ACTA INFOLOGICA 2022;6(2):175-188
dergipark.org.tr/acin

DOI: 10.26650/acin.1057060

RESEARCH ARTICLE

Gümrük Beyannamesi Sürecinde Öğrenmeye Dayalı Algoritmaların Etkinliğinin İncelenmesi

Examining the Efficiency of Learning-Based Algorithms in the Process of Declaring Customs

Mustafa Günerkan[1], **Ender Şahinaslan**[2], **Önder Şahinaslan**[3]

[1](Yüksek Lisans Öğrencisi) Maltepe Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye
[2](Dr.) EA Sağlık, Eğitim ve Bilişim Ltd. Şti, BT Yönetişim Danışmanı, İstanbul, Türkiye
[3](Dr. Öğr. Üyesi) Maltepe Üniversitesi, Bilişim Bölüm Başkanlığı, İstanbul, Türkiye

ORCID: M.G. 0000-0002-4202-2801;
E.Ş. 0000-0001-8519-7612;
Ö.Ş. 0000-0003-2695-5078

Corresponding author:
Önder ŞAHINASLAN
Maltepe Üniversitesi, Bilişim Bölüm Başkanlığı, İstanbul, Türkiye
E-mail address: ondersahinaslan@maltepe.edu.tr

Submitted: 13.01.2022
Revision Requested: 16.06.2022
Last Revision Received: 04.08.2022
Accepted: 02.08.2022
Published Online: 11.08.2022

Citation: Gunerkan, M., Sahinaslan, E., & Sahinaslan, O. (2019). Gümrük beyannamesi sürecinde öğrenmeye dayalı algoritmaların etkinliğinin incelenmesi. *Acta Infologica, 6*(2), 175-188.
https://doi.org/10.26650/acin.1057060

ÖZ

Gümrük işlemlerinde kullanılan beyannamelerin hatasız sunulması kritik önem taşır. Bu beyannamenin oluşturulmasında kullanılan yöntemlerin çeşitliliği, dinamizmi ve karmaşıklığı karşısında insan kaynaklı hatalı beyanname dosyaları üretilmektedir. Bunlar, iş gücü, müşteri ve para kaybı gibi birçok sorunun yanında sözleşme ve yasal uyum gibi hukuki sorunlara da neden olmaktadır. Bu sorunların çözümü için güncel bilgi teknolojileriyle desteklenen akıllı yapılara ihtiyaç duyulmaktadır. Bu amaçla lojistik sektöründe gümrük beyannamesi oluşturma alanında büyük veri üzerinden öğrenme algoritmalarının kullanılabilirliği önemlidir. Bu çalışmada, 4.005.343 beyanname verisi üzerinden gümrük beyannamesi sürecinde öğrenmeye dayalı algoritmaların etkinlik performansları değerlendirilmiştir. Performans ölçüm sonuçlarına göre %25 test oranı ile Train-test split yönteminde Karar Ağacı (%75.69) ve Torbalama (%75.70) algoritmalarında maksimum sonuç ulaşıldı. K değerinin 10 alındığı K-Fold yönteminde ise Karar Ağacı (%75.84) ve Torbalama (%75.83) benzer başarım oranları elde edildi. Bu sonuçlar, makine öğrenmesi algoritmalarının kullanımının bildirim hatalarını tespit etmek için etkili bir yöntem olduğunu ortaya koymuştur. Gümrük beyannamesi süreçlerinin iyileştirilmesine, akıllı kontrol yapılarının geliştirilmesine ve sahada yapılacak yeni çalışmalara kaynak teşkil edecektir.
Anahtar Kelimeler: Gümrük Beyannamesi, Öğrenme Algoritmaları, Lojistik, Büyük Veri

ABSTRACT

Having the declarations used in customs procedures be submitted without errors is critical. In the face of the diversity, dynamism, and complexity of the methods used in creating this declaration, human-induced declaration files are produced erroneously. These cause many problems such as loss of labor, customers, and money, as well as legal problems such as contract and legal compliance. Intelligent structures supported by current information technologies are needed to solve these problems. For this purpose, being able to use learning algorithms over big data is important in the field of customs declaration creation in the logistics industry. This study evaluates the efficiency performances of learning-based algorithms regarding the customs declaration process over 4,005,343 pieces of declaration data. According to the performance measurement results, the maximum result was achieved in the Decision Tree (75.69%) and Bagging (75.70%) algorithms with respect to the Train-test split method at a test rate of 25%. Regarding the K-Fold method, which assumes K to be equal to 10, similar success rates were obtained for the Decision Tree (75.84%) and Bagging (75.83%) algorithms. These results reveal the use of machine learning algorithms to be an effective method for detecting notification errors. This can be a resource for improving customs declaration processes and developing smart control structures, as well as for new studies to be carried out in the field.
Keywords: Customs Declaration, Learning Algorithms, Logistics, Big Data

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

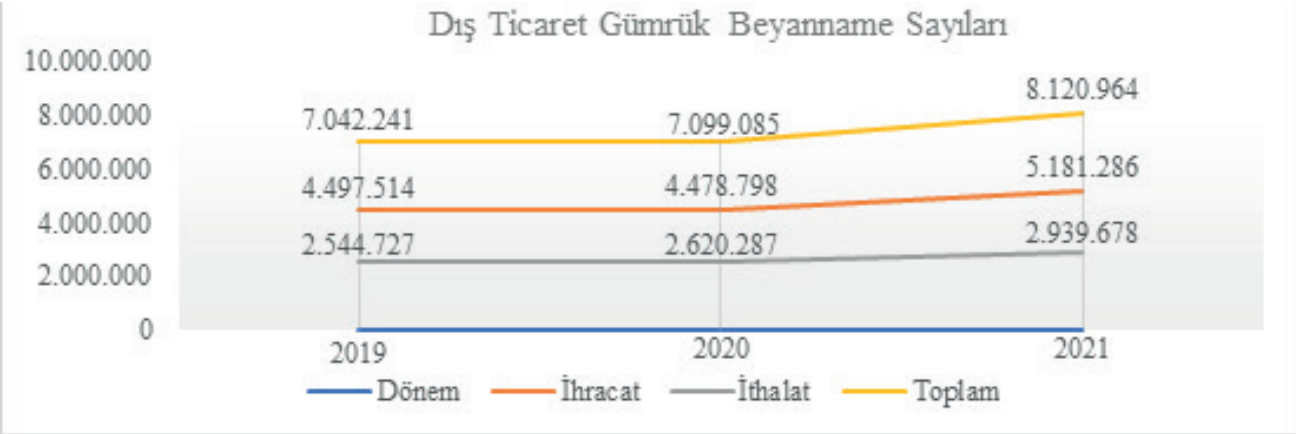
1. GİRİŞ

Gümrük beyannamesi birçok bileşenden oluşan karmaşık bir yapıya sahiptir. Gümrük işlemlerinde kilit role sahip gümrük beyannamesi titizlikle hazırlanması gereken, oluşturulması ciddi anlamda yetkin iş gücü ve zaman gerektiren bir süreçtir. Bu sürecin uygulama yazılım geliştirmesinde teknik bilgi ve beceri yanında ciddi iş süreci ve mevzuat bilgisi de gerekmektedir. Mevzuatlar ise her geçen gün yenilenen ve sürekli takip edilmesi gereken bir özelliktedir. İşin bu tür karmaşıklıkları yanında yetkin ve yeterli insan kaynağının olmaması beyanname yazımında ciddi hataların doğmasına sebep olmaktadır. Bu tür hatalar ise yüklü cezalarla karşı karşıya kalınmasına sebep olmaktadır. Bu durum aynı zamanda operasyon maliyet ve süreçlerini de olumsuz etkilemektedir. Ayrıca yasal ve hukuki bir takım olumsuz sonuçlarda doğurabilmektedir.

Günümüzde nesnelerin interneti(IoT), makineler arası iletişim (M2M), yapay zekâ, kuantum bilgisayar, blok zincir teknolojisi gibi birçok ileri gelişmeler vardır (Sahinaslan & Sahinaslan, 2019). Bilişim teknolojilerinde yaşanan böylesine hızlı gelişmeler yaşamın neredeyse her alanına dokunur olmuştur. Örneğin lojistik sektöründe bu yeni teknolojilerden birisi olan blok zincir teknolojisi değişim etkilerini göstermektedir (Koh, 2020). Dijitalleşme, dijital dönüşümler insan emek, iş gücü, çalışma koşul ve türlerinde de birtakım değişiklikler meydana getirmektedir (Şahinaslan Ö. , 2020). 7*24 açık dijital pazarlara çevik ve hızlı yanıt verebilecek dinamik altyapı ve uygulamalara ihtiyaç vardır (Şahinaslan E. , 2020). Yeni ihtiyaçlara göre şekillenen dijital teknolojiler ve uygulamalar ülkeler arası dış ticaret imkân ve kabiliyetini de geliştirmektedir. Bilgi ve iletişim teknolojisi kullanımı ticaret engellerini azaltmaya ve mesafeleri kısaltmaya yardımcı olarak ticaret verimliliğini ve hacmini de arttırmış, ticaret maliyetlerini ise önemli ölçüde azaltmıştır (Wang & Choi, 2018).

Dış ticaret ile büyüme arasında bağlantılar vardır (Uçan & Koçak, 2014). Dış ticaret büyümenin ana unsurlarından biri olarak değerlendirilmektedir. Ülkeler ürettikleri malların fazlasını ihraç ederek gelir elde edebilmektedir. Ayrıca kendinde mevcut olmayan veya üretemediği malları ithal ederek ihtiyaçlarını karşılayabilmektedir (Saçık, 2009). Dış ticareti her ülke kendi mevzuatlarına göre belli kurallara göre yapmaktadır ve bu kurallar sürekli gelişim halindedir. Ülkemizde 4458 sayılı Gümrük Kanunu ile dış ticaret işlemleri düzenlenmiştir. 4458 nolu Gümrük Kanunu 27.10.1999'da kabul edilmiştir. Bu kanunla ihracat ve ithalat amaçlı Türkiye Cumhuriyeti Gümrük Bölgesine giren ve çıkan her türlü mal, eşya ve taşıt araçlarına uygulanacak gümrük kuralları belirlenmiştir. Türkiye Cumhuriyeti Gümrük Bölgesi, Türkiye kara suları, iç suları ve hava sahası dahil Türkiye Cumhuriyeti topraklarını kapsar(Mevzuat Bilgi Sistemi, 2021).

