Vizyoner Dergisi

Süleyman Demirel Üniversitesi Vizyoner Dergisi, Yıl: 2026, Cilt: 17, Sayı: 51, 777-796.
Süleyman Demirel University Visionary Journal, Year: 2026, Volume: 17, No: 51, 777-796.

ISSN: 1308-9552
https://ror.org/04fjtte88

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

PERFORMANCE COMPARISON OF MACHINE LEARNING-BASED REGRESSION MODELS IN PREDICTING TÜRKİYE'S EXPORTS

TÜRKİYE İHRACATININ TAHMİNİNDE MAKİNE ÖĞRENMESİ TABANLI REGRESYON MODELLERİNİN PERFORMANS KARŞILAŞTIRMASI

Assoc. Prof. Dr. Bora ÖÇAL[1]

ABSTRACT

The study aims to identify the most suitable machine learning-based regression model that can be used to predict the future course of exports by analysing the macroeconomic indicators affecting Türkiye's export performance and through variables with a high correlation with exports. The study utilises 12 different macroeconomic variables related to the Turkish and global economies, and the relationship between these variables and exports are examined using correlation analysis. For the variables showing a strong relationship with exports, the linear model LASSO is applied alongside four different ensemble-based machine learning algorithms, using the Python programming language. To obtain unbiased and robust predictive accuracy, model performances are evaluated using a cross-validation framework. The results demonstrate that all models achieve high predictive performance; however, LASSO, in particular, provides the strongest overall accuracy with near-perfect explanatory power and a low error rate. Among the ensemble methods, Gradient Boosting emerges as the approach with the best performance. The findings demonstrate that regression methods can effectively capture Türkiye's export dynamics and provide a reliable foundation for export forecasts. The study contributes to the literature from both methodological and practical perspectives in understanding Türkiye's export dynamics, whilst also supporting data-driven policy development for decision-makers.

Keywords: International Trade, Export Prediction Models, Regression Analysis, Machine Learning, Correlation Analysis.

JEL Classification Codes: C82, C88, E00, F47.

ÖZ

Bu çalışma, Türkiye'nin ihracat performansını etkileyen makroekonomik göstergeleri analiz ederek, ihracatla yüksek korelasyona sahip değişkenler aracılığıyla ihracatın gelecekteki seyrini tahmin etmek için kullanılabilecek en uygun makine öğrenmesi tabanlı regresyon modelini tespit etmeyi amaçlamaktadır. Çalışmada, Türkiye ve küresel ekonomiyle ilgili 12 farklı makroekonomik değişken kullanılmış ve bu değişkenler ile ihracat arasındaki ilişki korelasyon analiziyle incelenmiştir. İhracatla güçlü ilişki gösteren değişkenlere Python programlama dili kullanılarak, dört farklı ensemble tabanlı makine öğrenimi algoritması ile birlikte doğrusal bir model olan LASSO uygulanmıştır. Öngörü doğruluğunu tarafsız ve sağlam şekilde elde etmek için model performanları çapraz doğrulama çerçevesi kullanılarak değerlendirilmiştir. Sonuçlar, tüm modellerin yüksek öngörü performansı elde ettiğini göstermekle birlikte özellikle LASSO, neredeyse mükemmel açıklayıcılık ve düşük hata seviyesi ile en güçlü genel doğruluğu sağlamıştır. Ensemble yöntemleri arasında ise Gradient Boosting, en iyi performans gösteren yaklaşım olarak öne çıkmıştır. Elde edilen bulgular, regresyon yöntemlerinin Türkiye'nin ihracat dinamiklerini etkili şekilde yakalayabildiğini ve ihracat tahminleri için güvenilir bir temel sağladığını göstermektedir. Bu çalışma, Türkiye'nin ihracat dinamiklerini anlamaya yönelik metodolojik ve pratik açıdan literatüre katkıda bulunurken karar vericiler için de veriye dayalı politika geliştirme konusunda önemli çıkarımlar sağlayacaktır.

Anahtar Kelimeler: Uluslararası Ticaret, İhracat Tahmin Modelleri, Regresyon Analizi, Makine Öğrenmesi, Korelasyon Analizi.

JEL Sınıflandırma Kodları: C82, C88, E00, F47.

[1] Süleyman Demirel University, School of Civil Aviation, Department of Aviation Electrical and Electronics, Türkiye, boraocal@sdu.edu.tr

Makale Geliş Tarihi / Received : 01.11.2025
Makale Kabul Tarihi / Accepted : 20.04.2026
https://doi.org/10.21076/vizyoner.1814982

Süleyman Demirel Üniversitesi Vizyoner Dergisi, Yıl: 2026, Cilt: 17, Sayı: 51, 777-796.
Süleyman Demirel University Visionary Journal, Year: 2026, Volume: 17, No: 51, 777-796.
ISSN: 1308-9552
https://ror.org/04fjtte88

EXTENDED SUMMARY

Purpose and Scope:

The study aims to identify the most suitable machine learning-based regression model that can be used to determine the macroeconomic factors affecting Türkiye's exports and to predict the future course of exports through variables with a high correlation with exports. Within the scope of the study, 12 different macroeconomic variables related to Türkiye and the global economy are selected. These variables include Türkiye's import volume and trade deficit, world export and import volumes, world trade volume index, dollar selling rate, CPI-based real effective exchange rate, total industrial production index, general CPI index, general domestic PPI index, average crude oil prices, and policy interest rates. Going beyond traditional econometric approaches, the study compares the performance of data-driven, high-predictive-power models: LASSO Regressor, Extra Trees Regressor, Random Forest Regressor, AdaBoost Regressor, and Gradient Boosting Regressor.

Design/methodology/approach:

Within the scope of the study, a correlation analysis is first conducted to determine the relationship between Türkiye's exports and the identified variables. The results of the correlation analysis reveal variables that are strongly and significantly related to export value. These results provide a statistical basis for understanding the data structure and the input data for the ensemble-based four regressors. Furthermore, all twelve identified variables are integrated into the model to enable the LASSO regressor to autonomously prioritize the most effective predictors. The second part of this statistical analysis rely on the application of algorithms of regression that are generated using the application of the machine learning method on the variables that are identified using the result of the correlation analysis. Five algorithms of regression are generated using the Python program based on the machine learning method with the intent of predicting the export value based on the variables that are identified with a great deal of precision. These five algorithms are LASSO Regressor Algorithm, Extra Trees Regression Algorithm, Random Forest Regression Algorithm, Ada Boost Regression Algorithm, and Gradient Boosting Algorithm. The four main comparison metrics used to determine the goodness of these five regression algorithms comprise four main simple analytical metrics: R^2 (coefficient of determination), RMSE (root mean square error), MSE (mean square error), and MAE (mean absolute error) to evaluate how these five regression algorithms perform to predict exports. The dataset spans 2006-2025 years on a monthly basis. The dataset comprises 80% variables to build on regression while 20% is utilized during the test phase. In addition, cross-validation is also implemented. The dataset is created after being closely monitored and compiled using reliable sources such as the Central Bank of the Republic of Türkiye, the World Bank, and CPB.

Findings:

The correlation analysis reveals a strong positive relationship between exports and import volume, global trade indicators, the dollar exchange rate, industrial production, CPI, and domestic PPI, while a strong negative correlation is observed with the CPI-based real effective exchange rate. These results indicate that Türkiye's exports are shaped by both domestic economic capacity and global trade dynamics. Regression findings further show that all machine learning models achieve high predictive accuracy. Among them, the LASSO Regressor provides the strongest performance under cross-validation ($R^2 = 0.999998$) with the lowest RMSE and MAE values, suggesting that a regularized linear specification captures export variation remarkably well. Among the ensemble approaches, Gradient Boosting performs best ($R^2 = 0.96846$), followed by Extra Trees and Random Forest, all of which maintain R^2 values above 0.95. AdaBoost produces comparatively larger errors, though still within an acceptable range. Graphical comparisons confirm that the leading models closely track both the overall export trend and short-term fluctuations, while error distributions remain centered around zero without systematic bias. Overall, these results support the robustness and reliability of the proposed regression methods for forecasting Türkiye's export dynamics.

Conclusion and Discussion:

The results obtained show that there is a very complex, multidimensional relationship between exports and both domestic economic factors and foreign trade relations. Specifically, they show that there is a direct relationship between export activity and domestic production capacity as well as foreign demands. The results obtained show that there is a very strong connection between export activity and both domestic and foreign factors such as demands and capacities. Among the five regression models employed within the study, it can be noted that the best results are obtained using the LASSO Regressor model. The result obtained shows that this method is very accurate and less error-prone. This implies that machine-learning approaches can also be very effectively employed within the prediction of exports. The paper presents both methodological and practical contributions to export literature. The methodological contribution entails a comparison of how regression approaches within machine-learning can gauge import and export within Turkish and various global economic approaches. On the other hand, within practical contributions, a more precise perspective on export dynamics within Turkish economy literature is presented. This implies that the study provides much more precise export information within Turkish economy literature than those previously available. The paper concludes that machine-learning approaches can serve as a very strong alternative within export prediction approaches due to their strong precision capabilities within economic approaches that entail much variability and complexity within variables' relations.

Süleyman Demirel Üniversitesi Vizyoner Dergisi, Yıl: 2026, Cilt: 17, Sayı: 51, 777-796.
Süleyman Demirel University Visionary Journal, Year: 2026, Volume: 17, No: 51, 777-796.
ISSN: 1308-9552
https://ror.org/04fjtte88

GENİŞLETİLMİŞ ÖZET

Amaç ve Kapsam:

Bu çalışma, Türkiye'nin ihracatını etkileyen makroekonomik faktörleri belirleyerek, ihracat ile yüksek korelasyona sahip değişkenler aracılığıyla ihracatın gelecekteki seyrini tahmin etmek için kullanılabilecek en uygun makine öğrenmesi tabanlı regresyon modelini tespit etmeyi amaçlamaktadır. Çalışma kapsamında, Türkiye ve küresel ekonomiye ilişkin 12 farklı makroekonomik değişken seçilmiştir. Bu değişkenler arasında Türkiye'nin ithalat hacmi ve dış ticaret açığı, dünya ihracat ve ithalat hacimleri, dünya ticaret hacmi endeksi, dolar satış kuru, TÜFE bazlı reel efektif döviz kuru, toplam sanayi üretim endeksi, genel TÜFE endeksi, genel yurt içi ÜFE endeksi, ortalama ham petrol fiyatları ve politika faiz oranları yer almaktadır. Çalışmada, geleneksel ekonometrik yaklaşımların ötesine geçilerek, veriye dayalı, yüksek tahmin gücüne sahip LASSO Regressor, Extra Trees Regressor, Random Forest Regressor, AdaBoost Regressor ve Gradient Boosting Regressor modellerinin performansları karşılaştırılmıştır.

