GÜMRÜK SİSTEMLERİNDE ÖĞRENME ALGORİTMALARI İLE DOĞRU BEYANNAME OLUŞTURMA VE KONTROL UYGULAMASI

Mustafa GÜNERKAN
211402103

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Önder ŞAHİNASLAN

İstanbul
T.C. Maltepe Üniversitesi
Lisansüstü Eğitim Enstitüsü
Ekim, 2022

GÜMRÜK SİSTEMLERİNDE ÖĞRENME ALGORİTMALARI İLE DOĞRU BEYANNAME OLUŞTURMA VE KONTROL UYGULAMASI

Mustafa GÜNERKAN
211402103
Orcid: 0000-0002-4202-2801

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Önder ŞAHİNASLAN

İstanbul
T.C. Maltepe Üniversitesi
Lisansüstü Eğitim Enstitüsü
Ekim, 2022



JÜRİ VE ENSTİTÜ ONAYI

Bu belge, Yükseköğretim Kurulu tarafından 19.01.2021 tarihli "*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*" ile bildirilen 6689 Sayılı Kişisel Verilerin Korunması Kanunu kapsamında gizlenmiştir.

ETİK İLKE VE KURALLARA UYUM BEYANI

Bu belge, Yükseköğretim Kurulu tarafından 19.01.2021 tarihli "*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*" ile bildirilen 6689 Sayılı Kişisel Verilerin Korunması Kanunu kapsamında gizlenmiştir.

TEŞEKKÜR

Hayat boyu her daim yanımda olan sevgili aileme, tez çalışmam sürecinde her türlü desteği veren yöneticim Sayın Serkan DİNGA'ya, değerli çalışma arkadaşlarıma, çalışmalarım sırasında bilgi ve deneyimleriyle yolumu aydınlatan, samimiyetle her türlü desteğini esirgemeyen kıymetli danışman hocam Dr.Öğr.Üyesi Önder ŞAHİNASLAN'a, çalışmamda verdiği katkı ve desteklerden dolayı Dr.Ender ŞAHİNASLAN'a teşekkür ederim.

Mustafa GÜNERKAN

Ekim, 2022

ÖZ

GÜMRÜK SİSTEMLERİNDE ÖĞRENME ALGORİTMALARI İLE DOĞRU BEYANNAME OLUŞTURMA VE KONTROL UYGULAMASI

Mustafa Günerkan
Yüksek Lisans Tezi
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Önder Şahinaslan
Maltepe Üniversitesi Lisansüstü Eğitim Enstitüsü, 2022

Gümrük sisteminde hatasız beyanname yazımı büyük önem taşımaktadır. Beyanname yazımında yapılan hatalar istenmeyen sonuçlara sebep olmaktadır. Özellikle yetkin ve yeterli insan kaynağının olmamasından dolayı beyanname yazımlarında hatalar yapılmaktadır. Beyannamelerin hatalı oluşturulmasından dolayı ciddi miktarlarda cezai durumlarla karşılaşılmaktadır. Bu durum firmaların operasyon maliyetlerini ve süreçlerini olumsuz etkilemektedir. Bu çalışma bu tür olumsuz sonuçların giderilmesi, operasyonel süreçlerin otomatize edilmesi, insan müdahalesinin asgari seviyeye çekilmesi, beyannamelerin hatasız ve doğru oluşturulması için yapılmıştır. Ayrıca gümrük operasyonlarındaki olası hataların doğurabileceği hukuki süreçlerin önüne geçilmesi de hedeflenmektedir. Bir gümrük beyannamesi birçok bileşenden oluşan karmaşık bir yapıya sahiptir. Bu bağlamda kontrol edilebilecek her bileşen için bir veri seti oluşturulmuştur. Çalışma kapsamında makine öğrenmesi ve derin öğrenme algoritmalarını modelleyerek oluşturulan büyük veri setleri üzerinde çalışılmıştır. Geliştirilen bu modeller sonucunda kullanıcıya her bileşen için önerilerde bulunulmaktadır. Öneri karşısında kullanıcının son kontrol ve onayı alınarak beyanname işlemi sonuçlandırılmaktadır. Böylece insan kaynaklı hataların tespiti sağlanarak hata oranlarının minimum seviyeye çekilmesi sağlanmıştır. Bu makine öğrenmesi tabanlı özgün yaklaşım ile gümrük beyannameleri oluşturularak dış ticarette önemli bir soruna çözüm sağlanmıştır.

Anahtar Sözcükler: Teknoloji ve Yenilik, Gümrük Beyannamesi, Öğrenme Algoritmaları, Veri Madenciliği, Sınıflandırma, Lojistik Sektör, Büyük Ve

ABSTRACT

APPLICATION OF DESIGNING AND REVIEWING CUSTOMS DECLARATION BY USING LEARNING ALGORITHMS

Mustafa Günerkan

Master Thesis

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Asst. Prof. Önder Şahinaslan

Maltepe University Graduate School, 2022

Error-free declaration writing is of great importance in the customs system. Errors made in the writing of the declaration lead to undesirable consequences. Mistakes are made in the writing of declarations, especially due to the lack of competent and adequate human resources. Due to the incorrect creation of declarations, serious criminal cases are encountered. This situation negatively affects the operational costs and processes of companies. This study was carried out to eliminate such negative results, to automate operational processes, to minimize human intervention, and to create accurate and error-free declarations. In addition, it is aimed to prevent legal processes that may result from possible errors in customs operations. A customs declaration has a complex structure consisting of many components. In this context, a data set has been created for each component that can be controlled. Within the scope of the study, large data sets created by modeling machine learning and deep learning algorithms were studied. As a result of these developed models, recommendations are made to the user for each component. In the face of the proposal, the declaration process is concluded by obtaining the final control and approval of the user. Thus, it was ensured that human-caused errors were detected and error rates were minimized. With this machine learning-based original approach, customs declarations were created and a solution to an important problem in foreign trade was provided.

Keywords: Technology and Innovation, Customs Declaration, Learning Algorithms, Data Mining, Classification, Logistics Sector, Big Data

İÇİNDEKİLER

JÜRİ VE ENSTİTÜ ONAYI i
ETİK İLKE VE KURALLARA UYUM BEYANI ii
TEŞEKKÜR iii
ÖZ iv
ABSTRACT vi
İÇİNDEKİLER vii
TABLOLAR LİSTESİ x
ŞEKİLLER LİSTESİ xii
KISALTMALAR xv
BÖLÜM 1. GİRİŞ 1
Literatür Taraması 7
Problem 12
Amaç 12
Önem 13
Varsayımlar 13
Sınırlıklar 13
Tanımlar 14
Gümrük Kavramı 14
Gümrük Mevzuatı 14
Gümrük Kanunu 15
Gümrük 16
Gümrükleme 16
Gümrük Müşaviri 16
Gümrüklerde Beyanların Kontrol Usulleri 17
Kırmızı Hat 18
Sarı Hat 18
Yeşil Hat 18
Mavi Hat 18
Gümrük Tarife İstatistik Pozisyonu 19

Gümrük Beyannamesi 20
Bilge Sistemi 24
Delphi Metodu 24
Makine Öğrenmesi 26
Öğrenme Yaklaşımları 28
Gözetimli Öğrenme 28
Gözetimsiz Öğrenme 28
Pekiştirmeli Öğrenme 29
Yarı Gözetimli Öğrenme 29
Öğrenmeyi Öğrenme 29
Makine Öğrenmesi Algoritma Türleri 30
Regression 30
Sınıflandırma (Classification) 32
Clustering 36
Makine Öğrenmesi Algoritmaları 37
Gaussian Naïve Bayes 38
Decision Tree Sınıflandırma Algoritması 40
K-Nearest Neighbors Sınıflandırma Algoritması 43
Random Forest Sınıflandırma Algoritması 47
Ridge Sınıflandırma Algoritması 49
Bagging Sınıflandırma Algoritması 49
AdaBoost Sınıflandırma Algoritması 50
Çok Katmanlı Algılayıcı 50
BÖLÜM 2. YÖNTEM VE METOT 51
Kullanılan Yazılım Teknolojileri 51
Delphi 51
SQL Server 52
Python 53
Anaconda 54
Spyder 55
Araştırma Modeli 56

Evren ve Örneklem 57
Veriler ve Toplanması 57
Verilerin Çözümlenmesi ve Yorumlanması 75
Akıllı Beyanname Uygulamasının Geliştirilmesi 85
BÖLÜM 3. BULGULAR VE YORUMLAR 96
Bulgular 96
Yorumlar 100
BÖLÜM 4. SONUÇ 103
Özet 103
Yargı 103
Öneriler 103
KAYNAKÇA 104
ÖZGEÇMİŞ 117

TABLOLAR LİSTESİ

Tablo 1. 2013-2021 Türkiye İhracat Rakamları 1

Tablo 2. 2013-2021 Türkiye İthalat Rakamları 1

Tablo 3. İstanbul Bölgesi Gümrük Müdürlükleri 17

Tablo 4. Taşıma Şekli Tablosu 57

Tablo 5. Taşıma Şekli Çeşitleri 58

Tablo 6. Taşıma Şekli Kayıt Sayıları 59

Tablo 7. Döviz Türü Tablosu 60

Tablo 8. Döviz Türü Çeşitleri 61

Tablo 9. Döviz Türü Kayıt Sayıları 62

Tablo 10. Sevk Ülke Tablosu 63

Tablo 11. Sevk Ülke Listesi 64

Tablo 12. Sevk Ülke Kayıt Sayıları 65

Tablo 13. Birim Tablosu 65

Tablo 14. Birim Çeşitleri 66

Tablo 15. Birim Kayıt Sayıları 68

Tablo 16. Uluslararası Anlaşma Tablosu 69

Tablo 17. Uluslararası Anlaşma Çeşitleri 70

Tablo 18. Uluslararası Anlaşma Kayıt Sayıları 71

Tablo 19. Ödeme Şekli Tablosu 72

Tablo 20. Ödeme Şekli Kodları Listesi 73

Tablo 21. Ödeme Şekli Kayıt Sayıları 74

Tablo 22. Taşıma Şekli Veri Seti Sayısallaştırması 76

Tablo 23. Döviz Türü Veri Seti Sayısallaştırılması 76

Tablo 24. Sevk Ülkesi Veri Seti Sayısallaştırması 77

Tablo 25. Birim Veri Seti Sayısallaştırılması 77

Tablo 26. Uluslararası Anlaşma Veri Seti Sayısallaştırılması 78

Tablo 27. Ödeme Şekli Veri Seti Sayısallaştırılması 78

Tablo 28. Declaration1 Özniteliğinin Sayısallaştırma Tablosu 90

Tablo 29. Taşıma Şekli Veri Seti Sonuçları 97

Tablo 30. Döviz Türü Veri Seti Sonuçları 97

Tablo 31. Sevk Ülke Veri Seti Sonuçları 98

Tablo 32. Birim Veri Seti Sonuçları 99

Tablo 33. Uluslararası Anlaşma Veri Seti Sonuçları 99

Tablo 34. Ödeme Şekli Veri Seti Sonuçları 100

ŞEKİLLER LİSTESİ

Şekil 1. İthalat/İhracat Süreci 3

Şekil 2. 2019-2021 Yılları Beyanname Sayıları 3

Şekil 3. 2019-2021 İhracat ve İthalat Yapan Firma Sayıları 4

Şekil 4. Gümrük Vergi Gelirlerinin Toplam Vergi Gelirleri İçindeki Payı 5

Şekil 5. Gümrük Beyanname Form Örneği 21

Şekil 6. Beyanname Program Örneği 22

Şekil 7. Beyanname Oluşturma Süreci 23

Şekil 8. Delphi Metodu Süreci 26

Şekil 9. Linear Regresyon Modeli 32

Şekil 10. Binary Sınıflandırma 34

Şekil 11. Multi-Class Sınıflandırma 35

Şekil 12. Imbalanced Sınıflandırma 36

Şekil 13. Clustering 37

Şekil 14. Navive Bayes Algoritma Türleri 39

Şekil 15. Decision Tree Mimarisi 41

Şekil 16. KNN Algoritması Örneği 44

Şekil 17. Euclidean Uzaklığı 45

Şekil 18. Manhattan Uzaklığı 46

Şekil 19. Random Forest Mimarisi 48

Şekil 20. Delphi Geliştirme Ortamı 52

Şekil 21. SQL Server Arayüzü 53

Şekil 22. Python Kod Örneği 54

Şekil 23. Anaconda Navigator Görünümü 55

Şekil 24. Spyder Geliştirme Ortamı 56

Şekil 25. 2010-2021 Taşıma Şekli Veri Seti Beyanname Sayıları 59

Şekil 26. 2010-2021 Döviz Türü Veri Seti Beyanname Sayıları 62

Şekil 27. 2010-2021 Sevk Ülke Veri Seti Beyanname Sayıları 64

Şekil 28. 2010-2021 Birim Veri Seti Beyanname Sayıları 67

Şekil 29. 2010-2021 Uluslararası Anlaşma Veri Seti Beyanname Sayıları 71

Şekil 30. 2010-2021 Ödeme Şekli Veri Seti Beyanname Sayıları 74

Şekil 31. Veri Önişleme Aşamaları 75

Şekil 32. Taşıma Şekli Sayıları 79

Şekil 33. Döviz Türü Sayıları 79

Şekil 34. Sevk Ülke Sayıları 80

Şekil 35. Birim Sayıları 80

Şekil 36. Uluslararası Anlaşma Sayıları 81

Şekil 37. Ödeme Şekli Sayıları 81

Şekil 38. Taşıma Şekli Modelinde Özniteliklerin Ağırlığı 82

Şekil 39. Döviz Türü Modelinde Özniteliklerin Ağırlığı 82

Şekil 40. Sevk Ülke Modelinde Özniteliklerin Ağırlığı 83

Şekil 41. Birim Modelinde Özniteliklerin Ağırlığı 84

Şekil 42. Uluslararası Anlaşma Modelinde Özniteliklerin Ağırlığı 84

Şekil 43. Ödeme Şekli Modelinde Özniteliklerin Ağırlığı 85

Şekil 44. Taşıma Şekli Tablo Yapısı 86

Şekil 45. Taşıma Şekli Tablosuna Veri Ekleme 87

Şekil 46. Döviz Türü Veri Setinde Sayısallaştırma 87

Şekil 47. Makine Öğrenmesi Python Eğitim Kodu 88

Şekil 48. Taşıma Şekli Modelinde Görselleştirme Python Kodu 89

Şekil 49. Taşıma Şekli Modelinde Özniteliklerin Ağırlığın Python Kodu 89

Şekil 50. Taşıma Şekli Veri Setinin Eğitim Sonuçlarının Saklanması 90

Şekil 51. Eğitim Sonuçları Binary Dosyaları 91

Şekil 52. Taşıma Şekli Modelinde Tahminleme Python Kodu 91

Şekil 53. Akıllı Beyanname Kontrol Uygulaması Başlatılması 92

Şekil 54. Tahminleme Yapan Python Programının Çalışma Süresi 93

Şekil 55. Akıllı Beyanname Kontrol Uygulaması Sonuçları 94

Şekil 56. Akıllı Beyanname Kontrol Uygulaması Log Kayıtları 95

KISALTMALAR

AB : Avrupa Birliği

BİLGE : Bilgisayarlı Gümrük Etkinlikleri

EX : İhracat

FTP : File Transfer Protocol

GTİP : Gümrük Tarife İstatistik Pozisyonu

IM : İthalat

KNN : K-Nearest Neighbors

MNIST : Modified National Institute of Standards and Technology

MLP : Multi-Layer Perceptron

OLAP : Online Analytical Processing

SFTP : Secure File Transfer Protocol

SMOTE : Synthetic Minority Oversampling Technique

XML : Extensible Markup Language

TPKK : Türk Parası Kıymetini Koruma Kanunu

BÖLÜM 1. GİRİŞ

Bilgi ve iletişim teknolojileri 1990'lı yıllardan itibaren hızla gelişmektedir. Bu gelişmeler her geçen gün insanları birbirine daha da yakınlaştırmaktadır. Bu yakınlaşma ülkeler için dış ticaret imkânlarını da arttırmaktadır. Bilgi ve iletişim teknolojisi kullanımı ticaret bilgisini iyileştirmeye, ticaret engellerini azaltmaya ve mesafeleri kısaltmaya yardımcı olarak ticaret verimliliğini ve hacmini arttırmıştır. Bununla birlikte ticaret maliyetlerini önemli ölçüde azaltmıştır (Wang ve Choi, 2018).

Dış ticaret ile büyüme arasında da bağlantı vardır (Uçan ve Koçak, 2014). Dış ticaret büyümenin ana unsurlarından biri olarak değerlendirilmektedir. Ülkeler ürettikleri malların fazlasını ihraç ederek gelir elde edebilmektedir. Ayrıca kendisinde mevcut olmayan veya üretemediği malları ithal ederek ihtiyaçlarını karşılayabilmektedir (Saçık, 2009).

Ülkemizde dış ticaret hacmi her geçen yıl artış göstermektedir. Artan dış ticaret imkanları ithalat ve ihracat rakamlarına yansımaktadır. Tablo 1'de (Türkiye İstatistik Kurumu, 2021) Türkiye'nin 2013 ile 2021 yılları arasındaki ihracat rakamları gösterilmiştir.

Tablo 1. 2013-2021 Türkiye İhracat Rakamları

Yıl	İhracat Dolar	İhracat Euro	İhracat TL
2013	161.480.914.702	121.863.850.941	308.103.246.552
2014	166.504.861.795	125.352.688.678	364.756.120.706
2015	150.982.113.766	135.980.782.870	410.764.687.932
2016	149.246.999.263	134.862.454.019	451.885.142.948
2017	164.494.619.316	145.881.210.828	601.271.453.500
2018	177.168.756.288	150.130.297.690	859.292.263.175
2019	180.832.721.702	161.569.183.727	1.028.025.283.663
2020	169.637.755.310	148.138.523.253	1.198.816.044.393
2021	225.264.314.222	190.826.030.074	2.023.525.242.535

Tablo 2'de (Türkiye İstatistik Kurumu, 2021) Türkiye'nin 2013 ile 2021 yılları arasındaki ithalat rakamları gösterilmiştir.

Tablo 2. 2013-2021 Türkiye İthalat Rakamları

Yıl	İthalat Dolar	İthalat Euro	İthalat TL

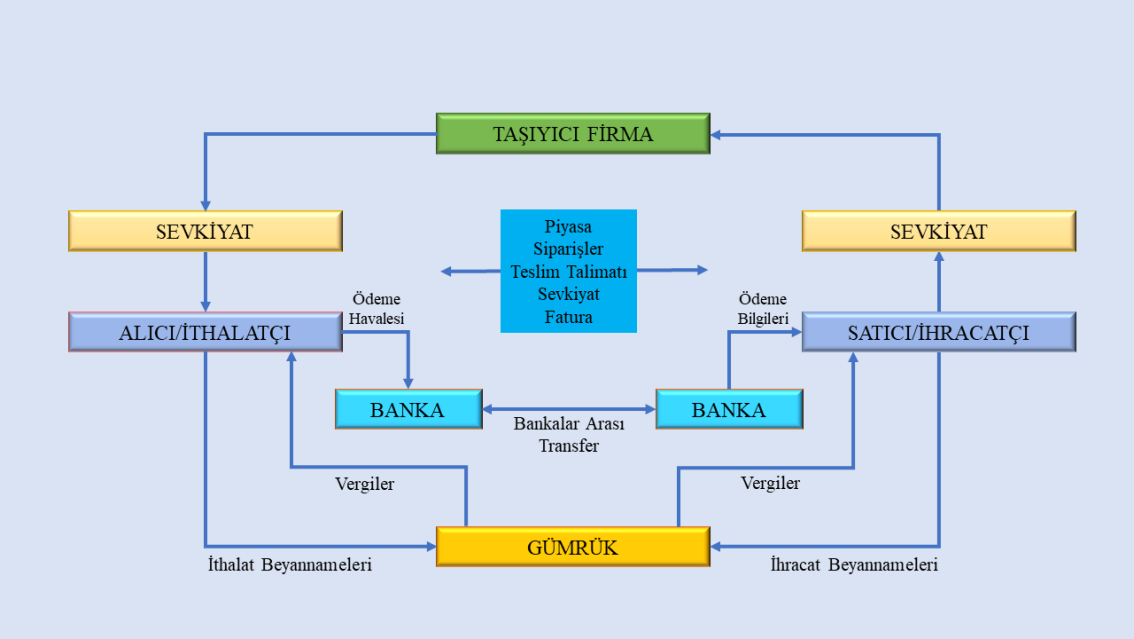
2013	260.822.803.002	196.680.580.542	496.377.064.101
2014	251.142.429.205	188.975.363.293	549.057.845.719
2015	213.619.211.455	192.213.730.127	579.198.244.287
2016	202.189.241.859	182.659.779.356	610.949.525.580
2017	238.715.127.912	211.397.408.269	870.796.829.167
2018	231.152.482.645	195.128.590.174	1.089.848.202.775
2019	210.345.202.552	187.857.716.019	1.194.475.897.628
2020	219.516.806.838	191.952.254.820	1.545.366.267.707
2021	271.422.757.506	229.820.849.052	2.437.710.854.349

Tablo 1 ve Tablo 2'de görüldüğü üzere ithalat ve ihracat rakamları her sene artış göstermiştir. Bu verilere göre Türkiye'nin ithalat rakamları ihracat rakamlarından fazladır ve bu yıllar arasında Türkiye dış ticaret açığı veren bir ülke olmuştur (Dağ ve Kızılkaya, 2019).

Dış ticareti her ülke kendi mevzuatları ışığında belli kurallara göre yapmaktadır ve bu kurallar sürekli gelişim halindedir. Ülkemizde 4458 sayılı Gümrük Kanunu ile dış ticaret işlemleri düzenlenmiştir. 4458 numaralı Gümrük Kanunu 27.10.1999'da kabul edilmiştir. 4.11.1999 tarihli ve 23866 sayılı Resmi Gazete 'de yayımlanmıştır (Aka ve Ürünal, 2018). Gümrük Kanunu ile Türkiye Cumhuriyeti Gümrük Bölgesine giren ve çıkan eşyaya ve taşıt araçlarına uygulanacak gümrük kuralları belirlenmiştir (Mevzuat Bilgi Sistemi, 2021). Türkiye Cumhuriyeti Gümrük Bölgesi, Türkiye Cumhuriyeti topraklarını kapsar. Türkiye kara suları, iç suları ve hava sahası gümrük bölgesine dâhildir (Mevzuat Bilgi Sistemi, 2021).

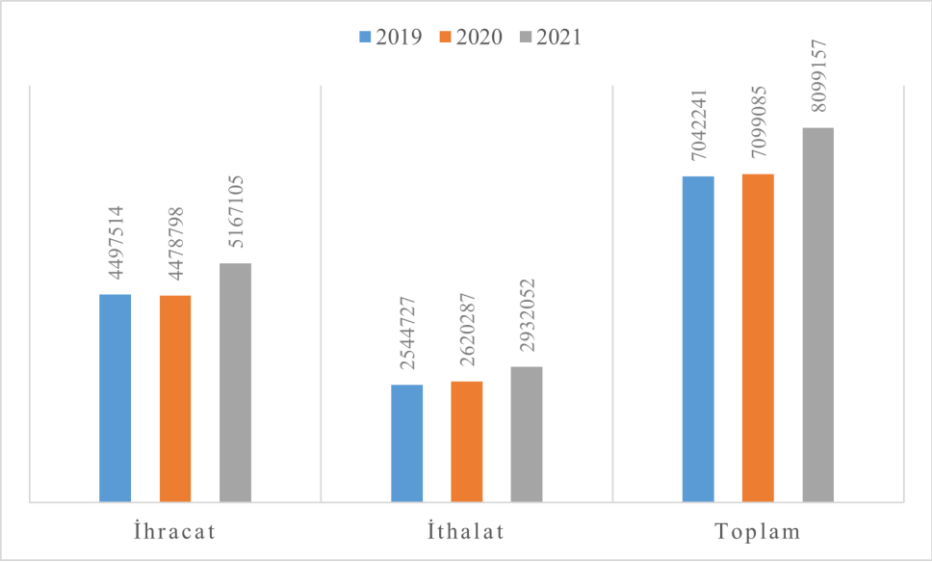
Gümrük işlemleri dış ticaretin önemli bir aşamasıdır. İşlemlerin yapılabilmesi için belli bilgileri içeren gümrük beyannamelerinin ilgili gümrük müdürlüğüne sunulması gerekmektedir. Teknolojik gelişmeler bu süreçlerde ciddi iyileşmeler sağlamıştır. İşlem sayılarının çok fazla olduğu günümüzde mevcut entegrasyon işlemleri ile ciddi sayıda beyanname işlemi kısa sürelerde sonuçlandırılabilmektedir (Güldüren ve Öztop, 2020).

Dış ticaret ithalat ve ihracat işlemlerini kapsar. Şekil 1'de görüldüğü gibi ithalat ve ihracat süreçleri birbirini izleyen işlemlerin bir bütünüdür. İthalat ve ihracat yapmak için gerekli şartların bilinmesi ve bunların eksiksiz olarak yapılması şarttır. İthalat ve ihracat süreçleri ciddi anlamda bilgi ve tecrübe gerektirmektedir.



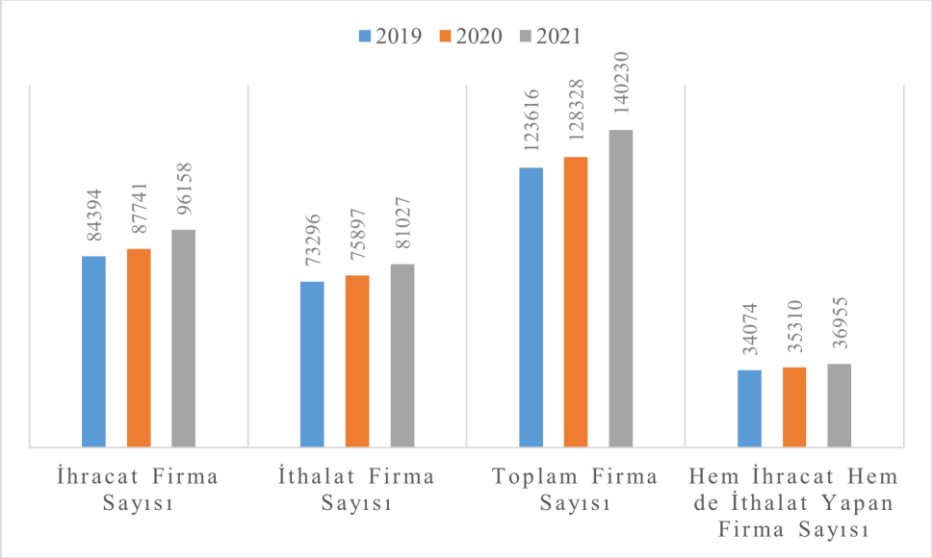
Şekil 1. İthalat/İhracat Süreci

Dış ticaret hacmi yıldan yıla artmaktadır ve bu süreçlerde de ciddi çalışmalar yapılmaktadır ve daha da yapılması gerekecektir. Şekil 2'de (T.C. Ticaret Bakanlığı, 2021) 2019, 2020 ve 2021 yıllarında ülkemizde yapılan beyanname sayıları verilmektedir. İhracat beyanname sayısı ithalat beyanname sayısından fazladır. Fakat ithalat beyanname sayısı az olsa da tutar olarak ihracattan fazladır. Bu fark dış ticaret açığına sebep olmaktadır.



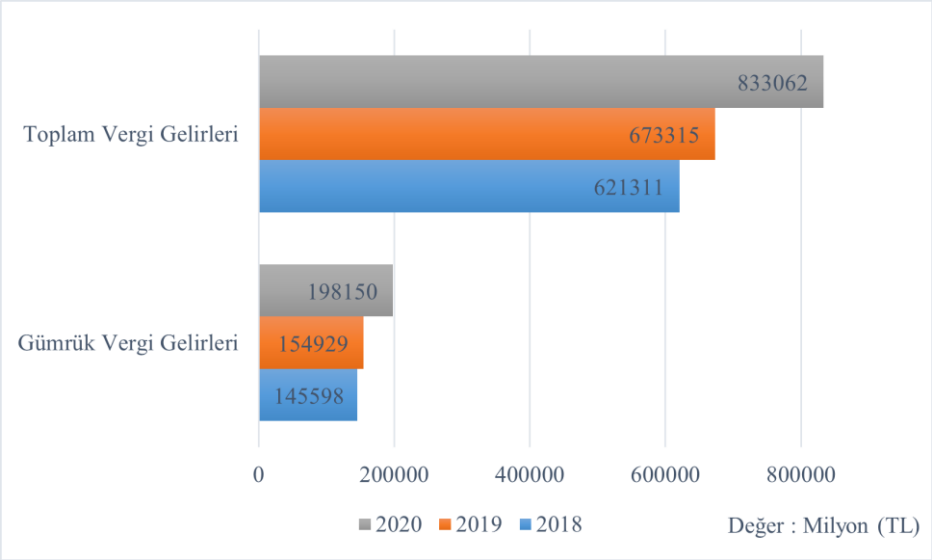
Şekil 2. 2019-2021 Yılları Beyanname Sayıları

Bugün birçok firma ithalat-ihracat yapmaktadır ve sayıları her geçen gün artmaktadır. Ülkemizde üretilen ürünlerin diğer ülkelere pazarlanması ve satılması ayrıca iç piyasanın ihtiyacı olan ürünlerin temininin yapılması için bu firmaların faaliyetleri çok önemlidir. Ticaret Bakanlığı tarafından yayınlanan 2019-2021 yılları arasında faaliyet gösteren ihracat ve ithalat yapan firma sayıları (T.C. Ticaret Bakalığı, 2021) Şekil 3'te sunulmaktadır. Bu verilere göre firma sayılarında artış olduğu yıllar bazında görülmektedir.



Şekil 3. 2019-2021 İhracat ve İthalat Yapan Firma Sayıları

İthalat-ihracat işlemlerinden alınan vergiler de işin önemli bir tarafıdır. Yapılan bu işlemlerden alınan vergiler devlet için ciddi bir gelir ortaya çıkarmaktadır. Kamu gelirleri içinde ithalat ve ihracat işlemlerinden alınan vergiler çok önemli bir yer tutmaktadır. Yıllık bazdaki rakamlara bakıldığında devletin ciddi bir gelir elde ettiği görülmektedir. Şekil 4'te (T.C. Ticaret Bakanlığı, 2021) 2018, 2019 ve 2020 yıllarında tahsil edilen toplam ve gümrük vergi gelirleri gösterilmektedir. Gümrük vergi gelirlerinin toplam vergi gelirleri içindeki payı 2018 yılında %23,4, 2019 yılında %23 ve 2020 yılında %23,8 olmuştur.



Şekil 4. Gümrük Vergi Gelirlerinin Toplam Vergi Gelirleri İçindeki Payı

Gümrük beyannamesinin hazırlanması, üzerinde titizlikle durulması gereken ve ciddi anlamda yetkin iş gücü gerektiren bir süreçtir. Gümrük beyannamesi birçok bileşenden oluşan karmaşık bir yapıya sahiptir. Bu süreç önemli derecede mevzuat bilgisi gerektirmektedir. Mevzuatlar ise her geçen gün yenilenen ve dikkatli şekilde takip edilmesi gereken bir özelliktedir. İşin karmaşıklığı dışında yetkin ve yeterli insan kaynağının olmaması beyanname yazımında istenmeyen hataların yapılmasına sebep olmaktadır. Bu hatalar ciddi cezalarla karşı karşıya kalınmasına sebep olmaktadır. Bu durum operasyon maliyetlerini ve süreçlerini olumsuz etkilemektedir. Ayrıca hukuki sonuçlar da doğurabilmektedir.

Makine öğrenmesi algoritmaları son yıllarda günlük yaşamımızın bir parçası haline geldi. Birçok alanda makine öğrenmesi algoritmaları kullanılmakta ve çok sayıda platformda karar vermemize yardımcı olarak karşımıza çıkmaktadır. Hatta bu platformlar çoklu makine öğrenmesi uygulaması içermektedir. Hangi ürünü satın alacağımızdan yazdığımız bir mesajda sonraki kelimeye kadar hayatımızın her alanında artık bize önerilerde bulunulmaktadır. Ticari uygulamaların dışında çeşitli bilim dallarında da makine öğrenmesi algoritmaları kullanılmaktadır.

Teknolojinin gelişmesiyle üretilen veride büyük bir artış oldu. Büyük veri denilen kavram ortaya çıktı ve bu büyük verinin yönetilmesi çok önemli bir hale geldi. Gelişmek ve

rekabet etmek için bireyler, kurumlar ve ülkeler bu büyük veriyi değerlendirmeye çalışmaktadırlar (Akter, Wamba, Gunasekaran, Dubey, ve Childe, 2016). Bu büyük veriyi değerlendirmek için çeşitli yöntemler geliştirildi. Makine öğrenmesi bu yöntemlerden birisi olup veriden bilgiye ulaşmaya çalışan bir bilimdir. Makine öğrenmesi verileri kullanarak akıllı aletlerin bir karar verme problemini kendi kendine çözmesine yardımcı olur. Yapay zekâ ve makine öğrenmesi günümüzde son derece popüler bir alandır. Makine öğrenimi, derin öğrenme algoritmaları kullanılarak insan zekâsını taklit eden bir bilgisayar teknolojisidir (Şahinaslan, Dalyan, ve Şahinaslan, 2022). Makine öğrenmesi aynı zamanda yapay zekânın bir alt dalı olup istatistik, yapay zeka ve bilgisayar bilimleriyle iç içedir.

Gümrük beyannamelerinin hatasız ve doğru oluşturulabilmesi için insan müdahalesinin asgari seviyeye çekilmesi ve olumsuz sonuçların giderilerek operasyonel süreçlerin iyileştirilmesi gerekmektedir. Bu da mevcut büyük verilerden yararlanılarak yapılabilecektir. Gümrük beyannamesinin bileşenlerinden veri setleri oluşturularak bu veri setleri modellenecektir. Modellenen bu veri setleri makine öğrenmesi ve öğrenme algoritmaları ile çalışılıp kullanıcıya her bileşen için öneride bulunulacaktır.

Kullanıcının son kontrolünden sonra onayı alınarak beyanname işlemi sonuçlandırılacaktır. Bu çalışmada hatasız beyanname yazımına destek olacak bir sistem kurmaya çalışılmıştır. Böylece insan kaynaklı operasyonel hataların tespiti yapılarak bu hata oranlarının minimum seviyeye çekilmesi sağlanacaktır. Bu hataları minimum seviyeye çekmek için büyük verilerden faydalanılarak makine öğrenmesi ve derin öğrenme algoritmaları ile hata tespit edilmeye çalışılmıştır. Makine öğrenmesi algoritmaları benzer birçok alanda kullanılmıştır ve başarılı sonuçlar elde edildiği görülmüştür.

Hatasız beyanname yazımı büyük önem taşımaktadır çünkü yapılan hatalar ciddi sorunlara sebep olmaktadır. Bu olumsuz sonuçların giderilmesi, operasyonel süreçlerin otomatize edilmesi, insan müdahalesinin asgari seviyeye çekilmesi, beyannamelerin hatasız ve doğru oluşturulması için bu çalışma yapılmıştır.

Literatür Taraması

Gümrük beyannamesinin makine öğrenme algoritmaları ile oluşturulmasının dış ticaret hizmetlerinin daha başarılı bir şekilde gerçekleştirilmesini sağlayacağını öngörmekteyiz. Bu bölümde öğrenme algoritmaları tabanlı gümrük beyannamelerinin oluşturulması için yapılan çalışmalar listelenmektedir. Önerdiğimiz çözümün ilgili çalışmalarla olan ilişkisi detaylı bir şekilde açıklanmıştır.

Günerkan ve arkadaşları yaptıkları çalışmada (Günerkan, Şahinaslan, ve Şahinaslan, 2022), gümrük beyannamesi oluşturma sürecinde makine öğrenme algoritmalarının etkinliğini incelemişlerdir ve beyanname oluşturma sürecinde hataların tespiti için kullanımı konusunda olumlu sonuçlar elde etmişlerdir.

İşler yaptığı çalışmada (İşler, 2018), idari para cezaları ile adli para cezalarını karşılaştırmıştır. İdari para cezalarında adli para cezalarında olduğu gibi yargı kararı gerekmediğini, idari kurumlar tarafından verildiğini ve tahsil edildiğini belirtmiştir. Ayrıca kamu düzenini sağlamak amacıyla uygulanan para cezalarının asıl amacının gelir sağlamaktan önce toplumsal düzeni, adaleti ve güvenliği sağlamak olduğunu vurgulamıştır. Böyle olmakla birlikte kamu gelirleri altında ceza gelirleri bulunduğunu söylemiştir. Olağanüstü kamu gelirleri arasında bulunan bu gelirler için bütçe tahminlerinde hedeflerin mevcut olması asıl hedefin yanında kamu geliri hedefinin de olduğu anlamına geldiğini belirtmiştir. Ceza gelirlerinin miktarının birçok vergiden daha fazla gelir sağladığının görüldüğünü tespit etmiştir. Bu gelirlerin merkezi yönetim bütçe açığına oranlarının son yıllarda %24 ile %35 arasında oluştuğunu belirterek sosyal düzen amacı ile toplanan para cezalarının kamu bütçesi için önemli bir gelir kalemi olduğunu söylemiştir.

Canrakerta ve arkadaşları yaptıkları çalışmada (Canrakerta, 2020), gümrük beyannamesinin belge denetçisi tarafından yeniden incelenmesi gerektiğini belirtmiştir. Gümrük beyannamesinin potansiyel hatalara sahip olabileceği belirtilmiştir. Bu hataların kasıtlı da olabileceğine değinilmiştir. İş zekası kullanılarak denetçiler için analiz imkanları sağlanarak sorun çözülmeye çalışılmıştır. Kimball metodolojisi ile veri ambarı geliştirmeye odaklanılmıştır. Araştırma sonuçları veri ambarı tasarımı şeklinde yapılmıştır. Veri ambarı daha sonra kontrol paneli, OLAP (Online Analytical Processing)

ve veri madenciliği için kullanılarak hata veya sahtekârlığın tespit edilmesi amaçlanmıştır. Karar ağacı, destek vektör makinesi, sinirsel ağ ve çeşitli topluluk yöntemleri gibi veri madenciliği algoritmalarını kullanmışlardır. Çalışma sonucunda SMOTE tekniğinin (Synthetic Minority Oversampling Technique), SMOTE olmayan tekniğe göre anlamlı derecede artmış bir duyarlılık skoruna sahip olduğunu tespit etmişlerdir.

Shao ve arkadaşları yaptıkları çalışmada (Shao, 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 kullanılabileceğini belirtmişlerdir.

Maruev ve arkadaşları yaptıkları çalışmada (Maruev, Stefanovskiy, Frolov, Troussov, ve Curry, 2014), giderek küreselleşen dünyamızda, uluslararası ticaretin önündeki engellerin incelenmesinin uluslararası ekonomi alanını ilgilendirdiğini belirtmişlerdir. Çalışmalarında gümrük beyannamelerinin hızlı ve doğru bir şekilde işlenmesi sorununa odaklanıp denetimli öğrenmeye dayalı ticari mallar için müşteri beyanlarının otomatik olarak işlenmesi için grafik tabanlı yayma etkinleştirme algoritmasının yeni bir kullanımını sunmuşlardır. Bu çalışmayla gümrük memurları, tüccarlar, taşıyıcılar ve sigortacılar tarafından kullanılmak üzere tavsiye sistemleri oluşturmuşlardır. Bu algoritma risk tabanlı geleneksel risk göstergelerinden ziyade yalnızca sevkiyat verileri üzerine eğitilmiştir. Bunun, gümrük yetkililerine tavsiyesinin 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.

Kaya ve Doğan yaptıkları çalışmada (Kaya ve Doğan, 2020), vergi tarife ve esaslarının, gümrük vergisinin tespitindeki parametrelerden sadece bir tanesi olduğunu belirtmişlerdir. Hesaplama bu tarife ve esaslara göre yapılsa da asıl belirleyici bilginin eşyanın gümrük kıymeti olduğunu belirtmişlerdir. Eşyanın gümrük kıymetinin bütün gümrük vergilerinin hesaplanmasında önemli olduğunu vurgulamışlardır. Bu açıdan, gümrük vergilerinin doğru hesaplanabilmesi ve tahsilatlarının eksiksiz yapılabilmesi için gümrük kıymetinin doğru tespitinin çok önemli olduğunu belirtmişlerdir.

