



**Türkiye'nin Kimyasal Madde İthalatının Gelecek Tahmini: Makine Öğrenmesi Ve Topluluk Öğrenme Yöntemleri Performans Analizi**

*Future Forecasting of Turkey's Chemical Imports: Performance Analysis of Machine Learning and Ensemble Learning Methods*

**Kâmil Abdullah EŞİDİR<sup>1</sup>**

<sup>1</sup>Dr., Fırat Kalkınma Ajansı, Elazığ, [abdullahesidir@yahoo.com](mailto:abdullahesidir@yahoo.com), [orcid.org/0000-0002-8106-1758](https://orcid.org/0000-0002-8106-1758)

*Araştırma Makalesi/Research Article*

**Makale Bilgisi**

Geliş/Received: 06.11.2024  
Kabul/Accepted:  
19.12.2024

**DOI:**

10.18069/firatsbed.1580620

**Anahtar Kelimeler**

Kimyasal Madde İthalatı,  
Makine Öğrenmesi,  
Topluluk Öğrenme,  
Ekonomik Tahmin,  
XGBoost.

**Keywords**

Chemicals Imports,  
Machine Learning,  
Ensemble Learning,  
Economic Forecasting,  
XGBoost.

**ÖZ**

Bu çalışma, Türkiye'nin kimyasal madde ithalatının gelecekteki değerlerini tahmin etmek amacıyla makine öğrenmesi ve topluluk öğrenme yöntemlerinin performansını değerlendirmektedir. Doğrusal Regresyon, Rastgele Orman, Rasyonel Kuadratik Regresyon, Destek Vektör Makinesi ve XGBoost modelleri kullanılmıştır. Veriler, TÜİK ve TCMB gibi güvenilir kaynaklardan elde edilmiştir. Makroekonomik değişkenler arasında Türkiye İthalatı, Kimyasallar ve Kimyasal Ürünler Üretim Endeksi, Aylık Ortalama Dolar Kuru, İmalat Sanayi Üretim Endeksi, Petrol Varil Fiyatları ve Kimyasal Madde İhracatı yer almaktadır. Analiz sonuçlarına göre, XGBoost modeli en yüksek doğruluk ve genelleme yeteneğine sahiptir. Model, eğitim, test ve çapraz doğrulama setlerinde en düşük hata oranları ve en yüksek R<sup>2</sup> değerleri ile başarılı performans göstermiştir. SHAP analizi, Türkiye İthalatı ve Kimyasallar Üretim Endeksi değişkenlerinin en yüksek etkiye sahip olduğunu ortaya koymuştur. XGBoost modeli ile yapılan projeksiyonlar, Türkiye'nin kimyasal madde ithalatının gelecekteki seyrine dair önemli bilgiler sunmakta, ekonomik planlama ve ticari stratejiler için kritik öneme sahiptir. Modelin yeteneği, politika yapımcılar ve iş dünyası için stratejik kararları kolaylaştırmaktadır.

**ABSTRACT**

This study evaluates the performance of machine learning and ensemble learning methods to predict the future values of Turkey's chemical imports. Linear Regression, Random Forest, Rational Quadratic Regression, Support Vector Machine and XGBoost models are used. Data are obtained from reliable sources such as TurkStat and CBRT. Macroeconomic variables include Turkey's Imports, Chemicals and Chemical Products Production Index, Monthly Average Dollar Exchange Rate, Manufacturing Industry Production Index, Oil Barrel Prices and Chemicals Exports. According to the analysis results, the XGBoost model has the highest accuracy and generalization ability. The model performed well in the training, test and cross-validation sets with the lowest error rates and the highest R<sup>2</sup> values. SHAP analysis reveals that Turkey Imports and Chemicals Production Index variables have the highest impact. The projections made with the XGBoost model provide important insights into the future course of Turkey's chemical imports and are critical for economic planning and trade strategies. The model's capability facilitates strategic decisions for policymakers and the business community.

**Atıf/Citation:** Eşidir, A. K. (2025). Türkiye'nin Kimyasal Madde İthalatının Gelecek Tahmini: Makine Öğrenmesi Ve Topluluk Öğrenme Yöntemleri Performans Analizi. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 35, 1, 1-18

**Sorumlu yazar/Corresponding author:** Kamil Abdullah, EŞİDİR, [abdullahesidir@yahoo.com](mailto:abdullahesidir@yahoo.com)

## 1. Giriş

Türkiye, büyüyen sanayi sektörü ve artan üretim kapasitesiyle kimyasal madde ithalatında önemli bir konuma sahiptir. Kimyasal maddeler; ilaç, gıda, tarım, otomotiv ve tekstil gibi sektörlerde temel hammadde olarak ekonomik kalkımda kritik rol oynamaktadır (Sevim vd., 2009). Özellikle plastikler, petrokimya ürünleri, ilaçlar ve tarım kimyasalları gibi geniş ürün yelpazesi öne çıkmaktadır. 2000 yılında 67 milyon dolar olan kimyasal madde ihracatı, 2023'te 1,3 milyar doları aşarken, ithalat değeri aynı dönemde 512 milyondan 3,4 milyar dolara yükselmiştir. Tablo 1'de yıllara göre kimyasal madde ihracat ve ithalat rakamları USD cinsinden verilmektedir.

**Tablo 1.** Kimyasal Madde Dış Ticaret Verileri

Yıllar	Kimyasal Madde İhracatı	Kimyasal Madde İthalatı	Yıllar	Kimyasal Madde İhracatı	Kimyasal Madde İthalatı
2000	66,773	512,449	2012	496,882	2,053,156
2001	64,299	484,489	2013	554,767	2,123,678
2002	76,817	579,860	2014	593,496	2,276,076
2003	95,566	743,088	2015	546,391	2,049,569
2004	121,883	960,110	2016	562,646	2,024,132
2005	160,869	1,131,470	2017	576,320	2,212,861
2006	220,295	1,299,243	2018	685,978	2,281,783
2007	280,222	1,537,152	2019	750,284	2,093,677
2008	364,258	1,732,921	2020	1,064,203	2,263,405
2009	320,298	1,473,007	2021	1,096,283	2,771,089
2010	399,775	1,795,445	2022	1,294,182	3,304,099
2011	451,639	2,204,854	2023	1,305,839	3,403,012

**Kaynak:** TÜİK, 2024.

Türkiye, sınırlı yerli enerji kaynakları nedeniyle enerji talebini büyük ölçüde ithalatla karşılamaktadır ve bu durum enerji güvenliği açısından stratejik planlamayı zorunlu kılmaktadır (Erdem, 2010). Kimyasal madde ithalatı, Türkiye'nin dış ticaretinde önemli bir yer tutmakla birlikte, bu sektör ekonomik büyüme, istihdam ve endüstriyel kalkınma açısından kritik bir role sahiptir (Emirhan ve Turgutlu, 2023; Şener vd., 2014). Devlet destekleri ve Ar-Ge yatırımları, sektörün rekabet gücünü artırarak küresel piyasalarda dinamik yapısını korumasına katkıda bulunmaktadır. Türkiye, Gümrük Birliği Anlaşması (1995) ile Avrupa ekonomik çerçevesine entegre olmuş, bu durum pazar erişimi ve rekabet avantajı açısından önemli fırsatlar sunmuştur (Crescimanno vd., 2014). Dolayısıyla, kimyasal madde ithalatının sürdürülebilirliği, dış ticaret politikalarının oluşturulmasında ve ekonomik stratejilerin belirlenmesinde büyük önem taşımaktadır.

Kimyasal madde ithalatının doğru tahmin edilmesi, dış ticaret dengesi, döviz rezervlerinin yönetimi ve dış ticaret açığının kontrolü açısından kritik önem taşımaktadır (Flyvbjerg, 2006). Ayrıca, hammadde tedarikinin sürekliliği için sanayi ve üretim süreçlerinin planlanmasına, tedarik zinciri yönetimine ve maliyet optimizasyonuna önemli katkılar sağlar (Ji vd., 2014; Zareipour vd., 2010). Bu tahminler, işletmeler ve politika yapımcılar için stratejik kararlar açısından yol gösterici niteliktedir. Makine öğrenmesi ve topluluk öğrenme yöntemleri, büyük veri setleri ve karmaşık ilişkiler içeren tahmin problemlerinde geleneksel yöntemlere kıyasla yüksek doğruluk sağlamaktadır (Medeiros, 2022; Palkovits ve Palkovits, 2019). Bu teknikler, verilerdeki örüntüleri otomatik olarak öğrenerek gelecekteki değerleri tahmin etme kapasitesine sahiptir (Huertas-Tato ve Brito, 2018). Örneğin, Rastgele Orman modeli emlak fiyatı tahmininde en başarılı sonuçları verirken, karar ağaçları daha düşük performans göstermiştir (Marzooqi, 2024). Bu nedenle, makine öğrenmesi yöntemleri, tahmin problemlerinde üstün performanslarıyla giderek daha fazla tercih edilmektedir. Bu çalışmada, Türkiye'nin kimyasal madde ithalatının tahmininde Doğrusal Regresyon, Rastgele Orman, Rasyonel Kuadratik Regresyon, Destek Vektör Makinesi (DVM) ve Extreme Gradient Boosting yöntemlerinin performansları karşılaştırılmıştır. Amaç, en doğru tahmin yöntemini belirleyerek gelecek tahminleri yapmaktır. Modellerin doğruluğu ve hata oranları detaylı şekilde incelenmiş, en uygun model seçilmiştir. Literatürde sınırlı olan bu alanda, veri odaklı makine öğrenmesi yöntemleri kullanılarak ekonomik etkilerin analizi gerçekleştirilmiştir.

Çalışma, daha yüksek doğruluk sağlayan tahminler sunarak politika yapıcılara karar destek mekanizmaları oluşturma potansiyeline sahiptir ve sektörel dinamiklerin kapsamlı anlaşılmasına katkı sağlamaktadır. Çalışmanın ilerleyen bölümlerinde, kullanılan veri seti, modelleme süreçleri ve elde edilen bulgular ayrıntılı bir şekilde incelenmiş ve bu alandaki literatüre katkıları değerlendirilmiştir. Bu kapsamda, kimyasal madde ithalatının tahmin edilmesine yönelik makine öğrenmesi ve topluluk öğrenme yöntemlerinin uygulanabilirliği ve etkinliği üzerine kapsamlı bir analiz sunulmuştur. Ayrıca, bu modellerin Türkiye'nin ekonomik stratejilerine ve sanayi politikalarına nasıl entegre edilebileceği tartışılmış ve gelecekteki çalışmalar için tavsiyelerde bulunulmuştur.

