Mesajlarının Belirlenmesi, *2012 20th Signal Processing and Communications Applications Conference (SIU)*, Muğla, 1-4.
- Yıldız, O. (2016). Metin Madenciliğinde Anahtar Kelime Seçimi: Bir Üniversite Örneği. *Yönetim Bilişim Sistemleri Dergisi*. 2(1), 29-50.
- Yılmaz, Ö. D. (2014). Tüketici Şikayetlerinin Örgütsel Öğrenme Aracı Olarak Değerlendirilmesi: Konaklama İşletmeleri ve Tur Operatörlerine Yönelik Şikayetler Üzerine Bir Araştırma. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 28(4), 131-148.