

Gümrük Kontrol Noktalarında Riskli Geçişlerin Belirlenmesine Yönelik Yapay Zekâ Temelli Bir Yaklaşım

Ezgi Zehra ŞEKER¹ , Ebru GEÇİCİ^{2*} , Alev TAŞKIN³ 

Öz

Gelişen teknoloji ve küreselleşmeyle birlikte ülkeler arasında insan ve ürün açısından giriş çıkışlar artmaya başlamıştır. Bu geçişlerde, ürünlerin ülkeler arasında aktarılması ile oluşan dış ticaret işlemlerinde ülkelerin belirli bölgelerinde yer alan sınır kapıları büyük önem taşımaktadır. Mal giriş çıkışının yapıldığı sınır kapıları gümrük olarak adlandırılmakta ve geçecek ürüne göre takip edilen süreçler farklılaşabilmektedir. Türkiye’de ise süreçlerin kontrol edilebilmesi için gümrük noktalarında üç farklı hat kullanılmaktadır: Kırmızı hat, sarı hat ve mavi hat. Kırmızı ve mavi hatlarda sırasıyla istisnasız tüm ürünler kontrol edilmekte ya da yetkilendirilmiş kişi sertifikasına sahip olanlar için kontrolsüz geçiş hakkı sağlanmaktadır. Sarı hatlarda ise ürünler gümrük memuru tarafından mevzuat ve yönergeye göre riskli ya da risksiz olarak sınıflandırılmakta ve bu sınıflandırma sonucuna göre gelen mallar kontrol edilmekte ya da edilmemektedir. Yapılan bu çalışmada sarı hat için ürünlerin riskli ya da risksiz olduğunu belirleyebilmek amacıyla makine öğrenmesi ve yapay sinir ağları yöntemleri kullanılarak model geliştirilecektir. Bu doğrultuda makine öğrenmesi başlığı altında yer alan k-en yakın komşu, lojistik regresyon, destek vektör makineleri, karar ağaçları, rassal orman ve naif Bayes yöntemleri ve yapay sinir ağları başlığı altında yer alan çok katmanlı algılayıcı (multi layer perceptron-MLP) yöntemi kullanılmıştır. Elde edilen sonuçlar incelendiğinde karar ağacı yönteminin mevcut veri seti için en iyi sonuçları verdiği gözlemlenmiştir.

Anahtar Kelimeler: Sınıflandırma, Makine öğrenmesi, Yapay sinir ağları, Yapay zekâ, Gümrük kontrol süreçleri.

An Artificial Intelligence Based Approach to Identify Risky Passes at Customs Control Points

Abstract

With the developing technology and globalization, the entrances and exits of people and products between countries have begun to increase. Border gates located in some certain areas of the countries have great importance in foreign trade transactions that occur by transferring products between countries during these transitions. Border gates where goods are entered and exited are called customs and the processes followed can be differ according to the product to be passed. In Turkey, three different lines are used at customs points to control the processes: red, yellow, and blue. On the red and blue lines, respectively, all products are controlled without exception or the right of uncontrolled passage is provided for those with an authorized person certificate. On the yellow lines, on the other hand, the products are classified as risky or risk-free by the customs officer according to the legislation and directive, and the incoming goods are controlled or not according to the results of this classification. In this study, a model will be developed for the yellow line using machine learning (ML) and artificial neural network (ANN) methods in order to determine whether the products are risky or risk-free. In this direction, k-nearest neighbor, logistic regression, support vector machines, decision trees, random forest, and naive Bayes methods under the ML title and multi-layer perceptron (MLP) method under the ANN title are used. When the results are examined, it is observed that the decision tree method gives the best results for the existing data set.

Keywords: Classification, Machine learning, Artificial neural networks, Artificial intelligence, Customs control processes.

^{1,2,3}Yıldız Teknik Üniversitesi, Makine Fakültesi, Endüstri Mühendisliği Bölümü, İstanbul, Türkiye, zehra.kadiz@yildiz.edu.tr
egecici@yildiz.edu.tr ataskin@yildiz.edu.tr

*Sorumlu Yazar/Corresponding Author

Geliş/Received: 28.09.2023

Kabul/Accepted: 17.05.2024

Yayın/Published: 18.06.2024

1. Giriş

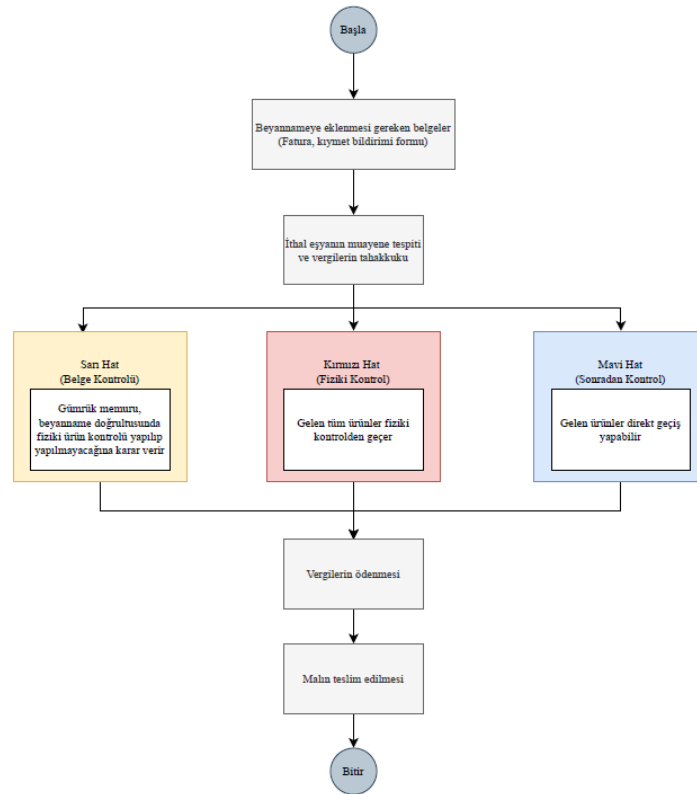
Küreselleşmeyle birlikte ülkelerin ekonomik anlamda bütünleşmeleri kaçınılmaz olmuştur. Dünya genelinde dış ticaretteki kısıtları ve kota sınırlarını azaltma ve kaldırma üzerine çalışmalar yapılmaktadır (Hatipler, 2011). Ülkeler arasındaki bu giriş çıkışların yapılması ve dış ticaret süreçlerinin yönetilebilmesi için ülkelerin sınırlarında yer alan belirli bölgelerde kontrol noktaları kurulmaktadır. Dış ticaret sürecinde ülke sınırlarından yapılan giriş çıkışlar sırasında mal akışının denetlendiği bu noktalara *gümrük* ya da *hudut kapısı* denilmektedir. Türkiye sınırlarında toplam 203 hudut kapısı bulunmaktadır ve Türkiye'nin sahip olduğu bu hudut kapıları Şekil 1'de gösterilmektedir. Bu hudut kapılarından 30'u kara hudutu, sekizi demiryolu hudutu, 101'i deniz hudutu ve 64'i hava hudutudur. Kara hudut kapılarında 29, demiryolu hudut kapılarında beş, deniz hudut kapılarında 43, hava hudut kapılarında ise 13 gümrük müdürlüğü hizmet vermektedir (Gümrük İşlemleri, 2023).



