ISSN: 2147-5121 / E-ISSN: 2717-7610

Yayın Aralığı: Yılda 2 Sayı - Başlangıç: 2013

İstanbul Nişantaşı Üniversitesi
Sosyal Bilimler Dergisi

Bilimsel Hakemli Dergi

Yıl: 2023 Cilt: 11 Sayı: 2

NUSBD

İstanbul Nişantaşı University
Journal of Social Sciences

Scientific Refereed Journal

Year: 2023 Vol.: 11 No: 2

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

DOI: 10.52122/nisantasisbd.1377642

COMPARISON OF PREDICTION PERFORMANCES OF REGRESSION MODELS IN MACHINE LEARNING: AN APPLICATION ON THE TURKISH MERCANTILE EXCHANGE WHEAT INDEX+

Dr. Öğr. Üyesi Hasan Arda BURHAN

Kütahya Dumlupınar Üniversitesi, İktisadi ve
İdari Bilimler Fakültesi, Ekonometri Bölümü

Kütahya/Türkiye

e-posta:arda.burhan@dpu.edu.tr

ORCID **0000-0003-4043-2652**

ABSTRACT

In recent years, assets in commodity markets have become popular. In this context, Turkey implemented a licensed warehousing system by founding the Turkish Mercantile Exchange (TMEX) to facilitate trades of Electronic Warehouse Receipts (EWRs). By employing machine learning (ML) methods to predict one of the most significant indices in the TMEX market, the TMEX Wheat Index (TMXWHT), using a daily dataset spanning from April 1, 2021, to February 20, 2023, the aim of this study is to assess various ML regression methods and propose an alternative approach for similar predictions. As the input variables, the US Dollar-Turkish Lira exchange rate (USD/TRY) and Brent crude oil prices, were selected. As verified by comparisons with the actual values and considering the Mean Absolute Percentage Error (MAPE), tree-based methods revealed better overall performance.

Keywords: Turkish Mercantile Exchange, Electronic Warehouse Receipts, Machine Learning Regression

Jel Codes: C40, C60, C80.

MAKİNE ÖĞRENMESİNDE REGRESYON MODELLERİNİN TAHMİN PERFORMANSLARININ KARŞILAŞTIRILMASI: TÜRKİYE ÜRÜN İHTİSAS BORSASI BUĞDAY ENDEKSİ ÜZERİNE BİR UYGULAMA

ÖZ

Son yıllarda emtia piyasalarındaki varlıkların popüler hale geldiği ifade edilebilmektedir. Bu bağlamda Türkiye'de kurulan Türkiye Ürün İhtisas Borsası (TÜRİB) aracılığıyla Elektronik Ürün Senedi (ELÜS) ticaretinin yapılabilmesi mümkün hale gelmiştir. TÜRİB piyasasındaki en önemli endekslerden biri olan TÜRİB Buğday Endeksi (TRBBGD) tahmininde makine öğrenmesi yöntemlerinin ve 01/04/2021-20/02/2023 dönemi günlük verilerinin kullanıldığı bu çalışmanın amacı, çeşitli makine öğrenmesi regresyon yöntemlerinin kullanılabilirliğini incelenmek ve benzer tahminler için alternatif bir yaklaşım sunmaktır. Girdi değişkenleri olarak ABD Doları-Türk Lirası kuru (USD/TRY), Brent ham petrol fiyatları, gecelik faiz oranı seçilmiştir. Gerçek TRBBGD değerleriyle yapılan karşılaştırma ve Ortalama Mutlak Yüzde Hata (OMYH) kullanılarak yapılan incelemelere göre ağaç tabanlı makine öğrenmesi yöntemlerinin diğer yaklaşımlara kıyasla daha iyi genel performans gösterdiği ifade edilebilmektedir.

Anahtar Kelimeler: Türkiye Ürün İhtisas Borsası, Elektronik Ürün Senetleri, Makine Öğrenmesi Regresyon

Jel Kodları: C40, C60, C80.

Geliş Tarihi/Received: 17.10.2023 | **Kabul Tarihi/Accepted:** 23.12.2023 | **Yayım Tarihi/Printed Date:** 30.12.2023

Kaynak Gösterme: Burhan, H. A., (2023). "Comparison of Prediction Performances of Regression Models in Machine Learning: An Application on the Turkish Mercantile Exchange Wheat Index". *İstanbul Nişantaşı Üniversitesi Sosyal Bilimler Dergisi*, 2(11) 602-623.

+ This study is the revised and edited version of the paper presented at the 7th International Conference on Economics Research and Financial Markets Congress (IERFM)

INTRODUCTION

Activities such as agriculture, energy, food, and animal husbandry play a critical role in survival and progress (Kern, 2002: 291; Araújo et al., 2021: 667). According to the United Nations Food and Agriculture Organization (FAO), the human population is expected to exceed 10 billion by 2050 (FAO, 2017: 10). The rapid increase in the human population, especially in the agricultural sector, necessitates emergency measures; therefore, it is recommended to increase global food production by approximately 60-70%. (Powell et al., 2012: 539; FAO, 2016: 11). However, there has been a remarkable instability in agricultural commodity prices in the last two decades which can be attributed to the rising integration of economies, interplay of supply and demand in the markets as well as the growing interconnectedness between commodity and security markets (Hernandez et al., 2021: 1326). In this sense, it is stated that the correlation between the economic growth of countries and the corresponding elevation of living standards cause an escalation in the demand for agricultural products (İlarslan and Yıldız, 2022: 88). Hence, the increases in population, biofuel production, and aforementioned changes in food demand in countries can be considered as demand-driven factors contributing to the rising agricultural commodity prices, whereas the rising costs of agricultural production can be attributed to factors such as adverse weather conditions, currency exchange rates, and increasing oil prices (Mollaahmetoğlu and Yaşar Akçalı, 2022: 46). On the other hand, adverse impacts on commodity prices are observed following negative monetary policy shocks (Gospodinov and Jamali, 2018: 239). Thereby, it can be clearly stated that in recent years, agricultural commodity prices have undergone cycles of escalation and decline (Tiwari et al., 2018: 470). Given this, ensuring stable prices and market equilibrium in the agricultural sector has become an increasingly pressing concern since agricultural commodities are also commonly utilized as a form of inflationary hedge investment which protects investors against inflation and other risks, and also offers the potential for hedging equity market exposures (Hernandez et al., 2021: 1328).

In addition, the estimation of agricultural commodity prices carries significant weight for all market participants. Fixing these prices prior to production is crucial for producers to avoid risk, and processors and exporters aim to fulfill their needs, while speculators seek opportunities for profit to the extent that the future predicts. (Xu and Zhang, 2022: 169). These can be addressed through the establishment of commodity markets in which multiple buyers and sellers engage in spot and derivative transactions by predetermined rules. Moreover, especially in developing countries, the lack of modern and effective trading mechanisms is considered a significant contributor to the high costs of trading (Nyarko and Pellegrina, 2022: 1). Hence, as a strategic component of the agricultural sector, the utility of commodity markets is also related to its capacity to mitigate these significant costs faced by the parties in commodity supply chains especially in these nations. Consequently, as a central clearing platform, commodity exchanges have garnered significance in tandem with commodity markets, particularly where high pre-transaction and post-transaction costs are in question. In this regard, to decrease transaction costs, and other specific purposes use of licensed warehousing and providing electronic warehouse receipts (EWRs) can be listed as the mechanisms provided by these institutions (Gün and Tahsin, 2019: 10). Therefore, it can be stated that development of licensed warehousing in the world has been accompanied by the development of organized commodity exchanges.