Yöntem:

Çalışma kapsamında öncelikle Türkiye'nin ihracatı ile tespit edilen değişkenler arasındaki ilişkiyi belirlemek amacıyla korelasyon analizi yapılmıştır. Korelasyon analizi sonuçları, ihracat değeri ile güçlü ve anlamlı bir ilişki içinde olan değişkenleri ortaya çıkarmıştır. Bu sonuçlar, veri yapısını anlamak ve ensemble tabanlı dört regresörün girdi verileri için istatistiksel bir temel sağlamıştır. Ayrıca belirlenen on iki değişkenin tümü, LASSO regresörünün en etkili tahmincileri özerk bir şekilde önceliklendirebilmesi için modele entegre edilmiştir. Analizin ikinci aşamasında, korelasyon analizi sonucu elde edilen verilere makine öğrenmesi tabanlı regresyon algoritmaları uygulanmıştır. Python programlama dili kullanılarak LASSO, Ekstra Trees, Random Forest, AdaBoost ve Gradient Boosting Regresörü olmak üzere beş farklı model, elde edilen verilere dayalı olarak, ihracat değerlerini tahmin etmek üzere eğitilmiştir. Bu modellerin performansları; R^2 (belirlilik katsayısı), RMSE (kök ortalama kare hata), MSE (ortalama kare hata) ve MAE (ortalama mutlak hata) olmak üzere dört temel istatistiksel metrik kullanılarak değerlendirilmiştir. Veri seti, 2006-2025 dönemine ait aylık verilerden oluşmaktadır. Verilerin 80%'lik kısmı modeli eğitmek için, geri kalan 20%'si ise test aşaması için kullanılmıştır. Ayrıca, çapraz doğrulama da uygulanmıştır. Bu veriler, Türkiye Cumhuriyeti Merkez Bankası (TCMB), Dünya Bankası ve Hollanda Ekonomi Politikası Analiz Bürosu (CPB) gibi güvenilir kaynaklardan derlenmiştir.

Bulgular:

Korelasyon analizi sonuçları, Türkiye'nin ihracatı ile ithalat hacmi, dünya ihracat ve ithalat hacmi, dünya ticaret hacmi endeksi, dolar satış kuru, toplam sanayi üretim endeksi, genel TÜFE endeksi ve genel yurt içi ÜFE endeksi arasında pozitif yönlü güçlü bir korelasyon olduğunu göstermiştir. Buna karşılık, TÜFE bazlı reel efektif döviz kuru değişkeninin ise ihracat ile negatif yönlü güçlü bir korelasyon gösterdiği tespit edilmiştir. Bu bulgular, Türkiye'nin ihracatının hem yurt içi ekonomik göstergeler hem de dış ticaret dinamikleri ile yakından ilişkili olduğunu ortaya koymuştur. Özellikle pozitif korelasyon gösteren değişkenler, ihracat performansının ulusal üretim kapasitesi ve küresel talep koşullarından doğrudan etkilendiğini göstermektedir. Regresyon bulguları ayrıca, tüm makine öğrenimi modellerinin yüksek tahmin doğruluğu elde ettiğini göstermektedir. Bunlar arasında LASSO Regresörü, çapraz doğrulama altında en güçlü performansı (R^2 = 0,999998) ve en düşük RMSE ve MAE değerlerini sunarak, düzenlenmiş doğrusal spesifikasyonun ihracat varyasyonunu oldukça iyi yakaladığını göstermektedir. Ensemble yaklaşımları arasında Gradient Boosting en iyi performansı gösterirken (R^2 = 0,9685), onu Extra Trees ve Random Forest izlemiş ve bu iki modelin R^2 değerlerinin 0,95'in üzerinde olduğu görülmüştür. AdaBoost nispeten daha büyük hatalar üretmiş ancak yine de bu değerleri kabul edilebilir bir aralık içinde yer almıştır. Grafiksel karşılaştırmalar, ele alınan modellerin hem genel ihracat eğilimini hem de kısa vadeli dalgalanmaları yakından takip ettiğini, hata dağılımlarının ise sistematik bir sapma olmaksızın sıfır etrafında merkezlendiğini doğrulamaktadır. Genel olarak, bu sonuçlar Türkiye'nin ihracat dinamiklerini tahmin etmek için önerilen regresyon yöntemlerinin sağlamlığını ve güvenilirliğini desteklemektedir.

Sonuç ve Tartışma:

Bulgular, Türkiye ihracatının hem iç ekonomik göstergeler hem de dış ticaret dinamikleri ile karmaşık ve çok boyutlu bir ilişki içinde olduğunu ortaya koymaktadır. Özellikle ihracat performansının ulusal üretim kapasitesi ve küresel talep koşullarından doğrudan etkilendiğini göstermektedir. Bu sonuçlar, Türkiye'nin ihracat performansının yalnızca üretim ve talep faktörlerinden değil, aynı zamanda döviz kuru hareketleri gibi makroekonomik değişkenlerden de güçlü bir şekilde etkilendiğini ortaya koymaktadır. Bu çalışmada kullanılan beş model arasında LASSO Regresörü modeli, ihracat tahmininde en başarılı sonuçları vermiştir. Bu modelin yüksek R^2 değeri ve düşük hata oranları, makine öğrenmesi tabanlı yaklaşımların Türkiye'nin ihracatını tahmin etmede etkili bir şekilde kullanılabileceğini göstermektedir. Çalışma hem metodolojik hem de uygulama açısından literatüre katkı sağlamaktadır. Metodolojik olarak, Türkiye ve dünya ekonomik verileri kullanılarak, makine öğrenmesi tabanlı regresyon modellerinin karşılaştırmalı performans analizi yapılmıştır. Uygulama açısından ise politika yapıcılar için ihracat dinamiklerini daha doğru tahmin edebilen veri odaklı bir yaklaşım önerilmiştir. Sonuç olarak makine öğrenmesi yöntemlerinin özellikle veri çeşitliliği ve değişkenler arasındaki karmaşık ilişkilerle karakterize edilen ekonomik sistemlerde yüksek tahmin doğruluğu sağlayan, ihracat tahminleri için güvenilir ve güçlü bir alternatif olduğu tespit edilmiştir.

Süleyman Demirel Üniversitesi Vizyoner Dergisi, Yıl: 2026, Cilt: 17, Sayı: 51, 777-796.
Süleyman Demirel University Visionary Journal, Year: 2026, Volume: 17, No: 51, 777-796.
ISSN: 1308-9552
https://ror.org/04fjtte88

1. INTRODUCTION

Exports are of strategic importance in terms of a country's economic growth, employment levels, foreign trade balance, and the sustainability of its foreign exchange reserves. In developing countries, improved export performance directly affects not only foreign trade revenues but also production capacity, technological advancement, and competitiveness within that country. Therefore, identifying the factors affecting exports and making accurate predictions about the future course of exports has become an important need for policymakers, investors, and industrial planners. When assessed over the last two decades for the case of Türkiye, exports have been one of the most dynamic components of economic growth, and changes in foreign trade policies and the exchange rate regime have substantially reshaped the position of the country in the global trading system.

The majority of research work on factors that affect exports published within the economic literature relies on macroeconomic variables such as exchange rate variables, productivity variables, variables expressing price level, variables expressing external demands, and variables expressing energy prices. These relationships tend to display features like being nonlinear, multifaceted, and time variant. Though standard econometric approaches such as OLS, ARDL, and VAR can handle these complexities to some extent, they can prove to be inefficient in situations where these relationships display nonlinearities. The growing use of machine learning approaches within econometrics over the past years has triggered a massive shift within this field to overcome these difficulties to some extent to gain more accurate results.

In particular, machine learning methods learn patterns or non-linear relationships in the data at hand without any prior assumptions. Among others, regularization-based models like LASSO and tree-based ensemble methods, such as Random Forest, Extra Trees, Gradient Boosting, and AdaBoost, generally model the interactions of economic indicators more accurately than others; hence, they often outperform traditional methods that have been built on mainly theoretical assumptions. This has resulted in a paradigm shift whereby there is a prediction accuracy within the realm of macroeconomic studies, and data-driven processes within decision-making are on the increase.

In this scenario, recent research work carried out by a number of researchers, such as Can and Gerşil (2025), Soydal et al. (2024), Al Farisy et al. (2024), and Zhu (2025), has identified that machine learning models have produced positive export predictions results not only in developed countries but also developing countries. For example, Can and Gerşil (2025) identified within their research work that there is no method that outperforms artificial neural networks on error rate on exports predicted by Türkiye. Similarly, Soydal et al. (2024) stated within their research work that they examined Turkish exports to Balkan countries using a trade volume employing a total of seven machine learning algorithms. They observed varying results concerning precision according to the dataset. This indicates that export results can also be predicted employing a wide range of macroeconomic variables.

The examination of export performance within the Turkish economy acquires considerable significance not only within the context of moving towards economic growth but also to diminish foreign dependencies and foreign exchange deficits to a considerable extent. The macroeconomic variables including exchange rate variability, energy prices, global demands, and other variables related to productions acquire considerable significance within this context because they directly influence export performance within the Turkish economy. Therefore, observing these factors within an experimental setting will aid within the context of understanding economic vulnerabilities within this setting.

The main objective of this study is to determine the macroeconomic factors affecting the export performance of Türkiye and to determine the machine learning model providing the highest prediction accuracy through these factors. In this respect, monthly data between 2006 and 2025 were used, and 12 different macroeconomic indicators were considered as independent variables, while exports of Türkiye were considered as dependent variables and analyzed. First, using correlation analysis, the relationships between the variables were examined. Then, the variables showing high positive or negative correlations were included in the Extra Trees Regressor, Random Forest Regressor, AdaBoost Regressor, and Gradient Boosting Regressor models for analysis Additionally, the LASSO regressor algorithm was applied to all variables considered within the scope of the study. Thus, the relative importance of the variables in terms of exports was determined and the prediction performance of the models was evaluated comparatively.