Şekil 1. Türkiye'deki gümrük noktaları (Türkiye'nin Sınır Kapıları, 2023).

Dış Ticaret Müsteşarlığı'nın görev ve yetki alanında bulunan ithalat politikalarının zaman zaman çeşitli yöntemlerle etkisiz kılınabilmektedir. Bu durum, uygulanan politikadan beklenen sonucun alınamamasına, vergi kaybı meydana gelmesine, oluşan haksız kazanç nedeniyle piyasa aktörleri arasında haksız rekabet doğmasına dolayısıyla piyasa başarısızlığına neden olmaktadır. İthalat politikalarını etkisiz kılma fiilleri başlıca Gümrük Tarife İstatistik Pozisyonu (GTİP), menşe ülke, eşyanın gümrük kıymeti, firma ve nihai kullanım alanı gibi politika parametrelerinde değişiklik yapmak suretiyle gerçekleştirilmektedir. Bu nedenle gümrük kapılarından geçecek ürünlere ait ilgili bilgilerin yer aldığı belgelerin ayrıntılı şekilde kontrol edilmesi gerekmektedir. Bu kapsamda öncelikli olarak gümrük kapısına gelen her mal belge kontrolüne alınır. Kontrol aşamalarının düzenli

bir şekilde yürütülebilmesi için var olan hudut noktalarından, o noktalar için tanımlanmış ürünler geçmekte ve geçişlerin sağlanabilmesi için Şekil 2’de yer alan akış diyagramındaki süreçler takip edilmektedir. Şekil 2’de belge kontrolleri sağlanan malların muayene işlemi için gidebileceği üç hat bulunmaktadır: Sarı hat, kırmızı hat ve mavi hat. *Kırmızı hattan* geçen ürünler riskli kategorisinde yer almaktadır. Bu hattan geçen tüm mallar belge kontrolüne ek olarak fiziksel kontrole tabi olmaktadır. Gerekirse mallara kimyevi tahliller de yapılmaktadır. *Sarı hat* ise eşyaya ait beyanname ve eklerinin doğruluğunun ve birbiriyle uygunluğunun kontrol edildiği hattır. Kırmızı hatların aksine burada her ürün fiziksel kontrolden geçmemekte, çalışan memurların beyanname doğrultusunda gerekli görmesi halinde malların kontrolü yapılmaktadır. *Mavi hat* ise T.C. Ticaret Bakanlığı tarafından belirlenen onaylanmış kişi ya da kurum statüsüne sahip kişilerin/kurumların ihracatta yararlandığı, eşyanın çıkış işlemlerinin tamamlanmasından önce belge kontrolüne veya muayeneye tabi tutulmadığı hattır. Başka bir deyişle bu hatta belgelerin kontrolü daha sonra gerçekleştirilmek üzere gümrük işlemleri sırasında yapılmaz. Bu hatta belirtilen işlemler T.C. Ticaret Bakanlığı tarafından belirlenen usul ve esaslar çerçevesinde gerçekleştirilmektedir.



Şekil 2. Gümrük süreçlerine ait akış diyagramı ve kontrol kodları.

Mavi hat ve kırmızı hatta takip edilecek işlemler başlangıç anında belirlidir. Eğer gelen ürünler mavi hattan geçecekse ürünlerin fiziki kontrolü yapılmadan ve belge kontrolü sonra gerçekleştirilmek üzere gelen ürünler direkt olarak hudut kapılarından geçebilmektedir. Kırmızı hatta yer alan ürünler

ise riskli kategoride oldukları için ürünlerin tamamı kontrol edilmektedir. Ancak sarı hatta yer alan kontrol işlemleri kırmızı ve mavi hattan farklı olarak o anda yapılan değerlendirmeye göre gerçekleşmektedir. Sarı hatta giriş çıkış sayılarının çokluğu nedeniyle tüm malların denetlenmesi mümkün olmamaktadır. Bu nedenle görevli memurun mevzuatlar dâhilinde gerekli gördüğü takdirde seçici inceleme ve denetim yapılmaktadır. Sarı hatta gelen ürünün fiziksel kontrolden geçip geçmeyeceğine, mevzuatlar doğrultusunda ürüne bağlı olarak gümrük memuru tarafından karar verilmektedir. Sarı hatlarda yapılacak kontroller T.C. Ticaret Bakanlığı tarafından yayınlanan mevzuat ve yönergeler dâhilinde görevli memurun kontrolüne bağlı olup yoğun dönemlerde doğru kararın verilmesi için gerekli süreç uzayabilir ve bu da kararın durumunu etkileyebilir. Alınan kararlar doğrultusunda bu hatta riskli görülen ürünler/geçişler ise kırmızı hatta yönlendirilmektedir. Bu çalışmanın kapsamı, sarı hatlarda yapılan geçişlerde hangi ürünlerin kontrol edilip edilemeyeceği ve kırmızı hatta yönlendirilmesi gerektiği konusunda görevli gümrük memuruna ön bilginin sunulması için bir karar destek sistemi oluşturmaktır. Bu doğrultuda memurun inisiyatifinde gerçekleşen muayene ve kontrol işlemine karar verilmesini sağlayan bir model geliştirilmesi amaçlanmaktadır. Bu kapsamda mevcut sistem için en uygun modelin geliştirilmesi amacıyla makine öğrenmesi ve yapay sinir ağları yöntemleri kullanılacaktır.

Yapay zekânın bir alt dalı olan makine öğrenmesi, var olan verileri kullanarak sistemi anlayıp cevap üretecek şekilde modelin eğitilmesine odaklanmaktadır. Yapay sinir ağları ise insan beyninin çalışma yapısından esinlenilerek oluşturulmuştur. Makine öğrenmesi modelleri ve yapay sinir ağları, var olan bilgileri kullanarak yeni çıktılar üretebilme ve sistemi eğitebilme özelliğine sahiptir. Bu nedenle çok çeşitli alanlarda gerekli bilgiyi elde etmek için kullanılmaktadır. Bu alanlardan biri de *sahteciliğin tespit edilmesidir*. Her iki yöntem de farklı alanlarda sahtecilik tespiti için kullanılmıştır. Kullanım alanlarından bazıları finansal durumda sahteciliğin tespiti (Perols, 2011), kredi kartı sahteciliğinin tespiti (Dornadula ve Geetha, 2019; Thennakoon ve ark., 2019; Maniraj ve ark., 2019), tıbbi bakım sigortasında sahteciliğin tespiti (Bauder ve Khoshgoftaar, 2017; Johnson ve Khoshgoftaar, 2019)'dir. Sahteciliğin tespiti amacıyla yapılan çalışmalar incelendiğinde, literatürde yer alan bu çalışmaların çoğunluğunun finans sektöründe olduğu gözlemlenmektedir. Literatürde, yazarların bilgisi dâhilinde gümrüklerde riskli geçişler kaynaklı sahteciliğin tespitine yönelik çalışmaya rastlanmamıştır. Az sayıda çalışmaya konu olan gümrük süreçleri için sahteciliğin belirlenmesi anlamında yapılan ilk çalışma olması bu çalışmanın literatürdeki katkısını göstermektedir. Başka bir deyişle akıllı tabanlı yapılar içeren ve gümrük kapılarında günün şartlarına ve mevzuata göre değişiklikleri tespit eden riske dayalı bir karar destek sistemi mevcut değildir. Bu nedenle geliştirilecek model ile denetleme alanında oluşturulacak ön bilgi sağlama sistemi bir ilk olma özelliği taşımaktadır. Bu makalenin amacı, makine öğrenmesi ve yapay sinir ağları yöntemleri kullanılarak en uygun model oluşturularak kısıtlı kaynakları en etkin şekilde kullanmayı sağlayacak

