



## TÜRKİYE'NİN BİLGİSAYAR, ELEKTRONİK VE OPTİK ÜRÜNLERİ DIŞ TİCARETİNİN MAKİNE ÖĞRENMESİ İLE TAHMİNLENMESİ

Tolga HASTAOĞLU<sup>1\*</sup>, Tolga AYDIN<sup>1</sup>

<sup>1</sup>Ataturk University, Faculty of Engineering, Department of Computer Engineering, 25100, Erzurum, Türkiye

**Özet:** Bu çalışmanın amacı, Türkiye'nin bilgisayar, elektronik ve optik ürünleri dış ticaretinin dinamiklerini anlamak ve makine öğrenmesi tekniklerini kullanarak bu ürünlerin dış ticaret verilerini tahmin etmektir. Çalışma, dış ticaret stratejilerinin geliştirilmesine katkıda bulunmayı ve Türkiye'nin uluslararası ticaret pazarındaki rekabet gücünü artırmak için veri odaklı karar verme süreçlerinin önemine vurgu yapmayı hedeflemektedir. Çalışma kapsamında öncelikle veri setindeki anomali gözlemler İzolasyon Ormanı (Isolation Forest) yöntemi ile tespit edilmiştir. Daha sonra Minimum Redundancy Maximum Relevance (MRMR) ve Recursive Feature Elimination (RFE) yöntemleri kullanılarak etkili değişkenler belirlenmiştir. Kümeleme aşamasında, veri setindeki örüntüleri belirlemek amacıyla K-Ortalamlar (K-Means) ve CLARANS algoritmaları kullanılmıştır. Son olarak, tahminleme modelleri olarak XGBoost, LightGBM ve Rastgele Orman Regresörü (Random Forest Regressor) uygulanarak dış ticaret verilerinin gelecekteki eğilimleri öngörülmüştür. Bu yöntemlerden üç farklı model oluşturulmuştur. İlk modelde, Isolation Forest algoritması ile anomali tespiti yapılmış, ardından özellik seçimi için MRMR ve RFE yöntemleri kullanılmış ve sonuç olarak Random Forest Regressor, XGBoost ve LightGBM regresyon modelleri uygulanmıştır. İkinci modelde, yine özellik seçimi yapılmış, sonrasında CLARANS ve K-means kümeleme algoritmaları uygulanmış ve ardından regresyon analizi gerçekleştirilmiştir. Üçüncü modelde ise, anomali tespiti için Isolation Forest kullanılmış, özellik seçimi yapılmış, kümeleme algoritmaları uygulanmış ve regresyon modelleri ile sonuçlar değerlendirilmiştir. Elde edilen sonuçlar karşılaştırılarak, dış ticaret verileri üzerinde makine öğrenmesinin etkinliği ortaya konulmuştur. Kullanılan metriklere göre en iyi sonuç model 2 (Clarans+MRMR+Random Forest) modeli ile  $R^2=0,8322$  olacak şekilde elde edilmiştir.

**Anahtar kelimeler:** Dış ticaret, Makine öğrenmesi, Regresyon


### Prediction of Türkiye's Foreign Trade of Computer, Electronic and Optical Products with Machine Learning


**Abstract:** The objective of this study is to understand the dynamics of Türkiye's foreign trade in computer, electronic, and optical products and to predict the trade data of these products using machine learning techniques. The study aims to contribute to the development of foreign trade strategies and emphasizes the importance of data-driven decision-making processes to enhance Türkiye's competitiveness in the international trade market. In this study, anomalous observations in the dataset were first identified using the Isolation Forest method. Subsequently, the most effective variables were determined using the Minimum Redundancy Maximum Relevance (MRMR) and Recursive Feature Elimination (RFE) methods. In the clustering phase, K-Means and CLARANS algorithms were employed to identify patterns within the dataset. Finally, prediction models, including XGBoost, LightGBM, and Random Forest Regressor, were applied to forecast future trends in foreign trade data. Three different models were developed using these methods. In the first model, anomaly detection was performed using the Isolation Forest algorithm, followed by feature selection using MRMR and RFE, and subsequently, Random Forest Regressor, XGBoost, and LightGBM regression models were applied. In the second model, feature selection was again conducted, followed by the application of CLARANS and K-Means clustering algorithms, and regression analysis was performed. In the third model, the Isolation Forest was used for anomaly detection, feature selection was conducted, clustering algorithms were applied, and results were evaluated using regression models. The results obtained were compared to demonstrate the effectiveness of machine learning on foreign trade data. According to the metrics used, the best result was achieved with Model 2 (CLARANS + MRMR + Random Forest) with an  $R^2$  value of 0.8322.

**Keywords:** Foreign trade, Machine learning, Regression

\*Sorumlu yazar (Corresponding author): Ataturk University, Faculty of Engineering, Department of Computer Engineering, 25100, Erzurum, Türkiye

E mail: tolga.hastaoğlu@icloud.com (T.HASTAOĞLU)

Tolga HASTAOĞLU  <https://orcid.org/0009-0001-0444-1167>

Tolga AYDIN  <https://orcid.org/0000-0002-8971-3255>

Gönderi: 12 Mayıs 2025

Kabul: 14 Ağustos 2025

Yayınlanma: 15 Eylül 2025

Received: May 12, 2025

Accepted: August 14, 2025

Published: September 15, 2025

Cite as: Hastaoğlu T, Aydın T. 2025. Prediction of Türkiye's foreign trade of computer, electronic and optical products with machine learning BSJ Eng Sci, 8(5): 1478-1492.

### 1.Giriş

Dış ticaret verileri, büyüme ve kalkınma süreci olarak ülkelerin ekonomik göstergelerinde çok önemli bir yere sahiptir. 1970'li yıllarda ortaya çıkan küreselleşme süreci kısa bir dönemde tüm dünyayı etkilemiştir. 1980 ve 1990'lı yıllarda ise küreselleşme finans piyasalarında çok

daha önemli bir boyuta ulaşmıştır (Aytekin, 2013). küreselleşmenin artışıyla birlikte, ülkelerin ihracat ve ithalat hacmi ekonomik istikrar ve sürdürülebilir büyüme açısından kritik hale gelmiştir. Dış ticaret bir ülkeden diğerine ihraç veya ithal edilen mal ve hizmetlerin değeridir, bu değer uluslararası ticaret politikaları ve

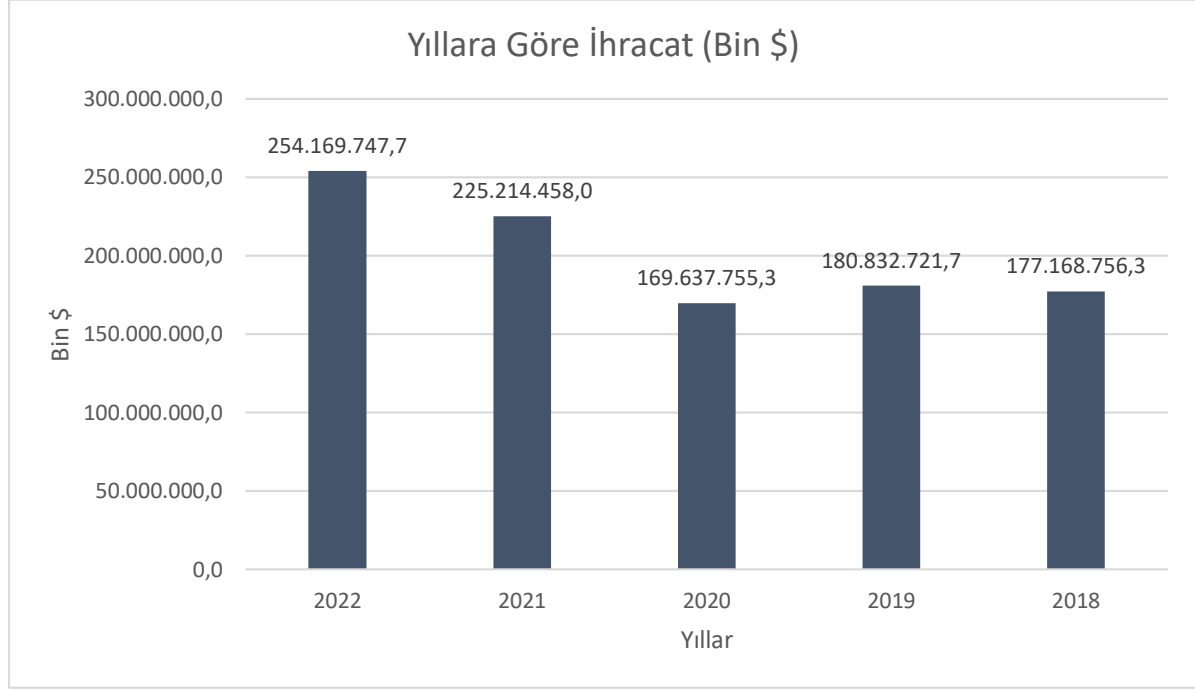


ticaretin yapıldığı her iki ülkenin iç ekonomi politikalarını etkiler. Ayrıca ülkelerdeki politika yapımcılar ve o ülkenin ekonomistleri için de gösterge niteliğindedir (Circlaeys vd., 2017).

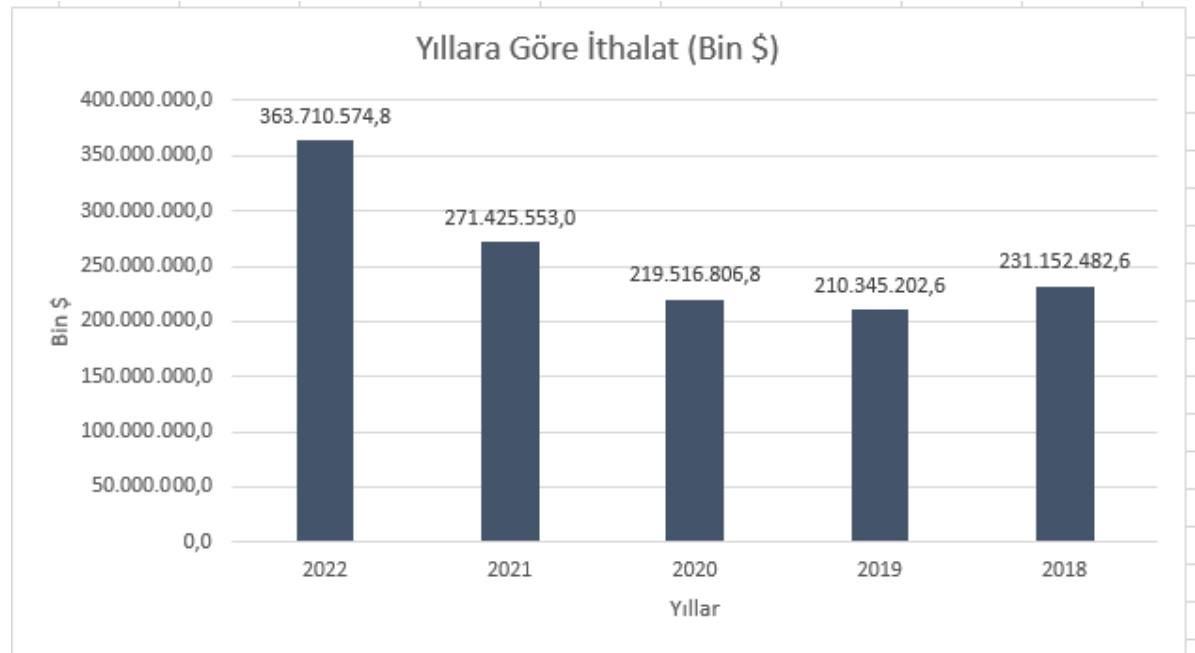
Türkiye, stratejik konumu ve üretim kapasitesi ile bölgesel bir ticaret merkezi konumunda olup, özellikle son yıllarda dış ticaret hacmini önemli ölçüde arttırmıştır. Bu nedenle, Türkiye'nin dış ticaret performansını tahmin etmek, ekonomik karar alıcılar ve politika yapımcılar için değerli bilgiler sunmaktadır. Şekil 1 ve Şekil 2'de Türkiye Dış Ticaret verilerinin yıllara göre ithalat ve ihracat

verileri verilmiştir. Şekillerde de görüldüğü üzere 2020 yılında bütün dünyayı etkileyen COVID 19 pandemisi ile dış ticaret hacmi düşmüş, 2021 ve 2022'de ise artışa geçmiştir.