Büyükgüral ve Türkoğlu yaptıkları çalışmada (Büyükgüral ve Türkoğlu, 2015), zaman aşımı sürelerinin 5326 sayılı Kabahatler Kanunu ile 4458 sayılı Gümrük Kanunu'nda farklı olduğunu belirtip, bunun Gümrük Kanunu'na göre uygulanacak para cezalarında zaman aşımı konusunda tereddütler yaşattığını belirtmiştir. Uygulanacak para cezalarında zaman aşımına ilişkin sürenin önemine değinerek çalışmalarında bu tereddütlerin giderilmesini amaçlamışlar ve gümrük vergileri nedeniyle oluşan para cezalarında zaman aşımı sürelerindeki tereddütleri ve bu tereddütlerin nasıl giderileceğini açıklamaya çalışmışlardır.

Özer yaptığı çalışmada (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 işlediğine ve ne gibi katkılar sağladığına değinmiştir.

İlgün yaptığı çalışmada (İlgün, 2020), vergi denetiminde büyük veri analitiğinin faydalarını ve zararlarını araştırmıştır. Büyük veri analitiğinin kullanılmasının bir tercih değil de kaçınılmaz bir sonuç olacağını vurgulamıştır. Gerekli yapısal dönüşüm ve altyapı çalışmalarının yeni teknolojilere uyum ve sonuçları konusundan bahsetmiştir.

Garcia ve Caballero yaptıkları çalışmada (García ve Caballero, 2021), birçok ülke için kritik bir konu olan gümrük dolandırıcılık kontrolünün optimizasyonunun ekonomik açıdan önemli bir problemini 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.

Ryzhova ve Sochenkov yaptıkları çalışmada (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 ile en iyi sonuca ulaşmışlardır.

Barua ve arkadaşları yaptıkları çalışmada (Barua, Zou, ve Zhou, 2020), uluslararası yük taşımacılığında makine öğrenmesi modelleri geliştirerek talep tahmini, operasyon ve varlık bakımı, araç güzergâhı ve zamanında teslimat performansı tahmini konularında kullanımını tartışmaktadırlar.

Li ve Li yaptıkları çalışmada (Li ve Li, 2019), hızla artan uluslararası ticaret miktarı ışığında, 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 evrişimli 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.

Mammadov yaptığı çalışmada (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 simülasyonlarını gerçekleştirmişlerdir.

Paula ve arkadaşları yaptıkları çalışmada (Paula, Ladeira, Carvalho, ve Marzagao, 2016), Brezilyalı ihracatçıları ihracatta dolandırıcılık yapma olasılığına göre sınıflandırmak için denetimsiz Derin Öğrenme modelinin uygulanmasında 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'ı kullanmışlardır. Çalışmalarını Brezilya Federal Gelir Sekreterliği tarafından sağlanan 2014 yılında Brezilya'da gerçekleşen mal ve ürün ihracatını verilerine göre yapmışlardır. İhracat şirketlerini karakterize eden niteliklerden kurdukları model ile en az yirmi ihracatçıdaki anormallikleri tespit edebilmişlerdir.

Günerkan ve arkadaşları yaptıkları çalışmada (Günerkan, Şahinaslan, ve Şahinaslan, 2022), oluşturdukları modelleri üzerinde Gaussian Naïve Bayes, Decision Tree, K-Nearest Neighbors, Random Forest, Ridge, Bagging, AdaBoost ve Multi-Layer Perceptron sınıflandırma algoritmalarını kullanarak %75,70 ile %99,70 arasında başarım oranları elde etmişlerdir.

Makine öğrenmesi ve derin öğrenme algoritmaları ile sağlık alanında da ciddi çalışmalar yapılmış ve hastalıkların teşhisi konusunda başarılı sonuçlar elde edilmiştir (Pamuk ve Kaya, 2021), (Akgül, Çelik, Aydın, ve Öztürk, 2020), (Saygın ve Baykara, 2021), (Gulia, Vohra, ve Rani, 2014), (Coşar ve Deniz, 2021) ve (Başer, Yangın, ve Sarıdaş, 2021). Pazarlama alanında ise tüketicilerin alışveriş davranışlarını belirlemek ve bu davranışlara ilişkin yeni öngörüler kazanmak, karar verme sürecine ışık tutmak, insan içgörüleriyle Pazarlama teorilerini ilişkilendirmek, Pazarlama alanının gelişen doğasını analiz etmek ve Pazarlama operasyonlarının performansını iyileştirmek için çalışmalar yapılmıştır (Sundsøy, Bjelland, Iqbal, “Sandy” Pentland, ve de Montjoye, 2014), (Siau ve Yang, 2017), (Ma ve Sun, 2020), (Cui, Wong, ve Lui, 2006), (Brei, 2020) ve (Cui ve Curry, 2005).

Veri giriş mekanizması, kullanıcı etkileşimi için ilk arayüz olduğu için insan-bilgisayar etkileşimi alanında yaygınlaşmaktadır (Salve, Bhutkar, ve Yammiyavar, 2021). İnsanların veri girişleri hatalı olabilir ve bu hatalı girişler olumsuz sonuçlara sebep olabilir (Barchard ve Pace, 2011). Veri kalitesi, modern veri tabanlarında kritik bir sorundur (Chen, Chen, Conway, Hellerstein, ve Parikh, 2011). Veri giriş hatalarını tespit etmek, azaltmak ve veri kalitesini iyileştirmek için otomatik yöntemler konusunda çok az araştırma yapılmıştır (Chen, Chen, Conway, Hellerstein, ve Parikh, 2011). Bu nedenle, araştırmacılar, veri girişini en aza indirecek çalışmaları yapmalı ve uygulamalılar (Barchard ve Pace, 2011).

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).

Makine öğrenmesi algoritmaları birçok alanda kullanılmıştır ve başarılı sonuçlar elde edilmiştir. Bu çalışmalar ışığında hatasız beyanname yazımına destek olacak bir sistem kurulmasına çalışılmıştır. Hatasız beyanname yazımı büyük önem taşımaktadır çünkü yapılan hatalar ciddi sorunlara sebep olmaktadır. Yapılan hataların olumsuz sonuçlarının giderilmesi, operasyonel süreçlerin otomatize edilmesi, insan müdahalesinin asgari seviyeye çekilmesi, beyannamelerin hatasız ve doğru oluşturulması için bu çalışma yapılmıştır.

Problem

Gümrük sisteminde hatasız beyanname yazımı büyük önem taşımaktadır. Yapılan hatalar olumsuz sonuçlara sebep olmaktadır. Özellikle yetkin ve yeterli insan kaynağının olmamasından dolayı beyanname yazımlarında hatalar yapılmaktadır. Ülkemizde beyannamelerin hatalı oluşturulmasından dolayı ciddi miktarlarda ceza ödenmektedir. Bu durum firmaların operasyon maliyetlerini ve süreçlerini olumsuz etkilemektedir.

Amaç

Yapılan hataların hukuki ve maddi sonuçları bu konuda yeni çalışmalar yapmayı gerekli kılmaktadır. Bu çalışma ile operasyon verimliliğinin artırılması, maliyetlerin düşürülmesi hedeflenmektedir. Mevcut büyük veri ve teknolojik imkânlar bu çalışmanın amacına ulaşmasında katkı sunmaktadır. Bu çalışma aşağıdaki temel hedefleri gerçekleştirme amacıyla yapılmıştır.

- Gümrük beyannamelerinin zamanında oluşmasının sağlanması ve operasyonların aksamaması,
- Gümrük beyannamelerinin daha doğru yazılması sağlanarak karşılaşılabilecek ciddi miktarlarda cezaların önüne geçilmesi,

- Operasyonel maliyet ve süreçlerinde iyileştirilmenin sağlanması,
- Gümrük beyannamelerinin oluşturulmasında en önemli parametrelerden bir tanesi olan yetkin ve yeterli personel eksikliğinin ortaya çıkardığı problemlerin azaltılması,
- Subjektif yaklaşımların doğurduğu olumsuzlukların azaltılması,
- Gümrük beyannameleri oluşturulurken yapılan hataların telafisi için harcanan kaynakların azaltılması,
- Hatalı gümrük beyannamesi oluşturulmasından dolayı doğan hukuki maliyetlerin azaltılması ve hukuki sonuçların önüne geçilmesi,
- Gümrük sisteminin tam entegre ve bir bütün olarak hareket etmesi,
- Hatasız beyanname yazım hedefinin gerçekleştirilmesi.

Önem

Gümrük beyannamesi birçok bileşenden oluşan karmaşık bir yapıya sahiptir. Elimizdeki büyük verileri kullanarak yapılacak doğruluk oranı yüksek tahminlemeler sayesinde hatalar zamanında tespit edilip olası olumsuzlukları giderilecektir.

Varsayımlar

Beyanname üzerindeki her bir bileşen için hatalı veri girişi yapılabileceği öngörülmektedir. Doğru veri setleri oluşturarak eğitim sonuçlarında belli bir başarım oranına ulaşılması hedeflenmektedir.

Sınırlıklar

Türkiye dış ticaretinin %8'ini yapan uluslararası bir lojistik firmasının 2010 ve 2021 yılları arasındaki 12 yıllık beyanname verileri ile çalışma yapılmıştır. Çalışmada kategorik veriler tahmin edildiği için sekiz tane sınıflandırma makine öğrenme algoritması kullanılmıştır.

Tanımlar

Bu bölümde Türkiye gümrük sistemi, gümrük beyannamesi, delphi metodu, makine öğrenmesi, öğrenme yaklaşımları, makine öğrenme algoritma türleri ve makine öğrenme algoritmaları ile ilgili bilgilere yer verilmiştir.

Gümrük Kavramı

Kelime itibariyle yer, işlem ve kurum anlamlarında kullanılan Gümrük kavramı, Türk Dil Kurumu Türkçe Sözlükte (Türk Dil Kurumu, 2022) şu şekillerde tanımlanmaktadır:

- Bir ülkeye giren veya ülkeden çıkan mal ve eşya üzerinden alınan vergi.
- Bir verginin alınması işlemiyle uğraşan devlet kuruluşu.
- Bir ülkenin giriş ve çıkışında gümrük denetim ve gözetiminin yapıldığı yer.

Bu tanımlardan hareketle, gümrük kavramının bir ülkeye mal giriş çıkışı ile ilgili düzenlemeler ve bu düzenlemeler doğrultusundaki uygulamalar olduğu söylenebilmektedir. Yapılan bu uygulamalar ülkelerin ilgili kanun/kanunları doğrultusunda hayata geçirilir. Bu noktada kanunlar, ortak bir dil bütünlüğü sağlamak adına yaptıkları düzenlemede geçen terimlerde kişiye göre yorumu ortadan kaldırmak ve işleyişte buna bağlı aksaklıkları bertaraf etmek için herkesçe aynı şekilde anlaşılacak biçimde, konuyla ilgili olan kavramları tanımlamaktadır. Ülkemizde gümrük işleyişini düzenleyen 4458 sayılı ve 5 Şubat 2000 tarihinde yürürlüğe giren Gümrük Kanunu, çeşitli konulara yönelik genelgeler ve tebliğlerde işleyişte karşımıza çıkan kavramlar tanımlanmıştır.

Gümrük Mevzuatı

Mevzuat kavramı, kamu kurum ve kuruluşları için değerlendirildiğinde, ilgili kurumun tüm işleyiş esaslarını ortaya koyan ve yazılı olarak kanun ya da kanuna bağlı kararname şeklinde düzenlenen kurallar bütünlüğüdür. Bu tanımdan hareketle Türk Gümrük Sistemi göz önüne alındığında, gümrüklerin ve gümrükleme işlemleri ile ilgili yasal düzenlemelerin bütününü gümrük mevzuatı olarak ifade etmek mümkündür.

Her ülkede dış ticaretin belli kuralları vardır. Dış ticaret bu kurallar zinciri ile denetlenmekte ve yönlendirilmektedir. Dış ticaret üzerindeki denetimde ana amaç sadece

vergi almak değildir. Dış ticaret üzerindeki bu denetim ile ülke ekonomisinin geliştirilmesi ve doğru yönlendirilmesi hedeflenmektedir. Gümrüklerde alınan vergiler ile devlet bir gelir elde etse de ana amaç ülke ekonomisini geliştirmek ve doğru yönlendirmektir.

Dış ticaretimizde belirleyici, yönlendirici, kısıtlayıcı ve yasaklayıcı hükümler yetkisine sahip kuruluşlar şunlardır: Dış Ticaret ve Hazine Müsteşarlıkları, Maliye, Tarım, Sağlık, Sanayi, Ulaştırma, Çevre, Kültür, İçişleri, Dışişleri ve Milli Savunma Bakanlıkları, Devlet Planlama Teşkilatı, Devlet İstatistik Enstitüsü. Gümrük Müsteşarlığı ise dış ticaretimizle ilgili mevzuatların uygulayıcısı konumundadır.

Ülkemizde dış ticaret belli mevzuatlar çerçevesinde yapılmaktadır. İthalat ve ihracat mevzuatı, kambiyo mevzuatı ve gümrük mevzuatı dış ticaretin düzenlenmesinde ve gerçekleştirilmesinde kullanılmaktadır.

Dış Ticaret Mevzuatı; bir Malın Türkiye Gümrük Bölgesine girebilmesine ve Türkiye Gümrük Bölgesinden çıkabilmesine ilişkin hükümler koyar. Kambiyo Mevzuatı; 1567 sayılı TPKK (Türk Parası Kıymetini Koruma Kanunu) da belirtilen kuralların uygulanmasıdır.

Gümrük Mevzuatı; Türkiye Gümrük Bölgesine giren ve Türkiye Gümrük Bölgesinden çıkan bir mal ile ilgili olarak devlet kuruluşlarının aldığı her tedbirin tam olarak uygulanmasını amaçlayan bir denetim sistemini içermektedir.

Gümrük Kanunu

1615 sayılı eski Gümrük Kanunu 1972 tarihinde yürürlüğe girmiştir ve 28 yıl yürürlükte kalmıştır. Gerek gümrük birliği süreci gerekse eski kanunun ihtiyaçları tam olarak karşılayamaması, gümrük uygulamalarında bir reform ihtiyacı hissettirmiştir. Bu reform 2000 yılında hayata geçirilmiş ve bu süreçte, Gümrük Kanunu'nda 27.10.1999 tarih ve 4458 sayılı yeni kanun 5 Şubat 2000 tarihi itibariyle yürürlüğe konmuştur. Böylelikle Türkiye'de gümrük ve gümrük işlemleri konusunda yeni bir dönem başlamıştır.

4458 sayılı yeni kanun 13 kısım, 247 esas ve 6 geçici maddeden oluşmuştur. Bu yeni kanunun iki hedefi vardır. Bunlardan birisi AB uyum sürecinde gümrük mevzuatının da AB şartları ile uyum içerisinde olmasını sağlamaktır ve ikincisi de Türkiye'de daha etkili,

daha işlevsel bir gümrük rejimini hayata geçirmektir. Sözü edilen bu yasa ile Türkiye'de yeni bir gümrük düzeni kurulmuş, gümrük yükümlülüğünün sınırları çizilmiş, gümrük idaresine yeni görevler verilmiş, gümrükte iş takibi bir düzene alınmış ve ceza uygulamaları ve hükümleri yeniden revize edilmiştir.

Gümrük

Ülke sınırlarından eşya ve malların giriş ve çıkışlarında denetimlerinin yapıldığı ve vergilerinin ödendiği kamu kurumudur. Tablo 3'te İstanbul bölgesine ait gümrük müdürlükleri listelenmiştir. Altı karakter ve rakamlardan oluşan gümrük kodu il plaka kodu ile başlamaktadır.

Gümrükleme

Ülkeler arasında gerçekleştirilen dış ticaret işlemlerinde, mal veya hizmetlerin gümrük bölgelerine girişi veya çıkışında uygulanacak zorunlu mevzuat işlemlerinin gerçekleştirilmesi gümrükleme kapsamına girmektedir. Gümrük bölgesi ülkenin topraklarını kapsar. Gümrük bölgesine kara suları, iç suları ve hava sahası da girmektedir.

Gümrük Müşaviri

27.10.1999 tarihli ve 4458 sayılı Gümrük Kanunu ile dış ticaret işlemlerini müşavirlik hizmeti kapsamında temsilci olarak yürüten meslektir. Dış ticaret ve gümrük işlemleri konularında yazılı görüş verebilirler, hakemlik veya bilirkişilik yapabilirler.

Tablo 3. İstanbul Bölgesi Gümrük Müdürlükleri

Gümrük Kodu	Gümrük Adı
340100	A.H.L. Yolcu Salonu Gümrük Müdürlüğü
340300	A.H.L. Kargo Gümrük Müdürlüğü
340600	İstanbul Posta Gümrük Müdürlüğü
340700	Beylikdüzü Akaryakıt Gümrük Müdürlüğü
340800	Karaköy Yolcu salonu Gümrük Müdürlüğü
340900	Haydarpaşa Gümrük Müdürlüğü
341200	Erenköy Gümrük Müdürlüğü
341300	Muratbey Gümrük Müdürlüğü
341453	İstanbul Havalimanı Gümrük Müdürlüğü
342000	Trakya Serbest Bölge Gümrük Müdürlüğü
342200	İstanbul Deri Serbest Bölge Gümrük
342400	Yeşilköy Gümrük Müdürlüğü
342800	A.H.L. Serbest Bölge Gümrük Müdürlüğü
343100	Ambarlı Gümrük Müdürlüğü
343200	Sabiha Gökçen Havalimanı Gümrük Müdürlüğü
343300	Zeytinburnu Gümrük Müdürlüğü
343400	Pendik Gümrük Müdürlüğü
343500	Halkalı Gar Gümrük Müdürlüğü

Gümrüklerde Beyanların Kontrol Usulleri

Gümrük işlemi uygulanan eşya için yapılan beyanın doğruluğunun kontrol edilmesi gümrük işlemlerinin önemli bir aşamasıdır. Gümrük yönetmeliğinin 180. maddesinde eşya için yapılan beyanın doğruluğunun kontrolü aşamasında, eşyanın muayene edilip edilmeyeceği, muayene edilecekse ne kadar eşyanın muayene edileceği gibi hususlar belirtilmiştir. Dört çeşit muayene hat kriteri bulunmaktadır.

Kırmızı Hat

Merkezdeki risk analizleri çalışmaları sonucunda BİLGE üzerinden sisteme tanıtılan kriterlere takılan sevkiyatların gümrük muayene memuru tarafından fiziki muayeneye tabi tutulduğu hattır. Bir beyanname dört sebeple kırmızı hatta yönlendirilebilir. Bunlar:

- Merkezdeki risk analizi biriminin sisteme tanıttığı profiller uyarınca;
- Gümrük mevzuatından kaynaklanan zorunlu yönlendirmeler. (Halihazırdaki uygulamada, kırmızı hat oranını artıran ve risk profilleriyle ilgisi bulunmayan bu kriter kırmızı hattan çıkarılmıştır.)
- Merkezdeki risk analizi biriminin belirlediği rastgele seçim kriteri sonucu;
- Gümrük muayene memurunun şüphe hâsıl olması üzerine şerh düşerek beyannameyi kırmızı hatta yönlendirmesidir.

Kırmızı hatta düşen eşyanın kontrolü mutlak suretle eşyanın görülmesiyle gerçekleştirilmektedir. Kırmızı hat kontrolü, ilgili gümrük muayene memurunca üç şekilde yapılabilmektedir. Haricen muayene; yani eşyanın dıştan görülmesi ve beyan ile uyuştuğunun tespit edilmesi halinde daha ileri gidilmemesi. Kısmi muayene; eşyanın bir kısmı itibariyle kapların açılması ve beyan ile uygunluğunun sağlanması halinde diğer kapları açılmaması. Tam muayene; eşyanın bütün kaplarının açılması suretiyle fiziki kontrolünün yapılması.

Sarı Hat

Fiziki muayeneye gerek görülmeksizin eşyaya ilişkin beyanname ile eklerinin doğruluğu ve birbiriyle uygunluğunun "sarı hat muayene memuru" tarafından kontrol edildiği hattır. Ancak belgelere göre tespitin yapılması mümkün bulunmuyorsa veya uygunsuzluk tespit edilirse idare amirinin uygun görmesi ile eşyanın fiziki kontrole tabi tutulması mümkün bulunmaktadır.

Yeşil Hat

Bu hat özelliği veya statüsü gereği bazı kişi veya eşyaya uygulanan hattır. Burada eşyanın hiçbir şekilde kontrolü yapılmaz. Çok spesifik, bekletilmesine taviz verilmesi mümkün olmayan eşya grubu için uygulanmaktadır.

Mavi Hat

Basitleştirilmiş Usul ve Sonradan Kontrol birimi tarafından, Risk Analizi Birimi verilerinden yararlanarak (örn: ihlallere ilişkin kayıtların incelenmesi) Gümrük mevzuatı çerçevesinde belirlenmiş çeşitli kriterleri sağlayan ve kendilerine "Onaylanmış Kişi Statü Belgesi" verilen firmalar vergi numaralarına göre bilgisayar sistemine tanıtılmakta ve bu hatta giren firma eşyasının gümrük muayenesi yapılmamaktadır. Bu hak firma bazında verilmektedir. Eşyanın kontrolleri bu hatta daha sonraki aşamada belge üzerinden yapılmakta, gerek görülürse firma kayıtlarına da gidilebilmektedir. Ancak Türk Gümrük idaresinin risk analizi kriterleri çerçevesinde bu firmaları da kırmızı hattan takip hakkı saklıdır.

Gümrük Tarife İstatistik Pozisyonu

Gümrük Tarife İstatistik Pozisyonunun kısaltması olarak kullanılan GTİP, Gümrük Tarife Cetveli üzerinde yer alan 12 haneli kodu temsil etmektedir. Uluslararası olarak aynı kriterlere dayanarak hazırlanan bu kodlar sayesinde taşınan mal Armonize Sisteme kaydedilmekte ve malın tanımı yapılmaktadır.

GTİP (Gümrük Tarife İstatistik Pozisyonu) numarası, gümrük tarife cetveli üzerinde yer alan koda verilen isimdir. Oluşturulan bu kod belirli standartlar çerçevesinde hazırlanmakta ve toplamda 12 rakam içermektedir. GTİP numaraları eşyanın pozisyonuna, eşyaya uygulanan çeşitli vergi uygulamalarına ve Dünya Gümrük Örgütü ya da Avrupa Birliği ülkelerinin kullandığı Kombine Nomanklatür kodları ile belirlenmektedir.

GTİP numaraları, Armonize Mal Tanımı ve Kodlama Sistemi kapsamında düzenlenmektedir. Tüm dünya genelinde kullanılan GTİP numaraları ve tarife cetvelleri de bu Armonize Sistem çerçevesinde oluşturulmaktadır. Yurtdışı ile yapılan eşya transferleri için hazırlanan tarife cetveli de bu esaslara dayanarak hazırlanmalıdır.

GTİP numaraları, eşyanın dâhil olduğu gruplara göre farklı rakamlardan oluşmaktadır. Eşyanın dâhil olduğu her ürün grubu için farklı GTİP kodları kullanılır. Herhangi bir ülkenin gümrüğüne giriş yapan eşya ise bu kodlar üzerinden işlem görerek ülkeye alınır. GTİP numaralarında yapılan yanlışlıklar ve ürünün farklı bir kategoriye kaydının yapılması ise ürün alımlarında çeşitli sorunlara yol açacağı için, GTİP hazırlanırken tüm detaylar dikkatli şekilde incelenmelidir.

Özellikle yurtdışı ile çalışan ve uluslararası alanda taşımacılık yapan firmalar için en büyük sorulardan bir tanesi de "GTİP kodu nedir?" sorusudur. GTİP kodu, tarife cetveli üzerinde ürünün dâhil olduğu gruptan, taşımacılığın yapıldığı ülke standartlarına kadar birçok detayı içeren kod olarak karşımıza çıkıyor. GTİP kodları üzerinde yer alan rakamlar da bu detayları belirtecek şekilde farklı kategorilere ayrılmıştır. Tarife cetveli (Armonize Sistem) toplamda 21 bölüme ve 96 fasıla ayrılmaktadır. Fasıllar ise 2'li kodlara, 4'lü pozisyonlara ve 6'lı alt pozisyonlara ayrılmaktadır.

Tüm dünya genelinde kullanılan kod, pozisyon ve alt pozisyon kodları birbiri ile aynıdır. Bu sayede taşımacılık yapılan ürünün kodlamasında ürünler uluslararası olarak tüm ülkelerde anlaşılır hale gelir. Kodların tüm dünyada aynı şekilde hazırlanması, belirtilen ürünlerin birbiri ile karışmasının önüne geçilmesini sağlar. GTİP koduna örnek verecek olursak, ülkemizde elmanın GTİP kodu 08.08.10.80.00.19 olarak kabul edilir.

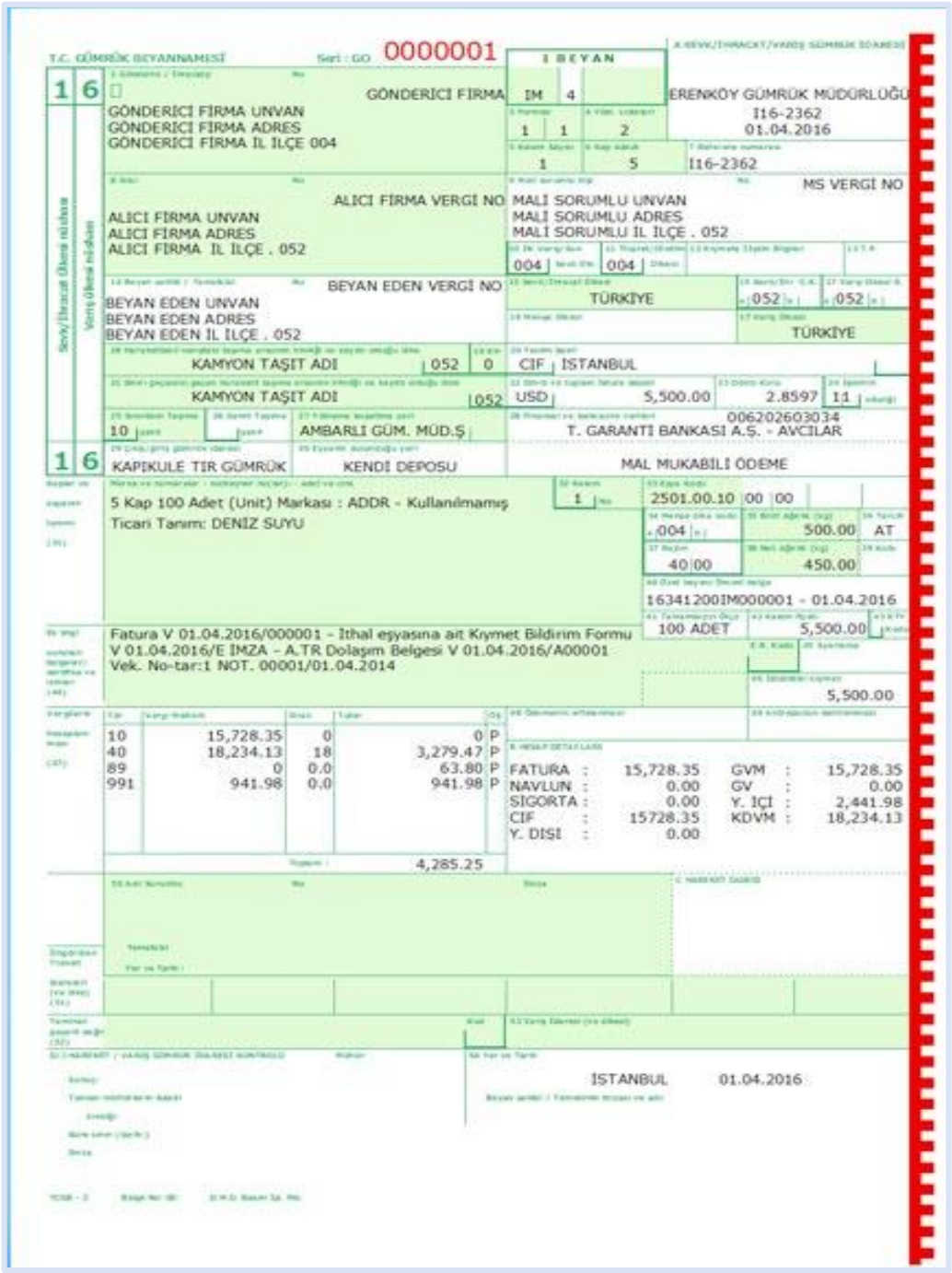
12'li kodlama sisteminde farklı kod grupları ürün hakkında farklı detayların belirtilmesinde kullanılır. Temel olarak bakacak olursak;

- İlk 4 rakam eşyanın pozisyon numarasını belirtmekte kullanılır,
- İlk 6 rakam ise eşya pozisyon numarası ile birlikte Dünya Gümrük Örgütü tarafından belirtilen ve tüm dünya ülkelerinde kullanılan Armonize Sistem Nomanklatür kodunu içerir,
- 7 ve 8. haneleri Avrupa Ülkeleri tarafından belirtilen Kombine Nomanklatür kodunu ifade eder,
- 9 ve 10. haneler, ifade edilen mala uygulanan farklı vergi uygulamalarını ifade etmek için kullanılır. Ürüne farklı bir vergi uygulaması yapılmıyor ise GTİP kodunda bu alan 00 ile belirtilir.
- 11 ve 12. haneler ise Gümrük Tarife İstatistik kodunu ifade etmek için kullanılır.

Gümrük Beyannamesi

Gümrük beyannamesi düzenlenerek yazılı gümrük beyanı yapılır. Gümrük beyannamesinin yapısı, hangi bilgilerden oluşacağı ve nasıl hazırlanacağı gümrük mevzuatı ile belirlenir. Gümrük beyannamesi birçok bilgi içerir. Örneğin gönderici-alıcı firma bilgileri, tarife, taşıma bilgileri, kıymet, menşe, mevcut ise ödenecek vergiler, eşya

için gerekli belgeler gibi bilgileri içerecektir. Avrupa Birliği üyesi ülkelerde kullanılmakta olan gümrük beyannamesi formu 1 Ocak 1996 tarihinden beri ülkemizde de kullanılmaktadır (Gümrük Rehberi, 2022). Şekil 5'te örnek bir gümrük beyanname formunun birçok bilgiden oluştuğu görünmektedir.

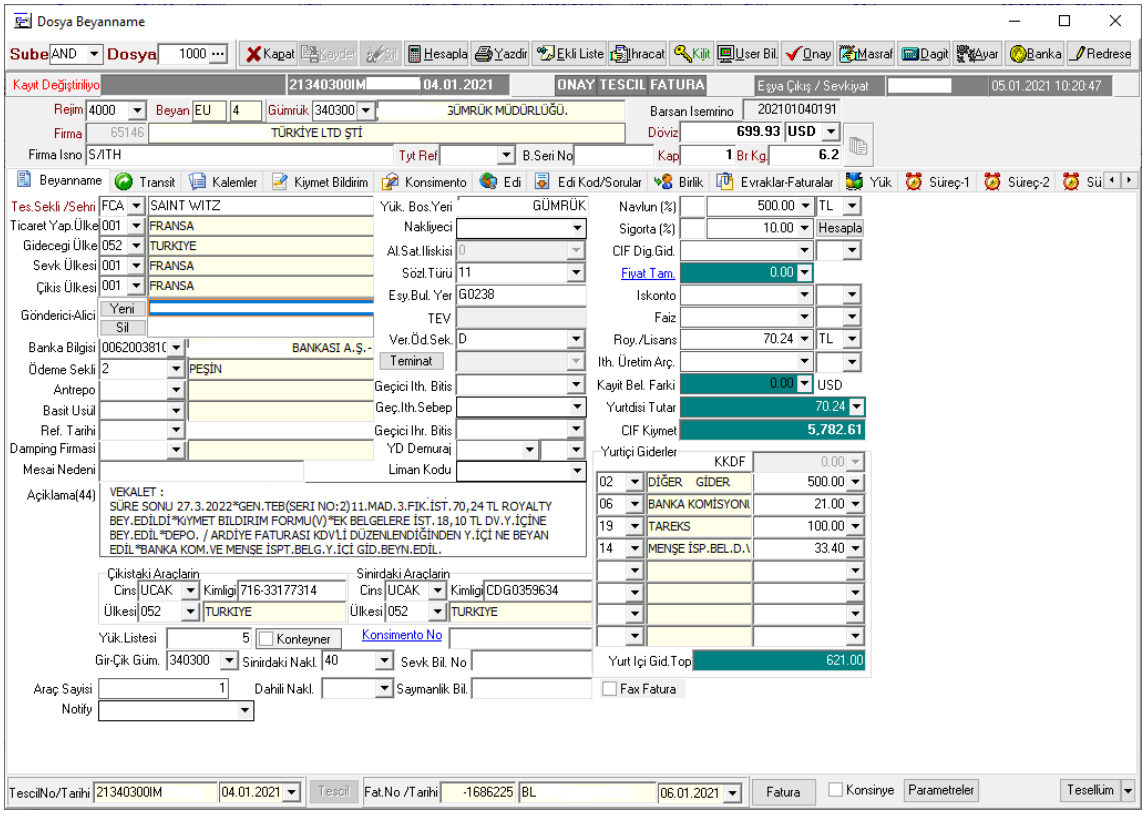


Şekil 5. Gümrük Beyanname Form Örneği

Gümrük beyannamesi bilgisayar sistemi üzerinde birden fazla ekrandan oluşur. Ekranlardaki alanlardan doldurulması zorunlu olanlara istenilen bilgiler girilerek beyanname gümrük bilgisayar sistemine kaydedilir. BİLGE sistemine kaydedilen gümrük beyannamesine tescil numarası diye adlandırılan değişmez ve tekil bir numara sistem tarafından verilir. Bu numaranın içeriğinde, içinde bulunulan yılın son iki hanesi, gümrük idaresinin kodu, yapılan işlemin ithalat (IM) veya ihracat (EX) olduğu bilgisi ve tescil sıra numarası bulunur (Gümrük Rehberi, 2022).

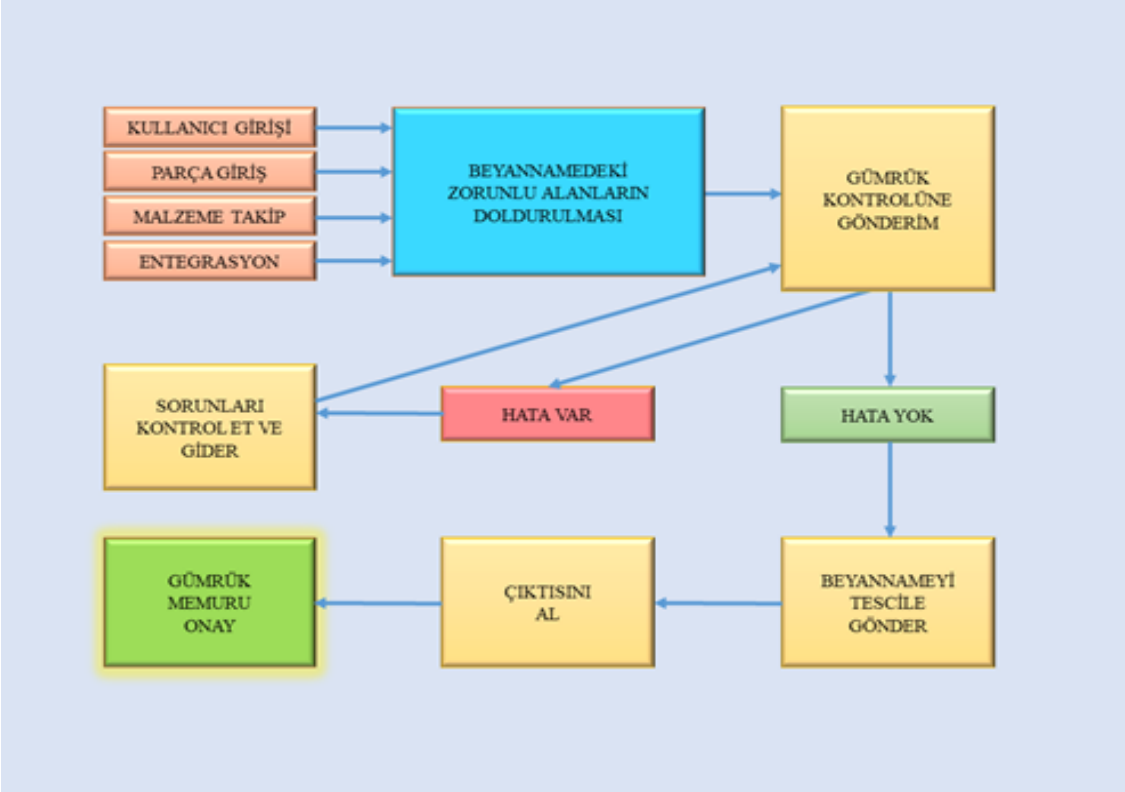
Beyanname tescili gümrük beyanının bu şekilde gümrük idaresi kayıtlarına aktarılmasına denir.

Uluslararası bir lojistik firmasının kendi bünyesinde geliştirilen beyanname program arayüz örneği Şekil 6'da görünmektedir. İşlemlerin bütün aşamaları bu program üzerinden gümrük sistemlerinin sağladığı servislerle yapılabilmektedir.



Şekil 6. Beyanname Program Örneği

Beyanname oluşturma süreci Şekil 7'de gösterilmektedir. Beyanname üzerindeki bilgilerin ilgili gümrük idarelerine sunulmadan önce titizlikle girişinin yapılması gerekmektedir. Normal süreçte bu bilgilerin hepsi kullanıcı tarafından girilmektedir fakat bu süreçte eğer imkân varsa bazı iyileştirmeler bulunmaktadır. İlgili müşterinin sağlayacağı text, excel, pdf, word ve xml (Extensible Markup Language) gibi elektronik ortamdaki dokümanların aktarımı ile bilgilerin önemli bir bölümünün girişi yardımcı programlar ile yapılmaktadır. Yine müşteri ile imkan varsa yapılan entegrasyon ile bilgiler ftp (File Transfer Protocol), sftp (Secure File Transfer Protocol) ve web servis gibi çeşitli platformlar üzerinden alınarak bilgi beyannameye otomatik olarak aktarılabilmektedir. Elektronik ortamdaki dokümanların kullanılması ve entegrasyon ile beyanname oluşturma süreci daha hızlı ve daha doğru bir şekilde yapılabilmektedir.



Şekil 7. Beyanname Oluşturma Süreci

Şekil 7'de görüldüğü üzere hazırlanan beyannameler ilgili gümrük idarelerine kontrol servisi ile sunulmaktadır ve hatasız gelen beyannameler tescil servisi kullanılarak sonuçlandırılmaktadır. Hata gelen kontrol sorgulamalarında beyannamedeki düzeltmeler

yapılarak işlem kontrol servisi ile başlanarak tekrarlanmaktadır. Beyanname kontrol ve tescil işlemlerinde kullanılan bu servisler BİLGE sistemi tarafından sağlanmaktadır.

Bilge Sistemi

Dış ticaret işlemlerinde Türkiye'ye giren veya çıkan eşyanın gümrük işlemlerinin yapılabilmesi için, bu eşya ve yapılan dış ticaret işlemi hakkında gümrük idaresine bir bildirimde bulunulması gereklidir. Gümrük beyanı bu bildirime denmektedir. Gümrük mevzuatına göre gümrük beyanları veri işleme yöntemiyle elektronik olarak yapılabilmektedir. Yazılı gümrük beyanı elektronik olarak yapılmak istendiğinde, BİLGE (Bilgisayarlı Gümrük Etkinlikleri) adı verilen sistem kullanılır. Gümrük işlemlerinin bilgisayar ortamında yürütülmesi amacıyla geliştirilen ve kullanılan uygulamaya verilen isim BİLGE sistemidir. BİLGE sistemi, eşyanın gümrük alanına girişinden çıkışına kadar tüm gümrük işlemlerinin gerçek zamanlı olarak bilgisayar ortamında yürütülmesini sağlar. Gümrük işlemleri yapılırken gümrük idaresine verilmesi gereken formlar BİLGE sistemi sayesinde elektronik ortamda doldurulabilir (Gümrük Rehberi, 2022).

BİLGE Sistemine giriş yapmak için bir kullanıcı koduna ve bir şifreye ihtiyaç vardır. BİLGE kullanıcı kodunu almak için herhangi bir gümrük idaresine başvurulması yeterlidir. Kullanıcı kodu talebinin onaylanması durumunda sisteme giriş için bir şifre belirlenmesi gerekir. Belirlenen şifre http://www.customs-edi.gov.tr adresinden aktif hale getirilerek BİLGE sistemine giriş yapılır. Şifre değişikliği yapılması gerekirse, bunun için de gümrük idaresine dilekçe ile başvurulması gerekir.