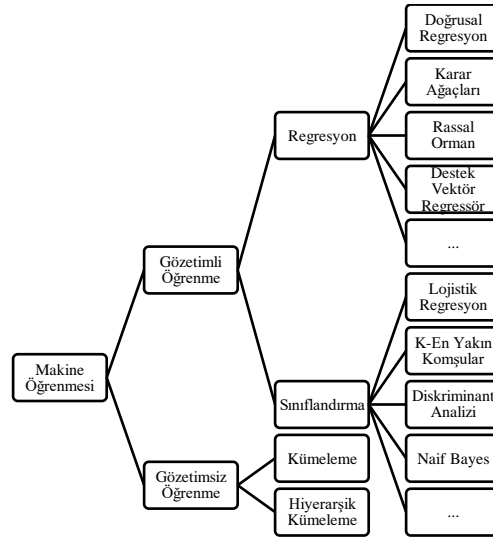
riske dayalı bir izleme ve denetim sistemi kurulmasıdır. Elde edilen model sayesinde sarı hatlara sonraki gelişlerde girişin riskli olup olmadığı bilgisi ile kontrolleri gerçekleştirecek gümrük memuru için ön bilgi oluşacaktır. Bu sayede gözden kaçabilecek riskli ürünlerin kontrol edilmeme durumu azalacaktır. Çalışmanın sonucunda ise riskli durumların gözden kaçmasının olabildiğince engellendiği, standardize olmuş ve sürdürülebilir bir yapı oluşturulması sağlanacaktır. Bu da ithalatçıların gümrük kanunlarına daha fazla uymasını sağlayacak ve yapılan izleme ve denetimler sonucunda yakalanan ihlallerin ve bunları önlemek için yapılacak politika değişikliklerinin sayısı azalacaktır.

Çalışmanın sonraki bölümlerinde takip edecek adımlar şu şekildedir: İkinci bölümde probleme ait modelin geliştirilmesinde kullanılacak olan metotlar anlatılacaktır. Üçüncü bölümde probleme ait detaylar sunulacaktır; veri seti, belirlenen yöntemlerin veri setine uygulanması ve nümerik çıktılar verilecektir. Son bölümde ise belirlenen yöntemler için sonuçlar ve değerlendirmeler sunulacaktır.

2. Teorik Metot

Yapılan bu çalışmada gümrük kapılarından yapılan geçişlerin riskli olup olmadığı konusunda bir ön bilginin oluşturularak gümrük memuruna sunulması amaçlanmaktadır. Problemin amacından da anlaşılacağı üzere problem bir sınıflandırma problemidir. Bu problemin çözülebilmesi ve gümrük memuruna doğruya yakın bilginin sunulabilmesi için en uygun modelin belirlenmesi gerekmektedir. Bu doğrultuda modellerin oluşturulmasında makine öğrenmesi yöntemleri ve yapay sinir ağlarından faydalanılacaktır.

Makine öğrenmesi temelde gözetimli ve gözetimsiz öğrenme olmak üzere ikiye ayrılmaktadır (Witten ve ark., 2013). Öğrenme yöntemlerinin gözetimli ve gözetimsiz olarak ikiye ayrılması, veri setinde yer alan ve bağımlı değişken olarak adlandırılan bir çıktı değerinin olup olmaması ile ilgilidir: (i) Eğer veri setinde bağımlı değişken, y , varsa *gözetimli öğrenme* ve (ii) veri setinde bağımlı değişken yoksa *gözetimsiz öğrenme* olarak adlandırılır. Bu ayrım ve kullanılan yöntemler Şekil 3'te özetlenmiştir. Regresyon ve tahminleme işlemi sürekli yapıya sahip olan bağımlı değişkenin tahmin edilmesi için kullanılırken sınıflandırma işlemi ise adından da belli olduğu gibi kesikli yapıya sahip olan bağımlı değişken kullanılarak verilerin sınıflandırılması işlemi kapsamaktadır. Gözetimsiz öğrenme başlığı altında yer alan kümeleme ise benzer özelliklere sahip olan bağımsız değişkenlerin bir araya gelmesi ile gruplandırma süreçlerini kapsamaktadır (Witten ve ark., 2013).



Şekil 3. Makine öğrenmesi tekniklerinin sınıflandırılması.

Bu çalışma, gümrük kapılarında yapılan geçişlerin riskli ya da risksiz olmalarına göre Şekil 3'te yer alan yöntemler kullanılarak sınıflandırılmalarına dayanmaktadır. Bu doğrultuda, gözetimli öğrenme başlığı altında yer alan sınıflandırma yöntemlerinden k-en yakın komşu, lojistik regresyon, naif Bayes, karar ağacı sınıflandırması, destek vektör sınıflandırması ve rassal orman yöntemleri kullanılacaktır. K-en yakın komşu (k-NN), sınıflandırma ve regresyon tahmini gerektiren konulara uygulanabilen denetimli bir makine öğrenme tekniğidir (Xiong ve Yao, 2021). En yakın komşu sayısı girdi olarak verilir ve uzaklığa göre sınıflandırma yapılır. Lojistik regresyon, olasılığa göre her bir bağımsız değişkenin istatistiksel önemini ölçmek için kullanılır. Bir sınıflandırma algoritması olarak büyük ölçüde olasılığa bağlıdır. Binom çıktı modelleyen güçlü bir makine öğrenmesi metodudur (Shah ve ark., 2020). Naif Bayes yöntemi ise, Bayes teoremine dayanan ve olasılıksal sınıflandırma yöntemi olarak adlandırılan yöntemlerden biridir (Aborisade ve Anwar, 2018). Naif varsayımı sınıf değişkenininin değeri verilen her özellik çifti arasında koşullu bağımsızlık varsayımına dayanmaktadır. Diğer karmaşık yöntemlere göre kullanımı daha kolay ve anlaşılır olması ve daha hızlı sonuçlar vermesi sebebiyle tercih edilen bir algoritmadır (Stephens ve ark., 2018). Karar ağacı sınıflandırma yöntemi, karar ağacı yapısını sınıflandırma için kullanan bir yöntemdir: Her düğüm bir özelliği, her dal bir kuralı ve her yaprak bir sonucu gösterir (Jadhav ve Channe, 2016). Karar ağaçları hiyerarşik bir yapıya sahiptir. Veri setinin küçük yapılara bölünmesi ile geliştirilir. Destek vektör sınıflandırması verileri birden çok sınıftan ayırmak için hiper düzlemler oluşturan, iyi bilinen bir makine öğrenme algoritmasıdır (Dong, 2022). Diğer sınıflandırma algoritmalarından farklı olarak oluşturulan kümeler arasındaki mesafeyi maksimize etmeye çalışmaktadır. Noktaların doğru veya düzlem ile ayrılması ve bu ayrım sonucunda ortaya çıkan mesafeyi dikkate almaktadır. Daha çok küçük ve orta büyüklükteki veri setleri için uygundur. Rassal orman algoritmasında ise birlikte çalışan çok sayıda karar ağacı oluşturulur ve tahmin için tüm ağaçların ortalaması kullanılır (Shah ve ark., 2020). Bu yapı, karar ağacı algoritma sonuçlarına göre daha tutarlı sonuçların elde edilmesini sağlar (Witten ve ark., 2013).