The study differs from existing research in the literature in three key ways: first, it addresses the macroeconomic indicators of Türkiye's exports holistically and applies machine learning-based models instead of traditional

econometric methods; second, several algorithms are systematically compared, hence identifying the most suitable model for export prediction. Finally, the dataset used was extended to include current macroeconomic indicators until July 2025, reflecting current dynamics in Türkiye's foreign trade.

This study makes a contribution to the literature in two different ways. First, the applications of machine learning for export predictions of Türkiye have generally been conducted using a limited number of variables or short-term data sets. The broad set of utilized macroeconomic variables and the long-term data structure of this study provide more reliable results in terms of predict accuracy. Second, in terms of revealing which model more effectively reflects macroeconomic dynamics and provides a methodological reference for future policy analysis, this paper compares the performance of different regression models. Thus, this is an original contribution to both the empirical literature and the development of data-driven economic predicting models.

2. LITERATURE REVIEW

The studies on determining export performance have received significant importance in the literature in terms of understanding the multidimensional relationships between economic growth, exchange rates, macroeconomic stability, and structural factors. In recent years, machine learning-based models have come to the fore in the relevant literature, offering higher predictive accuracy compared to classical econometric methods. However, econometric analyses conducted in Türkiye and other countries have comprehensively examined the macroeconomic, sectoral, and institutional determinants of exports. In this section, studies in the literature are classified under five main themes. The first one is export forecasting using machine learning, the second one is econometric analyses conducted on Türkiye, the third one is the impact of macroeconomic factors on exports, the fourth one is country-based international studies, and finally, findings related to institutional, sectoral, and structural factors.

2.1. Studies on Export Prediction Using Machine Learning

Recent years have witnessed an important trend within literature concerning an increasing usage of machine learning approaches for predicting foreign trade. Can and Gerşil (2025) applied time series, regression, and artificial neural networks (ANN) approaches to forecast exports, concluding that the results obtained by ANNs contained less error. To forecast trade volume between Türkiye and Balkan countries, Soydal et al. (2024) applied a total of seven different approaches to machine learning. They concluded that results varied according to both countries and dataset partitions. Random Forest algorithm outperformed other approaches to forecasting export volume figures of Indian SMEs, according to a relevant analysis performed by Kumar et al. (2025). Similarly, Gupta and Kumar (2024) applied an artificial neuronal network to successfully forecast exports of India's leather footwears and emphasized within results that implementation of SGD led to better outcomes. Focusing upon exports to Indonesia and employing Random Forest algorithm with 24 variables, Al Farisy et al. (2024) concluded that these algorithm-based results indicated better forecasting capability having 92.5% ratio within R^2 while being far better than those acquired employing other approaches. To forecast Pakistan's export volume, Yousfani et al. (2024) announced within their study that results obtained employing MLP algorithm indicated significantly better forecasting compared to approaches such ARIMA & ELM. Zhu (2025) concluded within research on forecasting Chinese imports & exports volume that ISSA-SVM optimized algorithm indicated significantly more precise forecasting compared to other approaches. Analyzing foreign trade volume, Alghushairy (2025) stated within results that error ratio within forecasting results obtained by this study indicated higher precision within forecasting results obtained by XGBoost algorithm if compared to those obtained employing other approaches such ExtraTrees & RandomForest. These studies on a whole display great prominence within literature on forecasting foreign trade volume due to higher degree of precision achieved within forecasting results rather than employing IMF econometric approaches.

2.2. Econometric Analysis Studies on Türkiye

Using differing research methodologies, research on Türkiye has investigated the complex interactions between exports and variables such as growth, exchange rate, and imports. Erdoğan (2016) examined the multiple regression of various variables that greatly affect the export of Türkiye and found that the exchange rate and EU's GDP are major determinants. Regarding the dynamics of export growth, Yapraklı (2007) showed that industrial exportation has a positive effect on growth. A relationship between exportation and RED exists over the long run, according to Balcılar et al. (2014). Yenisu (2019) used the ARDL approach to examine the cointegration of growth,

importation, and exportation, considering both the short run and the long run. Although there is a positive relationship between exportation, importation, and growth over the long run, based on the study of Aktaş (2009), the variability of the exchange rate negatively affects exportation and finds that the results of Kasman (2003) revealed this fact. Finally, based on the study of Çelik and İlkay (2016), an examination confirms that a positive relationship between exportation and importation exists. The above results confirm that exportation in the Turkish region depends on the dynamics of the exchange rate and importation.

2.3. Studies on the Impact of Macroeconomic Factors on Exports

The analysis of the export equation to identify the influence of macroeconomic variables on export performance finds that variables such as energy, industrial production, and macroeconomic stability are of major importance. The best results were found by ARDL-SVM. This can further be verified with the study of the foreign trade ratio of Türkiye performed by Özden (2022) on the variables of GDP, Exchange rate, CPI, PPI, and oil prices. While Bolat (2025) stated that energy consumption, domestic savings, and exports affect industrial output substantially, Gayaker (2025) stated that exchange rate and exchange rate volatility have opposite effects on exports. While Emsen (2018) stated export dependence on imports and exchange rate irrelevance on exports, Sarı (2010) stated exchange rate uncertainty to have negative effects on imports. The inflows of short-term capital have negative effects on exports and exchange rate variables, which is indicated by studies presented by Değirmen and Elmas (2008). Karaş and Karaş (2017) stated that there are unilateral relationships between exports and imports and two-way relationships between imports and exchange rate variables. The above studies make it apparent that macroeconomic stability is crucial to exports.

2.4. Country-Based International Studies

A wide range of studies using econometric and machine learning methods have identified export differences on a structural and regional level. Xu et al. (2023) identified export differences between KNN, SVM, RF, and ARIMA models on Chinese agricultural exports, highlighting export prediction ability using ML Models. Another study by Istiqlala and Kurdhi (2024) showed that SVM can generate export and import values with high accuracy. A research finding by Pamornmast et al. (2013) identified that exchange rate has negative effects on exports, but particularly on Thailand. Another two studies by Moschos (1989) and Bakari (2017) indicated positive effects on economic growth through exports but positively affected in a different manner because of export ability to developed countries. A research finding by Velnampy and Achchuthan (2013) identified positive effects on economic growth due to exports and imports but diversely identified exchange rate fluctuations by Zakaria (2013) on Malaysia. The study executed by Remeikienė et al. (2019) indicated tax and wages policy factors to positively affect export competitiveness on Lithuanian exports. Importantly, on an active global basis, export activities are noticed to shift according to their economic structure.

2.5. Studies on Institutional, Sectoral, and Structural Factors

More specific factors include studies examining how exports are affected by the institutional environment, sector outlook, and human capital. The export behavior observed within both textile industries is significantly affected by company sizes, educational background of textile managers, and marketing activities, according to Serra et al. (2012). The Turkish textile industry, examined by Genç et al. (2010), introduced a one-way causal relationship between exports and human capital. The impact on export performance is increased by corporate quality factors, according to a study by Ruzekova et al. (2020). Though Medineckiene and Kirdaite (2021) deduced that GDP and employment strongly influenced exports within the UK economy, Bista (2004) discovered that foreign direct investments positively influenced exports and growth within the Nepalese economy. Uysal and Mohamoud (2018) stated that East Africa exports were affected by human resources, industrialization, and exchange rates. These studies make it apparent that export-driven growth is significantly affected by institutional factors.

Thorough examination of the macroeconomic factors which affect export behavior in Türkiye ensures that this research provides something unique to the existing literature. The factors involved are more representative of T Türkiye's economy, including import and export values, balance of trade, exchange rate, industrial product index, prices, policy interest rate, world trade values, and energy costs, rather than considering GDP factors to forecast export. Rather than utilizing growth rate values, a methodological design that takes into account aspects concerning product, price, and market dynamics is involved. The study also utilizes correlation to identify variables, which ensures significantly relevant results. These results are comprised of five regression models on machine learning: Extra Trees, Random Forest, AdaBoost, and Gradient Boosting Machine. This indicates that both relevance factors

and the forecasting capability of machine learning technologies can be assessed using this methodological design. This work provides a more versatile methodological design ready to remove constraints faced by econometric studies, while its relevance to Turkish economy-specific information provides results very much applicable to policymakers.

3. PURPOSE AND LIMITATIONS OF THE STUDY

This study was conducted to determine the macroeconomic factors affecting Türkiye's exports through correlation analysis and to identify the most suitable model among the regression models determined for use in future export prediction using variables with high correlation with exports. The regression models used in the study were chosen because they represent diverse approaches, such as regularization-based, tree-based, gradient-based, and boosting methods.

The reason for considering the period 2006-2025 in the study is to comprehensively address Türkiye's economic and foreign trade dynamics over the last twenty years. However, since GDP data is only published quarterly or annually and the dataset includes indicators that directly reflect the production and foreign trade performance of the Turkish economy, the GDP variable was not included in the analysis. In addition, variables such as the total industrial production index, export and import volumes, exchange rates, and interest rates summarize the effects of economic growth, which are represented by the GDP, thus improving the overall explanatory capacity of the model.

On a related note with respect to model selection, although the AR/ARMA models are often used as a benchmark for univariate export forecasting, they are not used in this study. This is a result of the adoption of the multivariate regression framework with macroeconomic covariates. Since AR/ARMA models rely solely on the historical values of the dependent variable, they are not directly comparable in this multivariate setting; therefore, their integration has been left for future studies that may incorporate ARIMAX-type benchmarks within a dedicated time-series forecasting design.

4. DATA SET

This study examines 12 different macroeconomic variables that are thought to affect Türkiye's export activities and are frequently used in studies in the literature. Monthly data for these variables between January 2006 and July 2025 were used. These data have been compiled from reliable sources such as the Central Bank of the Republic of Türkiye (TCMB), the World Bank and the Netherlands Bureau for Economic Policy Analysis (CPB). The variables used in the study, the period of the data, and the sources from which they were obtained are shown in Table 1.