BİLGE sistemine 3 şekilde bağlantı sağlanabilir:

- Gümrük idarelerinde bulunan veri giriş salonlarındaki bilgisayar aracılığıyla
- Paket yazılımlar kullanarak
- İnternet üzerinden

Delphi Metodu

Grup yargılarını ortaya çıkarmak ve geliştirmek için kullanılan bir teknik olarak tanımlanmaktadır (Dalkey, 1969). Tekniğin amacına daha derinlemesine bakıldığında, Delphi metodunun kullanılmasının, karmaşık problemler, pahalı çabalar ve belirsizlikler karşısında uzman görüşlerinin bir fikir birliğini toplamak için yapılandırılmış grup

iletişimini kolaylaştırmak için olduğu görülmektedir (Grime ve Wright, 2016). Yapılandırılmış grup çalışmalarının yapılandırılmamışlara göre daha doğru tahminlere yol açması ilkelerine dayanır. Delphi yönteminin temel amacı, bu uzmanları karşılıklı bir anlaşmaya varmaya teşvik etmek ve bir grup uzlaşması oluşturmaktır. Birçok endüstri ve kuruluş, bu yöntemi iş tahminleri veya endüstri tahminleri, hükümet planlaması veya finansal stratejiler gibi yapısal kararlar için kullanabilir.

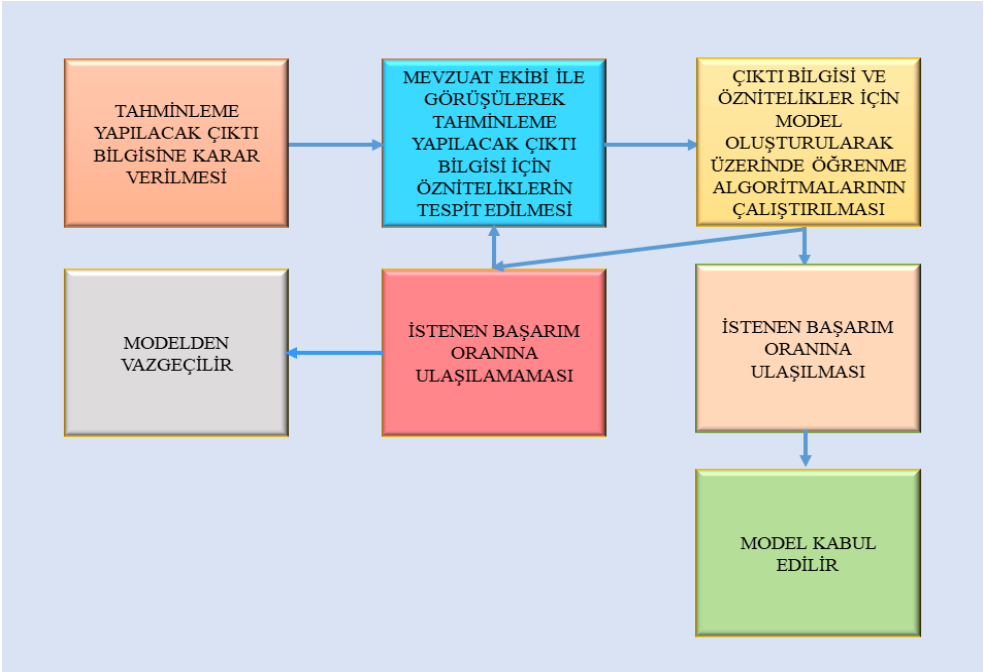
Delphi yöntemini kullanmanın avantajları şunlardır:

- Uzmanlardan bilgi edinme: Delphi yöntemi, herhangi bir tartışmayı çözmenin veya sizin ve diğer ekip üyelerinin sahip olabileceği karmaşık sorunları çözmenin harika bir yoludur. Uzmanların tümü, kuruluşunuz için faydalı olabilecek mantıklı bir anlaşmaya varabilir.
- Dürüst cevaplar almak: Cevapların tümü isimsiz olduğundan, panelistler yargı veya tepki almadan dürüst cevaplarını vermekte kendilerini daha rahat hissedeceklerdir. Aldığınız girdinin doğru olduğunu bilerek de kendinizi güvende hissedebilirsiniz.
- Güçlü bir grup fikir birliğine varmak: Uzmanlar belirli konularda başka düşünceler ve akıl yürütmeler duyduklarında, bu onların fikirlerini daha kolay etkileyebilir. Bu, uzmanların aynı kararı veya fikri daha fazla kişinin desteklediği bir grup kararına varmalarını sağlar ve bu da aldıkları kararı uygulama konusundaki güveninizi artırır.

Delphi yöntemini kullanmanın dezavantajları şunlar olabilir:

- Canlı bir tartışmanın oluşmasını engelleme: Katılımcılar hala geri bildirim sağlarken, girdilerinin çoğu anonim olarak bir ankete yazıldığından, orijinal fikirleri beyin fırtınası yapmıyor veya fikirlerini özgürce söylemiyorlar. Bu, profesyoneller arasında ileri geri iletilen dürüst düşüncelerin canlı bir tartışmasını önleyebilir.
- Yavaş yanıt verme süreleri: Bir soruna çözüm bulmanız veya hızlı bir şekilde karar vermeniz gerekiyorsa, takip edebileceğiniz daha iyi seçenekler olabilir. Delphi yöntemi, anketlerin dağıtılması, toplanması ve gözden geçirilmesinden oluşan birkaç turdan oluştuğu için, genellikle günün büyük bir bölümünü kullanır.
- Yanıtlar çok az değer sağlayabilir veya hiç değer sağlamayabilir: Katılımcıların bir fikir birliğine varamaması nedeniyle, aldığınız yanıtların çok az veya hiç değer vermeme olasılığı çalışma boyunca olabilir.

Şekil 8'de modellerin belirlenme sürecinde Delphi Metodunun uygulanma aşamaları gösterilmektedir. Süreç istenen başarım oranı yakalanıncaya kadar tekrarlanabilmektedir. Bu tekrarlar sonucu istenilen sonucun elde edilememesi sebebiyle modelden vazgeçilebilmektedir. Başarım oranı istenilen düzeyde olan model yapılan uygulamada kullanılmak üzere seçilmektedir. Seçilen modelin başarım oranı en yüksek olan öğrenme algoritması eğitim sonucu tahminlemede kullanılmak için ikili (binary) dosyada saklanmaktadır.



Şekil 8. Delphi Metodu Süreci

Makine Öğrenmesi

Makine öğrenimi veya makine öğrenmesi, bilgisayarların algılayıcı verisi ya da veri tabanları gibi veri türlerine dayalı öğrenimini mümkün hale getiren algoritmaların tasarım ve geliştirme süreçleriyle ilgilenen bir bilim dalıdır. Makine öğrenimi araştırmaları, bilgisayarlara karmaşık örüntüleri algılama ve veriye dayalı akılcı kararlar verebilme becerisi kazandırma konularına odaklanmaktadır. Makine öğrenimi bu bakımdan, istatistik, veri madenciliği, olasılık kuramı, örüntü tanıma, uyarlamalı denetim, yapay zekâ ve kuramsal bilgisayar bilimi gibi alanlarla ilintilidir.

Makine öğrenimi, bilgisayarların, açıkça programlanmadan görevleri nasıl yerine getirebileceklerini keşfetmelerini içerir. Belirli görevleri yerine getirmeleri için sağlanan verilerden öğrenen bilgisayarları kapsar. Bilgisayarlara atanan basit görevler için, makineye eldeki sorunu çözmek için gereken tüm adımları nasıl uygulayacağını bildiren algoritmalar programlamak mümkündür; bilgisayar tarafında öğrenmeye gerek yoktur. Daha gelişmiş görevlerde insan için gerekli algoritmaları elle oluşturmak zor olabilir. Uygulamada, insan programcıların gerekli her adımı belirlemesinden ziyade, makinenin kendi algoritmasını geliştirmesine yardımcı olmak daha etkilidir.

Makine öğrenimi disiplini, bilgisayarlara tam olarak tatmin edici bir algoritmanın bulunmadığı görevleri gerçekleştirmeyi öğretmek için çeşitli yaklaşımlar kullanır. Çok sayıda olası yanıtın olduğu durumlarda, doğru yanıtlardan bazılarını geçerli olarak etiketlemek bir yaklaşımdır. Bu, daha sonra bilgisayarın doğru yanıtları bulmak için kullandığı algoritmayı ya da algoritmaları geliştirmede eğitim verisi olarak kullanılabilir. Örneğin, sayısal karakter tanıma görevinde sistemi eğitmek için el yazısıyla yazılmış rakamların MNIST veri kümesi sıklıkla kullanılır.

Makine öğrenimi terimi 1959'da bilgisayar oyunları ve yapay zeka alanında öncü ve IBM çalışanı olan Amerikalı Arthur Samuel tarafından icat edildi (Kohavi ve Provost, 1998). 1960'larda makine öğrenimi araştırmasının temsili bir kitabı olan Nilsson'un Öğrenme Makineleri hakkındaki kitabında yer alan çoğunlukla örüntü sınıflandırması için makine öğrenimi ile ilgilidir. 1973'te Duda ve Hart tarafından yapılan model çalışması tanımlandığı gibi 1970'lerde de devam etti. 1981'de, bir sinir ağının bir bilgisayar terminalinden 40 karakteri (26 harf, 10 rakam ve 4 özel sembol) tanımayı öğrenmesi için öğretme stratejilerinin kullanımına ilişkin bir rapor verildi.

Tom M. Mitchell, makine öğrenimi alanında incelenen algoritmaların geniş ölçüde alıntılanan daha resmi bir tanımını yaptı: Bir bilgisayar programının performans ölçüsü "P" ve bazı "T" görev sınıflarıyla ilgili olarak "T" görevlerindeki performansı "E" deneyimiyle iyileşiyorsa "P" ile ölçüldüğü gibi E deneyiminden öğrendiği söylenir (Mitchell, 1997). Makine öğrenimi mevcut olan görevlerin bu tanımı, alanı bilişsel terimlerle tanımlamak yerine temelde operasyonel tanım sunar. Bu, Alan Turing'in "Bilgisayar Makineleri ve Zeka" (Turing, 2009) adlı makalesinde "Makineler düşünebilir

mi?" "Makineler bizim (düşünen varlıklar olarak) yapabildiğimizi yapabilir mi?" sorusuyla değiştirilir.

Günümüzün modern makine öğreniminde iki amaç bulunmaktadır. İlk amaç verileri geliştirilen modellere göre sınıflandırmaktır. Diğer amaç ise gelecekteki sonuçlar için bu modellere dayalı olarak tahminler yapmaktır. Verileri sınıflandırmaya özgü varsayımsal bir algoritma, kanserli benleri sınıflandırmada onu eğitmek için denetimli öğrenmeyle birleştirilen mollerin bilgisayar görüşü kullanabilir. Hal böyle olunca, hisse senedi ticareti için bir makine öğrenme algoritması, tüccara gelecekteki potansiyel tahminler hakkında bilgi verebilir. Makine öğrenimi günümüzde birçok farklı alanda kullanılmaktadır. Otomotiv sektöründe sensör verileri ile sürücüsüz araç çalışmalarında, konuşma tanıma teknolojisinde, bilgisayarlı görmede nesne tanıma, hastalık belirtilerinin tespiti ve risk hesaplamalarında, kredi kartı dolandırıcılığı denetiminde, arama motorlarında, pazarlama ve eğlence sektöründe, trafik tahmininde, duygu analizinde ve birçok farklı alanda kullanılmıştır.

İnsan sezgisine olan gereksinimi tümüyle ortadan kaldırmaya çalışan makine öğrenimi sistemleri mevcuttur. Bununla birlikte insan ve makine arasında iş birliğine dayalı bir yaklaşım benimseyen makine öğrenimi sistemleri de vardır. Sistemi tasarlayan kişinin verinin kodlanma şekli üzerinde tümüyle karar vermesi insan sezgisinin tümüyle ortadan kaldırılmasını imkânsız hale getirmektedir. Deneysel tekniklerin otomatikleştirilme çabası makine öğrenimi olarak görülmektedir.

Öğrenme Yaklaşımları

Makine öğrenimi algoritmaları ile veriler kullanılarak öğrenme süreci insan müdahalesi olmadan yapılır. Öğrenme sürecindeki deneyimler ile gelişim gerçekleştirilir. Hedeflenen sonuca göre makine öğrenimi algoritmaları birkaç sınıfa ayrılabilmektedir.

Gözetimli Öğrenme

Girdilerin hedef çıktılar ile eşlendiği öğrenme yöntemidir.

Gözetimsiz Öğrenme

Girdi kümelerinin modellendiği öğrenme yöntemidir.

Pekiştirmeli Öğrenme

Dünya algısına dayalı bir öğrenme yöntemidir. Her işlem ortamda bir etki oluşturmakta ve ortam, öğrenme algoritmasına yol gösteren ödüller biçiminde dönütler vermektedir.

Yarı Gözetimli Öğrenme

Uygun işlev ya da sınıflandırıcılar oluşturmak için etiketli ve etiketsiz örnekleri birlikte ele alan öğrenme yöntemidir.

Öğrenmeyi Öğrenme

Önceki deneyimlerden yararlanan öğrenme yöntemidir.

Makine öğrenimi algoritmaları genellikle denetimli veya denetimsiz olarak sınıflandırılır.

Denetimli makine öğrenimi algoritmaları, etiketli örnekler kullanarak gelecekteki olayları tahmin etmek için geçmişte öğrenilenleri yeni verilere uygulayabilir. Bilinen bir eğitim veri setinin analizinden başlayarak, öğrenme algoritması, çıktı değerleri hakkında tahminler yapmak için doğru bilinen bir önermeden yeni önermeler çıkararak bir fonksiyon üretir. Sistem, yeterince eğitildikten sonra herhangi bir yeni girdi için tahminler sağlayabilir. Öğrenme algoritması aynı zamanda çıktısını istenen çıktı ile karşılaştırabilir ve modeli buna göre uyarlamak için hata tespiti yapabilir.

Eğitmek için kullanılan bilgiler sınıflandırılmadığında veya etiketlenemediğinde denetimsiz makine öğrenimi algoritmaları kullanılır. Denetimsiz öğrenme, sistemlerin etiketlenmemiş verilerden bilinmeyen bir yapıyı tespit etmek için bir işlevi nasıl çıkardığını konu edinir. Sistem doğru çıktıyı bulmaz, ancak verileri araştırır ve etiketlenmemiş verilerden bilinmeyen yapıları tanımlamak için veri setlerinden çıkarımlar yapabilir.

Yarı denetimli makine öğrenimi algoritmaları, denetimli ve denetimsiz öğrenme arasında bir yere konumlanmaktadır. Çünkü eğitim için hem etiketli hem de etiketsiz verileri kullanmaktadır. Genellikle az miktarda etiketli veri ve büyük miktarda etiketsiz veri seçilerek uygulanmaktadır. Öğrenme doğruluğu bu yöntemi kullanan sistemlerde ciddi miktarlarda arttırılabilir. Genellikle, yarı denetimli öğrenme, edinilmiş etiketli veriler onu

eğitmek ve ondan öğrenmek için nitelikli ve ilgili kaynaklar gerektirdiğinde seçilir. Aksi durumda, etiketlenmemiş verilerin elde edilmesi için genellikle ek kaynak ihtiyacı olmaz.

Pekiştirmeli makine öğrenimi algoritmaları, eylemler üreterek ve hataları veya ödülleri keşfederek çevresiyle etkileşime giren bir öğrenme yöntemidir. Deneme yanılma yöntemi ve gecikmiş ödül, pekiştirmeli öğrenmenin en belirgin özellikleridir. Bu yöntem, makinelerin ve yazılımların performansını en üst seviyeye çıkarmak için belirli bir ölçüde ideal davranışın otomatik olarak belirlenmesine imkan tanır. Hangi eylemin en iyi olduğunu öğrenmesi için basit ödül geri bildirimine ihtiyaç vardır. Bu basit ödül geri bildirimi ihtiyacına takviye sinyali denmektedir.

Makine öğrenimi, büyük miktarda verinin analiz edilmesini sağlar. Fırsatların veya risklerin belirlenmesinde genellikle daha hızlı ve daha doğru sonuçlar verirken, doğru şekilde eğitilmesi için ilave zaman ve kaynak da gerektirebilir. Makine öğrenimini yapay zeka ve bilişsel teknolojilerle birleştirmek, büyük hacimli bilgilerin işlenmesinde daha da etkili olmasını sağlayabilir.

Makine Öğrenmesi Algoritma Türleri

Regression

Regresyon, bir diğer önemli ve yaygın olarak kullanılan istatistiksel ve makine öğrenimi aracıdır. Regresyona dayalı görevlerin temel amacı, verilen girdi verileri için sayısal değerler olan çıktı etiketlerini veya yanıtları tahmin etmektir. Çıktı, modelin eğitim aşamasında öğrendiklerine dayanacaktır. Temel olarak, regresyon modelleri, girdiler ve karşılık gelen çıktılar arasındaki belirli ilişkiyi öğrenmek için girdi veri özelliklerini (bağımsız değişkenler) ve bunlara karşılık gelen sürekli sayısal çıktı değerlerini (bağımlı veya sonuç değişkenleri) kullanır (Freedman, 2009).

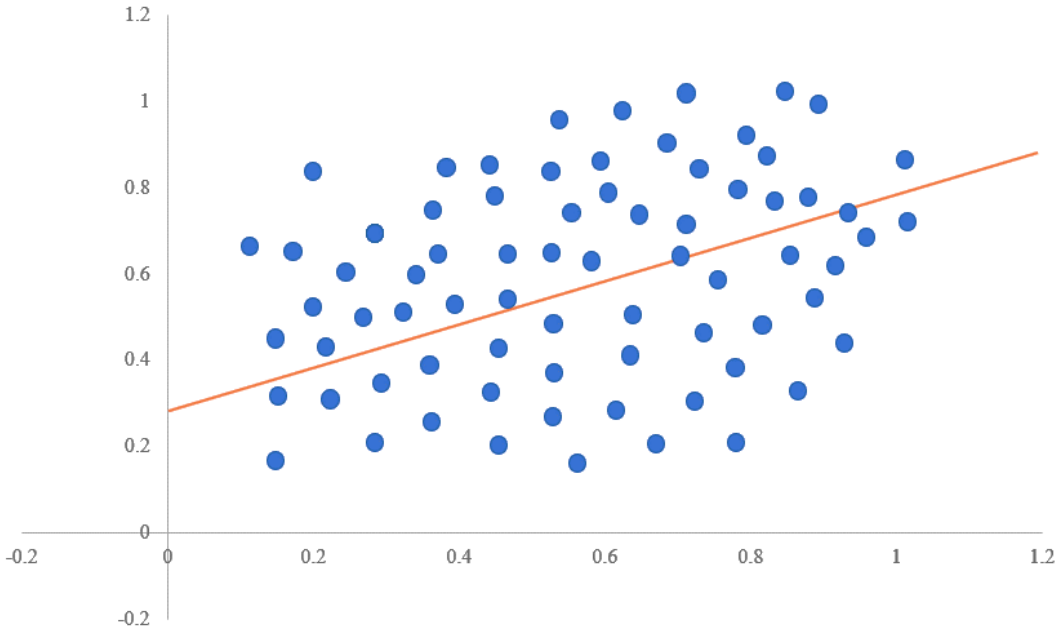
Regresyon modelleri şunlardır:

- Basit regresyon modeli - Bu, tahminlerin verilerin tek, tek değişkenli bir özelliğinden oluşturulduğu en temel regresyon modelidir.
- Çoklu regresyon modeli - Adından da anlaşılacağı gibi, bu regresyon modelinde tahminler, verilerin çoklu özelliklerinden oluşur.

Regresyon türleri şunlardır:

- Tahmine dayalı analiz için Linear regresyon kullanılır. Linear regresyon, kriter veya skaler yanıt ile çoklu tahmin ediciler veya açıklayıcı değişkenler arasındaki ilişkiyi modellemek için doğrusal bir yaklaşımdır. Linear regresyon, tahmin edicilerin değerleri verilen yanıtın koşullu olasılık dağılımına odaklanır. Linear regresyon için, fazla ezbere öğrenme tehlikesi vardır.
- Eğrisel veriler için Polynomial regresyon kullanılır. Polynomial regresyon en küçük kareler yöntemine uygundur. Regresyon analizinin amacı, bağımsız değişken x ile ilgili olarak bir bağımlı değişken y'nin beklenen değerini modellemektir.
- Regresyon modellerini tahmine dayalı modellerle uydurmak için Stepwise regresyon kullanılır. Otomatik olarak gerçekleştirilir. Her adımda değişken, açıklayıcı değişkenler kümesinden eklenir veya çıkarılır. Kademeli regresyon için yaklaşımlar, ileriye doğru seçim, geriye doğru eleme ve çift yönlü elemedir.
- Ridge regresyon, çoklu regresyon verilerini analiz etmek için bir tekniktir. Çoklu bağlantı oluştuğunda, en küçük kareler tahminleri yansızdır. Regresyon tahminlerine bir derece yanlılık eklenir ve sonuç olarak sırt regresyonu standart hataları azaltır.
- Lasso regresyonu hem değişken seçimi hem de düzenlileştirme gerçekleştiren bir regresyon analizi yöntemidir. Lasso regresyonu, yumuşak eşikleme kullanır. Lasso regresyonu, son modelde kullanılmak üzere sağlanan ortak değişkenlerin yalnızca bir alt kümesini seçer.
- ElasticNet regresyon, lasso ve ridge yöntemlerinin cezalarını doğrusal olarak birleştiren düzenli bir regresyon yöntemidir. ElasticNet regresyon, destek vektör makineleri, metrik öğrenme ve portföy optimizasyonu için kullanılır.

Örnek bir linear regresyon modeli, Şekil 9'da gösterilmektedir. Burada değişkenler arasındaki ilişki doğrusal eğimli (y=mx+c) düz bir çizgi olarak gösterilir. X'in (bağımsız değişken) değeri arttıkça, y'nin (bağımlı değişken) değeri de aynı şekilde artmaktadır. Doğrusal regresyonda veriler Şekil 9'da gösterildiği gibi doğrusal çizgi boyunca düzgün dağıtımlıdır.



Şekil 9. Linear Regresyon Modeli

Sınıflandırma (Classification)

Sınıflandırma algoritmaları denetimli makine öğrenimi algoritmalarıdır. Kategorik değerleri tahmin etmek için Sınıflandırma algoritmalarına ihtiyacımız vardır. Sınıflandırma algoritması, eğitim verileri temelinde yeni gözlemlerin kategorisini belirlemek için kullanılan bir Denetimli Öğrenme tekniğidir. Sınıflandırmada, bir program verilen veri setinden veya gözlemlerden öğrenir ve ardından yeni gözlemi bir dizi sınıfa veya gruba sınıflandırır. Sınıflandırmanın çıktı değişkeni kategoridir. Sınıflandırma algoritması Denetimli bir öğrenme tekniği olduğundan, etiketlenmiş girdi verilerini alır. Karşılık gelen çıktı ile girdi içerdiği anlamına gelir. Sınıflandırma algoritmasının temel amacı, belirli bir veri kümesinin kategorisini belirlemektir ve bu algoritmalar temel olarak kategorik verilerin çıktısını tahmin etmek için kullanılır (Gnanadesikan, 1977).

Binary, Multi-Class, Multi-Label ve Imbalanced olmak üzere 4 ana sınıflandırma türü vardır.

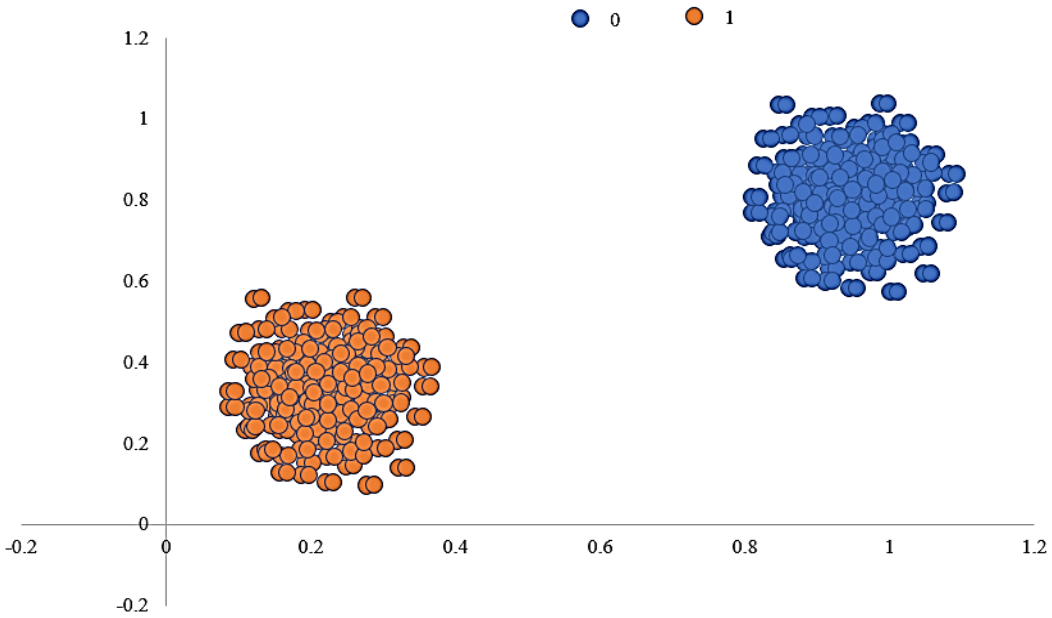
- Binary

Binary sınıflandırma, iki sınıf etiketine sahip sınıflandırma görevlerini ifade eder. E-posta spam algılama (spam veya değil), Churn tahmini (churn veya değil), dönüşüm tahmini (satın alın veya almayın) bu görevlere örnek verilebilir.

Binary sınıflandırma görevleri, normal durum olan bir sınıfı ve anormal durum olan başka bir sınıfı içerir. Örneğin, "spam değil" normal durumdur ve "spam" anormal durumdur. Başka bir örnek, "kanser tespit edilmedi", tıbbi bir test içeren bir görevin normal durumudur ve "kanser tespit edildi", anormal durumdur. Normal durum sınıfına 0 sınıf etiketi atanır ve anormal duruma sahip sınıfa sınıf etiketi 1 atanır.

Her örnek için bir Bernoulli olasılık dağılımını tahmin eden bir modelle ikili sınıflandırma görevini modellemek yaygındır. Bernoulli dağılımı, bir olayın 0 veya 1 olarak ikili bir sonuca sahip olacağı bir durumu kapsayan ayrı bir olasılık dağılımıdır. Sınıflandırma için bu, modelin sınıf 1'e ait bir örneğin veya anormal durumun olasılığını tahmin ettiği anlamına gelir.

Şekil 10’da Binary sınıflandırma ile iki sınıfa ait etiketlerin dağılımı gösterilmiştir.



Şekil 10. Binary Sınıflandırma

- Multi-Class

Multi-Class sınıflandırma, ikiden fazla sınıf etiketine sahip olan sınıflandırma görevlerini ifade eder. Yüz sınıflandırması, bitki türlerinin sınıflandırılması ve optik karakter tanıma örnek olarak verilebilir.

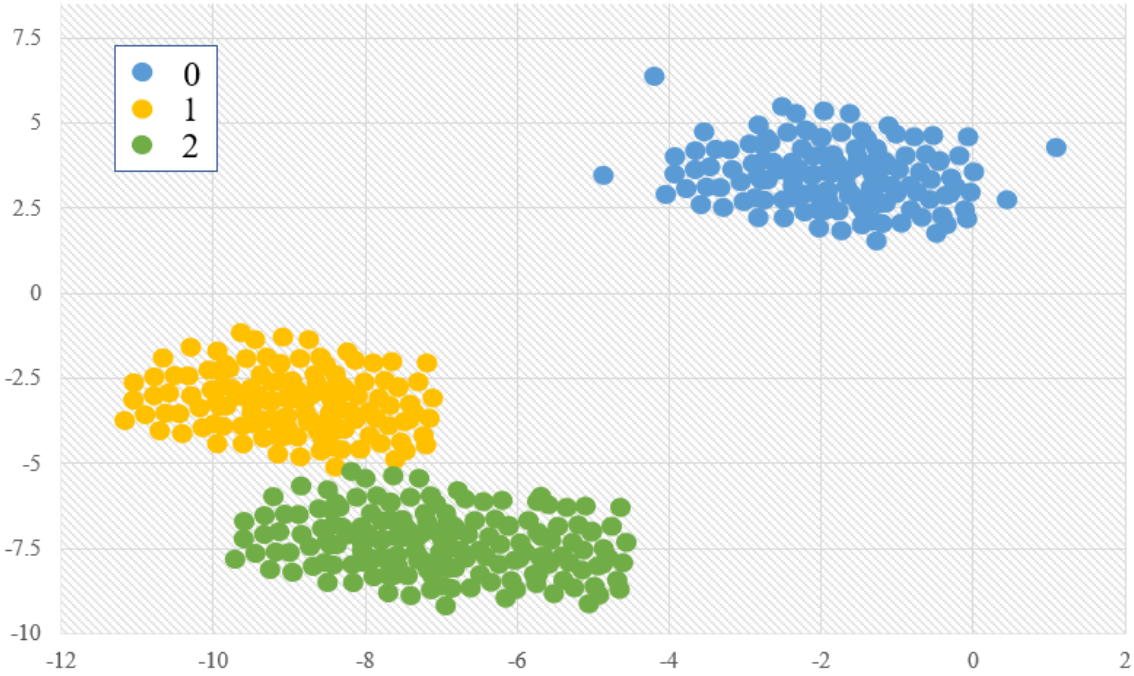
Binary sınıflandırmadan farklı olarak, Multi-Class sınıflandırma, normal ve anormal sonuçlar kavramına sahip değildir. Bunun yerine, örnekler bilinen bir dizi sınıf arasından birine ait olarak sınıflandırılır.

Bazı problemlerde sınıf etiketlerinin sayısı çok fazla olabilir. Örneğin, bir model bir fotoğrafın bir yüz tanıma sistemindeki binlerce veya on binlerce yüz arasından birine ait olduğunu tahmin edebilir. Metin çeviri modelleri gibi bir kelime dizisini tahmin etmeyi içeren problemler de Multi-Class sınıflandırmanın özel bir türü olarak kabul edilebilir. Tahmin edilecek kelime dizisindeki her kelime, kelime hazinesinin tahmin edilebilecek olası sınıfların sayısını tanımladığı ve boyut olarak on veya yüz binlerce kelime olabilen çok sınıflı bir sınıflandırma içerir.

Multi-Class bir sınıflandırma görevini, her örnek için bir Multinoulli olasılık dağılımını tahmin eden bir modelle modellemek yaygındır. Multinoulli dağılımı, bir olayın kategorik bir sonucu olacağı bir durumu kapsayan ayrı bir olasılık dağılımıdır.

Binary sınıflandırma için kullanılan birçok algoritma, Multi-Class sınıflandırma için kullanılabilir.

Şekil 11'de ikiden fazla sınıf etiketine sahip Multi-Class sınıflandırma dağılımı gösterilmiştir.



Şekil 11. Multi-Class Sınıflandırma

- Multi-Label

Multi-Label sınıflandırma, her örnek için bir veya daha fazla sınıf etiketinin tahmin edilebileceği, iki veya daha fazla sınıf etiketine sahip olan sınıflandırma görevlerini ifade eder.

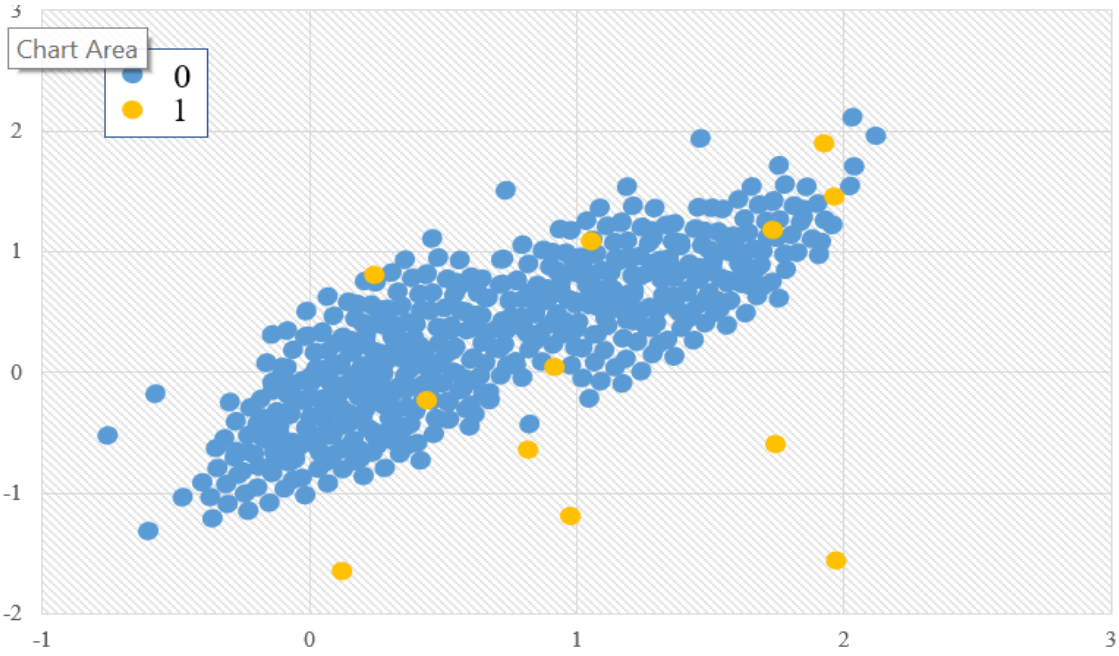
Multi-Label sınıflandırma görevlerini, her bir çıktının bir Bernoulli olasılık dağılımı olarak tahmin edildiği, birden çok çıktıyı tahmin eden bir modelle modellemek yaygındır. Bu aslında her örnek için çoklu Binary sınıflandırma tahminleri yapan bir modeldir.

Binary veya Multi-Class sınıflandırma için kullanılan sınıflandırma algoritmaları, doğrudan Multi-Label sınıflandırma için kullanılamaz. Algoritmaların çok etiketli versiyonları olarak adlandırılan standart sınıflandırma algoritmalarının özel versiyonları kullanılabilir. Bunlar Multi-Label Decision Trees, Multi-Label Random Forests ve Multi-Label Gradient Boosting algoritmalarıdır.

- Imbalanced

Imbalanced sınıflandırma, her sınıftaki örnek sayısının eşit olmayan şekilde dağıldığı sınıflandırma görevlerini ifade eder. Tipik olarak, Imbalanced sınıflandırma görevleri, eğitim veri kümesindeki örneklerin çoğunluğunun normal sınıfa ve örneklerin az bir kısmının anormal sınıfa ait olduğu Binary sınıflandırma görevleridir. Dolandırıcılık tespiti, aykırı değer tespiti ve tıbbi teşhis testleri örnek olarak verilebilir.

Bu problemler, özel teknikler gerektirebilse de ikili sınıflandırma görevleri olarak modellenmiştir. Şekil 12'de örneklerin çoğunluğunun normal sınıfa ve örneklerin az bir kısmının anormal sınıfa ait olduğu Imbalanced sınıflandırma dağılımı gösterilmiştir.



Şekil 12. Imbalanced Sınıflandırma

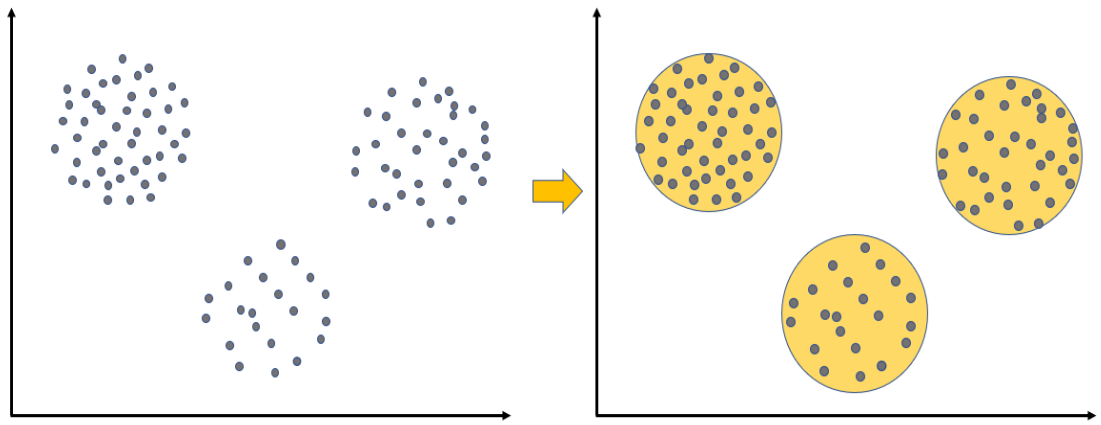
Clustering

Küme analizi veya kümeleme, denetimsiz bir makine öğrenimi görevidir. Verilerdeki doğal gruplandırmayı otomatik olarak keşfetmeyi içerir. Denetimli öğrenmenin (tahmini modelleme gibi) aksine, kümeleme algoritmaları yalnızca girdi verilerini yorumlar ve özellik alanında doğal grupları veya kümeleri bulur.

Kümeleme teknikleri, tahmin edilecek bir sınıf olmadığında, bunun yerine örneklerin doğal gruplara bölüneceği durumlarda uygulanır.

Bir küme, genellikle, etki alanındaki örneklerin (gözlemler veya veri satırları) kümeye diğer kümelerden daha yakın olduğu, özellik alanındaki bir yoğunluk alanıdır. Küme, bir örnek veya nokta özellik uzayı olan bir merkeze (merkez) sahip olabilir ve bir sınırı veya kapsamı olabilir. Bu kümeler, muhtemelen, örneklerin alındığı etki alanında, bazı örneklerin diğer örneklere göre daha güçlü bir benzerlik taşımasına neden olan bir mekanizmayı yansıtır. Belirlenen kümelerin değerlendirilmesi özneldir ve kümelenmeye özgü birçok nicel ölçü olmasına rağmen bir alan uzmanı gerektirebilir. Tipik olarak, kümeleme algoritmaları, bir algoritmanın keşfetmesi beklenen önceden tanımlanmış kümelere sahip sentetik veri kümeleri üzerinde akademik olarak karşılaştırılır (Estivill-Castro, 2002).

Kümeleme, denetimsiz bir öğrenme tekniğidir, bu nedenle herhangi bir yöntemin çıktısının kalitesini değerlendirmek zordur. Şekil 13'te kümeleme tekniği görsel olarak verilmektedir.



Şekil 13. Clustering

Makine Öğrenmesi Algoritmaları

Gaussian Naïve Bayes

Makine Öğreniminde, bilgi ve veri analizinde, kategorilerin sınıflandırılmasında, süreçleri tahmin etmedeki hızı ve doğruluğu ile karakterize edildiği için çok önemli algoritmalardır (Shang, Li, Deng, ve He, 2017). Bu yöntem Bayes teorisine (Efron, 2013) dayanmaktadır, çünkü bu teori ilk kez 1763'te İngiliz bilim adamı Thomas Bayes tarafından yayınlanan bir makalede ortaya çıkmıştır (Stigler, 2013). Bu algoritma, belirli bir sonucun olasılığını hesaplar ve karakterlerin ve özelliklerin birbirinden bağımsız olduğunu kabul eder.

Denetimli Öğrenme, Naive Bayes'in lineer sınıflandırma modeli ile ilgilidir. Naive Bayes, denetimli makine öğrenmesi algoritmalarından biridir ve diğer sınıflandırma algoritmalarından son derece hızlı bir sınıflandırma tekniğidir. Bilinmeyen veri kümelerinin sınıfını tahmin etmek için Bayes olasılık teoremini kullanır. Naive Bayes Teoremi, tüm tahmin edicilerin birbirinden bağımsız olduğu güçlü varsayımla uygulanır (Point, 2022). Başka bir deyişle, belirli bir örneğin kategorisini tahmin etmek için her bir özelliğin diğerlerinden bağımsız olması gerekir. Bu yaklaşım, bağımsızlığı nedeniyle Naive olarak bilinir.

Naive Bayes'in amacı, bir olayın koşullu olasılığını bağımsız olarak hesaplamaktır. Naive Bayes modelinin oluşturulması basittir ve çok büyük veri kümelerine uygulanabilir. Büyük verileri işlerken sürekli olarak geliştirilebilir. Ayrıca, son derece karmaşık sınıflandırma yöntemlerinden bile daha iyi performans gösterir (Abellán ve Castellano, 2017).