Yapay sinir ağlarının korelasyon tabanlı modellere göre avantajlarından biri büyük veri setini işleyebilmede daha yetenekli olmasıdır. Bağımlı ve bağımsız değişkenler arasındaki karmaşık doğrusal olmayan ilişkileri dolaylı olarak tespit etme kabiliyetine sahiptir (Pattanayak ve ark., 2021). Bu makalede makine öğrenmesi algoritmalarından sonra yapay sinir ağı uygulaması yapılacaktır. İnsan beyninin çalışma yapısından esinlenilerek oluşturulan bu yöntemde girdiler sırasıyla katmanlardan geçerek çıktı oluşmaktadır. Her bir katman nöronlardan oluşmakta ve buradan çıkan değerler aktivasyon fonksiyonundan geçerek elde edilmektedir.

Oluşturulan bu modellerden en iyi yöntemin seçilebilmesi için yöntemlerin performans göstergeleri kullanılarak değerlendirilmesi gerekmektedir. Kullanılan yöntemlerin belirlenmesinde olduğu gibi performans göstergelerinin de kullanılacak yöntemlere uygun olarak seçilmesi gerekmektedir. Bağımlı değişkenin yapısına bağlı olarak performans göstergeleri farklılık göstermektedir. Eğer ulaşılmaya çalışılan y değeri, kullanılan regresyon yöntemleri ile tahmin edilmeye çalışılıyorsa yani y değeri sürekli bir forma sahipse hata değerini hesaplamak için R^2 , ortalama kare hatası (mean squared error-MSE) gibi performans göstergeleri kullanılmaktadır. Eğer y değişkeni kesikli bir yapıya sahipse, bu sefer modelin doğruluğu, kesinliği gibi performans göstergelerine bakılmaktadır. Bu çalışmada kullanılacak olan veri seti incelendiği zaman y değişkeninin mevcudiyeti ve bu değişkenin 0 ya da 1 değerlerini alacak şekilde kesikli bir yapıya sahip olması nedeniyle performans göstergesi olarak *doğruluk oranı* (accuracy), *kesinlik değeri* (precision), *duyarlılık değeri* (recall) ve *F1-Skoru* kullanılacaktır. Bu performans göstergelerinin hesaplanması için Tablo 1’de yer alan *karişıklık matrisi* kullanılmaktadır. Karişıklık matrisi sınıflandırma sonucunda elde edilen değerlerin doğru ya da yanlış tahmin edilmesine göre oluşturulmaktadır. Bu matriste yer alan bilgiler Tablo 2’de yer aldığı şekilde kullanılarak ilgili performans göstergeleri hesaplanmaktadır (Alan ve Karabatak, 2020). Bu tablo, “Gösterge” sütunu başlığı altında sunulan performans göstergelerinin açıklamalarını ve nasıl elde edildiklerini vermektedir. Performans göstergelerine ait bilgiler, tabloda “Açıklama” sütununda yer alırken, performans göstergelerine ait matematiksel hesaplamalar “Formülasyon” sütunu altında verilmektedir.

Tablo 1. Karişıklık matrisi.

Gerçekleşen Değerler	Tahmin Değerleri	
	Doğru Pozitifler (DP)	Yanlış Negatifler (YN)
Yanlış Pozitifler (YP)	Doğru Negatifler (DN)	

Tablo 2. Sınıflandırma için kullanılan performans göstergeleri.

Gösterge	Açıklama	Formülasyon
Doğruluk Oranı	Doğru olarak sınıflandırılan örneklerin yüzdesini göstermektedir.	$Doğruluk = \frac{DP + DN}{DP + DN + YP + YN}$
Kesinlik Değeri	Pozitif olarak tahminlenen değerlerin gerçekten kaç adedinin doğru pozitif olduğunu göstermektedir.	$Kesinlik = \frac{DP}{DP + YP}$
Duyarlılık Değeri	Pozitif olarak tahmin edilmesi gereken işlemlerin ne kadarının pozitif olarak tahmin edildiğini gösteren performans göstergesidir.	$Duyarlılık = \frac{DP}{DP + YN}$
F1 Skoru	Testin doğruluğunun ölçüsüdür. Kesinlik ve duyarlılığın harmonik ortalaması şeklinde hesaplanmaktadır.	$F1 Skoru = 2 * \frac{Duyarlılık * Kesinlik}{Duyarlılık + Kesinlik}$

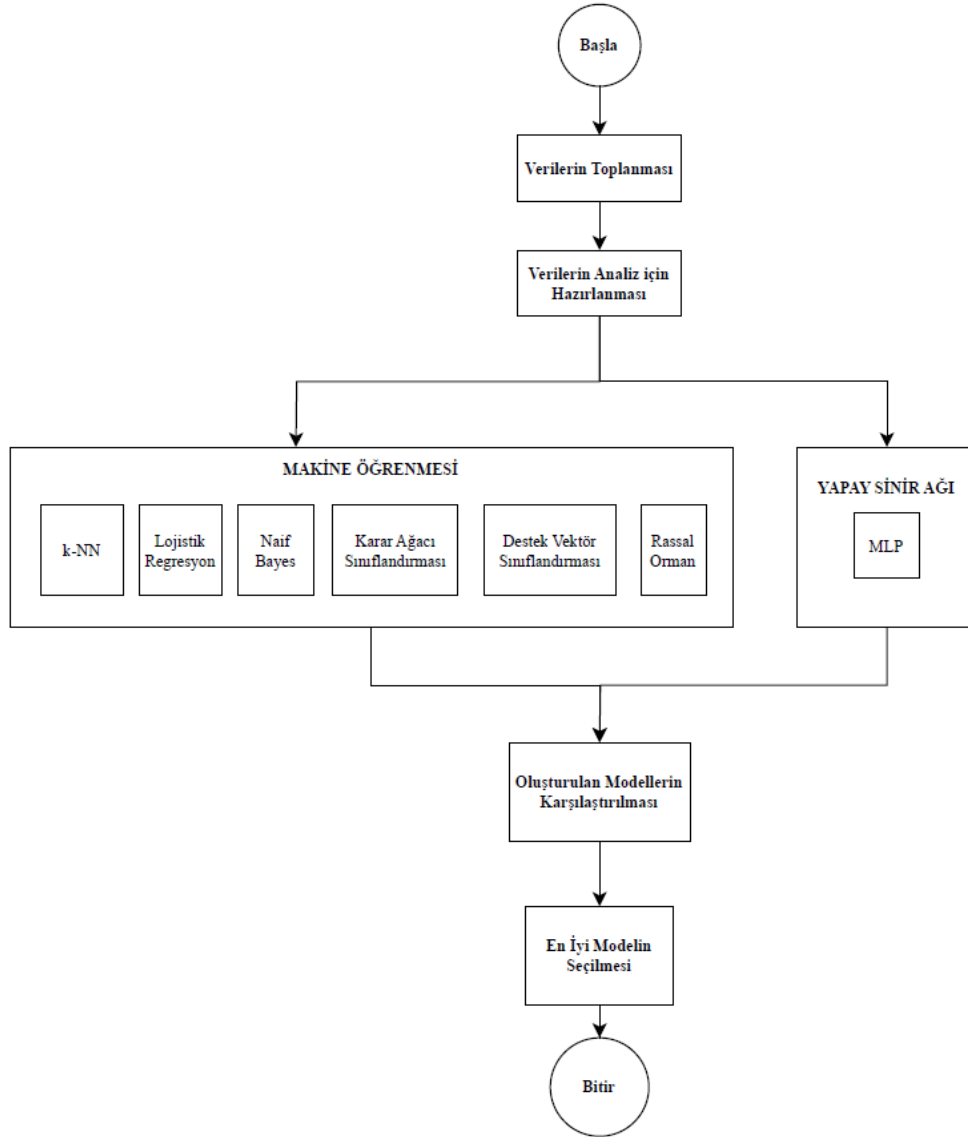
Bir sonraki bölümde yukarıda anlatılan yöntemler ile oluşturulan modeller, verilen performans göstergelerine göre değerlendirilerek mevcut veri seti için en uygun yöntemin belirlenmesi sağlanacaktır. Böylece geçişlerin riskli olup olmadığı değerlendirilecektir.