Table 1. Data Set

Variable Definition	Variable Type	Data Period (Monthly)	Data Source
Türkiye Exports (Million $)	Dependent	2006 -2025	TCMB (2025)
Türkiye Imports (Million $)	Independent	2006 -2025	TCMB (2025)
Türkiye Foreign Trade Deficit (Million $)	Independent	2006 -2025	TCMB (2025)
World Exports Volume (Million $)	Independent	2006 -2025	TCMB (2025)
World Imports Volume (Million $)	Independent	2006 -2025	TCMB (2025)
World Trade Volume Index	Independent	2006 -2025	CPB (2025)
Dollar Exchange Rate Sale ($/TL)	Independent	2006 -2025	TCMB (2025)
CPI-Based Real Effective Exchange Rate ($/TL)	Independent	2006 -2025	TCMB (2025)
Industrial Production Index Overall	Independent	2006 -2025	TCMB (2025)
CPI Index Overall	Independent	2006 -2025	TCMB (2025)
Domestic Producer Price Index Overall	Independent	2006 -2025	TCMB (2025)
Crude Oil Average Prices ($)	Independent	2006 -2025	Worldbank (2025)
Policy Interest Rates (%)	Independent	2006 -2025	TCMB (2025)

Since all data used in the study were collected monthly during the same years, no adjustments were necessary.

5. METHODOLOGY

In this study, correlation analysis was first applied to examine the interaction between Türkiye's exports and 12 selected macroeconomic variables. As a result of the correlation analysis, variables showing strong positive or negative correlations with exports were considered as independent variables in tree-based regression models. Using these identified variables, were incorporated into ensemble machine learning algorithms such as Extra Trees, Random Forest, AdaBoost, and Gradient Boosting models using the Python programming language. In addition, to fully leverage the model's variable selection capability and minimize potential multicollinearity issues, the LASSO regressor algorithm, a regularized linear approach, was applied to all independent variables in the study. These models were used to determine the most appropriate prediction model based on the macroeconomic variables affecting Türkiye's export performance and to comparatively analyze model performances.

Unlike traditional econometric techniques, where assumptions of stationarity must be satisfied to avoid spurious relationships, tree-based machine learning models, as used in this paper, do not require any assumptions regarding stationarity or distribution. Moreover, these models are capable of picking up trends and non-linear movements in the data, providing a viable alternative to traditional techniques for export prediction.

In addition to the conventional 80/20 train–test split, a 5-fold cross-validation (CV) procedure was employed to obtain an unbiased estimate of out-of-sample prediction errors. All regression algorithms, including LASSO, Extra Trees, Random Forest, AdaBoost, and Gradient Boosting, were evaluated under the same CV protocol. Missing values were handled within each training fold using median imputation implemented through a pipeline structure to avoid information leakage. The average performance metrics (R^2, RMSE, MAE, and MSE) across folds are reported.

5.1. Correlation Analysis

A correlation analysis was performed to better understand the mutual relationships between the 12 different macroeconomic independent variables in the data set and Türkiye's exports. The purpose of this analysis was to better understand the dynamics of international trade and to find highly correlated variables that might improve the precision of export forecasts in the future. A thorough visual evaluation of the dependencies and relationships between the numerical variables in the dataset is made possible by the correlation matrix heat map that is produced as a result of the correlation analysis. The correlation coefficient between two variables is represented by each cell in this matrix, which accepts values between -1 and +1. In this case, perfect negative correlation is represented by a value of -1, and perfect positive correlation is represented by a value of +1. Very little or no correlation between the variables is indicated by values near zero (Alghushairy, 2025, p. 6). Specifically, Pearson correlation is reported as an exploratory diagnostic to summarize pairwise linear associations and to inspect potential multicollinearity patterns among indicators. Since correlation coefficients are limited to linear, bivariate relationships, they are not used as a stand-alone criterion for determining predictive usefulness. Instead, the predictive modelling stage relies on tree-based ensemble regressors that can capture non-linear dependencies and joint effects among predictors. Furthermore, to enhance interpretability under the fitted models, permutation-based feature importance is additionally reported in the Findings section.

5.2. Machine Learning

The results that we obtain through the use of econometric calculations offer a very useful structure that helps identify relationships between variables and apply cause and effect relationships accordingly. This method of calculation is found to be highly effective in closely analyzing the relationships in question. However, it is important to note that the achievement of accurate results from the calculations is made possible with the satisfaction of a number of assumptions simultaneously. These are the homogeneity of variance between the groups of variables, the absence of autocorrelation between the residuals, the presence of low inter-variable correlations within the units of analysis, and the presence of a normal distribution of the error term. These are critical assumptions. If any of the above assumptions is violated or cannot be satisfied under certain conditions, alternative methods are employed to calculate the results. However, the application of the alternative methods are found to reduce the reliability of the findings. They even make the results of these calculations not so precise at times. Machine learning differs significantly in this context. This method combines different algorithms that help identify relationships between variables. This method further differentiates itself because these calculations happen without considering the strict assumptions of statistics. All these machine learning algorithms are very successful at identifying the non-linear relationships that other algorithms could potentially overlook. More precise results are achieved while handling highly dimensional datasets that contain a lot of variables. Furthermore, these algorithms

identify the complicated patterns of relationships used by data based on their records. Most machine learning algorithms can be divided into two broad categories. Regression algorithms rank under the first category. The other category consists of classification algorithms. Regression algorithms specialize in predicting quantitative results. This occurs by studying the quantitative relationships between the input variables. On the other hand, the other category of algorithms generates results that contain particular categories. This happens after arranging the corresponding points into distinct sections (Al Farisy et al., 2024, p. 2; Hastie et al., 2009, p. 9; Özden, 2022, p. 275).

This study aimed to pick out the best model for predicting Türkiye's exports from various macroeconomic factors. It used models like the LASSO Regressor, Extra Trees Regressor, the Random Forest Regressor, the AdaBoost Regressor, and the Gradient Boosting Regressor. These approaches help boost prediction accuracy. They also cut down on the risk that one decision tree might overfit the training data. The models combine multiple decision trees for that purpose. In the end, this setup lets researchers get solid results on export forecasts. It even handles the tricky connections between different variables pretty well (Breiman, 2001, p. 5-32; Friedman, 2001, p. 2-37; Geurts et al., 2006, p. 7-9).

5.2.1. Extra Trees Regressor (Extremely Randomized Trees Regressor)

This algorithm acts as an ensemble method used mainly for regression problems. This method combines the predictions made by several individual decision trees to arrive at the final prediction. Its major differentiating feature is the introduction of randomness while building the trees. Predictors are chosen randomly along with random samples. This increases the level of predictive generalization, thereby increasing the reliability of the model (Geurts et al., 2006). The Extra Trees Regressor has proven to be robust for problems involving complicated scenarios such as non-linear relationships and for datasets with higher dimensions. These factors increase the predictive capability of the model. To be more precise, this method makes predictions by considering random subsets of predictors rather than all predictors. Also, there are random splits made for each node to ensure that all trees are independent of each other. These factors have made the Extra Trees Regressor an attractive option for regression problems involving data science and machine learning. This algorithm is quick and easy to implement (Alghushairy, 2025, p. 3).

5.2.2. Random Forest Regressor

RandomForestRegressor represents a framework for ensemble learning, where a collective prediction is produced by combining the predictions of multiple individual decision trees (Çakmak et al., 2022, p. 1004). The distinguishing feature of the RandomForestRegressor, as compared with other approaches used for machine learning, is the high level of robustness it provides against overfitting, along with the high level of accuracy it attains as a consequence. The RandomForestRegressor minimizes the level of variance and hence overfitting, making it a very appropriate technique for the analysis of a dataset with multiple influencing factors.

5.2.3. AdaBoost Regressor (Adaptive Boosting Regressor)

The AdaBoost Regressor is an extension of the AdaBoost algorithm to regression-type problems and is an ensemble learning method involving the sequential training of a group of weak learners, such as decision trees. The algorithm updates its focus with each iteration by emphasizing misclassified instances from the previous iteration, allowing the next weak learner to focus on correcting the mistakes of the previous learner (Çakmak et al., 2022, p. 1004). As such, each subsequent weak decision tree is designed to correct the mistakes of the previous learner, and the final learner is a weighted aggregate of all the component learners. The AdaBoost Regressor is useful in modeling non-linear relationships and is known to produce high-performing predictions on datasets with low levels of noise. The level of sensitivity to performance, however, may be amplified in highly noisy datasets, requiring the adjustment of the AdaBoost Regressor's parameters.

5.2.4. Gradient Boosting Regressor

The Gradient Boosting Regressor, an implementation of the Gradient Boosting algorithm, is an ensemble learning model, specifically tailored for regression tasks. It involves the sequential training of multiple weak learners, typically decision trees, in an effort to reduce the residual error of previously trained models. Each tree added in succession improves the model's performance by learning from the gradient of the error of previously added trees. This allows the model to efficiently capture non-linear relationships, thus achieving high accuracy in predictions. However, in an effort to prevent overfitting, the model's ability to generalize is maintained by carefully selecting hyperparameters such as the learning rate, number of trees, and tree depth (Otchere et al., 2022, p. 1-3).

Süleyman Demirel Üniversitesi Vizyoner Dergisi, Yıl: 2026, Cilt: 17, Sayı: 51, 777-796.
Süleyman Demirel University Visionary Journal, Year: 2026, Volume: 17, No: 51, 777-796.
ISSN: 1308-9552
https://ror.org/04fjtte88

5.2.5. LASSO Regressor (Least Absolute Shrinkage and Selection Operator)

The LASSO Regressor is an L1-penalised (regularisation) based method that aims to increase prediction accuracy within the framework of linear regression while simultaneously controlling the complexity of the model. The basic approach involves adding a penalty term to the classical ordinary least squares (OLS) objective function, which includes the sum of the absolute values of the coefficients. This term pushes some coefficients towards zero, thereby enabling automatic variable selection. This feature contributes to producing more generalisable results by reducing the model's tendency to overfit, particularly in structures with explanatory variables that are interrelated (carrying the risk of multicollinearity), such as macroeconomic indicators (Tibshirani, 1996).