The commodity exchange system, initially consisting of exchanges that conducted spot transactions, has evolved into exchanges that also conduct futures trading with the development of licensed warehouses and based on this, the EWRs are started to be issued by these warehouses (Özsoy Çalış et al., 2022: 27). In the modern era of fast-paced transactions and swift communication, physical warehouse receipts are replaced by EWRs, which are electronic form of vouchers and proofs of possession of stored commodities (various agricultural products with long-term durability, such as wheat, barley, rice, rye, lentils, cotton, etc.) and provided by a digital platform to the goods' owner upon transportation of these goods to the delivery warehouse (Liu et al., 2013: 661; Pozez, 2016: 205; Aydın, 2021: 22). On the other hand, EWRs can be utilized as assurance for credit or sold directly for cash, but this is contingent upon the existence of a financial market, such as commodity exchanges, that recognizes these receipts as legitimate collateral for secured credit and provides a trading platform (Gabriel, 2012: 371).

Moreover, it is also stated that functional commodity exchanges have positive influences on enhancing agricultural markets, supporting the production of high-quality and standardized products, enabling the spread of supply over time, and encouraging price formation under free competitive conditions (Karabaş and Gürler, 2010: 200). Until 2019 in Turkey, electronic warehouse receipt (EWR) trading was being carried out regionally under the authority of ten commodity exchanges which are authorized by the Ministry of Trade. In 2019, with the establishment of the Turkish Mercantile Exchange (TMEX), these transactions were consolidated under a single umbrella to be conducted throughout the country (Doğan and Özulucan, 2023: 46). This development enabled the trade of EWRs for all investors across the country, thus taking an important step towards the widespread use of these receipts as an investment instrument and led to the establishment of sectoral depth (Özsoy Çalış et al., 2022: 27). Furthermore, the trading of EWRs in the TMEX also contributed to reducing transaction costs of market participants to a competitive level (Kılıçarslan and Sucu, 2021: 259). As stated above, EWRs linked to various agricultural commodities including barley, wheat, rye, paddy, hazelnut, lentil, corn, chickpea, cotton, etc. which are exchanged in the TMEX; in which, indices are created to gauge the price performance of these commodities or a group of them, based on the prices quoted in trades of EWRs. Examples of such indices include the TMEX Grain Index (TMXGRN), TMEX Barley Index (TMXBRL), TMEX Corn Index (TMEXCRN), and TMEX Wheat Index (TMXWHT) (TMEX, 2022).

As previously mentioned, commodity prices are subject to changes as a result of multifaceted interactions among macroeconomic variables, such as increasing commodity consumption, exchange rates, crude oil prices, as well as the policy decisions taken by countries (Gospodinov and Jamali, 2018: 239; Tiwari et al., 2018: 470). In the relevant literature, these mutual relationships have been investigated in various studies such as Nazlioglu et al. (2013), Paris (2018), Tiwari et al. (2018), Melichar and Atems (2019), Hung (2021), Chen et al. (2022) in the context of crude oil prices; by Frank and Garcia (2010), Rezitis (2015), Wang and Hu (2015), Hatzenbuehler et al. (2016), Ismail et al. (2017), Mollaahmetoğlu and Yaşar Akçalı (2022), Song et al. (2022) in terms of exchange and interest rates. Previous studies on commodity price forecasting have employed various econometric techniques, such as Vector Autoregressive (VAR), Autoregressive Integratedoving Average (ARIMA), and Vector Error Correction Models (VECM). In recent years, nonlinear prediction and machine learning (ML) methods have also emerged as promising approaches for this purpose. ML is a widely used approach for discovering relationships and patterns within datasets, and has been applied in various fields such as image recognition, fraud detection, customer relations, credit analysis, medicine, etc. (Liakos et al., 2018: 2674). Due to its data-centric nature, ML has also demonstrated significant potential in the agricultural sector, and there is a vast body of literature exploring its applications in this domain. Therefore, it can be concluded that ML methods can be accepted as effective approaches in dealing with agricultural datasets which may be linear or non-linear (Rashid et al., 2021: 63408). In line with the relevant literature, this study focuses on agricultural commodity prices by considering the EWRs in the TMEX. In this sense, the primary objective of this study can be stated as to evaluate the performance and applicability of ML regression methods, including Polynomial Regression (PR), Support Vector Regression (SVR), Decision Tree Regression (DTR), and Random Forest Regression (RFR) in predicting the TMEX Wheat Index (TMXWHT), which is one of the most significant indices in the TMEX market. To achieve this, a model that includes the USD/TRY exchange rate, crude-oil prices, and interest rate was utilized with a daily dataset of TMXWHT for the period spanning from 01/04/2021 to 20/02/2023.

This study makes a significant contribution to the literature in several ways. Firstly, this is the first study that employs ML regression methods for predicting TMEX indices. Also, a prediction model that incorporates macroeconomic variables, which are empirically validated in the literature is utilized in this study. In addition, this study can be considered a valuable reference for future studies by applying above stated ML regression methods, obtaining accurate results and these were confirmed by the related performance evaluation criteria. Furthermore, the successful utilization of ML regression methods, which are not commonly employed in Turkish agro-economic literature, contributes to the current relevant literature in Turkey. The structure of the paper is as follows: The next section presents a review of the literature. Brief explanations of the data and methods are provided in Section 2. This is followed by the applications and results, and the last section includes a discussion of the findings and the concluding remarks.

1. Literature Review

Agricultural commodity prices is a crucial concern for policy makers, market participants, including firms, consumers, and brokers. Thus, examining whether commodity prices can be predicted and investigating effective methods for this purpose can be regarded essential. As previously mentioned, econometric models have commonly been employed for predicting commodity prices. Nevertheless, in the last decade, several ML models have demonstrated effective performance due to advancements in ML algorithms and forecasting techniques (Xu and Zhang, 2022: 170). Given the substantial price volatilities and consequential impact on decision-making processes and welfare, the significance of commodity asset forecasting within the agricultural sector is self-evident. Thereby, using models that include macroeconomic and financial variables, empirical research has been undertaken to explore the predictability of these assets via the application of various approaches.