Dış ticaret ithalat ve ihracat işlemlerini kapsar. Ülkemizde üretilen ürünlerin diğer ülkelere pazarlanması ve satılması ayrıca iç piyasanın ihtiyacı olan ürünlerin temininin yapılması için ihracat ve ithalat yapan firmaların varlığı önemlidir. (T.C. Ticaret Bakanlığı, 2021) verilerine göre 87.741 adet ihracat, 75.897 adet ithalat, 35.310 adet hem ihracat hem de ithalat yapan firma bulunmaktadır. İthalat ve ihracat süreçleri birbirini izleyen işlemlerin bir bütünüdür. Dış ticareti yapmak için gerekli şartların bilinmesi ve bunların eksiksiz olarak yapılması şarttır. Bu sürecin performanslı bir şekilde yürütülebilmesi ciddi anlamda bilgi ve tecrübeye ihtiyaç duymaktadır. Aynı zamanda ülkemizde dış ticaret işlem hacmi yıldan yıla artmaktadır. 10.12.2021 tarihli son 3 yıla ait gümrük beyanname verileri Şekil-1'de yer almaktadır (T.C. Ticaret Bakanlığı, 2021).



Şekil 1. (T.C. Ticaret Bakanlığı, 2021) tarafından yayınlanan verilerden yazarlar tarafından oluşturulmuştur.

Bu verilere göre Ocak-Kasım 2021 tarihleri arasında 5.181.4286 adet ihracat, 2.939.678 adet ithalat olmak üzere toplam 8.120.964 adet beyanname işlemi gerçekleştirilmiştir. Bu kadar yoğun gerçekleşen işlemlerin gerçekleştirilmesinde insan kaynağından ya da zamanla yetersiz kalan uygulama ve güncel teknolojilerden yeterince yararlanılmadığı durumlarda bir takım hatalarla karşılaşılmaktadır. Bu durum ise süreçten beklenen hız, kalite, performans beklentilerini karşılayamamakta ve hatalı bildirimlere sebebiyet vermektedir. Bu hatalı bildirimler sonucunda işletmeler yüksek tutarlı cezalara maruz kalmakta, müşteri, itibar ve zaman kaybı gibi telafisi güç durumlarla karşılaşmaktadır. Diğer taraftan gümrük işlemleri dış ticaretin önemli bir aşamasıdır. İşlemlerin yapılabilmesi için belli bilgileri içeren gümrük beyannamelerinin ilgili gümrük müdürlüğüne sunulması gerekmektedir. Teknolojik gelişmeler bu süreçlerde ciddi iyileşmeler sağlamıştır. İşlem sayılarının çok fazla olduğu günümüzde mevcut entegrasyon işlemleri ile ciddi sayıda beyanname işlemi kısa sürelerde sonuçlandırılabilmektedir (Güldüren & Öztop, 2020).

Gümrük işlemlerinde kusursuz beyanname yazımı kritik öneme sahiptir. Eksik ya da hatalı oluşturularak gümrüğe iletilen beyannameler çok ciddi sorunlara sebep olduğu bilinmektedir. Bu hataları minimum seviyeye çekilmesinde bilişim teknolojilerinden büyük veri ve öğrenme algoritmaları kullanılarak geliştirilecek akıllı bir kontrolle beyanname hataların büyük oranda tespiti mümkündür. Böylece hatalı bildirimden kaynaklı olumsuz sonuçlarının doğmadan önlenmesine, operasyonel süreçlerin otomatize edilmesine, insan müdahalesinin asgari seviyeye çekilmesine ve beyannamelerin doğru oluşturulmasına katkı sunacaktır. Bu çalışma güncel bilişim teknolojilerinin sunduğu imkânlarla makine öğrenmesinin farklı alanlardaki başarılı uygulamalarından yola çıkarak hatasız beyanname oluşturulmasına destek olacak öğrenme tabanlı algoritmaların etkinlik başarım düzeyleri üzerinde gerçekleştirilmiştir.

2. LİTERATÜR TARAMA

Makine öğrenmesi algoritmaları Kotsiantis ve ark. (2007) bugüne kadar birçok alanda kullanılmıştır ve başarılı sonuçlar elde edilmiştir. Gümrük beyanname sürecinde başarılı bir şekilde kullanılabileceği varsayımından yola çıkarak bu çalışmaya karar verilmiştir. Bu amaçla bu alanda yapılan ilgili çalışmaların araştırması yapılmıştır. Önerdiğimiz çözümle ilgili temel akademik çalışmalara bu bölümde yer verilmiştir.

Canrakerta ve ark. (2020) gümrük beyannamesinin potansiyel hatalara sahip olabileceği ve bu hataların kasıtlı da yapılmış olabileceğinden yola çıkarak iş zekâsı ürünleriyle analiz yaparak olası sorunları ortaya çıkartmaya çalışmışlardır. Çalışmada, Kimball metodolojisi ile veri ambarı, OLAP ve veri madenciliği kullanılarak hata veya sahtekârlığın tespit edilmesi amaçlanmıştır. Karar ağaçları, destek vektör makineleri, sinirsel ağ ve çeşitli topluluk yöntemleri gibi veri madenciliği algoritmaları kullanmışlardır. Çalışma sonucunda geliştirilen SMOTE tekniğinin anlamlı derecede bir duyarlılık skoruna sahip olduğunu tespit etmişlerdir. Shao ve ark. (2002) bir veri madenciliği uygulaması ile gümrük beyannamesi verilerinde dolandırıcılık davranışını tespit etmeye çalışmıştır. Genişletilmesi kolay çok boyutlu bir veri modeli için uygun veri madenciliği teknolojisi kullanarak hibrit bir dolandırıcılık tespit stratejisi uygulamışlardır. Dolandırıcılık tespit uygulamalarında veri dağılımının özelliklerinden dolayı dolandırıcılık davranışını tahmin etmenin zor olduğunu belirtmişlerdir. Bunun yanında çalışmada açıkladıkları çok boyutlu kritere sahip genişletilmesi kolay veri modelinin hem modelin doğruluğunu hem de algoritmanın performansını iyileştirdiğini belirtmişlerdir. Model güçlü bir popülerleştirme yeteneğine sahip olduğundan, diğer benzer karmaşık uygulamalara referans olarak kullanılabileceği sonucuna varmışlardır.

Maruev ve ark. (2014) giderek küreselleşen dünyamızda, uluslararası ticaretin önündeki engellerin incelenmesinin uluslararası ekonomi alanını ilgilendirdiğini belirtmişlerdir. Çalışmalarında gümrük beyannamelerinin hızlı ve doğru bir şekilde işlenmesi sorununa odaklanmışlardır. Denetimli öğrenmeye dayalı ticari mallar için müşteri beyanlarının otomatik olarak işlenmesi için grafik tabanlı yayma etkinleştirme algoritmasının yeni bir kullanımını sunmuşlardır. Sundukları yöntem ile gümrük memurları, tüccarlar, taşıyıcılar ve sigortacılar tarafından kullanılmak üzere tavsiye sistemleri oluşturmuşlardır. Her zamanki risk temelli yaklaşımın aksine, bu algoritma geleneksel risk göstergelerinden ziyade yalnızca sevkiyat verileri üzerine eğitilmiştir. Bunun, gümrük yetkililerine tavsiyenin bir gönderinin içeriği açısından açıklanabilmesi ve gerçek zamanlı olarak doğrulanabilmesi açısından yararlı olduğunu belirtmişlerdir. Yaklaşımın fizibilitesini Rusya Federasyonu ile iki AB ülkesi arasındaki sekiz sınır kontrol noktasında bir ay boyunca kesintisiz olarak toplanan 2500 gümrük kaydına başvuru ile test etmişlerdir. Algoritma ile deneysel koşullar altında %100 doğruluk elde etmişlerdir.

Özer (2020) ülkemizde dış ticaret işlemlerinin sürelerini incelemiş ve bu süreleri gelişmiş ülkelerle karşılaştırmıştır. İşlemlerin daha hızlı ve tek bir yerden yapıldığı Tek Pencere Sistemi hakkında bilgi vererek buradaki sürecin nasıl işlediği ve ne gibi katkılar verdiğine değinmiştir. İlgün (2020) vergi denetiminde büyük veri analitiğinin olumlu ve olumsuz yanlarını araştırmıştır. Büyük veri analitiğinin kullanılmasının bir tercih değil de kaçınılmaz bir sonuç olacağını vurgulamıştır. Gerekli yapısal dönüşüm ve altyapı çalışmaları yeni teknolojilere uyum ve sonuçları konusundan bahsetmiştir.

García ve Caballero (2021) çoğu ülkeler için önemli konulardan biri olan gümrük dolandırıcılığına ilişkin kontrol sisteminin optimizasyonunun ekonomik açıdan önemline dair bir problemi incelemişlerdir. Makine öğrenimini ve çok amaçlı doğrusal programlamayı birleştiren Bayes tabanlı yeni bir hibrit yaklaşım önermişlerdir. Mevcut denetim sistemlerinin hassasiyetini iki katından fazlasına çıkarmanın (%237'lik bir artışla) mümkün olduğunu, insan kaynaklarının neredeyse %50'sini serbest bırakmanın ve ayrıca geçmiş sonuçların üzerinde performans göstermenin mümkün olduğunu göstermişlerdir. Paula ve ark. (2016) Brezilyalı ihracatçıların dolandırıcılık yapma eğilimlerini sınıflandırmada denetimsiz derin öğrenme modelinden elde edilen sonuçları sunmuşlardır. İhracatçıların büyük çoğunluğunun, standart bir şekilde birbiriyle ilişkili olan ihracat hacminin açıklayıcı özelliklerine sahip olduğunu varsayarak, veri modeliyle ilgili anormal durumları tespit etmek için 'AutoEncoder' metodunu kullanmışlardır. Çalışmalarını Brezilya Federal Gelir Sekreterliği tarafından sağlanan 2014 yılında Brezilya'da gerçekleşen mal ve ürün ihracat verilerine göre yapmışlardır. İhracat şirketlerini karakterize eden niteliklerden kurdukları model ile en az yirmi ihracatçıdaki anormallikleri tespit edebilmişlerdir.