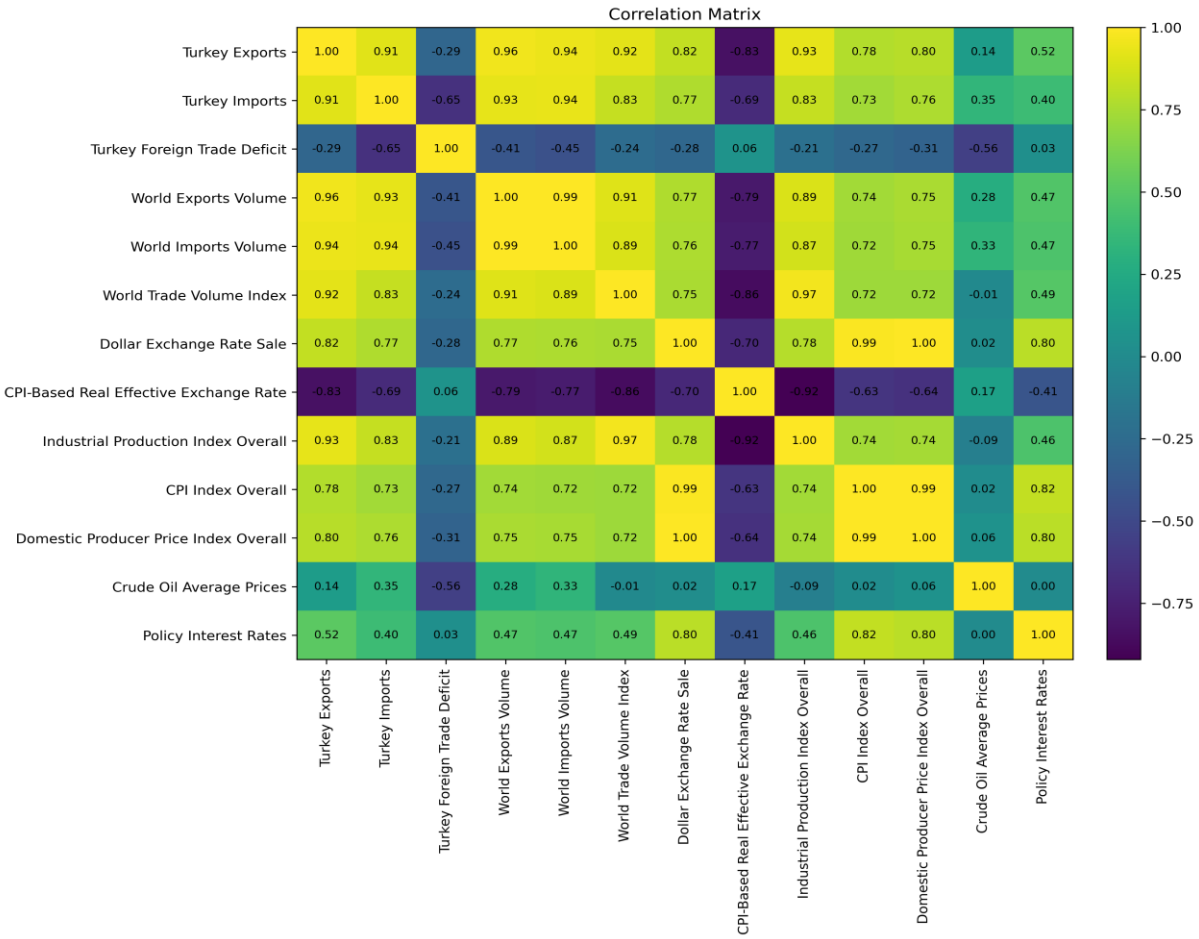
6. FINDINGS

The findings obtained as a result of the analyses conducted within the scope of the study are listed below.

6.1. Correlation Analysis Findings

In this section, a correlation analysis was performed to determine the strength and direction of the relationships between Türkiye's exports and 12 macroeconomic independent variables thought to affect exports. The correlation matrix obtained visually shows the positive or negative relationships between the variables and is presented in Figure 1. To complement correlation results and improve interpretability, permutation feature importance was reported based on the best-performing model on the held-out test set. This analysis reflects each predictor's contribution conditional on all other predictors; therefore, variables whose marginal correlation is modest but are informative jointly were revealed.

Figure 1. Correlation Matrix



The correlation matrix comprehensively shows the strength and direction of linear relationships between macroeconomic indicators affecting Türkiye's exports. According to Figure 1, exports are highly positively correlated with many key variables. In particular, variables with correlation values above +0.75 and below -0.75 can be considered strong associated factors of exports, and selecting these variables as independent variables in the model can be considered a statistically rational choice. However, to maintain the integrity of the data-driven approach, these variables were provided to the models without manual exclusion, allowing the LASSO regressor's L1 regularization to autonomously determine the optimal subset of predictors and address potential multicollinearity.

First, the strong positive correlation between Türkiye's imports and exports demonstrates the integrated structure of foreign trade. This strong relationship at the higher levels can be attributed to the fact that an increase in imports usually encourages production and the subsequent re-export of goods. Also, the strong positive relationship with world exports and world imports clearly manifests the fact that the Turkish export performance remains highly vulnerable to changes in world import and export levels. This trend clearly manifests the fact that the Turkish economy remains highly open to the outside world and that the external world supply-demand conditions remain highly decisive of Türkiye's export performance.

This particular export performance directly correlates with its production capacity, as made clear by the strong and positive relationship between export levels and the total industrial production index. A major catalyst that promotes an increase in export potential revolves around an increase in the industrial production levels. This particular fact becomes clear with the positive relationship that prevails between export levels and the General CPI and PPI indices. This manifests the fact that the increase in price levels marginally coincides with export earnings. This particular price differential due to the effects of the exchange rate or the increase due to the higher levels of inflation may be captured through statistical relationships under this particular scenario.

Real effective exchange rate results, when calculated based on the Consumer Price Index, reveal a significant inverse relation with exports when considering negative relationships. This unequivocally proves that a real appreciation of the exchange rate negatively affects the country's export competitiveness, whereas a depreciation of the Turkish Lira bolsters the country's export activity. The relation between trade deficits and exports is found to be weakly negative, proving that the country's trade balances improve with a rise in exports.

According to correlation analysis, Türkiye's exports show a strong negative relationship with the real exchange rate and a strong positive correlation with national production and price level indicators as well as global economic conditions. This result indicates that the selection of independent variables in the model is based on a statistically meaningful foundation and that the selected variables have high explanatory power for predicting Türkiye's exports. Therefore, using variables above +0.75 and below -0.75 as the correlation threshold will contribute to the model's strong predicting capacity while maintaining the balance of its multiple linear relationships.

6.2. Regression Analysis Findings

Of the 12 macroeconomic independent variables considered in the study, Türkiye's foreign trade deficit, crude oil prices, and policy interest rates were excluded from the regression analysis due to their weak correlation with the dependent variable, Türkiye's exports, and the analysis was performed with the other nine independent variables. Part of the data obtained was used for model training (80%), while the remainder was used for the testing phase (20%). This method was preferred in order to measure the generalization ability of the models and reduce the risk of overfitting.

6.2.1. Regression Model Metrics

The results of the five different regression models considered in the study are shown in Table 2. The table shows the performance of the five machine learning-based regression models applied in the study in predicting Türkiye's exports. The success of the models was evaluated based on explanatory power and error metrics; R^2, RMSE, MSE, and MAE values were used. The results obtained show the performance of different models in terms of prediction accuracy and error distribution.

Table 2. Metrics of Regression Models

Regression Model	R^2	RMSE	MSE	MAE
LASSO Regressor	0.9999	16.901689	285.66709400	14.06785
Extra Trees Regressor	0.9406	1099.5767	1209068.8875	874.0063
Random Forest Regressor	0.9400	1104.5900	1220119.1493	850.8244
AdaBoost Regressor	0.9369	1132.7289	1283074.8151	893.7773
Gradient Boosting Regressor	0.9296	1197.2696	1433454.5297	870.6714

The values in Table 2 indicate that the models exhibit high predictive performance; however, the results clearly separate into two tiers. The LASSO Regressor achieves the highest R^2 (0.9999) together with the lowest RMSE (16.90), MSE (285.67), and MAE (14.07), indicating the best overall fit and the smallest average error among all compared models. Among the tree-based methods, the Extra Trees Regressor yields the highest R^2 (0.9406) and the lowest RMSE (1099.58) and MSE (1,209,068.89), making it the strongest performer in terms of squared-error-based criteria. In contrast, the Random Forest Regressor provides the lowest MAE (850.82) among the tree-based models, suggesting the best performance in terms of average absolute deviation. The AdaBoost Regressor shows weaker performance than Extra Trees and Random Forest across the reported metrics, while the Gradient Boosting Regressor presents the highest RMSE and MSE, indicating comparatively larger error magnitudes in some instances. Overall, the pattern suggests that most prediction errors are moderate, but occasional larger deviations inflate the RMSE and MSE in the tree-based models.

From a model-selection perspective, LASSO should be preferred when the goal is to achieve the lowest overall error and strongest explanatory power in this dataset. If the comparison is restricted to tree-based approaches, the Extra Trees Regressor is preferable when the priority is to minimize squared-error/outlier sensitivity (RMSE/MSE), whereas the Random Forest Regressor is preferable when the goal is to minimize the average absolute error (MAE). Table 3 reports the cross-validation performance metrics of the regression models.