Göleç et al., (2012) employed classical least square estimation, SVR, Adaptive Neuro-Fuzzy Inference Systems (ANFIS), GENFIS3:Mamdani, Fuzzy System Modeling approach based on Improved Fuzzy Functions (FSMIFF) and Fuzzy Functions with Standard Fuzzy C-Means techniques to predict the Commodity Research Bureau (CRB) index using Shanghai and CRB index values. After determining the relationship between these indexes by the Granger causality test and VECM, predicitons were obtained. According to the results, FSMIFF method yielded more suitable results. Gargano and Timmermann (2014) investigated the predictability of commodity prices using macroeconomic and financial indicators in quarterly, monthly, and annual periods. The study utilized a multivariate regression analysis by incorporating 16 indicators and the SandP Goldman Sachs Commodity Index values. The findings revealed that the most successful results were achieved for quarterly periods. Etienne (2015) used three models, namely the no-change, autoregressive (AR), and Mixed-Data Sampling (MIDAS) models to predict monthly corn prices. The analysis focused on five high-frequency data series, including the Baltic Dry Index (BDI), the Standard and Poor's (SandP) 500 market index, the 3-month US Treasury bill interest rate, the trade-weighted US dollar index against major currencies, and the nearby crude oil futures prices. Findings revealed that the no-change and AR models exhibit much better forecasting performance compared to the MIDAS model. Rezitis (2015) investigated the interplay between a set of 30 international agricultural prices (e.g. barley, corn, rice, sorghum, wheat,. etc.), five international fertilizer prices, US dollar exchange rates, and crude oil prices by using panel VAR methods and Granger causality tests with a monthly dataset spanning from June 1983 to June 2013. The findings suggested that US dollar exchange rates and crude oil prices have an impact on international agricultural commodity and fertilizer prices. In addition, bidirectional panel causality effects between international agricultural prices and crude oil prices and, also between international agricultural prices and US exchange rates were observed. Bork et al. (2019) focused on the relationship between commodity prices and exchange rates. The study included a dataset comprising 19 commodities, including nine agricultural commodities, and seven exchange rates and Granger causality approach is used. However, the findings did not provide conclusive evidence in support of the hypothesis that exchange rates can be used to predict movements in the commodity markets. Roy and Bhar (2020) utilized a time-varying parameter model and examined the association between fluctuations in the export commodity prices and Australian dollar-US dollar exchange rate. Through the analysis of monthly data, the reationship between exchange rate and commodity prices was discovered. Wang et al. (2020) conducted an analysis of the predictability of changes in several commodity prices by using regression analysis with a set of eight variables, such as exchange rates, inflation, in addition to technical indicators such as Momentum Rule (MOM), Filtering Rule (FR), Moving Average (MA), Oscillator Trading Rule (OSLT), and the Support-Resistance Rule (SR). The findings suggested that technical information resulted in stronger and more robust predictability than traditional economic information which is widely used. As a recent one that did not include a prediction but closeley related to the scope of this study, Mollaahmetoğlu and Yaşar Akçalı (2022) investigated the causal relationship between the EWR indices of the TMEX, the US Dollar Index (DXY), and the USD/TRY exchange rate, utilizing daily datasets of 01.04.2021 to 09.05.2022 period. The findings of the study indicated a bi-directional causality among the USD/TRY exchange rate and the grain index, and a uni-directional causality from the Dollar USD/TRY to the corn, barley, and wheat index. Furthermore, it was stated that the wheat index exhibits a positive response to a negative shock in the DXY and USD/TRY.

In accordance with the scope of this study, related literature was also reviewed by considering wheat prices and the use of ML methods in particular. Özçelik and Özer (2006) conducted a study to investigate the relationship between wheat production and its price in Turkey. The Koyck model was emplyed and analysis included a yearly dataset covering the period from 1973 to 2004. Based on the findings, a significant correlation between wheat production and prices, with the effect of prices on production was observable up to three years in the past. Dörtok and Aksoy (2018) used Generalized Method of Moments (GMM) estimation method and modelled a simultaneous framework including the wheat product's supply, demand, foreign trade, and price dimensions between 1961 and 2013. As a result of the analysis, stock, fertilizer price, and export were determined as the variables that affect the wheat prices. Çayır (2019) investigated the influence of EWR prices on the local commodity exchange prices, and farmer-to-merchant deal. With a dataset for the period from July 2014 to October 2018 and by utilizng the Granger causality test, it was seen that EWR prices did not have influence on the exchange prices and off-exchange market prices. On the other hand, both exchange prices and off-exchange market prices affect the EWR prices. To forecast wheat prices by using past values, Dias and Rocha (2019) conducted a comparative analysis of several modelling techniques, including Support Vector Machines (SVM), ARIMA, Multivariate Adaptive Regression Splines (MARS), Classification and Regression Trees (CART), and Random Forest (RF). Main objective of the study was to predict monthly wheat prices for the upcoming six-months and, the MARS-based approach appeared to be the method that yielded most consistent outcomes. The objective of Amin's (2020) study was to forecast the prices of seven daily commodities, including wheat, by utilizing ensemble ML models such as Light Gradient Boosted Machine (LighGBM), GradientBoost, eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), Bagging, and SVR. Based on the findings, ensemble models exhibited excellent performance on medium to large datasets and provided better performance than the SVR in accuracy. Sun et al. (2021) examined the predictability of China's wheat futures prices by using financial variables such as the futures price of wheat in Russia, the futures price of rice, the actual interest rate, and the premise and a daily dataset from June 15, 2010, to October 20, 2020. Use of a traditional model to predict the wheat futures price, is followed by the evaluation of the robustness of the results with various appraoches such as ARIMA, Artificial Neural Network (ANN) with one hidden layer, the random walk, and others. The findings suggest that ANN showed better performance in forecasts considering all the models examined in this research and proved its potential to predict China's wheat futures returns. By taking the contribution of mercantile exchanges into account, İlter Küçükçolak (2022) studied the market factors that have influences on the formation of agricultural commodity prices,. The model incorporated USD/TRY exchange rate, Brent oil price, overnight interest rate, wheat and maize prices. By utilizing GARCH methodology and two daily datasets from January 4, 2016 to July 25, 2019, and July 26, 2019 to December 31, 2020, the price volatility was analyzed. The findings indicate that, the USD/TRY exchange rate and overnight interest rate have a positive effect on wheat prices, while Brent oil price has a negative effect. Additionally, the USD/TRY exchange rate has a positive effect on maize prices. Kayral and Aksoy (2022) analyzed the potential presence of the day-of-the-week effect on the TMXBRL, TMXWHT, and TMXCRN indices of the TMEX by employing the GARCH (1,1) model and using the return series between 01.04.2021 and 24.01.2022. According to the findings, on Mondays and Thursdays abnormal returns may be obtained in the TMXWHT and TMXCRN, and on Mondays and Tuesdays this will be the case in the TMXBRL index. To predict the prices of eight agricultural commodities, such as, wheat, soybean oil, coffee, cotton, etc., Xu and Zhang (2022) conducted a study to examine the efficacy of integrated nonlinear autoregressive (NAR) and ANN models with different variations and a dataset includes more than a 50 year data. Accurate results were obtained and therefore, it is stated that the integrated NAR and ANN models may be useful for forecasting the prices of agricultural commodities. By using daily data spanning from January 2002 to December 2020, Ben Ameur et al. (2023) examined the applipcability of several deep learning (DL) algorithms for predicting the prices of certain commodities,. The findings reveal that effective tool for forecasting commodity prices is the the Long Short-Term Memory (LSTM) approach and DL methods enable the monitoring of commodity markets that are subject to structural changes in relation to other traded assets.

2. Methodology

2.1. Data Description and Preprocessing

This study employed a model that include multiple variables and a daily dataset spanning from 01/04/2021 to 20/02/2023. Specifically, the dataset included the US Dollar-Turkish Lira exchange rate (USD/TRY) and Brent crude-oil prices, as well as the overnight interest rate, all of which were sourced from the Central Bank of the Republic of Turkey (CBRT). Additionally, the dataset incorporated the wheat index of the TMEX (TMXWHT), which was obtained from the TMEX database. Figure 1 shows the index values over the handled period:

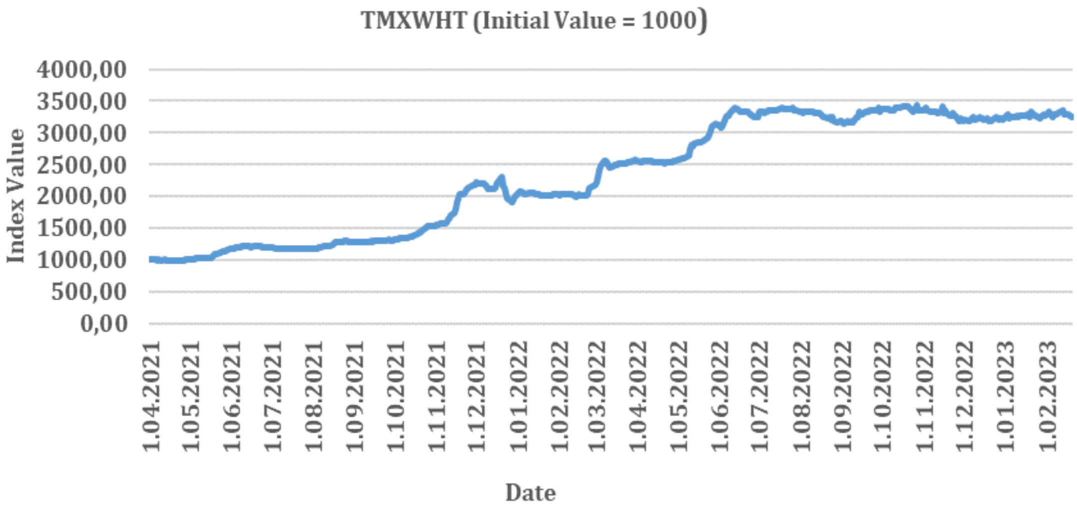

Figure 1. TMEXWHT Index
Source: Turkish Mercantile Exchange

The TMEX indices are created to assess the price performance of an individual or a set of agricultural commodities through the employment of EWR trade quotations in this exchange's EWR market, with an initial value of 1000 and a first calculation date of 01/04/2021; while the TMXWHT index, as a composite commodity index, incorporates both liquidity and production factors in its calculation method (TMEX, 2022). According to Figure 1, increases and stable periods can be observed during the handled period. In addition, time series graphs of the USD/TRY, Brent crude-oil prices, and overnight interest rate are shown in Figure 2 in which a certain increase after 2022 is seen in USD/TRY; however, decreases in oil prices and interest rates were observed during this year.

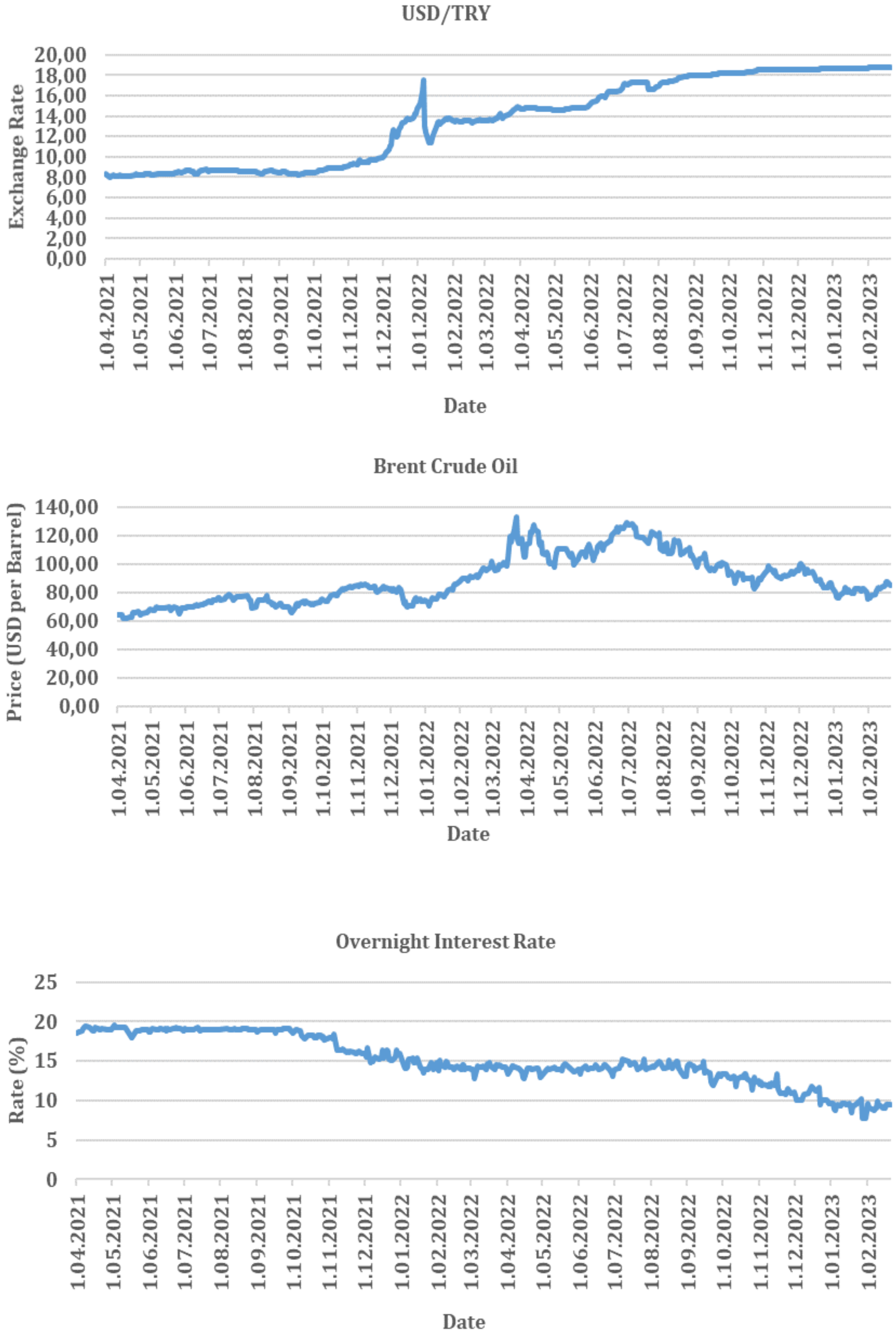

Figure 2. USD/TRY, Brent Crude-Oil Prices, Overnight Interest Rate
Source: CBRT Electronic Data Delivery System

Firstly, in the data pre-processing stage where Python programming language is used, the Simple Imputer class was used from the Scikit-learn library to take care of the missing values in the dataset. Afterwards, the data underwent feature scaling for Support Vector Regression (SVR) application as it was stated that this procedure improves the performance of the SVR method (Lin et al., 2018: 123). Unlike SVR, other methods in this study did not necessitate feature scaling.

2.2. Methods

Machine learning regression methods form a diverse set of techniques with unique applications and theoretical foundations in predictive modeling. Among these, SVR, Polynomial Regression (PR), Decision Tree Regression (DTR), and Random Forest Regression (RFR) stand out as prominent approaches, each offering distinct advantages and insights into relationships within datasets. Brief explanations of these methods are given in the following sub-sections.

2.2.1. Polynomial Regression

In a regression model such as; $y = \phi^T(x)\theta + \varepsilon,\ x \in R$, and the input space of one-dimension space turns into a feature space of K-dimensions as $\theta \in R^K$ (Deisenroth et al. 2020: 295). When data represents a more complex nature than a straight line, this is mainly the case and by using powers of features in the PR method and performing the training, this nonlinear relationship can still be shown as a multiple linear regression (Géron, 2019: 188; Raschka et al., 2022: 294). Therefore, it can be stated that the interaction between dependent and independent variables is provided within a Kth degree polynomial modelling (Maulud and Abdulazez, 2020: 141). As an example, a mapping $\psi : R \rightarrow R^{n+1}$ can be identified such that $\psi(x) = (1, x, x^2, ..., x^n)$; hence:

$$p(\psi(x)) = a_0 + a_1x + a_2x^2 + ... + a_nx^n = (a, \psi(x))$$

can be used to obtain optimal coefficients vector a by using the least squares algorithm (Shalev-Shwartz and Ben-David, 2014: 126). In addition, it can also be stated that as a special case of linear regression, above mentioned predictions of PR are also known as interactive features (Miller, 2017: 121).