Model, veri tabanında bulunan bilgilere göre eğitilir, ardından form yeni kayıtların türünü belirler ve daha önce mevcut olan bilgilere göre bunları sınıflandırır. Bu yöntem farklı sistemlerde kullanılmaktadır. Resimlerdeki belirli bir yüzün özelliklerini tanımak veya o içeriğe ve diğer kullanımlara yönelik duyguları analiz ederek metnin içeriğini (olumsuz, olumlu, iyimser) belirlemek gibi (Abellán ve Castellano, 2017).

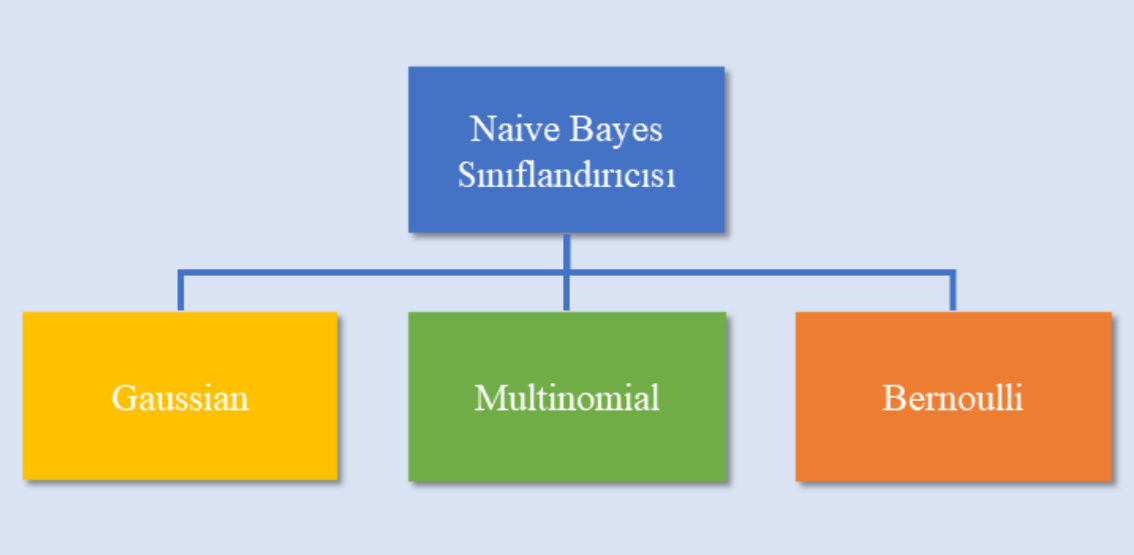
Bayes teoremi, olayla ilgili olabilecek koşulların ön bilgisine dayanarak bir olayın olasılığını tanımlar. Aşağıda Bayes Teoremini türeten koşullu olasılık formülü gösterilmektedir (Triola, 2010).

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

Naive Bayes formülündeki parametrelerin açıklamaları:

- P(A) = A olayının bağımsızlık olasılığı
- P(B) = B olayının bağımsızlık olasılığı
- P(B|A) = A olayı bilindiğinde B olayının olasılığı
- P(A|B) = B olayı bilindiğinde A olayının olasılığı

Şekil 14'te görüldüğü gibi Naive Bayes algoritmalarının 3 ana türü vardır. Bunlar Gaussian, Multinomial ve Bernoulli'dir.



Şekil 14. Navive Bayes Algoritma Türleri

- Gaussian Naive Bayes

Bu özel algoritma türü, normal yeniden dağıtımı izleyen sürekli değerler ve özellikler durumunda kullanılır. Bu algoritma, sınıflandırma ve regresyon için karmaşık makine öğrenimi algoritmalarının temeli olarak kullanılır (Ontivero-Ortega, Lage-Castellanos, Valente, Goebel, ve Valdes-Sosa, 2017).

- Multinomial Naive Bayes

Belge sorunlarını çözmek ve bunları dosyanın belirli bir kategoriye ait olmasına göre sınıflandırmak için kullanılır ve belge içindeki sözcüklerin tekrarlanması ilkesine bağlıdır (Kibriya, Frank, Pfahringer, ve Holmes, 2004).

- Bernoulli Naive Bayes

İkili kavram üzerinde çalışır ve belirli bir değeri tahmin eden mantıksal değişkenlerdir (Singh, Kumar, Gaur, ve Tyagi, 2019). Örneğin yalnızca (evet veya hayır).

Naive Bayes'in başlıca avantajları:

- Naive Bayes modeli geleneksel sayısal hipotezden ortaya çıkmış ve istikrarlı bir gruplama yeterliliğine sahiptir.
- Bu algoritma, küçük verilerle bile uğraştığı ve gerekli görevleri artımlı bir şekilde ele aldığı için büyük verilere ihtiyaç duymaz.
- Bu algoritma sembolleri kategorize etmek amacıyla kullanılır ve veri kaybolduğunda etkilenmez ve Naive Bayes basit bir algoritmadır.

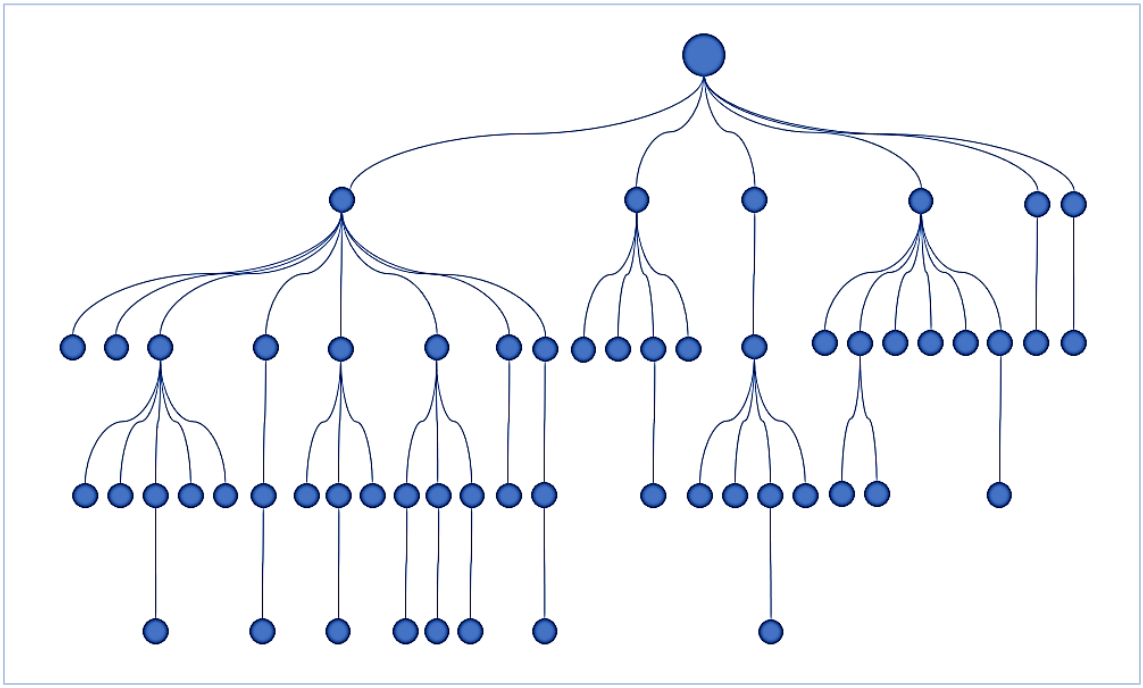
Naive Bayes'in ana dezavantajları:

- Naive Bayes modeli, işlediği verilerin özelliklerinin belirli bir çıktı sınıfı için çoğunlukla birbirinden bağımsız olduğunu düşünmektedir. Çünkü bu model diğer sınıflandırma yöntemleriyle karşılaştırıldığında en düşük hata oranını içermektedir. Veri düzenlemesinin özellikleri arasındaki korelasyon ne kadar küçük olursa, Naive Bayes algoritması o kadar doğru olmalıdır.
- Naive Bayes algoritması önceki ve varsayılan tahminlere dayanır ve önceki varsayımsam modele geri dönmeden algoritma kötü tahmin sonuçları verir.
- Önceki sonuçlara olan güven nedeniyle Naive Bayes algoritmasını kullanan bir hata oranı vardır.
- Algoritma, veri girişinin ifadesine çok duyarlıdır.

Decision Tree Sınıflandırma Algoritması

Decision Tree, kümeleme ve tahmin problemlerinde sıkça kullanılan algoritmalardan biridir. Eğitim sürecinde yukarıdan aşağıya veya genelden özele doğru ilerleyen bir strateji sunan algoritmalardır. Ağaç yapısı gibi bir tür akış şeması olan bu stratejide her

düğümün öznitelik değeri ölçülür ve elde edilen sonuçlarla dallar oluşturulur (Shalev-Shwartz ve Ben-David, 2014). Tipik bir karar ağacı mimarisi Şekil 15'te gösterilmektedir.



Şekil 15. Decision Tree Mimarisi

Decision Tree, bilgi kazancı gibi ayırma ölçütlerinin eşlik ettiği bir kök düğümle başlatılır. Seçilen kriterlere göre, bu kök düğüm dallara ayrılmıştır. Bu ayırma, yaprak düğümleri tarafından sonlandırılana kadar devam eder. Bu yapıda verilerin izlediği yol, if-then-else algoritmalarına bağlı olarak ikili (evet/ hayır), kategorik (sıcak/soğuk), sayısal (SPEI> 1.50) ayrımlarda verilen cevapla ilerlemektedir. Ağaç, kök düğümdeki tüm eğitimli vakaları içerir ve kümelenme olup olmadığını kontrol eder. Kök düğümdeki tüm durumlar tek bir kümeyle çakışırsa, çözüm elde edilir. Aksi takdirde, kök düğüm dallara bölünür ve dal doğrudan karar verecek kadar basit olana kadar tekrarlanır. Veri kümesinin boyutuna bağlı olarak, ağaç dalları zor bir görev olabilir. Ezbere öğrenmeyi önlemek için, karar ağacından az sayıda nesne içeren yaprak düğümlerinin çıkarılmasını ifade eden budama algoritmaları kullanılır (Rokach ve Maimon, 2014).

Decision Tree algoritması beklenen bilginin hesaplanmasıyla başlar. Bu Entropi olarak adlandırılır. Entropi teorisi, rastgele bir değişkenle ilişkili belirsizliği ölçmek için kullanılır. Decision Tree modellemesinde, ağaç dallarını ve düğümlerini ayrıştırmak için

kullanılabilir. Bir veri kümesindeki safsızlığın karakterize edilmesini sağlar. Daha yüksek entropi daha yüksek belirsizlik anlamına gelirken, daha düşük entropi daha düşük belirsizlik anlamına gelir (Witten, Frank, ve Hall, 2011).

Aşağıda Entropi Denklemi (Shannon, 1948) gösterilmektedir. Burada, p(x) belirli bir sınıfa ait grubun yüzdesini ve H ise entropiyi belirtmektedir.

$$\boldsymbol{H = -\sum p(x)\, log\, p(x)}$$

En çok kullanılan Decision Tree algoritmaları:

- ID3
- C4.5
- CHAID
- CART

Decision Tree Algoritmalarının avantajları:

- Daha kullanışlıdır çünkü anlaşılması ve yorumlanması kolaydır.
- Çok büyük veri kümelerine uygulanması kolaydır.
- Çıkarımsal modellerin aksine ağaç modelleri araştırmacılardır.
- Karar ağaçları parametrik değildir. Diğer teknikler gibi istatistiksel varsayımlarla da ilgilenmez.
- Hiyerarşik ayrışma, mevcut özelliklerinin ilişki yönünü ve önem sırasını ortaya koymaktadır.
- Ağaç yapısı nedeniyle karışık bağlantılar ortaya çıkarılabilir.
- Sürekli ve kategorik niteliklerle çalışabilir.

Decision Tree algoritmalarının dezavantajları:

- Özellikle büyük veri kümelerinde çok sayıda sürekli değişken varsa, hesaplama maliyetleri diğer sınıflandırma yöntemlerinden daha maliyetli olabilir.
- En uygun ağacı bulmak bazen zor olabilir. Yararlı bir karar ağacı oluşturmaya çalışırken karmaşık ve büyük ağaçlar üretilebilir.

- Girdi sayısı düşük olduğunda ağaç yeterli bilgiyi yakalayamayabilir.
- Bazı algoritmalar yalnızca kategorik verilerle çalışır ve sürekli değişkenlerin sınıflandırılmasında başarılı olmayabilir.
- Öğrenme kümesinin veri ve değişken sayısı büyükse veya karışıksa, zaman ve ağaç karmaşıklıkları değişir.

K-Nearest Neighbors Sınıflandırma Algoritması

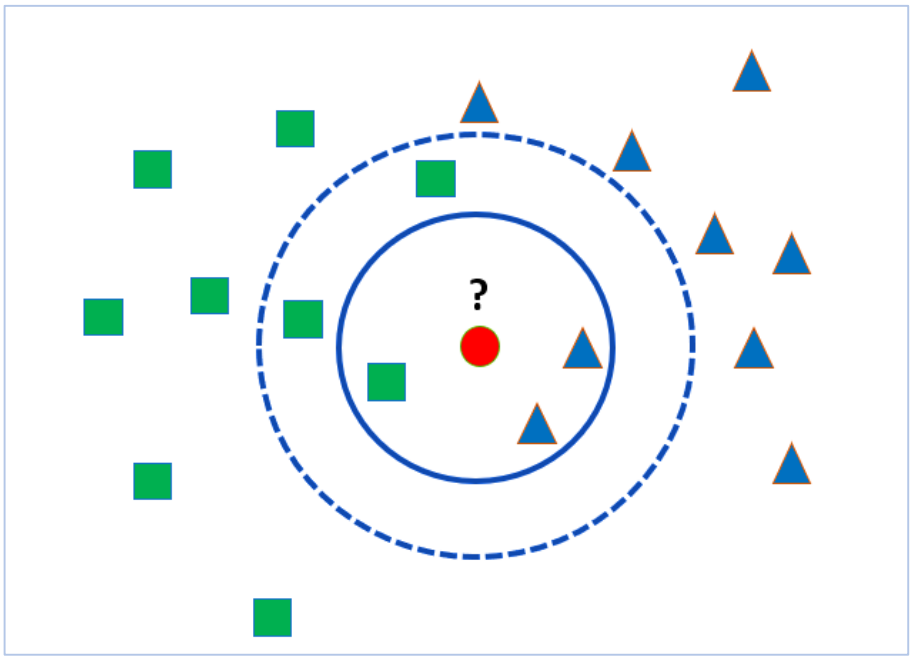
K-Nearest Neighbors (KNN) algoritması, bir veri noktasının kendisine en yakın veri noktalarının hangi gruba ait olduğuna bağlı olarak bir grubun veya başka bir grubun üyesi olma olasılığını tahmin etmek için kullanılan bir veri sınıflandırma yöntemidir. K-Nearest Neighbors algoritması, sınıflandırma ve regresyon problemlerini çözmek için kullanılan denetimli bir makine öğrenme algoritması türüdür. Bununla birlikte, esas olarak sınıflandırma problemleri için kullanılır (Piryonesi ve El-Diraby, 2020).

KNN tembel bir öğrenme ve parametrik olmayan bir algoritmadır. Buna tembel öğrenme algoritması veya tembel öğrenci denir, çünkü eğitim verilerini sağladığınızda herhangi bir eğitim gerçekleştirmez. Bunun yerine, verileri yalnızca eğitim süresi boyunca saklar ve herhangi bir hesaplama yapmaz. Veri kümesinde bir sorgu gerçekleştirilinceye kadar bir model oluşturmaz. Bu, KNN'yi veri madenciliği için ideal kılar (Samworth, 2012).

Parametrik olmayan bir yöntem olarak kabul edilir, çünkü temel veri dağılımı hakkında herhangi bir varsayımda bulunmaz. Basitçe söylemek gerekirse, KNN etrafındaki veri noktalarına bakarak bir veri noktasının hangi gruba ait olduğunu belirlemeye çalışır.

Bir veri noktasının A grubunda mı yoksa B grubunda mı olduğunu belirlemek için algoritma, yakınındaki veri noktalarının durumlarına bakar. Veri noktalarının çoğunluğu A grubundaysa, söz konusu veri noktasının A grubunda olması muhtemeldir ve bunun tersi de geçerlidir. Kısaca KNN, en yakın komşu olarak da bilinen en yakın açıklamalı veri noktasına bakarak bir veri noktasının sınıflandırılmasını içerir (Beyer, Goldstein, Ramakrishnan, ve Shaft, 1999).

Şekil 16'da KNN algoritmasında K'nın alacağı değere göre kırmızı dairenin hangi sınıfa dâhil olacağı gösterilmiştir. Bu örnekte kırmızı daire, K değeri 3 alınırsa mavi, 5 alınırsa yeşil gruba dâhil olacaktır.



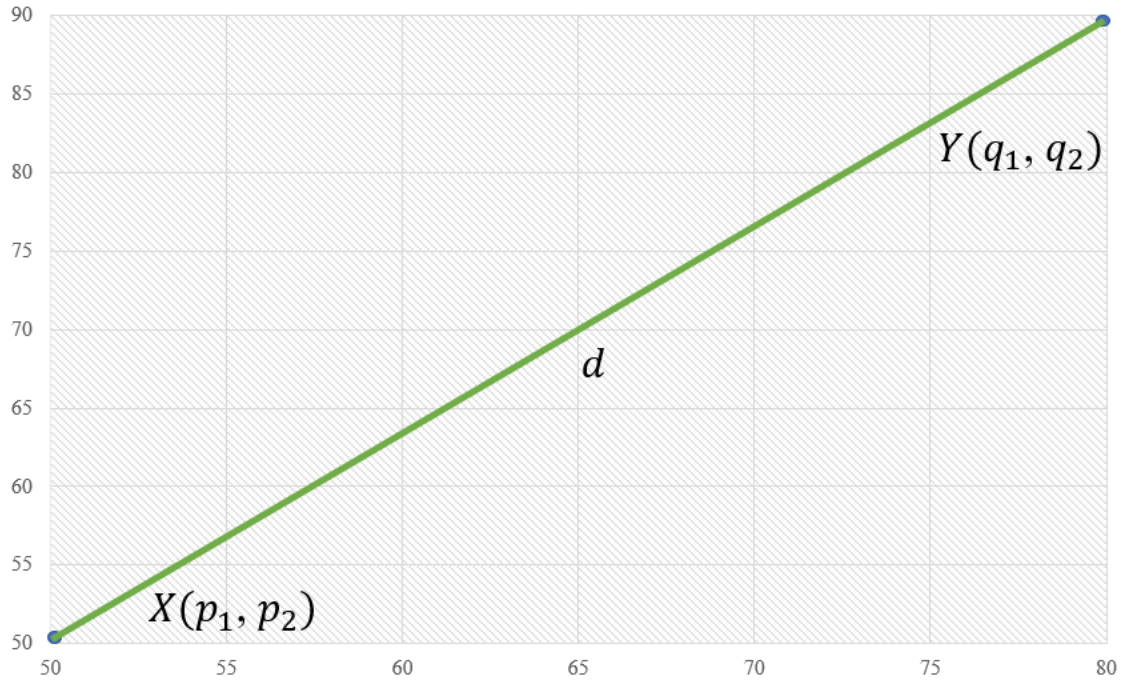
Şekil 16. KNN Algoritması Örneği

KNN algoritmasının etkinliğinde uzaklık ölçüsünün seçimi önemlidir. Manhattan, Mahalanobis, Minkowski ve Euclidean gibi çeşitli ölçüm teknikleri bulunmaktadır. Euclidean uzaklık ölçüm en çok kullanılandır. Euclidean uzaklık ölçüm tekniğinin diğerlerinden daha fazla kullanılmasının sebeplerinden biri ise KNN algoritmasında uygulanan diğer faklı ölçüm tekniklerine göre daha iyi ve etkin sonuçlar veriyor olmasıdır (Jaskowiak ve Campello, 2011).

- Euclidean Uzaklığı

Euclidean Uzaklığı iki nokta arasındaki en kısa mesafeyi temsil eder. Çoğu makine öğrenme algoritması, gözlemler arasındaki benzerliği ölçmek için bu mesafe metriğini kullanır.

Şekil 17'de X ve Y noktaları arasındaki Euclidean uzaklığı gösterilmiştir.



Şekil 17. Euclidean Uzaklığı

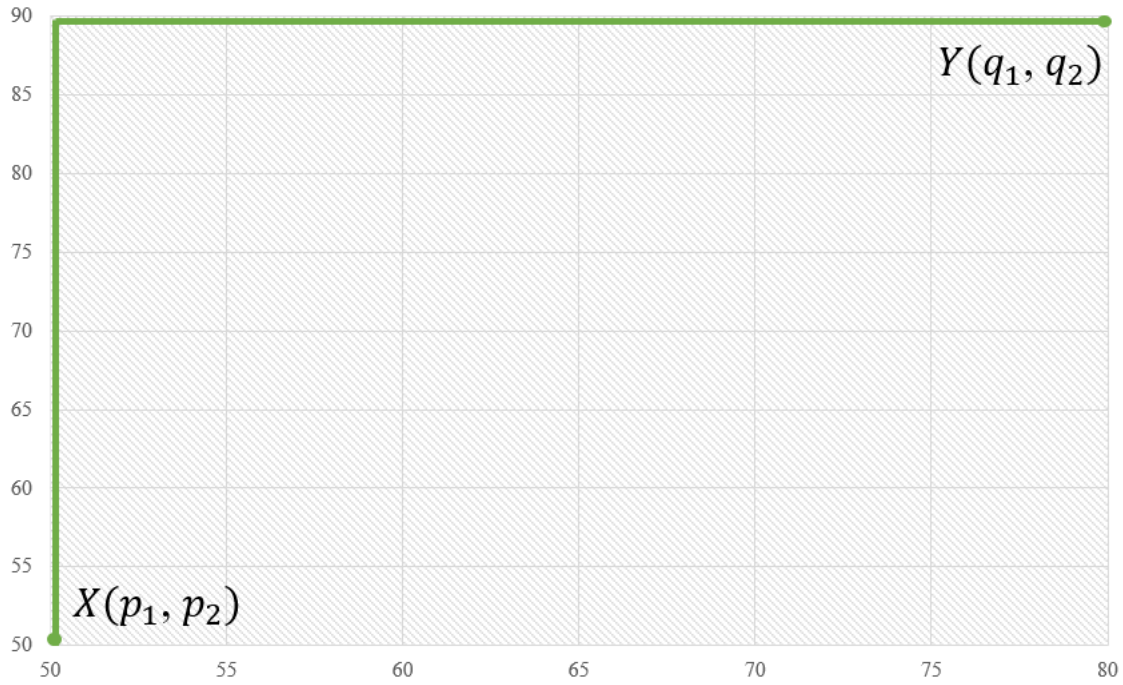
Aşağıda Euclidean uzaklığı formülü gösterilmiştir.

$$d = ((p_1 - q_1)^2 + (p_2 - q_2)^2)^{1/2}$$

- Manhattan Uzaklığı

Manhattan Uzaklığı, tüm boyutlardaki noktalar arasındaki mutlak farkların toplamıdır.

Şekil 18'de Manhattan Uzaklığını hesaplamak için mutlak mesafelerin toplamının hem x hem de y yönlerinde alındığı gösterilmiştir.



Şekil 18. Manhattan Uzaklığı

Aşağıda Manhattan uzaklığı formülü gösterilmiştir.

$$d = |p_1 - q_1| + |p_2 - q_2|$$

- Minkowski Uzaklığı

Minkowski Mesafesi, Euclidean ve Manhattan Mesafesinin genelleştirilmiş şeklidir.

Aşağıda Minkowski uzaklığı formülü gösterilmiştir.

$$D = \left(\sum_{i=1}^{n} |p_i - q_i|^p \right)^{1/p}$$

- Hamming Uzaklığı

Hamming Uzaklığı, aynı uzunluktaki iki dizi arasındaki benzerliği ölçer. Aynı uzunluktaki iki dizi arasındaki Hamming Uzaklığı, karşılık gelen karakterlerin farklı olduğu konumların sayısıdır.

KNN algoritmasının avantajları:

- KNN Tembel Öğrenen (Örnek tabanlı öğrenme) olarak adlandırılır. Eğitim döneminde hiçbir şey öğrenmez. Eğitim verilerinden ayırt edici bir işlev türetmez. Başka bir deyişle, bunun için eğitim süresi yoktur. Eğitim veri kümesini saklar ve yalnızca gerçek zamanlı tahminler yaparken ondan öğrenir. Bu, KNN algoritmasını eğitim gerektiren diğer algoritmalardan çok daha hızlı hale getirir.
- KNN algoritması tahmin yapmadan önce herhangi bir eğitim gerektirmediğinden, algoritmanın doğruluğunu etkilemeyecek yeni veriler sorunsuz bir şekilde eklenebilir.
- KNN'nin uygulanması çok kolaydır. KNN'yi uygulamak için gereken sadece iki parametre vardır. Bunlar K değeri ve mesafe fonksiyonudur.

KNN algoritmasının dezavantajları:

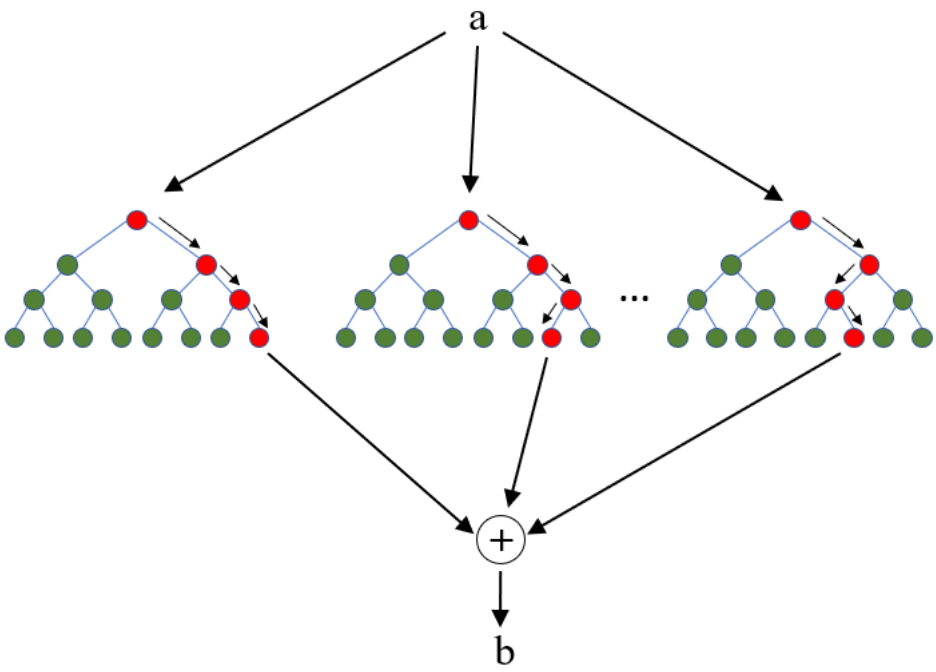
- Büyük veri kümelerinde, yeni nokta ile mevcut her nokta arasındaki mesafeyi hesaplama maliyeti büyüktür ve bu da algoritmanın performansını düşürür.
- KNN algoritması yüksek boyutlu verilerle iyi çalışmaz, çünkü çok sayıda boyutta algoritmanın her boyuttaki mesafeyi hesaplaması zorlaşır.
- KNN algoritmasını herhangi bir veri kümesine uygulamadan önce özellik ölçeklendirme (standardizasyon ve normalleştirme) yapmamız gerekir. Bunu yapmazsak, KNN yanlış tahminler üretebilir.
- KNN, veri kümesindeki gürültüye duyarlıdır. Eksik değerleri manuel olarak atamamız ve aykırı değerleri kaldırmamız gerekir.

Random Forest Sınıflandırma Algoritması

Son 20 yılda, Random Forest sınıflandırıcısı, başarılı sınıflandırma sonuçları ve elde edilen işlem hızı nedeniyle büyük ilgi görmüştür (Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, ve Rigol-Sanchez, 2012). Random Forest sınıflandırıcı, bir dizi karar ağacından gelen tahminleri kullanarak güvenilir sınıflandırmalar sağlar (Breiman, 2001). Ek olarak, bu sınıflandırıcı, hedef sınıflar arasında ayrım yapma konusunda en yüksek yeteneğe sahip özellikleri seçmek ve sıralamak için kolayca kullanılabilir. Bu önemli bir varlıktır çünkü uzaktan algılanan verilerin yüksek boyutu nedeniyle en alakalı değişkenleri seçmek zaman alır (Belgiu ve Dragut, 2016).

Random Forest, hem sınıflandırma hem de regresyon için kullanılan denetimli bir öğrenme algoritmasıdır. Ancak, esas olarak sınıflandırma problemleri için kullanılır. Bildiğimiz gibi bir orman ağaçlardan oluşur ve daha fazla ağaç daha sağlam orman demektir. Benzer şekilde, rastgele orman algoritması veri örnekleri üzerinde karar ağaçları oluşturur ve ardından her birinden tahmin bilgisini alarak son aşamada oylama yoluyla en iyi çözümü seçer. Tek bir karar ağacından daha iyi olan bir topluluk yöntemidir, çünkü sonucun ortalamasını alarak aşırı uyumu azaltır.

Şekil 19 genel Randon Forest mimarisini göstermektedir.



Şekil 19. Random Forest Mimarisi

Random Forest algoritmasının avantajları:

- Karar ağaçlarında ezbere öğrenmeyi azaltır ve doğruluğun artırılmasına yardımcı olur.
- Hem sınıflandırma hem de regresyon problemlerine karşı esnektir.
- Hem kategorik hem de sürekli değerlerle iyi çalışır.
- Verilerde bulunan eksik değerleri otomatikleştirir.

- Kural tabanlı bir yaklaşım kullandığı için verilerin normalleştirilmesi gerekli değildir.

Random Forest algoritmasının dezavantajları:

- Çıktılarını birleştirmek için çok sayıda ağaç oluşturduğundan, kaynakların yanı sıra çok fazla hesaplama gücü gerektirir.
- Sınıfı belirlemek için çok sayıda karar ağacını birleştirdiği için eğitim için de çok zaman gerektirir.
- Karar ağaçları topluluğu nedeniyle, yorumlanabilir olmaktan zarar görür ve her bir değişkenin önemini belirleyemez.

Ridge Sınıflandırma Algoritması

Çoklu regresyon modellerinin katsayılarını tahmin etmek için bağımsız değişkenlerin yüksek oranda korelasyon gösterdiği senaryolarda kullanılan yöntemdir. Kimya, mühendislik ve ekonometri gibi birçok alanda kullanılmıştır. Ridge regresyon, lineer regresyon modellerinin bazı çoklu bağlantılı (yüksek derecede korelasyonlu) bağımsız değişkenlere sahip olduğu durumlarda en küçük kareler tahmin edicilerinin belirsizliğine olası bir çözüm olarak geliştirilmiştir. Varyansı ve ortalama kare tahmincisi genellikle daha önce türetilen en küçük kareler tahmincilerinden daha küçük olduğundan, bu daha kesin bir ridge parametreleri tahmini sağlar (Chang ve Fan, 2002).

Bagging Sınıflandırma Algoritması

Bootstrap Aggregation (Bagging), temel sınıflandırıcıların her birine orijinal veri kümesinin rastgele alt kümelerine uyan ve daha sonra nihai bir tahmin oluşturmak için bireysel tahminlerini (oylama veya ortalamayla) toplayan bir topluluk meta tahmincisidir. Böyle bir meta-tahminci tipik olarak, bir kara kutu tahmincisinin (örneğin, bir karar ağacı) varyansını azaltmanın bir yolu olarak, yapım prosedürüne randomizasyon getirerek ve daha sonra bir topluluk oluşturarak kullanılabilir. Veri kümesinin rastgele alt kümeleri örneklerin rastgele alt kümeleri olarak çizildiğinde, bu algoritma Yapıştırma olarak bilinir. Numuneler değiştirilerek çizilirse, yöntem Torbalama olarak bilinir. Veri kümesinin rasgele alt kümeleri özelliklerin rasgele alt kümeleri olarak çizildiğinde, yöntem Rasgele Alt Alanlar olarak bilinir. Son olarak, temel tahminciler hem örneklerin

hem de özelliklerin alt kümeleri üzerine kurulduğunda, yöntem Rastgele Yamalar olarak bilinir (Sutton, 2005).

AdaBoost Sınıflandırma Algoritması

AdaBoost, hatalarından ders alarak zayıf sınıflandırıcıları geliştirmek için yinelemeli bir yaklaşım kullanan bir topluluk öğrenme sürecidir. Bu algoritma Yoav Freund ve arkadaşları tarafından geliştirilmiştir ve "meta-öğrenme" olarak da bilinir. Paralel birleştirme kullanan rastgele ormanın aksine, Adaboost "sıralı birleştirme" kullanır. Yüksek doğrulukta iyi bir sınıflandırıcı elde etmek için kötü performans gösteren birçok sınıflandırıcıyı birleştirerek güçlü bir sınıflandırıcı oluşturur. Bu anlamda, AdaBoost, sınıflandırıcının verimliliğini önemli ölçüde artırarak uyarlanabilir bir sınıflandırıcı olarak adlandırılır, ancak bazı durumlarda aşırı uyumları tetikleyebilir. AdaBoost, ikili sınıflandırma problemlerinde karar ağaçlarının, temel tahmin edicinin performansını artırmak için en iyi şekilde kullanılır, ancak gürültülü verilere ve aykırı değerlere karşı hassastır (Schapire, 2013).

Çok Katmanlı Algılayıcı

İleri beslemeli yapay sinir ağı olarak da bilinen derin öğrenmenin temel mimarisine çok katmanlı algılayıcı (Multi-Layer Perceptron-MLP) denir. Tipik bir MLP, bir giriş katmanı, bir veya daha fazla gizli katman ve bir çıkış katmanından oluşan tamamen bağlı bir ağdır. Bir katmandaki her düğüm, bir sonraki katmandaki her düğüme belirli bir ağırlıkta bağlanır. MLP, modeli oluştururken ağırlık değerlerini dâhili olarak ayarlamak için bir sinir ağındaki en "temel yapı taşı" olan "Geri Yayılım" tekniğini kullanır. MLP ölçekleme özelliklerine duyarlıdır ve gizli katmanların, nöronların ve yinelemelerin sayısı gibi çeşitli hiper parametrelerin ayarlanmasına izin verir, bu da hesaplama açısından maliyetli bir modelle sonuçlanabilir (Gardner ve Dorling, 1998).

BÖLÜM 2. YÖNTEM VE METOT

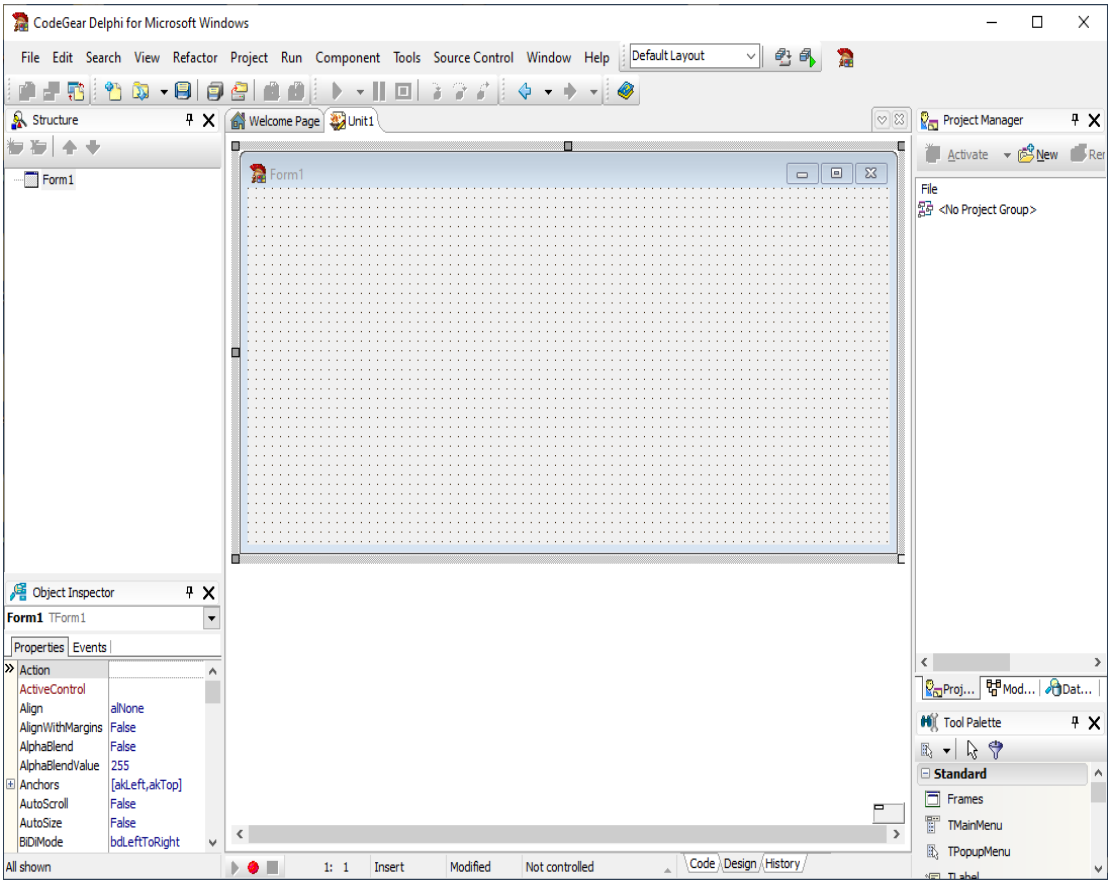
Gümrük beyannamesi yüzlerce bilgi alanından oluşmaktadır (Gümrük Rehberi, 2021) ve (İhracatta Kullanılan Uluslararası Dokümanlar, 2021). Her bir bilgi alanının tek tek doğru bir şekilde doldurulması gerekmektedir. Bu bağlamda geliştirecek modele yönelik her bir bilgi alanı için tahmin yapmalıdır. Bu nedenle her bir veri alanı için bir veri seti oluşturulacaktır. Her bir veri setinin öznitelikleri (Dash ve Liu, 1997) belirlenecektir. Öznitelikleri belirlemede ise Delphi metodu kullanılmıştır (Ameyaw, Hu, Shan, Shan, ve Le, 2016), (Zartha Sossa, Halal, ve Zarta, 2019), (Lund, 2020) ve (Sourani ve Sohail, 2014). Delphi metodunda yapılmak istenen geleceğe yönelik tahmin yapmak, bunun için uzmanlardan yararlanmak ve bir uzlaşma veya karara varmaktır. Burada da gümrük mevzuatı bilgisi olan kişilere bilgi alanı tahmininde kullanılacak özniteliklerin tespiti için sorular sorularak modeller belirlenmiştir. Modellerin ilk eğitim sonuçları değerlendirilerek özniteliklerin üzerinden geçilerek değişiklikler yapılmıştır. Bu süreç kabul edilebilir başarım oranlarına ulaşıncaya kadar tekrarlanarak modellere son hali verilmiştir.

Kullanılan Yazılım Teknolojileri

Delphi

Delphi Object Pascal programlama dilini temel alan bir programlama dilidir. Masaüstü, mobil, web ve konsol için hızlı uygulama geliştirme imkânı sunmaktadır. Şu anda Embarcadero Technologies tarafından geliştirilmekte ve sürdürülmektedir. Delphi derleyicileri Microsoft Windows, MacOS, IOS, Android ve Linux (x64) için yerel kod üretir. Delphi bir kod editörü, bir görsel tasarımcı, entegre bir hata ayıklayıcı, bir kaynak kodu kontrol bileşeni ve üçüncü taraf eklentileri için destek içerir. Kod düzenleyicide kod tamamlama, gerçek zamanlı hata denetimi ve yeniden düzenleme bulunur. Görsel bileşen kütüphanesi, platformlar arası geliştirme seçeneğine ve veri tabanı desteği önemli özellikleridir. Delphi, derleme hızı, yerel kod ve geliştirici üretkenliği sağlar. Tam nesne yönelimli bir dildir (Delphi (software) - Wikipedia, 2022).

Aşağıda yer alan Şekil 20'de Delphi geliştirme ortamı gösterilmektedir.