Table 3. Metrics of Regression Models with Cross-Validation

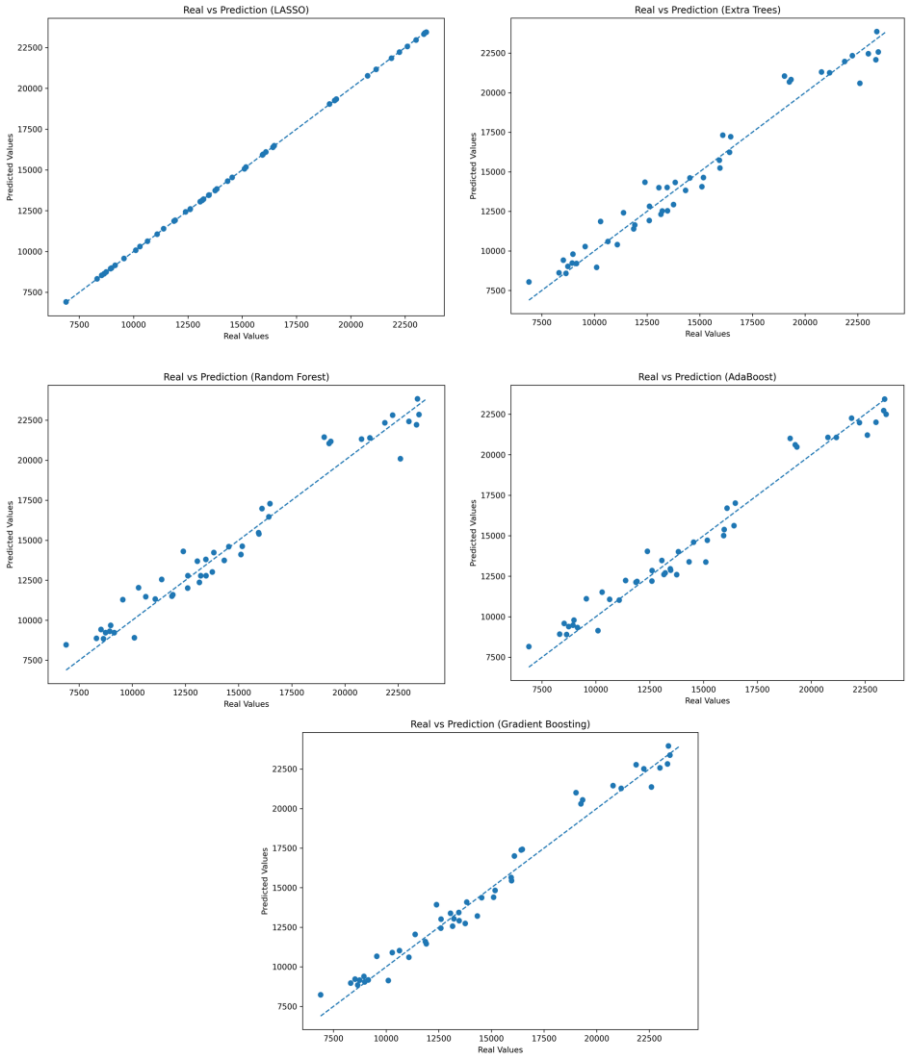
Regression Model	R^2	RMSE	MSE	MAE
LASSO Regressor	0.99998	6.20349	39.37447	4.81809
Gradient Boosting Regressor	0.96846	777.089	607004.2	626.185
Extra Trees Regressor	0.95632	919.981	847293.7	714.460
Random Forest Regressor	0.95260	963.040	931242.8	745.614
AdaBoost Regressor	0.94406	1042.73	1095153	827.547

The results indicate that the predictive accuracy of all models remains stable across folds, confirming the robustness of the analysis. Among the evaluated approaches, the LASSO regressor achieves the highest performance with an almost perfect average coefficient of determination ($R^2 = 0.99998$) and the lowest error values (RMSE = 6.20, MAE = 4.82), suggesting a strong linear structure in the selected macroeconomic indicators. Tree-based ensemble models, such as Gradient Boosting, Extra Trees, and Random Forest, also provide high predictive power with R^2 values above 0.95, demonstrating their ability to capture non-linear relationships in foreign trade dynamics. Overall, the cross-validation findings support the reliability of the model comparisons and show that the main conclusions are not driven by a particular train–test split.

6.2.2. Analysis of the Relationship Between Real and Prediction Values of the Models

In this section, the relationships between the real and prediction values of the models are examined, and the prediction accuracy and reliability of the models are analyzed comparatively. In this way, an attempt is made to determine which model is more successful in export predictions. The real and prediction graphs of the models are shown in Figure 2.

Figure 2. Real and Prediction Graphs of Regression Models



In Figure 2, the relationship between the observed (real) and predicted export values is evaluated for each model using the "Real vs Prediction" scatter plots. The proximity of the points to the 45-degree reference line indicates how accurately each model reproduces the observed values. Overall, all models capture the general increasing trend between real and predicted values; however, differences in dispersion around the diagonal are visible across methods. These visual patterns are consistent with the cross-validated performance metrics reported in Table 3. The LASSO Regressor exhibits an almost perfect alignment of points along the diagonal, indicating exceptionally strong agreement between predictions and observed values across the entire range. Together with its near-perfect cross-validated performance (Table 3), this suggests that a regularized linear specification provides an excellent approximation for export variation in this dataset. Among the non-linear approaches, the Gradient Boosting Regressor shows a tight clustering of points around the diagonal and a limited degree of dispersion, reflecting strong predictive performance, which is also supported by its superior cross-validation metrics relative to the other tree-based models. The Extra Trees and Random Forest regressors likewise demonstrate a dense concentration around the reference line, confirming robust predictive capability over most of the value range. In comparison, the AdaBoost Regressor displays relatively wider scatter, indicating larger deviations for some observations, consistent with its comparatively higher cross-validated error levels. In summary, Figure 2 indicates high predictive accuracy for all models, with LASSO providing the closest real–prediction agreement. Among the ensemble methods, Gradient Boosting emerges as the strongest performer under cross-validation, followed by Extra Trees and Random Forest, while AdaBoost appears relatively less stable in some ranges.

Süleyman Demirel Üniversitesi Vizyoner Dergisi, Yıl: 2026, Cilt: 17, Sayı: 51, 777-796.
Süleyman Demirel University Visionary Journal, Year: 2026, Volume: 17, No: 51, 777-796.
ISSN: 1308-9552
https://ror.org/04fjtte88

6.2.3. Error Distribution Analysis of Models

The distribution of prediction errors within the regression models utilized is examined within this section. The error distribution provides a more comprehensive indication concerning the effectiveness of regression models because it shows how precise each regression model is and its tendencies to err on the sides of over- or under-prediction. The error distribution graphs are shown within Figure 3.

Figure 3. Distribution of Prediction Error of Regression Models

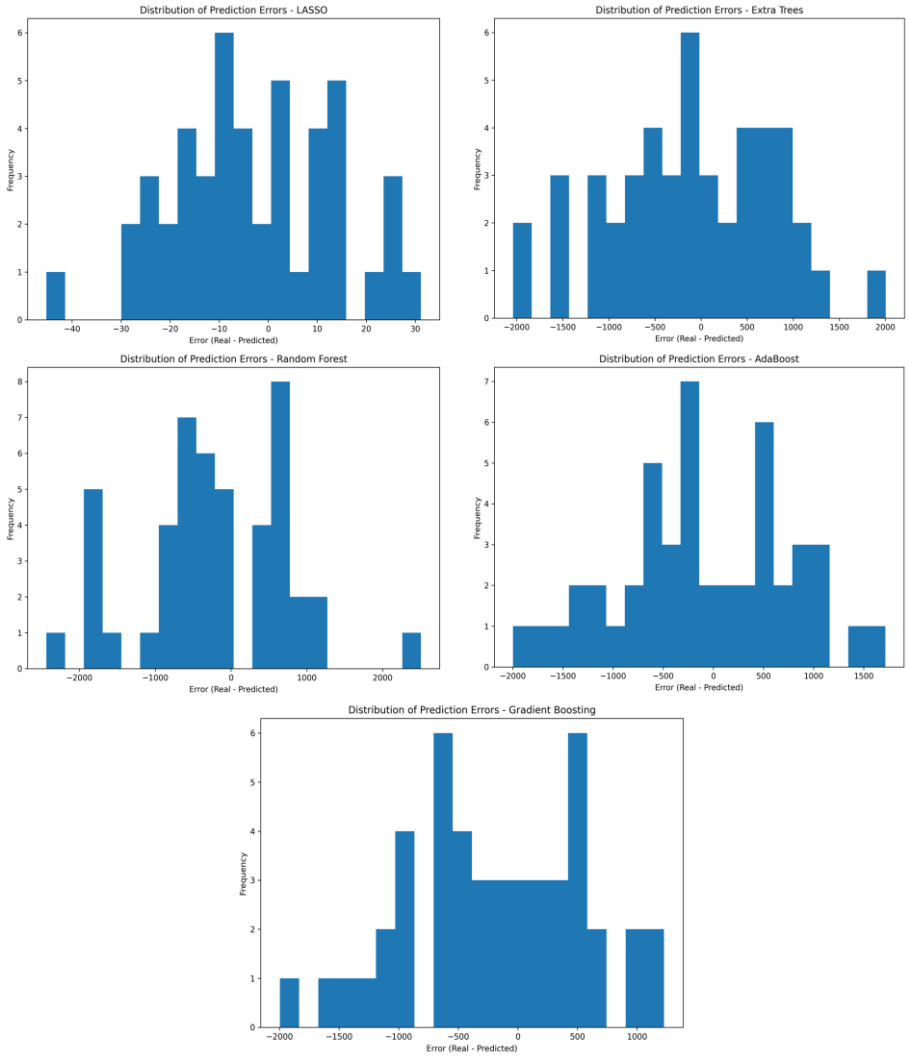


Figure 3 presents the distributions of prediction errors (Real − Predicted) for LASSO and the four tree-based models. Overall, the error histograms for all models are centered close to zero, suggesting that the models do not exhibit a strong systematic bias across the majority of observations. Nevertheless, the spread and tail behavior differ across methods, and these patterns are consistent with the cross-validated error metrics reported in Table 3. The LASSO Regressor shows a very narrow error distribution concentrated tightly around zero, indicating highly consistent predictions and very small deviations. This behavior aligns with the near-perfect agreement observed in the Real vs Prediction plot and is consistent with its extremely low cross-validated error levels in Table 3. Among the tree-based models, the Gradient Boost Regressor has a relatively more concentrated distribution with most of the errors being concentrated around zero and with limited dispersion in the tails of the distribution relative to the other ensemble models. This indicates a high degree of robustness and consistency in the model's predictions. The Extra Trees Regressor and the Random Forest Regressor have a generally balanced and symmetric distribution of errors with most of the errors concentrated around zero, though with a higher degree of dispersion

in the tails of the distribution than the Gradient Boost Regressor, suggesting occasional large errors though with high average performance. The AdaBoost Regressor has a more dispersed and irregular distribution with a higher degree of tail risk, suggesting a high degree of variability in response to certain observations and a higher chance of large prediction errors.

In summary, Figure 3 indicates that LASSO yields the most concentrated (lowest-variance) error profile. Among the non-linear models, Gradient Boosting demonstrates the most stable error behavior, followed by Extra Trees and Random Forest, while AdaBoost shows comparatively higher variability, consistent with the cross-validation results reported in Table 3.

6.2.4. Analysis of Real and Prediction Series of Models

In this section, the regression models' real export data and prediction values are visually compared. Examining the real and prediction series allows for evaluating the models' performance over time and more clearly demonstrating the accuracy of the predictions. The graph of the real and prediction series for the models is shown in Figure 4.