Li ve Li (2019) hızla artan uluslararası ticarete konu malların sınıflandırılmasının zorluğunu belirtmişlerdir. Gelişmiş makine öğrenimi tekniklerinin malları verimli bir şekilde sınıflandırması için bir fırsat sağladığını söyleyerek gümrük sınıflandırma sürecini kolaylaştırmak için metin-görüntü uyarlamalı bir sinir ağı önermişlerdir. Önerdikleri model, biri metin, diğeri görüntü için olmak üzere iki bağımsız alt model içermektedir. Alt modeller, model eğitim sonucuna göre parametrelerin değerini ayarlayabilen yeni bir yöntemle birleştirmişlerdir. Son olarak, bir grup gümrük tarife koduna ve bir e-ticaret sitesinden alınan bir veri setine dayalı bir vaka çalışması ve karşılaştırma deneyleri yapmışlardır. Deney sonuçlarına göre modellerinin gümrük sınıflandırılmasında başarılı bir şekilde uygulandığını görmüşlerdir. Ryzhova ve Sochenkov (2019) metinsel açıklamalarına dayalı olarak malların gümrük sınıflandırmasına odaklanmışlardır. Kopyalı ve kopyasız olmak üzere iki tür veri kümesi kullanarak farklı makine ve derin öğrenme modelleri önermişlerdir. Lojistik regresyon (Sperandei, 2014) ile başarılı sonuca ulaşmışlardır. Barua ve ark. (2020) uluslararası yük taşımacılığında makine öğrenmesi modelleri geliştirerek talep tahmini, operasyon ve varlık bakımı, araç güzergâhı ve zamanında teslimat performansı tahmini konularında kullanımını tartışmaktadırlar.

Mammadov (2020) gümrük tarafında kullanılan teknolojileri incelemiştir. Makine öğrenmesi de bunlardan biridir. X-ray'in daha iyi uygulanması için bilgisayar tabanlı eğitim yazılımı geliştirerek denetleme mekanizmalarının simulasyonlarını gerçekleştirmiştir. Makine öğrenmesi ve derin öğrenme algoritmalarıyla sağlık alanında da ciddi çalışmalar yapılmış ve hastalıkların teşhisi konusunda başarılı sonuçlar elde edilmiştir (Pamuk & Kaya, 2021; Akgül ve ark., 2020; Saygın & Baykara, 2021; Gulia ve ark., 2014; Coşar & Deniz, 2021; Başer ve ark., 2021).

Pazarlama alanında yapılan çalışmalar incelendiğinde; tüketicilerin alış-veriş alışkanlıklarının belirlemesi ve pazarlama davranışlarına ilişkin yeni öngörüler kazandırmada Sundsoy ve ark. (2014) karar verme sürecine ışık tutmada Cui ve ark. (2006), insan iç görüleriyle pazarlama teorilerini ilişkilendirmede Ma & Sun (2020), pazarlama alanının gelişen doğasını analiz etmede Siau & Yang (2017) ve pazarlama operasyonlarının performansını iyileştirmede Brei (2020) ve Cui & Curry (2005) akademik çalışmaları gerçekleştirilmiştir.

Modern veri girişi teknolojileri doğrudan veri girişi programlarına uygulanarak giriş hataları büyük ölçüde azaltılmıştır (Mullooly, 1990). Rastgele seçilen formların doğrulanması ile kritik alanların %100 doğrulanması, tüm alanların %100 doğrulanmasına uygun maliyetli bir alternatif sağlamış ve %100 doğrulamayı gereksiz kılmıştır (Mullooly, 1990). Veri giriş mekanizması, kullanıcı etkileşimi için ilk ara yüz olduğu için insan-bilgisayar etkileşimi alanında yaygınlaşmaktadır (Salve ve ark., 2021). İnsanların veri girişleri hatalı olabilir ve bu hatalı girişler olumsuz sonuçlara sebep olabilir (Barchard & Pace, 2011). Veri kalitesi, modern veri tabanlarında kritik bir sorundur. Veri giriş hatalarını tespit etmek, azaltmak ve veri kalitesini

iyileştirmek için otomatik yöntemler konusunda çok az araştırma yapılmıştır (Chen ve ark., 2011). Bu nedenlerle araştırmacılar veri girişlerini en aza indirecek çalışmaları yaparak bu alanda yeni saha uygulamalarını geliştirmelidir (Barchard & Pace, 2011).

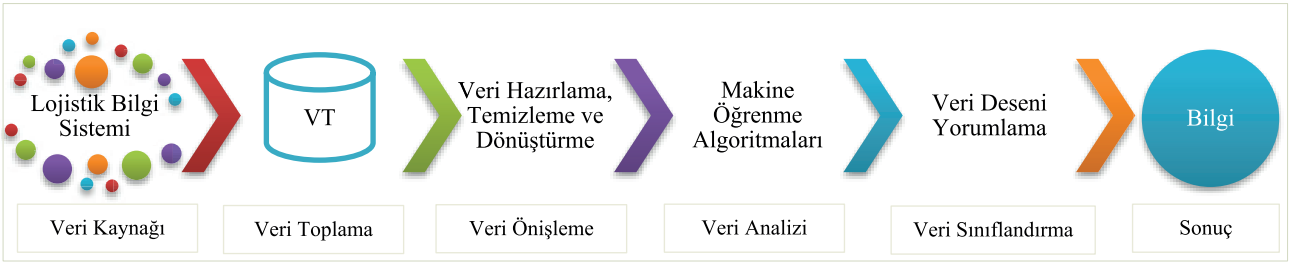
Makine öğrenmesi algoritmaları banka, sigorta, sağlık, eğitim, güvenlik, telekomünikasyon, havacılık ve uzay, savunma gibi pek çok sektörlerde farklı alanda kullanılmakta olup, yeni kullanım alanlarına ilişkin araştırmalarsa artan bir hızda devam etmektedir. Tüm bu alanda yapılan başarılı çalışmalar göz önüne alındığında lojistik sektöründe de kullanım alanlarının araştırılması sektöre büyük katkı sağlayacaktır. Örneğin gümrük beyannamesinde yaşanan sorunların en asgariye indirgenmesinde ve hatasız beyanname yazımına destek olacak akıllı bir sistemin kurulmasında öğrenme algoritmalarının etkinliğinin test edilmesine ihtiyaç vardır. Bu amacı karşılamak için bu çalışmada öncelikle gümrük beyanname sürecinde öğrenmeye dayalı algoritmaların etkinliğinin araştırılması hedeflenmiştir.

3. MATERYAL VE METOT

Gümrük beyannamesi yüzlerce bilgi alanından oluşmaktadır (Gümrük Rehberi, 2021) ve (İhracatta Kullanılan Uluslararası Dökümanlar, 2021). Her bir bilgi alanının tek tek doğru bir şekilde doldurulması gerekmektedir. Bu bağlamda geliştirdiğimiz modelde her bir bilgi alanı tahmininde her bir veri alanı için bir veri seti oluşturulmuştur. Bunların özniteliklerini (Dash & Liu, 1997) belirlemede Delphi metodu kullanılmıştır (Ameyaw ve ark., 2016), (Zartha ve ark., 2019), (Lund, 2020) ve (Sourani & Sohail, 2014). Delphi yönteminde yapılmak istenen geleceğe yönelik tahmin yapmak, bunun için uzmanlardan yararlanmak ve bir uzlaşma veya karara varmaktır. Burada da gümrük mevzuatı bilgisi olan kişilere veri/bilgi alanı tahmininde kullanılacak özniteliklerin tespiti için sorular sorularak modeller belirlenmiştir. Modellerin ilk eğitim sonuçları değerlendirilmesinde özniteliklerin üzerinden geçilerek değişiklikler yapılmıştır. Bu süreç kabul edilebilir başarım oranlarına ulaşıncaya kadar tekrarlanarak modellere son hali verilmiştir.

3.1. Süreç Akışı

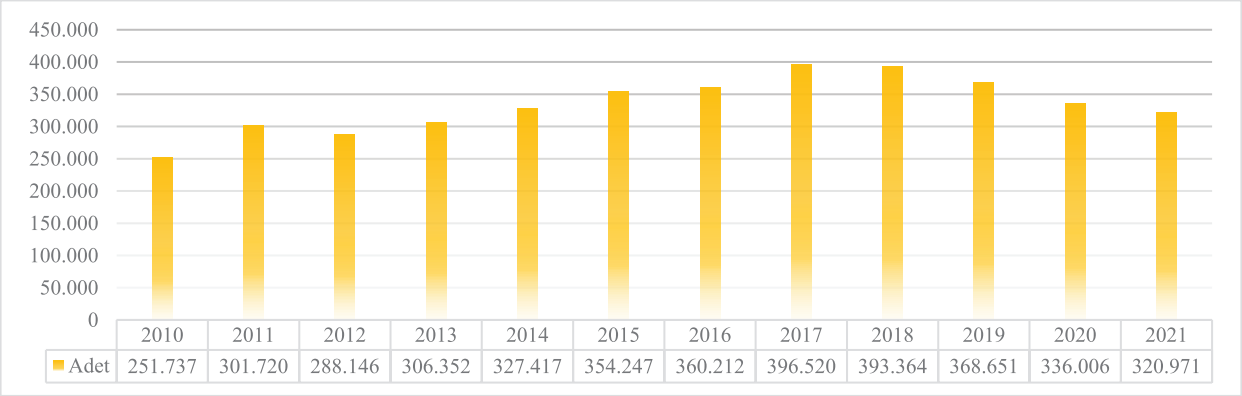
Bu çalışma da uygulanan yönteme ilişkin ana süreç aşamaları Şekil 2'de gösterilmektedir. Lojistik bilgi sistemi kaynaklarından beyanname ve teslim şekli verisine ait ham veri kaynağı ve alanlarının belirlenmesi, seçimi, toplaması, verinin hazırlaması, veri analizi, veri sınıflandırması ve sonuçların değerlendirme aşamalarından oluşmaktadır.