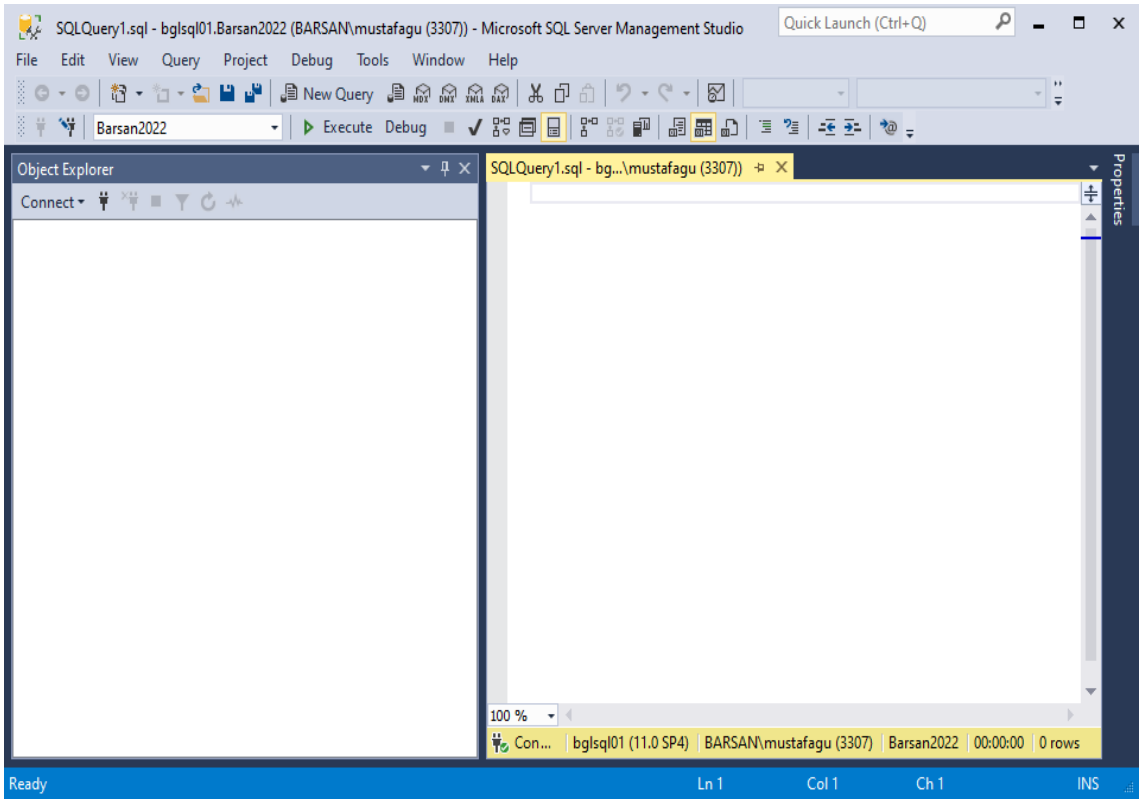


Şekil 20. Delphi Geliştirme Ortamı

SQL Server

Microsoft SQL Server, Microsoft tarafından geliştirilen ilişkisel bir veri tabanı yönetim sistemidir. Bir veri tabanı sunucusu olarak çalışır. Aynı bilgisayarda veya bir ağ üzerinden başka bir bilgisayarda çalışabilen diğer yazılım uygulamaları tarafından talep edilen verileri depolamak ve almak birincil işlevidir. Küçük tek makineli uygulamalardan çok sayıda eşzamanlı kullanıcılı büyük uygulamalara kadar değişen iş yüklerine yönelik farklı sürümleri bulunmaktadır (Microsoft SQL Server - Wikipedia, 2022). MS SQL sunucusunda veriler kullanıcılar tarafından görülebilecek şekilde mantıksal bileşenler halinde VT üzerinde tutulur (Şahinaslan ve Şahinaslan, 2022)

Aşağıda yer alan Şekil 21'de SQL Server arayüzü gösterilmektedir.



Şekil 21. SQL Server Arayüzü

Python

Python üst düzey, genel amaçlı bir programlama dilidir. Tasarım felsefesi, önemli girinti kullanımıyla kod okunabilirliğini vurgular. Dil yapıları ve nesne yönelimli yaklaşımı, programcıların küçük ve büyük ölçekli projeler için net, mantıksal kod yazmasına yardımcı olmayı amaçlamaktadır. Veri bilimi, makine öğrenimi uygulamaları, büyük ölçekli veri işleme, tahmine dayalı analitik için gerekli kütüphanelere sahiptir (Python (programming language) - Wikipedia, 2022).

Şekil 22'de Python kod örneği gösterilmektedir.

```python
# -*- coding: utf-8 -*-
"""
Created on Tue Apr 05 09:51:02 2022

@author: mustafagu
"""

from sklearn.naive_bayes import GaussianNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.linear_model import RidgeClassifier
from sklearn.ensemble import AdaBoostClassifier
from sklearn.neural_network import MLPClassifier

import pandas as pd
import pypyodbc
from sklearn.model_selection import train_test_split

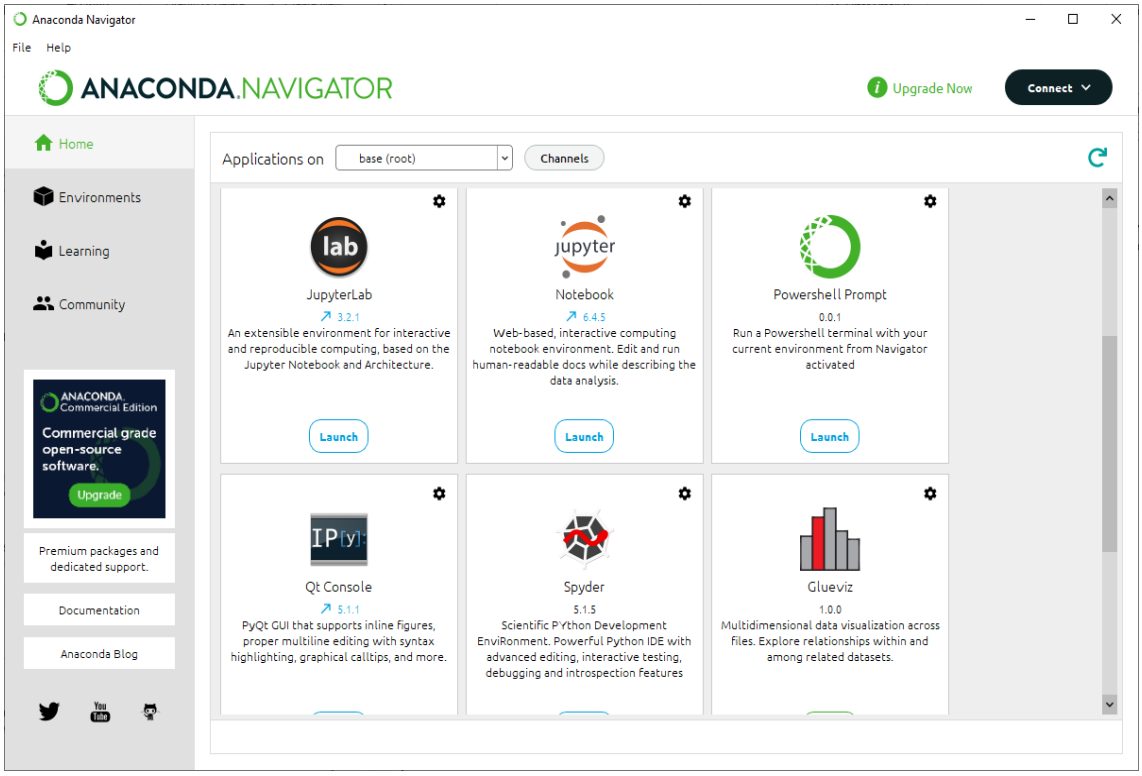
connection=pypyodbc.connect('DRIVER=SQL SERVER;SERVER=ServerName;DATABASE=DBName;UID=UId;PWD=Password')
c = connection.cursor()
c.execute('''
          SELECT M.firmnumber
               , customsoffice = CONVERT(INT, P1.aciklama)
               , regime = CONVERT(INT, P2.aciklama)
               , declaration1 = CONVERT(INT, P3.aciklama)
               , declaration2 = CONVERT(INT, P4.aciklama)
               , entryoffice = CONVERT(INT, P5.aciklama)
               , internalvehiclemode = CONVERT(INT, P6.aciklama)
               , bordervehiclemode = CONVERT(INT, P7.aciklama)
               , bordervehiclecode
          FROM MLGumrukDosyaTasimaSekli M
          LEFT OUTER JOIN gnlparm1 P1 ON (P1.modul = 'ML' AND P1.grup = 'CUSTOFFICE' AND P1.kod = M.customsoffice)
          LEFT OUTER JOIN gnlparm1 P2 ON (P2.modul = 'ML' AND P2.grup = 'REGIME' AND P2.kod = M.regime)
          LEFT OUTER JOIN gnlparm1 P3 ON (P3.modul = 'ML' AND P3.grup = 'DECL1' AND P3.kod = M.declaration1)
          LEFT OUTER JOIN gnlparm1 P4 ON (P4.modul = 'ML' AND P4.grup = 'DECL2' AND P4.kod = M.declaration2)
          LEFT OUTER JOIN gnlparm1 P5 ON (P5.modul = 'ML' AND P5.grup = 'CUSTOFFICE' AND P5.kod = M.entryoffice)
          LEFT OUTER JOIN gnlparm1 P6 ON (P6.modul = 'ML' AND P6.grup = 'VEHMODE' AND P6.kod = M.internalvehiclemode)
          LEFT OUTER JOIN gnlparm1 P7 ON (P7.modul = 'ML' AND P7.grup = 'VEHMODE' AND P7.kod = M.bordervehiclemode)
          ''')
```

Şekil 22. Python Kod Örneği

Anaconda

Anaconda, bilimsel bilgi işlem için Python ve R programlama dillerinin paket yönetimini ve dağıtımını basitleştirmeyi amaçlayan ücretsiz ve açık kaynaklı bir dağıtımıdır. Dağıtım, Windows, Linux ve macOS için uygun veri bilimi paketlerini içerir. Anaconda'daki paket sürümleri, paket yönetim sistemi conda tarafından yönetilir. Anaconda, otomatik olarak yüklenen 250'den fazla paketle birlikte gelir ve 7.500'den fazla ek açık kaynak paketi yüklenebilir. Anaconda Navigator, kullanıcıların komut satırı komutlarını kullanmadan uygulamaları başlatmasına ve conda paketlerini, ortamlarını ve kanallarını yönetmesine olanak tanıyan Anaconda dağıtımında bulunan bir masaüstü grafik kullanıcı arabirimidir. Spyder, Anaconda Navigator'da varsayılan uygulamalardan biridir (Anaconda (Python distribution) - Wikipedia, 2022).

Şekil 23'te Anaconda Navigator görünümü gösterilmektedir. Şekil 23'te Anaconda Navigator'da varsayılan olarak bulunan bazı uygulamalar görünmektedir.

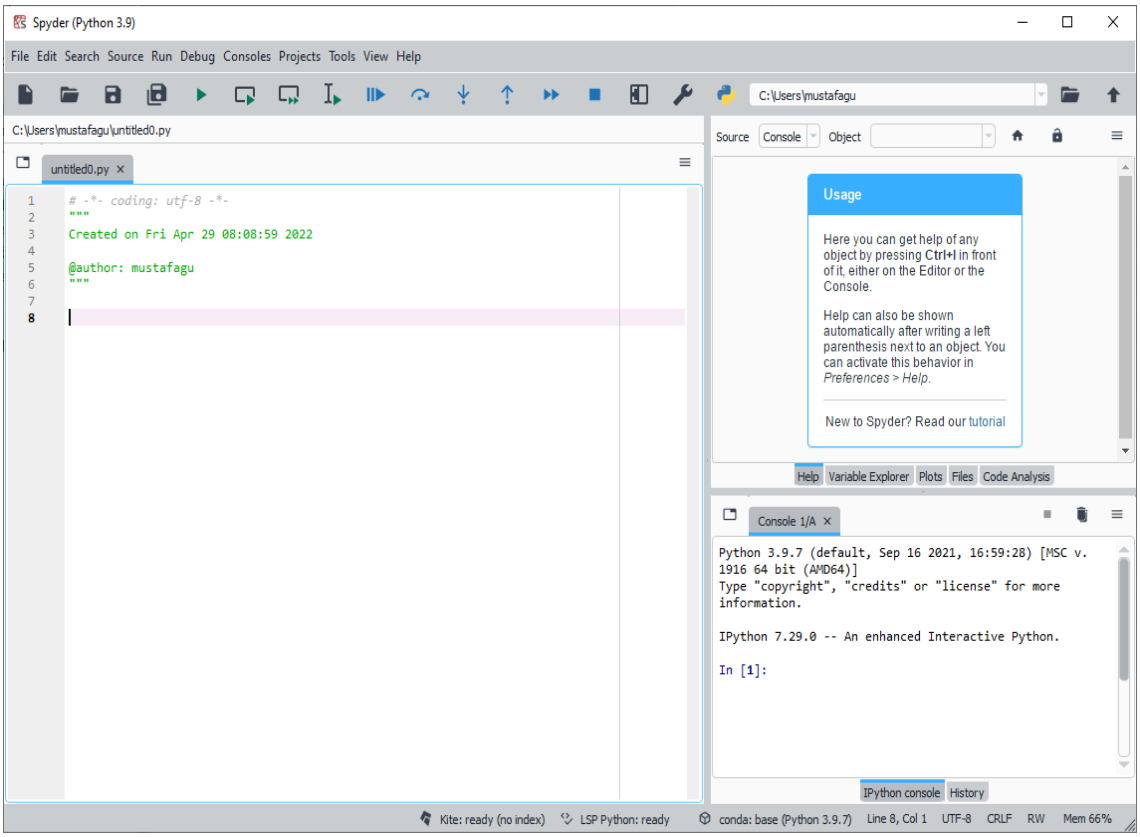


Şekil 23. Anaconda Navigator Görünümü

Spyder

Spyder, Python dilinde bilimsel programlama için açık kaynaklı bir platformlar arası tümleşik geliştirme ortamıdır. Spyder, NumPy, SciPy, Matplotlib, pandas, IPython, SymPy, Cython ve diğer açık kaynaklı yazılımlar dâhil olmak üzere bilimsel uygulamalarda kullanılabilecek bir dizi önemli paketi içerir. Eklentilerle genişletilebilir (Spyder (software) - Wikipedia, 2022).

Şekil 24'te Spyder geliştirme ortamı gösterilmektedir.



Şekil 24. Spyder Geliştirme Ortamı

Araştırma Modeli

Gümrük beyanname bileşenlerinin tahminlemesi için oluşturulan veri setleri üzerinde makine öğrenmesi ve derin öğrenme algoritmaları çalıştırılmıştır. Veri setleri uluslararası bir lojistik firmasının 2010 ve 2021 yılları arasındaki 12 yıllık gümrük beyanname verilerinden oluşturulmuştur. Veri setleri içerisindeki çıktılara uygun öznitelikler mevzuat ekibi ile Delphi yöntemi kullanılarak belirlenmiştir. Belirlenen veri seti eğitim sonuçlarına göre mevzuat ekibi ile öznitelikler iyileştirilerek başarım oranları iyileştirilmiş ve kabul edilebilir oranlara ulaşılmıştır.

Evren ve Örneklem

Gümrük beyannamesinden veri tahminlemesi yapabilmek için makine öğrenmesi modelleri oluşturulmuştur. Bu modellerin verileri için SQL Server veri tabanında tablolar açılmıştır. Bu tablolar aynı veri tabanındaki beyanname bilgileri ile doldurulmuştur. Beyanname verileri 2010 ve 2021 yılları arasındaki 12 yıllık süredeki işlemleri içermektedir.

Veriler ve Toplanması

Tablo 4'te taşıma şekli bilgisinin tahminlemesi için kullanılacak verilerden her yıla ait bir örnek gösterilmektedir ve veri setinde 4.064.131 adet kayıt mevcuttur. Bordervehiclecode çıktı bilgisidir. Firmnumber, customsoffice, regime, declaration1, declaration2, entryoffice, internalvehiclemode ve bordervehiclemode öznitelik bilgileridir.

Tablo 4. Taşıma Şekli Tablosu

id	Key			Öznitelikler								Çıktı
	year	department	filenumber	firmnumber	customsoffice	regime	declaration1	declaration2	entryoffice	internalvehiclemode	bordervehiclemode	bordervehiclecode
1	2010	AVR	7010	1	341300	1040	EU	1	220200	TIR	TIR	30
277237	2011	AND	136	9200	343200	1000	EU	1	343200	UCAK	UCAK	40
582715	2012	AND	84	3870	343100	4000	IM	4	343100	GEMI	GEMI	10
1139021	2013	TRK	16514	132	590400	7100	ANT	7	220400	VAGON	VAGON	20
1177333	2014	GMR	182	5	160200	7100	ANT	7	220200	KAMYON	KAMYON	30
1728758	2015	GEB	4565	10218	410300	1000	EX	1	410300	GEMI	GEMI	10
2118563	2016	ISN	7871	57803	341300	4071	EU	4	220200	TIR	TIR	30
2598235	2017	EGE	33456	62041	350300	1000	EU	1	350300	GEMI	GEMI	10
3008710	2018	ISN	5037	57803	341300	7100	ANT	7	343400	KAMYON	KAMYON	17
3009107	2019	GEB	577	62849	410500	7100	ANT	7	410300	TIR	GEMI	10
3587757	2020	TRK	22091	132	590400	4071	EU	4	220400	VAGON	VAGON	20
4064129	2021	MRS	7588	61263	330100	3151	EU	3	330100	TIR	GEMI	17

firmnumber : Gümrük beyannamesinin hangi firmaya ait olduğu bilgisini içerir.

customsoffice : Gümrük beyannamesinin hangi gümrükte işleme alındığı bilgisini içerir.

regime : Gümrük beyannamesinde hangi rejime göre işlem yapıldığı bilgisini içerir.

declaration1 : Rejim bilgisine ait detay bilgidir.

declaration2 : Rejim bilgisine ait detay bilgidir.

entryoffice : Giriş çıkış gümrük bilgisini içerir.

internalvehiclemode : Sınırdaki aracın türü bilgisini içerir.

bordervehiclemode : Çıkıştaki aracın türü bilgisini içerir.

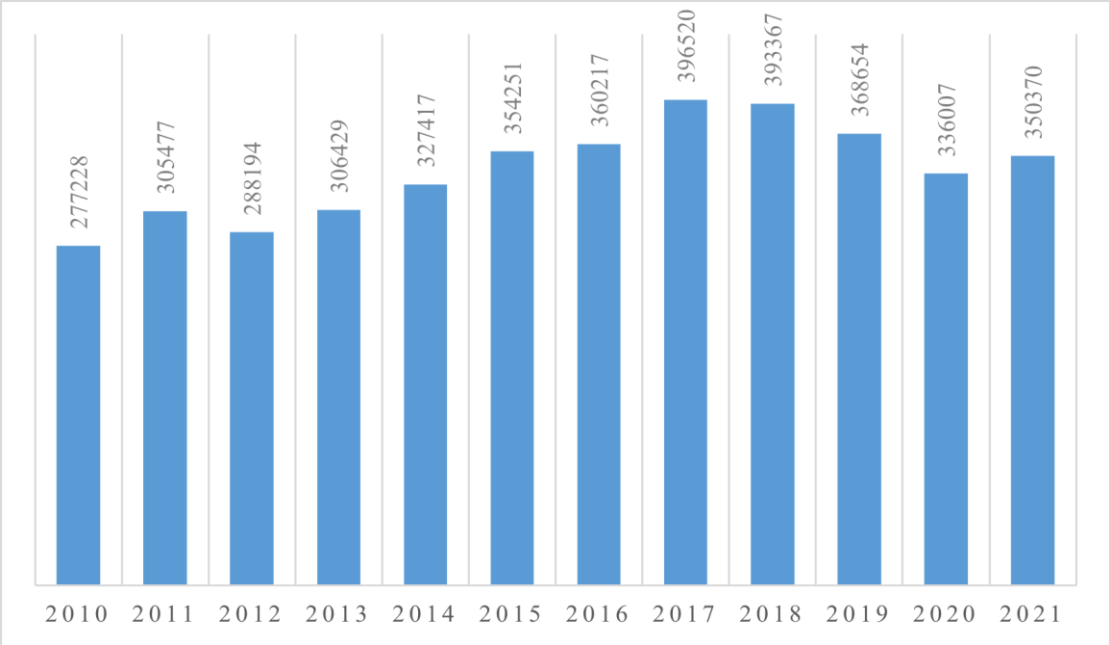
bordervehiclecode : Çıktı bilgisidir ve taşıma şekli kodu bilgisini içerir.

Tablo 5'te taşıma şekli çeşitleri gösterilmektedir.

Tablo 5. Taşıma Şekli Çeşitleri

Kod	**Açıklama**
10	Deniz yolu ile taşıma
12	Deniz aracında demiryolu vagonu
16	Deniz aracında motorlu karayolu taşıma aracı
17	Deniz aracında römork veya yarı römork
18	Deniz aracında, iç sularda taşımaya özgü araç
20	Demiryolu ile yapılan taşıma
23	Demiryolu vagonunda karayolu taşıma aracı
30	Karayolu ile yapılan taşıma
40	Havayolu ile yapılan taşıma
50	Posta ile yapılan taşıma
70	Sabit taşıma tesisatı
80	İç sularda yapılan taşıma
90	Kendinden hareketli

Şekil 25'te taşıma şekli veri setinde yıllara göre kayıt sayıları gösterilmektedir.



Şekil 25. 2010-2021 Taşıma Şekli Veri Seti Beyanname Sayıları

Tablo 6'da çıktı bilgisi olan taşıma şekli kayıt sayıları azalan sırada listelenmiştir.

Tablo 6. Taşıma Şekli Kayıt Sayıları

Taşıma Şekli Kodu	Kayıt Sayısı
30	2064538
10	830793
40	625132
17	346499
16	110061
20	86385
50	245
90	199
12	139
23	87
18	50
70	3

Tablo 7'de döviz türü bilgisinin tahminlemesi için kullanılacak verilerden her yıla ait bir örnek gösterilmektedir ve 4.065.936 adet kayıt mevcuttur. Currencytype çıktı bilgisidir. Firmnumber, customsoffice, regime, declaration1, declaration2 ve tradecountry öznitelik bilgileridir.

Tablo 7. Döviz Türü Tablosu

id	Key			Öznitelikler						Çıktı
	year	department	filenumber	firmnumber	customsoffice	regime	declaration1	declaration2	tradecountry	currencytype
267068	2010	AND	42256	15233	350300	7100	ANT	7	006	GBP
582864	2011	MRS	8976	10149	330100	3151	EU	3	006	GBP
607127	2012	AND	22657	8	341200	4071	EU	4	001	TL
1100825	2013	ISN	1012	45614	330100	4000	IM	4	039	EUR
1461578	2014	TRK	37891	42641	590400	7100	ANT	7	001	EUR
1798258	2015	AND	9232	41493	340900	4000	IM	4	720	USD
2220857	2016	AVR	19802	6877	410500	4071	IM	4	732	JPY
2421174	2017	MRS	28807	10149	340300	3151	EU	3	006	GBP
2918335	2018	GEB	59906	8090	410300	5100	IM	5	005	TL
3010920	2019	AVR	1509	4500	340300	4000	IM	4	720	USD
3517913	2020	TRK	9795	132	590400	4000	EU	4	005	EUR
3715569	2021	GMR	92066	109	160100	1000	EU	1	001	EUR

firmnumber : Gümrük beyannamesinin hangi firmaya ait olduğu bilgisini içerir.

customsoffice : Gümrük beyannamesinin hangi gümrükte işleme alındığı bilgisini içerir.

regime : Gümrük beyannamesinde hangi rejime göre işlem yapıldığı bilgisini içerir.

declaration1 : Rejim bilgisine ait detay bilgidir.

declaration2 : Rejim bilgisine ait detay bilgidir.

tradecountry : Ticaret yapılan ülkesi bilgisini içerir.

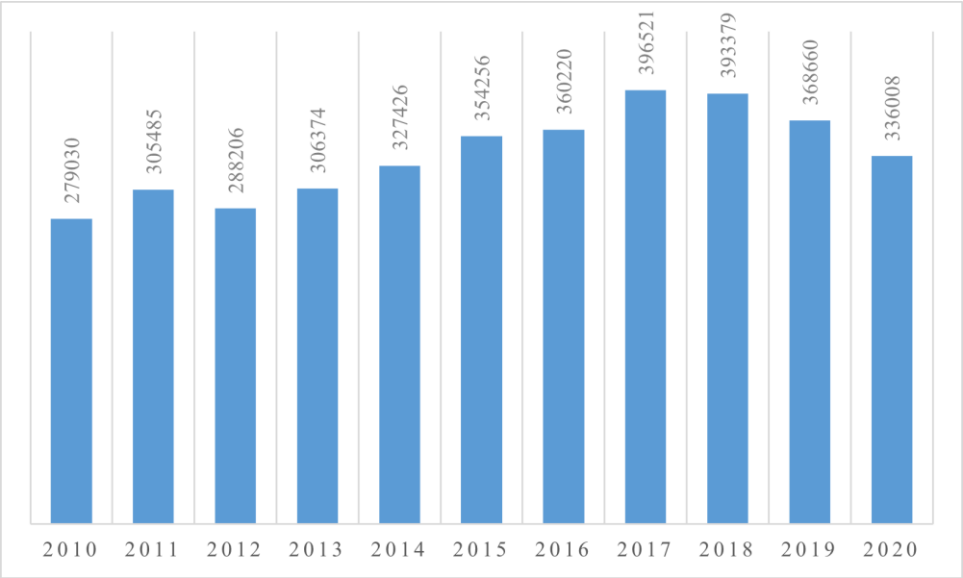
currencytype : Çıktı bilgisidir ve döviz türü bilgisini içerir.

Tablo 8'de döviz türü çeşitleri gösterilmektedir.

Tablo 8. Döviz Türü Çeşitleri

Kod	Açıklama
AED	Birleşik Arap Emirlikleri Dirhemi
AUD	Avustralya Doları
BGN	Bulgar Levası
BRL	Brezilya Reali
BYR	Beyaz Rusya Rublesi
CAD	Kanada Doları
CHF	İsviçre Frangı
CNY	Çin Yuan Renminbisi
CZK	Çek Kronu
DKK	Danimarka Kronu
EGP	Mısır Paundu
EUR	Euro
GBP	İngiliz Sterlini
GEL	Gürcistan Larisi
HKD	Hong Kong Doları
HUF	Macar Forinti
ILS	İsrail Yeni Şekeli
INR	Hindistan Rupisi
IQD	Irak Dinarı
IRR	İran Riyali
JOD	Ürdün Dinarı
JPY	Japon Yeni
KRW	Kore Vonu
KWD	Kuveyt Dinarı
KZT	Kazakistan Tengesi
MAD	Fas Dirhemi
MXN	Meksika Pezosu
MYR	Malezya Ringiti
NOK	Norveç Kronu
PKR	Pakistan Rupisi
PLN	Polonya Zilotisi
RON	Yeni Romen Leyi
RUB	Rus Rublesi
SAR	Suudi Arabistan Riyali
SEK	İsveç Kronu
SGD	Singapur Doları
THB	Tayland Bahtı
TL	Türk Lirası
TND	Tunus Dinarı
TWD	Tayvan Yeni Doları
USD	ABD Doları
VND	Vietnam Dongu
ZAR	Güney Afrika Randı

Şekil 26'da döviz türü veri setinde yıllara göre kayıt sayıları gösterilmektedir.



Şekil 26. 2010-2021 Döviz Türü Veri Seti Beyanname Sayıları

Tablo 9'da çıktı bilgisi olan döviz türü kayıt sayıları azalan sırada listelenmiştir.

Tablo 9. Döviz Türü Kayıt Sayıları

Döviz Türü Kodu	Kayıt Sayısı
EUR	3163907
USD	657430
TL	141410
GBP	44269
JPY	31734
CHF	10874
SEK	5298
RUB	2352
DKK	1701
GEL	1350
CNY	1339
RON	930
AED	926
PLN	794
NOK	726
AUD	444
CAD	401
SAR	18
INR	12
ZAR	12
CZK	7
HUF	2

Tablo 10'da sevk ülke bilgisinin tahminlemesi için kullanılacak verilerden her yıla ait bir örnek gösterilmektedir ve 4.031.130 adet kayıt mevcuttur. Dispatchcountry çıktı bilgisidir. Firmnumber, customsoffice, regime, declaration1, declaration2, entryoffice, incotermcode, exitcountry, tradecountry ve destinationcountry öznitelik bilgileridir.

Tablo 10. Sevk Ülke Tablosu

id	Key			Öznitelikler										Çıktı
	year	department	filenumber	firmnumber	customsoffice	regime	declaration1	declaration2	entryoffice	incoterm code	exitcountry	tradecountry	destinationcountry	dispatchcountry
1	2010	AVR	291	1	341300	1040	EU	1	220200	FCA	052	003	003	003
274931	2011	AND	25	29724	341200	4071	EU	4	220200	CIP	004	004	052	004
862306	2012	TRK	54195	1614	341600	7100	ANT	7	343100	DAP	003	003	052	003
1123955	2013	MRS	15185	10239	330100	1000	EU	1	330100	CIF	052	004	004	004
1166449	2014	GMR	220	5	160200	4071	EU	4	220200	FCA	004	004	052	004
1815030	2015	ISN	9882	56619	330900	4000	EU	4	340300	DAP	032	032	052	032
1842685	2016	GMR	1295	5	340300	1000	EU	1	340300	FCA	052	004	004	004
2559102	2017	TRK	38696	64542	341300	1000	EU	1	221300	DAP	052	063	063	063
2977073	2018	ANK	865	61504	341300	7100	ANT	7	220500	EXW	009	009	052	009
2977116	2019	GEB	902	10581	410500	4071	EU	4	343400	EXW	011	011	052	001
3345416	2020	EGE	62089	15	350300	1000	EX	1	350300	FOB	052	624	624	624
3680848	2021	GMR	2739	5	341453	1000	EX	1	341453	FCA	052	728	728	728

firmnumber : Gümrük beyannamesinin hangi firmaya ait olduğu bilgisini içerir.

customsoffice : Gümrük beyannamesinin hangi gümrükte işleme alındığı bilgisini içerir.

regime : Gümrük beyannamesinde hangi rejime göre işlem yapıldığı bilgisini içerir.

declaration1 : Rejim bilgisine ait detay bilgidir.

declaration2 : Rejim bilgisine ait detay bilgidir.

entryoffice : Giriş çıkış gümrük bilgisini içerir.

incotermcode : Teslim şekli bilgisini içerir.

exitcountry : Çıkış ülke bilgisini içerir.

tradecountry : Ticaret yapılan ülke bilgisini içerir.

destinationcountry : Gideceği ülke bilgisini içerir.

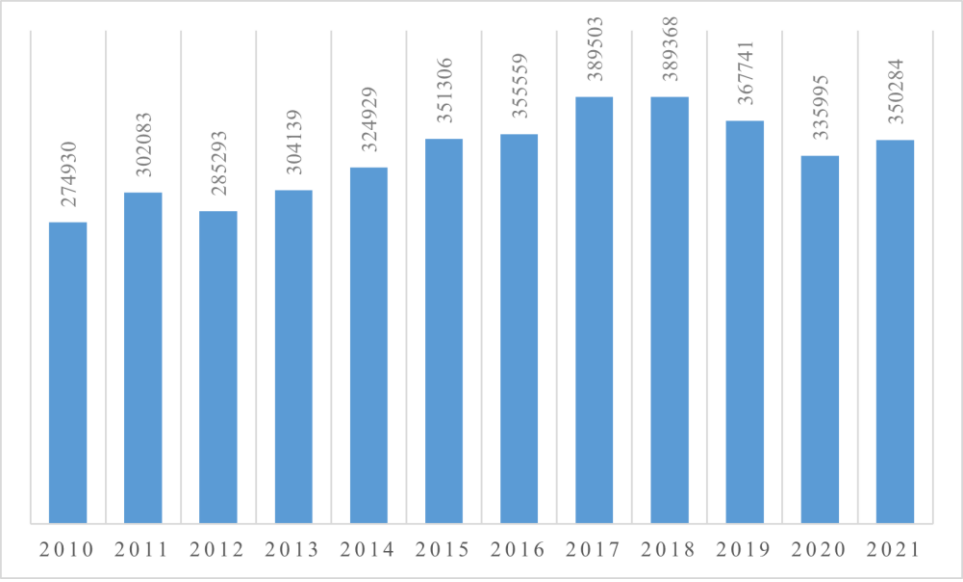
dispatchcountry : Çıktı bilgisidir ve sevk ülke bilgisini içerir.

Tablo 11'de en çok kullanılan 14 sevk ülkesi listelenmektedir.

Tablo 11. Sevk Ülke Listesi

Kod	Adı
001	Fransa
003	Hollanda
004	Almanya
005	İtalya
006	Birleşik Krallık
011	İspanya
017	Belçika
060	Polonya
061	Çek Cumhuriyeti
064	Macaristan
066	Romanya
400	Birleşik Devletler
720	Çin (Halk Cumhuriyeti)
732	Japonya

Şekil 27'de sevk ülke veri setinde yıllara göre kayıt sayıları gösterilmektedir.



Şekil 27. 2010-2021 Sevk Ülke Veri Seti Beyanname Sayıları

Tablo 12'de çıktı bilgisi olan sevk ülke kayıt sayıları azalan sırada listelenmiştir. Tablo 12'de en çok kullanılan 14 ülke koduna yer verilmiştir.

Tablo 12. Sevk Ülke Kayıt Sayıları

Sevk Ülke Kodu	Kayıt Sayısı
004	1243835
001	288758
005	276405
066	167664
060	146053
006	137482
011	131600
720	128899
017	107943
003	90597
400	78760
732	76374
061	67500
064	60967

Tablo 13'te birim bilgisinin tahminlemesi için kullanılacak verilerden her yıla ait bir örnek gösterilmektedir ve veri setinde 22.015.778 adet kayıt mevcuttur. Unit çıktı bilgisidir. Firmnumber, customsoffice, regime, declaration1, declaration2, hscode ve statisticalunit öznitelik bilgileridir.

Tablo 13. Birim Tablosu

	Key				Öznitelikler							Çıktı
id	year	department	filenumber	linenumber	firmnumber	customsoffice	regime	declaration1	declaration2	hscode	statisticalunit	unit
8	2010	AND	127	1	5999	341200	4071	EU	4	8708.99.97.99.19	C62	C62
1340911	2011	AVR	105	1	11	343100	7100	ANT	7	4810.19.90.90.00	KGM	KGM
2851220	2012	GMR	51681	5	5	340300	4000	EU	4	8542.31.90.00.00	C62	C62
4349837	2013	ANK	2395	4	48266	341200	1000	EX	1	3917.39.00.80.00	MTR	MTR
6186407	2014	GMR	70080	2	5	160200	6321	EU	6	8466.93.70.00.00	SET	SET
8395124	2015	MRS	7750	1	10149	341300	3151	EU	3	5516.41.00.90.00	MTK	MTK
9241746	2016	AVR	40794	19	10084	341300	4071	EU	4	2813.10.00.00.00	GRM	GRM
10725703	2017	AND	9	15	47852	340900	4000	IM	4	8467.92.00.00.00	SET	SET
15076242	2018	ISN	19299	2	46762	440200	1000	EU	1	6004.10.00.00.19	KGM	KGM
17304015	2019	TRK	7018	1	132	343100	5100	EU	5	7306.30.11.90.00	MTR	MTR
18595047	2020	EGE	55840	2	60522	352200	4000	IM	4	7601.20.80.00.00	KGM	KGM
21698216	2021	GMR	101876	49	5	341200	7100	ANT	7	8708.30.91.00.00	SET	SET

firmnumber : Gümrük beyannamesinin hangi firmaya ait olduğu bilgisini içerir.

customsoffice : Gümrük beyannamesinin hangi gümrükte işleme alındığı bilgisini içerir.

regime : Gümrük beyannamesinde hangi rejime göre işlem yapıldığı bilgisini içerir.

declaration1 : Rejim bilgisine ait detay bilgidir.

declaration2 : Rejim bilgisine ait detay bilgidir.

hscode : Tarife bilgisini içerir.

statisticalunit : İstatistiki birim bilgisini içerir.

unit : Çıktı bilgisidir ve birim kodu bilgisini içerir.

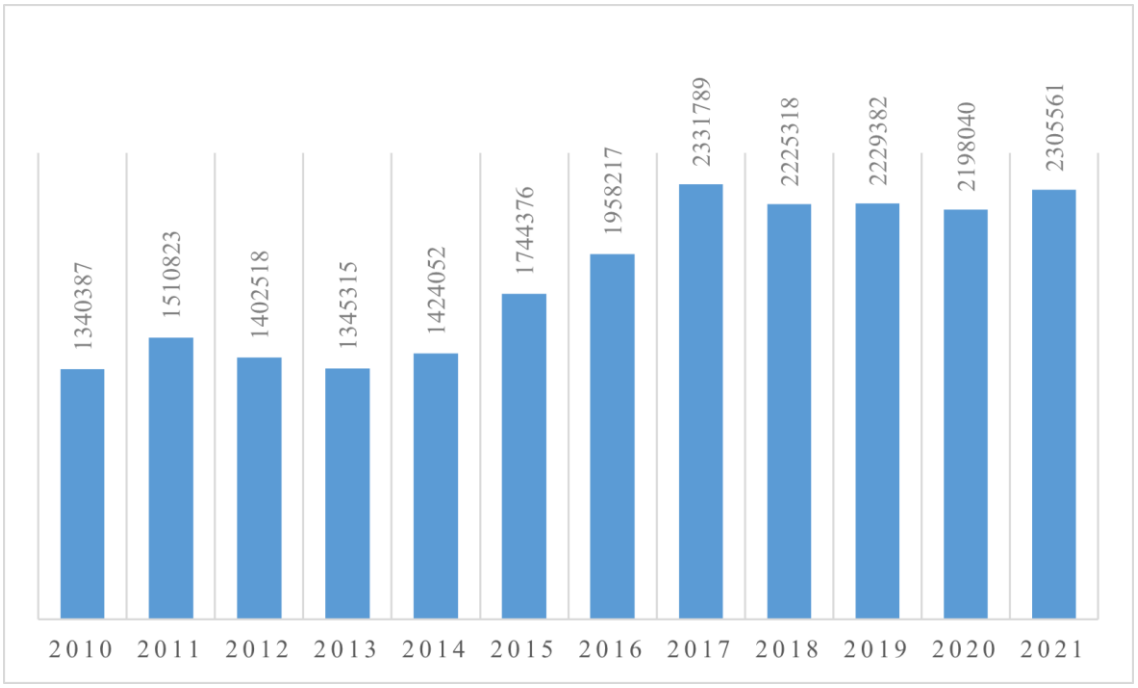
Tablo 14'te birim çeşitleri gösterilmektedir.

Tablo 14. Birim Çeşitleri

Kod	**Açıklama**
ADET	Adet (Dikkat Sadece Serbest Bölge İşlem Formu İçindir)
AYR	Altın Ayarı
B32	KG-Metre Kare
BAS	Baş
BRUTKG	Brüt Kilogram
C62	Adet (Unit)
CCT	Ton Başına Taşıma Kapasitesi
CIFT	Cift (Dikkat Sadece Gözetim İçin)
CPR	Adet-Çift
D30	Brüt Kalori Değeri
D40	Bin Litre
GFI	Fissile İzotop Gramı
GMS	Gümüş
GRM	Gram
GT	Gross Ton
H62	Yüz Adet
K20	Kilogram Potasyum Oksit
K58	Kurutulmuş Net Ağırlıklı Kilogramı
K62	Kilogram-Adet
KFO	Difosfor Pentaoksit Kilogramı
KGM	KGM
KH6	Kilogram-Baş
KHO	Hidrojen Peroksit Kilogramı
KMA	Metil Aminlerin Kilogramı
KNI	Azotun Kilogramı
KOH	Kilogram Potasyum Hidroksit
KPH	Kg Potasyum Oksid
KPR	Kilogram-Çifr
KSD	%90 Kuru Ürün Kilogramı

KSH	Sodyum Hidroksit Kilogramı
KUR	Uranyum Kilogramı
KWH	Kilowatt Saat
KWT	Kilowatt
LPA	Saf Alkol Litresi
LTR	Litre
MTK	Metre Kare
MTQ	Metre Küp
MTR	Metre
NCL	Hücre Adedi
NCR	Karat
NETKG	Net Kilogram
OMV	ÖTV Maktu Vergi
OTB	ÖTV Ölçü Birimi
PR	Çift
R9	Bin Metre Küp
RLO	Rulo
SET	Set
T3	Bin Adet
TWH	Bin Kilowatt Saat

Şekil 28'de birim veri setinde yıllara göre kayıt sayıları gösterilmektedir.



Şekil 28. 2010-2021 Birim Veri Seti Beyanname Sayıları

Tablo 15'de çıktı bilgisi olan birim kayıt sayıları azalan sırada listelenmiştir.