Figure 4. Real and Prediction Series of Regression Models

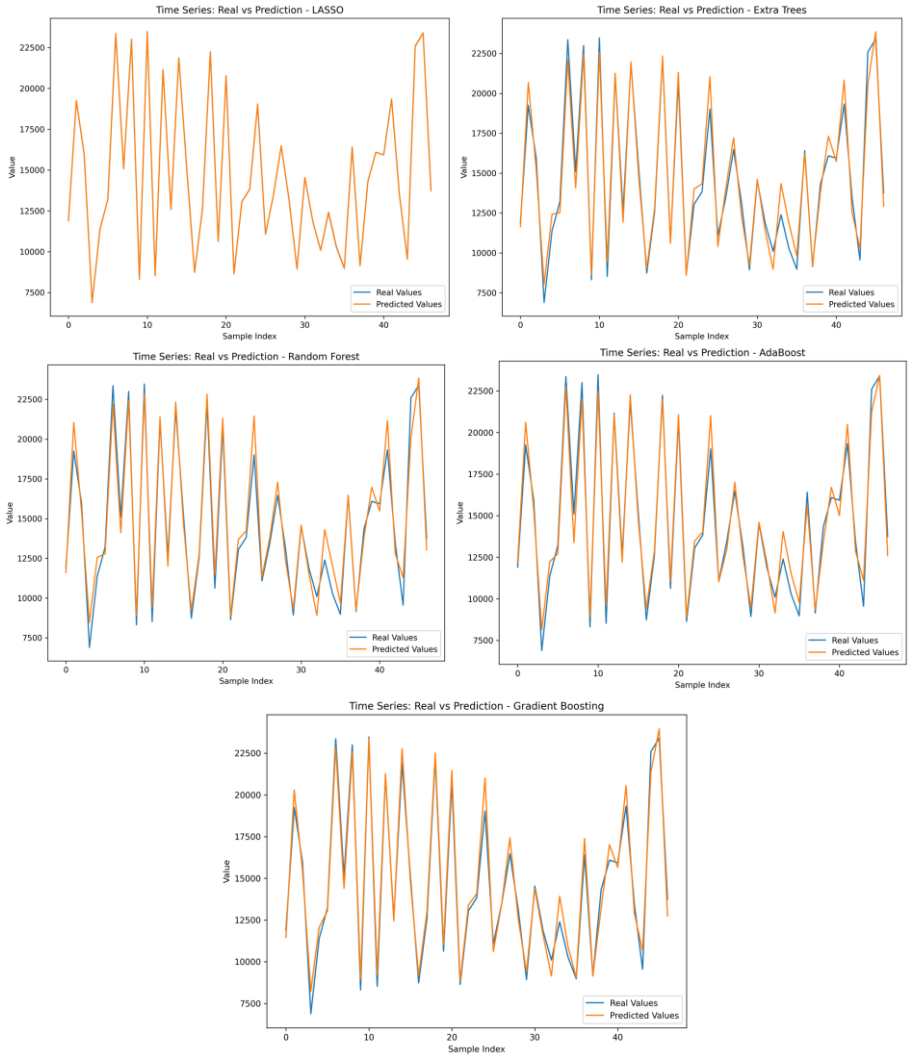


Figure 4 compares the actual and predicted export series for five regression models according to the sample order in the validation set. The substantial overlap between the actual and predicted curves across the indices indicates that the models successfully capture both level differences between observations and local fluctuations. Overall,

the close tracking of the predicted series suggests strong generalization performance, consistent with the cross-validated results reported in Table 3. Upon detailed examination, the LASSO regressor exhibits nearly perfect alignment with the actual values and closely follows the full set of fluctuations in the series. This visual fit is consistent with its very high cross-validated explanatory power and extremely low prediction errors. Among the ensemble methods, Gradient Boosting, Extra Trees, and Random Forest also demonstrate strong overlap with the actual series, with only limited deviations in certain segments. These randomized and boosting-based tree ensembles appear to learn non-linear relationships effectively and maintain stable predictive performance across the validation samples, in line with their cross-validation metrics. For AdaBoost, the predicted curve shows relatively larger deviations from the actual values in some portions of the series, indicating comparatively more conservative predictions and higher sensitivity to specific observations. Nevertheless, the overall phase alignment is preserved across models: even when minor amplitude differences occur, the direction of increases and decreases is generally captured.

Consequently, Figure 4 indicates that all models explain intra-sample variation at a reliable level, while LASSO provides the closest tracking of the actual series. Among non-linear approaches, Gradient Boosting emerges as the strongest performer, followed by Extra Trees and Random Forest, whereas AdaBoost shows relatively larger deviations, consistent with the cross-validation results in Table 3.

7. CONCLUSION AND DISCUSSION

This study is carried out with the aim of identifying the macroeconomic factors that affect export performance in Türkiye and selecting the most appropriate regression model for export prediction using those factors. For this purpose, monthly data covering the period 2006-2025 were used, and the relationship between exports and a range of macroeconomic indicators for Türkiye was first explored using correlation analysis. Following this exploratory phase, all significant variables were integrated into the models, allowing the LASSO regressor's inherent shrinkage capability to prioritize the most influential predictors. For the evaluation of the prediction performance of the developed regression models using a contemporary data science perspective, a regularized linear regression model (LASSO), as well as ensemble-based machine learning regressors (Extra Trees, Random Forest, AdaBoost, and Gradient Boosting), were also developed. Most importantly, apart from the regular evaluation using a hold-out set, the developed regression model's prediction performance is also evaluated using a cross-validation framework (Table 3), which is a less biased estimate of the generalization error. The correlation analysis suggests that the export performance of Türkiye is affected by both domestic and global economic factors. Exports is strongly and positively related to those indicators that reflect external demand conditions (global trade volumes, world import/export) as well as domestic production conditions (industrial production, prices). This indicates that changes in international trade conditions and domestic production capacity are closely related to changes in Türkiye's export conditions. Finally, the negative relationship between exports and the real exchange rate based on the CPI implies that an appreciation of the real value of the Turkish Lira is related to weaker export competitiveness, thus underlining the importance of price competitiveness in foreign trade.

The results of cross-validation of the regression models indicate that they attain a high predictive accuracy (Table 3), thus suggesting that macroeconomic indicators convey significant information regarding export outcomes. In terms of the approaches considered here, LASSO has the highest overall performance with near-perfect explanatory ability and lowest error rates under cross-validation. This indicates that, for the dataset considered here, the export–macroeconomy relationship has a very clear and significant linear structure that can be modeled very well with a regularized linear model. This exceptional performance is attributed to the inherent high correlation between trade-weighted indicators and total export volume, suggesting that the selected variables exhibit a very strong linear cointegration with Türkiye's export trajectory. Among the non-linear models considered here, Gradient Boosting has the highest cross-validation performance among the ensemble models, followed by Extra Trees and Random Forest, while AdaBoost has larger errors and variability. This is in accordance with the behavior of the models as indicated in the graphical analysis: the scatter plots of the real and predicted values and the comparison of the actual and predicted series show that LASSO closely follows the actual series, while Gradient Boosting, Extra Trees, and Random Forest follow very closely as well, with AdaBoost having larger variability in the comparisons.

One of the key methodological contributions of the current paper is the comparison between ensemble-based machine learning models and the regularized linear approach (LASSO) in the context of export forecasting. Tree-

based ensemble methods, such as Random Forest and Extra Trees, rely on multiple randomized decision trees and are therefore more stable in their performance. Boosting-based ensemble methods, such as AdaBoost and Gradient Boosting, rely on sequential improvement of the accuracy of the model by focusing on correcting past errors; however, their sensitivity might vary depending on the specific data set. In the current study, Gradient Boosting is more stable and accurate, while AdaBoost is more sensitive and generates larger deviations. These results are also in line with the findings of error distribution analysis, where errors are concentrated around zero, indicating that there is no significant bias in these models, but rather differences in terms of variability. In terms of interpretation of results from a policy perspective, these results indicate that exports from Türkiye are highly responsive to global trade conditions as well as domestic production-related indicators in the short and medium term. This is because trade activity-related variables and industrial production had significant explanatory power in these models. At the same time, the results also indicate that export competitiveness is closely related to relative price conditions, as captured by the real effective exchange rate. Therefore, it is suggested that policies designed to support production capacity, cost competitiveness, and adaptability to global demand conditions will be more effective in supporting export performance.

The findings of this current study are generally consistent with the emerging empirical literature on the strong prediction properties of machine learning models for export forecasting. For example, Can and Gerşil (2025) find that their artificial intelligence-based models outperform traditional models in predicting export values for Türkiye, while Soydal et al. (2024) similarly document the prediction properties of ensemble algorithm-based models for forecasting trade volume between Türkiye and Balkan countries. In an international context, Al Farisy et al. (2024) also emphasized that high explanatory power is obtained for export prediction models by models of the type of Random Forest. Similarly, to the above-mentioned studies, our study's findings verified that it is possible to successfully model Türkiye's export dynamics with data-based regression models. A distinguishing feature of our study is the outstanding performance of the LASSO regressor with cross-validation, implying that there is a strong linear component to the relationship between exports and macroeconomic variables in the long dataset (2006-2025), in addition to the non-linear component identified with Gradient Boosting. This study further reinforces the body of research confirming that machine learning models offer a strong complementary approach to more traditional econometric methods for export forecasting.

This study differs from a part of the existing literature in its use of a data-driven predictive approach and a comparative evaluation of various regression algorithms, both machine learning ensemble and regularized linear methods, on a Türkiye-centric macroeconomic data set. Furthermore, rather than depending on GDP as an indicator of other economic factors, a more direct method has been used by analyzing the dynamics of exports, such as industrial production, trade volume, exchange rate indicators, and interest rates. This provides a clearer view of the mechanisms involved. Ultimately, the reliability of the results obtained from cross-validation increases the credibility of the evaluation of the different models, suggesting that the results are not due to any particular train-test split. Therefore, the study provides both methodological and practical contributions.

DECLARATION OF THE AUTHOR

Declaration of Contribution Rate: The author contributes the study on his own.

Declaration of Support and Thanksgiving: No support is taken from any institution or organization.

Declaration of Conflict: There is no potential conflict of interest in the study.