Şekil 2. Çalışma Ana Süreç Aşamaları

3.2. Veri Toplama

Sürecin bu aşamasında lojistik bilgi sisteminden çalışmada kullanılacak anlamlı sayılan gümrük beyanname verilerinin seçimi yapılarak toplanmıştır. Çalışmada, kullanılacak veri setinin oluşturulmasında ülkemizde büyük işlem hacmine sahip öncü bir uluslararası lojistik firmanın 2010 yılından sonra üretilen toplam 4.498.069 adet beyanname verileri belirlenmiştir. Bu veriler, teslim şekli bilgisi tahmininde kullanılmak üzere MS SQL veri tabanı üzerinde oluşturulan bir tablo üzerinde toplanmıştır. Çalışmada kullanılan beyanname verilerinin yıl bazında adetleri ve buna ilişkin grafik Şekil 3'te verilmiştir.



	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Adet	251.737	301.720	288.146	306.352	327.417	354.247	360.212	396.520	393.364	368.651	336.006	320.971

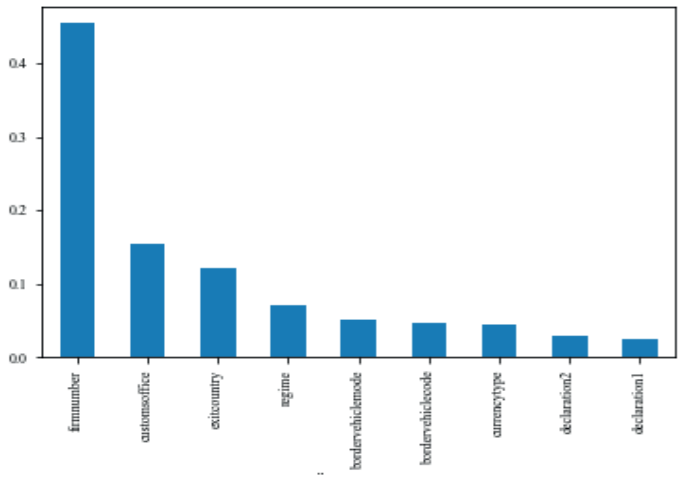
Şekil 3. 2010-2021 yılları arasında yıllık bazında beyanname sayıları

Son 12 yılın beyanname verileri incelendiğinde 2010 yılında en az 251.737 adet ve 2017 yılında ise en fazla 396.520 adet beyanname verisinin olduğu gözlemlenmektedir. Grafikte yer alan beyanname sayılarında 2018 yılına kadar artış yaşanırken 2019 yılından itibaren kısmi azalış vardır. Bu azalışın COVID-19 pandemi sonrasına denk gelmesi dikkat çekici bulunmuştur.

Çalışma için oluşturulan veri tabanı tablosu üzerine toplanan beyanname verilerine ait her yıldan bir adet ve rasgele seçilen (n=24) adet örnek veri seti bilgisi Tablo-1'de yer almaktadır. Bunlar anahtar alanlar, öznitelik belirten alanlar ve çıktı alanı olarak üç bölüm olarak ele alınıp değerlendirilebilir.

Öznitelik veri setinin belirlenmesinde uzman görüşlerine dayanan Delphi yöntemi kullanılmıştır. Tahminlemede gümrük beyannamesi veri alanını doğrudan etkileyecek alanların belirlenmesinde mevzuat ekibinden destek alınmıştır. Seçilen alanlar üzerinde ayrı ayrı çalışılarak elde edilen sonuçlar yine bu uzmanlarla birlikte değerlendirilmiştir. Değerlendirmeler sonucunda varsa iyileştirmeler mevzuat ekibiyle birlikte çalışma yeniden tekrarlanarak nihai modele karar verilmiştir. Gümrük beyannamesi üzerinden belirlenen öznitelikler; firma numarası (firmnumber), beyannamenin işleme alındığı gümrük ofisi(customsoffice), beyannamesinde işlemine ait rejim bilgisi(regime), rejim detay bilgisi(declaration1), rejim detay bilgisine ait ek bilgi(declaration2), işlem döviz türü(currencytype), çıkış ülke bilgisi(exitcountry), taşıma nakliye tür bilgisi(bordervehiclemode) ve nakliye tür kodu (bordervehiclecode) alanlarından oluşmaktadır. Çıktı bilgisi olarak teslim şekli kodu(incotermcode) veri alanından oluşmaktadır.

Modeli oluşturan özniteliklerin ağırlıklarının tespitine yönelik Python uygulaması üzerinden görselleştirme kodu yazılmıştır. Kullanılan özniteliklerin ağırlıklarına ilişkin elde edilen grafik Şekil 4'de gösterilmektedir. Elde edilen sonuçlara göre lojistik firmanın işlem yaptığı 9.350 farklı firmaya verilen firma numarası en etkili öznitelik olarak bulunmuştur.



Şekil 4. Öznitelik ağırlıkları

Tablo 1

Teslim Şekli Veri Tablosu Örnek Veri Seti

Örnek No	Anahtar Alanlar				Öznitelik Alanları								Çıktı
	Year	Depart-ment	File-number	Firm-number	Customs-office	Regime	Declar-ation1	Declar-ation2	Currency-type	Exit-country	B.vehicle-mode	B.vehicle-code	Incoterm-code
1	2010	AVR	2226	1	341300	1040	EU	1	USD	52	TIR	30	FCA
2	2010	GMR	23198	5	340300	4000	EU	4	EUR	3	UCAK	40	FOB
3	2011	AND	25	29724	341200	4071	EU	4	EUR	4	TIR	30	CIP
4	2011	TRK	4335	132	343100	4000	IM	4	EUR	728	GEMI	10	CIF
5	2012	GMR	16120	5	340300	4000	EU	4	EUR	4	UCAK	40	CPT
6	2012	AND	26245	8	341200	7100	ANT	7	TL	1	TIR	30	FCA
7	2013	EGE	920	14	340300	4000	EU	4	EUR	4	UCAK	40	DAP
8	2013	MRS	16677	10239	60600	1000	EX	1	USD	52	KAMYON	30	EXW
9	2014	GMR	2255	5	340300	3153	EU	3	EUR	52	UCAK	40	CPT
10	2014	EGE	210	101	351900	3151	EU	3	EUR	52	GEMI	10	EXW
11	2015	GEB	36647	60068	410500	4071	EU	4	EUR	5	TIR	30	DAP
12	2015	GEB	25548	47939	341200	1000	EU	1	EUR	52	TIR	30	DAP
13	2016	MRS	8047	10149	341300	3151	EU	3	EUR	52	KAMYON	30	CPT
14	2016	AND	7759	62093	330100	4071	IM	4	USD	616	KAMYON	30	CFR
15	2017	GEB	3804	10172	410300	4071	EU	4	GBP	6	GEMI	10	DAP
16	2017	TRK	31346	2308	343100	5100	IM	5	USD	664	GEMI	10	FOB
17	2018	ISN	31430	45616	330100	3151	EU	3	USD	52	GEMI	10	FOB
18	2018	GEB	1002	62255	340300	4000	EU	4	EUR	17	UCAK	40	EXW
19	2019	ANK	87	50917	341200	4010	EU	4	EUR	4	KAMYON	30	EXW
20	2019	EGE	30921	60522	350300	4010	EU	4	EUR	11	TIR	17	EXW
21	2020	GEB	26883	10582	341200	4071	EU	4	EUR	63	TIR	30	FCA
22	2020	GMR	2159	5	160200	1000	EU	1	EUR	52	KAMYON	30	FCA
23	2021	EGE	20268	60522	352200	5100	IM	5	EUR	52	TIR	30	EXW
24	2021	EGE	271	14	351700	3151	EU	3	TL	52	TIR	30	DAP

Bu anahtar veri alanları; yıl(year), bölüm bilgisi(department) ve beyanname belge numarası(filenumber) alanından oluşmaktadır.

3.3. Veri Önişleme

Verilerin önişleme süreç aşaması modelin tahminlemesinde önemli bir yer tutmaktadır (Alexandropoulos ve ark., 2019). Verinin entegrasyonu, temizlenmesi, dönüştürülmesi, azaltılması ve boyutunun küçültülmesi bu aşamada yapılmaktadır (Alasadi & Bhaya, 2017). Bu aşamada toplanan veriler üzerinde belirli veri kontrolleri gerçekleştirilmiştir. Çalışmaya hiçbir katkısı olmayacak, hatta veri analiz ve değerlendirmesinde olumsuz yansıması olabilecek öznitelik içeriği boş ('NULL') , ya da sembol (@, #, %, &, *) gibi işaretleri barındıran, çalışma amaç ve kapsamına girmeyen veriler tespit edilerek ilgili veri temizliği yapılmıştır. Bu işlem sonucunda verilerden 492.726 adeti çalışma kapsamından çıkartılmıştır.

Kategorik verilerin çoğunluğunun sayısal olmadığı durumda 'encoding' yöntemi kullanılarak veriler sayısallaştırılıp kullanılır (Dahouda & Joe, 2021). Bu nedenle çalışmada kullanılan beyanname teslim şekli verilerin sayısallaştırılması sağlanmıştır. Sayısallaştırma işleminde; 'Label Encoder' metodu kullanılmıştır (Jackson & Agrawal, 2019). Çalışmada ele alınan verilerin 'encoding' yöntemi uygulanarak sayısallaştırılması sağlanmıştır. Bu verilerden biri olan teslim şekli veri setinin sayısallaştırılma sonrası oluşan yeni veri setine ilişkin örnek veriler (n=9) Tablo 2'de gösterilmektedir.