3. Uygulama: Gümrükte Riskli Geçişlerin Belirlenmesi

Gümrük süreçlerinde yapılan kontrol işlemleri sarı hat olarak adlandırılan kontrol noktalarında memurların değerlendirmesi doğrultusunda gerçekleştirilmektedir. Ayrıca kontrol sürecinde günün şartlarına ve mevzuata göre değişiklikleri tespit edebilen riske dayalı bir denetleme sistemi bulunmamaktadır. Kontrol eden kişiye bağlı olan bu süreçte gözden kaçabilen kontroller, ithalat politikalarının ihmal edilmesine, güvenlik sorunlarına, vergi kaçakçılığına ve haksız kazançlara sebebiyet verebilmektedir. İhlaller devam ettiği sürece devamlı politika değişikliğine gidilmesi gerekmekte ve bu durum ülkeler açısından maddi ve manevi kayıplara neden olmaktadır. Kişiye bağımlı olarak devam eden kontrol süreçlerinin otomatize edilerek bir standart oluşturulması hem güvenlik açısından fayda sağlayacak hem de ithalatçıların haksız kazanç ve rekabet avantajı kazanmalarını engellemiş olacaktır. Bu doğrultuda çalışmanın bu kısmında öncelikle kullanılacak veriler hakkında bilgi verilecek sonrasında ise en uygun modelin seçilmesi için yapılan analizler ve elde edilen sonuçlar sunulacaktır. Bu kapsamda Şekil 4'te yer alan akış diyagramındaki süreç takip edilecektir. Öncelikle verilerin incelenmesi ve analiz için hazırlanması süreçlerine yer verilecektir. Sonrasında ise makine öğrenmesi yöntemlerinden olan k-NN, lojistik regresyon, naif Bayes, karar ağacı sınıflandırması, destek vektör sınıflandırması ve rassal orman algoritması ile yapay sinir ağlarına ait model karşılaştırılarak problem için en uygun yöntemin belirlenmesi sağlanacaktır.

3.1. Veri Seti ve Verinin Analizler için Hazırlanması

Gümrük kapılarında geçişlerin riskli olup olmadığını belirlenmesi amacıyla en uygun sınıflandırma algoritmasının belirlenmesi için tescil numarası (geçiş yapılacak ürünlere ait belgenin tescil edilme numarası), tescil tarihi (belgenin tescil edildiği tarih), alıcı (ülkede hangi kişi/kurumun alacağı bilgisi), sevk edildiği ülke (ürünün ulaşacağı son noktanın ülke bilgisi), menşei ülke (ürünün geldiği ülke), gümrük GTİP (ürünün uluslararası olarak sınıflandırılabilmesini sağlayan 12’li koda ait ürün kodu), toplam ürün miktarı, ürünün toplam kıymeti, ürün birim fiyatı ve işçi sayısı (ürünü gönderen kurumda çalışan kişi sayısı) değişkenlerinin yer aldığı veri seti kullanılmıştır. Toplam 10 değişkenin yer aldığı veri seti 201.572 gözlem değerinden oluşmaktadır.



Şekil 4. Geçişlerin riskli olup olmadığının belirlenmesi için kullanılacak olan modelin elde edilmesinde kullanılan akış diyagramı.

Veri setinin analizlere uygun hale gelmesi için aşağıdaki adımlar takip edilmiştir.

1. Veri setinde yer alan eksik veriler incelenmiştir. İşçi sayısı değişkeninde eksik değerler tespit edilerek, bu satırlar veri setinden çıkarılmıştır.
2. Veri seti yalnızca bağımsız değişkenlerden oluşmaktadır. Sınıflandırma işleminin yapılabilmesi için bağımlı değişken olan y değişkenine ihtiyaç duyulmaktadır. Bu doğrultuda Tablo 3 ve Tablo 4'teki dönüşümler kullanılarak her bir geçiş için risk değeri oluşturulmuştur. Bu tablolar uzmanlardan alınan bilgiler kullanılarak elde edilmiştir. Her bir gözlem değerine tablodaki değerler uygun şekilde atanmış ve menşei ülke, birim fiyat, toplam miktar ve işçi sayısı değişkenlerinden elde edilen dört risk değeri toplanarak ilgili gözlem değeri için toplam risk değeri oluşturulmuştur.

Tablo 3. Bağımlı değişkenin oluşturulmasında kullanılacak olan dönüşümler.

Sıra	Menşei Ülke		Birim Fiyat	
	Seviye	Risk Değeri	Seviye	Risk Değeri
1	Yüksek Riskli	100	0-5	100
2	Riskli	50	5-10	75
3	Risksiz	0	10-15	50
4	-	-	15>	0

Tablo 4. Bağımlı değişkenin oluşturulmasında kullanılacak olan dönüşümler.

Sıra	Toplam Miktar		İşçi Sayısı	
	Seviye	Risk Değeri	Seviye	Risk Değeri
1	1-250	50	0-10	100
2	251-500	75	11-25	50
3	500>	100	25>	0

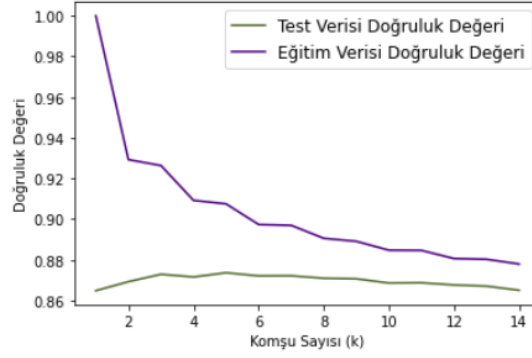
3. Tüm gözlem değerleri için toplam risk puanı hesaplandıktan sonra, risk puanlarına göre 0 (risksiz geçiş) ya da 1 (riskli geçiş) ataması yapılmıştır. Bu atamanın yapılabilmesi için toplam risk bağımlı değişkenine ait çeyreklik dilimleri kullanılmıştır. Çeyreklik dilimler, veri setinin küçükten büyüğe sıralandıktan sonra dörde bölünmesiyle oluşturulmaktadır. İlk çeyrek verinin %25'lik kısmına denk gelmektedir. Üçüncü çeyrek olarak adlandırılan kısım sıralanmış verinin %75'lik kısmını içermektedir ve bağımlı değişkenin oluşturulması için eşik değer olarak belirlenmiştir. 225'e karşılık gelen eşik değerinden yüksek risk puanına sahip gözlem değerlerine 1 ataması yapılarak geçişler riskli olarak değerlendirilmiştir. Böylece ikili yapıya sahip bağımlı değişken, y , elde edilmiştir.
4. Bağımlı değişken oluşturulduktan sonra veri setindeki aykırı (outlier) değerler analiz edilmiş, birim fiyat değişkenindeki aykırı değerler veri setinden uzaklaştırılmıştır.