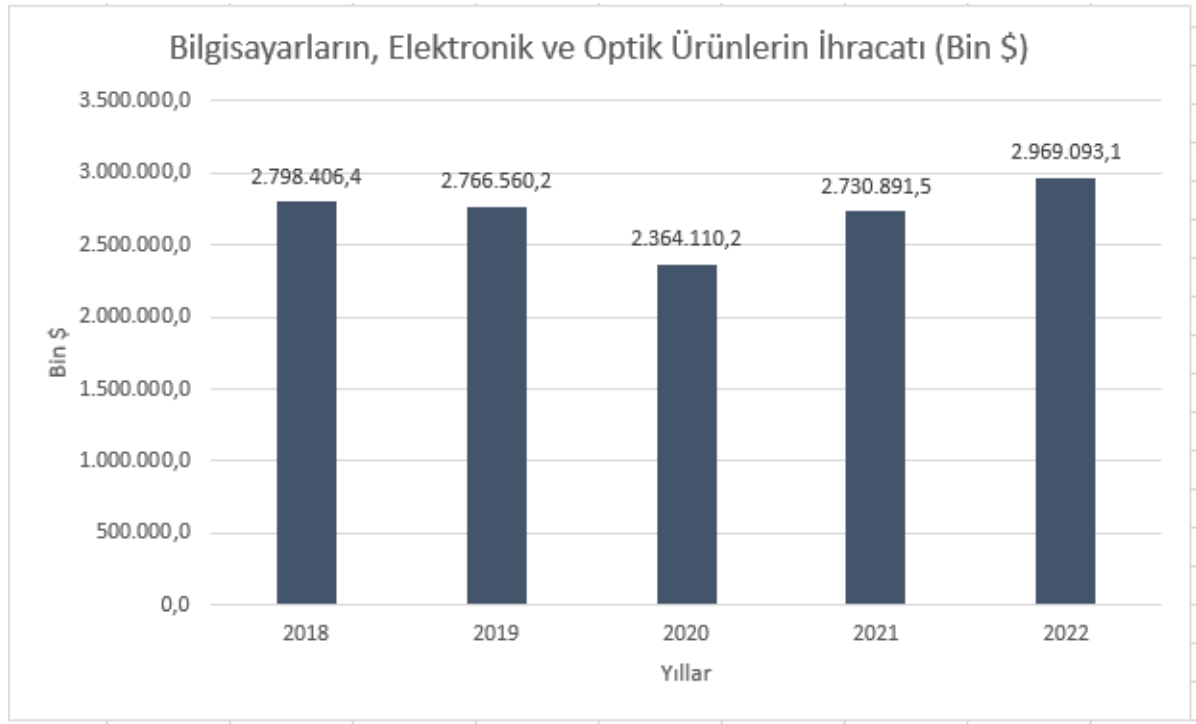
Şekil 3 ve Şekil 4 ise çalışma kapsamında ele alınan Bilgisayarların, Elektronik ve Optik Ürünlerin imalatı için ithalat ve ihracat Şekilleri verilmiştir. Bu tür teknolojik ürünlerin ithalatının, ihracatına göre yaklaşık 6 kat daha büyük olduğu görünmekte ve ilgili sektörde dış ticaret açığı olduğu açıkça görünmektedir.



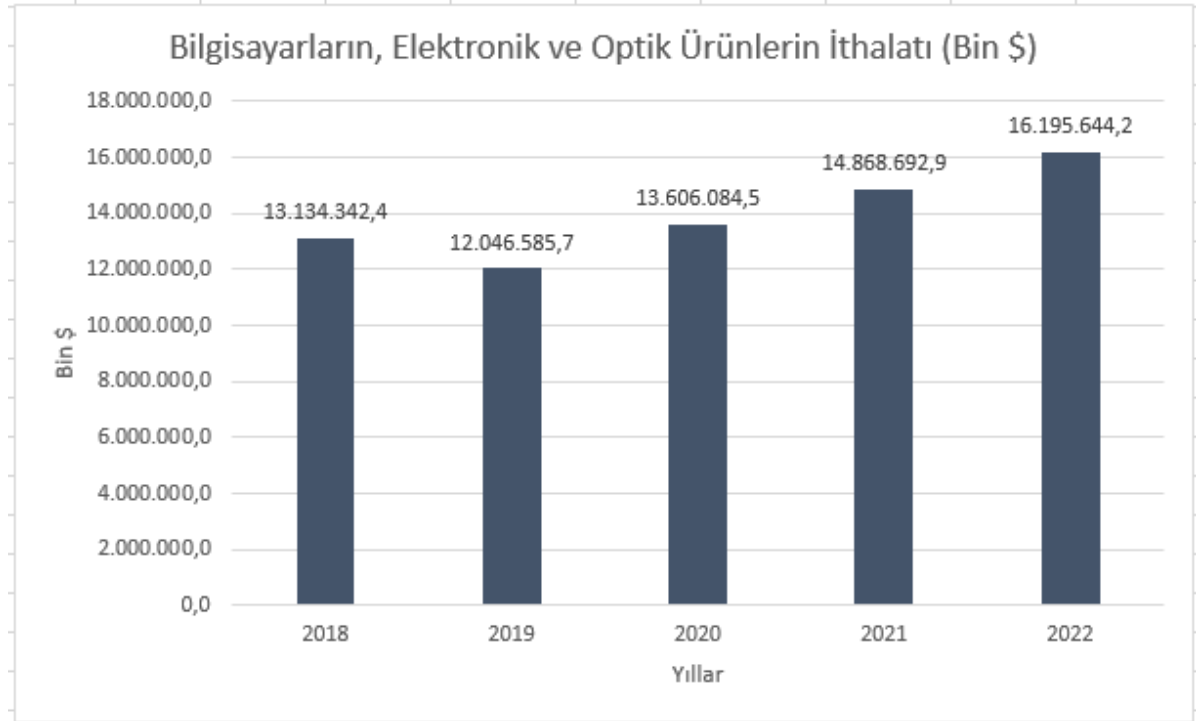
Şekil 1. Türkiye toplam ihracat grafiği



Şekil 2. Türkiye toplam ithalat grafiği



Şekil 3. Bilgisayarların, elektronik ve optik ürünlerin imalatı ihracat grafiği



Şekil 4. Bilgisayarların, elektronik ve optik ürünlerin imalatı ithalat grafiği

Dış ticaret verilerinde geleneksel olarak yer çekimi modeli kullanılmaktadır. Newton'un yerçekimi yasasından yola çıkılan bu yöntemde Reçus ve Timbergen (1964) iki ülke arasındaki ticaret akışını ülkelerin Gayri Safi Yurt İçi Hasılası ve ülkeler arasındaki mesafe ile kurmuştur. Bu model ampirik olarak güçlü olduğu için çok yoğun bir şekilde yıllarca kullanılmıştır. Anderson (2021) bu model için teorik ve çok ayrıntılı bir literatür hazırlamıştır. Ancak geleneksel tahmin yöntemleri, dış ticaret gibi büyük

hacimli ve karmaşık verilerle çalışırken sınırlı kalabilmektedir (Aktan, 2018). Bu nedenle büyük hacimli ve karmaşık verilerin çözümü için makine öğrenmesi tabanlı modeller ön plana çıkmaktadır.

Torkul vd. (2017)'ne göre "Makine öğrenmesi mantıksal işlemler gerçekleştirebilen makinelerin, gözlem ve ölçüm yöntemleriyle elde edilen verileri tecrübe olarak kabul etmesi ve bu tecrübelerden matematiksel algoritmalar aracılığıyla anlamlı ilişkiler üretmesi sürecidir".

Öğrenmenin amacı elimizde bulunan veriler ile bilgisayarların karar verebilme yeteneğinin kullanılmasıdır. Makine öğrenmesi uygulamalarında üç çeşit öğrenme yöntemi vardır. Bu yöntemler denetimli öğrenme, denetimsiz öğrenme ve yarı denetimli öğrenme şeklindedir.

Sınıflandırma, regresyon, karar ağaçları için denetimli öğrenme kullanılır. Bu öğrenme algoritmasında hem giriş hem de çıkış verileri sisteme verilir. Giriş verileri kullanılarak çıkış verileri tahminlenmeye ya da öğrenmeye çalışılır. Araştırmada kullanılan giriş ve çıkış verileri bu sistemdeki algoritmaya işlenerek performansı ölçülür ve araştırmada kullanılan çıkış verileri tahmin edilmeye çalışılır. Denetimsiz öğrenmede sisteme giriş verileri verilerek çıktıyı kendi keşfedebilmektedir. Veriler arasındaki örüntüyü makinenin keşfetmesi ve ilişki ağını kendisinin bulması beklenir. Boyut azaltma, kümeleme gibi çalışmalarda denetimsiz öğrenme kullanılır. Son olarak yarı denetimli öğrenme ise kullanılan verilerin içerisinde az sayıda işaretlenmiş veri olduğu ve işaretlenmemiş çok sayıda verinin tahmini ve sınıflandırılması yapılmaktadır (Gökalp, 2022).