## **2. Literatür İncelemesi**

Bu çalışma, Türkiye'nin kimyasal madde ithalatını etkileyen makroekonomik değişkenleri bütüncül bir yaklaşım ile ele alarak, literatürde önemli bir boşluğu doldurmaktadır. Ayrıca, farklı makine öğrenmesi modellerinin performanslarının sistematik bir şekilde karşılaştırılması ve bu modellerin ekonomik politika ile ticari stratejilere uygulanabilir tahminler sunacak biçimde optimize edilmesi, mevcut literatüre özgün katkılar sağlamaktadır. Bu kapsamda, çalışma yalnızca literatürdeki metodolojik eksiklikleri gidermekle kalmamakta, aynı zamanda ekonomik karar alma süreçlerine yönelik tahminlerin stratejik önemine de dikkat çekmektedir. Makroekonomik değişkenler, kimyasal madde ithalat ve ihracatını önemli ölçüde etkilemektedir. Moshiri ve Kheirandish (2023), petrol fiyatı şoklarının küresel ticaret ve makroekonomik değişkenler üzerindeki yayılma etkilerini ortaya koyarak, bu şokların farklı sektörlerdeki etkilerini vurgulamıştır. Döviz kurlarının ithalat üzerindeki etkisi ise uluslararası ekonomi literatüründe önemli bir konudur. Döviz kuru oynaklığı, ithalat fiyatlarında dalgalanmalara yol açarak ithalat talebini etkilerken, Sercu ve Uppal (2003) bu dalgalanmaların uluslararası ticaret ve ulusal refah üzerindeki etkilerini genel denge analiziyle incelemiştir. Bir ülkenin genel ithalatı ile belirli sektörlerin ithalatı arasındaki ilişki, teknoloji ve sermaye bağımlılığı gibi faktörlerden etkilenmektedir. Erduman vd. (2020), sermaye ve teknoloji yoğun sektörlerin, örneğin petrol ürünleri ve temel metallerin, daha yüksek ithalat gereksinimleri olduğunu belirtmiştir. Wang vd. (2021) ise imalat ithalatının, yerli sanayinin toplam faktör verimliliğini artırabileceğini vurgulamıştır. Kimyasalların ithalatı üzerinde ise üretim endeksi önemli bir rol oynar; yıllık kimyasal üretim miktarı çevresel etkiler açısından kritik bir faktördür (Kishi ve Sekine, 2003). Ayrıca, üretimdeki hammaddelerin seçimi ve kimyasal sentez süreçlerinin karmaşıklığı, kimyasal ürünlerin çevresel performansını belirlemede hayati önem taşır (Fantke ve Ernststoff, 2017). Kimyasal ürünlerin yüksek saflığı, kaliteyi artırırken üretim maliyetlerini de düşürmektedir (Zolfıgol, 2024). Ayrıca, büyük ölçekli kimyasal üretim ve ithalat, özellikle bir milyon poundu aşan hacimlerde önem kazanmaktadır (Walker vd., 2003). Bu çalışmada, kimyasal madde ithalatını etkileyen dolar kuru, Türkiye genel ithalatı, imalat sanayi üretim endeksi, Kimyasal Ürünlerin Üretimi Endeksi, petrol varil fiyatları ve kimyasal madde ihracatı gibi makroekonomik değişkenler analize dahil edilmiştir. Bu değişkenler, ekonomik ve endüstriyel dinamikleri daha doğru analiz ederek ithalat tahminlerinin doğruluğunu artırmaktadır.

Makine öğrenmesi ve topluluk öğrenme yöntemleri, tahmin problemlerinde artan doğruluk ve verimlilikleriyle öne çıkmaktadır. Şimşek vd. (2024), Batı Teksas Petrolü (WTI) fiyatlarının tahmini için LSTM tabanlı özellik çıkarımı ve XGBoost regresöründen oluşan bir hibrit model geliştirmiştir. Ocak 1986 - Mayıs 2023 dönemine ait verilerle yapılan çalışmada, altın fiyatı, ABD 10 yıllık tahvil getirisi, küresel ekonomik aktivite endeksi ve ABD Dolar Endeksi bağımsız değişkenler olarak kullanılmış; model, MAE, MSE, RMSE, MAPE ve R<sup>2</sup> ölçütlerine göre üstün performans göstermiştir. Gür (2024) ise Türk Hava Yolları (THYAO) hisse fiyatlarını tahmin etmek için SVM, XGBoost ve LSTM algoritmalarını kullanarak, Ocak 2010 - Eylül 2023 dönemi günlük verileriyle modelleri değerlendirmiştir. Bulgular, LSTM modelinin en düşük hata oranlarıyla en iyi performansı sunduğunu göstermektedir. Bu çalışmalar, makine ve derin öğrenme yöntemlerinin finansal ve ekonomik tahminlerde etkinliğini ortaya koymaktadır. Gür (2024), ABD Tüketici Fiyat Endeksi (CPI) tahmini için LSTM, MARS, XGBoost ve hibrit modeller (LSTM-MARS, LSTM-XGBoost) kullanmış, Ocak 1974 - Ekim 2023 dönemi verilerinde genetik algoritma ve Bayes optimizasyonu ile hiperparametre ayarı yapmıştır. LSTM-XGBoost modeli, ekonomik değişkenler arasındaki karmaşık ilişkileri ele alarak en yüksek tahmin doğruluğunu sağlamıştır. Şimşek (2024) ise Türkiye TÜFE tahmini için DDPG, XGBoost, SVR, KNN ve CNN-BiLSTM yöntemlerini uygulamış; Ocak 2003 - Ağustos 2023 verilerinde üretici fiyat endeksi, M1 para arzı, altın fiyatı, dolar kuru ve faiz oranı gibi değişkenler kullanmıştır. SVR modeli, en iyi performansı göstermiştir. Ayrıca, Gür (2024), Türkiye'de aylık inek peyniri üretimini tahmin etmek için LSTM, GRU,

MLP, SVR ve kNN modellerini incelemiş ve LSTM modeli en yüksek doğruluğu sağlamıştır. Ocak 2010 - Eylül 2023 dönemi verilerine dayalı olarak yapılan tahminlerde, süt üretimi, yem fiyatı ve peynir ihracatı bağımsız değişken olarak kullanılmış; Ekim 2023 - Eylül 2024 için artış trendi öngörülmüştür. Bu çalışmalar, makine öğrenmesi ve derin öğrenme yöntemlerinin ekonomik tahminlerdeki etkinliğini vurgulamaktadır. İthalat ve ihracat tahminlerinde makine öğrenmesi ve topluluk öğrenme yöntemleri yaygın olarak kullanılmaktadır. Jošić ve Žmuk (2022), Gauss süreçleri, doğrusal regresyon ve çok katmanlı perceptron ile Hırvatistan'ın 2020 yılı ikili ticaretini tahmin etmiştir. Özemre ve Kabadurmuş (2020), ihracat tahmininde büyük veri ve makine öğrenimi algoritmalarını kullanırken, Brabenec ve Šuleř (2020) Çek Cumhuriyeti ile Çin arasındaki ticaret dengesini modellemiştir. Wang (2022), PSO ile optimize edilmiş RVM modeliyle blok zinciri güvenliğine dayalı ticaret tahmini geliştirmiştir. Meng vd. (2018), enerji tahmininde hibrit makine öğrenimi algoritmalarını ele almış, Bayrak (2020) ise tıbbi cihaz dış ticaretini tahmin etmek için döviz kuru ve enflasyonu içeren DVM tabanlı bir topluluk modeli önermiştir. Bu çalışmalar, makine öğrenmesi yöntemlerinin ticaret tahmininde doğruluğu artırarak ticaret dinamiklerine önemli katkılar sunduğunu göstermektedir. Kimyasal ürün sektörüne yönelik tahmin çalışmalarında, Broeren vd. (2014), temel kimyasal üretimi için maliyet odaklı bir küresel tahmin modeli geliştirmiş ve 2010-2030 döneminde bölgesel üretim, enerji verimliliği, emisyonlar ve politikaların etkilerini incelemiştir. Bulgular, yeni kimyasal üretim kapasitesinin %60'ının OECD dışı bölgelerde olacağını ve CO2 emisyonlarının %50 artacağını, ancak enerji politikalarıyla bu artışın %30-40'a indirilebileceğini göstermektedir. Li ve Li (2022) ise PSO-BP sinir ağı modeliyle Çin'in kimyasal ihracatını tahmin ederek, modelin geleneksel yöntemlere kıyasla daha yüksek doğruluk sağladığını ortaya koymuştur. Pisuttinart vd. (2022), Tayland'ın gübre ithalat talebini SARIMA modelleriyle analiz etmiş ve 2022 tahminlerine göre azotlu gübre talebinde %5.12 artış, potasyumlu gübrede %8.74 azalış, bileşik gübrede ise %4.74 artış tespit etmiştir. Bu çalışmalar, kimyasal ürün sektöründe tahmin modellerinin doğruluğu ve uygulama alanlarının önemini vurgulamaktadır.

Gür ve Eşidir (2024), Türkiye'nin hurda demir-çelik ithalatını tahmin etmek için LSTM, MLP, Random Forest, SVM, XGBoost ve Doğrusal Regresyon modellerini karşılaştırmış ve LSTM modelinin en iyi performansı gösterdiğini belirlemiştir. RMSE, MSE, MAE, MAPE ve R<sup>2</sup> metrikleriyle değerlendirilen çalışmada, LIME analizi ile "YÜFE" ve "Aylık Demir Çelik İthalatı" değişkenlerinin tahmin sonuçlarına yüksek etkisi tespit edilmiştir. Bu çalışma, makine öğrenmesi yöntemlerinin ithalat tahminlerindeki etkinliğini vurgulayarak ekonomik planlama ve strateji geliştirme süreçlerine katkı sağlamaktadır. Özellikle Türkiye gibi gelişmekte olan ülkeler için kimyasal madde sektörü için tahmin çalışmalarının daha önce hiç uygulanmamış olması, çalışmanın önemini ve yenilikçi yönünü artırmaktadır. Türkiye'nin ekonomik stratejilerine yönelik bu tahminler, politika yapıcılar ve iş dünyası için değerli bilgiler sunarak karar süreçlerine rehberlik edecektir. Bu çalışmada kullanılan modellerin geniş bir yelpazesi, farklı makroekonomik değişkenlerin dahil edilmesi ve elde edilen sonuçların kapsamlı analizi, kimyasal madde ithalatı tahminleri konusundaki literatüre önemli katkılar sağlayacaktır. Ayrıca, bu çalışmanın bulguları, benzer ekonomik ve endüstriyel dinamiklere sahip diğer ülkeler için de uygulanabilirlik sunmaktadır ve bu alandaki gelecekteki çalışmalar için bir temel oluşturacaktır.

### 3. Araştırmanın Metodolojisi

Bu bölümde, Türkiye'nin kimyasal madde ithalatının tahmininde kullanılan yöntemler ve teknikler açıklanmaktadır. Çalışmada, doğrusal regresyon, Rastgele Orman, Destek Vektör Makinesi, XGBoost ve Rasyonel Kuadratik Regresyon gibi modeller, makroekonomik veriler kullanılarak değerlendirilmiştir. Doğrusal regresyon, basitliği ve yorumlanabilirliği nedeniyle temel kıyaslama modeli olarak seçilmiştir. Rastgele Orman ve Destek Vektör Makinesi, doğrusal olmayan ilişkileri modelleme kapasitesiyle dahil edilmiş, XGBoost ise güçlü genelleme yeteneği ve SHAP analiziyle değişken önemini dengeli sunmasıyla öne çıkmıştır. Rasyonel Kuadratik Regresyon ise karmaşık ilişkileri modelleme yeteneğiyle tercih edilmiştir. Bu modeller, ekonomik tahminlerde çeşitlilik sağlamak ve en iyi performansı belirlemek amacıyla bir arada değerlendirilmiştir.

### 3.1. Veri Seti

Analizlerde kullanılan veriler, Türkiye İstatistik Kurumu (TÜİK) ve Türkiye Cumhuriyet Merkez Bankası (TCMB) kaynaklarından elde edilmiştir. Ocak 2005 ile Mart 2024 arasındaki 231 aylık dönemi kapsayan veri setinde, 6 bağımsız ve 1 bağımlı değişken için her birine ait 231 veri bulunmaktadır. Veri setlerinde herhangi bir kayıp veya eksik bilgi yoktur. Modelin bağımlı değişkeni aylık kimyasal madde ithalatı (USD) olup, bağımsız değişkenler ise; ortalama Amerikan Dolar Kuru (TL), aylık Türkiye ithalatı, İmalat Sanayi Üretim Endeksi, Kimyasallar ve Kimyasal Ürünler Üretim Endeksi, aylık petrol varil fiyatı (USD) ve kimyasal ürünlerin aylık ihracatıdır (USD). Her bir değişkenin aylık verileri .xlsx formatında bir dosyada birleştirilerek tahmin modellerinde kullanılmak üzere toplanmıştır.