2.2.2. Support Vector Regression

Obtaining a function that maximizes the discrepancy between the actual value y and the predicted values by using all training can be stated as the main objective of Support Vector Regression (SVR), since it establishes linear decision boundaries and high-dimensional feature spaces which are used for data sets that are not linearly separable (Basak et al., 2007: 203). For the squared differences as errors:

$$e_2(r^t, f(x^t)) = [r^t - f(x^t)]^2$$

a loss function that is ε sensitive in SVR can be stated as:

$$e_\varepsilon(r^t, f(x^t)) = \begin{cases} 0 & if\ |r^t - f(x^t)| < \varepsilon \\ |r^t - f(x^t)| - \varepsilon & otherwise \end{cases}$$

meaning errors up to ε are tolerated (see Alpaydın (2004)). However, it is stated that applying such mappings can be computationally inefficient, therefore the Kernel Trick (Polynomial, Gaussian, Gaussian Radial Basis Function (RBF), Laplace RBF, etc.) is used to address this challenge (Pajankar and Joshi, 2022: 157-158).

2.2.3. Decision Tree Regression

The use of binary split allows decision trees (DTs) to serve both as a regression model and a classification tool. In classification, the method predicts a class in each node, however, a numerical value is obtained at each node in regression (Géron, 2019: 248). Utilizing a binary split, the algorithm divides the data into two segments, aiming to minimize the sum of squared deviations from the mean within each segment. This process continues until all nodes reach a minimum node size, a parameter set by the user (Xu et al., 2005: 323). Furthermore, one of the important benefits of the Decision Tree Regression (DTR) can be stated as its ability to work without feature scaling either the dataset is linear or not (Raschka et al., 2022: 300). DTs are known for their high sensitivity to minor fluctuations in the training data, however, in Random Forest Regression (RFR) this can be solved by performing predictions across multiple trees (Géron, 2019: 250).

2.2.4. Random Forest Regression

The RFR, similar to DTR, is an ensemble technique that can be used for classification as well as regression (Paper, 2020: 120). The method modifies the learning process for sub-trees by utilizing bagging, and reduces the correlation between the predictions generated by these structures (Brownlee, 2017: 129). During the prediction phase, each tree evaluates a test item, with a label assigned by the decision tree and the labels obtained in return are then combined through an aggregation process and subjected to a voting mechanism and the label receiving the highest number of votes across the these trees considered as the final output (Pajankar and Joshi, 2022: 170).
In terms of generalization performance, this algorithm requires minimal parameter tuning, and enables less effect of outliers on the outcomes, thereby outperforming a single decision tree (Raschka et al., 2022: 301). Unlike linear regression, it is also asserted that RFR exhibits robustness in handling nonlinearity and overfitting (Khanal et al., 2018: 217).

2.2.5. Performance Evaluation Metrics

The criterion for comparing regressors and assessing the quality of fitting primarily relies on the level of error in predicting unseen datasets. In this context, Root Mean Square Error (RMSE) quantifies the typical disparity between values predicted by a model and the actual values (Maabreh and Almasabha, 2023):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(error)^2}{n}}, where\ error = y_i - \hat{y}_i$$

Conversely, as argued by Willmott and Matsuura (2005), RMSE exhibits certain drawbacks when compared to Mean Absolute Error (MAE). Therefore, as a more effective evaluation metric, MAE can be employed to capture the absolute variability between predicted and actual values in a dataset (Maabreh and Almasabha, 2023):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Additionally, Mean Absolute Percentage Error (MAPE) signifies the relative absolute deviation in per unit values, as discussed by Díaz et al. (2019: 616):

$$MAPE = \frac{100}{n}\sum_{j}^{n}\frac{y_j - y_j{}'}{y_j}$$

Consequently, it can be stated that when aiming to evaluate accuracy in percentage terms, the use of MAPE emerges as preferable option to facilitate enhanced interpretability and applicability across diverse scales.

APPLICATIONS AND RESULTS

In this study, Polynomial Regression (PR), Support Vector Regression (SVR), Decision Tree Regression (DTR), and Random Forest Regression (RFR) Machine Learning (ML) regression methods were evaluated by using Python programming language and related tools in the NumPy, Pandas and Sci-kit libraries. In ML applications, it is common practice to divide the dataset into the training and the test set and although there are different approaches in the literature, the dataset has been split into 80% training and 20% testing data.
As previously mentioned, in this analysis, feature scaling was only applied to the SVR model, and after this step, all of models were trained by using the training sets. In the training phase, the model undergoes optimization using a set of parameters, specifically hyperparameters, which are external configuration settings for the model and are not acquired from the data during the training process, and then the test set results are predicted. In the final step, model performances were evaluated on the test set based on the Error Measurement Criteria.

The cross-validation (CV) method is one of the recommended approaches in the literature to evaluate the success (Matloff, 2017: 26). To train and test a model k-times on different subsets of training data, and to build an estimate of the model's performance on unseen data, CV involves further splitting the training set into subsets or folds (Brownlee, 2017: 24). Thereby, in this process, the training data is divided into k parts (usually 5 or 10), with the model containing the first part used as the test set and the remaining parts used as the training set, thus allowing for performance evaluation of the first part and this is followed by the application for all other parts (Murphy, 2012: 24). Brief explanations of the steps of CV is given below (Song et al., 2021):

- Splitting the Dataset: Divide the dataset into k folds of equal size based on the selected CV method.
- Model Training and Testing: Train the model iteratively k times, utilizing k-*1* folds for training and the remaining fold for testing.
- Evaluation of Performance: Assess the model's performance by measuring prediction accuracy in each iteration.
- Performance Averaging: Combine all prediction accuracy results by calculating the average estimates, providing a more dependable evaluation of the model's overall performance.

In addition, the R-squared values were also calculated in this study to provide an alternative model evaluation approach. The obtained CV and R-squared values are given in Table 1.

Table 1. Average CV Scores of the Models

Models	CV Scores	R-Squared
Polynomial Regression	0.98	0.99
Support Vector Regression	0.99	0.98
Decision Tree Regression	0.98	0.99
Random Forest Regression	0.99	0.99

According to Table 1, the all models yielded high CV scores, meaning a good level of accuracy (high percentage) since the CV score represents the highest achievable performance of an algorithm. In addition, it is also stated that obtaining a high CV score can provide confidence that the appropriate ML technique has been selected for the related analysis (Barnard and Opletal, 2020: 12). When R-squared values are considered, it is also seen that models provided significantly high values, which may be due to overfitting. Therefore, to address this concern, the analysis proceeded by utilizing all the models and the whole dataset to estimate TMEX Wheat Index values. Initially, the next ten week days after the last day on the dataset, spanning from 20/02/2023 to 03/03/2023 were taken as a sample and utilized as the validation values. After performing feature scaling just for the SVR model, the prediction steps followed for each model are illustrated in Figure 3:

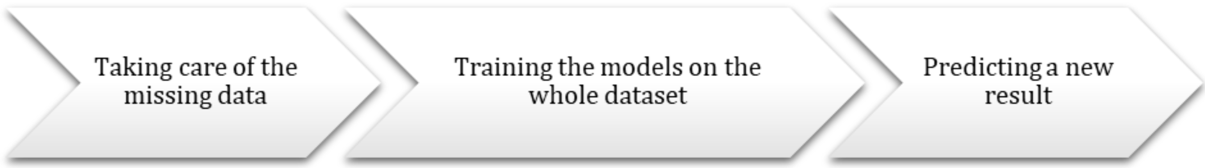

Figure 3. Prediction Steps of the Models

This step is followed by training the models on the whole dataset and finally, predictions were obtained to compare with the actual validation values. In Table 2, the performances of the utilized models are evaluated by comparing acquired predictions with actual index values:

Table 2. Prediction Performance of the Models

Date	PR Prediction	SVR Prediction	DTR Prediction	RFR Prediction	Actual Index Value
20/02/2023	3312.81	3271.53*	3286.50	3286.31	**3253.80**
21/02/2023	3283.41	3239.93	3343.50	3298.49*	**3306,95**
22/02/2023	3218.70	3241.56	3330.16*	3299.23	**3333,52**
23/02/2023	3304.52	3271.83*	3286.50	3285.83	**3262,02**
24/02/2023	3297.11	3256.74*	3286.50	3285.20	**3268,85**
27/02/2023	3285.01*	3185.33	3343.50	3298.19	**3266,74**
28/02/2023	3301.03	3247.69	3343.50	3297.30*	**3284,08**
01/03/2023	3330.81	3272.96*	3286.50	3288.28	**3239,91**
02/03/2023	3346.62	3281.79*	3283.36	3289.50	**3236,54**
03/03/2023	3342.32	3273.51	3301.90	3324.62*	**3329,18**

*denotes the most accurate prediction on the given date

Based on the findings presented in Table 2, it can be asserted that all models yielded predictions in close proximity to the actual index values. For instance, on 20 February 2023, the actual index value was 3253.80. The SVR method provided the closest prediction at 3271.53, while the least accurate prediction was observed with the PR method at 3312.81. On the other hand, on 21 February 2023, with an actual index value of 3306.95, the RFR method produced the closest prediction at 3298.49, whereas the least accurate prediction was obtained with the SVR method at 3239.93. Overall, the most accurate number of results for the handled period belonged to SVR and this is followed by the RFR. Figure 4 provides a comparison between the predicted and actual values over the validation period to offer a clearer perspective:

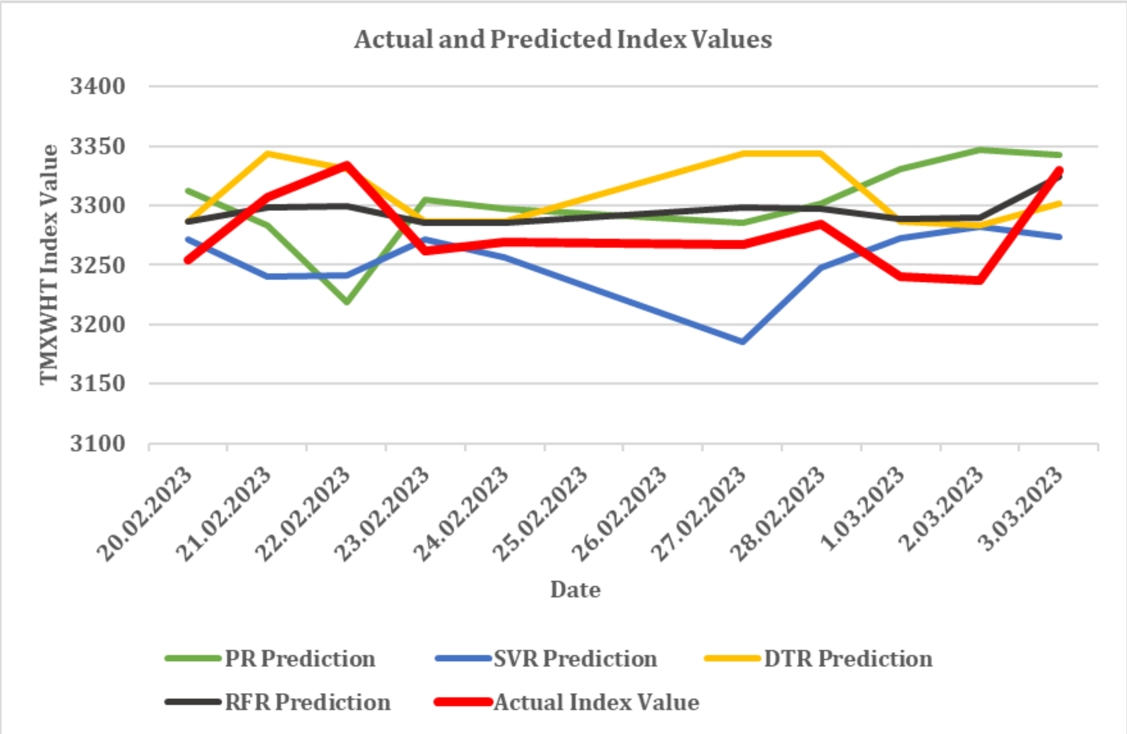

Figure 4. Comparison of Actual and Predicted Values of the TMXWHT Index

When Figure 4 is considered, it is evident that all predictions almost closely align with the actual index values. However, the DTR and RFR models exhibit a more similar trend to the actual state when compared to the other models. In conclusion, prediction performances were evaluated and as a commonly used performance metric that expresses the prediction errors as a percentage, Mean Absolute Percentage Error (MAPE) was utilized for this purpose (De Myttenaere et al. 2016: 38). Functioning as an assessment metric, Mean Absolute Error (MAE) serves as a performance metric by capturing the absolute variance between predicted and actual values in a dataset, measuring the average absolute difference, while Mean Absolute Percentage Error

(MAPE) assesses the average percentage difference (Díaz et al., 2019: 616; Maabreh and Almasabha, 2023). In this sense, calculated MAPE values are presented in Table 3:

Table 3. MAPE Scores

Models	MAPE Scores
Polynomial Regression	0.0278
Support Vector Regression	0.0349
Decision Tree Regression	0.0168
Random Forest Regression	0.0157

In general, it is widely acknowledged that the accuracy of the prediction is considered to be high if the obtained MAPE value is below 10%; in addition, a value calculated between 10% and 20% is indicative of a good level of accuracy, while if it is obtained between 20% and 50% , the predictions can still be considered reasonable (Giraka and Selvaraj, 2020: 488). By considering the scores in Table 3, it can be clearly stated that in this analysis, a high level of accuracy was evident for all models. However, tree-based models yielded much better overall performance compared to other approaches.

DISCUSSION

Agriculture is widely acknowledged as the primary source that provides sustenance and livelihood for people. Nonetheless, over the past two decades, agricultural commodity prices have exhibited considerable volatility, because of various factors, such as the rising integration of economies, increasing world population, biofuel production and agricultural production costs, as well as the complex mutual relationship of supply and demand in the markets, and the distinct link between commodity and security markets. In addition, it is stated that the active participation of global investors, traders, and speculators in agricultural commodity markets is an additional factor that contributes to the boost in agricultural prices, as these players increasingly take commodity markets as alternative investment areas for hedging and portfolio diversification purposes in the global financial markets (Nazlioglu et al., 2013: 659). Therefore, prediciting these prices can be considered as a crucial task both for policy-makers and for the participants of these markets, since producers aim to fix these prices before production, while exporters and processors seek to satisfy the requirements and speculators look for profit opportunities (Xu and Zhang, 2022: 169). Moreover, it can be stated that accurate forecasts is also valuable for hedging decisions as these strategies are used for minimizing volatilities in the market (Vancsura et al., 2023: 27). Along with the impact of globalization and these asset allocations, called *"financialization of commodities"*, commodity markets were emerged where buyers and sellers are engaged for spot and derivative transactions. And finally, the conventional barter sessions that took place in these marketplaces have transformed into complex and advanced futures and derivatives markets, also known as commodity exchanges that have particular locations, procedures, rules, and regulations (Eleje et al., 2008: 133). Forwards, futures, and spot transactions in commodity exchanges lead to price formations, which play a significant role for economies as prices include information about real supply/demand outlook and this is used in investement and production decisions (UNCTAD, 2009: 18). It is also stated that commodity exchanges also ensure lower transactions cost, provide help in managing the price volatility related risks (Worku et al., 2016: 81). In this context, licensed warehousing and electronic warehouse receipts (EWRs) are offered in these markets (Gün and Tahsin, 2019: 10). Simply, an electronic warehouse receipt (EWR) can be defined as an electronic voucher provided by an e-commerce platform upon the deposit of goods by the owner (Liu et al., 2013: 661). To operate a market for trading EWRs and futures contracts based on these receipts, the Turkish Mercantile Exchange (TMEX) was established in Turkey (Mollaahmetoğlu and Yaşar Akçalı, 2022: 46). The TMEX provides a countrywide market for agricultural commodities such wheat, barley, rice, rye, lentil, chickpea, corn, hazelnut, cotton, soybean and olive, etc. by consolidating them in EWRs (Aydın, 2021: 26; İlter Küçükçolak, 2022: 326). In addition, indices are generated utilizing the prices quoted in trades of these receipts. such as the TMEX Wheat Index (TMXWHT), TMEX Grain Index (TMXGRN), TMEX Barley Index (TMXBRL), (TMEX, 2022).