Tablo 2

Sayısallaştırma Sonrası Teslim Şekli Veri Seti Örneği

index	firm number	customs office	regime	declaration1	declaration2	currency type	exit country	bordervehicle mode	bordervehicle code	incoterm code
0	1	64	4	2	0	20	29	7	7	FCA
1	1	64	4	2	0	20	29	7	7	FCA
2	1	64	4	2	0	20	29	7	7	FOB
3	1	64	4	2	0	20	29	7	7	FOB
4	1	64	4	2	0	20	29	7	7	FCA
5	1	64	4	2	0	20	29	7	7	FCA

6	1	59	8	2	1	20	29	14	8	FOB
7	1	59	24	4	3	20	126	14	8	DDU
8	1	59	24	2	3	20	9	14	8	DDP
9	1	64	24	2	3	20	2	7	7	DDU

3.4. Veri Analizi

Çalışmada beyanname verisi teslim şekli üzerinden veri analiz ve tahmininde bulunmak amacıyla başarılı bulunan makine öğrenme yaklaşımlarından yararlanılmıştır. Makine öğreniminde Karar Ağaçları (Decision Tree), veri seti değişkenlerini bir ağaç modelinde kullanarak istenilen değerin tahmin edilmesinde kullanılır. Daha isabetli tahmin sonuçlarının elde edilmesinde karar ağaçları temelinde geliştirilen Torbalama(Bagging), Rastgele Orman (Random Forest) ve Artırma (Boosting) türünde yaklaşımlarda vardır. Veri analiz ve sınıflandırma algoritmalarının seçiminde bu alanda gerçekleştirilen ve başarım düzeyi yüksek bulunan; Chen ve ark. (2019) Naive Bayes, Chandrasekar ve ark. (2017) Decision Tree Classification, Zhang ve ark. (2018) Nearest Neighbors Classification, Abdulkareem & Abdulazeez (2021) Random Forest Classification, Dobriban & Wager (2018) Ridge Classification, Roshan & Asadi (2020), Zhang ve ark. (2019) Bagging Classification AdaBoost Classification ve Tang ve ark. (2016) Multi-Layer Perceptron çalışmaları da dikkate alınmıştır. Yine eğitim verileri temelinde yeni gözlemlerin kategorisini belirlemede kullanılan sınıflandırma algoritmalarının veri setine uygunluğu da göz önünde tutulan diğer bir husus olmuştur.

3.4. Veri Sınıflandırma

Veri bilimi ve makine öğrenimi için en yaygın kullanılan Python programlama diline ait 'scikit-learn' kütüphanesi 'spyder editörü' ile kullanılmaktadır. Bu pakette modellerin eğitimi için standart test verisi oranı %25 olarak alınır (Paper, 2020). Bu çalışmada kullanılan modellerin eğitiminde test verisi oranı %25 olarak dikkate alınarak eğitimler gerçekleştirilmiştir. Öncesinde farklı algoritma da denenmiş ancak özellikle regresyon algoritmalarındaki Huang ve ark. (2019) olumsuz sonuçlardan dolayı diğer algoritmalar tercih edilmemiştir. Sınıflandırmada Naive Bayes, Decision Tree, K-Nearest Neighbors, Random Forest, Ridge Classifier, Bagging Classifier, Ada Boost Classifier, MLP Classifier öğrenme algoritmaları kullanılmıştır.

4. BULGULAR

Bu çalışmada, 2010-2021 yılları arasında oluşan 4.498.069 adet veriden çalışmaya elverişli olmayan 492.726 adet (%10.95) veri çalışma kapsamından çıkartılmıştır. Öğrenme algoritmalarının etkinliğini ölçmede toplam 4.005.343 adet beyanname teslim şekli verisi kullanıldı. Veri eğitim ve sınıflandırmasında daha önce çalışmalarda kullanılan ve başarılı bulunan makine öğrenme algoritmaları test edildi. Veri setinin eğitiminde Train-test split yöntemi kullanımı ile elde edilen sınıflandırma algoritma başarım oranlarına ait sonuç değerleri Tablo 3'de gösterilmektedir.

Tablo 3

Trian-Test Split Yönteminde Teslim Şekli Veri Seti Sınıflandırma Algoritma Başarım Oranları

Algoritma	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**Tüm**
Naive Bayes - NB	**42.10**	35.29	30.87	31.76	32.65	31.05	31.70	35.36	35.73	36.85	38.04	38.28	30.58
Decision Tree Classifier-CART	80.87	79.20	77.42	77.45	78.96	79.03	78.03	79.31	79.92	80.28	82.86	**83.19**	**75.69**
K-Nearest Neighbors - KNN	78.81	76.29	74.54	73.60	74.92	76.26	75.03	76.46	77.69	77.71	**80.22**	80.00	72.21
Random Forest - RFC	80.76	78.97	77.40	77.28	78.72	78.87	77.79	79.21	79.70	80.08	82.60	**82.91**	75.51
Ridge Classifier - RC	**43.46**	38.82	37.46	36.14	36.98	35.29	32.65	34.04	33.88	35.09	37.07	36.76	32.14
Bagging Classifier - BC	80.93	79.26	77.85	77.47	79.08	79.11	78.08	79.42	79.94	80.32	82.92	**83.13**	**75.70**
Ada Boosting Classifier - ABC	13.17	7.68	6.93	39.40	18.13	17.19	29.60	21.18	13.75	9.27	**25.08**	5.40	26.59
MLP Classifier - MLPC	**55.00**	51.26	49.68	44.47	45.51	45.08	40.73	41.24	42.29	43.12	41.38	44.71	39.17

Tablo 3'de 'Tüm' sütunundaki sonuçlar çalışmada kullanılan yıl bağımsız bütün veriler üzerinden elde edilen sonuçları göstermektedir. Bu sonuçlara göre Naive Bayes(%30.58), Decision Tree Classification(%75.69), Nearest Neighbors Classification(%72.21), Random Forest Classification (%75.51), Ridge Classification (%32.14), Bagging Classification (%75.70), Ada Boosting Classification algoritmasında(%26.59) ve Multi-Layer Perceptron (%39.17) başarım sonuçları elde edilmiştir. Sınıflandırma algoritmalarında Decision Tree Classification, Nearest Neighbors Classification, Random Forest Classification ve Bagging Classification algoritmaları diğerlerine göre daha başarılı sonuç üretmiştir. En başarılı orana ise

%75.70 ile Decision Tree Classification ve Bagging Classification algoritmalarında elde edilmiştir. Diğer taraftan gümrük operasyonlarında mevzuat ve uygulamalarda zamanla değişiklikler, ciddi işleyiş farklılıkları olabilmektedir. Bu farklılıkların sonuçlar üzerinde bir etkisinin olup olmadığını araştırmak için çalışma yıllar bazında da gerçekleştirilmiştir. Bu sonuçlar incelendiğinde Ada Boosting Classifier algoritması dışında diğer algoritmalarda yıllar bazında ciddi farklılıklar göstermediği, tüm veriler üzerinden elde edilen sonuçlara göre genelde daha başarılı sonuçlar elde edildiği gözlemlenmiştir. Ada Boost algoritmasının performans sağlayabilmesi için yeterli miktarda veriye ihtiyacı vardır ve daha az veride başarım oranı düşük çıkmaktadır (Schapire, 1999). Veriler yıllar bazında incelendiğinde 2021 yılında Decision Tree algoritması %83.11'lik bir oranla en iyi başarım oranına sahiptir.

Train-test split yönteminde eğitim için %75 ve test için %25 veri ayrılmaktadır. Bu yöntemde veri parçalanırken verinin dağılımına göre modelin eğitim ve testinde sapma ya da aşırı öğrenme problemi ortaya çıkabilmektedir (Refaeilzadeh ve ark., 2016). Kullandığımız veri seti eğitiminde bu tür bir problemin var olup olmadığını anlamak için ayrıca K-Fold yöntemi ile de çalışıldı. K-Fold yönteminde veri K adet alt kümeye bölünerek bir alt küme test verisi olarak alınır ve K-1 adet alt kümenin eğitiminde kullanılır. Bu eğitim sonuçlarının ortalaması alınarak algoritma başarım oranı tespit edilir. Çalışmamızda K-Fold yönteminde K değeri genel olarak en performanslı çıkan 10 değeri kullanılmıştır. Veri setinin eğitiminde K-Fold yöntemi kullanımı ile elde edilen sınıflandırma algoritmalarına ait başarım oranları Tablo 4'de gösterilmektedir.

Tablo 4

K-Fold Yönteminde Teslim Şekli Veri Seti Sınıflandırma Algoritma Başarım Oranları

Algoritma	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**Tüm**
Naive Bayes - NB	**40.56**	33.79	34.22	31.35	33.36	32.75	29.77	33.33	33.57	32.83	33.64	32.90	30.33
Decision Tree Classifier-CART	81.16	79.24	77.86	77.92	79.33	79.23	78.14	79.64	80.14	80.73	82.78	**83.31**	**75.84**
K-Nearest Neighbors - KNN	78.85	76.32	74.43	73.95	75.82	75.92	74.88	76.06	76.98	77.61	80.35	**80.92**	71.63
Random Forest - RFC	81.13	79.10	77.58	77.59	79.04	79.05	77.86	79.40	79.87	80.51	82.56	**83.11**	75.63
Ridge Classifier - RC	42.61	37.53	36.08	34.80	34.51	34.27	32.88	34.15	33.51	34.47	34.73	34.58	32.47
Bagging Classifier - BC	81.19	79.30	77.93	77.93	79.38	79.29	78.20	79.71	80.19	80.77	82.77	**83.34**	**75.83**
Ada Boosting Classifier - ABC	**37.90**	20.42	9.39	14.23	17.27	16.29	32.51	25.83	11.85	17.56	34.07	11.82	34.56
MLP Classifier - MLPC	**54.96**	49.31	46.75	43.87	44.08	43.82	40.51	39.46	39.18	40.45	38.82	38.70	39.14

Çalışılan her iki yöntemle elde edilen sonuçlar tüm algoritmalarda benzer hatta birebir sayılan başarım değeri üretmiştir.