Hem denetimli hem de denetimsiz öğrenmenin kullanıldığı bu çalışmada, makine öğrenmesi yöntemlerinin bu tür çok boyutlu ve dinamik veriler üzerindeki güçlü tahmin performansından faydalanarak, Türkiye'nin dış ticaret verileri üzerinde üç farklı tahmin modeli önerilmiştir. Önerilen modeller, dış ticaret verilerindeki anomalilerin tespiti, öznitelik seçim yöntemleri ile veri boyutunun azaltılması, veri kümelerinin yapısına uygun kümeleme yöntemleri ve çeşitli regresyon teknikleri ile elde edilen sonuçların analizini kapsamaktadır. Bu kapsamda çalışma literatüre katkı sağlamaktadır. Çalışma dört bölüme ayrılmıştır. Birinci bölümde teorik çerçeve ele alınmıştır. İkinci bölümde konu kapsamında literatür araştırması ile veri seti ve modelin tahmin yöntemi tanıtılmaktadır. Üçüncü bölümde tahmin edilen modelden elde edilen bulgular kısmı, dördüncü bölümde bulgular ışığında elde edilmiş sonuçları ile çalışmanın kısıtları ve gelecek çalışmalar için tavsiyelerden bahsedilmiştir.

## 2. Materyal ve Yöntem

Veri Seti Türkiye İstatistik Kurumu A grubu mikro veri setlerinden dış ticaret istatistikleri mikro veri setidir. Veriler TÜİK Erzurum Bölge Müdürlüğü Veri Araştırma Merkezi'nde kullanılabilir. Veriler Veri Araştırma Merkezi dışına çıkarılmadığından burada sağlanmış olan bilgisayar ile algoritmalar çalıştırılmıştır. Bütün sektörleri içeren kayıtlar incelenmiş ve çalışmanın konusu olacak şekilde Nace Rev2 26 (Bilgisayarların, elektronik ve optik ürünlerin imalatı) sektörü verileri içerisinde 2018-2022 yılları için tam sayım olacak şekilde 817500 veri işleme alınmıştır.

Bu çalışma kapsamında elimizde bulunan veri için en iyi algoritma bulunmaya çalışılmıştır. 3 farklı model ile yapılan çalışmalarda 30 farklı sonuç elde edilmiş bu

sonuçlar ışığında en iyi model seçilmiştir.

### 2.1. Literatür Taraması

Dış ticaret akışlarının modellenmesi ve öngörülmesinde makine öğrenmesi yöntemlerinin kullanımı, son yıllarda artan bir ilgi görmektedir. Bu bağlamda Nuroğlu (2014) geleneksel yerçekimi yöntemi ile yapay sinir ağları modellerini karşılaştırmıştır. Bu modellerde yapay sinir ağlarının girdi ve çıktılarının arasındaki ilişkileri öğrenme ve açıklama yeteneğinin daha iyi olduğu gözlemlenmiştir. Bayrak (2020) ise medikal cihazların dış ticareti için makine öğrenme yöntemlerini kullanmış ve DVM yöntemi ile  $R^2=0,974$  performans yakalamıştır. Nummelin ve Hanninen (2016) Birleşmiş Milletler FAO verilerini kullanarak kütük dış ticareti tahminlemesi için makine öğrenmesi yöntemleri kullanmışlardır.

Bu çalışmaya benzer olarak yurtdışında Filimnova vd. (2021) makalesinde Rusya'nın dış ticaret verilerini kullanırken kümeleme analizinin tahminleme gücüne etkisini ölçmeye çalışmıştır. Huang vd. (2020) Çin ve 5 merkez Asya ülkelerinin ihracat hacmi tahminlemesinde çok fazla faktör olduğunda klasik istatistik yöntemleri ile karar ağaçları gibi çok değişkenli regresyon analizini kullanmışlardır. Jošić ve Žmuk (2022) Hırvatistan Dış Ticaret verileri için gaussian yöntemler, lineer regresyon ve çok katmanlı perceptron kullanarak çalışmışlardır.

Gopinath vd. (2021) Amerika Birleşik Devletlerinde tarım dış ticareti için yapay sinir ağları kullanarak makine öğrenmesi modelleri kullanmıştır. Jia vd. (2019) Latin Amerikada yapılan yakıt ihracatının yönünü belirlemek için karar ağaçları, rasgele ormanlar ve Boosting ağaç yöntemini kullanmışlardır. Shen vd. (2021) ise 10 ülkenin dış ticaretini ölçmek için zaman serileri verileri ile LSTM modeli kullanmıştır.

Baxter ve Srisaeng (2018) yapay sinir ağları kullanarak Avusturalya'nın hava kargoculuğu ile yapacağı ihracatı tahmin çalışması yapmıştır. CHAN vd. (2021) çalışmasında Amerika Birleşik Devletleri ile 15 Asya ülkesi arasında yapılan kıyafet ticaretini çoklu regresyon ve yapay sinir ağları modelleri ile tahminlemeye çalışmıştır. Burada klasik çoklu regresyona göre yapay sinir ağları modelinin daha yüksek  $R^2$  değerine sahip olduğu görülmüştür.

Dumor ve Yao (2019) çalışmasında Çin'in Belt and Road inisiyatifi ile ortak ülkeleri olan Kenya, Burundi, Rwanda, Tanzanya, Uganda ve Cibuti ile yaptığı ticareti yapay sinir ağları ile tahminlemeye çalışmıştır. Çalışma kapsamında yerçekimi modeli sonuçları ile yapay sinir ağları modeli karşılaştırılmış ve yapay sinir ağları modelinin daha düşük RMSE sahip olduğu görülmüştür.

Micocci ve Rungi (2023) derin öğrenme algoritmaları olan LOGIT, LOGIT-LASSO, CART, Random Forest, BART ve BART-MIA kullanarak Fransa'nın dış ticaret verileri hesaplamaya çalışılmış ve en iyi sonuç BART-MIA modeli ile alınmıştır.

Gupta ve Kumar (2022) Hindistan'ın dış ticaretini ölçmek için Doğrusal Regresyon, DVM, Karar Ağaçları ve Rasgele Orman algoritmalarını kullanmıştır. Bu algoritmalar arasında en iyi sonucu Karar Ağacı algoritmaları vermiştir.

Eşidir (2025) Türkiye'nin kimyasal madde ithalatını ölçmek için doğrusal regresyon, rasgele orman, rasyonel kuadratik regresyon, DVM ve XGBOOST regresyon modellerini kullanmıştır. Burada XGBOOST algoritması en iyi sonuçları vermiştir.

Ay vd. (2024) Türkiye ve Türk Cumhuriyetleri (Azerbaycan, Kazakistan, Kırgızistan, Özbekistan ve Türkmenistan) arasındaki ticareti ölçmek için lineer regresyon, gauss süreç regresyonu ve çok katmanlı algılayıcılar gibi farklı makine öğrenme algoritmaları kullanmış, en iyi sonuca çok katmanlı algılayıcılar algoritması ile ulaşmıştır.

Gür ve Eşidir (2024) Türkiye'nin hurda demir çelik ithalatını ölçmek için LTSM, MLP, Random Forest, DVM, XGBOOST ve doğrusal regresyon modellerini kullanmış, LTSM modeli ile en iyi sonuçlara ulaşmıştır.

Soydal vd (2024) Türkiye ve Balkan ülkeleri üzerine çekim modeli uygulaması yapmış ve her ülke değişkeni için en iyi sonuçları veren yöntemleri çalışmasında yayınlamıştır.

## 2.2. Veri Seti

Öncelikle kullanılacak sektör için veriler SAS Enterprise uygulaması ile 5 yıllık çekilmiştir. Daha sonrasında aynı paket program aracılığıyla birleştirilmiştir. Veri işleme için Jupyter Notebook kullanılmıştır. Önce veri temizleme yapılmıştır. Veri temizleme yapılırken eksik ve tekrarlayan veriler silinmiştir. Bu işlem sonucunda 15 satır mükerrer veri ve 22 satır veri ürün kodu ve parasal değeri eksik olduğu için yanlışlık olmaması adına silinmiştir, Silinen verinin toplam veriye oranı 0,00005'dir. Daha sonra metin verileri kodlanmıştır. Bu işlemler ithalat ve ihracat verileri için ayrı ayrı yapılmış, veriler birleştirilmiştir.

5 yıllık veri seti 18 öznitelikten oluşmaktadır. Bu öznitelikler aşağıdaki şekilde açıklanmıştır

IHRITH: İhracat ve İthalat ayrımını gösterir.

YIL: Yıl bilgisini içerir.

ISTPOZ: Gümrük Tarifesi İstatistik Pozisyonu 12 dijital (GTİP)

ULKE: İhracat/İthalat yapılan ülke kodlarını içerir. (Ülke bilgisi ithalatta menşei ülke, ihracatta malın gideceği bilinen son ülkeyi ifade eder)

OLCU: GTİP'lerin ölçü kodlarını içerir.

IL\_KODU: Dış ticaret yapan firmanın merkezinin bulunduğu il kodu

ODEME\_SEKLI: Alıcı ve satıcı arasındaki ödemenin ne şekilde yapıldığını ifade eder.

DOVIZ\_KODU: Yapılan dış ticaretin para biriminin kodu

YOL: Taşıma şeklinin kodu

GCGUMRUK: Giriş çıkış gümrük kodu

SOZLESME\_KODU: Alıcı ve satıcı arasındaki ticari sözleşme türlerini ifade eder.

KONTEYNER: Malın konteynerle mi konteynersiz mi taşındığını belirten kodlardır.

BEC: Geniş ekonomik gruplar sınıflaması

NACE: Avrupa Topluluğunda Ekonomik Faaliyetlerin İstatistik Sınıflaması, Rev.2 (4 dijitaldir)

CPA: Avrupa Topluluğunda Ekonomik Faaliyete Göre

Ürünlerin İstatistik Sınıflaması, 2008

versiyonu (4 dijitaldir)

MIKTAR\_1: Ürünlerin kilogram bilgisini ifade eder. (doğalgaz, elektrik hariç)

MIKTAR\_2: Yardımcı ölçü birimini ifade eder. (adet m<sup>3</sup>, m<sup>2</sup> gibi)

DOLAR: ABD para birimi cinsinden dış ticareti ifade eder.

EURO: Avrupa Birliği para birimi cinsinden dış ticareti ifade eder.

Çalışma bütününde test veri setleri %20, %25 ve %30 olarak çalıştırılmış en iyi sonuç %20 test, %80 train veri setinde ulaşılmış bu nedenle bütün çalışmada bu şekilde çalışılmıştır.

Burada uygulanan modele göre öznitelik seçimi algoritmaları uygulanmış ve bu algoritmalar sonucunda çıkan öznitelik önemliliğine göre seçilerek model sonuçları alınmıştır. Öznitelik önemliliği için MRMR ve RFE algoritmaları uygulanmıştır.

## 2.3. Anomali Tespit

### 2.3.1. Isolation forest

Literatürde, "Anomali Tespiti" ile aynı veya benzer anlamlara sahip farklı terimler kullanılmaktadır. Bunlar arasında olay tespiti, yenilik tespiti, (nadir) olay tespiti, sapma keşfi, değişim noktası tespiti, arıza tespiti, izinsiz giriş tespiti ve kötüye kullanım tespiti yer almaktadır (DA, 2013). Esasında farklı görünen bu terimler aynı amacı anlatmak istemektedir. Veri dağılımından belirgin bir şekilde sapmış veri noktalarını tespit etmek. Anomali gücünün bir ölçüsü genellikle sapma miktarı üzerinden değerlendirilir veya stokastik olarak, bir anomali olma ihtimali anlamına gelen "anomali skoru" olarak tanımlanır.