### 3.2. Veri Ön İşleme

Verilerin analizden önce temizlenmesi ve dönüştürülmesi gerekmektedir. Elde edilen resmi verilerde eksik değer bulunmamaktadır. Veriler, Python'un Pandas paketi kullanılarak bir Excel dosyasından içe aktarılmıştır. Daha sonra, veri çerçevesi bağımsız değişkenler (X) ve bağımlı değişken (Y) olarak ayrılmıştır. Ön işleme aşamasının bir parçası olarak, veriler min-max ölçeklendirme kullanılarak standartlaştırılmıştır. Bu, makine öğrenimi ve veri analizinde özelliklerin belirli bir aralıkta, genellikle 0 ile 1 arasında yeniden ölçeklendirilmesi için yaygın olarak kullanılan bir tekniktir. Bu ölçeklendirme yöntemi, tüm özelliklerin analize eşit şekilde katkıda bulunmasını sağlar ve belirli bir özelliğin ölçeği nedeniyle modele hakim olmasını engeller. İşlem, özelliğin minimum değerinin çıkarılması ve ardından maksimum ve minimum değerler arasındaki fark olan özelliğin aralığına bölünmesini içerir (Ampomah vd., 2021). Denklem 1'de bu ölçeklendirme yöntemi gösterilmektedir.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

### 3.3. Yazılım ve Donanım

Çalışmada, Python 3.6 ile veri analizi, modelleme ve görselleştirme işlemleri gerçekleştirilmiştir. Veri işlemede Pandas ve NumPy, modelleme süreçlerinde Scikit-Learn ve XGBoost, görselleştirmede ise Matplotlib ve Seaborn kütüphaneleri kullanılmıştır. Hızlı veri işleme ve model eğitimi için yüksek işlem gücüne sahip bir bilgisayar tercih edilmiştir. Kullanılan bilgisayar sisteminde AMD Ryzen 9 5950X işlemci, 32 GB RAM, 1 TB SSD ve 10 GB VRAM'e sahip NVIDIA GeForce RTX 3080 GPU bulunmakta olup, bu altyapı yapılan analiz ve işlemlerin etkinliğini ve başarısını doğal olarak artırmıştır.

### 3.4. Modeller

Bu çalışmada, Türkiye'nin kimyasal madde ithalatının gelecekteki değerlerini tahmin etmek amacıyla çeşitli makine öğrenmesi ve topluluk öğrenme modelleri kullanılmıştır. Modellerin seçimi, farklı tahmin yöntemlerinin performanslarını karşılaştırarak en iyi sonucu elde etmek için yapılmıştır. Her bir model, veri setindeki karmaşık ilişkileri ve örüntüleri yakalamak için farklı matematiksel ve istatistiksel teknikler kullanılmaktadır. Aşağıda, çalışmada kullanılan başlıca modeller ve bu modellerin tahmin sürecindeki matematiksel temelleri detaylı bir şekilde açıklanmaktadır.

#### 3.4.1. Doğrusal Regresyon (Linear Regression)

Doğrusal regresyon, bağımlı değişken (y) ile bir veya daha fazla bağımsız değişken (x) arasındaki doğrusal ilişkiyi modellemek için kullanılır (Marill, 2004). Tek bir bağımsız değişken söz konusu olduğunda, doğrusal bir ilişki kurmak veya bir korelasyon katsayısı elde etmek için basit doğrusal regresyon kullanılabilir (MacCallum vd., 2002). Birden fazla bağımsız değişken söz konusu olduğunda, bu değişkenlere dayalı olarak bağımlı değişkeni tahmin etmek için çoklu doğrusal regresyon kullanılır (Eberly, 2007). Bu teknik, sürekli bir sonuçla ilişkili olarak birkaç değişkenin aynı anda değerlendirilmesine olanak tanımaktadır (Eberly, 2007). Doğrusal regresyon modeli Denklem 2'de gösterildiği gibidir:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (2)$$

Burada,  $y$  bağımlı değişkeni,  $x_i$  bağımsız değişkenleri,  $\beta_i$  regresyon katsayılarını,  $\beta_0$  sabit terimi ve  $\epsilon$  hata terimini temsil etmektedir. Regresyon katsayıları ( $\beta_i$ ), en küçük kareler yöntemi ile belirlenir. En küçük kareler yöntemi, hata terimini minimize ederek bağımlı değişkenin en iyi şekilde tahmin edilmesini sağlar. Katsayıların hesaplanması Denklem 3'te gösterilmektedir:

$$\beta = (X^T X)^{-1} X^T y \quad (3)$$

Burada,  $X$  bağımsız değişkenlerin matris formu,  $y$  bağımlı değişken vektörü ve  $\beta$  katsayı vektörünü temsil etmektedir.

### 3.4.2. Rastgele Orman (Random Forest)

Random Forest, birden fazla karar ağacının birleşiminden oluşan bir topluluk öğrenme yöntemidir. Her bir karar ağacı, veri setinin rastgele bir alt kümesi üzerinde eğitilir ve nihai tahmin, tüm ağaçların tahminlerinin ortalaması alınarak yapılır (Svetnik vd., 2003; Biau ve Scornet, 2016). Rastgele Orman, sağlamlığı ve yüksek boyutlu verilerle iyi başa çıkma kabiliyetiyle bilinir, bu da onu bireysel karar ağaçlarına kıyasla aşırı uyum gibi sorunlara karşı daha toleranslı hale getirir. Algoritma, karar ağaçları topluluğu aracılığıyla doğal olarak bir belirsizlik ölçüsü sağlar ve bu da tahminlerde güvenilirliğine katkıda bulunur (Dutschmann ve Baumann, 2021; Li, 2023). Bu modelin matematiksel ifadesi Denklem 4'teki gibidir:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (4)$$

Burada,  $\hat{y}$  nihai tahmini,  $T_i(x)$  her bir karar ağacının tahminini ve  $N$  toplam ağaç sayısını göstermektedir. Her bir karar ağacı, belirli bir derinliğe kadar büyütülür ve bu süreçte veri setindeki en iyi bölünmeleri (splits) belirler. Her bir düğümde (node) veri, belirli bir özelliğe göre iki alt kümeye bölünür. Bölünme, genellikle Gini impüriti veya bilgi kazancı gibi bir kriter kullanılarak yapılır.

### 3.4.3. Rasyonel Kuadratik Regresyon (Rational Quadratic Regression)

Rasyonel Kuadratik Regresyon, doğrusal ve doğrusal olmayan bileşenleri birleştirerek bağımlı ve bağımsız değişkenler arasındaki karmaşık ilişkileri modelleyen esnek bir tekniktir. Bu yöntem, çoklu regresyon modeline ikinci dereceden terimler ekleyerek doğrusal olmayan etkileri incelemeyi sağlar (Wen vd., 2021). İkinci dereceden terimlerin ve etkileşimlerin entegrasyonu, hesaplama verimliliğini artırarak daha doğru modelleme ve nedensel içgörüler sunar (Bhagwat vd., 2012). Özellikle büyük veri kümelerinde hata azaltma avantajı sunan bu model, karmaşık ilişkilerin güvenilir şekilde analiz edilmesini sağlar (Polat vd., 2023). Doğrusal regresyonun yetersiz kaldığı durumlarda, kuadratik regresyon gibi modeller daha iyi uyum sağlayabilir (Hui vd., 2012).

$$y = \frac{ax^2 + bx + c}{dx^2 + ex + f} + \epsilon \quad (5)$$

Burada,  $a, b, c, d, e,$  ve  $f$  katsayıları,  $\epsilon$  hata terimini temsil etmektedir.

### 3.4.4. Destek Vektör Makinesi (Support Vector Machine)

Destek Vektör Regresyonu (DVR), Destek Vektör Makineleri (DVM) prensiplerini regresyon problemlerine uyarlayan bir makine öğrenme tekniğidir. DVR, tüm veri noktalarının hiper düzlemden toplam sapmasını en aza indiren optimal bir regresyon hiper düzlemi oluşturmayı amaçlar (Cao vd., 2022; Wen vd., 2019). Model, ikinci dereceden optimizasyon problemi çözerek hiper düzlem ile veri noktaları arasındaki marjı en üst düzeye çıkarır ve noktaları belirli bir hata toleransı dahilinde kapsar (Mishra vd., 2023; Wie vd., 2020). DVR, doğrusal olmayan ilişkileri ele almak için verileri düşük boyutlu uzaydan yüksek boyutlu bir özellik uzayına dönüştürerek, bu uzayda optimum hiper düzlemi arar (Zhang vd., 2021). Böylece model, veri dağılımını etkili bir şekilde yakalayıp regresyon hatalarını en aza indirir.

$$y = \sum_{i=1}^n (a_i - a_i^*) K(x_i, x) + b \quad (6)$$

Burada,  $y$  tahmin edilen değeri,  $a_i$  ve  $a_i^*$  Lagrange çarpanlarını,  $K(x_i, x)$  çekirdek fonksiyonunu,  $b$  bias terimini ve  $x_i$  eğitim verisini temsil etmektedir. SVR, doğrusal olmayan ilişkileri modellemek için çekirdek fonksiyonları (lineer, polinomial, radyal bazlı fonksiyonlar gibi) kullanır. Çekirdek fonksiyonu  $K(x_i, x)$ , yüksek boyutlu uzayda veri noktaları arasındaki benzerlikleri ölçer.

### 3.4.5. Extreme Gradient Boosting (XGBoost)

XGBoost, gradyan artırma algoritmasına dayanan bir topluluk öğrenme yöntemidir ve hata fonksiyonunu en aza indirmek için zayıf öğrencileri sıralı şekilde birleştirerek tahmin doğruluğunu artırır (Chen, 2016). Ölçeklenebilirliği ve esnekliği sayesinde popülerlik kazanan XGBoost, gradyan artırma kütüphaneleri arasında öne çıkmaktadır (Ge vd., 2021). Boosting kategorisinde yer alan bu algoritma, zayıf öğrencileri güçlü öğrencilere dönüştürmeyi amaçlar (Asaad, 2024) ve birden fazla karar ağacını entegre ederek sağlam bir sınıflandırıcı oluşturur (Wang vd., 2020; Liang vd., 2020). Modelin tahmin fonksiyonu Denklem 7'de gösterilmektedir:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (7)$$

Burada,  $\hat{y}_i$  tahmin edilen değerleri,  $F$  karar ağaçlarının uzayını,  $K$  toplam ağaç sayısını ve  $f_k$  her bir ağacı temsil etmektedir. XGBoost, her iterasyonda hatayı minimize etmek için gradyan iniş algoritmasını kullanır ve bu sayede modelin doğruluğunu artırır. Modelin genel kayıp fonksiyonu Denklem 8'deki gibidir:

$$L(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (8)$$

Burada,  $L(\phi)$  genel kayıp fonksiyonunu,  $l(y_i, \hat{y}_i)$  tahmin hatasını,  $\Omega(f_k)$  karmaşıklık cezasını,  $y_i$  gerçek değeri ve  $\hat{y}_i$  tahmin edilen değeri göstermektedir.

Bu bölümde anlatılan modellerin her biri, veri setindeki karmaşık ilişkileri ve örüntüleri yakalamak için farklı matematiksel ve istatistiksel teknikler kullanmaktadır. Bu sayede, kimyasal madde ithalatının gelecekteki değerlerini daha doğru bir şekilde tahmin etmek için en iyi modelin seçimi amaçlanmıştır. Tüm modeller için hiperparametre ayarları Random Search (Rastgele Arama) algoritması ile yapılmıştır ve Tablo 2'de her bir model için belirlenen ayarlar gösterilmiştir.