Veri seti eğitiminde 'Train-test split' ve 'K-Fold' olmak üzere iki farklı yaklaşım test edilmiştir. Her iki yaklaşımın çalışılan algoritmalar üzerinden elde edilen başarım oranlarına ait grafik Şekil 5'de gösterilmektedir. Burada da görüleceği üzere AdaBoost sınıflandırma algoritması hariç diğer sınıflandırma algoritmalarının başarım sonuçları birebir örtüştüğü görülmektedir.



5. TARTIŞMA VE SONUÇ

Gümrük işlemlerinde beyannamelerin kusursuz yazımı çok önemlidir. Bu çalışma gümrük sisteminde beyanname hazırlanırken

yapılabilecek olası hataların tespiti için makine öğrenme algoritmalarının kullanımının etkinliğini ortaya çıkarmak ve akıllı kontrol mekanizmalarının kurulmasına bir altyapı oluşturmak amacıyla yapılmıştır. Çalışma ülkemizdeki önde gelen Türkiye dış ticaretinin yaklaşık %8'lik hacminin gerçekleştiği uluslararası bir lojistik bir firmanın son 12 yıllık gümrük beyanname verileri üzerinden gerçekleştirilmiştir. Çalışmada bilinen ve birçok alanda başarılı bulunan makine öğrenme algoritmalarının etkinliği 4.005.343 adet gümrük beyanname teslim şekli verisi üzerinden test edilmiştir.

Veri setinin belirlenmesinde sahada ilgili teknik kişilerin görüşlerinden yararlanılarak Delphi yöntemi tercih edilmiştir. Belirlenen ilk veri seti üzerinden yapılan çalışmalarda en iyi sonuç elde edilen algoritmalarda daha düşük seviyelerde başarım oranları elde edilmişken sahada uzmanlarla birlikte yapılan iyileştirme çalışmaları sonucunda çalışmadaki başarım oranlarına kadar yükseltilmiştir. Eğitim çalışmasında Train-test split ve K-Fold yöntemleri kullanılmıştır. Çalışmamızda kullandığımız veri setinin eğitiminde 'Train-test split' yönteminden kaynaklı bilinen herhangi bir olumsuzla karşılaşılmamıştır. Her iki yöntemin kullanılması ile tüm veriler üzerinden elde edilen sonuçlar kıyaslandığında başarım oranlarının AdaBoost algoritması hariç birebir örtüştüğü görülmüştür. Sınıflandırma algoritmaları başarım oranları bakımından değerlendirildiğinde 'Decision Tree Classification', 'Nearest Neighbors Classification', 'Random Forest Classification' ve 'Bagging Classification' algoritmalarında diğerlerine göre daha başarılı sonuçlar elde edilmiştir. Train-test split yönteminde 'Decision Tree Classification' (%75.69) ve 'Bagging Classification' (%75.70) algoritmalarında en başarılı sonuç elde edilmiştir. K-Fold yönteminde ise 'Decision Tree Classification' algoritması (%75.84) ve 'Bagging Classification' (%75.83) ile en başarılı sonuçlar bulunmuştur. Bu ölçekte büyük bir veri setinden elde edilen sonuçlar yeterli seviyede başarılı bulunmuştur.

Makine öğrenimi algoritmaları bir çok alanda tahminleme yapmak için kullanılmaktadır. Akademik başarı tahmininde öğrenme yönetim sistemi log kayıtları üzerinde K-En Yakın Komşuluk, Naïve Bayes, Destek Vektör Makineleri, CART Karar Ağacı ve C5.0 Karar Ağacı sınıflandırma algoritmalarının kullanıldığı çalışmada başarım oranları %80'in üzerinde çıkmıştır (Yavuzarslan & Erol, 2022). Çalışmamızda Karar Ağaçları ve K-En Yakın Komşuluk algoritmalarında %80 oranına çok yakın sonuçlara ulaştık. Çok dilli duygu analizini YouTube verileri üzerinden Naïve Bayes sınıflandırma algoritması kullanarak yapan çalışmada başarım oranı %65.56 bulunmuştur (Şahinaslan ve ark., 2022). Centroid tabanlı sınıflayıcıların yüksek başarım gösterdiği çalışmada meme kanseri teşhisinde %99.04 oranı Euclidian tabanlı sınıflayıcı ile elde edilmiştir (Takcı, 2016). Makine öğrenme yöntemlerinin kan vermeye elverişli donörlerin tespitinde kullanıldığı çalışmada Karar Ağaçları, Destek Vektör Makineleri ve K-En Yakın Komşuluk algoritması kullanılmış ve en başarılı yöntemin Destek Vektör Makineleri olduğu sonucu bulunmuştur (Karadağ, 2021). Satış tahmininde makine öğrenme algoritalarını kullanan çalışmada en iyi r^2 skoru, hipertune edildikten sonra 0.9726 ile Rastgele Orman makine öğrenme algoritması ile tespit edilmiştir (Nacar & Erdebilli, 2021). Makine öğrenimi ile hisse senedi değerinin tahmin edildiği çalışmada derin öğrenme modellerinden LSTM (Long Short Term Memory) mimarisi ile %95 doğruluk oranı elde edilmiştir (Gavcar & Metin, 2021). Kablosuz Sensör Ağlarına dağıtık servis reddi saldırılarının makine öğrenme metotları kullanılarak tespit edildiği çalışmada tüm öğrenme modellerinde %99.72 ile en yüksek doğruluk oranı Rastgele Orman algoritmasında gerçekleşmiştir (Okur & Dener, 2021). Makine öğrenmesi yöntemleriyle iş başvurularının değerlendirildiği çalışmada artırım topluluk metodunu kullanan XGBoost modeli en yüksek başarım oranına sahip olmuştur (Ereken & Tarhan, 2021). Makine öğrenmesi yöntemleri ile otel rezervasyon iptallerinin tahmin edildiği çalışmada %73 doğruluk oranı ile C4.5 karar ağacı makine öğrenme algoritması en iyi sonucu vermiştir (Boz ve ark., 2018). Trol hesapların tespiti için makine öğrenme algoritmaları kullanılan çalışmada Twitter üzerinden elde edilen 238.925 mesaj ile çalışılmış ve Lojistik Regresyon makine öğrenmesi algoritmasında %93.93 ile en iyi sonuca ulaşılmıştır (Erdi ve ark., 2021). Prostat kanseri tümör oluşumunun makine öğrenmesi algoritmaları ile incelendiği çalışmada kullanılan 7 adet sınıflandırıcı arasında %85.37 ile Gradyan artırma algoritması en başarılı sonuçları vermiştir (Aydın Atasoy & Demiröz, 2021). Makine öğrenimi algoritmalarının gümrük işlemlerindeki etkinliğini incelediğimiz çalışmamızda elde ettiğimiz sonuçları bir çok alanda yapılan benzer amaçlı çalışmalar ışığında değerlendirdiğimizde süreci iyileştirme ve ileri götürme potansiyeli görülmektedir. İncelediğimiz çalışmaların çoğunda bir makine öğrenmesi ile yetinilmeyip çeşitli makine öğrenme algoritmaları ile çalışılıp en iyi sonucu veren algoritmalar belirlenmiştir. Çalışmamızda modelimiz üzerinde sekiz adet makine öğrenmesi algoritması kullanılarak en iyi sonucu veren algoritmalar tespit edilmiştir. En iyi sonuca sahip makine öğrenme algoritmaları kullanılarak oluşturulacak tahminleme modelleri ile kurulacak öneri sistemiyle beyanname oluşturma sürecini hızlandıracak ve kalitesini arttıracak yeni çalışmalar yapılabilecektir.

Gümrük sistemlerinde beyanname yazımında makine öğrenme algoritmalarıyla ulaşılan veri tahminleme sonuçları sektörün sürekli gelişen ve değişen dinamik yapısı da dikkate alındığında kabul edilebilir bir başarılı düzeyindedir. Bu çalışmada kullanılan öğrenme algoritmalarından elde edilen başarım oranları bir yandan sektörün bu konudaki ihtiyacını karşılama potansiyeline sahipken diğer yandan makine öğrenmesi konusunda birçok yeni çalışmaların yapılmasına da vesile olacaktır. Böyle bir yöntemin kullanılmasıyla her bir beyanname gönderimi öncesinde sistemin üreteceği tahmin sonuçlarını kullanarak kullanıcılara daha isabetli önerilerin sunulması sağlanacaktır. Böylece yapılacak tahminleme sonrası düzeltilen her veri yapılan hata oranlarını aşağıya çekmeye yardımcı olacaktır ve olası birçok problemin önüne geçecektir. Bu durum operasyon maliyetlerine ve süreçlerine de olumlu yansıyacaktır. Olası hataların doğurabileceği hukuki sonuçlar dâhil olası birçok olumsuz sonuçların da önüne geçilebilecektir.

Her sektör elindeki büyük veriyi kullanarak öğrenme temelli daha zeki sistemler, mimariler kurmak zorundadır. Dış ticaret sektörü gümrük süreçleri de bilişim teknolojilerinin sunduğu bu yeni yaklaşım ve imkânlardan istifade etmelidir. Bu çalışmadan elde edilen başarım oranları bu alanda çalışma yapmayı teşvik edecek seviyededir Bu yeni teknolojik buluş ve yöntemler lojistik sektörün gelişiminde pozitif katkı sunmaktadır. Gümrük beyanname oluşturma sürecine destek olacak öğrenme tabanlı bir kontrolün kurulmasına bu alanda başarılı bulunan öğrenme algoritmalarının katkısı büyük olacaktır. Bu çalışmada elde edilen başarılı sonuçlar hem sektör için hem de makine öğrenme teknolojisi bu alanda kullanımı açısından Türkiye'de öncü olacak niteliktedir.