İzolasyon ormanı algoritması ile verideki aykırı değerleri tespit etmek için rastgele oluşturulmuş izolasyon ağaçlarından oluşan bir orman oluşturulur. Her ağaç, analiz edilen veri kümesinin bir alt kümesi olan bağımsız olarak alınmış örneklerle dayanarak büyütülür. Bu ağaçlar, analiz edilen uzayı özyinelemeli olarak bölerek ayrı noktalar yapraklarında izole edilene veya derinlik sınırına ulaşılan kadar işleyen "izole edici" ağaçlar olarak adlandırılır. Aykırı değer tespiti, verilerin oluşturulan ormana gönderilmesi ve her bir veri noktası için bir anomali ölçüsünün üretilmesiyle gerçekleştirilir. Bu çalışmada Python programlama dilinde Sklearn Kütüphanesinde IsolationForest kullanılmıştır.

## 2.4. Öznitelik Seçimi

### 2.4.1. MRMR öznitelik seçim algoritması

MRMR algoritması sınıf etiketleriyle ilişkili öznitelikleri seçmeye çalışırken, seçilen öznitelikler arasındaki artıklığı, fazlalığı minimize etmeye çalışan bir literatür yöntemidir (Ding ve Peng, 2005). Öznitelik kümesini seçerken, seçilmek istenilen kümenin seçilebilecek en iyi küme olması için iki koşul vardır, birincisi minimum ortaklık, ikincisi maksimum ilişkidir. Bu çalışmada Python programlama dilinde MRMR kütüphanesinden mrmr\_classif yöntemi kullanılmıştır.

### 2.4.2. RFE öznitelik seçim algoritması

Geriye dönük öznitelik seçimi olan RFE öznitelik

seçiminde öncelikle bir model oluşturulur. Bu modelde her özneliğe bir önem puanı vermektedir. Sonraki aşamalarda en düşük önem puanındaki öznelikler silinerek yeniden hesaplanır ve istenilen sayıda öznelik sayısına ulaşılır. (Emanet vd., 2021)

Çalışma kapsamında Python programlama dilinde Sklearn Kütüphanesinde RFE kullanılmıştır.

## 2.5. Kümeleme Algoritmaları

### 2.5.1. Clarans algoritması

K-medoids, her kümenin merkezini küme içerisindeki gerçek bir gözlem noktası (medoid) olarak belirler. CLARANS ise K-Medoids'in daha gelişmiş versiyonudur ve klasik k-medoids algoritması, veri seti büyüdükçe daha maliyetlidir. CLARANS ise küresel en iyi çözüme daha yakın bir sonuç elde etmeyi amaçlayan rastgeleleştirilmiş bir arama prosedürü uygular.

CLARANS'ın temel fikri, küme merkezlerinin komşuluk uzayında rastgele araştırılması ve bu süreçte maliyeti düşüren komşuların kabul edilmesidir. Bu yönüyle CLARANS, kısmi bir rastgele optimizasyon algoritması olarak da değerlendirilebilir. (Ng ve Han, 2002)

### 2.5.2. K-Means algoritması

K-Means algoritması, gözetimsiz öğrenme (unsupervised learning) kapsamında en yaygın kullanılan merkez-temelli (centroid-based) kümeleme yöntemlerinden birisidir. Amaç, gözlemleri, her biri kendi merkezine (centroid) en yakın olacak şekilde k adet kümeye ayırmaktır. Bu merkezler, kümelerin geometrik temsilcileridir ve kümeler arası farklılıkları en üst düzeye çıkarırken, kümeler içi benzerlikleri en üst düzeye çıkarmayı hedefler. (Demiralay ve Çamurcu, 2005)

K-Means algoritmasının en önemli özelliği büyük veri setlerinde hızlı ve etkili olmasıdır. Bunun yanında çok boyutlu veri kümelerine uygulama kolaylığı nedeniyle de çalışma kapsamında seçilmiştir.

## 2.6. Regresyon Algoritmaları

### 2.6.1. XGBOOST algoritması

Çok boyutlu büyük veri analizinde etkin olarak kullanılabilen XGBOOST algoritmasının bir çok regresyon modeline göre kullanımı ve skorları daha yüksek çıkmaktadır. Model, ardışık ağaçlar inşa ederek bu hatayı minimize etmeye çalışır. Ayrıca, erken durdurma (early stopping), çapraz doğrulama (cross-validation) ve hiperparametre optimizasyonu ile yüksek doğruluk ve genelleme sağlanabilir. (Abar, 2020)

Ağaç tabanlı yöntemler arasında en yüksek skorları veren ve Gradyan Boosting temeline kurulmuş XGBOOST algoritması aşırı öğrenmeyi kontrol etmek için L1 (lasso) ve L2 (ridge) cezalarını kullanır. (Carmona vd., 2019)

### 2.6.2. LightGBM algoritması

LightGBM modeli XGBOOST ile benzer prensiplerle çalışmasıyla birlikte büyük ölçekli veri setlerinde daha hızlı modelleme ve daha düşük bellek kullanımı yapısal optimizasyon içerir. Bu optimizasyonları ağaç dallarını dikey bir şekilde değil, yatay bir şekilde dallanma yolunu gitmesi ile gerçekleştirir. Bu sayede daha düşük ram kullanımı sağlar. Ancak LightGBM regresyon modelinin zaafı aşırı öğrenmeye (overfitting) meyilli olmasıdır. Bu

nedenle model kurulduktan sonra hiperparametre optimizasyonuna çok büyük rol düşmektedir. (Ke vd., 2017)

### 2.6.3. Random forest regresyon algoritması

Random Forest Regresyonu hem sınıflandırma hem de regresyon problemlerinde kullanılan topluluk öğrenme yöntemlerinden bir tanesidir. Veri seti rastgele parçalanarak küçük parçalara ayrılır. Bu küçük parçalardan oluşmuş tahminlerin ortalaması alınarak sonuç ortaya çıkarılır. Model küçük parça sonuçları üzerine kurulu olduğu için hem aşırı öğrenmenin (overfitting) önüne geçmekte hem de veri gürültüsüne karşı dayanıklılığı artmaktadır. Zaman serilerinde de yüksek başarılı sonuçlar sağlamaktadır. (Akşehir ve Kılıç, 2019)

Veri setimiz 5 yıllık verilerden oluşmakta ve random forest algoretimeisi daha başarılı sonuçlar verdiği için random forest regresyon yöntemi de modellerimizde kullanılacaktır.

## 2.7. Hiperparametre Optimizasyonu

### 2.7.1. GridSearchCv ile hiperparametre optimizasyonu

Grid Search (Izgara Arama), hiperparametre ayarlaması için kullanılan bir yöntemdir ve hiperparametrelerin tüm olası kombinasyonlarını sistematik olarak değerlendirir. Bu yöntem, önceden belirlenmiş hiperparametre değerlerinin Kartesyen çarpımını oluşturarak her bir kombinasyon için modeli eğitir. Model performansı, çapraz doğrulama yöntemi kullanılarak ölçülür; bu sayede modelin veri kümesindeki anlamlı desenleri öğrenmesi sağlanır.

Grid Search yöntemi şu adımları içerir:

1. Her bir hiperparametre için makul bir değer kümesi tanımlanır.
2. Model, hiperparametrelerin tüm olası kombinasyonları ile eğitilir.
3. Her kombinasyonun performansı, çapraz doğrulama veya ayrı bir doğrulama seti kullanılarak değerlendirilir.
4. En iyi sonuç veren kombinasyon seçilerek nihai modelde kullanılır.

Her ne kadar Grid Search, en uygun hiperparametre kombinasyonunu bulmayı garanti etse de, yüksek hesaplama maliyeti önemli bir dezavantajdır. Hiperparametre sayısının artmasıyla birlikte kombinasyon sayısı üstel olarak çoğalır,  $O(n^k)$  karmaşıklığına sahip olduğu için geniş arama alanları için verimsiz hale getirmektedir (Ahmad vd., 2022)

### 2.7.2. RandomSearchCv ile hiperparametre optimizasyonu

Önceki bölümde belirtildiği üzere, Grid Search (Izgara Arama), model seçimi için kapsamlı bir arama yöntemidir. Bu yöntemde, hiperparametre değerlerinden oluşan bir izgara kurulur; her kombinasyon için model eğitilir ve test verileri üzerinde performans skorları hesaplanır. Bu süreç, özellikle yüksek boyutlu arama alanlarında ciddi verimsizliklere yol açabilir. Örneğin, beş hiperparametre için her biri 10 farklı değer test edilmesi durumunda 100.000 deneme gerçekleştirilir. Eğer 10 katlı çapraz doğrulama kullanılırsa, bu durum 1.000.000 model

eđitimi ve 1.000.000 tahmine denk gelir; dolayısıyla hem hesaplama g¼c¼ hem de zaman aısından önemli maliyetler doğurmaktadır.

Buna karşılık, Random Search arama alanını örnekleyerek belirli bir olasılık dağılımından hiperparametre setlerini değerlendirir. Kısaca, bu yöntem, üzerinde çalışılan model için en uygun çözüml¼ bulmak amacıyla hiperparametrelerin rastgele kombinasyonlarını kullanır. Örneđin, tüm 100.000 denemenin değerlendirilmesi yerine yalnızca 1000 rastgele seçilen hiperparametre seti incelenir. Random Search'te gerçekleştirilecek deneme sayısı, hiperparametre optimizasyon süreci başlamadan önce belirlenir; bu nedenle, n denemenin yapıldığı

durumlarda yöntem, O(n) hesaplama karmaşıklığına sahiptir. Ancak, Random Search algoritmasının önemli bir dezavantajı, önceki denemelerden elde edilen bilgileri kullanarak bir sonraki denemeyi yönlendirmemesi ve sonraki denemeyi öngörmek için stratejik bir yaklaşım benimsememesidir (Elgeldawi vd., 2021).

alışma kapsamında hem GridSearchCv hem de RandomSearchCv hiperparametre optimizasyonu algoritmaları uygulanmıştır. Her iki optimizasyon algoritması sonuçları regresyon modellerine uygulanmış ve en yüksek R<sup>2</sup> ve en düşük MSE ve RMSE sonuçlarına göre ortaya çıkan en iyi hiperparametre sonuçları aşağıda bulunan tablolarda her alt kırılım için verilmiştir.