**Tablo 2.** Random Search Algoritması ile Optimizasyon Sonucu Modellerin Belirlenen Hiperparametreleri

Random Forest	Linear Regression	SVR	XGBoost	Rational Quadratic Regression
n_estimators:400	Preset: Linear	Kernel: Linear	Sub_sample:0.7	Preset:
max depth: 40	Terms: Linear	Epsilon: 0.1	n_estimators: 200	Kernel
min samples split: 3	Robust option: Off	C: 10	max_depth: 3	Tol: 1e-3
min samples leaf: 2	Fit_intercept: True	Max_iter: 1000	learning rate: 0.2	Shrinking:
max featutes: log2	Random state: 42	Tol: 1e-3	colsample_bytree: 1.0	True
bootstrap: True		Shrinking: True	min_child_weight: 1	Cache_size:
criterion: gini		Cache_size: 300	gamma: 0	200
Random state: 42		Verbose: False	reg_alpha: 0	Verbose:
		Random state: 42	reg_lambda: 1	False
			scale_pos_weight: 1	Random
			Random state: 42	state: 42

### 3.5. Model Performans Değerlendirme

Veri seti, modellerin eğitimi ve test edilmesi için %80 eğitim ve %20 test olarak ayrılmıştır. Random state kullanımı, veri bölünmesinin tutarlı olmasını sağlayarak modellerin performanslarının karşılaştırılabilirliğini artırmıştır (Ferrat vd., 2018). Modelin genelleme yeteneğini değerlendirmek için 15 katlı k-fold çapraz doğrulama uygulanmıştır; bu yöntemde veri seti k eşit parçaya bölünerek her parça sırayla test seti, kalan k-1 parça ise eğitim seti olarak kullanılmıştır (Moon vd., 2020). Süreç, toplamda 15 kez tekrarlanarak modelin farklı veri bölümlerindeki performansı ölçülmüştür (Naseem vd., 2019). Modellerin eğitim, test ve çapraz doğrulama performanslarını değerlendirmek için MSE, MAE, MAPE, RMSE ve R<sup>2</sup> gibi ölçütler kullanılmıştır. Bu sonuçlar, modellerin tahmin doğruluğu ve verimliliğini kapsamlı bir şekilde değerlendirmek amacıyla analiz edilmiştir. Bu ölçütlerin matematiksel hesaplamaları sırasıyla Denklem 9-13'te sunulmuştur.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (9)$$

$$MAE = \frac{\sum_{i=1}^n |t_i - g_i|}{n} = \frac{\sum_{i=1}^n e_j}{n} \quad (10)$$

$$MAPE = \frac{\sum_{t=1}^n \frac{u_t}{\bar{y}_t}}{n} \times 100 \quad (11)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (12)$$

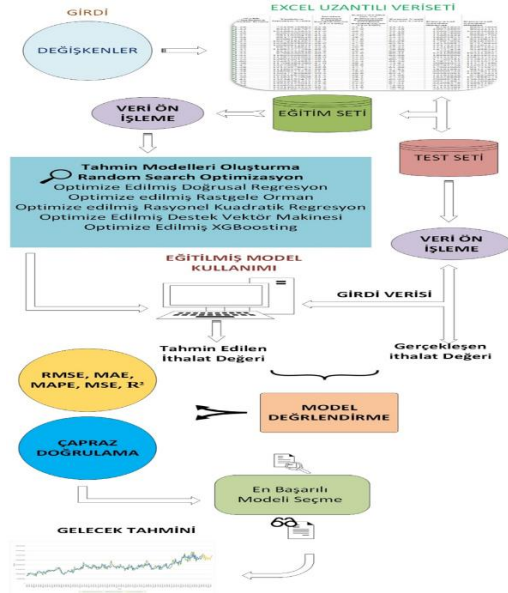
$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \mu)^2} \quad (13)$$

MSE, tahmin edilen ve gerçek değerler arasındaki karesel farkların ortalamasını hesaplayarak tahmin hatalarının büyüklüğünü ölçer (Tyass vd., 2023). MAE, hataların ortalamasını alır ve sezgisel bir anlayış sunar çünkü farkların karesi alınmaz (Bilgili vd., 2022). MAPE, tahmin hatalarını yüzde cinsinden ifade ederek göreceli bir doğruluk ölçüsü sağlar (Metlek vd., 2021). RMSE ise MSE'nin karekökü olarak büyük hatalara daha fazla ağırlık verip, hataların standart sapmasını ölçer (Truong vd., 2021). R<sup>2</sup>, bağımsız değişkenlerin bağımlı değişkendeki varyansı ne ölçüde açıkladığını gösteren belirleme katsayısıdır (Elias vd., 2011).

### 3.6. Önerilen Yaklaşım

Bu çalışmada, Türkiye'nin kimyasal madde ithalatını tahmin etmek için makroekonomik değişkenler kullanılarak makine öğrenmesi ve topluluk öğrenme modelleri uygulanmıştır. Modeller, veri seti üzerinde eğitilip test edildikten sonra çapraz doğrulama ile genel performansları değerlendirilmiştir. En başarılı model kullanılarak gelecekteki ithalat değerleri tahmin edilmiştir. Bulgular, politika yapıcılar ve iş dünyası için yol gösterici bilgiler sunarak ekonomik strateji belirlemede önemli katkı sağlamaktadır. Çalışmanın genel organizasyonu ve yöntemi, kapsamlı analizi özetleyen bir şema ile Şekil 1'de sunulmaktadır.



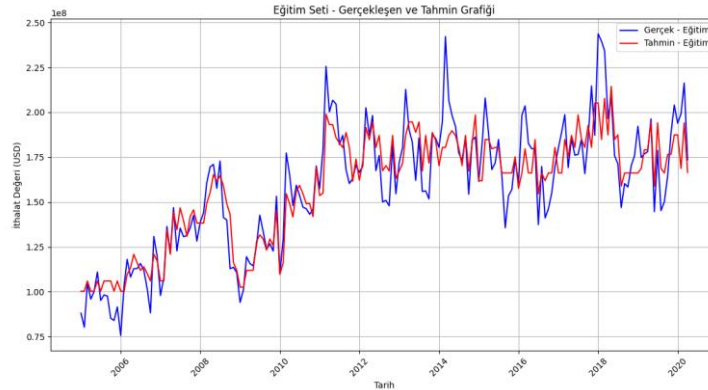


Şekil 1. Önerilen yaklaşım

Şekil 1'deki şemaya göre, Türkiye'nin kimyasal madde ithalatını tahmin etmek için çok aşamalı bir süreç izlenmiştir. İlk olarak, ortalama dolar kuru, Türkiye ithalat verileri, sanayi üretim endeksleri, petrol fiyatı ve kimyasal ürün ihracatını içeren veri seti oluşturulmuş ve %80 eğitim, %20 test olarak ayrılmıştır. Doğrusal Regresyon, Rastgele Orman, Destek Vektör Makinesi ve XGBoost modelleri kullanılarak, random search ile en iyi hiperparametre değerleri belirlenmiştir. Modeller, RMSE, MAE, MAPE, MSE ve  $R^2$  metrikleriyle değerlendirilmiş; çapraz doğrulama ile genelleme yetenekleri incelenmiştir. En iyi performansı gösteren model seçilerek, gelecek 12 aylık kimyasal madde ithalat tahmini yapılmıştır.

#### 4. Çalışmanın Bulguları

Bu bölümde, kimyasal madde ithalatı tahmininde makine öğrenimi ve topluluk öğrenme tekniklerinin performansları hakkında bir değerlendirme sunulmaktadır. Doğrusal Regresyon, Rastgele Orman, Rasyonel Kuadratik Regresyon, Destek Vektör Makinesi ve XGBoost modelleri, eğitim setindeki veriler kullanılarak eğitilmiştir. Eğitim performansları  $R^2$ , MAE, RMSE, MAPE ve MSE gibi metrikler kullanılarak değerlendirilmiştir. Eğitim sonuçlarının tahmin grafiği Şekil 2'de gösterilirken, Tablo 3'te XGBoost modelinin eğitim sonuçları gösterilmiştir.



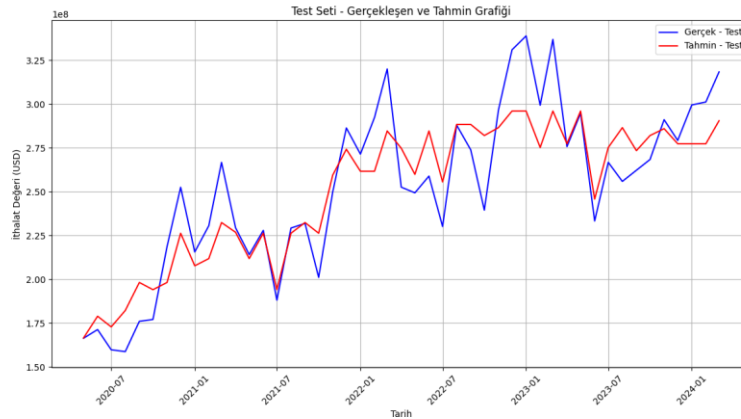
Şekil 2. XGBoost modelinin eğitim sonuçları tahmin grafiği

Şekil 2'de sunulan eğitim sonuçlarına ilişkin tahmin grafiklerine bakıldığında, XGBoost modelinin tahminleri, gerçek değerleri oldukça iyi takip etmektedir ve düşük hata oranlarına sahiptir. Model, büyük sapmaları minimize etmekte ve veri setinin trendini başarılı bir şekilde yakalamaktadır. XGBoost modelinin performansı, eğitim setinde yüksek bir uyum göstermektedir. Şekil 2'deki tahmin grafikleri, XGBoost modelinin eğitim verisi üzerinde yüksek performans gösterdiğini ve tahminlerin gerçek değerlere çok yakın olduğunu açıkça ortaya koymaktadır. Model, veri setindeki trendleri doğru bir şekilde takip etmekte ve büyük sapmaları başarılı bir şekilde minimize etmektedir. Bu durum, modelin eğitim setine oldukça iyi uyum sağladığını ve verinin genel yapısını doğru şekilde yakaladığını göstermektedir. Ancak, modelin performansının test verisinde nasıl bir seyir izleyeceği de önemlidir; bu nedenle, test setinde de benzer başarıyı sağlamak, modelin genelleme yeteneğini doğrulamak açısından kritik olacaktır.