Teşekkür: Çalışmamızda önerilen yapının uygulanabilirliğinin test edilmesinde verdiği destekten dolayı Barsan Global Lojistik firmasına teşekkürlerimizi sunarız.
Hakem Değerlendirmesi: Dış bağımsız.
Çıkar Çatışması: Yazarlar çıkar çatışması beyan etmemişlerdir.
Finansal Destek: Yazarlar finansal destek beyan etmemişlerdir.
Yazar Katkıları: Çalışma Konsepti/Tasarım- Ö.Ş., E.Ş., M.G.; Veri Toplama- M.G.; Veri Analizi/Yorumlama- M.G., E.Ş., Ö.Ş.; Yazı Taslağı- M.G., E.Ş.; İçeriğin Eleştirel İncelemesi- Ö.Ş., E.Ş., M.G.; Son Onay ve Sorumluluk- Ö.Ş., E.Ş., M.G.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.
Author Contributions: Conception/Design of Study- Ö.Ş., E.Ş., M.G.; Data Acquisition- M.G.; Data Analysis/Interpretation- M.G., E.Ş., Ö.Ş.; Drafting Manuscript- M.G., E.Ş.; Critical Revision of Manuscript- Ö.Ş., E.Ş., M.G.; Final Approval and Accountability- Ö.Ş., E.Ş., M.G.

Kaynaklar/References

Abdulkareem, N. M., & Abdulazeez, A. M. (2021). Machine Learning Classification Based on Radom Forest Algorithm: A Review. *International Journal of Science and Business, IJSAB International, 5(2)*, 128-142. https://ideas.repec.org/a/aif/journl/v5y2021i2p128-142.html adresinden alındı

Aka, A., & Ürünal, A. A. (2018). Türkiye'de Dış Ticaret Uygulamaları: 4458 Sayılı Gümrük Kanunu Özelinde. *Balkan Sosyal Bilimler Dergisi*, 7(13), 154-170. https://dergipark.org.tr/en/pub/bsbd/issue/34559/336447 adresinden alındı

Akgül, G., Çelik, A. A., Aydın, Z. E., & Öztürk, Z. K. (2020). Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı. *Bilişim Teknolojileri Dergisi, 13 (3)*, 255-268. doi:10.17671/gazibtd.710728

Alasadi, S. A., & Bhaya, W. S. (2017). Review of Data Preprocessing Techniques in Data Mining. *Journal of Engineering and Applied Sciences, 12(16)*, 4102-4107 ISBN: 1816-949X. https://d1wqtxts1xzle7.cloudfront.net/54509277/4102-4107-with-cover-page-v2.pdf Expires adresinden alındı

Alexandropoulos, S.-A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data preprocessing in predictive data mining. *The Knowledge Engineering Review, 34*, 1-33. doi:10.1017/s026988891800036x

Ameyaw, E. E., Hu, Y., Shan, M., Shan, S. P., & Le, Y. (2016). Application Of Delphi Method In Construction Engineering And Management Research: A Quantitative Perspective. *Journal Of Civil Engineering And Management, 22(8)*, 991–1000. doi:10.3846/13923730.2014.945953

Aydın Atasoy, N. & Demiröz, A. (2021). Makine öğrenmesi algoritmaları kullanılarak prostat kanseri tümör oluşumunun incelenmesi. Avrupa Bilim ve Teknoloji Dergisi , Ejosat Özel Sayı 2021 (ISMSIT) , 87-92 . DOI: 10.31590/ejosat.1018897

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior, 27(5)*, 1834–1839. doi:10.1016/j.chb.2011.04.004

Barua, L., Zou, B., & Zhou, Y. (2020). Machine learning for international freight transportation management: A comprehensive review. *Research in Transportation Business & Management, 100453.* doi:10.1016/j.rtbm.2020.100453

Başer, B. Ö., Yangın, M., & Sarıdaş, E. S. (2021). Makine öğrenmesi teknikleriyle diyabet hastalığının sınıflandırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 25 (1)*, 112-120. doi:10.19113/sdufenbed.842460

Boz, M. , Canbazoğlu, E. , Özen, Z. & Gülseçen, S. (2018). Otel rezervasyon iptallerinin makine öğrenmesi yöntemleri ile tahmin edilmesi. Veri Bilimi , 1 (1) , 7-14 . Retrieved from https://dergipark.org.tr/tr/pub/veri/issue/41532/490816

Brei, V. A. (2020). Machine Learning in Marketing: Overview, Learning Strategies, Applications, and Future Developments. *Foundations and Trends® in Marketing, 14(3)*, 173–236. doi:10.1561/1700000065

Büyükgüral, A., & Türkoğlu, Y. (2015). Gümrük Kanunu'na Göre Tatbik Edilen Para Cezalarında Zamanaşımı Sorunu. *Gümrük ve Ticaret Dergisi, (6)*, 74-85. https://dergipark.org.tr/tr/pub/gumrukticaretdergisi/issue/53324/708737 adresinden alındı

Canrakerta, &. H. (2020). Application of Business Intelligence for Customs Declaration: A Case Study in Indonesia. *Journal of Physics*. https://www.researchgate.net/publication/339011997_Application_of_Business_Intelligence_for_Customs_Declaration_A_Case_Study_in_Indonesia adresinden alındı

Chandrasekar, P., Qian, K., Shahriar, H., & Bhattacharya, P. (2017). Improving the Prediction Accuracy of Decision Tree Mining with Data Preprocessing. *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC).* doi:10.1109/compsac.2017.146

Chen, K., Chen, H., Conway, N., Hellerstein, J. M., & Parikh, T. S. (2011). Usher: Improving Data Quality with Dynamic Forms. *IEEE Transactions on Knowledge and Data Engineering, 23(8)*, 1138–1153. doi:doi:10.1109/tkde.2011.31

Chen, S., Webb, G. I., L. L., & Ma, X. (2019). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems, 105361*. doi:10.1016/j.knosys.2019.105361

Coşar, M., & Deniz, E. (2021). Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi. *Avrupa Bilim ve Teknoloji Dergisi, Ejosat Özel Sayı 2021 (ICAENS)*, 1112-1116. doi:10.31590/ejosat.1012986

Cui, D., & Curry, D. (2005). Prediction in Marketing Using the Support Vector Machine. *Marketing Science, 24(4)*, 595–615. doi:10.1287/mksc.1050.0123

Cui, G., Wong, M. L., & Lui, H.-K. (2006). Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming. *Management Science, 52(4)*, 597–612. doi:10.1287/mnsc.1060.0514

Dahouda, M. K., & Joe, I. (2021). A Deep-Learned Embedding Technique for Categorical Features Encoding. *IEEE Access, 9*, 114381–114391. doi:10.1109/access.2021.3104357

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1(1-4)*, 131–156. doi:10.1016/s1088-467x(97)00008-5

Dobriban, E., & Wager, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *The Annals of Statistics, 46(1)*, 247–279. doi:10.1214/17-aos1549

Erdi, B. , Şahin, E. A. , Toydemir, M. S. & Dökeroğlu, T. (2021). Makine öğrenmesi algoritmaları ile trol hesapların tespiti. Düzce Üniversitesi Bilim ve Teknoloji Dergisi , 9 (1) , 430-442 . DOI: 10.29130/dubited.748366

Ereken, Ö. & Tarhan, Ç. (2021). İş başvurularının makine öğrenmesi yöntemleriyle değerlendirilmesi. Yönetim Bilişim Sistemleri Dergisi , 7 (2) , 65-85. Retrieved from https://dergipark.org.tr/tr/pub/ybs/issue/67451/991689

García, I. G., & Caballero, A. M. (2021). A Multi-Objective Bayesian Approach with Dynamic Optimization. A Hybrid of Decision Theory and Machine Learning Applied to Customs Fraud Control in Spain. *Mathematics, 9(13)*, 1529. doi:10.3390/math9131529

Gavcar, E. & Metin, H. M. (2021). Hisse Senedi Değerlerinin Makine Öğrenimi (Derin Öğrenme) ile Tahmini . Ekonomi ve Yönetim Araştırmaları Dergisi , 10 (2) , 1-11 . Retrieved from https://dergipark.org.tr/tr/pub/eyad/issue/68049/1056795

Gulia, A., Vohra, R., & Rani, P. (2014). Liver Patient Classification Using Intelligent Techniques. *International Journal of Computer Science and Information Technologies, 5 (4)*, 5110-5115 . https://www.semanticscholar.org/paper/Liver-Patient-Classification-Using-Intelligent-Gulia-Vohra/798856e3c30ed88661d0aa596cc23e12410181f5#references adresinden alındı

Gümrük Rehberi. (2021). 2021 tarihinde T.C. Ticaret Bakanlığı: https://gumrukrehberi.gov.tr/sayfa/g%C3%BCmr%C3%BCk-beyannamesinde-hangi-bilgiler-yer-al%C4%B1r adresinden alındı

Huang, J.-C., Ko, K.-M., Shu, M.-H., & Hsu, B.-M. (2019). Application and comparison of several machine learning algorithms and their integration models in regression problems. *Neural Computing and Applications*. doi:10.1007/s00521-019-04644-5

İhracatta Kullanılan Uluslararası Dökümanlar. (2021). 2021 tarihinde Mevzuat.Net: https://www.mevzuat.net/fayda/dokumanlar.aspx adresinden alındı

İlgün, M. F. (2020). Vergi Denetim Sürecinde Büyük Veri Analitiği. *Siyaset, Ekonomi ve Yönetim Araştırmaları Dergisi, 8 / 1*, 1-24. https://dergipark.org.tr/en/pub/seyad/issue/55429/698700 adresinden alındı

Jackson, E., & Agrawal, R. (2019). Performance Evaluation of Different Feature Encoding Schemes on Cybersecurity Logs. *2019 SoutheastCon*. doi:10.1109/southeastcon42311.2019.9020560

Karadağ, K. (2021). Kan vermeye elverişli donörlerin makine öğrenme yöntemleri ile tespiti . Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi , 8 (15) , 508-514 . DOI: 10.54365/adyumbd.993772

Kaya, M., & Doğan, A. (2020). Dış Ticarete Konu Eşyanın Vergilendirilmesinde Gümrük Kıymetinin Rolü, Beyanı ve Kontrolü. *Gümrük ve Ticaret Dergisi, 7 (19)*, 10-24. https://dergipark.org.tr/tr/pub/gumrukticaretdergisi/issue/53766/695384 adresinden alındı

Koh, L. D. (2020). Blockchain in transport and logistics – paradigms and transitions. *International Journal of Production Research*, 58:7, 2054-2062,. doi:10.1080/00207543.2020.1736428

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering, 160(1)*, 3-24.