**Tablo 1.** Modellere göre XGBOOST en iyi hiperparametreler

Model Seçimi	Öznitelik Seçimi	Kümeleme Modeli	Regresyon Seçimi	Hiperparametre
Model1	MRMR	-	XGBOOST	colsample_bytree': 0.535, 'gamma': 0.387, 'learning_rate': 0.0413, 'max_depth': 11, 'min_child_weight': 6, 'n_estimators': 878, 'reg_alpha': 0.451, 'reg_lambda': 0.113, 'subsample': 0.992
Model2	MRMR	CLARANS	XGBOOST	'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 200, 'subsample': 1.0
Model2	MRMR	K-Means	XGBOOST	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8
Model3	MRMR	CLARANS	XGBOOST	'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 200, 'subsample': 1.0
Model3	MRMR	K-Means	XGBOOST	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8
Model1	RFE	-	XGBOOST	colsample_bytree': 0.878, 'gamma': 0.234, 'learning_rate': 0.0912, 'max_depth': 9, 'min_child_weight': 2, 'n_estimators': 537, 'reg_alpha': 0.772, 'reg_lambda': 0.520, 'subsample': 0.926
Model2	RFE	CLARANS	XGBOOST	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8
Model2	RFE	K-Means	XGBOOST	'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 200, 'subsample': 1.0
Model3	RFE	CLARANS	XGBOOST	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8
Model3	RFE	K-Means	XGBOOST	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8

**Tablo 2.** Modellere göre LightGBM en iyi hiperparametreler

Model Seçimi	Öznitelik Seçimi	Kümeleme Modeli	Regresyon Seçimi	Hiperparametre
Model1	MRMR	-	LightGBM	colsample_bytree': 0.745, 'gamma': 0.667, 'learning_rate': 0.0175, 'max_depth': 11, 'min_child_weight': 6, 'n_estimators': 878, 'reg_alpha': 0.41, 'reg_lambda': 0.153, 'subsample': 0.992
Model2	MRMR	CLARANS	LightGBM	boosting_type': 'gbdt', 'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 31, 'subsample': 0.7}
Model2	MRMR	K-Means	LightGBM	{'boosting_type': 'gbdt', 'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 100, 'subsample': 0.7}
Model3	MRMR	CLARANS	LightGBM	{'boosting_type': 'gbdt', 'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 100, 'subsample': 0.7}
Model3	MRMR	K-Means	LightGBM	boosting_type': 'gbdt', 'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 100, 'subsample': 0.7
Model1	RFE	-	LightGBM	colsample_bytree': 0.836, 'gamma': 0.453, 'learning_rate': 0.0221, 'max_depth': 8, 'min_child_weight': 6, 'n_estimators': 658, 'reg_alpha': 0.141, 'reg_lambda': 0.533, 'subsample': 0.889
Model2	RFE	CLARANS	LightGBM	'boosting_type': 'gbdt', 'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 31, 'subsample': 0.7
Model2	RFE	K-Means	LightGBM	'boosting_type': 'gbdt', 'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 50, 'subsample': 0.7
Model3	RFE	CLARANS	LightGBM	boosting_type': 'dart', 'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 31, 'subsample': 0.7
Model3	RFE	K-Means	LightGBM	'boosting_type': 'gbdt', 'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 100, 'subsample': 0.7

**Tablo 3.** Modellere göre random forest en iyi hiperparametreler

Model Seçimi	Öznitelik Seçimi	Kümeleme Modeli	Regresyon Seçimi	Hiperparametre
Model1	MRMR	-	Random Forest Regressor	colsample_bytree': 0.786, 'gamma': 0.578, 'learning_rate': 0.0963, 'max_depth': 9, 'min_child_weight': 6, 'n_estimators': 864, 'reg_alpha': 0.248, 'reg_lambda': 0.312, 'subsample': 0.910
Model2	MRMR	CLARANS	Random Forest Regressor	'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 200, 'subsample': 1.0
Model2	MRMR	K-Means	Random Forest Regressor	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8
Model3	MRMR	CLARANS	Random Forest Regressor	'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 200, 'subsample': 1.0
Model3	MRMR	K-Means	Random Forest Regressor	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8
Model1	RFE	-	Random Forest Regressor	colsample_bytree': 0.965, 'gamma': 0.479, 'learning_rate': 0.0847, 'max_depth': 11, 'min_child_weight': 6, 'n_estimators': 744, 'reg_alpha': 0.146, 'reg_lambda': 0.132, 'subsample': 0.787
Model2	RFE	CLARANS	Random Forest Regressor	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8
Model2	RFE	K-Means	Random Forest Regressor	'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 200, 'subsample': 1.0
Model3	RFE	CLARANS	Random Forest Regressor	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8
Model3	RFE	K-Means	Random Forest Regressor	'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200, 'subsample': 0.8



## 2.8. Çalışmada Kullanılan Modeller

Bu çalışmada kullanılan algoritmaların seçimi, veri setinin büyüklüğü, karmaşıklığı ve yapısal çeşitliliği dikkate alınarak yapılmıştır. Dış ticaret verileri; yüksek boyutlu, öznelik bakımından heterojen ve yıl, ülke, ürün ve firma gibi çok sayıda kategorik ve sayısal değişken içeren bir yapı arz etmektedir. Bu nedenle hem anlamlı özneliklerin seçilmesi, hem de karmaşık örüntülerin çıkarılması açısından farklı aşamalarda farklı algoritmaların birlikte kullanılması uygun görülmüştür.

Regresyon aşamasında XGBoost, LightGBM ve Random Forest Regressor algoritmaları tercih edilmiştir. Bu üç algoritma, ağaç tabanlı yöntemlerin farklı varyasyonlarını temsil etmekte olup, yüksek doğruluk ve genelleme gücü ile literatürde başarıları kanıtlanmış modellerdir. Özellikle: XGBoost, gradyan artırma temelli bir yöntem olarak karmaşık örüntüleri modelleme başarısı ile öne çıkmaktadır.

LightGBM, benzer prensiplere dayanmakla birlikte daha düşük bellek kullanımı ve hız avantajı nedeniyle yüksek hacimli verilerde tercih edilmiştir.

Random Forest, çok sayıda karar ağacının ortalaması ile overfitting'e karşı dayanıklı sonuçlar üretmektedir. Özellikle zaman serilerine benzer dönemsel dış ticaret verilerinde başarılı performansı sayesinde bu çalışmaya dahil edilmiştir.

Ayrıca, özellik seçimi (MRMR ve RFE) ile modelin boyutu azaltılarak hesaplama yükü minimize edilmiş, yalnızca anlamlı değişkenlerin kullanılması sağlanmıştır.

Kümeleme algoritmaları olarak ise K-Means ve CLARANS kullanılmıştır. K-Means'in hızlı ve etkili olması ile CLARANS'ın küresel minimumlara daha yakın çözümler sunması bu seçimi desteklemiştir.

Model eğitimi süreci, hiperparametre optimizasyonu da içerecek şekilde, GridSearchCV ve RandomSearchCV yöntemleri ile sistematik olarak gerçekleştirilmiştir. Eğitim ve test veri setleri 80:20 oranında ayrılmış, farklı model bileşenlerinin etkisini gözlemleyebilmek adına üç temel model mimarisi oluşturulmuştur. Böylece her algoritmanın, farklı ön işleme aşamalarıyla nasıl performans sergilediği detaylı biçimde test edilmiştir.

Çalışma boyunca 3 farklı model oluşturulmuş ve sonuçlar bu modellerden çıkan R<sup>2</sup>, MSE ve RMSE değerlerine göre karşılaştırılmıştır.

### 2.8.1. Model 1: Anomali tespiti, öznelik seçimi ve regresyon

Bu modelde, önce Isolation Forest yöntemi kullanılarak anomali tespiti yapılmıştır. Anomaliler çıkarıldıktan sonra öznelik seçimi için Minimum Redundancy Maximum Relevance (MRMR) ve Recursive Feature Elimination (RFE) yöntemleri kullanılmıştır. Son olarak, tahmin için XGBoost, LightGBM ve Random Forest Regressor yöntemleri uygulanmıştır. Modelin detaylı Şekil 5'de de gösterildiği üzere aşağıdaki gibidir.

- Anomali Tespiti: Isolation Forest
- Öznelik Seçimi: MRMR ve RFE
- Regresyon: XGBoost, LightGBM, Random Forest Regressor.

Model 1		
Veri Alınması	Veri Önleme	Anomali Tespiti
TÜİK VAM ile 2018-2022 yılları arası 26 Bilgisayarların, elektronik ve optik ürünlerin imalatı sektörü verileri 817500 veri	*Veri Temizliği *Kayıp Veri analizi *Veri Sınıflamaları, Kodlamaları *Uluslararası Sınıflamalara Göre Faaliyet Kodlaması yapılması	Isolation Forest Algoritması kullanılarak anomali tespiti yapılması
Öznelik Seçimi		Regresyon Modelleri
*MRMR (Minimum Redundancy Maximum Relevance) *RFE (Recursive Feature Elimination)		*XGBOOST *LightGBM *Random Forest Regressor

Şekil 5. Model 1 akış diyagramı.

### 2.8.2. Model 2: Kümeleme, öznelik seçimi ve regresyon

Bu modelde, veri kümeleme işlemi ile analiz edilmiştir. K-means ve CLARANS algoritmaları ile kümeler oluşturulmuş, ardından öznelik seçimi için MRMR ve RFE yöntemleri kullanılmıştır. Kümelenmiş veri setleri

üzerinden regresyon uygulanmıştır. Akış diyagramı Şekil 6'de gösterilmiştir.

- Kümeleme: K-means, CLARANS
- Öznelik Seçimi: MRMR ve RFE
- Regresyon: XGBoost, LightGBM, Random Forest Regressor.

Model 2		
Veri Alınması	Veri Önışleme	Öznitelik Seçimi
TÜİK VAM ile 2018-2022 yılları arası 26 Bilgisayarların, elektronik ve optik ürünlerin imalatı sektörü verileri 817500 veri	*Veri Temizliđi *Kayıp Veri analizi *Veri Sınıflamaları, Kodlamaları *Uluslararası Sınıflamalara Göre Faaliyet Kodlaması yapılması	*MRMR (Minimum Redundancy Maximum Relevance) *RFE (Recursive Feature Elimination)
Kümeleme Algoritması		Regresyon Modelleri
*CLARANS Kümeleme Algoritması *K-Means Kümeleme Algoritması		*XGBOOST *LightGBM *Random Forest Regressor

Şekil 6. Model 2 akış diyagramı.

### 2.8.3. Model 3: Anomali tespiti, kümeleme, öznitelik seçimi ve regresyon

Bu modelde ise, önce anomali tespiti yapılmış, ardından kümeleme uygulanmış ve son olarak öznitelik seçimi ile regresyon uygulanmıştır. Modelin detaylı Şekil 7’te de gösterildiđi üzere aşağıdaki gibidir.

- Anomali Tespiti: Isolation Forest
- Öznitelik Seçimi: MRMR ve RFE
- Kümeleme: K-means, CLARANS
- Regresyon: XGBoost, LightGBM, Random Forest Regressor.