**Tablo 3.** Modellerin Eğitim Sonuçlarının Metrik Ölçümleri

	Doğrusal Regresyon	Rastgele Orman	Rasyonel Kuadratik Regresyon	Destek Vektör Makinesi	XGBoost
<b>RMSE</b>	0.0632	0.0270	0.0001	0.0654	0.0046
<b>MSE</b>	0.0040	0.0007	0.0002	0.0042	0.0002
<b>MAE</b>	0.0503	0.0200	0.0001	0.0539	0.0035
<b>MAPE</b>	0.0753	0.0290	0.0001	0.0841	0.0057
<b>R<sup>2</sup></b>	0.9046	0.9825	0.9998	0.8978	0.9994

Tablo 3'teki sonuçlara göre, modellerin performansları karşılaştırılmıştır. Doğrusal Regresyon modeli, 0.0632 RMSE, 0.0040 MSE, 0.0503 MAE ve 0.0753 MAPE değerleriyle orta düzey performans sergilemiş ve 0.9046 R<sup>2</sup> ile varyansın %90'ını açıklamıştır. Ancak hata oranlarının yüksekliği, doğrusal regresyonun karmaşık veri setlerinde yetersiz kaldığını göstermektedir. Rastgele Orman modeli, 0.0270 RMSE, 0.0007 MSE, 0.0200 MAE ve 0.0290 MAPE ile iyi bir performans göstermiştir. 0.9825 R<sup>2</sup> değeriyle varyansın %98'den fazlasını açıklayarak, karmaşık yapıyı başarılı şekilde modellemiştir. Rasyonel Kuadratik Regresyon modeli, 0.0001 RMSE, 0.0002 MSE, 0.0001 MAE ve 0.0001 MAPE değerleriyle en iyi performansı sergilemiş, 0.9998 R<sup>2</sup> ile varyansın neredeyse tamamını açıklamıştır. Bu sonuçlar, modelin son derece düşük hata oranlarıyla üstün başarı gösterdiğini ortaya koymaktadır. Destek Vektör Makinesi, 0.0654 RMSE, 0.0042 MSE, 0.0539 MAE ve 0.0841 MAPE ile orta düzeyde kalmış ve 0.8978 R<sup>2</sup> değeriyle varyansın %90'a yakını açıklamıştır. Ancak hata oranlarının yüksekliği modelin diğer modellere göre daha düşük performansını göstermektedir. XGBoost modeli, 0.0046 RMSE, 0.0002 MSE, 0.0035 MAE ve 0.0057 MAPE değerleriyle yüksek doğruluk sağlamış, 0.9994 R<sup>2</sup> ile varyansın %99'undan fazlasını açıklamıştır. Sonuç olarak, Rasyonel Kuadratik Regresyon ve XGBoost modelleri, en düşük hata oranları ve en yüksek R<sup>2</sup> değerleri ile en başarılı performansı göstermiştir. Rastgele Orman modeli ise güvenilir bir alternatif olarak öne çıkarken, Doğrusal Regresyon ve Destek Vektör Makinesi modelleri daha düşük performans sergilemiştir. Aşağıdaki Şekil 3, modellerin test seti üzerindeki tahmin grafiklerini göstermektedir. Tablo 4'te ise test sonuçlarının metrik ölçümleri sunulmaktadır.



**Şekil 3.** XGB modelinin test sonuçları tahmin grafiği

Şekil 3'te sunulan grafiğe göre, XGBoost modeli, hem eğitim hem de test setlerinde yüksek performans göstermiş olup, en iyi genel doğruluğa sahip model olarak öne çıkmaktadır. Bu bulgular, model seçimi ve değerlendirmesinde eğitim ve test setleri arasındaki performans farklarının dikkatle incelenmesi gerektiğini vurgulamaktadır. Grafikteki bulgular, XGBoost modelinin eğitim verisi üzerinde çok iyi performans gösterdiği gibi test verisi üzerinde de benzer başarıyı sürdürebildiğini göstermektedir. Bu durum, modelin aşırı uyum yapmadığını ve daha geniş veri setlerinde de etkili olabileceğini düşündürmektedir. Modelin yüksek genel doğruluğu, doğru hiperparametre ayarları ve modelin yapısal doğruluğu sayesinde elde edilmiştir.

**Tablo 4.** Modellerin Test Sonuçlarının Metrik Ölçümleri

	Doğrusal Regresyon	Rastgele Orman	Rasyonel Kuadratik Regresyon	Destek Vektör Makinesi	XGBoost
<b>RMSE</b>	0.0686	0.0671	0.0756	0.0687	0.0653
<b>MSE</b>	0.0047	0.0045	0.0057	0.0047	0.0042
<b>MAE</b>	0.0529	0.0519	0.0587	0.0553	0.0511
<b>MAPE</b>	0.0775	0.0766	0.0850	0.0869	0.0774
<b>R<sup>2</sup></b>	0.8866	0.8915	0.8625	0.8864	0.8975

Tablo 4'teki test sonuçlarına göre, XGBoost modeli 0.0653 RMSE, 0.0042 MSE, 0.0511 MAE ve 0.0774 MAPE değerleriyle en iyi performansı göstermiş ve 0.8975 R<sup>2</sup> ile veri setindeki varyansın %89.75'ini açıklamıştır. Rastgele Orman modeli, 0.0671 RMSE, 0.0045 MSE, 0.0519 MAE ve 0.0766 MAPE ile XGBoost'a yakın bir performans sergileyerek 0.8915 R<sup>2</sup> değerine ulaşmıştır. Doğrusal Regresyon modeli, 0.0686 RMSE, 0.0047 MSE, 0.0529 MAE ve 0.0775 MAPE değerleriyle orta düzeyde performans göstermiştir (R<sup>2</sup> = 0.8866). Destek Vektör Makinesi ise benzer hata oranları (0.0687 RMSE, 0.0047 MSE) ve 0.8864 R<sup>2</sup> ile doğrusal regresyona yakın bir sonuç elde etmiştir. Rasyonel Kuadratik Regresyon modeli ise en yüksek hata oranlarına (0.0756 RMSE, 0.0057 MSE, 0.0587 MAE) ve en düşük 0.8625 R<sup>2</sup> değerine sahip olup test setinde en zayıf performansı göstermiştir. Bu bulgular, test sonuçlarının eğitim performanslarından farklılık gösterebileceğini ve genelleme yeteneklerinin değerlendirilmesinin önemini vurgulamaktadır. XGBoost ve Rastgele Orman modelleri, düşük hata oranları ve yüksek R<sup>2</sup> değerleriyle öne çıkarken, Rasyonel Kuadratik Regresyon beklenen performansı sağlayamamıştır. Çalışmada, model performansını daha güvenilir biçimde değerlendirmek için 15 katlı çapraz doğrulama uygulanmıştır. Çapraz doğrulama, modelin yeni verilere karşı performansını test ederek aşırı uyumu belirlemeye ve veri setinin tamamından yararlanmaya olanak tanımaktadır. Bu yöntem sayesinde, modellerin ortalama performansı ve istikrarı ölçülerek objektif bir karşılaştırma sağlanmıştır. Modellerin çapraz doğrulama sonuçları Tablo 5'te sunulmuştur.

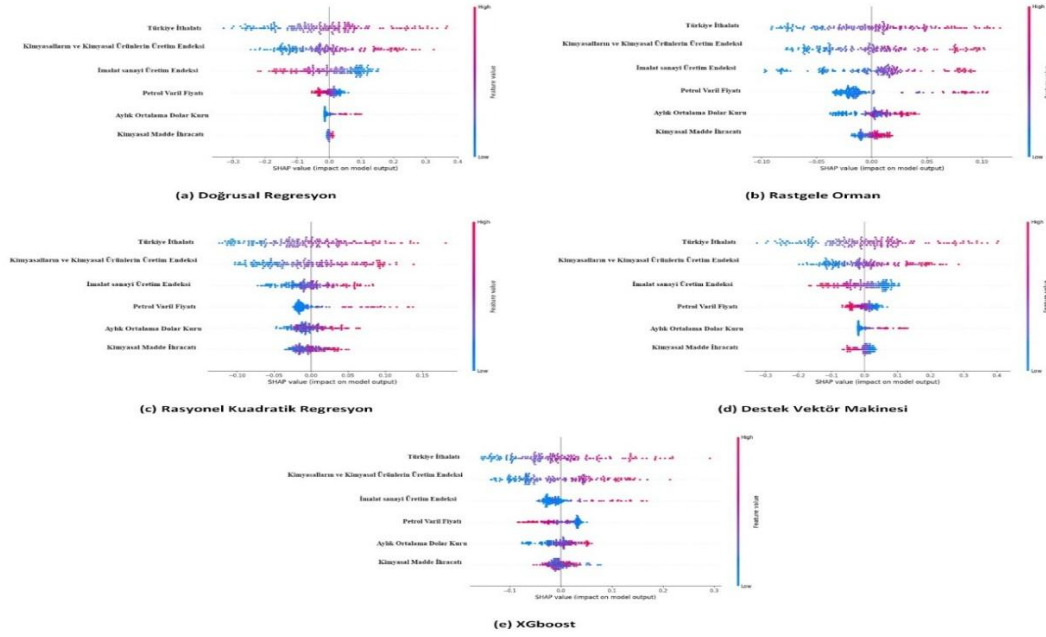
**Tablo 5.** Modellerin 15 Katlı Çapraz Doğrulama Sonuçlarının Metrik Ölçüm Sonuçları

	Doğrusal Regresyon	Rastgele Orman	Rasyonel Kuadratik Regresyon	Destek Vektör Makinesi	XGBoost
<b>RMSE</b>	0.0729	0.0718	0.0767	0.0656	0.0646
<b>MSE</b>	0.0054	0.0053	0.0061	0.0044	0.0043
<b>MAE</b>	0.0551	0.0549	0.0575	0.0548	0.0524
<b>MAPE</b>	0.0800	0.0853	0.0820	0.0797	0.0780
<b>R<sup>2</sup></b>	0.8552	0.8635	0.8399	0.8814	0.8820

Tablo 5'teki 15 katlı çapraz doğrulama sonuçlarına göre, XGBoost modeli en düşük hata oranlarına (0.0646 RMSE, 0.0043 MSE, 0.0524 MAE, 0.0780 MAPE) ve 0.8820 R<sup>2</sup> ile en iyi performansı sergilemiştir. Destek Vektör Makinesi, 0.0656 RMSE, 0.0044 MSE ve 0.8814 R<sup>2</sup> ile XGBoost'u yakından takip ederken, Rastgele Orman modeli 0.0718 RMSE, 0.0053 MSE ve 0.8635 R<sup>2</sup> ile güçlü bir performans göstermiştir. Doğrusal Regresyon, 0.0729 RMSE ve 0.8552 R<sup>2</sup> ile orta düzeyde performans sunarken, Rasyonel Kuadratik Regresyon en yüksek hata oranlarına (0.0767 RMSE, 0.0061 MSE) ve 0.8399 R<sup>2</sup> ile en düşük performansa sahiptir. Bu bulgular, XGBoost'un doğruluk ve genelleme yeteneği açısından en üstün model olduğunu göstermektedir.

Rastgele Orman ve Destek Vektör Makinesi modelleri güvenilir alternatifler sunarken, Rasyonel Kuadratik Regresyon modelinin düşük test ve çapraz doğrulama performansı aşırı uyum sorununu işaret etmektedir. Bu sonuçlar, model değerlendirmelerinde eğitim, test ve çapraz doğrulama yöntemlerinin önemini vurgulamaktadır. XGBoost, diğer modellere kıyasla gradyan artırma algoritmasının hataları iteratif olarak düzeltme yeteneği ve hiperparametre optimizasyonundaki esnekliği sayesinde üstün performans göstermiştir. Özellikle doğrusal olmayan ilişkileri etkili bir şekilde öğrenmesi ve kritik değişkenlerin etkisini dengeli bir şekilde değerlendirmesi, onu Rastgele Orman ve Destek Vektör Makinesi gibi güçlü alternatiflerden ayırmıştır. Ayrıca, eğitim ve test setlerindeki düşük hata oranları, modelin genelleme yeteneğini ve aşırı uyum riskine karşı dayanıklılığını ortaya koymuştur.

Bu çalışmada, modellerin tahminlerinde bağımsız değişkenlerin önemini belirlemek için SHAP (SHapley Additive exPlanations) analizi uygulanmıştır. SHAP, her bir değişkenin model tahminlerine katkısını açıklayarak karar mekanizmalarını şeffaf hale getiren bir tekniktir (Cao, 2024). Bu analiz, değişkenlerin göreceli etkilerini ortaya koyarak model değerlendirme sürecinde önemli avantajlar sağlar ve yalnızca performans metrikleriyle değil, modelin karar süreçleriyle de güvenilirliğini artırır (Broeck vd., 2022; Yap vd., 2021). Özellikle karmaşık veri setlerinde en etkili değişkenleri belirlemek, model performansını iyileştirme açısından kritik ipuçları sunar. Ayrıca, SHAP analizi, açıklanabilirlik ve şeffaflık sağlayarak modellerin gerçek dünya uygulamalarında daha güvenilir ve kabul edilebilir olmasına katkıda bulunur (Zhu vd., 2024). Şekil 4'te tüm modellerin test seti üzerindeki SHAP analizi sonuçları gösterilmektedir.