Tablo 15. Birim Kayıt Sayıları

Birim Kodu	Kayıt Sayısı
C62	20274915
KGM	1047388
MTR	264474
MTK	228062
SET	79198
GRM	45762
PR	39467
LTR	22429
MTQ	8424
KSD	2600
NCL	776
KFO	761
KHO	396
K58	277
T3	224
CPR	184
KSH	178
KNI	79
BAS	76
B32	29
KPH	26
KMA	20
ADET	9
H62	7
GT	5
D30	4
NCR	4
K62	2
KH6	1
CCT	1

Tablo 16'da uluslararası anlaşma bilgisinin tahminlemesi için kullanılacak verilerden her yıla ait bir örnek gösterilmektedir ve veri setinde 22.000.635 adet kayıt mevcuttur.

Prefrequestcode çıktı bilgisidir. Firmnumber, customsoffice, regime, declaration1, declaration2 ve exitcountry öznitelik bilgileridir.

Tablo 16. Uluslararası Anlaşma Tablosu

id	Key				Öznitelikler						Çıktı
	year	department	filenumber	linenumber	firmnumber	customsoffice	regime	declaration1	declaration2	exitcountry	prefrequestcode
9	2010	AND	22070	1	96	351900	3151	EX	3	052	DI
1705036	2011	AVR	46800	4	502	341300	5100	EU	5	005	AT
2903529	2012	TRK	3325	1	132	341600	4071	EU	4	011	AKCT
4367233	2013	AVR	5520	2	15415	343100	4000	IM	4	700	GYU
6980154	2014	TRK	48777	1	132	343100	4000	IM	4	728	KORE
8729961	2015	MRS	2703	1	38834	330100	4000	IM	4	624	ISR
10638222	2016	TRK	42968	1	132	590400	5171	EU	5	039	EFTA
12899836	2017	TRK	48399	1	2278	343100	4000	IM	4	701	MY
13950437	2018	EGE	31095	1	15	341300	4000	IM	4	098	SRB
17382935	2019	GMR	12653	2	5	341200	4071	IM	4	212	TNS
18783360	2020	GMR	736	2	5	160200	4071	EU	4	039	PAAMK
20738593	2021	AVR	47515	59	6877	410500	5100	EU	5	006	BK

firmnumber : Gümrük beyannamesinin hangi firmaya ait olduğu bilgisini içerir.

customsoffice : Gümrük beyannamesinin hangi gümrükte işleme alındığı bilgisini içerir.

regime : Gümrük beyannamesinde hangi rejime göre işlem yapıldığı bilgisini içerir.

declaration1 : Rejim bilgisine ait detay bilgidir.

declaration2 : Rejim bilgisine ait detay bilgidir.

exitcountry : Çıkış ülkesi bilgisini içerir.

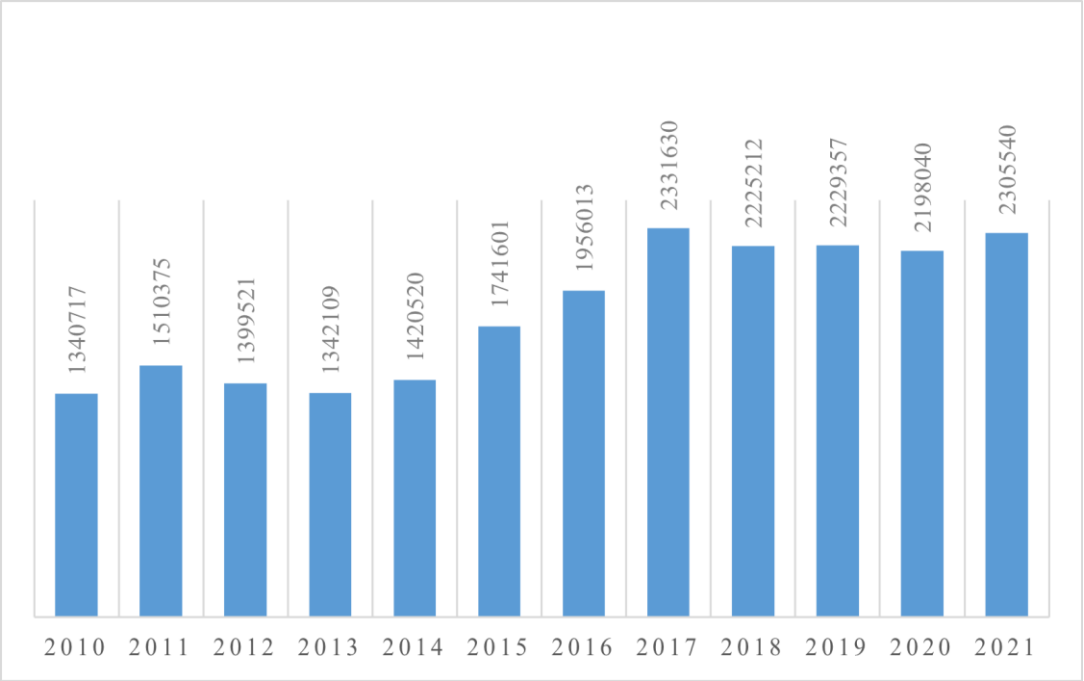
prefrequestcode : Çıktı bilgisidir ve uluslararası anlaşma bilgisini içerir.

Tablo 17'de uluslararası anlaşma çeşitleri gösterilmektedir.

Tablo 17. Uluslararası Anlaşma Çeşitleri

Kod	Açıklama
AKCT	Türkiye-AB Avrupa Kömür Çelik Topluluğu Ürünleri STA'sı
ARN	Arnavutluk STA
AT	Türkiye-AB Ortaklık Konseyinin Gümrük Birliği 1/95 sayılı Kararı
ATRM	Türkiye-AB Ortaklık Konseyi-tarım ürünleri ticareti 1/98 sayılı Kararı
BBMK	Batı Balkan Menşe Kümülasyonu Sistemi
BK	Birleşik Krallık STA
BSH	Bosna-Hersek STA
DI	Diğer
EAZG	En Az Gelişmiş Ülkeler
EFTA	EFTA STA
FAS	Fas STA
FIL	Filistin STA
GUR	Gürcistan STA
GYU	Gelişme Yolundaki Ülkeler
IRI	İran TTA
ISR	Israil STA
KAR	Karadağ STA
KORE	G. Kore STA
KOS	Kosova Tek Taraflı Taviz
MD	Moldova STA
MKE	Makedonya STA
MORIT	Morityus STA
MSR	Mısır STA
MY	Malezya STA
OTDU	Özel Teşvik Düzenlemesi Ülkeleri
PAAMK	Pan-Avrupa-Akdeniz Menşe Kümülasyonu Sistemi
SIL	Şili STA
SIN	Singapur STA
SRB	Sırbistan STA
SUR	Suriye STA
TNS	Tunus STA

Şekil 29'da uluslararası anlaşma veri setinde yıllara göre kayıt sayıları gösterilmektedir.



Şekil 29. 2010-2021 Uluslararası Anlaşma Veri Seti Beyanname Sayıları

Tablo 18'de çıktı bilgisi olan uluslararası anlaşma kayıt sayıları azalan sırada listelenmiştir.

Tablo 18. Uluslararası Anlaşma Kayıt Sayıları

Uluslararası Anlaşma Kodu	Kayıt Sayısı
AT	14248579
DI	7513126
KORE	49458
AKCT	47948
EFTA	41329
GYU	38224
EAZG	26008
SRB	7293
MSR	5081
BK	4806
ISR	4030
TNS	3511
MY	2387
BSH	2280
FAS	2144
PAAMK	1812

ATRM	1392
BBMK	527
SUR	202
SIN	175
MKE	92
GUR	84
OTDU	70
ARN	33
MORIT	19
SIL	11
MD	5
KAR	5
KOS	3
FIL	1

Tablo 19'da ödeme şekli bilgisinin tahminlemesi için kullanılacak verilerden her yıla ait bir örnek gösterilmektedir ve 19.941.787 adet kayıt mevcuttur. Paymentcode çıktı bilgisidir. Firmnumber, customsoffice, regime, declaration1, declaration2, bankcode ve featurecode öznitelik bilgileridir.

Tablo 19. Ödeme Şekli Tablosu

id	Key				Öznitelikler							Çıktı
	year	department	filenumber	linenumber	firmnumber	customsoffice	regime	declaration1	declaration2	bankcode	featurecode	paymentcode
1	2010	AND	400	1	192	341200	4071	IM	4	999999999	02	7
1340201	2011	EGE	14	2	441	350300	4000	EU	4	006700682034	01	1
2850522	2012	MRS	13	1	10157	330100	4000	EU	4	020600003038	01	3
4264860	2013	AND	1056	53	8	341200	7100	ANT	7	999999999	01	7
5601685	2014	AND	362	1	39140	341200	4071	EU	4	999999999	02	7
7047984	2015	ANK	2359	1	38636	343100	1000	EX	1	020500007042	01	2
8290583	2016	AVR	29559	1	15415	343100	4000	IM	4	004600788034	02	2
10013579	2017	EGE	50840	54	38343	341300	4000	EU	4	004600408035	01	2
12117135	2018	GEB	29737	262	69611	410500	4071	EU	4	020500134034	01	2
14534960	2019	EGE	45847	1	60522	350300	3151	EU	3	013403020035	01	1
17551279	2020	MRS	1219	8	10281	330100	3151	EX	3	001209492038	01	1
19886450	2021	TRK	9608	1	132	343100	4000	EU	4	003200018034	01	1

firmnumber : Gümrük beyannamesinin hangi firmaya ait olduğu bilgisini içerir.

customsoffice : Gümrük beyannamesinin hangi gümrükte işleme alındığı bilgisini içerir.

regime : Gümrük beyannamesinde hangi rejime göre işlem yapıldığı bilgisini içerir.

declaration1 : Rejim bilgisine ait detay bilgidir.

declaration2 : Rejim bilgisine ait detay bilgidir.

bankcode : Banka kodu bilgisini içerir.

featurecode : Özellik kodu bilgisini içerir.

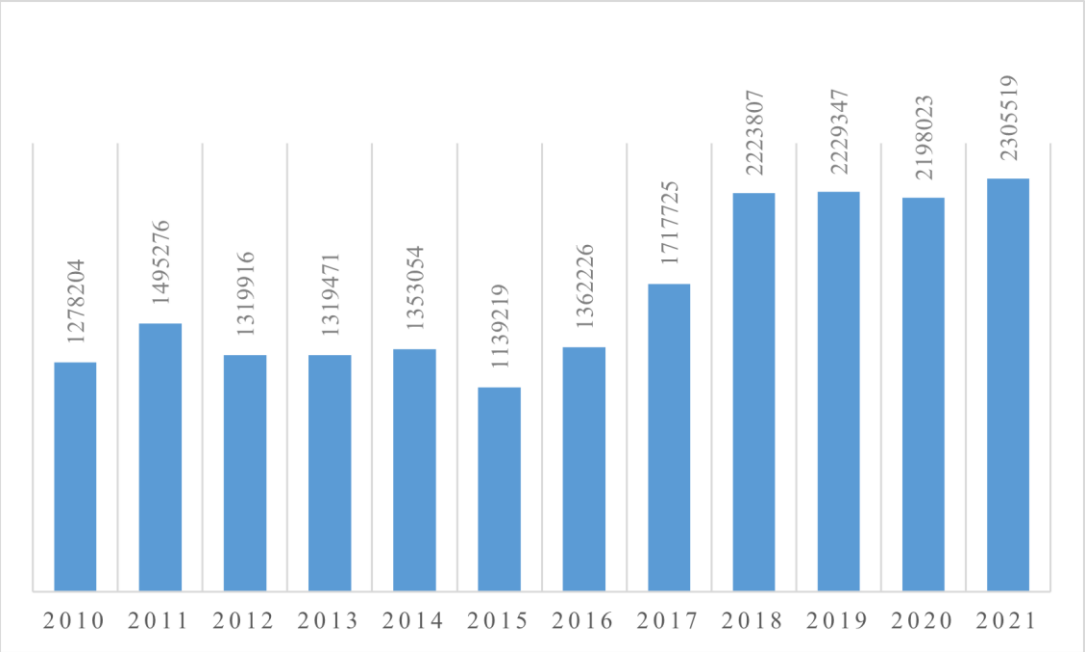
paymentcode : Çıktı bilgisidir ve ödeme şekli bilgisini içerir.

Tablo 20'de ödeme şekli çeşitleri gösterilmektedir.

Tablo 20. Ödeme Şekli Kodları Listesi

Kod	**Adı**
1	Mal Mukabili
11	Özel Takas
12	Akreditif
13	Bağlı Muamele
14	Kabul Kredili Akreditif
15	Kabul Kredili Vesaik Mükabili
16	Kabul Kredili Mal Mükabili
17	Özel Hesap
2	Peşin
3	Vesaik Mükabili
6	Vadeli Akreditif
7	Bedelsiz
4	Peşin (Kağıt Ortamı)

Şekil 30'da ödeme şekli veri setinde yıllara göre kayıt sayıları gösterilmektedir.



Şekil 30. 2010-2021 Ödeme Şekli Veri Seti Beyanname Sayıları

Tablo 21'de çıktı bilgisi olan ödeme şekli kayıt sayıları azalan sırada listelenmiştir.

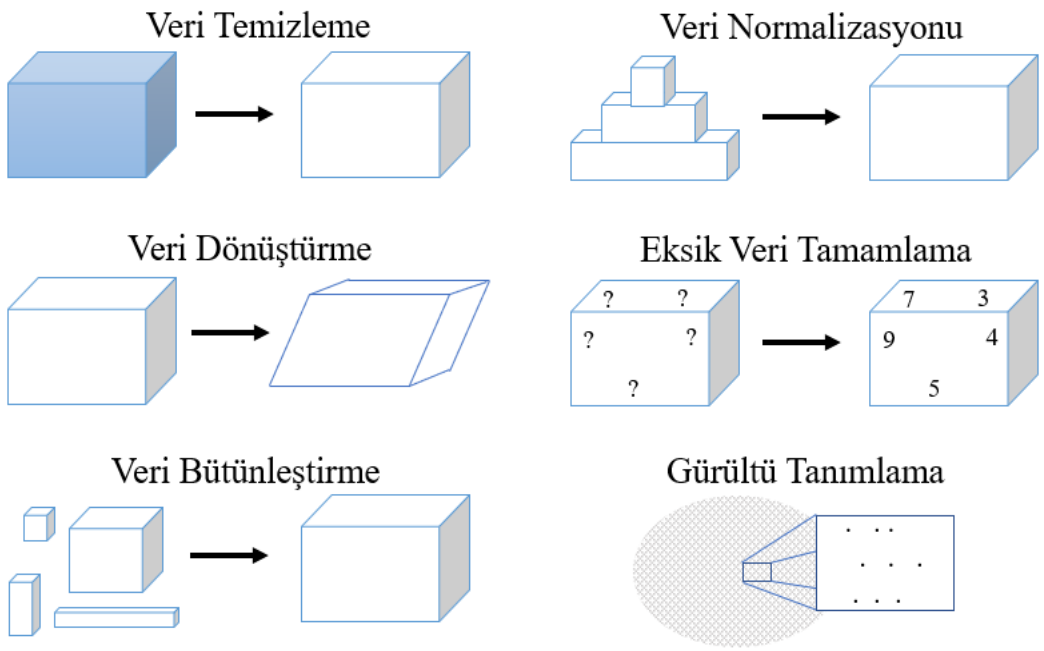
Tablo 21. Ödeme Şekli Kayıt Sayıları

Ödeme Şekli Kodu	Kayıt Sayısı
2	13163952
1	4277779
7	1619856
3	728121
12	99504
6	34283
4	10260
15	7347
14	416
16	248
13	17
17	4

Verilerin Çözümlenmesi ve Yorumlanması

Sınıflandırma algoritmalarının başarımlarının yüksekliği nedeniyle çalışmalar Gaussian Naïve Bayes (Chen, Webb, Linyuan, ve Ma, 2019), Decision Tree sınıflandırma algoritması (Chandrasekar, Qian, Shahriar, ve Bhattacharya, 2017), K-Nearest Neighbors sınıflandırma algoritması (Zhang, Li, Zong, Zhu, ve Wang, 2018), Random Forest sınıflandırma algoritması (Abdulkareem ve Abdulazeez, 2021), Ridge sınıflandırma algoritması (Dobriban ve Wager, 2018), Bagging sınıflandırma algoritması (Roshan ve Asadi, 2020), AdaBoost sınıflandırma algoritması (Zhang, et al., 2019) ve Multi-Layer Perceptron (Tang, Deng, ve Huang, 2016) ile yapılmıştır.

Verilerin önişleme süreci modelin tahminlemesinde önemli bir yer tutmaktadır (Alexandropoulos, Kotsiantis, ve Vrahatis, 2019). Verinin entegrasyonu, temizlenmesi, dönüştürülmesi, azaltılması ve boyutunun küçültülmesi bu aşamada yapılmaktadır (Alasadi ve Bhaya, 2017). Şekil 31'de veri önişleme yöntemleri görsel olarak verilmiştir.



Şekil 31. Veri Önişleme Aşamaları

Tabloların içeriği doldurulurken boş bilgi içeren ve uygun olmayan kayıtlar seçilmeyerek önişleme adımlarına bu aşamada başlanmıştır. Kategorik verilerin çoğu sayısal olmadığı için encoding yöntemi kullanılarak veriler sayısallaştırılarak dönüştürülmüş ve

kullanılabilir hale getirilmiştir (Dahouda ve Joe, 2021). Sayısallaştırma işleminde LabelEncoder metodu (Jackson ve Agrawal, 2019) kullanılmıştır.

Tablo 22'de encoding işlemi sonrasında taşıma şekli veri setinin sayısallaştırılmış hali gösterilmiştir.

Tablo 22. Taşıma Şekli Veri Seti Sayısallaştırması

index	firmnumber	customsoffice	regime	declaration1	declaration2	entryoffice	internalvehiclemode	bordervehiclemode	bordervehiclecode
0	1	3	7	1	2	84	1	1	30
1	1	3	7	1	2	84	1	1	30
2	1	3	7	1	2	84	1	1	30
3	1	3	7	1	2	84	1	1	30
4	1	3	7	1	2	84	1	1	30
5	1	2	16	1	6	2	4	4	40
6	1	2	1	4	1	2	4	4	40
7	1	2	1	1	1	2	4	4	40
8	1	3	1	1	1	84	1	1	30
9	1	2	1	1	1	2	4	4	40

Tablo 23'te encoding işlemi sonrasında döviz türü veri setinin sayısallaştırılmış hali gösterilmiştir.

Tablo 23. Döviz Türü Veri Seti Sayısallaştırılması

index	firmnumber	customsoffice	Regime	declaration1	declaration2	tradecountry	currencytype
0	1	64	4	2	0	2	USD
1	1	64	4	2	0	2	USD
2	1	64	4	2	0	2	USD
3	1	64	4	2	0	2	USD
4	1	64	4	2	0	2	USD
5	1	64	4	2	0	2	USD
6	1	59	9	2	1	40	USD
7	1	59	25	4	3	175	USD
8	1	59	25	2	3	10	USD
9	1	64	25	2	3	26	USD

Tablo 24'te encoding işlemi sonrasında sevk ülkesi veri setinin sayısallaştırılmış hali gösterilmiştir.

Tablo 24. Sevk Ülkesi Veri Seti Sayısallaştırması

index	firmnumber	customsoffice	reg ime	declaration1	declaration2	entry office	incotermcode	exitcountry	tradecountry	destinationcountry	dispatch country
0	1	3	7	1	2	84	2	1	12	19	003
1	1	3	7	1	2	84	2	1	12	19	003
2	1	3	7	1	2	84	5	1	12	19	003
3	1	3	7	1	2	84	5	1	12	19	003
4	1	3	7	1	2	84	5	1	12	19	003
5	1	3	1	1	1	84	10	10	4	1	003
6	1	3	1	1	1	2	10	10	4	1	003
7	1	2	1	1	1	2	10	10	4	1	003
8	1	2	1	1	1	2	10	10	4	1	003
9	1	2	1	4	1	2	10	16	11	1	400

Tablo 25’te encoding işlemi sonrasında birim veri setinin sayısallaştırılmış hali gösterilmiştir.

Tablo 25. Birim Veri Seti Sayısallaştırılması

index	firmnumber	customs office	regime	declaration1	declaration2	hscode	statisticalunit	unit
0	192	33	24	4	3	4921	24	MTR
1	5999	33	24	2	3	4291	5	C62
2	5999	33	24	2	3	1605	5	C62
3	5999	33	24	2	3	1195	5	C62
4	5999	33	24	2	3	5051	5	C62
5	5999	33	24	2	3	4334	5	C62
6	5999	33	24	2	3	1350	5	C62
7	5999	33	24	2	3	5093	5	C62
8	8065	38	39	0	6	2620	5	C62
9	8065	38	39	0	6	3096	5	C62

Tablo 26’da ise encoding işlemi sonrasında uluslararası anlaşma veri setinin sayısallaştırılmış hali gösterilmiştir.

Tablo 26. Uluslararası Anlaşma Veri Seti Sayısallaştırılması

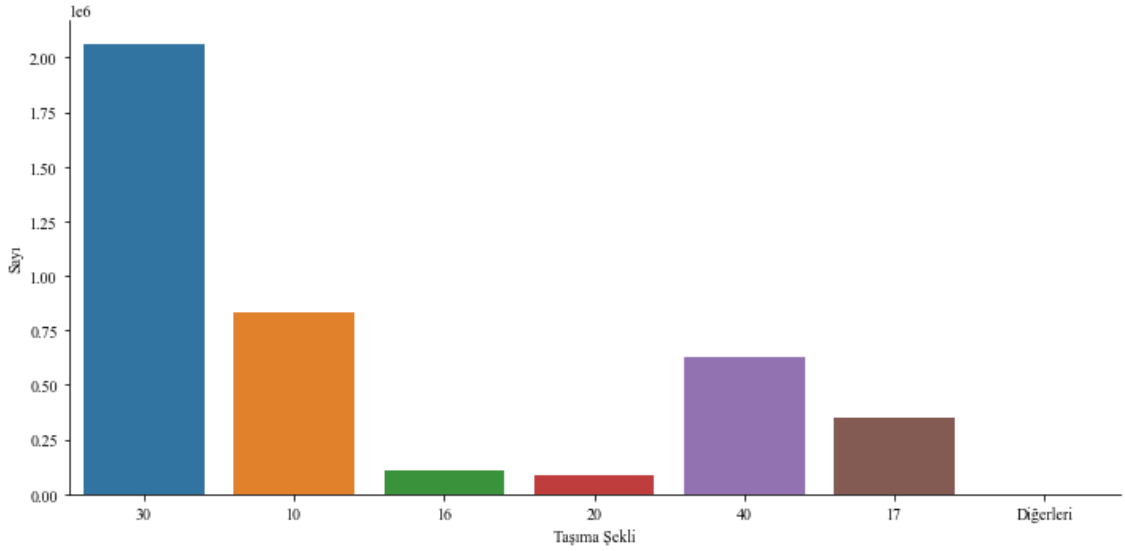
index	firmnumber	customsoffice	regime	declaration1	declaration2	exitcountry	prefrequestcode
0	5999	33	40	0	6	9	AT
1	5999	33	25	2	3	3	AT
2	5999	33	25	2	3	2	DI
3	7375	33	3	3	0	20	DI
4	8	33	25	2	3	0	AT
5	8	33	25	2	3	0	AT
6	8	33	40	0	6	0	AT
7	8	33	40	0	6	0	AT
8	96	50	13	3	2	20	DI
9	5719	34	13	3	2	20	DI

Tablo 27'de encoding işlemi sonrasında ödeme şekli veri setinin sayısallaştırılmış hali gösterilmiştir.

Tablo 27. Ödeme Şekli Veri Seti Sayısallaştırılması

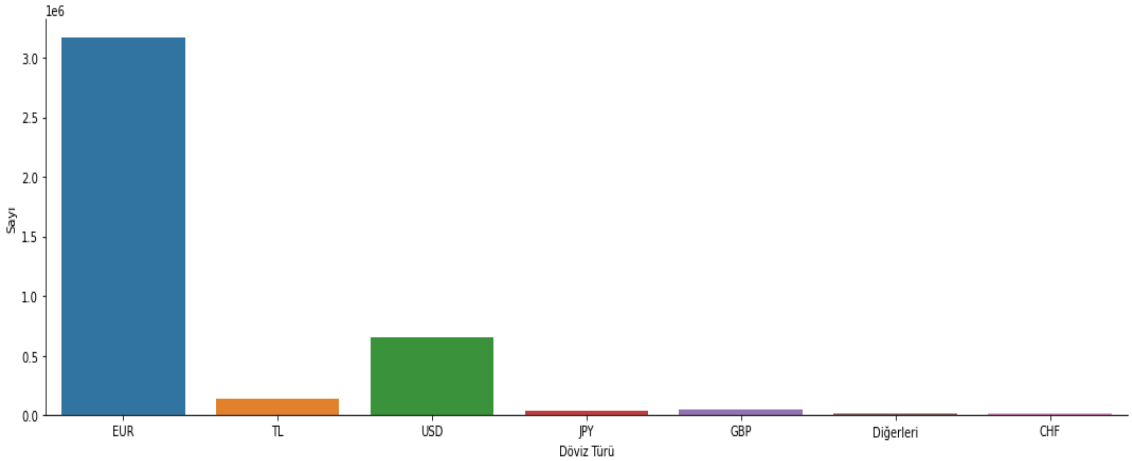
index	firmnumber	customs office	regime	declaration1	declaration2	bankcode	featurecode	payment code
0	192	4	3	4	1	1	2	7
1	192	3	6	1	5	20	1	1
2	192	3	6	1	5	20	1	1
3	192	3	6	1	5	20	1	1
4	192	3	6	1	5	20	1	1
5	192	3	6	1	5	20	1	1
6	192	3	6	1	5	20	1	1
7	10156	11	1	4	1	627	1	2
8	192	3	6	1	5	20	1	1
9	192	3	6	1	5	20	1	1

Büyük veri setleri için uygun olan seaborn kütüphanesi (Waskom, 2021) ile taşıma şekli veri seti görselleştirilerek Şekil 32 elde edilmiştir. 30 kodlu karayolu ile yapılan taşıma en çok kullanılan taşıma şekli olarak görünmektedir.



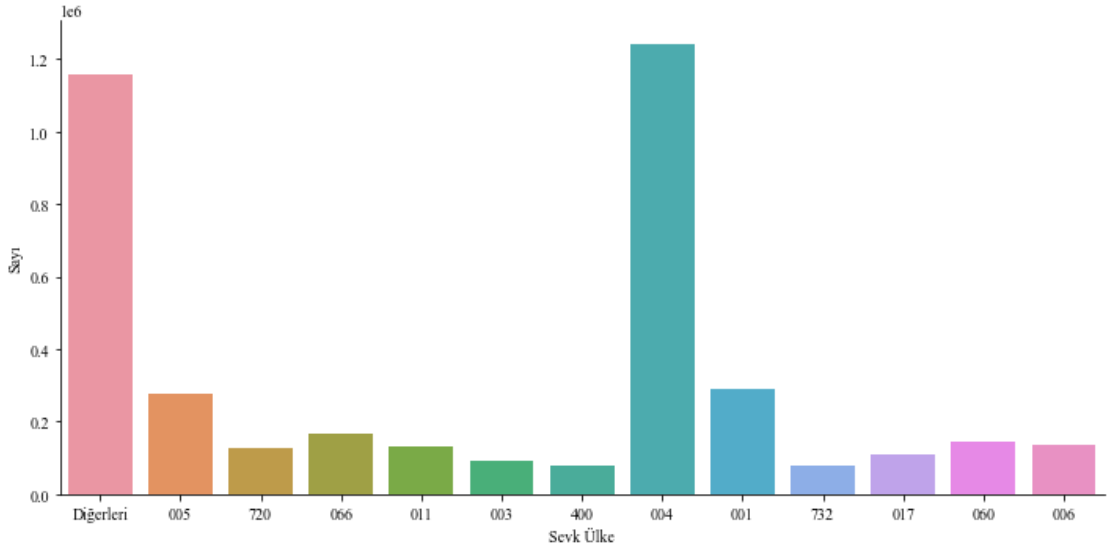
Şekil 32. Taşıma Şekli Sayıları

Seaborn kütüphanesi ile döviz türü veri seti görselleştirilerek Şekil 33 elde edilmiştir. EUR kodlu Euro (Avrupa Ortak Para Birimi) en çok kullanılan döviz türü olarak görünmektedir.



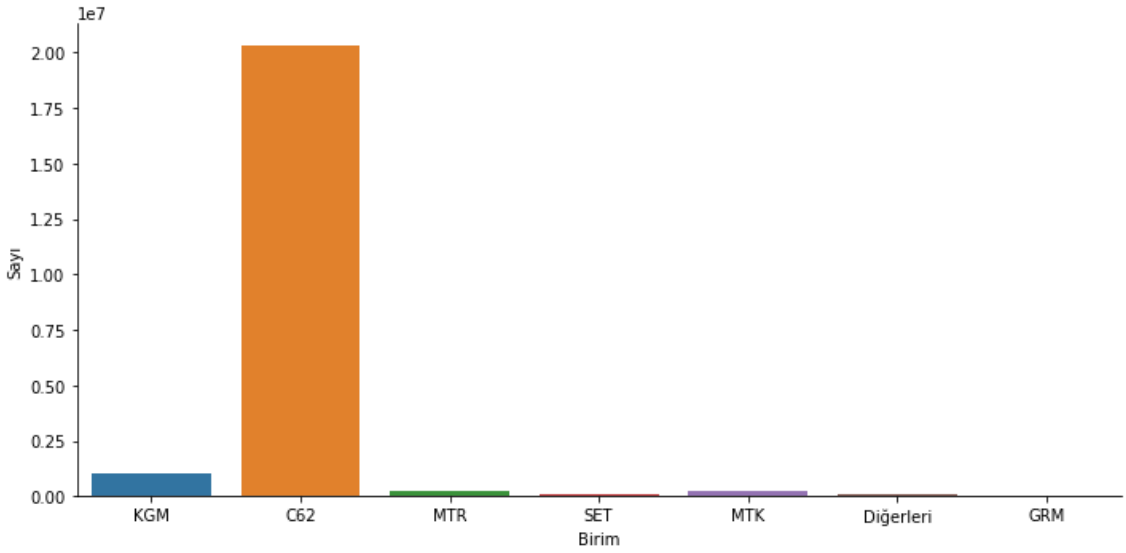
Şekil 33. Döviz Türü Sayıları

Sevk ülke veri seti görselleştirilerek Şekil 34 elde edilmiştir. 004 kodlu Almanya en çok kullanılan sevk ülke olarak görünmektedir.



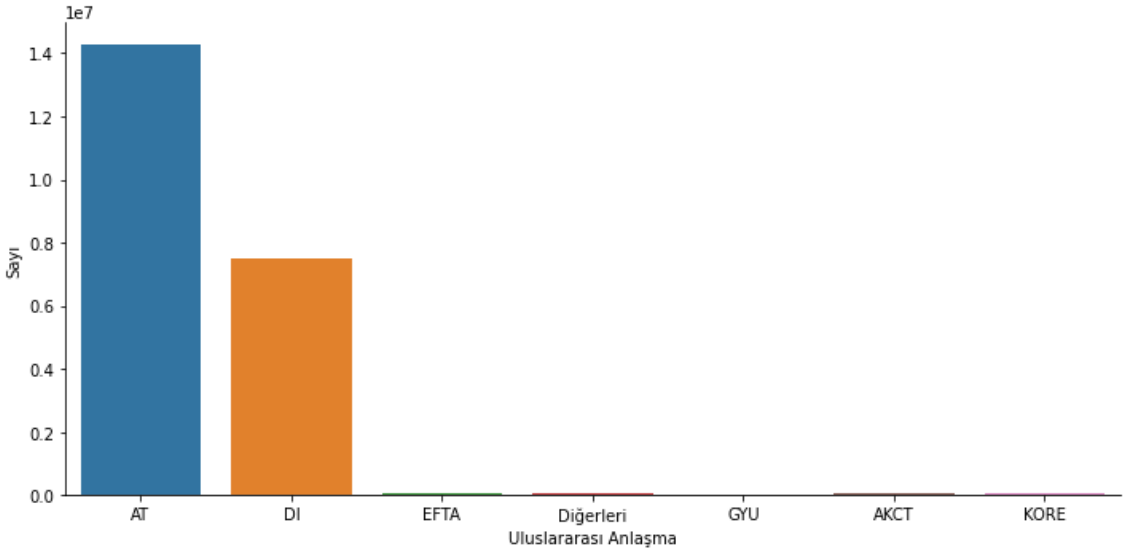
Şekil 34. Sevk Ülke Sayıları

Birim veri seti görselleştirilerek Şekil 35 elde edilmiştir. C62 kodlu Adet (Unit) en çok kullanılan birim olarak görünmektedir.



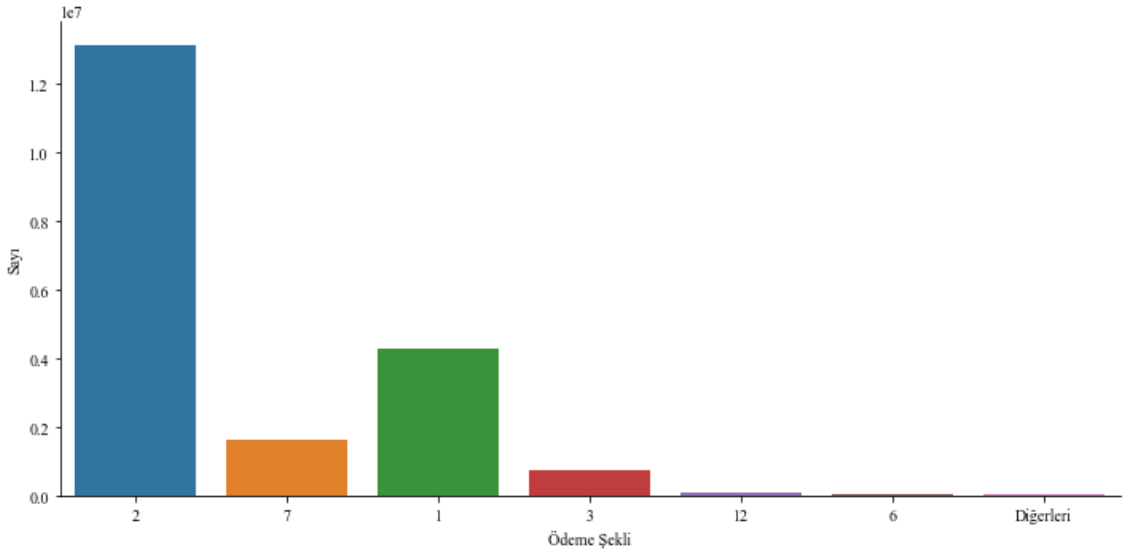
Şekil 35. Birim Sayıları

Uluslararası anlaşma veri seti görselleştirilerek Şekil 36 elde edilmiştir. AT kodlu Türkiye-AB Ortaklık Konseyinin Gümrük Birliği 1/95 sayılı Kararı en çok kullanılan uluslararası anlaşma olarak görünmektedir.



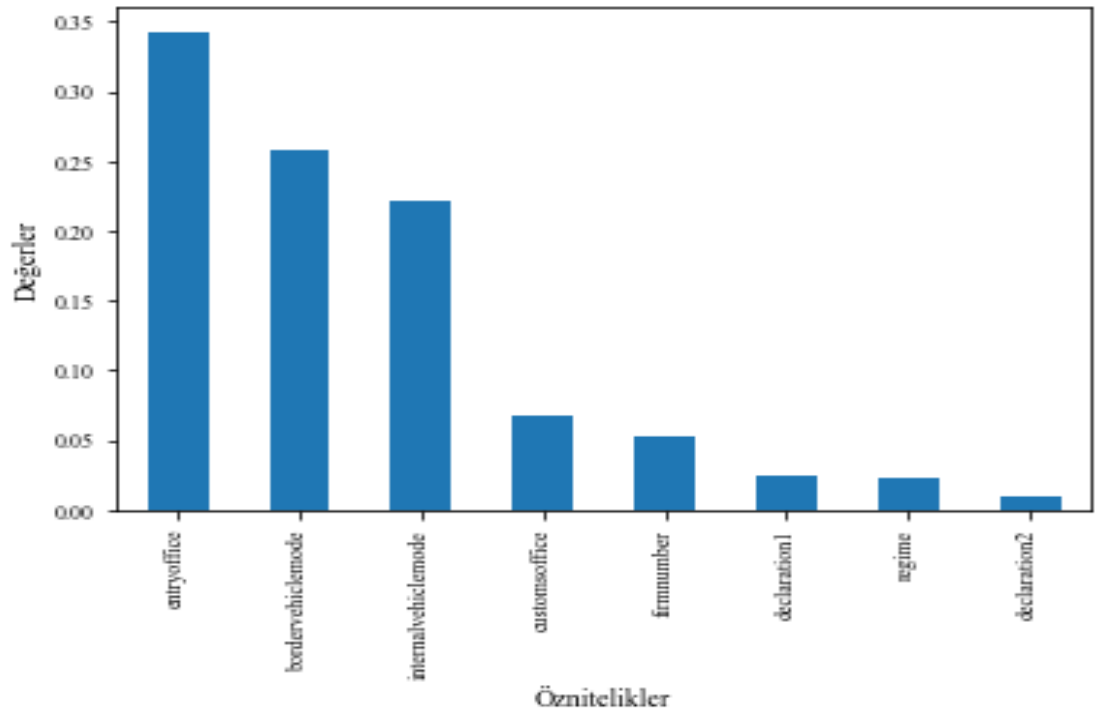
Şekil 36. Uluslararası Anlaşma Sayıları

Ödeme şekli veri seti görselleştirilerek Şekil 37 elde edilmiştir. 2 kodlu peşin en çok kullanılan ödeme şekli olarak görünmektedir.



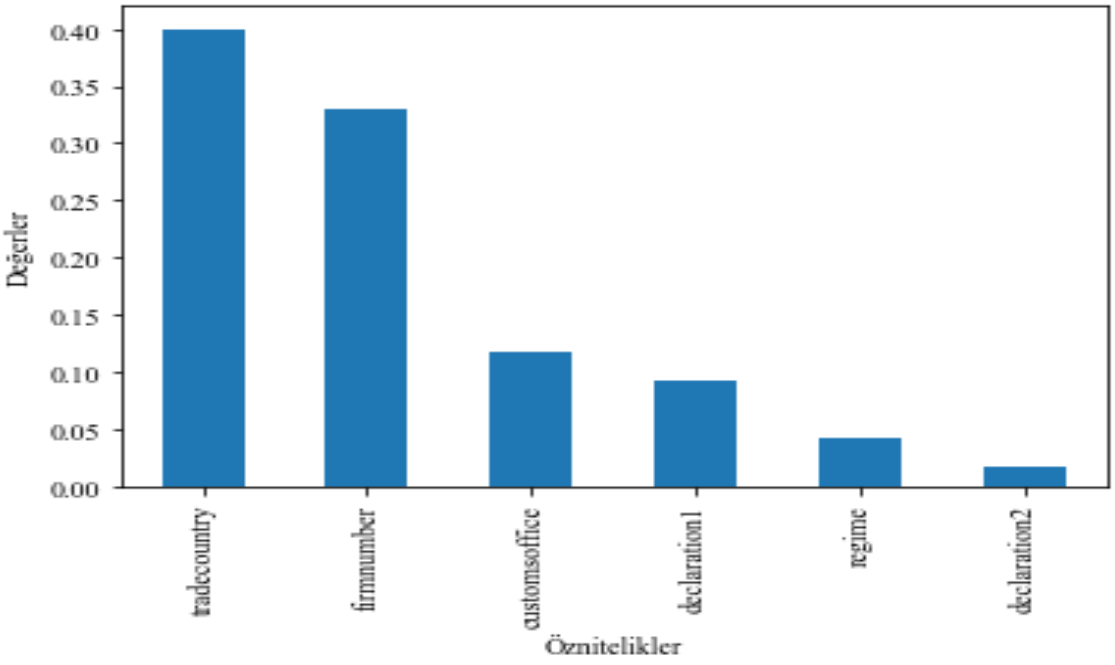
Şekil 37. Ödeme Şekli Sayıları

Özniteliklerin seçimi algoritmaların başarım oranlarında çok etkilidir. Şekil 38'de taşıma şekli modeli için seçilen özniteliklerin etkisi (Zien, Krämer, Sonnenburg, ve Rätsch, 2009) gösterilmiştir. Giriş çıkış gümrüğü etkisi en çok olan öznitelik olarak görünmektedir. Giriş çıkış gümrüğü özniteliğinden sonra çıkıştaki aracın türü etkisi en çok olan ikinci öznitelik olarak görünmektedir.