Model 3			
Veri Alınması	Veri Önışleme	Anomali Tespiti	Öznitelik Seçimi
TÜİK VAM ile 2018-2022 yılları arası 26 Bilgisayarların, elektronik ve optik ürünlerin imalatı sektörü verileri 817500 veri	*Veri Temizliđi *Kayıp Veri analizi *Veri Sınıflamaları, Kodlamaları *Uluslararası Sınıflamalara Göre Faaliyet Kodlaması yapılması	Isolation Forest Algoritması kullanılarak anomali tespiti yapılması	*MRMR (Minimum Redundancy Maximum Relevance) *RFE (Recursive Feature Elimination)
Kümeleme Algoritması		Regresyon Modelleri	
*CLARANS Kümeleme Algoritması *K-Means Kümeleme Algoritması		*XGBOOST *LightGBM *Random Forest Regressor	

Şekil 7. Model 3 Akış Diyagramı

### 2.9. Çalışmada Kullanılan Metrikler

Çalışmada kullanılan modeller  $R^2$ , MSE ve RMSE metriklerine göre karşılaştırılmıştır.

#### 2.9.1. Determinasyon katsayısı ( $R^2$ )

Yağmur (2022) Makalesinde  $R^2$  tanımını “Korelasyon katsayısının karesi olan determinasyon katsayısı, bağımlı deđişkendeki deđişkenliđin ne kadarının bağımsız deđişkenlerdeki deđişkenlikle açıklanabildiđini ifade eder. İki deđişken arasındaki doğrusal ilişkinin gücünü

belirleyen determinasyon katsayısı 0-1 aralığında deđer almakta olup ilişkinin gücü 1’e yaklaştıkça artmaktadır.  $R^2$ ’nin 1’e eşit olması regresyon tahminlerinin gerçekleşen deđerlere tam olarak uyduđu anlamına gelmektedir.” şeklinde yapmaktadır (eşitlik 1).

$$R^2 = \frac{\sum_{i=1}^x (y_i - \bar{y})^2 - \sum_{i=1}^x e_i^2}{\sum_{i=1}^x (y_i - \bar{y})^2} \quad (1)$$

$Y_i$  :  $Y$  gerçek değerlerini göstermektedir.

$\bar{Y}$  :  $Y_i$  değerlerinin ortalama değerlerini göstermektedir.

$e_i$  :  $Y$  gerçek değerleri ile tahmin değerleri arasındaki farkı göstermektedir

### 2.9.2. Mean squared error MSE (karesel ortalama hata)

Modeller sonucu algoritmanın elde ettiği değerler ile gerçek değerler arasındaki farkın karesinin ortalaması MSE değerini vermektedir. Burada yapılan tahminin gerçek değerden ne kadar saptığı elde edilmektedir (eşitlik 2).

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (2)$$

$Y_i$  :  $Y$  gerçek değerlerini göstermektedir.

$\bar{Y}$  :  $Y_i$  değerlerinin ortalama değerlerini göstermektedir.

$e_i$  :  $Y$  gerçek değerleri ile tahmin değerleri arasındaki farkı göstermektedir.

### 2.9.3. Root mean squared error rmse (karesel ortalama hataların kökü)

Burada kullanılan metrik ise elde edilen MSE değerinin karekökünü almaktadır. MSE değerinde karesel olarak hesaplanan için gerçek veriden sapma değeri RMSE ile hesaplanmaktadır (eşitlik 3).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (3)$$

## 3. Bulgular ve Tartışma

Bu makale Türkiye'nin dış ticaret verileri kullanılarak bir model ortaya çıkarmak amacıyla yazılmıştır. Yapılan çalışmada veriler A grubu mikro veriler olduğu için kurum dışına çıkartılamamış, TÜİK VAM bilgisayarı üzerinde kullanılmıştır. Burada kullanılan bilgisayarın CPU ve ram özellikleri yeterli olmadığı için bütün veriler kullanılamamış, 2018-2022 yılları arasında NACE 26 "Bilgisayarların, elektronik ve optik ürünlerin imalatı" faaliyetinde bulunulan 817500 veri kullanılmıştır.

Çalışmamızda Türkiye'nin bilgisayar, elektronik ve optik ürünler dış ticaretine ilişkin verileri kullanarak üç farklı makine öğrenmesi modeli oluşturulmuş ve performansları değerlendirilmiştir. Model 1'de Anomali Tespiti (Isolation Forest), Özellik Seçimi (MRMR, RFE) ve Regresyon (XGBoost, LightGBM, Random Forest Regressor) yöntemleri uygulanmıştır. Model 2'de Kümeleme (K-Means, CLARANS), Özellik Seçimi (MRMR, RFE) ve Regresyon (XGBoost, LightGBM, Random Forest Regressor) süreçleri gerçekleştirilmiştir. Model 3'te ise Anomali Tespiti, Kümeleme, Özellik Seçimi ve Regresyon yöntemleri entegre edilerek daha kapsamlı bir

tahminleme süreci yürütülmüştür.

Öncelikle, veri seti üzerinde gerçekleştirilen izolasyon ormanı yöntemi ile anomali tespiti sonucunda belirlenen aykırı gözlemler çıkarılmış ve bu işlem veri kalitesini artırmıştır. Anomali tespit işlemi sonrası, MRMR ve RFE yöntemleri ile en etkili değişkenler seçilmiş ve modelin hesaplama yükü azaltılmıştır. Modelin tahmin başarısını artıran ve aşağıda detaylıca açıklanan özneliklerin yıl, ülke, il, NACE Rev2 dördlüsü, CPA 2.1 dördlüsü, ürün miktarı, ithalat-ihracat ayrımı ve gümrük sınıflamaları olduğu görülmüştür.

Yıl: Makroekonomik faktörler (resesyonlar, pandemi etkisi, döviz kuru dalgalanmaları vb.) ve politika değişiklikleri (gümrük vergileri, teşvikler vb.) ile doğrudan ilgilidir, ayrıca yıllara göre ürün fiyatı dalgalanmaları olacağından yıl özneliğinin modelde bulunması hayati önem taşımaktadır.

Ülke: Bazı ülkelerle ticaret yapısı teknoloji ithalatı ağırlıklı iken, bazılarıyla hammadde ihracatı gibi farklılıklar gösterebilmektedir. Örneğin Avrupa Birliği ülkelerine yapılan ihracat, katma değeri yüksek ürünler açısından farklı bir dinamik sunar. Bu farklılıkların modelin tahmin doğruluğunu artırmada önemli rol oynayacağı düşünülmektedir.

İl: İller ve bölgeler arası üretim kapasitesi, teşvik politikaları ve liman erişimi gibi faktörler dış ticaret hacmini ve ticaret yapılan ürün cinsini doğrudan etkilemektedir. Bu nedenle il bazlı ayırımın kümeleme ve regresyon aşamalarında belirleyici rol oynadığı düşünülmektedir.

Nace Rev2: Ekonomik faaliyet kodu olan NACE bazında 26 sektörü seçilmiştir. Ancak 26 sektöründe 10 farklı dördlü bulunmaktadır. Bu ayırım çalışılan sektör içinde farklı faaliyet alanları için tahmin yapılabilmesini sağlamaktadır.

CPA 2.1: Ürün sınıflaması olan CPA bazında da aynı NACE'de olduğu gibi 26 sektörü alınmaktadır. Burada 116 farklı altılı bulunmakta ve 116 farklı ürün için tahminleme yapılabilmektedir.

Ürün Miktarı: Birim fiyat değişimlerinin izlenmesi, dış ticaret karlılığı ve verimliliği gibi önemli göstergelerin hesaplanmasını sağlar.

İthalat-İhracat Ayrımı: Model için temel belirleyici unsurdur. Bu ayırım ile hem Türkiye'nin ithalat hem de ihracat değeri tahminlenecektir. 26 sektörü için dış ticaret açığı ya da fazlası tahmini bu değişken ile kontrol edilebilecektir.

Gümrük Sınıflamaları: Burada kullanılan sınıflama CPA sınıflamasına benzer şekilde ürün sınıflamasıdır. En net farkı ise sekizli sınıflama olduğundan daha alt kırılımda ürünler için hesaplama yapılabilme yeteneğini modele kazandırmasıdır.

Kümeleme aşamasında, K-Means ve CLARANS yöntemleri kullanılarak veri gruplara ayrılmıştır. K-Means yönteminde en uygun küme sayısını belirlemek için Dirsek Yöntemi ve Silhouette Skoru kullanılmıştır. CLARANS yöntemi ise büyük veri kümelerinde daha esnek bir kümeleme sağladığı için tercih edilmiştir. Kümeleme

sonuçları, farklı dış ticaret eğilimleri gösteren grupların belirlenmesine yardımcı olmuştur.

Regresyon modellerinin performanslarını karşılaştırmak amacıyla R-Kare ( $R^2$ ), Karese Ortalama Hata (MSE) ve Kök Ortalama Kare Hata (RMSE) gibi değerlendirme metrikleri kullanılmıştır. Yapılan analizler sonucunda, Model 2'ün (Kümeleme + Özellik Seçimi + Regresyon) en yüksek doğruluk oranına sahip olduğu görülmüştür. Özellikle Random Forest algoritmasının diğer regresyon yöntemlerine kıyasla daha düşük hata oranlarına ve daha yüksek determinasyon katsayısına sahip olduğu belirlenmiştir. Bununla birlikte, kümeleme yöntemlerinin tahmin başarısını önemli ölçüde artırdığı ve belirli kümelerde regresyon modellerinin daha başarılı tahminler ürettiği gözlemlenmiştir.

Ayrıca, ithalat ve ihracat verilerinin yıllık bazda tahmin edilmesi sonucunda, Türkiye'nin bilgisayar, elektronik ve optik ürünler sektöründe belirli yıllarda ihracatın arttığı, ancak küresel ekonomik krizler ve tedarik zinciri problemleri gibi dış faktörlerin dış ticaret üzerinde dalgalanmalara yol açtığı belirlenmiştir. Özellikle 2020 yılında COVID-19 pandemisinin ticaret hacmi üzerinde olumsuz etkiler yarattığı, ancak sonraki yıllarda ihracatta toparlanmalar yaşandığı görülmüştür. Makine öğrenmesi modelleri ile yapılan tahminler, gelecekteki dış ticaret eğilimlerini anlamada önemli bilgiler sağlamaktadır.

Her bir regresyon algoritması için Model1 3 fold ve 100 deneme için çalıştırılmış ve sonuç olarak optimum sonuçlar alınmıştır. Model2 ve Model3 için ise her regresyon modeli ayrı ayrı kümeleme algoritmaları için 3 fold 34 deneme için çalıştırılmış, çıkan sonuçlar kümelere göre ayrı ayrı alınmıştır. Kümelere göre alınan sonuçlar küme ağırlıkları kullanılarak ağırlıklı ortalama yöntemine göre hesaplanarak  $R^2$ , MSE ve RMSE sonuçları

hesaplanmıştır.