Şekil 4. Tüm modellerin test seti üzerindeki SHAP analizi sonuçları

Şekil 4'teki SHAP analizi sonuçlarına göre, Türkiye İthalatı ve Kimyasallar ve Kimyasal Ürünler Üretim Endeksi değişkenleri tüm modellerde en yüksek etkiye sahip olup, tahminlere en büyük katkıyı sağlamaktadır. Doğrusal Regresyon ve Rastgele Orman modellerinde İmalat Sanayi Üretim Endeksi ve Petrol Varil Fiyatı da önemli etki gösterirken, XGBoost modeli tüm değişkenlerin etkisini dengeli dağıtarak özellikle Türkiye İthalatı ve Kimyasal Ürünler Üretim Endeksi için yüksek SHAP değerleri elde etmiştir. Destek Vektör Makinesi modelinde Aylık Ortalama Dolar Kuru değişkeninin etkisi belirginleşirken, Rasyonel Kuadratik Regresyon modeli yine Türkiye İthalatı ve Kimyasal Ürünler Üretim Endeksi değişkenlerini öne çıkarmıştır. Bu bulgular, söz konusu değişkenlerin hurda demir çelik ithalatı tahminlerinde kritik rol oynadığını ve SHAP analizinin modellerin karar mekanizmalarını şeffaflaştırarak güvenilirliklerini artırdığını göstermektedir. Çalışmanın bir sonraki aşamasında, kimyasal madde ithalatı tahmini için en iyi model olarak belirlenen XGBoost kullanılmıştır. Tahminlerin yapılabilmesi için bağımsız değişkenlerin gelecekteki değerleri de tahmin

edilmiştir. Bu değerler, XGBoost modeli ile tahmin edilip veri setine dahil edilerek, gelecek 12 ayın ithalat değerleri tahmin edilmiştir. Gelecek 12 aylık tahmin değerleri Tablo 6'da yer almaktadır.

**Tablo 6.** XGBoost Modeli ile Bağımsız Değişkenlerin Gelecek 12 Aylık Tahmin Sonuçları

Dönem	Aylık Dolar Kuru	Türkiye İthalatı	İmalat Sanayi Üretim Endeksi	Kimyasalların ve Kimyasal Ürünlerin Üretimi Endeksi	Petrol Varil Fiyatı	Kimyasal Madde İhracatı
2024-4	32,41	25,804,759,060	96,6	84,74	93,4	110,381,882
2024-5	33,27	31,719,102,429	111,9	75,47	100,4	116,719,793
2024-6	34,52	24,462,028,943	103,4	74,84	90,9	109,600,980
2024-7	35,63	29,683,466,070	106,6	80,11	89,9	99,806,107
2024-8	36,70	28,288,467,798	108,5	86,15	96,2	116,993,348
2024-9	37,85	25,720,133,585	113	93,72	97	108,452,445
2024-10	38,76	27,552,210,172	116,8	90,6	98,1	112,929,900
2024-11	39,11	27,474,951,128	113,8	82,94	97	110,752,726
2024-12	39,77	27,314,796,667	125,1	77,76	94,7	106,238,677
2025-1	40,68	24,771,192,383	101,2	80,12	98,4	94,022,509
2025-2	41,47	26,298,371,418	105,1	83,48	100,2	112,601,415
2025-3	41,83	28,004,749,771	113,5	85,41	108,2	118,507,309

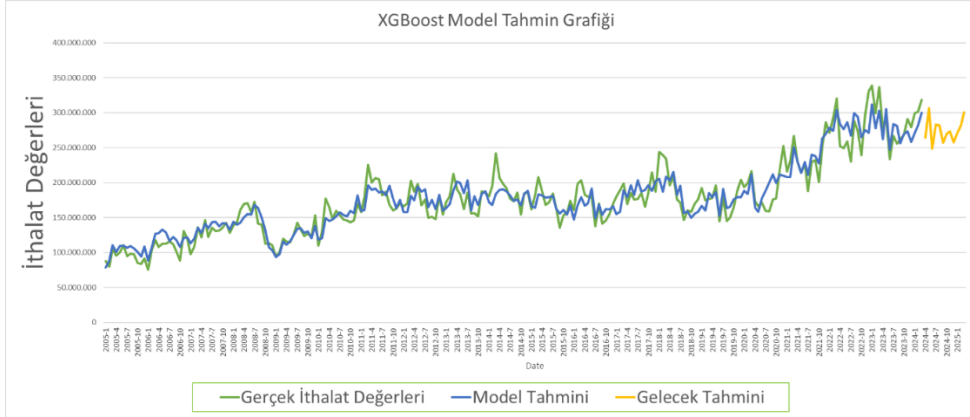
Tablo 6, XGBoost modeli ile yapılan gelecek 12 aylık tahmin sonuçlarını sunmaktadır. Aylık Dolar Kuru 2024 Nisan'da 32,41 olarak tahmin edilirken, 2025 Mart'ta 41,83'e yükselmesi öngörülmektedir. Türkiye İthalatı, Nisan 2024'te 25,8 milyar ABD doları seviyesinden Mayıs 2024'te 31,7 milyar ABD doları'na çıkması beklenmektedir. İmalat Sanayi Üretim Endeksi Nisan 2024'te 96,6 olarak tahmin edilirken, Aralık 2024'te 125,1 seviyesine ulaşacağı öngörülmektedir. Kimyasalların ve Kimyasal Ürünlerin Üretimi Endeksi, Nisan 2024'te 84,74'ten Eylül 2024'te 93,72'ye yükselmesi beklenmektedir. Petrol Varil Fiyatı, Nisan 2024'te 93,4 ABD doları seviyesinden 2025 Mart'ta 108,2 ABD doları'na çıkacağı tahmin edilmiştir. Kimyasal Madde İhracatı ise Nisan 2024'te 110,4 milyon ABD doları'ndan Mart 2025'te 118,5 milyon ABD doları'na yükselmesi öngörülmektedir. Bu tahminler, döviz kuru, üretim endeksleri ve petrol fiyatlarındaki değişimlerin ekonomik ve sektörel dinamikleri anlamak için önemli bilgiler sunmaktadır. XGBoost model kullanımı ile kimyasal madde ithalatının gelecek 12 ayına yönelik tahminler yapılmıştır. Bu amaçla, bağımsız değişkenlerin tahmini değerleri veri setine eklenmiş ve modelin öğrenme kapasitesinden yararlanılarak daha tutarlı tahminler elde edilmiştir. Bu yaklaşım, gelecekteki ithalat değerleri hakkında daha güvenilir öngörüler sunmaktadır. Tablo 7'de bu tahmin sonuçları gösterilmektedir.

**Tablo 7.** Kimyasal Madde İthalatı Tahmin Değerleri

Dönem	Kimyasal İthalatı (USD)	Madde	Dönem	Kimyasal İthalatı (USD)	Madde
2024-4	263,923,947		2024-10	270,043,710	
2024-5	306,606,515		2024-11	273,491,311	
2024-6	248,317,483		2024-12	257,422,155	
2024-7	283,379,514		2025-1	271,035,758	
2024-8	281,767,028		2025-2	281,955,669	
2024-9	256,774,041		2025-3	300,555,026	

Tablo 7, XGBoost modeli ile Türkiye'nin kimyasal madde ithalatına yönelik 12 aylık tahminleri sunmaktadır. 2024 Nisan ayında 263,9 milyon ABD doları olarak tahmin edilen ithalat, Mayıs ayında 306,6 milyon ABD doları ile belirgin bir artış göstermekte, Haziran ayında ise 248,3 milyon ABD doları'na düşmektedir. Temmuz ve Ağustos aylarında ithalat sırasıyla 283,4 milyon ve 281,8 milyon ABD doları olarak tahmin edilirken, Eylül ayında 256,8 milyon ABD doları'na gerilemektedir. Ekim ve Kasım aylarında ithalat 270 milyon ve 273,5 milyon ABD doları seviyelerine yükselerek yılın en yüksek değerlerine ulaşmaktadır. 2025 Ocak ayında 271 milyon ABD doları olarak öngörülen ithalat, Şubat ayında 282 milyon ABD doları, Mart ayında ise 300,6 milyon ABD doları'na çıkarak tekrar artış göstermektedir. Bu tahminler, ithalatta dalgalı bir seyir izleneceğini ve özellikle Mayıs ile Mart aylarında belirgin artışlar yaşanacağını göstermektedir. Dalgalanmalar, döviz kuru

değişimleri, küresel talep artışı ve yerel ekonomik koşullara bağlı olabilir. Şekil 5'te, yeşil renk gerçek değerleri, mavi renk geçmiş tahminleri, sarı renk ise gelecekteki 12 aylık ithalat tahminlerini göstermektedir. Bu tahminler, politika yapıcılar ve işletmeler için stratejik planlama açısından önemli bilgiler sunmaktadır.



Şekil 5. XGBoost modeli ile kimyasal madde ithalat değerlerinin tahminleri ve gelecek 12 aylık öngörüler

### 5.Sonuç, Tartışma ve Öneriler

Bu çalışma, Türkiye'nin kimyasal madde ithalatını tahmin etmek için makroekonomik değişkenler kullanarak Doğrusal Regresyon, Rastgele Orman, Rasyonel Kuadratik Regresyon, Destek Vektör Makinesi (DVM) ve XGBoost modellerini değerlendirmiştir. Analizlerde XGBoost modeli, en düşük hata oranları ve en yüksek R<sup>2</sup> değerleri ile en başarılı performansı sergilemiştir. Rastgele Orman modeli de yüksek doğruluk ve düşük hata oranlarıyla güvenilir sonuçlar sunarken, Rasyonel Kuadratik Regresyon modeli aşırı uyum nedeniyle test setinde beklenen performansı gösterememiştir. Destek Vektör Makinesi ve Doğrusal Regresyon ise orta düzeyde performans sergileyerek karmaşık veri setlerinde yetersiz kalmıştır. SHAP analizi sonuçları, Türkiye İthalatı ve Kimyasallar ve Kimyasal Ürünler Üretim Endeksi değişkenlerinin tüm modellerde en yüksek etkiye sahip olduğunu ortaya koymuştur. Bu bulgular, makine öğrenmesi yöntemlerinin kimyasal madde ithalatı tahmininde geleneksel yöntemlere kıyasla üstün performans sağladığını ve elde edilen sonuçların ekonomik stratejiler için yol gösterici nitelikte olduğunu göstermektedir.

Bu çalışmanın bulguları, literatürdeki benzer çalışmalarla uyumludur. Broeren vd. (2014), temel kimyasal üretiminde bölgesel maliyetler ve talep artışını dikkate alarak enerji verimliliği ve iklim politikalarının önemini vurgulamıştır. Benzer şekilde, bu çalışmada Türkiye İthalatı ve Kimyasallar ve Kimyasal Ürünler Üretim Endeksi'nin kimyasal madde ithalatı üzerindeki kritik etkisi ortaya konmuştur. Li ve Li (2022), PSO-BP sınır ağı modeliyle Çin'in kimyasal ihracat tahmininde yüksek doğruluk elde etmiş; bu çalışmada da XGBoost modeli benzer şekilde karmaşık ithalat tahminlerinde üstün performans göstermiştir. Pisuttinuart vd. (2022) ise Tayland'ın kimyasal gübre ithalatını SARIMA modeliyle analiz etmiş ve talep değişkenliklerini incelemiştir. Bu çalışmada ise daha geniş bir makroekonomik değişken seti kullanılarak modeller karşılaştırılmış ve kimyasal madde ithalatı tahmin edilmiştir. Modellerin test ve çapraz doğrulama sonuçları karşılaştırıldığında, XGBoost modelinin eğitim, test ve çapraz doğrulama setlerinde tutarlı performansı, genelleme yeteneğini ve güvenilirliğini kanıtlamaktadır. Buna karşın, doğrusal regresyon ve Rasyonel Kuadratik Regresyon gibi modellerde görülen test ve çapraz doğrulama farklılıkları, bu modellerin aşırı uyum riskine ve karmaşık ilişkileri yeterince öğrenememesine işaret etmektedir. Bu durum, modellerin genelleme kapasitesini değerlendirirken test ve çapraz doğrulama sonuçlarının kritik önemini vurgulamaktadır.