5. Son olarak bağımsız değişkenler arasındaki korelasyon incelenerek ilişkili olan değişkenler belirlenmiştir. Aralarında korelasyon olan değişkenler, uzman görüşleri alınarak veri setinden çıkarılmıştır. Bu işlem sonucunda *alıcı*, *GTİP*, *toplam miktar*, *birim fiyat* ve *işçi sayısından* oluşan bağımsız değişkenler ve bağımlı değişken olan *risk durumunu* kapsayan 164.282 satırlık veri seti analizlere hazır hale getirilmiştir.
6. Veri seti oluşturulduktan sonra, öncelikle minimum-maksimum ölçeği (veri setinde yer alan her bir değişken içinde yer alan değerlerin [0,1] arasında olacak şekilde ölçeklendirilmesi) kullanılarak ölçeklendirme işlemi yapılmıştır.
7. Sonrasında modelin eğitilmesi için veri setinin %80'i eğitim seti ve test edilebilmesi için veri setinin %20'si test seti olarak ayrıştırılmıştır (Akbiyık ve Arı, 2022).

3.2. Modellerin Oluşturulması ve Elde Edilen Sonuçlar

Çalışmanın bu kısmında makine öğrenmesi başlığı altında incelediğimiz k-NN, lojistik regresyon, naif Bayes, karar ağacı sınıflandırması, destek vektör sınıflandırması ve rassal orman algoritması sonuçları ve yapay sinir ağları kapsamında incelenen çok katmanlı algılayıcı (multi-layer perceptron, MLP) modeline ait sonuçlar sırasıyla sunulacaktır. Bahsedilen modeller python programlama dili kullanılarak python paketleri ile oluşturulmuştur. Python paketleri scikit-learn tarafından sunulan makine öğrenmesi sınıflandırma algoritmalarını (sırasıyla NearestNeighbors, LogisticRegression, GaussianNB, DecisionTreeClassifier, svm ve RandomForestClassifier) ve MLP algoritmasını içermektedir. Oluşturulan modeller, Intel® Core™ i5-10210U CPU işlemcili 8 GB RAM ve 64 bit işletim sistemi özelliklerine sahip bilgisayar kullanılarak çalıştırılmıştır.

k-NN yönteminde k olarak ifade edilen komşuluk sayısının belirlenmesi gerekmektedir ve bu işlem için en çok tercih edilen yöntemlerden biri *dirsek grafiği* olarak adlandırılan yöntemdir. Şekil 5'te k-NN için oluşturulan dirsek grafiği yer almaktadır. Eğitim ve test hata değerlerinin azalmaya başladığı ve her ikisinin kabul edilebilir seviyede doğruluk değeri aldığı nokta, modele uygulanabilecek k adet yakın komşu sayısını göstermektedir (Witten ve ark., 2013; Alpaydin, 2020). Bu doğrultuda k değeri 6 olarak seçilmiştir.



Şekil 5. K-NN yöntemine ait dirsek grafiği.

k-NN algoritması için komşuluk sayısı belirlendikten sonra bu algoritma için çalıştırılan fonksiyonda yer alan farklı metrik değerleri kullanılarak modeller oluşturulmuştur. Bu kapsamda k-NN yönteminde en yakın komşuların hesaplanması için iki farklı algoritma kullanılmıştır: “auto (var olan verilere göre en uygun yöntemin seçilmesini sağlayan parametre değeri)” ve “ball_tree (eğitim verilerinin çoklu dairesel bloklara bölünmesi ve test verileri için mesafelerin tüm veriler yerine sadece bu bloklar içinde hesaplanır)”. Komşular arasında mesafeyi hesaplamak için “Minkowski” ve “Öklid” uzaklık hesaplama fonksiyonları kullanılmıştır. Son olarak bu yöntem için ağırlık parametresinde “mesafelerin tersi ile ağırlık noktaları (distance)” ve “eşit ağırlıklar (uniform)” metrikleri kullanılmıştır. Her parametre kullanılarak k-NN algoritması için farklı modeller oluşturulmuştur. Oluşturulan bu modellerden eğitim verisinde en iyi performansı veren modele ait çıktılar Tablo 5’te sunulmuştur. Model başlığı altında modeli oluşturmak için kullanılan parametrelerin neler olduğu belirtilmiştir. Diğer yöntemler açıklanırken Tablo 5’te yer aldığı şekilde o yöntem için en iyi modele ait çıktıların tüm yöntemler açıklandıktan sonra özet tablo olarak Tablo 6 şeklinde verileceğini hatırlatmak isteriz.

Tablo 5. k-NN modelinin test verisine ait performans göstergesi sonuçları ($k = 6$).

Model \ Performans Göstergesi	Doğruluk	Duyarlılık	Kesinlik	F1-Skoru
Model (Algoritma = “auto”, metrik = “Minkowski”, ağırlık = “distance”)	%87,87	%87,87	%87,61	%88

Kullanılan bir diğer yöntem ise lojistik regresyondur. Bu yöntemin uygulaması aşamasında ceza parametresine ait farklı metrikler kullanılmıştır. Bu doğrultuda “12” bu parametrenin yer almadığı “none” metriği kullanılarak iki model oluşturulmuş ve eğitim veri setinde en iyi sonucu veren modelin ceza parametresi verilmeyen model olduğu görülmüştür. Lojistik regresyona benzer şekilde naif Bayes yönteminde de ceza parametresinde değişik metrikler kullanılmış ve en iyi sonuçlar “none” metriği ile elde edilmiştir. Karar ağacı sınıflandırmasında ise lojistik regresyon ve naif Bayes yöntemlerinden farklı olarak birden fazla parametre için farklı metrikler kullanılmıştır. Bu doğrultuda