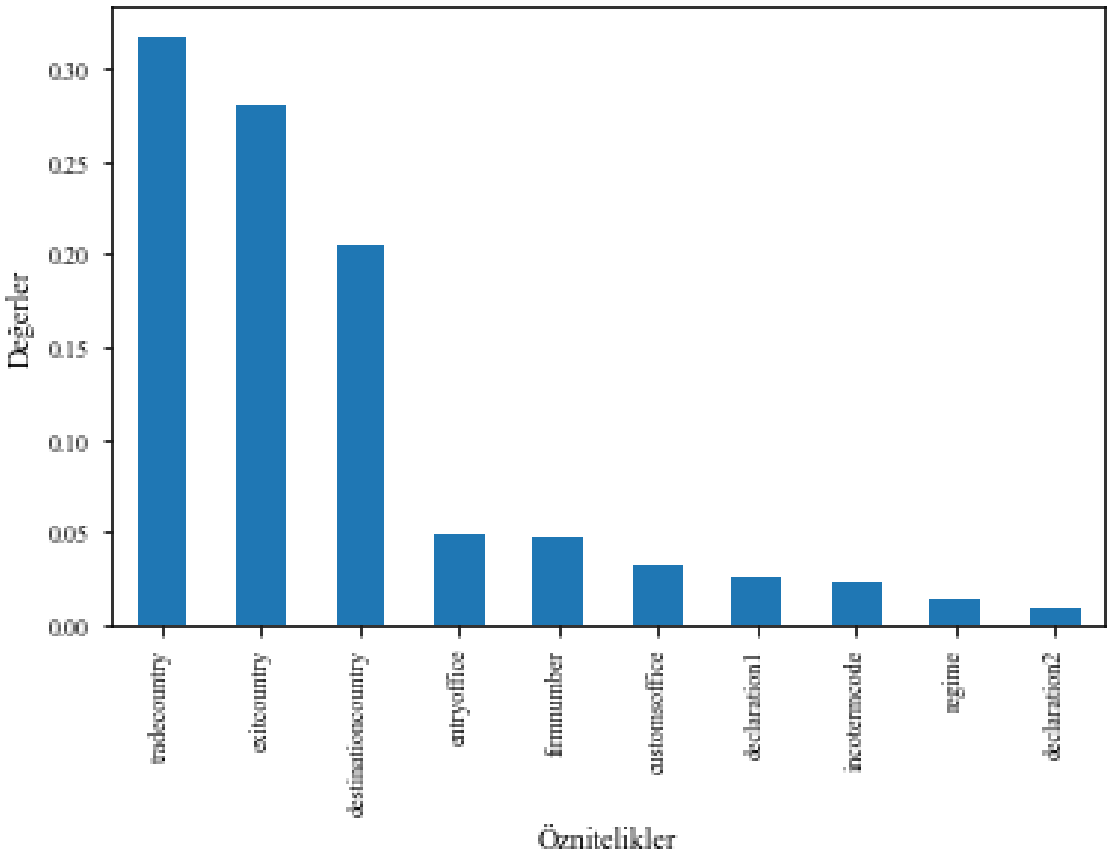
Şekil 38. Taşıma Şekli Modelinde Özniteliklerin Ağırlığı

Şekil 39'da döviz türü modelinde ticaret yapılan ülke etkisi en çok olan öznitelik olarak görünmektedir. Ticaret yapılan ülke özniteliğinden sonra firma numarası etkisi en çok olan ikinci öznitelik olarak görünmektedir.



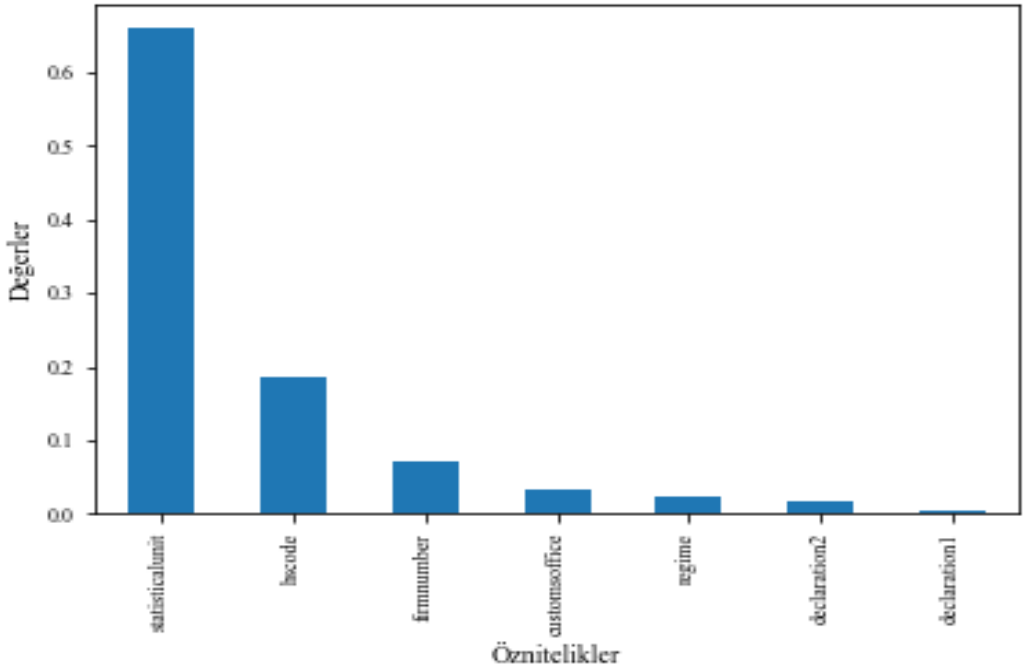
Şekil 39. Döviz Türü Modelinde Özniteliklerin Ağırlığı

Şekil 40’da sevk ülke modelinde ticaret yapılan ülke etkisi en çok olan öznitelik olarak görünmektedir. Ticaret yapılan ülke özniteliğinden sonra çıkış ülkesi etkisi en çok olan ikinci öznitelik olarak görünmektedir.



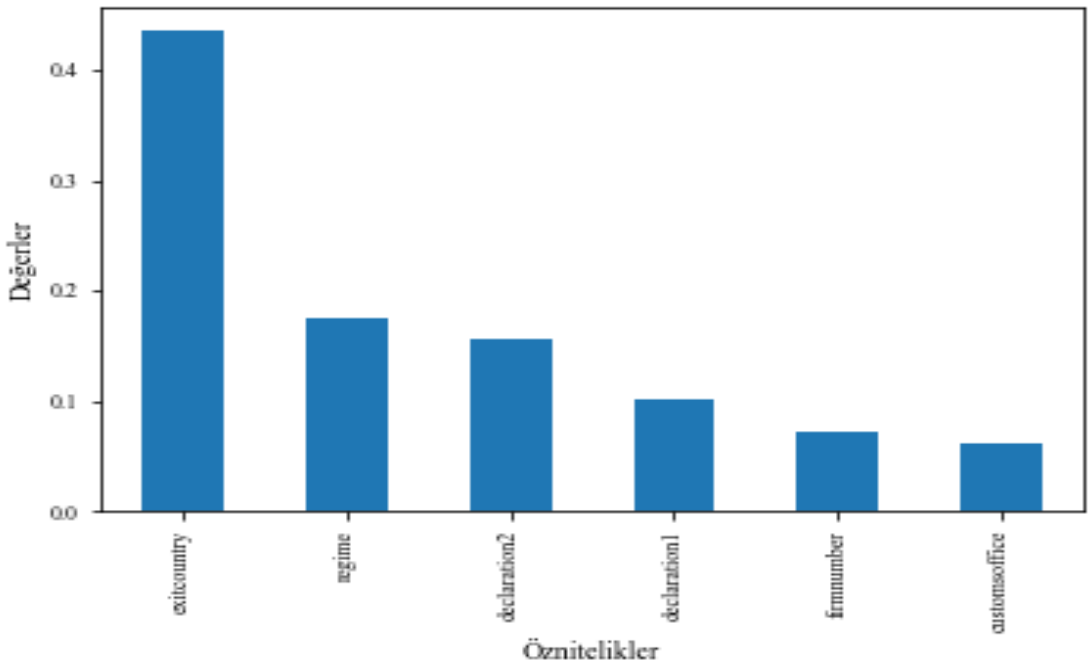
Şekil 40. Sevk Ülke Modelinde Özniteliklerin Ağırlığı

Şekil 41’de birim modelinde istatistiki birim etkisi en çok olan öznitelik olarak görünmektedir. İstatistiki birim özniteliğinden sonra tarife etkisi en çok olan ikinci öznitelik olarak görünmektedir.



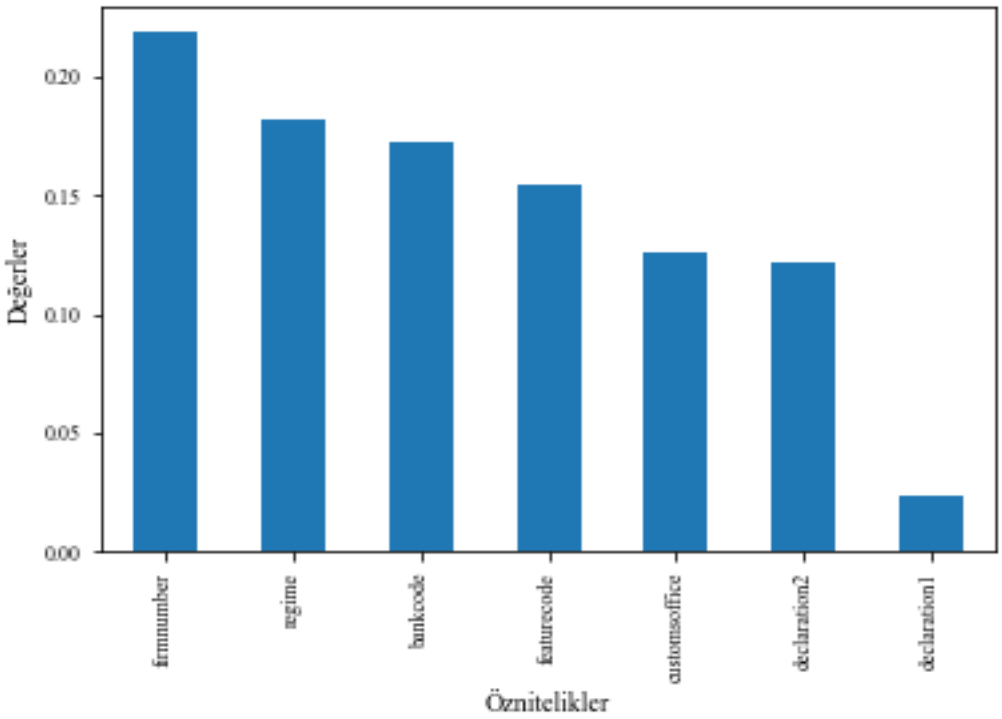
Şekil 41. Birim Modelinde Özniteliklerin Ağırlığı

Şekil 42'de uluslararası anlaşma modelinde çıkış yapılan ülke etkisi en çok olan öznitelik olarak görünmektedir. Çıkış yapılan ülke özniteliğinden sonra rejim etkisi en çok olan ikinci öznitelik olarak görünmektedir.



Şekil 42. Uluslararası Anlaşma Modelinde Özniteliklerin Ağırlığı

Şekil 43'te ödeme şekli modelinde firma numarası etkisi en çok olan öznitelik olarak görünmektedir. Firma numarası özniteliğinden sonra rejim etkisi en çok olan ikinci öznitelik olarak görünmektedir.



Şekil 43. Ödeme Şekli Modelinde Özniteliklerin Ağırlığı

Akıllı Beyanname Uygulamasının Geliştirilmesi

Akıllı beyanname uygulaması ile gümrük beyanname oluşturma aşamasında yapılan hataların minimum seviyeye çekilmesi hedeflenmektedir. Gümrük beyannamesindeki veri seti belirlenebilen bileşenler üzerinde makine öğrenmesi ve derin öğrenme algoritmalarıyla modelleme yapılarak çalışılmıştır. Başarım oranı yüksek olan veri setleri uygulamaya dâhil edilmiştir. Uygulamaya dâhil edilen modeller için kullanıcıya önerilerde bulunulacaktır. Kullanıcının son kontrol ve onayı alınarak beyanname işlemi sonuçlandırılacaktır.

Çalışma uluslararası bir lojistik firmasının 12 yıllık beyanname verileri ile yapılmıştır. Uygulama Delphi 2007 geliştirme ortamı ile yazılan gümrük programı üzerinde çalışacaktır. Sektör için ilk ve tek örnek olacak bu çalışma ile gümrük işlemlerinde önemli bir soruna çözüm sağlanacaktır.

Uygulamada mevcut gümrük beyanname verilerinin de bulunduğu Microsoft SQL Server platformundaki veri tabanı kullanılmıştır. 2010 ve 2021 yılları arasındaki beyanname verilerinde veri setleri T-SQL sorgulama dili kullanılarak oluşturulmuştur. Veri setlerinin eğitime hazır hale getirilmesi bu aşamada yapılmıştır. Öznitelik verileri boş olan ve uygun veri içermeyen kayıtlar bu aşamada elenmiştir.

Makine öğrenmesi algoritmalarının veri setleri üzerinde çalıştırılması, verinin sayısallaştırılması, veri setlerinin görselleştirilmesi, özniteliklerin ağırlıklarının bulunması, eğitim sonuçlarının binary olarak saklanması ve saklanan binary dosyalar kullanılarak tahminleme yapılması işlemleri için Python programlama dilinde kodlar yazılmıştır. Editör olarak Spyder seçilmiştir. Python 3.9 versiyonu kullanılmıştır.

Makine öğrenmesi algoritmalarının veri setleri üzerinde çalıştırılması için Sklearn kütüphanesi kullanılmıştır. Sayısal olmayan veriler eğitim öncesinde LabelEncoder kütüphanesi kullanılarak sayısallaştırılmıştır. Veri setlerinin görselleştirilmesi için büyük veri setlerinde başarılı sonuçlar veren Seaborn kütüphanesi kullanılmıştır. Özniteliklerin ağırlığının görselleştirilmesi için MatPlotLib kütüphanesi kullanılmıştır. Eğitim sonuçlarının binary olarak saklanması ve binary dosyadan okunması için Pickle kütüphanesi kullanılmıştır. Parametre kullanımı için Sys kütüphanesi kullanılmıştır. Veri setlerinin veri tabanından okunması için Pypyodbc ve listelere alınarak kullanılması için Pandas kütüphaneleri kullanılmıştır.

Uygulama için taşıma şekli, döviz türü, sevk ülke, birim, uluslararası anlaşma ve ödeme şekli için öznitelikler tespit edilerek birer veri seti oluşturulmuştur. Her bir veri seti için SQL Server'da tablo açılmıştır. Şekil 44'te veri setlerinden taşıma şekli için oluşturulan tablonun yapısı görülmektedir.

```
CREATE TABLE [dbo].[MLGumrukDosyaTasimaSekli](
    [id] [int] IDENTITY(1,1) NOT NULL,
    [year] [int] NOT NULL,
    [department] [varchar](3) NOT NULL,
    [filenumber] [int] NOT NULL,
    [firmnumber] [int] NOT NULL,
    [customsoffice] [varchar](6) NOT NULL,
    [regime] [varchar](5) NOT NULL,
    [declaration1] [varchar](3) NOT NULL,
    [declaration2] [varchar](3) NOT NULL,
    [entryoffice] [varchar](6) NOT NULL,
    [internalvehiclemode] [varchar](6) NOT NULL,
    [bordervehiclemode] [varchar](6) NOT NULL,
    [bordervehiclecode] [varchar](2) NOT NULL,
 CONSTRAINT [PK_MLGumrukDosyaTasimaSekli] PRIMARY KEY CLUSTERED
(
    [id] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
```

Şekil 44. Taşıma Şekli Tablo Yapısı

Veri setleri için oluşturulan tablolar 2010 ve 2021 yıları arası gümrük beyanname bilgileri ile doldurulmuştur. Bu işlem yapılırken öznitelik verilerinin dolu olup olmadığı kontrol edilerek uygun olmayan kayıtlar bu aşamada elenmiştir. Bu işlem ile veri setleri makine öğrenme algoritmalarının uygulanmasına hazır hale getirilmiştir. Şekil 45'te taşıma şekli veri seti için açılmış tabloya kayıtları ekleyen SQL kodunun bir kısmı görünmektedir.

```
insert into MLGumrukDosyaTasimaSekli (year, department, filenumber, firmnumber, customsoffice, regime, declaration1,
                                      declaration2, entryoffice, internalvehiclemode, bordervehiclemode, bordervehiclecode)
select d.yil, d.departman, d.dosyano, d.firmano, d.gumruk, d.rejim, d.beyan1, d.beyan2,
d.girikgumruk, d.ciktasitcinsi, d.sinirtasitcinsi, d.sinirnak
from Barsan2010..ggumrukdosya d
where isnull(d.iptal,0)=0 and d.tesciltar is not null and left(d.rejim,1) in ('1','2','3','4','5','6','7')
and isnull(d.rejim,'')<>'' and isnull(d.ciktasitcinsi,'')<>'' and isnull(d.sinirtasitcinsi,'')<>'' and isnull(d.sinirnak,'')<>''
and isnull(d.beyan1,'')<>'' and isnull(d.beyan2,'')<>'' and isnull(d.girikgumruk,'')<>'' and isnull(d.gumruk,'')<>''
and exists (select null from gnlparm1 p where p.modul='BEYAN' and p.grup='SINIRNAK' and p.kod=d.sinirnak)
union all
select d.yil, d.departman, d.dosyano, d.firmano, d.gumruk, d.rejim, d.beyan1, d.beyan2,
d.girikgumruk, d.ciktasitcinsi, d.sinirtasitcinsi, d.sinirnak
from Barsan2011..ggumrukdosya d
where isnull(d.iptal,0)=0 and d.tesciltar is not null and left(d.rejim,1) in ('1','2','3','4','5','6','7')
and isnull(d.rejim,'')<>'' and isnull(d.ciktasitcinsi,'')<>'' and isnull(d.sinirtasitcinsi,'')<>'' and isnull(d.sinirnak,'')<>''
and isnull(d.beyan1,'')<>'' and isnull(d.beyan2,'')<>'' and isnull(d.girikgumruk,'')<>'' and isnull(d.gumruk,'')<>''
and exists (select null from gnlparm1 p where p.modul='BEYAN' and p.grup='SINIRNAK' and p.kod=d.sinirnak)
```

Şekil 45. Taşıma Şekli Tablosuna Veri Ekleme

Veri setlerine makine öğrenme algoritması uygulanmadan önce sayısal olmayan verilerin sayısallaştırılması işlemi yapılmıştır. Bu işlem için LabelEncoder kütüphanesi kullanılmıştır. Şekil 46'da döviz türü veri setinde sayısal olmayan verilerin sayısallaştırılması işlemini yapan kod bölümü gösterilmiştir.

```
le = LabelEncoder()
dataframe.customsoffice=le.fit_transform(dataframe.customsoffice)
dataframe.regime=le.fit_transform(dataframe.regime)
dataframe.declaration1=le.fit_transform(dataframe.declaration1)
dataframe.declaration2=le.fit_transform(dataframe.declaration2)
dataframe.tradecountry=le.fit_transform(dataframe.tradecountry)
```

Şekil 46. Döviz Türü Veri Setinde Sayısallaştırma

Hazırlanan veri setlerine makine öğrenme algoritmalarının uygulanması aşamasına geçilerek 8 tane makine öğrenme algoritması ile çalışma yapılmıştır. Bunlar Gaussian Naïve Bayes, Decision Tree sınıflandırma algoritması, K-Nearest Neighbors sınıflandırma algoritması, Random Forest sınıflandırma algoritması, Ridge sınıflandırma

algoritması, Bagging sınıflandırma algoritması, AdaBoost sınıflandırma algoritması ve Multi-Layer Perceptron makine öğrenme algoritmalarıdır. Makine öğrenme algoritmaları yıllar için ayrı ayrı ve bütün yılları içeren veriye yapılarak sonuçları değerlendirilmiştir. Bu aşamada sonuçların iyileştirilmesi yapılarak kabul edilebilir başarım oranlarına ulaşabilmek için işlem en baştan yeni öznitelikler ve yeni tablo yapıları ile tekrarlanmıştır. Başarım oranı kabul edilebilir seviyede olan veri setleri ile devam edilmiştir. Başarım oranında kabul edilebilir seviyede olmayan veri setleri uygulama dışında tutulmuştur.

Yıllar için ayrı ayrı bulunan sonuçlar mevzuat değişikliklerinin eğitim sonuçlarına ne oranda yansıdığını tespit etmek için yapılmıştır. Bütün yıllar için bulunan sonuçlar ise uygulamada hangi makine öğrenme algoritmasının kullanılacağının tespiti için yapılmıştır. Şekil 47'de makine öğrenmesi eğitim çalışmalarının yapıldığı Python kodunun bir bölümü görünmektedir.

```
models = []
models.append(('NB', GaussianNB()))
models.append(('CART', DecisionTreeClassifier()))
models.append(('KNN', KNeighborsClassifier()))
models.append(('RFC', RandomForestClassifier()))
models.append(('RC', RidgeClassifier()))
models.append(('BC', BaggingClassifier()))
models.append(('ABC', AdaBoostClassifier()))
models.append(('MLPC', MLPClassifier()))
test_size = 0.25
seed = 7
results = []
names = []
for name, model in models:
    X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=test_size, random_state=seed)
    model.fit(X_train, Y_train)
    result = model.score(X_test, Y_test)

    results.append(result)
    names.append(name)
    msg = "%s: %f" % (name, result)
    print(msg)
```

Şekil 47. Makine Öğrenmesi Python Eğitim Kodu

Çalışmada veri setleri ile görselleştirme çalışmaları da yapılmıştır. Şekil 48'de taşıma şekli modeli için çıktı sayılarını gösteren görselleştirme Python kodunun bir kısmı görünmektedir.

```
dataframe = pd.DataFrame(c.fetchall(), columns = ['firmnumber', 'customsoffice', 'regime',
                                                  'declaration1', 'declaration2', 'entryoffice',
                                                  'internalvehiclemode', 'bordervehiclemode', 'bordervehiclecode'])
sns.set_style({'font.family':'Times New Roman'})
g=sns.catplot(x="bordervehiclecode", kind="count", data=dataframe, height=5, aspect=2)
g.set(title = "Taşıma Şekli Sayıları")
g.set_axis_labels("Taşıma Şekli","Sayı")
```

Şekil 48. Taşıma Şekli Modelinde Görselleştirme Python Kodu

Veri setlerinin eğitim sonuçlarında özniteliklerin ağırlığı birbirinden farklıdır. Özniteliklerin ağırlığının tespiti için çalışma yapılarak görsel olarak sonuçlar elde edilmiştir. Şekil 49'da taşıma şekli modelinde özniteliklerin ağırlığını görsel olarak veren Python kodunun bir kısmı görünmektedir.

```
test_size = 0.25
seed = 7
results = []
names = []
for name, model in models:
    X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=test_size, random_state=seed)
    model.fit(X_train, Y_train)
    result = model.score(X_test, Y_test)

    results.append(result)
    names.append(name)

    feature_names = ['firmnumber', 'customsoffice', 'regime', 'declaration1', 'declaration2',
                     'entryoffice', 'internalvehiclemode', 'bordervehiclemode']
    df_feature_all = pd.DataFrame([tree.feature_importances_ for tree in model.estimators_])
    df_feature_long = pd.melt(df_feature_all,var_name='feature name', value_name='values')
    df_feature_importance = pd.DataFrame(model.feature_importances_,
                                         index=feature_names,
                                         columns=['feature importance']).sort_values('feature importance',
                                                                                     ascending=False)

    plt.rcParams.update({'font.family':'Times New Roman'})
    pd.Series(model.feature_importances_,
              index=feature_names).nlargest(8).plot(kind='bar', fontsize=8,
                                                    ylabel="Değerler",
                                                    xlabel="Öznitelikler",
                                                    title="Taşıma Şekli Modelinde Özniteliklerin Ağırlığı")
```

Şekil 49. Taşıma Şekli Modelinde Özniteliklerin Ağırlığın Python Kodu

Çalışılan ve başarım oranı istenilen seviyede olan veri setlerinin eğitiminin saklanması işlemi için Pickle kütüphanesi kullanılmıştır ve eğitim sonuçları binary dosyalarda saklanmıştır. Bütün veri setlerinde Bagging sınıflandırma makine öğrenme algoritması

eğitim sonuçları tahminleme yapmak için saklanmıştır. Bunun sebebi bu algoritmanın başarım oranının yapılan çalışmada hep en iyi sonuç olarak çıkmasıdır.

Eğitim sonuçlarının saklanması sürecinde yine sayısal olmayan verilerin sayısallaştırılması gerekmektedir. Sayısal veya sayısallaştırılmış öznitelik verileri ile çıktı bilgisi saklanan bu binary dosyadan okunarak tahminleme yapılırken sayısal olmayan özniteliğin bir verisinin hangi sayısal değeri aldığını bilmek gerektiği için LabelEncoder kütüphanesi kullanılmamıştır. Sayısallaştırma işlemi için sayısal olmayan her öznitelik ayrı ayrı çalışılarak en çok geçen bilgi 1'den başlayacak şekilde artan numara verilerek veri tabanına kaydedilmiştir. Eğitim sonucu ise sayısal olmayan bu özniteliğin karşılığı olan ve veri tabanında kayıtlı sayısal değer ile saklanmıştır. Bu işlem ile tahminleme yaparken ilgili sayısal olmayan öznitelik için hangi sayısal değeri parametre olarak geçeceğimiz belirlenmiştir.

Tablo 28'de taşıma şekli veri setine göre declaration1 özniteliğinin sayısallaştırma değerleri gösterilmektedir.

Tablo 28. Declaration1 Özniteliğinin Sayısallaştırma Tablosu

Declaration1	Sayısal Değer	Kayıt Sayısı
ANT	3	607419
BR	5	6
EU	1	2403872
EX	2	609174
IM	4	443660

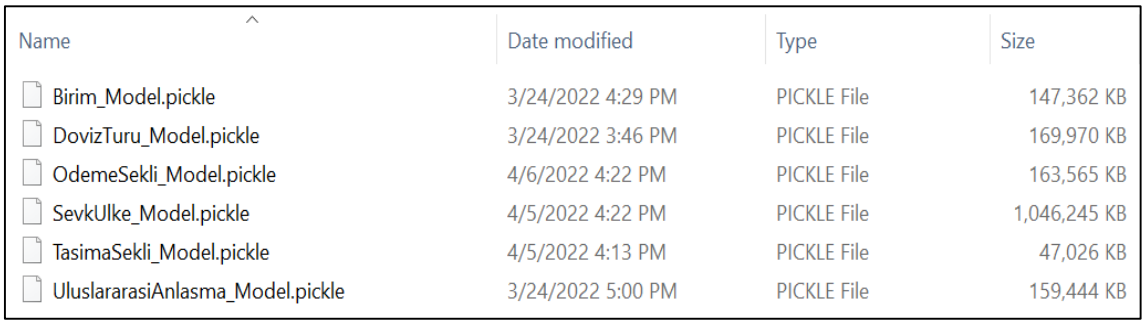
Şekil 50'de taşıma şekli veri setinin eğitim sonuçlarının binary dosyada saklanma işlemi için yazılan Python kodunun bir kısmı görünmektedir

```
dataframe = pd.DataFrame(c.fetchall(), columns = ['firmnumber', 'customsoffice', 'regime',
                                                  'declaration1', 'declaration2', 'entrysoffice',
                                                  'internalvehiclemode', 'bordervehiclemode', 'bordervehiclecode'])
array=dataframe.values
X = array[:,0:8]
Y = array[:,8]

model=BaggingClassifier()
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.25, random_state=7)
model.fit(X_train, Y_train)
filename='TasimaSekli_Model.pickle'
dump(model, open(filename, 'wb'))
```

Şekil 50. Taşıma Şekli Veri Setinin Eğitim Sonuçlarının Saklanması

Şekil 51'de binary dosyada saklanmış eğitim sonuçları listesi ve boyutları gösterilmektedir.



Name	Date modified	Type	Size
Birim_Model.pickle	3/24/2022 4:29 PM	PICKLE File	147,362 KB
DovizTuru_Model.pickle	3/24/2022 3:46 PM	PICKLE File	169,970 KB
OdemeSekli_Model.pickle	4/6/2022 4:22 PM	PICKLE File	163,565 KB
SevkUlke_Model.pickle	4/5/2022 4:22 PM	PICKLE File	1,046,245 KB
TasimaSekli_Model.pickle	4/5/2022 4:13 PM	PICKLE File	47,026 KB
UluslararasiAnlasma_Model.pickle	3/24/2022 5:00 PM	PICKLE File	159,444 KB

Şekil 51. Eğitim Sonuçları Binary Dosyaları

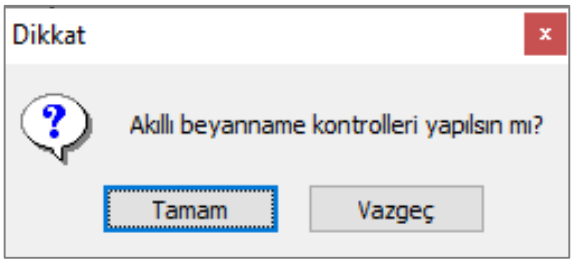
Şekil 52'de taşıma şekli modeli için tahminleme yapan Python kod bölümü gösterilmektedir. Gümrük uygulamasından işlem yapılan kayıt id bilgisi, o işlem ile yapılan kaçıncı işlem olduğuna ait versiyon bilgisi ve hangi kullanıcının yaptığı bilgisi uygulamaya parametre ile geçirilmektedir. Sistemdeki veriler ile tahminleme yapılmaktadır. Mevcut veri girişi ile karşılaştırılarak fark var ise kullanıcıya öneri olarak sunacak şekilde kaydedilmektedir. Tahminleme işlemi binary olarak kaydedilen eğitim dosyası okunarak yapılmaktadır.

```
tasimasekli_model=BaggingClassifier()
filename='TasimaSekli_Model.pickle'
tasimasekli_model = load(open(filename, 'rb'));
predictedbordervehiclecode = tasimasekli_model.predict([[dfTT.get('firmnumber')[0],
                                                         dfTT.get('customsoffice')[0],
                                                         dfTT.get('regime')[0],
                                                         dfTT.get('declaration1')[0],
                                                         dfTT.get('declaration2')[0],
                                                         dfTT.get('entryoffice')[0],
                                                         dfTT.get('internalvehiclemode')[0],
                                                         dfTT.get('bordervehiclemode')[0]]])[0]
if predictedbordervehiclecode != dfTT.get('bordervehiclecode')[0] :
    insertsql = '''INSERT INTO MLIslemSonuc (islemid, islemversiyon, islem, sirket, yil, departman,
                                             dosyano, girilendeger, tahmindeger, instime, insuser)
                   VALUES (?, ?, 5, ?, ?, ?, ?, ?, ?, GETDATE(), ?)
                '''
    c.execute(insertsql, [islemid, islemver, int(dfI.get('sirket')[0]),
                          int(dfI.get('yil')[0]), dfI.get('departman')[0],
                          int(dfI.get('dosyano')[0]), dfTT.get('bordervehiclecode')[0],
                          predictedbordervehiclecode, user])
    c.commit()
```

Şekil 52. Taşıma Şekli Modelinde Tahminleme Python Kodu

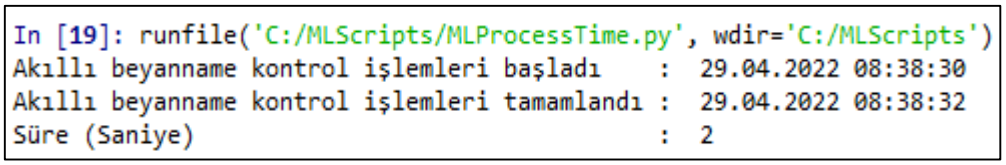
Gümrük beyanname bilgilerinin girişi yapıldıktan sonra öncelikle hesaplama işlemi yapılır. Hesaplama işlemi ile girilen bilgilere göre beyannamenin vergisi hesaplanır. Hesaplaması yapılan gümrük beyannamesi gümrük sistemlerinin web servisi kullanılarak kontrole gönderilir. Gümrükten kontrol sonucu onay alan beyanname işlemi tescil ile sonlandırılacak aşamaya gelir. Akıllı beyanname işlemi bu aşamada devreye girecek şekilde çalışacaktır. Tescil işlemi yine gümrüğün web servisi kullanılarak yapılmaktadır ve işlem olarak geri dönüşü yoktur. Önceki her aşamadaki hata giderilebilir fakat tescil alınması durumunda cezai yaptırımlarla karşı karşıya kalınmaktadır. Bu kritik işlem öncesi modellediğimiz veri setleri ile hazırladığımız Akıllı Beyanname Kontrol Uygulaması ile mevcut çıktı verileri, uygulamanın öznitelikleri kullanarak yaptığı tahminleme sonucu bulduğu değer ile karşılaştırılarak farklı oldukları durumda sonucu kullanıcıya öneri olarak sunacaktır. Yapılan öneri kullanıcıya son bir kontrol fırsatı tanıyarak olası hatanın önüne geçme fırsatı sunacaktır. Yüksek başarım oranlarına rağmen kullanıcı olağan akışın dışında bir işlem yapıyorsa geçerli sebebi ile mevcut bilgi girişiyle işlemine devam edebilecektir.

Tescil işlemi başlatıldığında Şekil 53'te görüldüğü gibi kullanıcıdan Akıllı Beyanname Kontrol Uygulaması kontrollerinin yapılıp yapılmaması için onay alınmaktadır. Kullanıcı bu işlemi isterse atlayabilir.



Şekil 53. Akıllı Beyanname Kontrol Uygulaması Başlatılması

Kullanıcının onay vermesi durumunda tahminleme yapan ve Python programlama dili ile yazılan program çalışmakta ve Şekil 54'te görülüğü gibi çok kısa sürede işlemi sonuçlandırmaktadır. Bu süreye bütün veri tabanı işlemleri ve tahminleme yapılırken eğitim sonuç dosyalarını kullanma dâhildir.



Şekil 54. Tahminleme Yapan Python Programının Çalışma Süresi

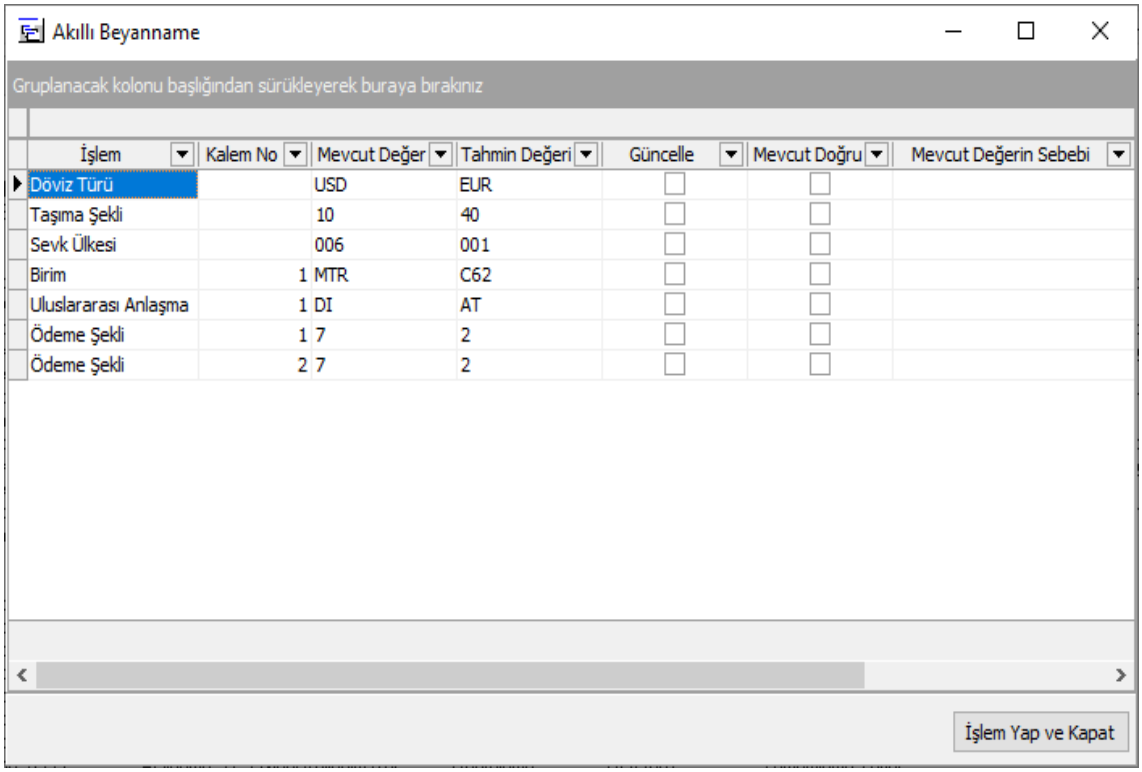
Program 3 parametre alarak çalışmaktadır. İlk parametre kontrol işlemi numarasıdır. Bu numara ile bütün beyanname verilerine ulaşılabilmektedir. İkinci parametre versiyon numarasıdır. Kontrol işlemi ile birden fazla tescile gönderme yapılabileceği için her denemede 1'den başlayarak artan numara şeklinde bilgi içerir. Üçüncü parametre ise kullanıcı bilgisidir ve işlemi yapan kullanıcı bilgisi alanlarını doldurmak içindir.

Şekil 55'te Akıllı Beyanname Kontrol Uygulaması çalıştıktan sonra çalışılan modeller için yapılan tahminleme değerleri ile mevcut değerler arasında fark olanların listesi gösterilmektedir.

Şekil 55'teki işlem bilgisi hangi alanın tahminlemesinin yapıldığını göstermektedir. Kalem No bilgisi ise beyanname kalemlerindeki bir alanın tahminlemesinde doldurulur. Beyanname ana ekran bilgilerinin tahminlemesinde Kalem No bilgisi olmaz. Beyanname kalemlerinde ise tahminleme yaptığımız alan için fark olan kalemler listeye getirilir. Bu örnekte iki kalemli bir beyanname kullanıldı ve Birim modeli için sadece birinci kalemde tahminleme verisi ile mevcut veri farklı çıkmasına rağmen Ödeme Şekli modelinde iki kalemde de fark çıktı. Mevcut Değer bilgisi beyannamede kullanıcının girdiği veridir. Tahmin Değeri ise modellerimizin öznitelikleri kullanılarak yapılan tahminleme değeridir.

Kullanıcı Akıllı Beyanname Kontrol Uygulamasını kullandığında Şekil 55'teki gibi bir liste ile karşılaştığında mevcut verileri uygulamanın önerdikleri ile karşılaştırarak son bir kontrol ile eğer varsa hataları düzeltme fırsatı bulacaktır. Burada hatalı olduğunu düşündüğü kayıtlar için Güncelle seçeneğini seçmelidir. Hatalı olmadığını düşünüyorsa Mevcut Doğru seçeneğini seçmeli ve Mevcut Değerin Sebebi bilgisini girmelidir. Mevcut Doğru seçeneği seçilmiş kayıtlar için Mevcut Değerin Sebebi bilgisi zorunludur. Güncelle ve Mevcut Doğru listedeki bir kayıt için aynı anda seçilememektedir. İşlem Yap ve Kapat ile işlemine devam etmeden önce listedeki her bir kayıt için Güncelle veya

Mevcut Doğru seçeneği seçilmiş olması kontrolü eklenmiştir. İşlem Yap ve Kapat seçeneği çalıştığında Güncelle seçilmiş bir kayıt yok ise işleme devam edilmektedir. Güncelle seçilmiş kayıtlar var ise ilgili bilgi beyannamede otomatik olarak güncellenmekte ve süreci hesaplama yapma öncesine çekmektedir. Normal işlemlerde de beyannamede yapılan bir değişiklik süreci hesaplama aşamasına çekmektedir. Kullanıcı yine hesaplama ve kontrol aşamalarından sonra tescil aşamasına gelmelidir.