Bu çalışma, Türkiye'nin kimyasal madde ithalatını tahmin etmek için makine öğrenmesi ve topluluk öğrenme yöntemlerini kullanarak literatüre önemli bir katkı sağlamaktadır. Türkiye özelinde geniş bir veri seti ve çeşitli makroekonomik değişkenlerin kullanılması, çalışmanın yenilikçi yönüdür. Daha önceki çalışmalar genellikle tek bir model veya belirli kimyasal ürünlere odaklanırken, bu çalışma farklı yöntemleri karşılaştırarak en iyi tahmin modelini belirlemiştir. COVID-19 pandemisi (2020-2021), enerji fiyatlarındaki artış (2021-2022) ve Rusya-Ukrayna savaşı (2022) gibi olağandışı olaylar, ithalat hacimlerinde ani değişimlere yol açmış ve XGBoost modeli bu karmaşık desenleri başarıyla yakalamıştır. Modelin gradyan artırma algoritması sayesinde ani değişimlere uyum sağlama yeteneği ve geniş makroekonomik değişken setini dikkate alması, tahmin

doğruluğunu artırmıştır. Ayrıca, hiperparametre optimizasyonu ile aşırı uyum sorunu önlenmiş, modelin genel performansı iyileştirilmiştir. Bu bulgular, XGBoost'un esnekliği ve karmaşık veri ilişkilerini öğrenme kapasitesini vurgularken, çalışma politika yapıcılar için stratejik planlamada önemli bir referans niteliği taşımaktadır. Bu çalışmanın bazı kısıtlılıkları bulunmaktadır. Rasyonel Kuadratik Regresyon gibi bazı modeller, aşırı uyum sorununa yatkın olup genelleme yeteneğini zayıflatabilir, özellikle küçük veri setlerinde bu durum daha belirgin hale gelir. Ayrıca, modellerin eğitimi yüksek hesaplama gücü gerektirdiğinden, donanım ve yazılım kaynaklarının yetersizliği büyük veri setleri üzerinde kısıt oluşturabilir. Çalışma, XGBoost ve Rastgele Orman modellerinin üstün performansını vurgularken, model seçiminde dikkatli olunması ve hiperparametre optimizasyonunun titizlikle yapılması gerektiğini önermektedir. Türkiye İthalatı ve Kimyasallar ve Kimyasal Ürünler Üretimi Endeksi gibi kritik değişkenlerin dahil edilmesi önemlidir. Gelecekteki çalışmalar, daha geniş veri setleri kullanarak performansı artırmalı, bağımsız değişken çeşitliliğini artırmalı ve farklı sektör verilerini entegre etmelidir. Ani ekonomik ve politik değişimlere karşı senaryo analizleri ve stres testleri yapılmalı, esnek ve dayanıklı modeller geliştirilmelidir. Ayrıca, veri analitiği konusunda sürekli eğitim, uluslararası işbirliği ve veri paylaşımı teşvik edilmeli; büyük veri setlerinin işlenmesi için yüksek performanslı bilgisayar sistemlerine ve bulut tabanlı çözümlere yatırım yapılmalıdır. Bu öneriler, Türkiye'nin kimyasal madde ithalatı tahminlerinde model doğruluğunu ve güvenilirliğini artırmaya katkı sağlayacaktır. Kimyasal madde ithalat tahminleri; döviz rezerv yönetimi, dış ticaret dengesi ve stratejik ithalat politikalarının oluşturulmasında önemli rol oynamaktadır. Ayrıca, bu tahminler sanayi planlaması, hammadde tedariki ve maliyet optimizasyonu süreçlerinde işletmelere rehberlik edebilir. Elde edilen tahmin sonuçları, ekonomik ve ticari planlamalarda kullanılmalıdır. Politika yapıcılar, bu tahminleri dikkate alarak, dış ticaret politikalarını ve stratejik planlamalarını geliştirebilirler. İş dünyası ise tedarik zinciri yönetimi, stok kontrolü ve maliyet optimizasyonu konularında bu tahminlerden faydalanabilir. Sonuç olarak, bu çalışma, makine öğrenmesi ve topluluk öğrenme yöntemlerinin kimyasal madde ithalatı tahminlerinde etkin bir şekilde kullanılabileceğini ve bu yöntemlerin doğru tahminler sağladığını göstermektedir. Bu bulgular, politika yapıcılar ve iş dünyası için değerli bilgiler sunmakta ve gelecekteki çalışmalar için sağlam bir temel oluşturmaktadır.

#### **Kaynaklar**

- Al Marzooqi, F. I. and Redouane, A. (2024). Predicting real estate prices using machine learning in abu dhabi. *Iraqi Journal of Science*, 1689-1706. <https://doi.org/10.24996/ij.s.2024.65.3.40>
- Bayrak, T. (2020). A machine-learning-based model for forecasting medical device foreign trade. *Eskişehir Technical University Journal of Science and Technology a - Applied Sciences and Engineering*, 21(4), 477-485. <https://doi.org/10.18038/estubtda.803546>
- Bhagwat, A., Baets, B. D., Steen, A., Vlaeminck, B., & Fievez, V. (2012). Prediction of ruminal volatile fatty acid proportions of lactating dairy cows based on milk odd- and branched-chain fatty acid profiles: new models, better predictions. *Journal of Dairy Science*, 95(7), 3926-3937. <https://doi.org/10.3168/jds.2011-4850>
- Biau, G. and Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197-227. <https://doi.org/10.1007/s11749-016-0481-7>
- Bilgili, M., Keiyinci, S., & Ekinci, F. (2022). One-day ahead forecasting of energy production from run-of-river hydroelectric power plants with a deep learning approach. *Scientia Iranica*, 0(0), 0-0. <https://doi.org/10.24200/sci.2022.58636.5825>
- Brabenc, T. and Šuleř, P. (2020). Machine learning forecasting of cr and prc balance of trade. *SHS Web of Conferences*, 73, 01004. <https://doi.org/10.1051/shsconf/20207301004>
- Broeck, G. V. d., Lykov, A., Schleich, M., & Suci, D. (2022). On the tractability of shap explanations. *Journal of Artificial Intelligence Research*, 74, 851-886. <https://doi.org/10.1613/jair.1.13283>
- Broeren, M. L. M., Saygin, D., & Patel, M. K. (2014). Forecasting global developments in the basic chemical industry for environmental policy analysis. *Energy Policy*, 64, 273-287. <https://doi.org/10.1016/j.enpol.2013.09.025>
- Cao, S. and Hu, Y. (2024). Interpretable machine learning framework to predict gout associated with dietary fiber and triglyceride-glucose index. *Nutrition & Metabolism*, 21(1). <https://doi.org/10.1186/s12986-024-00802-2>
- Cao, T., She, D., Zhang, X., & Yang, Z. (2022). Understanding the influencing factors and mechanisms (land use changes and check dams) controlling changes in the soil organic carbon of typical loess watersheds in

- china. *Land Degradation & Development*, 33(16), 3150-3162. <https://doi.org/10.1002/ldr.4378>
- Chen, T. and Guestrin, C. (2016). Xgboost: a scalable tree boosting system.. <https://doi.org/10.48550/arxiv.1603.02754>
- Crescimanno, M., Galati, A., & Bal, T. (2014). The role of the economic crisis on the competitiveness of the agri-food sector in the main mediterranean countries. *Agricultural Economics (Zemědělská Ekonomika)*, 60(2), 49-64. <https://doi.org/10.17221/59/2013-agricecon>
- Dutschmann, T. and Baumann, K. (2021). Evaluating high-variance leaves as uncertainty measure for random forest regression. *Molecules*, 26(21), 6514. <https://doi.org/10.3390/molecules26216514>
- Eberly, L. E. (2007). Multiple linear regression. *Topics in Biostatistics*, 165-187. [https://doi.org/10.1007/978-1-59745-530-5\\_9](https://doi.org/10.1007/978-1-59745-530-5_9)
- Elias, R., Fang, L., & Wahab, M. (2011). Electricity load forecasting based on weather variables and seasonalities: a neural network approach. *Icsssm11*. <https://doi.org/10.1109/icsssm.2011.5959472>
- EMİRHAN, P. N. and TURGUTLU, E. (2023). İşgücü talebi uluslararası ticarete tepki veriyor mu? Türk imalat sektöründen kanıtlar. *Öneri Dergisi*, 18(59), 187-201. <https://doi.org/10.14783/marueri.1075714>
- Erdem, Z. B. (2010). The assessment of coal's contribution to sustainable energy development in Turkey. *Energy Exploration & Exploitation*, 28(2), 117-129. <https://doi.org/10.1260/0144-5987.28.2.117>
- Erduman, Y., Eren, O., & Gül, S. (2020). Import Content of Turkish Production And Exports: A Sectoral Analysis. *Central Bank Review*, 20(4), 155-168. <https://doi.org/10.1016/j.cbrev.2020.07.001>
- Fantke, P. and Ernstoff, A. (2017). Lca of chemicals and chemical products. *Life Cycle Assessment*, 783-815. [https://doi.org/10.1007/978-3-319-56475-3\\_31](https://doi.org/10.1007/978-3-319-56475-3_31)
- Ferrat, L. A., Goodfellow, M., & Terry, J. R. (2018). Classifying dynamic transitions in high dimensional neural mass models: a random forest approach. *PLOS Computational Biology*, 14(3), e1006009. <https://doi.org/10.1371/journal.pcbi.1006009>
- Flyvbjerg, B. (2006). From nobel prize to project management: getting risks right. *Project Management Journal*, 37(3), 5-15. <https://doi.org/10.1177/875697280603700302>
- Ge, X., Ding, J., Jin, X., Wang, J., Chen, X., Li, X., ... & Xie, B. (2021). Estimating agricultural soil moisture content through uav-based hyperspectral images in the arid region. *Remote Sensing*, 13(8), 1562. <https://doi.org/10.3390/rs13081562>
- Gür, Y. E., & Eşidir, K. A. (2024). Türkiye Hurda Demir Çelik İthalatının Gelecek Değerlerinin Derin Öğrenme, Makine Öğrenmesi ve Topluluk Öğrenme Yöntemleri ile Öngörülmesi. *Alanya Akademik Bakış*, 8(3), 885-908. <https://doi.org/10.29023/alanyaakademik.1497646>
- Gür, Y. E. (2024). Stock Price Forecasting Using Machine Learning and Deep Learning Algorithms: A Case Study for the Aviation Industry. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 36(1), 25-34. <https://doi.org/10.35234/fumbd.1357613>
- Gür, Y. E. (2024). Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach. *Data Science in Finance and Economics*, 4(4), 469-513. <https://doi.org/10.3934/DSFE.2024020>
- Gür, Y. E. (2024). Innovation in the dairy industry: forecasting cow cheese production with machine learning and deep learning models. *International Journal of Agriculture Environment and Food Sciences*, 8(2), 327-346. <https://doi.org/10.31015/jaefs.2024.2.9>
- Huertas-Tato, J. and Brito, M. (2018). Using smart persistence and random forests to predict photovoltaic energy production. *Energies*, 12(1), 100. <https://doi.org/10.3390/en12010100>
- Hui, D., Wang, J., Le, X., Shen, W., & Ren, H. (2012). Influences of biotic and abiotic factors on the relationship between tree productivity and biomass in china. *Forest Ecology and Management*, 264, 72-80. <https://doi.org/10.1016/j.foreco.2011.10.012>
- Ji, H., Fu, X., & Cheng, K. (2014). Engineering drawing man-hour forecasting based on bp-ga in design of chemical equipment. *2014 20th International Conference on Automation and Computing*. <https://doi.org/10.1109/iconac.2014.6935487>
- Jošić, H. and Žmuk, B. (2022). A machine learning approach to forecast international trade: the case of croatia. *Business Systems Research Journal*, 13(3), 144-160. <https://doi.org/10.2478/bsrj-2022-0030>
- Kishi, H. and Sekine, Y. (2003). Simple prediction of atmospheric concentration of hydrophilic compounds based on the classification of industrial uses. *QSAR & Combinatorial Science*, 22(3), 396-398. <https://doi.org/10.1002/qsar.200390029>
- Li, G., Yu, Z., Zheng, B., Qi, B., Su, Z., & Wang, D. (2023). Voltage sag source location based on the random forest. *Journal of Physics: Conference Series*, 2584(1), 012144. <https://doi.org/10.1088/1742-6596/2584/1/012144>
- Li, N., & Li, M. (2022). Forecast of Chemical Export Trade Based on PSO-BP Neural Network Model. *Journal*