ayırma kalitesini ölçmek için için “gini” ve “entropy” metrikleri kullanılmıştır. Her iki parametre de sınıfların nasıl bölündüğünü değerlendirmek için kullanılmaktadır. Karar ağacında yer alması istenen maksimum değişken sayısını belirtmek için kullanılan bileşen için de “none”, “auto”, “sqrt” ve “log2” metrikleri kullanılmıştır. Modeller için elde edilen performans göstergelerine göre eğitim verisi için en iyi sonuçlar, ayırma amacıyla “entropy” metriğinin kullanıldığı ve maksimum değişken sayısının “none” olarak belirtildiği modelde elde edilmiştir. Destek vektör sınıflandırmasında ise oluşturulacak olan kümeler arasında bir doğrunun oluşturulması ve bu oluşturulan doğru mesafesinin sınıflar arasında olabildiğince maksimum olması sağlanmaya çalışılmaktadır. Bu kapsamda kernel tipinin “rbf”, “linear”, “poly” ve “sigmoid” olarak farklılaştırıldığı modeller oluşturulmuştur. Burada bahsedilen kernel, verilerin doğrusal yapıda olmayan sistemlerin hiper düzlem kullanılarak sınıflandırılmasında kullanılan isimdir ve bu parametre ile sınıflandırmada kullanılacak olan yöntemin belirtilmesi sağlamaktadır. Bu parametredeki değişime sınıfların ağırlıklarının da değiştirilmesi eklenmiştir. Elde edilen sonuçlara göre mevcut veri seti için en iyi sonuçlar “rbf” kernel fonksiyonunun sınıflarda ağırlık kullanılmadan oluşturulması ile elde edilmiştir. Burada bahsedilen “rbf”, radyal temel fonksiyonunun (radial basis function) kısaltması olarak kullanılmaktadır ve noktaların hangi kümeye ait olduğunun belirlenmesinde normal dağılımdan faydalanmaktadır. Destek vektör sınıflandırma yöntemi genellikle küçük ve orta büyüklükte veriler için daha iyi sonuçlar vermektedir (Li ve Yu, 2011). Yapılan çalışmada büyük bir veri seti kullanılmasından dolayı diğer yöntemlere kıyasla performans göstergeleri açısından eğitim veri setinde daha kötü sonuçlar elde edilmiştir. Makine öğrenmesi başlığı altında incelenen son yöntem olan rassal orman tekniği ise birden fazla karar ağacının aynı anda çalıştırılıp ortalamasının elde edilmesi ile oluşmaktadır (Shah ve ark., 2020). Bu yöntemde ormandaki ağaç sayısı (n), ayırmanın kalitesini ölçen fonksiyon ve sınıfların ağırlıkları ile ilgili farklı parametre değerleri değiştirilerek modeller oluşturulmuştur. Bu değişiklikler ile oluşturulan en iyi model, tahminleyici sayısının 1000, ayırma yönteminin “entropy” olduğu ve ağaç derinliğinin üç olarak ifade edildiği ağırlıksız sınıflı modelde elde edilmiştir.

Yapay sinir ağları temelde beş bölümden oluşmaktadır: Girdiler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktılar (Gardner ve Dorling, 1998). Ağırlıklar, toplama fonksiyonu ve aktivasyon fonksiyonu kısmında yer alan parametrelerin değiştirilmesi ile farklı modeller oluşturulabilmektedir. Böylece veri seti için en uygun yapay sinir ağı modeli belirlenebilmektedir. Python’da MLP modeli oluşturulabilmesi için en yaygın kullanılan kütüphaneler tensorflow ve keras’tır. Bu kütüphaneler kullanılarak yapay sinir ağı modelleri kurulmuş ve MLP modelinde yer alan aktivasyon fonksiyonu, hücre sayısı, katman sayısı, epoch sayısı, parti boyutu, doğrulama ayırımı, en iyileme motoru, kayıp parametreleri değiştirilmiştir. Yapılan işlemler sonucunda en iyi sonucun “adam” en iyileme motoru ile “aktivasyon fonksiyonu =

relu, relu, relu, sigmoid”, “katman sayısı = 3”, “epoch = 100”, “nod sayıları = 16, 64, 64, 1” metrikleri kullanılarak elde edilmiştir.

Yedi farklı yöntem kullanılarak gümrük noktalarındaki geçişlerin riskli ya da risksiz şekilde sınıflandırılması için oluşturulan modellere ait sonuçlar ise Tablo 6’da yer almaktadır. Her bir yöntem için oluşturulan modellerden en iyi sonucun elde edildiği modellere ait bilgilerin yer aldığı Tablo 6 aşağıdaki şekilde özetlenmiştir.

Tablo 6. Yöntemler için elde edilen sonuçlara ait en iyi modellerin yer aldığı özet tablo.

Yöntem	Model \ Performans Göstergesi	Doğruluk	Duyarlılık	Kesinlik	F1-Skoru
k-NN	algoritma = “auto”, metrik = “Minkowski”, ağırlık = “distance”	%87,87	%87,87	%87,61	%88
Lojistik Regresyon	ceza = none	%78,08	%78,08	%76,33	%76
Naif Bayes	ceza= none	%74,81	%74,81	%73,18	%67
Karar Ağacı Sınıflandırması	critierion= entropy, max_features=none	%93,43	%93,43	%93,41	%93
Destek Vektör Sınıflandırması	kernel = rbf, class_weight = None	%80,57	%80,57	%79,79	%80
Rassal Orman	n_estimators=1000, criterion = entropy, class_weight= None, max_depth = 3	%84,31	%84,31	%83,94	%84
Yapay Sinir Ağı	Optimizer = adam, Aktivasyon Fonksiyonu = Relu, Relu, Relu, Sigmoid, Katman Sayısı = 3, Epochs = 100, Node Sayıları = 16, 64, 64, 1	%84,43	%85	%85	%85

Tablo 6’da yer alan değerler incelendiğinde oluşturulan modellere göre en kötü performansın naif Bayes ve sonrasında lojistik regresyon yöntemleriyle elde edildiği görülmektedir. En iyi performansın ise k-NN ve karar ağacı sınıflandırması yöntemlerinde elde edildiği gözlenmiştir. Ancak k-NN’deki test – eğitim verisi performans göstergelerinin karşılaştırılması yapıldığında k-NN algoritmasının karar ağacı yöntemine göre daha düşük performans gösterdiği görülmektedir. Bu durumda mevcut veri seti için en iyi yöntem karar ağacı sınıflandırması olarak ortaya çıkmaktadır. Buna ilaveten her model için elde edilen performans değerlerinin literatürde yer alan minimum kullanılabilirlik seviyesinin üstünde olduğunu ve her modelin karar verme amacıyla kullanılabilceğini söylenebilir. Makine öğrenmesi modellerinde bir modelin kullanılabilmesi için en az %60 tercihen %70’in üzerinde doğru tahmin yapma yeteneğine sahip olması gerekmektedir.

Modelin tahmin gücünün belirlenmesi için gümrük kapılarından yapılan geçişlere ait farklı veriler modele verilmiştir. Bu kapsamda veri setinden rassal olarak 10 veri seçilmiş (metin içinde gösterim kolaylığı olması nedeniyle 10 olarak seçilmiştir) ve sonrasında karar ağacı sınıflandırması modeli kullanılarak seçilen gözlem değerleri için bağımlı değişken olan riskli geçiş olup olmadığı kontrol edilmiştir. Bu doğrultuda elde edilen sonuçlar Tablo 7’de verilmiştir. Tabloda yer alan

“Gözlem” sütunu o gözlem değerinin veri setindeki sırasını göstermektedir. Bu değerlerin veri seti içinden rassal olarak seçildiği buradan takip edilebilmektedir. “Gerçek Bağımlı Değişken Değeri” sütunu veri setinde o gözlem değerinin sahip olduğu bağımlı değişken değerini gösteriyorken “Tahmini Bağımlı Değişken Değeri” sütunu ise gözlem değerine ait ilgili verilerin modele girilerek tahmin edilmesi sonucunda elde edilen model çıktısını ifade etmektedir.