Li, G., & Li, N. (2019). Customs classification for cross-border e-commerce based on text-image adaptive convolutional neural network. *Electronic Commerce Research*. doi:doi:10.1007/s10660-019-09334-x

Lund, B. D. (2020). Review of the Delphi method in library and information science research. *Journal of Documentation, 76(4)*, 929–960. doi:10.1108/jd-09-2019-0178

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing – Connecting computing power to human insights. *International Journal of Research in Marketing*. doi:10.1016/j.ijresmar.2020.04.005

Mammadov, F. (2020). Customs system and using advanced technologies in customs [Master Thesis, Technische Universität Wien]. reposiTUm. https://resolver.obvsg.at/urn:nbn:at:at-ubtuw:1-138802 adresinden alındı

Maruev, S., Stefanovskiy, D., Frolov, A., Troussov, A., & Curry, J. (2014). Deep Mining of Custom Declarations for Commercial Goods. *Procedia*

Economics and Finance, 12, 397–402. doi:10.1016/S2212-5671(14)00360-8

Mevzuat Bilgi Sistemi. (2021). 12 22, 2021 tarihinde Mevzuat Bilgi Sistemi: https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=4458&MevzuatTur=1&MevzuatTertip=5 adresinden alındı

Mullooly, J. P. (1990). The effects of data entry error: An analysis of partial verification. *Computers and Biomedical Research, 23(3)*, 259–267. doi:https://doi.org/10.1016/0010-4809(90)90020-D

Nacar, E. N. & Erdebilli (b.d.rouyendegh), B. (2021). Makine öğrenmesi algoritmaları ile satış tahmini . Endüstri Mühendisliği , 32 (2) , 307-320 . DOI: 10.46465/endustrimuhendisligi.811183

Okur, C. & Dener, M. (2021). Makine öğrenme metotları kullanılarak ksa ddos saldırıları tespiti . El-Cezeri , 8 (3) , 1550-1564 . DOI: 10.31202/ecjse.971592

Özer, A. C. (2020). Türkiye'nin Gümrük Prosedür Uygulamaları ve Etkinliği Üzerine Bir Değerlendirme. *Bitlis Eren Üniversitesi Sosyal Bilimler Dergisi, 9 (2)*, 341-345. doi:10.47130/bitlissos.842154

Pamuk, Z., & Kaya, C. (2021). Classification of Type 2 Diabetes Using Machine Learning Techniques. *Avrupa Bilim ve Teknoloji Dergisi, Ejosat Özel Sayı 2021 (ICAENS)*, 1265-1268. doi:10.31590/ejosat.1014878

Paper, D. (2020). *Hands-on Scikit-Learn for Machine Learning Applications.* doi:10.1007/978-1-4842-5373-1

Paula, E. L., Ladeira, M., Carvalho, R. N., & Marzagao, T. (2016). Deep Learning Anomaly Detection as Support Fraud Investigation in Brazilian Exports and Anti-Money Laundering. *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA).* doi:10.1109/icmla.2016.0172

Refaeilzadeh, P., Tang, L., & Liu, H. (2016). Cross-Validation. Encyclopedia of Database Systems, 1–7. doi:10.1007/978-1-4899-7993-3_565-2

Roshan, S. E., & Asadi, S. (2020). Improvement of Bagging performance for classification of imbalanced datasets using evolutionary multi-objective optimization. *Engineering Applications of Artificial Intelligence, 87*, 103319. doi:10.1016/j.engappai.2019.103319

Ryzhova, A., & Sochenkov, I. (2019). Deep Learning for Customs Classification of Goods Based on Their Textual Descriptions Analysis. *2019 Ivannikov Ispras Open Conference*, *2019*, s. pp. 55-59. doi:10.1109/ispras47671.2019.00014

Saçık, S. (2009). Dış Ticaret Politikası ve Ekonomik Büyüme İlişkisi. *Karamanoğlu Mehmetbey Üniversitesi Sosyal ve Ekonomik Araştırmalar Dergisi*, 2009(1), 162-171. https://dergipark.org.tr/tr/pub/kmusekad/issue/10220/125634 adresinden alındı

Sahinaslan, O., & Sahinaslan, E. (2019). "Cross-object information security: A study on new generation encryption". *AIP Conference Proceedings*(2086, 030034 (2019)). doi:https://doi.org/10.1063/1.5095119.

Salve, S., Bhutkar, G., & Yammiyavar, P. (2021). Can Dynamic Widgets Improve Data Entry Efficiency? *In: Muzammil M., Khan A.A., Hasan F. (eds) Ergonomics for Improved Productivity. Design Science and Innovation. Springer, Singapore.* doi:https://doi.org/10.1007/978-981-15-9054-2_90

Saygın, E., & Baykara, M. (2021). Karaciğer Yetmezliği Teşhisinde Özellik Seçimi Kullanarak Makine Öğrenmesi Yöntemlerinin Başarılarının Ölçülmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 33 (2)*, 367-377. doi:10.35234/fumbd.832264

Schapire, R. E. (1999). Theoretical Views of Boosting and Applications. *Algorithmic Learning Theory*, 13–25. doi:10.1007/3-540-46769-6_2

Shao, H. &.-R. (2002). Applying data mining to detect fraud behavior in customs declaration. https://www.researchgate.net/publication/3995840_Applying_data_mining_to_detect_fraud_behavior_in_customs_declaration adresinden alındı

Siau, K. L., & Yang, Y. (2017). Impact of Artificial Intelligence, Robotics, and Machine Learning on Sales and Marketing. *MWAIS 2017 Proceedings*, 48. http://aisel.aisnet.org/mwais2017/48 adresinden alındı

Sourani, A., & Sohail, M. (2014). The Delphi Method: Review and Use in Construction Management Research. *International Journal of Construction Education and Research, 11(1)*, 54–76. doi:10.1080/15578771.2014.917132

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 12-14. doi:10.11613/bm.2014.003

Sundsoy, P., Bjelland, J., Iqbal, A. M., "Sandy" Pentland, A., & de Montjoye, Y.-A. (2014). Big Data-Driven Marketing: How Machine Learning Outperforms Marketers' Gut-Feeling. *Lecture Notes in Computer Science*, 367–374. doi:10.1007/978-3-319-05579-4_45

Şahinaslan, E. (2020). Endüstri 4.0 Dönüşümünde Öne Çıkan Teknolojiler,. A. &. Hayaloğlu içinde, *Mühendislik Alanında Akademik Çalışmalar* (s. 235-252). Ankara: Gece Kitaplığı.

Şahinaslan, Ö. , Dalyan, H. & Şahinaslan, E. (2022). Naive bayes sınıflandırıcısı kullanılarak youtube verileri üzerinden çok dilli duygu analizi .Bilişim Teknolojileri Dergisi , 15 (2) , 221-229 . DOI: 10.17671/gazibtd.999960

Şahinaslan, Ö. (2020). "Yeni Nesil Teknolojiler". G. &. Telli içinde, *Digital Dönüşüm* (s. 48-49). İstanbul: Maltepe University Books.

T.C. Ticaret Bakanlığı. (2021, Aralık). 2021 tarihinde Yıllara Göre Beyanname Sayıları: https://ticaret.gov.tr/data/5d63d89d13b8762f7c43a738/20-Yillara%20Gore%20Beyanname%20Sayilari.pdf adresinden alındı

Takcı, H. (2016). Centroid sınıflayıcılar yardımıyla meme kanseri teşhisi . Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi , 31 (2) , 0-0 . DOI: 10.17341/gummfd.50403

Tang, J., Deng, C., & Huang, G.-B. (2016). Extreme Learning Machine for Multilayer Perceptron. *IEEE Transactions on Neural Networks and Learning Systems, 27(4)*, 809–821. doi:10.1109/tnnls.2015.2424995

Uçan, O., & Koçak, E. (2014). Türkiye'de dış ticaret ve ekonomik büyüme arasındaki ilişki. *Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 7(2), 51. https://dergipark.org.tr/tr/pub/niguiibfd/issue/19755/211482 adresinden alındı

Wang, M. L., & Choi, C. H. (2018). How information and communication technology affect international trade: a comparative analysis of BRICS countries. *Information Technology for Development*, 1–20. doi:10.1080/02681102.2018.1493675

Yavuzarslan, M. & Erol, Ç. (2022). Öğrenme yönetim sistemi log kayıtlarının akademik başarı tahmininde kullanılması . Bilişim Teknolojileri Dergisi, 15 (2) , 199-207 . DOI: 10.17671/gazibtd.837884

Zartha Sossa, J. W., Halal, W., & Zarta, R. H. (2019). Delphi method: analysis of rounds, stakeholder and statistical indicators. *Foresight, 21(5)*, 525–544. doi:10.1108/fs-11-2018-0095

Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2018). Efficient kNN Classification With Different Numbers of Nearest Neighbors. *IEEE Transactions on Neural Networks and Learning Systems, 29(5)*, 1774–1785. doi:10.1109/tnnls.2017.2673241

Zhang, Y., Ni, M., Zhang, C., Liang, S., Fang, S., Li, R., & Tan, Z. (2019). Research and Application of AdaBoost Algorithm Based on SVM. *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC).* doi:10.1109/itaic.2019.8785556

Zien, A., Krämer, N., Sonnenburg, S., & Rätsch, G. (2009). The Feature Importance Ranking Measure. *Lecture Notes in Computer Science*, 694–709. doi:10.1007/978-3-642-04174-7_45