- of Mathematics*, 2022(1), 1487746. <https://doi.org/10.1155/2022/1487746>
- Liang, W., Luo, S., Zhao, G., & Wu, H. (2020). Predicting hard rock pillar stability using gbdt, xgboost, and lightgbm algorithms. *Mathematics*, 8(5), 765. <https://doi.org/10.3390/math8050765>
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19-40. <https://doi.org/10.1037/1082-989x.7.1.19>
- Marill, K. A. (2004). Advanced statistics: linear regression, part ii: multiple linear regression. *Academic Emergency Medicine*, 11(1), 94-102. <https://doi.org/10.1111/j.1553-2712.2004.tb01379.x>
- Medeiros, M. C. (2022). Forecasting With Machine Learning Methods. *Advanced Studies in Theoretical and Applied Econometrics*, 111-149. [https://doi.org/10.1007/978-3-031-15149-1\\_4](https://doi.org/10.1007/978-3-031-15149-1_4)
- Meng, F., Weng, K., Shallal, B., Chen, X., & Mourshed, M. (2018). Forecasting algorithms and optimization strategies for building energy management & demand response. Sp 2018. <https://doi.org/10.3390/proceedings2151133>
- Metlek, S., Kandilli, C., & Kayaalp, K. (2021). Prediction of the effect of temperature on electric power in photovoltaic thermal systems based on natural zeolite plates. *International Journal of Energy Research*, 46(5), 6370-6382. <https://doi.org/10.1002/er.7575>
- Mishra, S., Srivastava, R., Muhammad, A., Amit, A., Chiavazzo, E., Fasano, M., ... & Asinari, P. (2023). The impact of physicochemical features of carbon electrodes on the capacitive performance of supercapacitors: a machine learning approach. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-33524-1>
- Moon, J., Ke, F., & Sokolikj, Z. (2020). Automatic assessment of cognitive and emotional states in virtual reality-based flexibility training for four adolescents with autism. *British Journal of Educational Technology*, 51(5), 1766-1784. <https://doi.org/10.1111/bjet.13005>
- Moshiri, S. and Kheirandish, E. (2023). Global impacts of oil price shocks: the trade effect. *Journal of Economic Studies*, 51(1), 126-144. <https://doi.org/10.1108/jes-08-2022-0455>
- Nas, S., Akboz Caner, A., & Ergin Ünal, A. (2024). Türkiye'de Enflasyon Oranlarının Makine Öğrenme Yöntemi ile Tahmini. *Gaziantep Üniversitesi Sosyal Bilimler Dergisi*, 23(3), 1029-1045. <https://doi.org/10.21547/jss.1371005>
- Naseem, M., Chaudhary, K., Sharma, B. N., & Lal, A. G. (2019). Using ensemble decision tree model to predict student dropout in computing science. *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. <https://doi.org/10.1109/csde48274.2019.9162389>
- Özemre, M. and Kabadurmuş, Ö. (2020). A big data analytics based methodology for strategic decision making. *Journal of Enterprise Information Management*, 33(6), 1467-1490. <https://doi.org/10.1108/jeim-08-2019-0222>
- Palkovits, R. and Palkovits, S. (2019). Using artificial intelligence to forecast water oxidation catalyts. *ACS Catalysis*, 9(9), 8383-8387. <https://doi.org/10.1021/acscatal.9b01985>
- Pisutinusart, C., Jatuporn, C., Suvanvihok, V., & Seerasarn, N. (2022). Forecasting the import demand for chemical fertilizer in Thailand. *The EURASEANs: journal on global socio-economic dynamics*, (3 (34)), 61-70. [https://doi.org/10.35678/2539-5645.3\(34\).2022.61-70](https://doi.org/10.35678/2539-5645.3(34).2022.61-70)
- Polat, K., Şentürk, Ü. K., & Arıcan, M. (2023). A novel cuffless blood pressure prediction: uncovering new features and new hybrid ml models. *Diagnostics*, 13(7), 1278. <https://doi.org/10.3390/diagnostics13071278>
- Renas Rajab Asaad and M. Abdulazeez, A. (2024). Comprehensive classification of iris flower species: a machine learning approach. *Indonesian Journal of Computer Science*, 13(1). <https://doi.org/10.33022/ijcs.v13i1.3717>
- Sercu, P. and Uppal, R. (2003). Exchange rate volatility and international trade: a general-equilibrium analysis. *European Economic Review*, 47(3), 429-441. [https://doi.org/10.1016/s0014-2921\(01\)00175-1](https://doi.org/10.1016/s0014-2921(01)00175-1)
- Sevim, A., Demir, İ., Höfte, M., Humber, R. A., & Demirbağ, Z. (2009). Isolation and characterization of entomopathogenic fungi from hazelnut-growing region of Turkey. *BioControl*, 55(2), 279-297. <https://doi.org/10.1007/s10526-009-9235-8>
- Svetnik, V., Liaw, A., Tong, C., Culberson, J., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and qsar modeling. *Journal of Chemical Information and Computer Sciences*, 43(6), 1947-1958. <https://doi.org/10.1021/ci034160g>
- Şener, S., Savrul, M., & Aydın, O. (2014). Structure of small and medium-sized enterprises in turkey and global competitiveness strategies. *Procedia - Social and Behavioral Sciences*, 150, 212-221. <https://doi.org/10.1016/j.sbspro.2014.09.119>
- Şimşek, A. I., Bulut, E., Gur, Y. E., & Tarla, E. G. (2024). A novel approach to Predict WTI crude spot oil price: LSTM-based feature extraction with Xgboost Regressor. *Energy*, 309, 133102. <https://doi.org/10.1016/j.energy.2024.133102>

- Şimşek, A. I. (2024). Forecasting consumer price index using macroeconomic variables: a comparative analysis of machine learning and deep learning approaches. *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (28), 15-29. <https://doi.org/10.29029/busbed.1394983>
- Truong, N., Ngo, N., & Pham, A. (2021). Forecasting time-series energy data in buildings using an additive artificial intelligence model for improving energy efficiency. *Computational Intelligence and Neuroscience*, 2021, 1-12. <https://doi.org/10.1155/2021/6028573>
- Türkiye Cumhuriyet Merkez Bankası Elektronik Veri Dağıtım Sistemi (TCMB-EVDS), <https://evds2.tcmb.gov.tr/index.php?/evds/serieMarket>, , Erişim Tarihi: 22.05.2024.
- Türkiye İstatistik Kurumu (2024), [www.tuik.gov.tr](http://www.tuik.gov.tr), Erişim Tarihi: 15.05.2024.
- Tyass, I., Khalili, T., Rafik, M., Bellat, A., Raihani, A., & Mansouri, K. (2023). Wind speed prediction based on statistical and deep learning models. *International Journal of Renewable Energy Development*, 12(2), 288-299. <https://doi.org/10.14710/ijred.2023.48672>
- Walker, J. D., Dimitrov, S., & Mekenyan, O. G. (2003). Using hpv chemical data to develop qsars for non-hpv chemicals: opportunities to promote more efficient use of chemical testing resources. *QSAR & Combinatorial Science*, 22(3), 386-395. <https://doi.org/10.1002/qsar.200390028>
- Wang, L., Wang, X., Chen, A., Jin, X., & Che, H. (2020). Prediction of type 2 diabetes risk and its effect evaluation based on the xgboost model. *Healthcare*, 8(3), 247. <https://doi.org/10.3390/healthcare8030247>
- Wang, N., Liu, W., Sun, S., & Wang, Q. (2021). The influence of complexity of imported products on total factor productivity. *Mobile Information Systems*, 2021, 1-7. <https://doi.org/10.1155/2021/3384068>
- Wang, Y. (2022). Import and export trade forecasting algorithm based on blockchain security and pso optimized hybrid rvm model. *Security and Privacy*, 6(2). <https://doi.org/10.1002/spy2.218>
- Wen, B., Li, R., Zhao, X., Ren, S., Chang, Y., Zhang, K., ... & Zhu, X. (2021). A quadratic regression model to quantify plantation soil factors that affect tea quality. *Agriculture*, 11(12), 1225. <https://doi.org/10.3390/agriculture11121225>
- Wen, J., Zhang, Y., Yang, G., He, Z., & Zhang, W. (2019). Path loss prediction based on machine learning methods for aircraft cabin environments. *IEEE Access*, 7, 159251-159261. <https://doi.org/10.1109/access.2019.2950634>
- Wie, Y. M., Lee, K. G., Lee, K. H., Ko, T. S., & Lee, K. H. (2020). The experimental process design of artificial lightweight aggregates using an orthogonal array table and analysis by machine learning. *Materials*, 13(23), 5570. <https://doi.org/10.3390/ma13235570>
- Yap, M., Johnston, R. L., Foley, H., MacDonald, S., Kondrashova, O., Tran, K., ... & Waddell, N. (2021). Verifying explainability of a deep learning tissue classifier trained on rna-seq data. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-81773-9>
- Zareipour, H., Cañizares, C. A., & Bhattacharya, K. (2010). Economic impact of electricity market price forecasting errors: a demand-side analysis. *IEEE Transactions on Power Systems*, 25(1), 254-262. <https://doi.org/10.1109/tpwrs.2009.2030380>
- Zhang, Z., Xin, Q., & Li, W. (2021). Machine learning-based modeling of vegetation leaf area index and gross primary productivity across north america and comparison with a process-based model. *Journal of Advances in Modeling Earth Systems*, 13(10). <https://doi.org/10.1029/2021ms002802>
- Zhu, C., Liu, X., & Chen, D. (2024). Prediction of digital transformation of manufacturing industry based on interpretable machine learning. *Plos One*, 19(3), e0299147. <https://doi.org/10.1371/journal.pone.0299147>
- Zolfigol, M. A., Azizian, S., Torabi, M., Yarie, M., & Notash, B. (2024). The importance of nonstoichiometric ratio of reactants in organic synthesis. *Journal of Chemical Education*, 101(3), 877-881. <https://doi.org/10.1021/acs.jchemed.3c00530>

---

### Etik, Beyan ve Açıklamalar

---

#### 1. Etik Kurul izni ile ilgili;

Bu çalışmanın yazarı, Etik Kurul İznine gerek olmadığını beyan etmektedir.

#### 2. Bu çalışmanın yazarı, araştırma ve yayın etiği ilkelerine uyduklarını kabul etmektedir.

3. Bu çalışmanın yazarı kullanmış oldukları resim, şekil, fotoğraf ve benzeri belgelerin kullanımında tüm sorumlulukları kabul etmektedir.

#### 4. Bu çalışmanın benzerlik raporu bulunmaktadır.

---