Şekil 55. Akıllı Beyanname Kontrol Uygulaması Sonuçları

Şekil 56'da Akıllı Beyanname Kontrol Uygulaması çalıştığında tutulan log kayıtları gösterilmektedir.

```
select *
from MLIslemSonuc
where islemid=513735 and islemversiyon=1
```

Results | Messages

id	islemid	islemversiyon	islem	sirket	yil	departman	dosyano	kalemno	girilendeger	tahmindeger	guncelle	guncellendi	mevcutdogru	mevcutsebep	instime	insuser
1	513735	1	2	1	2022	AND	100	NULL	USD	EUR	NULL	NULL	NULL	NULL	2022-04-18 17:17:40.867	3676
2	513735	1	5	1	2022	AND	100	NULL	10	40	NULL	NULL	NULL	NULL	2022-04-18 17:17:40.947	3676
3	513735	1	6	1	2022	AND	100	NULL	006	001	NULL	NULL	NULL	NULL	2022-04-18 17:17:42.307	3676
4	513735	1	3	1	2022	AND	100	1	MTR	C62	NULL	NULL	NULL	NULL	2022-04-18 17:17:42.550	3676
5	513735	1	4	1	2022	AND	100	1	DI	AT	NULL	NULL	NULL	NULL	2022-04-18 17:17:42.817	3676
6	513735	1	7	1	2022	AND	100	1	7	2	NULL	NULL	NULL	NULL	2022-04-18 17:17:43.070	3676
7	513735	1	7	1	2022	AND	100	2	7	2	NULL	NULL	NULL	NULL	2022-04-18 17:17:43.070	3676

Şekil 56. Akıllı Beyanname Kontrol Uygulaması Log Kayıtları

Akıllı Beyanname Kontrol Uygulaması ile olası hataların tespit edilmesi hedeflenmektedir. Olağan olmayan veri girişlerinin tespiti ve kullanıcıya sunulması ile bu konuda olumlu sonuçlar beklenmektedir.

Yapılan uygulamanın altyapısı ile log kayıtları incelenerek hataların azalmasına ne derece katkı yapıldığı tespit edilebilecektir. Ayrıca hatalı beyannamelerde log kayıtları incelenerek bu konuda uygulama önerilerinin dikkate alınıp alınmadığı ve daha hangi bilgiler için öneri yapmak gerektiği tespit edilebilecektir. Bu tespitler sonucu yeni modeller eklenerek uygulama daha ileri bir seviyeye çıkarılacaktır.

BÖLÜM 3. BULGULAR VE YORUMLAR

Bulgular

Scikit-Learn kütüphanesinde modellerin eğitimi için standart test verisi oranı %25 olarak alınır (Paper, 2020). Modellerin eğitiminde test verisi oranı %25 alınarak eğitimler gerçekleştirilmiştir. Birçok algoritma denenmiş ve özellikle regresyon algoritmalarındaki (Huang, Ko, Shu, ve Hsu, 2019) olumsuz sonuçlardan dolayı bu algoritmaların kullanılmamasına karar verilmiştir. Sınıflandırma algoritmalarında (Kotsiantis, Zaharakis, ve Pintelas, 2007) yüksek başarımlar elde edilmiştir.

Gümrük operasyonlarında mevzuatın sürekli değişmesi nedeniyle 2010 ile günümüz arasında ciddi bir işleyiş farkı olacağı göz önüne alınarak yıllar bazında da bu eğitim çalışması yapılmıştır. Yıl bazında daha yüksek başarımlar elde edilmiştir.

Tablo 29'da gösterildiği gibi taşıma şekli modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %86.41, Decision Tree sınıflandırma algoritmasında %98.76, K-Nearest Neighbors sınıflandırma algoritmasında %98.36, Random Forest sınıflandırma algoritmasında %98.74, Ridge sınıflandırma algoritmasında %79.82,

Bagging sınıflandırma algoritmasında %98.76, AdaBoost sınıflandırma algoritmasında %42.29 ve Multi-Layer Perceptron algoritmasında %94.09 başarım elde edilmiştir. Sınıflandırma algoritmalarında Decision Tree, K-Nearest Neighbors, Random Forest ve Bagging diğerlerine göre daha başarılı sonuç vermiştir. En başarılı orana ise %98.76 ile Decision Tree sınıflandırma ve Bagging sınıflandırma algoritmaları sahiptir.

Tablo 29. Taşıma Şekli Veri Seti Sonuçları

Algoritma	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Tüm Yıllar
Naïve Bayes - NB	93.09	91.14	89.93	88.82	89.63	85.83	88.43	85.97	80.19	83.99	83.68	84.09	86.41
Decision Tree Classifier - CART	97.95	98.35	99.21	99.28	99.23	99.30	99.32	99.42	99.14	98.96	99.06	99.30	98.76
K-Nearest Neighbors - KNN	97.32	97.80	98.50	98.37	98.39	98.55	98.44	98.67	98.28	98.03	98.11	98.43	98.36
Random Forest - RFC	97.95	98.35	99.23	99.21	99.15	99.26	99.24	99.37	99.08	98.95	99.04	99.30	98.74
Ridge Classifier - RC	82.91	79.52	80.88	80.07	79.37	78.25	79.82	80.37	79.55	80.94	79.95	78.48	79.82
Bagging Classifier - BC	97.99	98.38	99.24	99.29	99.22	99.30	99.31	99.43	99.13	98.96	99.07	99.32	98.76
Ada Boost Classifier - ABC	38.79	75.10	70.74	83.40	87.58	85.84	82.30	84.37	82.41	83.80	58.47	89.08	42.29
MLP Classifier - MLPC	95.45	94.36	94.21	94.02	93.49	92.44	91.64	91.28	90.57	91.45	76.95	89.70	94.09

Tablo 30'da gösterildiği gibi döviz türü modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %82.01, Decision Tree sınıflandırma algoritmasında %95.96, K-Nearest Neighbors sınıflandırma algoritmasında %95.40, Random Forest sınıflandırma algoritmasında %95.97, Ridge sınıflandırma algoritmasında %81.36, Bagging sınıflandırma algoritmasında %96.00, AdaBoost sınıflandırma algoritmasında %82.95 ve Multi-Layer Perceptron algoritmasında %84.43 başarım elde edilmiştir. Sınıflandırma algoritmalarında Decision Tree, K-Nearest Neighbors, Random Forest ve Bagging diğerlerine göre daha başarılı sonuç vermiştir. En başarılı orana ise %96.00 ile Bagging sınıflandırma algoritması sahiptir.

Tablo 30. Döviz Türü Veri Seti Sonuçları

Algoritma	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Tüm Yıllar
Naïve Bayes - NB	82.26	79.68	83.10	84.91	83.71	83.44	83.47	81.36	80.71	79.01	77.68	77.57	82.01
Decision Tree Classifier - CART	97.08	97.09	96.99	97.07	97.34	96.91	96.88	96.04	96.18	96.33	96.17	96.45	95.96
K-Nearest Neighbors - KNN	96.54	96.34	96.16	96.42	96.60	96.27	96.27	95.26	95.33	95.64	95.38	95.80	95.40
Random Forest - RFC	97.15	97.14	97.03	97.12	97.39	96.99	96.97	96.09	96.24	96.35	96.27	96.47	95.97
Ridge Classifier - RC	83.82	83.61	83.35	84.64	84.04	83.53	83.42	81.34	80.91	79.62	77.97	78.33	81.36
Bagging Classifier - BC	97.09	97.12	97.07	97.14	97.35	96.96	96.96	96.11	96.29	96.40	96.25	96.47	96.00
Ada Boost Classifier - ABC	68.12	85.22	95.74	85.21	82.33	84.87	84.10	71.24	65.95	77.54	78.68	80.15	82.95
MLP Classifier - MLPC	88.89	88.66	88.42	87.36	86.48	87.27	85.83	80.82	83.24	81.63	81.92	79.11	84.43

Tablo 31'de gösterildiği gibi sevk ülke modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %50.03, Decision Tree sınıflandırma algoritmasında %97.89, K-Nearest Neighbors sınıflandırma algoritmasında %94.71, Random Forest sınıflandırma algoritmasında %97.83, Ridge sınıflandırma algoritmasında %31.48, Bagging sınıflandırma algoritmasında %97.92, AdaBoost sınıflandırma algoritmasında %35.30 ve Multi-Layer Perceptron algoritmasında %70.07 başarım elde edilmiştir. Sınıflandırma algoritmalarında Decision Tree, K-Nearest Neighbors, Random Forest ve Bagging diğerlerine göre daha başarılı sonuç vermiştir. En başarılı orana ise %97.92 ile Bagging sınıflandırma algoritması sahiptir.

Tablo 31. Sevk Ülke Veri Seti Sonuçları

Algoritma	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Tüm Yıllar
Naïve Bayes - NB	53.71	54.54	56.04	54.56	55.40	53.71	51.66	52.40	51.23	52.59	50.59	51.09	50.03
Decision Tree Classifier - CART	97.57	97.89	97.87	97.73	98.34	98.39	98.06	98.06	98.13	98.06	98.01	97.95	97.89
K-Nearest Neighbors - KNN	92.36	92.57	92.03	91.39	91.98	92.08	91.80	92.00	91.67	91.78	92.05	92.13	94.71
Random Forest - RFC	97.42	97.89	97.85	97.65	98.22	98.28	97.96	98.03	98.05	97.94	97.98	97.93	97.83
Ridge Classifier - RC	32.78	33.59	36.55	34.11	32.96	33.13	32.13	30.36	30.65	30.10	29.66	30.05	31.48
Bagging Classifier - BC	97.62	97.94	97.94	97.76	98.39	98.43	98.12	98.16	98.19	98.07	98.08	97.98	97.92
Ada Boost Classifier - ABC	37.02	38.30	39.12	37.24	36.90	37.36	35.23	33.66	34.02	33.60	32.05	32.80	35.30
MLP Classifier - MLPC	85.46	79.86	79.34	65.29	66.21	66.74	56.44	53.60	56.99	48.27	52.42	55.48	70.07

Tablo 32'de gösterildiği gibi birim modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %97.85, Decision Tree sınıflandırma algoritmasında %99.69, K-Nearest Neighbors sınıflandırma algoritmasında %99.51, Random Forest sınıflandırma algoritmasında %99.70, Ridge sınıflandırma algoritmasında %91.98, Bagging sınıflandırma algoritmasında %99.70, AdaBoost sınıflandırma algoritmasında %96.34 ve Multi-Layer Perceptron algoritmasında %97.99 başarım elde edilmiştir. Sınıflandırma algoritmalarında Decision Tree, K-Nearest Neighbors, Random Forest ve Bagging diğerlerine göre daha başarılı sonuç vermiştir. En başarılı orana ise %99.70 ile Random Forest sınıflandırma ve Bagging sınıflandırma algoritmaları sahiptir.

Tablo 32. Birim Veri Seti Sonuçları

Algoritma	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Tüm Yıllar
Naïve Bayes - NB	95.45	96.87	97.11	96.24	97.19	97.42	97.85	98.22	98.92	98.97	99.07	99.03	97.85
Decision Tree Classifier - CART	99.66	99.61	99.63	99.44	99.70	99.67	99.64	99.78	99.83	99.88	99.92	99.92	99.69
K-Nearest Neighbors - KNN	99.36	99.33	99.28	99.01	99.23	99.26	99.25	99.48	99.51	99.61	99.67	99.69	99.51
Random Forest - RFC	99.67	99.61	99.64	99.48	99.70	99.67	99.65	99.79	99.84	99.89	99.92	99.93	99.70
Ridge Classifier - RC	91.79	92.02	91.33	90.15	89.50	90.89	91.28	92.48	93.23	93.15	93.30	93.19	91.98
Bagging Classifier - BC	99.67	99.62	99.66	99.47	99.71	99.68	99.65	99.79	99.84	99.89	99.92	99.93	99.70
Ada Boost Classifier - ABC	94.04	95.31	94.81	94.22	95.41	95.60	95.25	96.65	97.54	98.02	98.00	98.05	96.34
MLP Classifier - MLPC	97.46	97.16	96.87	96.85	97.43	97.03	97.85	97.89	98.38	98.58	97.82	98.93	97.99

Tablo 33'te gösterildiği gibi uluslararası anlaşma modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %86.41, Decision Tree sınıflandırma algoritmasında %96.44, K-Nearest Neighbors sınıflandırma algoritmasında %96.08, Random Forest sınıflandırma algoritmasında %96.44, Ridge sınıflandırma algoritmasında %89.88, Bagging sınıflandırma algoritmasında %96.44, AdaBoost sınıflandırma algoritmasında %81.97 ve Multi-Layer Perceptron algoritmasında %94.29 başarım elde edilmiştir. Sınıflandırma algoritmalarında Decision Tree, K-Nearest Neighbors, Random Forest ve Bagging diğerlerine göre daha başarılı sonuç vermiştir. En başarılı orana ise %96.44 ile Decision Tree sınıflandırma, Random Forest sınıflandırma ve Bagging sınıflandırma algoritmaları sahiptir.

Tablo 33. Uluslararası Anlaşma Veri Seti Sonuçları

Algoritma	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Tüm Yıllar
Naïve Bayes - NB	89.64	88.81	86.14	83.12	84.61	86.18	83.84	82.49	85.77	86.26	88.67	85.81	86.41
Decision Tree Classifier - CART	97.43	97.83	97.96	97.91	98.01	97.77	97.96	97.69	98.14	96.47	98.08	97.02	96.44
K-Nearest Neighbors - KNN	97.19	97.61	97.81	97.67	97.83	97.39	97.80	97.50	98.00	96.27	97.53	96.61	96.08
Random Forest - RFC	97.43	97.83	97.98	97.92	98.02	97.78	97.97	97.69	98.15	96.47	98.08	97.02	96.44
Ridge Classifier - RC	91.78	90.46	90.30	86.95	88.88	91.89	90.48	90.13	91.35	91.04	93.13	89.39	89.88
Bagging Classifier - BC	97.41	97.83	97.97	97.90	98.01	97.76	97.96	97.68	98.14	96.47	98.08	97.01	96.44
Ada Boost Classifier - ABC	33.28	34.72	92.89	91.51	92.04	88.84	28.25	87.02	33.71	87.22	81.87	31.57	81.97
MLP Classifier - MLPC	94.85	95.04	96.01	93.70	93.96	95.21	93.49	94.86	95.44	93.18	96.11	93.85	94.29

Tablo 34'te gösterildiği gibi ödeme şekli modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %63.75, Decision Tree sınıflandırma algoritmasında %94.09, K-Nearest Neighbors sınıflandırma algoritmasında %93.34, Random Forest sınıflandırma algoritmasında %94.09, Ridge sınıflandırma algoritmasında %72.97, Bagging sınıflandırma algoritmasında %94.10, AdaBoost sınıflandırma algoritmasında %79.18 ve Multi-Layer Perceptron algoritmasında %78.62 başarım elde edilmiştir. Sınıflandırma algoritmalarında Decision Tree, K-Nearest Neighbors, Random Forest ve Bagging diğerlerine göre daha başarılı sonuç vermiştir. En başarılı orana ise %94.10 ile Bagging sınıflandırma algoritması sahiptir.

Tablo 34. Ödeme Şekli Veri Seti Sonuçları

Algoritma	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Tüm Yıllar
Naïve Bayes - NB	78.48	66.61	55.47	48.70	69.76	84.20	79.18	74.27	62.62	64.55	67.11	65.98	63.75
Decision Tree Classifier - CART	97.70	97.57	97.11	97.52	97.28	94.98	94.89	94.54	94.50	93.46	93.68	95.01	94.09
K-Nearest Neighbors - KNN	96.86	97.17	96.66	97.07	96.77	94.18	94.33	93.98	93.99	91.04	92.58	93.67	93.34
Random Forest - RFC	97.70	97.58	97.14	97.53	97.31	95.01	94.91	94.57	94.52	93.48	93.70	95.02	94.09
Ridge Classifier - RC	77.94	69.44	69.87	69.87	71.74	87.76	82.05	76.07	67.18	69.96	72.16	67.94	72.97
Bagging Classifier - BC	97.69	97.57	97.12	97.52	97.28	94.98	94.89	94.55	94.51	93.46	93.70	95.00	94.10
Ada Boost Classifier - ABC	77.68	69.96	56.86	57.19	28.82	59.97	81.80	48.80	63.39	69.49	41.44	32.88	79.18
MLP Classifier - MLPC	86.13	92.04	88.88	81.66	86.34	87.79	82.74	77.51	74.51	72.34	74.28	72.45	78.62

Yorumlar

Gümrük sistemlerinde beyanname yazımında makine öğrenme algoritmaları ile veri tahminlemesi yapmak elde edilen sonuçlara göre uygun görünmektedir. Yapılan bu çalışma sektördeki büyük bir açığa çare olacak sonuçlar vermiştir. Elde edilen sonuçlar sektörün sürekli gelişen ve değişen yapısına göre başarılı kabul edilebilir. Algoritma başarım oranları ve sektörün bu konudaki ihtiyacı makine öğrenmesi konusunda daha çok çalışma yapılmasına vesile olacaktır.

Yapılan tahminleme sonrası düzeltilen her veri, yapılan hata oranlarını aşağıya çekmeye yardımcı olacaktır ve olası birçok problemin önüne geçilebilecektir. Bu durum operasyon

maliyetlerine ve süreçlerine olumlu yansıyacaktır. Olası hataların doğurabileceği hukuki sonuçların önüne geçilebilecektir.

Çalışmada 8 tane makine öğrenmesi algoritması kullanılmıştır. Bunlar Gaussian Naïve Bayes, Decision Tree sınıflandırma, K-Nearest Neighbors sınıflandırma, Random Forest sınıflandırma, Ridge sınıflandırma, Bagging sınıflandırma, AdaBoost sınıflandırma ve Multi-Layer Perceptron makine öğrenme algoritmalarıdır. Sınıflandırma algoritmalarında Decision Tree, K-Nearest Neighbors, Random Forest ve Bagging diğerlerine göre daha başarılı sonuç vermiştir.

Taşıma şekli modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %86.41, Decision Tree sınıflandırma algoritmasında %98.76, K-Nearest Neighbors sınıflandırma algoritmasında %98.36, Random Forest sınıflandırma algoritmasında %98.74, Ridge sınıflandırma algoritmasında %79.82, Bagging sınıflandırma algoritmasında %98.76, AdaBoost sınıflandırma algoritmasında %42.29 ve Multi-Layer Perceptron algoritmasında %94.09 başarım elde edilmiştir. En başarılı oran ise %98.76 ile Decision Tree sınıflandırma ve Bagging sınıflandırma algoritmalarında elde edilmiştir.

Döviz türü modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %82.01, Decision Tree sınıflandırma algoritmasında %95.96, K-Nearest Neighbors sınıflandırma algoritmasında %95.40, Random Forest sınıflandırma algoritmasında %95.97, Ridge sınıflandırma algoritmasında %81.36, Bagging sınıflandırma algoritmasında %96.00, AdaBoost sınıflandırma algoritmasında %82.95 ve Multi-Layer Perceptron algoritmasında %84.43 başarım elde edilmiştir. En başarılı oran ise %96.00 ile Bagging sınıflandırma algoritmasında elde edilmiştir.

Sevk ülke modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %50.03, Decision Tree sınıflandırma algoritmasında %97.89, K-Nearest Neighbors sınıflandırma algoritmasında %94.71, Random Forest sınıflandırma algoritmasında %97.83, Ridge sınıflandırma algoritmasında %31.48, Bagging sınıflandırma algoritmasında %97.92, AdaBoost sınıflandırma algoritmasında %35.30 ve Multi-Layer Perceptron algoritmasında %70.07 başarım elde edilmiştir. En başarılı oran ise %97.92 ile Bagging sınıflandırma algoritmasında elde edilmiştir.

Birim modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %97.85, Decision Tree sınıflandırma algoritmasında %99.69, K-Nearest Neighbors sınıflandırma algoritmasında %99.51, Random Forest sınıflandırma algoritmasında %99.70, Ridge sınıflandırma algoritmasında %91.98, Bagging sınıflandırma algoritmasında %99.70, AdaBoost sınıflandırma algoritmasında %96.34 ve Multi-Layer Perceptron algoritmasında %97.99 başarım elde edilmiştir. En başarılı oran ise %99.70 ile Random Forest sınıflandırma ve Bagging sınıflandırma algoritmalarında elde edilmiştir.

Uluslararası anlaşma modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %86.41, Decision Tree sınıflandırma algoritmasında %96.44, K-Nearest Neighbors sınıflandırma algoritmasında %96.08, Random Forest sınıflandırma algoritmasında %96.44, Ridge sınıflandırma algoritmasında %89.88, Bagging sınıflandırma algoritmasında %96.44, AdaBoost sınıflandırma algoritmasında %81.97 ve Multi-Layer Perceptron algoritmasında %94.29 başarım elde edilmiştir. En başarılı oran ise %96.44 ile Decision Tree sınıflandırma, Random Forest sınıflandırma ve Bagging sınıflandırma algoritmalarında elde edilmiştir.

Ödeme şekli modelinde tüm yıllar için yapılan eğitimlerde Gaussian Naïve Bayes algoritmasında %63.75, Decision Tree sınıflandırma algoritmasında %94.09, K-Nearest Neighbors sınıflandırma algoritmasında %93.34, Random Forest sınıflandırma algoritmasında %94.09, Ridge sınıflandırma algoritmasında %72.97, Bagging sınıflandırma algoritmasında %94.10, AdaBoost sınıflandırma algoritmasında %79.18 ve Multi-Layer Perceptron algoritmasında %78.62 başarım elde edilmiştir. En başarılı oran ise %94.10 ile Bagging sınıflandırma algoritmasında elde edilmiştir.

Çalışmada uluslararası bir lojistik firmasının 2010 ve 2021 yılları arasındaki 12 yıllık gümrük beyanname verileri kullanılmıştır. Kullanılan beyanname verileri Türkiye dış ticaretinin %8'ini oluşturmaktadır ve çalışma için yeterli veriye ulaşmaya olanak sağlamıştır. Bu veriler üzerinde yapılan çalışma ile elde edilen sonuçlar sektör için makine öğrenme teknolojisi kullanımı açısından Türkiye'de öncü olacak niteliktedir.

BÖLÜM 4. SONUÇ

Bu çalışma gümrük sisteminde beyanname hazırlanırken yapılabilecek hataların tespiti için makine öğrenme algoritmalarının kullanımının etkinliğini ortaya çıkarmak amacıyla yapılmıştır. Elde edilen algoritma başarım oranları bu konuda çalışma yapmayı teşvik edecek derecede olumludur. Her sektör elindeki büyük veriyi kullanmak zorundadır ve gümrük sektörü de bu sürecin dışında kalmamalıdır.

Özet

Gümrük sisteminde hatasız beyanname yazımı büyük önem taşımaktadır. Yapılan hatalar olumsuz sonuçlara sebep olmaktadır. Yapılacak çalışma ile hataların azaltılması, operasyon verimliliği ve maliyetlerin düşürülmesi hedeflenirken aynı zamanda daha çok ve verimli iş yapılması planlanıyor. Elimizdeki büyük verileri kullanarak yapılacak doğruluk oranı yüksek tahminlemeler sayesinde hatalar zamanında tespit edilip olası olumsuzlukları giderilecektir. Modellerin doğruluk oranları çalışmada yapılacak tahminleme miktarında belirleyici olacaktır. Elde edilen algoritma başarım oranları bu konuda çalışma yapmayı teşvik edecek derecede olumludur.

Yargı

Çalışmada elde edilen sonuçlar makine öğrenmesi ve derin öğrenme algoritmalarının gümrük sistemlerinde de etkin olarak kullanılabileceğini göstermektedir. Çalışma, sektör için makine öğrenme teknolojisi kullanımı açısından Türkiye'de öncü olacak niteliktedir.

Öneriler

Farklı veri setleri ve farklı makine öğrenme algoritmaları üzerinde yeni çalışmalar yapılarak hataların tespitinde daha başarılı sonuçlar elde edilebilir. Daha büyük veri kullanılarak algoritmaların doğrulukları değerlendirilebilir. Özellikle başarılı sonuçların gümrük sisteminde beyanname oluşturma aşamasında kullanılmasının sürece çok ciddi katkılar yapacağı düşünülmektedir.

KAYNAKÇA

(2021, 12 31). 03 21, 2022 tarihinde Türkiye İstatistik Kurumu: www.tuik.gov.tr adresinden alındı

Abdulkareem, N. M., ve 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. 02 03, 2022 tarihinde https://ideas.repec.org/a/aif/journl/v5y2021i2p128-142.html adresinden alındı

Abellán, J., ve Castellano, J. G. (2017). Improving the Naive Bayes Classifier via a Quick Variable Selection Method Using Maximum of Entropy. *Entropy, 19(6)*, 247. doi:10.3390/e19060247

Aka, A., ve Ürünal, A. A. (2018). Türkiye'de Dış Ticaret Uygulamaları: 4458 Sayılı Gümrük Kanunu Özelinde. *Balkan Sosyal Bilimler Dergisi*, 7(13), 154-170. 02 03, 2022 tarihinde https://dergipark.org.tr/en/pub/bsbd/issue/34559/336447 adresinden alındı

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

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., ve Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics, 182*, 113–131. doi:10.1016/j.ijpe.2016.08.018

Alasadi, S. A., ve Bhaya, W. S. (2017). Review of Data Preprocessing Techniques in Data Mining. *Journal of Engineering and Applied Sciences, 12(16)*, 4102-4107. 03 02, 2022 tarihinde https://d1wqtxts1xzle7.cloudfront.net/54509277/4102-4107-with-cover-page-v2.pdf?Expires=1640953672veSignature=clJAc2YdAfYUYGjRA8SlL~IvdlbA

Cm0klxxPI3EUOAkvuCMBK-L4u53COsllZQZM3s6U7R-2qGg3lPhfR3QjivPR1U2aTI0pNeNWCmCMkslwb~zpsDx9~uc7dNc0kyHP-3KKpNktg9o6l-KAqD adresinden alındı

Alexandropoulos, S.-A. N., Kotsiantis, S. B., ve 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., ve 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

Anaconda (Python distribution) - Wikipedia. (2022, 4 18). 4 18, 2022 tarihinde Wikipedia: https://en.wikipedia.org/wiki/Anaconda_(Python_distribution) adresinden alındı

Barchard, K. A., ve 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., ve Zhou, Y. (2020). Machine learning for international freight transportation management: A comprehensive review. *Research in Transportation Business and Management, 100453*. doi:10.1016/j.rtbm.2020.100453

Başer, B. Ö., Yangın, M., ve 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

Belgiu, M., ve Dragut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24–31. doi:10.1016/j.isprsjprs.2016.01.011

Beyer, K., Goldstein, J., Ramakrishnan, R., ve Shaft, U. (1999). When is “nearest neighbor” meaningful? *In International conference on database theory* (s. 217-235). Berlin, Heidelberg: Springer.

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

Breiman, L. (2001). Machine Learning. *45(1)*, 5–32. doi:10.1023/a:1010933404324

Büyükgüral, A., ve Türkoğlu, Y. (2015). Gümrük Kanunu’na Göre Tatbik Edilen Para Cezalarında Zamanaşımı Sorunu. *Gümrük ve Ticaret Dergisi, (6)*, 74-85. 03 03, 2022 tarihinde https://dergipark.org.tr/tr/pub/gumrukticaretdergisi/issue/53324/708737 adresinden alındı

Canrakerta, ve. H. (2020). Application of Business Intelligence for Customs Declaration: A Case Study in Indonesia. *Journal of Physics*. 01 05, 2022 tarihinde 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., ve 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

Chang, J. H., ve Fan, K. C. (2002). A new model for fingerprint classification by ridge distribution sequences. *Pattern Recognition, 35(6)*, 1209-1223.

Chen, K., Chen, H., Conway, N., Hellerstein, J. M., ve 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., ve Ma, X. (2019). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems, 105361*. doi:10.1016/j.knosys.2019.105361

Coşar, M., ve 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., ve 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., ve 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

Dağ, M., ve Kızılkaya, F. (2019). Türkiye'de Dış Ticaret Açığı Sürdürülebilir mi? Ekonometrik Bir Analiz. *Yaşar Üniversitesi E-Dergisi , Jourrnal of Yasar / Special Issue on Applied Economics and Finance*, 148-153. 03 01, 2022 tarihinde https://dergipark.org.tr/en/pub/jyasar/issue/44178/521430 adresinden alındı

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

Dalkey, N. C. (1969). *The Delphi Method: An Experimental Study of Group Opinion.* RAND Corp Santa Monica Calif. 03 01, 2022 tarihinde https://apps.dtic.mil/sti/citations/AD0690498 adresinden alındı

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

Delphi (software) - Wikipedia. (2022, 4 18). 4 18, 2022 tarihinde Wikipedia: https://en.wikipedia.org/wiki/Delphi_(software) adresinden alındı

Dobriban, E., ve 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

Efron, B. (2013). Bayes' Theorem in the 21st Century. *Science, 340(6137)*, 1177–1178. doi:10.1126/science.1236536

Estivill-Castro, V. (2002). Why so many clustering algorithms. *ACM SIGKDD Explorations Newsletter, 4(1)*, 65–75. doi:10.1145/568574.568575

Freedman, D. A. (2009). *Statistical Models: Theory and Practice.* Cambridge University Press.

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

Gardner, M. W., ve Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment, 32(14-15)*, 2627-2636.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* Wiley.

Grime, M. M., ve Wright, G. (2016). Delphi Method. *Wiley StatsRef: Statistics Reference Online*, 1–6. doi:10.1002/9781118445112.stat07879

Güldüren, B., ve Öztop, S. (2020). Dış Ticaret İşlemlerinde Bilişim Sistemlerinin Önemi. *Süleyman Demirel Üniversitesi Vizyoner Dergisi, Cilt: 11 Sayı: Ek*, 169-184. doi:10.21076/vizyoner.649512

Gulia, A., Vohra, R., ve Rani, P. (2014). Liver Patient Classification Using Intelligent Techniques. *International Journal of Computer Science and Information Technologies, 5 (4)*, 5110-5115. 01 05, 2022 tarihinde https://www.semanticscholar.org/paper/Liver-Patient-Classification-Using-Intelligent-Gulia-Vohra/798856e3c30ed88661d0aa596cc23e12410181f5#references adresinden alındı

Gümrük Rehberi. (2021). 12 01, 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

Gümrük Rehberi. (2022, 5 12). 5 12, 2022 tarihinde T.C. Ticaret Bakanlığı: https://gumrukrehberi.gov.tr/sayfa/g%C3%BCmr%C3%BCk-beyannamesi-nedir adresinden alındı

Gümrük Rehberi. (2022, 5 12). 5 12, 2022 tarihinde T.C. Ticaret Bakanlığı: https://gumrukrehberi.gov.tr/sayfa/ithalatta-g%C3%BCmr%C3%BCk-i%C5%9Flemlerini-nas%C4%B1l-yapacaks%C4%B1n%C4%B1z adresinden alındı

Gümrük Rehberi. (2022, 5 12). 5 12, 2022 tarihinde T.C. Ticaret Bakanlığı: https://gumrukrehberi.gov.tr/kategori/ticari-slemler/blge-sistemi-hakkinda adresinden alındı

Günerkan, M., Şahinaslan, E., ve Şahinaslan, Ö. (2022). Gümrük Beyannamesi Sürecinde Öğrenmeye Dayalı Algoritmaların Etkinliğinin İncelenmesi. *Acta Infologica*. doi:10.26650/acin.1057060

Günerkan, M., Şahinaslan, Ö., ve Şahinaslan, E. (2022). The Use and Success of Machine Learning Algorithms in Improving the Customs Declaration Process. *International Conference of Mathematical Sciences (ICMS 2022), ISBN 9786052124291, Maltepe University.* Istanbul.

Huang, J.-C., Ko, K.-M., Shu, M.-H., ve 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 Dokümanlar. (2021). 12 02, 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. 02 01, 2022 tarihinde https://dergipark.org.tr/en/pub/seyad/issue/55429/698700 adresinden alındı

İşler, İ. (2018). Türkiye'de Kamu Geliri Olarak Para Cezalarının Mali Boyutunun Değerlendirilmesi. *Gümrük ve Ticaret Dergisi, 13*, 48-57. 02 03, 2022 tarihinde

https://dergipark.org.tr/tr/pub/gumrukticaretdergisi/issue/53314/708577 adresinden alındı

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

Jaskowiak, P. A., ve Campello, R. G. (2011). Comparing correlation coefficients as dissimilarity measures for cancer classification in gene expression data. *In Proceedings of the Brazilian symposium on bioinformatics*, (s. 1-8).

Kaya, M., ve Doğan, A. (2020). Dış Ticarete Konu Eşyanın Vergilendirilmesinde Gümrük Kıymetinin Rolü, Beyanı ve Kontrolü. *Gümrük ve Ticaret Dergisi, 7 (19)*, 10-24. 01 04, 2022 tarihinde https://dergipark.org.tr/tr/pub/gumrukticaretdergisi/issue/53766/695384 adresinden alındı

Kibriya, A. M., Frank, E., Pfahringer, B., ve Holmes, G. (2004). Multinomial Naive Bayes for Text Categorization Revisited. *AI 2004: Advances in Artificial Intelligence*, 488–499. doi:10.1007/978-3-540-30549-1_43

Kohavi, R., ve Provost, F. (1998). Glossary of Terms. *Machine Learning, 2-3*, 271–274. doi:10.1023/A:1017181826899

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

Li, G., ve 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., ve 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. 02 01, 2022 tarihinde https://resolver.obvsg.at/urn:nbn:at:at-ubtuw:1-138802 adresinden alındı

Maruev, S., Stefanovskiy, D., Frolov, A., Troussov, A., ve 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=4458veMevzuatTur=1veMevzuatTertip=5 adresinden alındı

Microsoft SQL Server - Wikipedia. (2022, 4 18). 4 18, 2022 tarihinde Wikipedia: https://en.wikipedia.org/wiki/Microsoft_SQL_Server adresinden alındı

Mitchell, T. (1997). *Machine Learning.* McGraw Hill.

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

Ontivero-Ortega, M., Lage-Castellanos, A., Valente, G., Goebel, R., ve Valdes-Sosa, M. (2017). Fast Gaussian Naïve Bayes for searchlight classification analysis. *NeuroImage, 163*, 471–479. doi:10.1016/j.neuroimage.2017.09.001

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., ve 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., ve 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

Piryonesi, S. M., ve El-Diraby, T. E. (2020). Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. *Journal of Transportation Engineering, Part B: Pavements*, 146(2). doi:10.1061/jpeodx.0000175

Point, T. (2022). *Introduction to Naive Bayes Algorithm*. 01 05, 2022 tarihinde https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_naive_bayes.htm adresinden alındı

Python (programming language) - Wikipedia. (2022, 4 18). 4 18, 2022 tarihinde Wikipedia: https://en.wikipedia.org/wiki/Python_(programming_language) adresinden alındı

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., ve Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*, 93–104. doi:10.1016/j.isprsjprs.2011.11.002

Rokach, L., ve Maimon, O. (2014). *Data mining with decision trees: theory and applications, 2nd Edition.* World Scientific Pub Co Inc.

Roshan, S. E., ve 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., ve Sochenkov, I. (2019). Deep Learning for Customs Classification of Goods Based on Their Textual Descriptions Analysis. *2019 Ivannikov Ispras*

Open Conference (ISPRAS), 2019, s. pp. 55-59. doi:10.1109/ISPRAS47671.2019.00014

Saçık, S. Y. (2009). Dış Ticaret Politikası ve Ekonomik Büyüme İlişkisi. *Karamanoğlu Mehmetbey Üniversitesi Sosyal ve Ekonomik Araştırmalar Dergisi*, 2009(1), 162-171. 02 02, 2022 tarihinde https://dergipark.org.tr/tr/pub/kmusekad/issue/10220/125634 adresinden alındı

Şahinaslan, Ö., Dalyan, H., ve 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

Salve, S., Bhutkar, G., ve 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

Samworth, R. J. (2012). Optimal weighted nearest neighbour classifiers. *The Annals of Statistics, 40(5)*, 2733–2763. doi:10.1214/12-aos1049

Saygın, E., ve 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. (2013). Explaining AdaBoost. *Empirical Inference*, 37–52. doi:10.1007/978-3-642-41136-6_5

Shalev-Shwartz, S., ve Ben-David, S. (2014). *"18. Decision Trees"*. *Understanding Machine Learning.* Cambridge University Press.

Shang, F., Li, Y., Deng, X., ve He, D. (2017). Android malware detection method based on naive Bayes and permission correlation algorithm. *Cluster Computing, 21(1)*, 955–966. doi:doi:10.1007/s10586-017-0981-6

Shannon, C. (1948).

Shao, H. ve.-R. (2002). Applying data mining to detect fraud behavior in customs declaration. 04 01, 2022 tarihinde https://www.researchgate.net/publication/3995840_Applying_data_mining_to_detect_fraud_behavior_in_customs_declaration adresinden alındı

Siau, K. L., ve Yang, Y. (2017). Impact of Artificial Intelligence, Robotics, and Machine Learning on Sales and Marketing. *MWAIS 2017 Proceedings*, 48. 02 01, 2022 tarihinde http://aisel.aisnet.org/mwais2017/48 adresinden alındı

Singh, G., Kumar, B., Gaur, L., ve Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. *International Conference on Automation, Computational and Technology Management (ICACTM).* doi:10.1109/icactm.2019.8776800

Sourani, A., ve 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

Spyder (software) - Wikipedia. (2022, 4 18). 4 18, 2022 tarihinde Wikipedia: https://en.wikipedia.org/wiki/Spyder_(software) adresinden alındı

Stigler, S. M. (2013). The True Title of Bayes's Essay. *Statistical Science, 28(3)*, 283–288. doi:10.1214/13-sts438

Sundsøy, P., Bjelland, J., Iqbal, A. M., "Sandy" Pentland, A., ve 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

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of statistics, 24*, 303-329.

T.C. Ticaret Bakalığı. (2021, 12 31). 03 22, 2022 tarihinde Yillara Gore Beyanname Sayilari: https://ticaret.gov.tr/data/5d63d89d13b8762f7c43a738/20-Yillara%20Gore%20Beyanname%20Sayilari.pdf adresinden alındı

T.C. Ticaret Bakanlığı. (2021, 12 31). 03 22, 2022 tarihinde Yillara Gore Dis Ticaret: https://ticaret.gov.tr/data/5d63d89d13b8762f7c43a738/1-Yillara%20Gore%20Dis%20Ticaret.pdf adresinden alındı

T.C. Ticaret Bakanlığı. (2021, 12 31). 03 22, 2022 tarihinde Yillara Gore Tahsil Edilen Gumruk Vergileri ve Toplam Icindeki Payi: https://ticaret.gov.tr/data/5d63d89d13b8762f7c43a738/27-Yillara%20Gore%20Tahsil%20Edilen%20Gumruk%20Vergileri%20ve%20Toplam%20Icindeki%20Payi.pdf adresinden alındı

Tang, J., Deng, C., ve 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

Triola, M. F. (2010). *Bayes' theorem.* 03 23, 2022 tarihinde PDF: http://faculty.ccp.edu/faculty/sgramlich/152/bayes.pdf adresinden alındı

Turing, A. M. (2009). Computing Machinery and Intelligence. *Parsing the Turing Test*, 23–65. doi:10.1007/978-1-4020-6710-5_3

Türk Dil Kurumu. (2022, 4 18). 4 18, 2022 tarihinde Türk Dil Kurumu: https://sozluk.gov.tr adresinden alındı

Uçan, O., ve Koçak, E. (2014). Türkiye'de dış ticaret ve ekonomik büyüme arasındaki ilişki. *Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 7(2), 51. 02 01, 2022 tarihinde https://dergipark.org.tr/tr/pub/niguiibfd/issue/19755/211482 adresinden alındı

Wang, M. L., ve 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

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software, 6(60)*, 3021. doi:https://doi.org/10.21105/joss.03021

Witten, I., Frank, E., ve Hall, M. (2011). *Data Mining.* Burlington, MA: Morgan Kaufmann.

Zartha Sossa, J. W., Halal, W., ve 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., ve 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., ve 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., ve Rätsch, G. (2009). The Feature Importance Ranking Measure. *Lecture Notes in Computer Science*, 694–709. doi:10.1007/978-3-642-04174-7_45

ÖZGEÇMİŞ

Mustafa GÜNERKAN

Eğitim

Y.Ls.	2020 - 2022	Maltepe Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği (Tezli)
Ls.	1998 - 2004	Anadolu Üniversitesi, İktisat Fakültesi, İktisat
Ön. Li.	1994 - 1998	Boğaziçi Üniversitesi, Meslek Yüksek Okulu, Bilgisayar Programcılığı
Lise	1990 - 1993	Ceyhan Lisesi

İş/İstihdam

2007 - Halen	Barsan Global Lojistik – Yazılım Müdürü
1998 - 2006	Logo Yazılım – Yazılım Uzmanı

Yayınlar ve Diğer Bilimsel/Sanatsal Faaliyetler

"The Use and Success of Machine Learning Algorithms in Improving the Customs Declaration Process", International Conference of Mathematical Sciences (ICMS 2022), ISBN 9786052124291, Maltepe University, Istanbul.

Günerkan, M., Şahinaslan, E. ve Şahinaslan, Ö. (2022). Gümrük Beyannamesi Sürecinde Öğrenmeye Dayalı Algoritmaların Etkinliğinin İncelenmesi. Acta Infologica. DOI: 10.26650/acin.1057060