Elde edilen sonuçlar IBM SPSS versiyon 20 paket program (Statistical Package for Social Sciences v.21, IBM, Chicago, IL) paket programı kullanılarak gruplar arası fark testine tabi tutulmuştur. Üç grup için Kruskal Wallis ve İkili gruplar için Mann-Whitney U fark testi uygulanmış, bütün P değerler 0,05 değerinden büyük olduğu için gruplar arası istatistiki olarak anlamlı fark olmadığı görülmüştür. Gruplar arası anlamlı bir fark olmadığından en yüksek  $R^2$  değeri kullanılarak en yüksek tahmin değeri bulunulmak istenilmiştir.

Kullanılan verileri göre en yüksek  $R^2$  0,8322 ile MRMR öznitelik seçimi, Clarans kümelemesi ve Random Forest Regressor kullanılan ve makalede model 2 olarak adlandırılan modelde elde edilmiştir. Aynı modelde MSE değeri 21683120914,9233 ve RMSE değeri 144269,018 çıkmıştır.

En düşük MSE değeri 26317190,429, RMSE değeri 4627,3104 ile Model 3 modellemesinde olmuştur. Bu model 3 uygulamasında öncelikle Isolation Forest ile anomali tespiti yapılmış, K-Means ile kümeleme yapılmış, MRMR yöntemi ile öznitelik seçilmiş ve LightGBM Regresyon algoritması ile model tamamlanmıştır. Düşük hata değeri olsa dahi bu modelin  $R^2$  değeri 0,7857'dir.

Çıkan sonuçlar karşılaştırıldığında Model3'te daha düşük hatalar olmasına rağmen Model2'nin daha yüksek  $R^2$ 'ye sahip olması, izolasyon ormanı ve kümeleme algoritması kullanımının varyansı azalttığı ancak determinasyon katsayısını sınırladığı şeklinde yorumlanabilmektedir.

Sonuç olarak, bu çalışma, Türkiye'nin bilgisayar, elektronik ve optik ürünler dış ticaretinin makine öğrenmesi ile tahmin edilmesinde farklı yöntemlerin karşılaştırmalı analizini sunarak, en başarılı model kombinasyonlarını ortaya koymuştur.

**Tablo 4.** Modellere Göre XGBOOST Algoritma Sonuçları

Model Seçimi	Öznitelik Seçimi	Kümeleme Modeli	Regresyon Seçimi	$R^2$	MSE	RMSE
Model1	MRMR	-	XGBOOST	0,7861	227.725.704,3	15.090,6
Model2	MRMR	CLARANS	XGBOOST	0,6704	74.339.256.508,0	242.739,3
Model2	MRMR	K-Means	XGBOOST	0,7589	51.721.079.447,0	162.545,1
Model3	MRMR	CLARANS	XGBOOST	0,7605	198.861.827,3	13.876,3
Model3	MRMR	K-Means	XGBOOST	0,7436	210.449.067,8	14.478,6
Model1	RFE	-	XGBOOST	0,7925	166.904.775,1	12.919,2
Model2	RFE	CLARANS	XGBOOST	0,7357	166.550.000.000,0	300.934,4
Model2	RFE	K-Means	XGBOOST	0,7602	60.180.218.688,0	165.537,5
Model3	RFE	CLARANS	XGBOOST	0,7417	130.135.895,9	10.371,7
Model3	RFE	K-Means	XGBOOST	0,7454	209.001.603,9	14.431,0

**Tablo 5.** Modellere Göre LightGBM Algoritma Sonuçları

Model Seçimi	Öznitelik Seçimi	Kümeleme Modeli	Regresyon Seçimi	R <sup>2</sup>	MSE	RMSE
Model1	MRMR	-	LightGBM	0,7925	179.618.006,5	13.402,2
Model2	MRMR	CLARANS	LightGBM	0,6401	88.900.885.650,0	258.405,1
Model2	MRMR	K-Means	LightGBM	0,7642	40.932.954.877,0	155.975,1
Model3	MRMR	CLARANS	LightGBM	0,7927	171.511.277,3	12.894,9
Model3	MRMR	K-Means	LightGBM	0,7857	26.317.190,4	4.627,3
Model1	RFE	-	LightGBM	0,7904	171.185.487,4	13.083,8
Model2	RFE	CLARANS	LightGBM	0,7278	180.678.000.000,0	303.949,6
Model2	RFE	K-Means	LightGBM	0,7804	41.065.577.372,0	152.402,4
Model3	RFE	CLARANS	LightGBM	0,7832	112.450.066,9	9.705,3
Model3	RFE	K-Means	LightGBM	0,7894	172.676.279,9	13.116,0

**Tablo 6.** Modellere Göre Random Forest Algoritma Sonuçları

Model Seçimi	Öznitelik Seçimi	Kümeleme Modeli	Regresyon Seçimi	R <sup>2</sup>	MSE	RMSE
Model1	MRMR	-	Random Forest Regressor	0,7593	192.266.701,4	13.866,0
Model2	MRMR	CLARANS	Random Forest Regressor	0,8322	21.683.120.915,0	144.269,0
Model2	MRMR	K-Means	Random Forest Regressor	0,8142	16.350.213.943,0	123.714,6
Model3	MRMR	CLARANS	Random Forest Regressor	0,7618	197.867.193,7	13.838,8
Model3	MRMR	K-Means	Random Forest Regressor	0,7522	203.010.840,9	14.221,8
Model1	RFE	-	Random Forest Regressor	0,8018	189.454.067,4	13.764,2
Model2	RFE	CLARANS	Random Forest Regressor	0,8225	21.683.120.915,0	144.269,0
Model2	RFE	K-Means	Random Forest Regressor	0,8145	16.572.497.458,0	124.399,2
Model3	RFE	CLARANS	Random Forest Regressor	0,7503	127.765.522,2	10.340,8
Model3	RFE	K-Means	Random Forest Regressor	0,7548	201.162.104,9	14.155,7

#### 4. Sonuç

Bu çalışmada, 2018-2022 yılları arası Türkiye Dış Ticaret verileri kullanılarak Türkiye için Nace Rev2 26 (Bilgisayarların, elektronik ve optik ürünlerin imalatı) sektörü ihracat ve ithalat verileri makine öğrenmesi teknikleriyle tahmin edilmesine yönelik bir analiz gerçekleştirilmiştir. Anomali tespiti, değişken seçimi, kümeleme ve regresyon yöntemlerinin entegre edilmesiyle üç farklı model geliştirilmiş ve karşılaştırmalı performans analizi yapılmıştır. Elde edilen bulgular, dış ticaret tahminlerinde makine öğrenmesi yöntemlerinin etkinliğini ortaya koymuş ve en başarılı model kombinasyonlarını belirlemeye çalışılmıştır.

Tablo 4 'de XGBOOST regresyon modeli kullanılarak oluşturulmuş modeller görülmektedir. Burada en yüksek R<sup>2</sup> 0,7925 ile model 1 RFE öznitelik seçimi ile olmuştur. Tablo 5'de ise LightGBM regresyon modeli için sonuçlar listelenmiştir. Burada en yüksek R<sup>2</sup> 0,7925 ile model 1 MRMR özniteliği seçeneğindedir.

Kullanılan metriklere göre en iyi R<sup>2</sup> değeri veren sonuç

Tablo 6'da görüldüğü üzere model 2 (Clarans + MRMR + Random Forest) modeli ile elde edilmiştir. İleride yapılacak çalışmalarda sonuçlar diğer NACE ikililerinde de test edilerek dış ticaret verileri için en uygun model tespit edilebilecek ve ekonomik anlamda ülkenin ihtiyaç duyduğu projeksiyon rakamları elde edilebilecektir. Tablo 6 'da görüldüğü üzere en yüksek R<sup>2</sup> en düşük RMSE eşleşmesinde ise en iyi sonuç model 1 (Isotaliton Forest+ RFE+Random Forest) ile elde edilmiştir. Ulaşılmaya çalışılan verilerde hata oranında düşüklükte istenildiği için bu modelin kullanımının da uygun olacağı düşünülmüştür.

Çalışmanın bulguları, mevcut literatürle karşılaştırıldığında birden fazla açıdan özgün katkılar sunmaktadır. Mevcut çalışmalarda çoğunlukla dış ticaret tahminlerinde klasik regresyon ya da yapay sinir ağı tabanlı tekil modellemeler kullanılmışken, bu çalışmada çok aşamalı ve modüler bir makine öğrenmesi mimarisi önerilmiştir. Özellikle, anomali tespiti, kümeleme ve öznitelik seçimi gibi işlemlerin tahmin başarısına etkisi

detaylı biçimde analiz edilmiştir.

En yüksek  $R^2$  değeri Model 2'de (CLARANS + MRMR + Random Forest) elde edilmiştir ( $R^2 = 0,8322$ ). Bu durum, aşağıdaki nedenlerle açıklanabilir:

CLARANS algoritması, geleneksel K-Means'e kıyasla daha esnek ve küresel optimuma yakın kümeler üretmiş; bu da alt gruplar üzerinde uygulanan regresyon modellerinin daha homojen ve doğru tahminler yapmasına imkan tanımıştır.

Random Forest algoritması, ensemble yapısı sayesinde örüntü farklılıklarını etkili şekilde modellemiş, veri setindeki bölgesel, yıllık ve ürün bazlı farklılıkları yakalamada avantaj sağlamıştır.

MRMR öznelik seçimi, gereksiz ve korelasyonlu değişkenleri filtreleyerek sadece tahmini güçlendiren değişkenleri modele dahil etmiştir. Bu durum, özellikle Random Forest gibi çok sayıda karar ağacı kullanan modellerde model karmaşıklığını azaltarak performansı artırmıştır.

Model 3 ise, düşük MSE ve RMSE değerlerine ulaşmasına rağmen (MSE=26.317.190, RMSE=4.627,3),  $R^2$  değerinin görece düşük kalması ( $R^2=0,7857$ ) dikkat çekicidir. Bu durum, modelin bazı veri noktalarını oldukça isabetli tahmin ettiğini (düşük hata), ancak veri setinin tamamı üzerindeki varyansı yeterince iyi açıklayamadığını göstermektedir. Bunun başlıca nedeni:

Anomali tespit ve kümeleme sonrası, modelin daha "sadeleşmiş" bir veri yapısı üzerinden eğitilmesi ve veri çeşitliliğinin kısmen azaltılmasıdır. Bu da modelin varyansı daha dar bir aralıkta öğrenmesine neden olmuş olabilir.

Ayrıca LightGBM modelinin aşırı öğrenmeye meyilli yapısı, düşük hata verirken genelleme kabiliyetini azaltmış olabilir.

Bu bulgular, yalnızca hata oranlarının değil, determinasyon katsayısının da tahmin kalitesinin çok boyutlu bir ölçüsü olarak değerlendirilmesi gerektiğini ortaya koymaktadır.