Tablo 7. Seçilen model kullanılarak yapılan tahmin sonuçları.

Gözlem	Gerçek Bağımlı Değişken Değeri	Tahmini Bağımlı Değişken Değeri
155118	0	0
72894	0	0
93893	0	0
18860	1	1
11404	0	0
22704	1	0
171504	0	1
156312	0	0
30410	0	0
107111	0	0

Tablo 7’de yer alan değerler için karışıklık matrisi oluşturulduğunda yanlış pozitif ve yanlış negatif olarak adlandırılan ve bizim azaltmaya çalıştığımız değerlerin 1 olarak elde edildiği görülmektedir. Başka bir deyişle verilen örnek veri setinde 1 gözlem değeri gerçekte riskli bir geçiş olarak sınıflandırılmışken, model tarafından risksiz geçiş olarak değerlendirilmiştir. Aynı şekilde başka bir geçiş ise gerçekte riskli bir geçişe olarak sınıflandırılmamışken model riskli geçiş olarak adlandırmıştır.

4. Sonuçlar ve Öneriler

Küreselleşen dünya ile birlikte ülkeler arasındaki hareketlilik artmış, ticaret süreçleri gelişmiş, hacim ve çeşitlilik artışı karşısında teknolojik gelişmelere de paralel olarak akıllı kontrol sistemleri kurulma ihtiyacı ortaya çıkmıştır. Ülkeler arasında mal giriş çıkışlarının kontrollerinin yapıldığı noktalar gümrük kapısı ya da hudut noktası olarak adlandırılmaktadır. Türkiye’de yer alan 203 gümrük kapısında da süreçlerdeki kontroller kırmızı hat, mavi hat ve sarı hat olarak ayrılan renk kodlarına göre devam etmektedir. Kırmızı hatlara gelen tüm ürünler fiziki kontrolden geçiyorken mavi hatlardan yapılan geçişlerde fiziki kontrol olmaksızın direkt geçiş sağlanmaktadır. Sarı hatlarda ise gümrük memurunun görüşlerine göre ürünler kontrol edilmekte ya da edilmemektedir. Ancak kişiye bağlı olan bu durumda riskli geçişlerin kontrol edilmeden geçme ihtimali az da olsa bulunmaktadır. Bu çalışmanın amacı gümrük süreçlerinde yapılacak geçişlerin riskli olup olmadığı hakkında bir bilginin oluşturularak gümrük memuruna kontrol öncesi ön bilginin verilmesini sağlayacak otomatize bir sistemin oluşturulmasıdır. Bu doğrultuda makine öğrenmesi yöntemleri olan

k-NN, lojistik regresyon, naif Bayes, karar ağacı sınıflandırması, destek vektör sınıflandırması ve rassal orman algoritması ve yapay sinir ağları kapsamında incelenen MLP yöntemi kullanılarak modeller oluşturulmuş ve doğruluk değeri, kesinlik değeri, duyarlılık ve F1-Skoru performans göstergelerine göre oluşturulan modeller karşılaştırılmıştır. Elde edilen sonuçlara göre mevcut modeller arasında en iyi sınıflandırmanın karar ağacı sınıflandırması ile elde edildiği görülmüştür.

Bu çalışmada var olan değişkenler, makine öğrenmesi ve yapay sinir ağları kullanılarak oluşturulmuş ve bu alanda yapılan ilk çalışma olma özelliği göstermektedir. Sonraki çalışmalarda ise veri setinde yer alan değişkenler için değişken seçimi algoritmaları uygulanarak, kullanılan yöntemler geliştirilebilir.

Yazarların Katkısı

Tüm yazarlar çalışmaya eşit katkıda bulunmuştur.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- Aborisade, O., Anwar, M., (2018). Classification for Authorship of Tweets by Comparing Logistic Regression and Naive Bayes Classifiers. IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 2018, 269-276
- Akbıyık, A., & Arı, O. (2022). Lojistik Regresyon İle Faydalı Müşteri Yorumlarını Tahminleme. Journal of Research in Business, 7 (IMISC 2021 Special Issue), 15-32.
- Alan A., & Karabatak, M. (2020). Veri Seti-Sınıflandırma İlişkisinde Performansa Etki Eden Faktörlerin Değerlendirilmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 32(2), 531-540.
- Alpaydin, E. (2020). Introduction to machine learning. MIT press.
- Bauder, R. A., & Khoshgoftaar, T. M. (2017, December). Medicare fraud detection using machine learning methods. In 2017 16th IEEE international conference on machine learning and applications (ICMLA) (pp. 858-865). IEEE.
- Dong, S. (2022, January). Virtual currency price prediction based on segmented integrated learning. In 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA) (pp. 549-552). IEEE.
- Dornadula, V. N., & Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. Procedia computer science, 165, 631-641.
- Gardner, M. W., & 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.

- Gümrük İşlemleri, Erişim Tarihi: 02.01.2023, <https://ticaret.gov.tr/gumruk-islemleri/sikca-sorulan-sorular/ticari/gumruk-islemleri>
- Hatıplı, M. (2011). Türkiye AB Gümrük Birliği Antlaşması ve Antlaşmanın Türkiye Ekonomisine Etkileri, Trakya Üniversitesi Sosyal Bilimler Dergisi, 13(1), 14-32.
- Jadhav, S. D., & Channe, H. P. (2016). Comparative study of K-NN, naive Bayes and decision tree classification techniques. International Journal of Science and Research (IJSR), 5(1), 1842-1845.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Medicare fraud detection using neural networks. Journal of Big Data, 6(1), 1-35.
- Li, X. ve Yu, W. (2011). Fast Support Vector Machine Classification for Large Data Sets. International Journal of Computational Intelligence Systems 7(2), 197-212.
- Maniraj, S. P., Saini, A., Ahmed, S., & Sarkar, S. (2019). Credit card fraud detection using machine learning and data science. International Journal of Engineering Research, 8(9), 110-115.
- Pattanayak, S., Loha, C., Hauchhum, L., Sailo, L. (2021). Application of MLP-ANN Models for Estimating the Higher Heating Value of Bamboo Biomass, Biomass Conversion and Biorefinery 11, 2499–2508.
- Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. Auditing: A Journal of Practice & Theory, 30(2), 19-50.
- Shah, K., Patel, H., Sanghvi, D., Shah, M., (2020). A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification. Augment Hum Res 5, 12
- Stephens, C.R., Huerta, H.F. & Linares, A.R., (2018). When is the Naive Bayes approximation not so naive?. Mach Learn 107, 397–441.
- Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., & Kuruwitaarachchi, N. (2019, January). Real-time credit card fraud detection using machine learning. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 488-493). IEEE.
- Türkiye'nin Sınır Kapıları (Gümrükler), Erişim Tarihi: 02.01.2023, <https://www.tarihselbilgi.com/sinir-kapilari/>
- Xiong L., Yao, Y. (2021). Study on an adaptive thermal comfort model with K-nearest-neighbors (KNN) algorithm. Building and Environment, 202,108026