Since a wide range of econometric and machine learning (ML) based appraches have been used in commodity price related forecasts, the principal objective of this study can be stated as to assess the accuracy of some of the widely used ML regression techniques in predicting commodity market indices. For this purpose the wheat index (TMXWHT) which is one of the key indices of TMEX is chosen for the analysis. The main reason of this choice is that wheat accounts for 44% of the total cultivated grain area in Turkey and it is also stated that starting from 2019 wheat and its related products witnessed a significant price increase trend due to the rising demand resulting from the impact of the COVID-19 pandemic (Polat, 2022: 12). In this respect, this study evaluated the effectiveness of Polynomial Regression (PR), Support Vector Regression (SVR), Decision Tree Regression (DTR), and Random Forest Regression (RFR) methods within a model including TMXWHT index, US Dollar-Turkish Lira exchange rate (USD/TRY), crude-oil prices, and overnight interest rate variables and by utilizing a daily dataset spanning from 01/04/2021 to 20/02/2023 to determine the potential applicability of these methods in similar prediciton scenarios, and to provide an alternative approach for future research in this area. Following the necessary data preparation and pre-processing steps, the performance of the employed algorithms was evaluated by the cross-validation (CV) and R-squared metrics. With regard to this, CV involved training and testing a model on various subsets of training data to create an estimate of the model's performance on unseen data. Since, obtained high scores indicated a high level of accuracy, the analysis proceeded by utilizing all the models in the prediciton of TMXWHT. The initial step involved selecting the ten consecutive weekdays following the final day of the dataset to serve as the validation sample. Subsequently, the predicted values were compared to the actual values. Based on the results, while all models produced close predictions to the actual values, the SVR model provided more number of accurate results; however, the DTR and RFR models showed a much more similar trend to the actual values when compared to the other utilized models. In the final step, the performance of the models was assessed using Mean Absolute Percentage Error (MAPE). As high accuracy in prediction is generally indicated by MAPE values falling within the range of 0% to 10%, and since all values in this study fall within this range, it can be asserted that all models demonstrate a high level of accuracy. However, the tree-based models (DTR and RFR) exhibited significantly better overall performance in comparison to the other ML methods.

As stated above in Section 2, the application of ML techniques, including artificial neural networks (ANNs), yield promising results in predicting time series data and offers an appealing alternative to traditional methods. When related studies in the literature is revisited to compare obtained results in this study, it was seen that the applicability of utilized ML algorithms is verified, similar to the studies of Dias and Rocha (2019), Amin (2020), Sun et al. (2021), and Xu and Zhang (2022) that also predicted wheat prices. The literature presents successful predictions for not only wheat prices but also other grains and food prices, as evidenced by studies such as Harris (2017), Liu and Yu (2019), Nagarajan et al. (2021) and it can be stated that the findings of the current study are also consistent with these prior works. In terms of the variables included in the model, this study can also be considered in line with the works of İlter Küçükçolak (2022), in which the relationship between wheat prices and the same independent variables used in this study -USD/TRY, Brent crude-oil price, and overnight interest rate- was examined; additionally with the study of Mollaahmetoğlu and Yaşar Akçalı (2022) in which the relationship between USD/TRY and the TMEX Wheat Index was analyzed.

The contribution of this study can be stated as follows: After reviewing the relevant literature, it can be stated that this study is the first work to employ ML techniques in forecasting TMEX indices which are based on EWRs. Secondly, it utilizes a prediction model that includes empirically validated macroeconomic variables that have influence on commodity indices and proposes a feasible approach for similar predictions since accurate results were obtained with the application of four different ML algorithms which are also confirmed by the performance evaluation criteria. In addition, successful utilization of above mentioned ML regression methods can be stated as another contribution especially when Turkish literature is considered, since these methods are not commonly used in Turkey’s existing agro-economic literature. In terms of policy implications, obtained results can both be a reference in predictions of wheat and other agricultural commodity prices which are subject to remarkable instabilities especially in the last two decades. Therefore, policymakers might explore the adoption of risk management

strategies informed by the findings of this study to alleviate the adverse effects of fluctuating commodity prices. Policy planners can also leverage the findings of this study to guide strategic development for the agricultural sector. This underscores the importance of adaptable policies that consider the dynamic relationships between the input variables employed in this study and market performance. In addition, the results offer valuable insights for those involved in the commodity markets, encompassing producers, exporters, and traders, since policymakers can aid in distributing this information to assist participants in making well-informed decisions amid evolving market conditions. Taking into account the predictive insights offered by ML methods, TMEX participants, in particular, can develop strategies to adapt to evolving market conditions. This may involve diversifying portfolios according to anticipated trends in various indices and staying well-informed about the potential influence of crucial variables, including exchange rates and commodity prices. In conclusion, the study's results can assist policymakers in formulating specific policies to improve risk management, foster market stability, and offer valuable insights for individuals engaged in commodity markets. Finally, researchers, as well as policy and decision-makers, people in managerial levels, government agencies, commodity chambers, and exchanges can benefit from this study by utilizing and improving similar methodology.
This study has some important limitations: First of all, from a variety of indices, only the wheat index of TMEX is selected. Moreover, although several other factors have effect on commodity prices, only three of these were included in the model. Another limitation is that the training of the algorithms and predictions were performed on a limited number of datasets due to the availability of TMXWHT index data; however it is a well-known fact that, a more comprehensive dataset would have a favorable effect on training and prediction capabilities of utilized algorithms. Finally, just four of the several ML algorithms were used in this study. For future studies, other available commodity indices and macroeconomic variables that are validated by the related literature within more extensive datasets can be included in predictions. In addition alternate and integrated ML methodologies can be utilized to expand the scope of this study.

REFERENCES

Alpaydin, E. (2004). *Introduction to Machine Learning.* Cambridge: The MIT Press.

Amin, M.N. (2020). "Predicting Price of Daily Commodities Using Machine Learning", *International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)* (s. 1-6). Bahrain: IEEE.

Araújo, S.O., Peres, R.S., Barata, J., Lidon, F., and Ramalho, J.C. (2021). "Characterising the Agriculture 4.0 Landscape—Emerging Trends, Challenges and Opportunities", *Agronomy* , 11(4), 667-704.

Aydın, A. (2021). "Elektronik Ürün Senedi (ELÜS) ve İslam Hukuku Açısından Tahlili", *Ticaret ve Fikri Mülkiyet Hukuku Dergisi*, 7(1), 21-36.

Barnard, A., and Opletal, G. (2020). "Selecting Machine Learning Models for Metallic Nanoparticles", *Nano Futures*, 4(3),035003,1-13.

Basak, D., Pal, S., and Patranabis, D.C. (2007). "Support Vector Regression", *Neural Information Processing*, 11(10), 203-224.