Declaration of Ethics: The study does not necessitate an approval of ethical committee.

YAZARIN BEYANI

Katkı Oranı Beyanı: Yazar, çalışmanın tümüne tek başına katkı sağlamıştır.

Destek ve Teşekkür Beyanı: Çalışmada herhangi bir kurum ya da kuruluştan destek alınmamıştır.

Çatışma Beyanı: Çalışmada herhangi bir potansiyel çıkar çatışması söz konusu değildir.

Etik Beyanı: Etik kurul belgesi gerektiren bir çalışma değildir.

REFERENCES

Aktaş, C. (2009). Türkiye'nin ihracat, ithalat ve ekonomik büyüme arasındaki nedensellik analizi. *Kocaeli Üniversitesi Sosyal Bilimler Dergisi,* (18), 35–47.

Al Farisy, A. A., Mayangsari, C., Al Amany, S. U., Sartika, D., & Harlan, F. B. (2024, October). The determinants of Indonesia export performance: Gravity model and random forest model. In *2024 Ninth International Conference on Informatics and Computing (ICIC),* (p. 1–6). IEEE. Medan, Indonesia.

Alghushairy, O. (2025, May). Enhancing global logistics forecasting: Leveraging machine learning approaches for predicting import and export flows. In *2025 International Conference on Innovation in Artificial Intelligence and Internet of Things (AIIT),* (p. 1–15). IEEE. Jeddah, Saudi Arabia.

Bakari, S. (2017). *The nexus between export, import, domestic investment and economic growth in Japan* (MPRA Paper No. 76110). *Munich Personal RePEc Archive (MPRA)*. https://mpra.ub.uni-muenchen.de/76110/

Balcılar, M., Bal, H., Algan, N., & Demiral, M. (2014). Türkiye'nin ihracat performansı: İhracat hacminin temel belirleyicilerinin incelenmesi (1995–2012). *Ege Akademik Bakış, 14*(3), 451–462.

Bista, R. B. (2004). FDI in Nepal: Regression analysis in aspect of GDP growth and export. *Economic Journal of Development Issues, 5*(2), 1–10.

Bolat, İ. (2025). Türkiye'de sanayi üretim endeksini etkileyen faktörlerin analizi. *Ekonomik ve Sosyal Araştırmalar Dergisi, 21*(1), 20–35.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Can, Ş., & Gerşil, M. (2025). İhracat tutarı tahminlerinin R ile modellenmesi ve model performanslarının karşılaştırılması. *Alanya Akademik Bakış, 9*(1), 131–145. https://doi.org/10.29023/alanyaakademik.1507145

Çakmak, M. A., Kurt, M. E., & Çakmak, C. (2022). Makine öğrenmesi algoritmaları ile Covid-19 hastalarının mortalite risklerinin hesaplanması. *Süleyman Demirel Üniversitesi Vizyoner Dergisi, 13*(35), 994-1011. https://doi.org/10.21076/vizyoner.1074212

CPB. (2025). *CPB World Trade Monitor*. Retrieved August 2025, from https://www.cpb.nl/en/wtm/cpb-world-trade-monitor-july-2025

Çelik, T., & İlkay, S. Ç. (2016). Türkiye'de ihracat ve ithalat arasında uzun dönemli ilişkinin 1989–2015 dönemi için test edilmesi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 21*(3), 961-969.

Değirmen, S., & Elmas, F. (2008). Türkiye'de faiz ve kur politikalarının ihracat üzerindeki etkisi. *Ekonomik Yaklaşım, 19*(69), 47–66. https://doi.org/10.5455/ey.10675

Emsen, Ö. (2018). Türkiye ekonomisinde ihracat ile ithalat ve döviz kuru arasındaki ilişkiler. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi, 19*(1), 78–91.

Erdoğan, A. (2016). Türkiye'nin ihracatını etkileyen faktörler: Çoklu regresyon analizi. *Sosyal Bilimler Araştırma Dergisi, 5*(2), 1–8.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gayaker, S. (2025). Döviz kuru politikası ve döviz kuru oynaklığının ihracat üzerindeki etkisi: Türkiye örneği. *Üçüncü Sektör Sosyal Ekonomi Dergisi, 60*(3), 3032–3052. https://doi.org/10.63556/tisej.2025.1524

Genç, M. C., Değer, M. K., & Berber, M. (2010). Beşeri sermaye, ihracat ve ekonomik büyüme: Türkiye ekonomisi üzerine nedensellik analizi. *Bilgi Ekonomisi ve Yönetimi Dergisi, 5*(1), 29–41.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning,* (63), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Gupta, V., & Kumar, E. (2024, May). Predicting India's high export goods trade: A machine learning approach. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE),* (p. 263–267). IEEE. Gautam Buddha Nagar, Hindistan.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Istiqlala, H. Y., & Kurdhi, N. A. (2024, October). Forecasting import-export volume and value using support vector machines and random forests. In *Proceeding International Conference on Mathematics and Learning Research,* (p. 37–49). Surakarta, Indonesia.

Karaş, G., & Karaş, E. (2017). Reel efektif döviz kuru, ihracat ve ithalat arasındaki ilişki: Türkiye özelinde ekonometrik bir değerlendirme. *Uşak Üniversitesi Sosyal Bilimler Dergisi, 10*(IASOS Özel Sayısı), 27–46.

Kasman, A. (2003). Türkiye'de reel döviz kuru oynaklığı ve bunun ihracat üzerine etkisi: Sektörel bir analiz. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 22*(2), 169–186.

Kumar, S., Basavaraj, M. S., & Pachava, V. (2025). Machine learning-driven export forecasting: A comparative analysis for MSME growth. *Journal of Economics, Innovative Management and Entrepreneurship, 3*(1), 14-24. https://doi.org/10.59652/jeime.v3i1.401

Medineckiene, M., & Kirdaite, V. (2021). Evaluation of influencing factors on Great Britain's export values. *Economics and Culture, 18*(1), 59–69. https://doi.org/10.2478/jec-2021-0005

Moschos, D. (1989). Export expansion, growth and the level of economic development: An empirical analysis. *Journal of Development Economics, 30*(1), 93–102. https://doi.org/10.1016/0304-3878(89)90052-7

Otchere, D. A., Ganat, T. O. A., Ojero, J. O., Tackie-Otoo, B. N., & Taki, M. Y. (2022). Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *Journal of Petroleum Science and Engineering, 208*, 109244. https://doi.org/10.1016/j.petrol.2021.109244

Özden, E. (2022). The dynamics affecting the export-import ratio in Turkey: A hybrid model proposal with econometrics and machine learning approach. *İktisat Politikası Araştırmaları Dergisi, 9*(2), 265–291. https://doi.org/10.26650/JEPR1088322

Pamornmast, C., Jermsittiparsert, K., & Sriyakul, T. (2013). Exchange rate and country's export competitiveness: An empirical discourse analysis. *Asian Social Science, 9*(13), 147–156.

Remeikienė, R., Gasparėnienė, L., & Sadeckas, A. (2019). The determinants of the competitiveness of Lithuanian export: Macroeconomic approach. *Business: Theory and Practice, 20*, 170–178. https://doi.org/10.3846/btp.2019.16

Ruzekova, V., Kittova, Z., & Steinhauser, D. (2020). Export performance as a measurement of competitiveness. *Journal of Competitiveness, 12*(1), 145–160. https://doi.org/10.7441/joc.2020.01.09

Sarı, A. (2010). Döviz kuru oynaklığının ithalata etkileri: Türkiye örneği. *Istanbul University Econometrics and Statistics e-Journal*, (11), 31–44.

Serra, F., Pointon, J., & Abdou, H. (2012). Factors influencing the propensity to export: A study of UK and Portuguese textile firms. *International Business Review, 21*(2), 210–224. https://doi.org/10.1016/j.ibusrev.2011.02.006

Soydal, H., Ay, M., & Koç, S. (2024). Makine öğrenimi ile ticaretin öngörülmesi: Türkiye ve Balkan ülkeleri üzerine çekim modeli analizi. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi, 27*(2), 746–765. https://doi.org/10.29249/selcuksbmyd.1545687

TCMB. (2025). *Elektronik veri dağıtım sistemi (EVDS)*. Retrieved August 2025, from https://evds2.tcmb.gov.tr/index.php?/evds/serieMarket

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Uysal, Ö., & Mohamoud, A. S. (2018). Determinants of export performance in East Africa countries. *Chinese Business Review, 17*(4), 168–178. https://doi.org/10.17265/1537-1506/2018.04.002

Velnampy, T., & Achchuthan, S. (2013). Export, import and economic growth: Evidence from Sri Lanka. *Journal of Economics and Sustainable Development, 4*(9), 147–155.

World Bank. (2025). *Commodity markets outlook*. Retrieved August 2025, from https://www.worldbank.org/en/research/commodity-markets

Xu, H., Chaiboonsri, C., & Chokethaworn, K. (2023, February). Research on export prediction model based on machine learning: Application of China's agricultural products export. In *Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing,* (p. 259–263). Sanya, China. https://doi.org/10.1145/3592686.3592733

Yapraklı, S. (2007). İhracat ile ekonomik büyüme arasındaki nedensellik: Türkiye üzerine ekonometrik bir analiz. *ODTÜ Gelişme Dergisi,* (34), 97–112.

Yenisu, E. (2019). Türkiye'de ihracat, ithalat ve ekonomik büyüme ilişkisi: ARDL sınır testi yaklaşımı. *Gaziantep Üniversitesi Sosyal Bilimler Dergisi, 18*(3), 1175–1193. https://doi.org/10.21547/jss.418935

Yousfani, K., Qureshi, M., Daniyal, M., & Ismail, M. (2024). Comparative study of machine learning (ML) and conventional time series methodologies in modelling the exports trade of Pakistan. *Indus Journal of Social Sciences, 2*(2), 349–367.

Zakaria, Z. (2013). The relationship between export and exchange rate volatility: Empirical evidence based on the trade between Malaysia and its major trading partners. *Journal of Emerging Issues in Economics, Finance and Banking, 2*(2), 668–684.

Zhu, Q. (2025). Predicting the total export-import volume of China's economic trade through machine learning. *International Journal of Mathematical, Engineering and Management Sciences, 10*(1), 63-75. https://doi.org/10.33889/IJMEMS.2025.10.1.004