Elde edilen sonuçlar, politika yapımcılar ve sektör temsilcileri için önemli karar destek mekanizmaları sunmaktadır. Veri kaynağı olan TÜİK aynı zamanda Türkiye'de resmi istatistik programı kapsamında veri toplayan, yayınlayan ve dağıtımını yapan kamu kurumudur. Çalışma sonucunda ortaya çıkan projeksiyonlar resmi istatistik programı kapsamında yayınlanan dış ticaret istatistikleri için katkı sağlayacaktır. Ayrıca çalışma neticesinde çıkan sonuçlar bütün sektörler yayılarak bütün sektörler için dış ticaret tahminleri oluşturabilecektir. Bununla birlikte Ticaret Bakanlığı sitesinde bu tür istatistikler için TÜİK bağlantısı kullanılmaktadır. Yapılacak projeksiyonlar ile ürün ve faaliyet bazlı dış ticaret teşviği, vergi iade politikaları ve dış ticaret açığına neden olan eksik parçalar tahmin edilebilecektir.

Makine öğrenmesi tabanlı tahmin modelleri, ihracat ve ithalat politikalarının planlanmasında, kaynak tahsisinde ve sektörel stratejilerin geliştirilmesinde kullanılabilir. Özellikle hızlı değişen küresel ticaret dinamiklerine uyum

sağlamak adına, tahminleme süreçlerinin sürekli güncellenmesi ve yeni veri kaynaklarıyla desteklenmesi önerilmektedir.

#### 4.1 Çalışmanın Kısıtları ve Önerileri

Gelecek çalışmalarda aynı modellerin Türkiye'nin en çok dış ticaretine konu olan hazır giyim ve tekstil, otomotiv ve gıda sektörlerinde uygulanmasını yapılması tavsiye edilmektedir. Bu çalışmada en büyük kısıt kullanılan verilerin TÜİK Veri Araştırma Merkezi dışına çıkarılamaması ve bu nedenle yüksek işlemcili bilgisayarlar kullanılmaması olmuştur. Sektör değişiklikleri ile farklı yöntemler de denenerek bu kısıt altında daha iyi sonuçlara ulaşmak mümkün olacaktır.

Sonuç olarak bu çalışma Türkiye'nin bilgisayar, elektronik ve optik ürünler dış ticaret tahmininde makine öğrenmesi yöntemlerinin kullanımına ilişkin kapsamlı bir analiz sunmaktadır. Anomali tespiti, kümeleme, değişken seçimi ve regresyon aşamalarının birlikte ele alınması, tahmin performansını artırmada aldığı rol bu çalışma ile herüstik olarak incelenmiştir. Gelecekteki çalışmaların, daha geniş veri setleri ve farklı makine öğrenmesi algoritmalarını içerecek şekilde genişletilmesi, dış ticaret tahminlerinde daha yüksek doğruluk oranlarına ulaşılmasını sağlayacağı düşünülmektedir.

#### Katkı Oranı Beyanı

Yazarların katkı yüzdeleri aşağıda verilmiştir. Yazarlar makaleyi incelemiş ve onaylamıştır.

	T.H.	T.A.
K	60	40
T	60	40
Y	40	60
VTI	60	40
VAY	60	40
KT	60	40
YZ	60	40
KI	60	40
GR	60	40
PY	50	50

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= Yazım, KI= kritik inceleme, GR= gönderim ve revizyon, PY= proje yönetimi

#### Çalışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

#### Etik Onay Beyanı

Bu araştırmada hayvanlar ve insanlar üzerinde herhangi bir çalışma yapılmadığı için etik kurul onayı alınmamıştır.

#### Teşekkür ve Bilgilendirme

Bu çalışma Doç. Dr. Tolga AYDIN danışmanlığında Tolga HASTAOĞLU'nun hazırladığı "Türkiye'nin Dış Ticaret Verilerinin Makine Öğrenmesi Yöntemleri ile Analizi ve

Tahmini" isimli yayınlanmamış yüksek lisans tezinden türetilmiştir. Bu makalede elde edilen bulgu ve değerlendirmeler yazarın kişisel görüşleri olup, verilerin temin edildiği Türkiye İstatistik Kurumunu hiçbir şekilde bağlamamaktadır.

## Kaynaklar

- Abar H. 2020. Xgboost ve Mars yöntemleriyle altın fiyatlarının kestirimi. *Ekev Akad Derg*, 2020(83): 427-446.
- Ahmad G N, Fatima H, Ullah S, Saidi A S. 2022. Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV. *IEEE Access*, 10: 80151-80173.
- Akşehir Z D, Kılıç E. 2019. Makine öğrenmesi teknikleri ile banka hisse senetlerinin fiyat tahmini. *Türkiye Bilisim Vakfı Bilgisayar Bilim Müh Derg*, 12(2):30-39.
- Aktan E. 2018. Büyük veri: Uygulama alanları, analitiği ve güvenlik boyutu. *Bilgi Yönetimi*, 1(1): 1-22.
- Anderson J E. 2011. The gravity model. *Annu Rev Econ*, 3(1):133-160.
- Ay A, Soydal H, Ay M. 2024. Çekim modeli çerçevesinde ticaret tahmininde makine öğrenmesi yöntemlerinin performans karşılaştırması: Türkiye ve Türk Cumhuriyetleri örneği. *Yönetim Ekon Derg*, 31(3): 439-459.
- Aytekin İ. 2013. Küreselleşme ve ekonomik küreselleşme. *Bitlis Eren Üniv Sos Bilim Derg*, 1(2): 123-134.
- Baxter G, Srisaeng P. 2018. The use of an artificial neural network to predict Australia's export air cargo demand. *Int J Traffic Transp Eng*, 8(1): 15-30.
- Bayrak T. 2020. A machine-learning-based model for forecasting medical device foreign trade. *Eskisehir Tech Univ J Sci Technol A-Appl Sci Eng*, 21(4): 477-485.
- Carmona P, Climent F, Momparler A. 2019. Predicting failure in the US banking sector: An extreme gradient boosting approach. *Int Rev Econ Finance*, 61: 304-323.
- Chan E M, Ho D C, Tsang C W. 2021. Estimating United States-Asia clothing trade: Multiple regression vs. artificial neural networks. *J Asian Finance Econ Bus*, 8(7): 403-411.
- Circlaeys S, Kanitkar C, Kumazawa D. 2017. Bilateral trade flow prediction. Unpublished manuscript. Available at: <http://cs229.stanford.edu/proj2017/final-reports/5240224.pdf> (accessed date: September 28, 2024).
- Da M. 2014. Outlier detection for temporal data. Berlin, Germany, pp: 45-65.
- Demiralay M, Çamurcu A Y. 2005. Cure, agnes ve k-means algoritmalarındaki kümeleme yeteneklerinin karşılaştırılması. *İstanbul Tic Univ Fen Bilim Derg*, 4(8): 1-18.
- Ding C, Peng H. 2005. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol*, 3(2): 185-205.
- Dumor K, Yao L. 2019. Estimating China's trade with its partner countries within the Belt and Road Initiative using neural network analysis. *Sustainability*, 11(5): 1449.
- Elgeldawi E, Sayed A, Galal A R, Zaki A M. 2021. Hyperparameter tuning for machine learning algorithms used for Arabic sentiment analysis. *Informatics*, 8(4): 79.
- Emanet S, Baydoğmuş G K, Demir Ö. 2021. Öznitelik seçme yöntemlerinin makine öğrenmesi tabanlı saldırı tespit sistemi performansına etkileri. *Dicle Univ Müh Fak Müh Derg*, 12(5): 743-755.
- Eşidir K A. 2025. Türkiye'nin kimyasal madde ithalatının gelecek tahmini: Makine öğrenmesi ve topluluk öğrenme yöntemleri performans analizi. *Fırat Univ J Soc Sci*, 35(1): 261-278.
- Filimonova M, Kislyakov A, Tikhonyuk N. 2021. Structural and dynamic modelling of the regions' foreign trade profile based on graph cluster analysis. *Strategica: Shaping the future of business and economy*, Bucharest, Romania, pp: 34-49.
- Gopinath M, et al. 2021. International agricultural trade forecasting using machine learning. *Data Policy*, 3:e1.
- Gökalp Ö M. 2022. Makine öğrenmesi. *Gazi Univ, Gazi Bilisim Enstitüsü, Adli Bilisim Bölümü, Ankara, Türkiye*, pp: 1-14.
- Gupta V, Kumar E. 2022. Evaluation of machine learning models for predicting exports of India, *Proc 7th Int Conf Comput Eng Technol (ICCET)*, Goa, February 2022, India. IET, pp: 47-52.
- Gür Y E, Eşidir K A. 2024. Türkiye hurda demir çelik ithalatının gelecek değerlerinin derin öğrenme, makine öğrenmesi ve topluluk öğrenme yöntemleri ile öngörülmesi. *Alanya Akad Bakış*, 8(3): 885-908.
- Huang R, et al. 2020. Forecasting trade potential between China and the five Central Asian countries: under the background of Belt and Road Initiative. *Comput Econ*, 55: 1233-1247.
- Jia H, Adland R, Wang Y. 2019. Latin American oil export destination choice: A machine learning approach, *Proc IEEE Int Conf Ind Eng Eng Manag (IEEM)*, Macao, December 2019, pp: 345-348.
- Jošić H, Žmuk B. 2022. A machine learning approach to forecast international trade: The case of Croatia. *Bus Syst Res*, 13(3): 144-160.
- Ke G, et al. 2017. LightGBM: A highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*, 30: 3146-3154.
- Micocci F, Rungi A. 2023. Predicting exporters with machine learning. *World Trade Rev*, 22(5): 584-607.
- Ng R T, Han J. 2002. CLARANS: A method for clustering objects for spatial data mining. *IEEE Trans Knowl Data Eng*, 14(5): 1003-1016.
- Nummelin T, Hänninen R. 2016. Model for international trade of sawnwood using machine learning models. Helsinki, Finland, pp: 154-168.
- Nuroğlu E. 2014. Estimating and forecasting trade flows by panel data analysis and neural networks. *İstanbul Univ İktisat Fak Mec*, 64(1): 85-111.
- Reçus L, Tinbergen J. 1962. Shaping the world economy. Suggestions for an international economic policy. The Twentieth Century Fund, New York. *Rech Econ Louvain*, 30(2): 157.
- Shen M L, Lee C F, Liu H H, Chang P Y, Yang C H. 2021. Effective multinational trade forecasting using LSTM recurrent neural network. *Expert Syst Appl*, 182: 115199.
- Soydal H, Ay M, Koç S. 2024. Makine öğrenimi ile ticaretin öngörülmesi: Türkiye ve Balkan ülkeleri üzerine çekim modeli analizi. *Selçuk Univ Sos Bil Meslek Yüksekokulu Derg*, 27(2): 746-765.
- Torkul O, Gülseçen S, Uyaroğlu Y, Çağal G, Uçar M K. 2017. Mühendislikte yapay zeka ve uygulamaları. Ankara, Türkiye, pp: 156-158.
- Yağmur E Ç. 2022. Atmosferik partikül maddelerin makine öğrenmesi ile tahmini: Beşiktaş, İstanbul örneği. *Konya J Eng Sci*, 10(4): 807-826.