Ben Ameur, H., Boubaker, S., Ftiti, Z., Louhichi, W., and Tissaoui, K. (2023). "Forecasting Commodity Prices: Empirical Evidence Using Deep Learning Tools", *Annals of Operations Research*, 1-19.

Bork, L., Kaltwasser, P.R., and Sercu, P. (2019). "Commodity Index Construction and the Predictive Power of Exchange Rates", *Journal of Banking and Finance*, 1-49.

Brownlee, J. (2017). *Master Machine Learning Algorithms.* Retrieved from: https://machinelearningmastery.com/master-machine-learning-algorithms/.

Chen, Z., Yan, B., and Kang, H. (2022). "Dynamic Correlation Between Crude Oil and Agricultural Futures Markets", *Review of Development Economics*, 26(3), 1798-1849.

Çayır, C. (2019). Elektronik Ürün Senetlerinin Tarımsal Fiyatlara Etkisi: Türkiye Örneği. (Yayınlanmamış) Yüksek Lisans Tezi, İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

De Myttenaere, A., Golden, B., Le Grand, B., and Rossi, F. (2016). "Mean Absolute Percentage Error for Regression Models", *Neurocomputing*, 192, 38-48.

Deisenroth, M.P., Faisal, A.A., and Ong, C.S. (2020). *Mathematics for Machine Learning.* Cambridge: Cambridge University Press.

Dias, J., and Rocha, H. (2019). "Forecasting Wheat Prices Based on Past Behavior: Comparison of Different Modelling Approaches", in Computational Science and Its Applications–ICCSA 2019: 19th International Conference. *Computational Science and Its Applications–ICCSA 2019: 19th International Conference Proceedings, Part III* (s. 167-182). Saint Petersburg: Springer.

Díaz, G., Coto, J., and Gómez-Aleixandre, J. (2019). "Prediction and Explanation of the Formation of the Spanish Day-Ahead Electricity Price Through Machine Learning Regression", *Applied Energy*, 239, 610-625.

Doğan, Z., and Özulucan, A. (2023). "Lisanslı Depoculuk Sistemi ve Elektronik Ürün Senetleri ile İlgili Sağlanan Vergi ve Diğer Mali Destekler", *Vergi Raporu,*, 281, 42-56.

Dörtok, A., and Aksoy, A. (2018). "Türkiye Buğday Sektörünün Eşanlı Model Yöntemiyle Tahmini", *Kahramanmaraş Sütçü İmam Üniversitesi Tarım ve Doğa Dergisi*, 21(4), 580-586.

Eleje, E. O., Josaphat, U.O., and Nwokeji, N.N. (2008). "Commodity Exchange Markets and Economic Development", *Nigerian Journal Of Banking And Finance*, 8, 132-146.

Etienne, X.L. (2015). "Financialization Of Agricultural Commodity Markets: Do Financial Data Help To Forecast Agricultural Prices?", *2015 Agricultural and Applied Economics Association and Western Agricultural Economics Association Annual Meeting*, (pp. 1-38). San Francisco.

FAO (2016). The State of Food and Agriculture: Climate Change, Agriculture and Food Security, Rome.

FAO (2017). The Future of Food and Agriculture: Trends and Challenges, Rome.

Frank, J., and Garcia, P. (2010). "How Strong are the Linkages Among Agricultural, Oil, and Exchange Rate Markets?", *Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis.* St. Louis : NCCC.

Gabriel, H.D. (2012). "Warehouse Receipts and Cecuritization in Agricultural Finance", *Uniform Law Review* , 17(1-2). 369-376.

Gargano, A., and Timmermann, A. (2014). "Forecasting Commodity Price Indexes Using Macroeconomic and Financial Predictors", *International Journal of Forecasting*, 30(3), 825-843.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-learn, Keras, and Tensorflow.* Sebastopol: O'Reilly Media, Inc..

Giraka, O., and Selvaraj, V.K. (2020). "Short-term Prediction of Intersection Turning Volume Using Seasonal ARIMA Model", *Transportation Letters*, 12(7), 483-490.

Gospodinov, N., and Jamali, I. (2018). "Monetary Policy Uncertainty, Positions of Traders and Changes in Commodity Futures Prices", *European Financial Management*, 24(2), 239-260.

Göleç, A., Murat, A., Tokat, E., and Türkşen, İ.B. (2012). "Forecasting Model of Shanghai and CRB Commodity Indexes", *Expert Systems with Applications*, 39(10), 9275-9281.

Gün, N., and Tahsin, E. (2019). "Role of Electronic Warehouse Receipt System in Development of Commodity Exchanges: An Assessment For Turkey", *Tarım Ekonomisi Araştırmaları Dergisi*, 5(1), 9-24.

Harris, J. (2017). *A Machine Learning Approach To Forecasting Consumer Food Prices.* Nova Scotia: Halifax.

Hatzenbuehler, P. L., Abbott, P. C., and Foster, K.A. (2016). "Agricultural Commodity Prices and Exchange Rates Under Structural Change", *Journal of Agricultural and Resource Economics*, 41(2), 204-224.

Hernandez, J.A., Kang, S.H., and Yoon, S.-M. (2021). "Spillovers and Portfolio Optimization of Agricultural Commodity and Global Equity Markets", *Applied Economics*, 53(12), 1326-1341.

Hung, N.T. (2021). "Oil Prices and Agricultural Commodity Markets: Evidence from Pre and During COVID-19 Outbreak", *Resources Policy* , 73, 102236.

Ismail, A., Ihsan, H., Khan, S. A., and Jabeen, M. (2017). "Price Volatility of Food and Agricultural Commodities: A Case Study of Pakistan", *Journal of Economic Cooperation and Development*, 38(3), 77-120.

İlarslan, K., and Yıldız, M. (2022). "Do International Agricultural Commodity Prices Have an Effect on the Stock Market Index? A Comparative Analysis Between Poland and Turkey", *Sosyoekonomi*, 30(52), 87-107.

İlter Küçükçolak, N. (2022). "Ürün İhtisas Borsacılığının Gıda Fiyat İstikrarına Katkısı", *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 49, 325-339.

Karabaş, S., and Gürler, A.Z. (2010). "Lisanslı Depoculuk Sisteminin İşleyişi ve Türkiye'de Uygulanabilirliği", *Sosyal Bilimler Araştırmaları Dergisi*, 5(1), 196-210.

Kayral, İ.E., and Aksoy, L. (2022). "Yeni Gıda Endekslerinde Haftanın Günü Etkisi Var Mı?", *IBAD Sosyal Bilimler Dergisi*, 12, 461-476.

Kern, M. (2002). "Food, Feed, Fibre, Fuel and Industrial Products of the Future: Challenges and Opportunities, Understanding the Strategic Potential of Plant Genetic Engineering", *Journal of Agronomy and Crop Science*, 188, 291-305.

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., and Shearer, S. (2018). "Integration of High Resolution Remotely Sensed Data and Machine Learning Techniques for Spatial Prediction of Soil Properties and Corn Yield", *Computers and Electronics in Agriculture*, 153, 213-225.

Kılıçarslan, A., and Sucu, M.Ç. (2021). "Ticaret Borsası ve Ürün İhtisas Borsasında İşlem Maliyetleri: Karşılaştırmalı Bir Uygulama", *Sakarya İktisat Dergisi*, 10(3), 257-274.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). "Machine Learning in Agriculture: A Review", *Sensors*, 18(8), 2674-2703.

Lin, A.Y., ZHANG, M., and SELPI, S. (2018). "Using Scaling Methods to Improve Support Vector Regression's Performance for Travel Time and Traffic Volume Predictions", in I. Rojas, H. Pomares and O. Valenzuela (Eds.). *Time series analysis and forecasting ITISE 2017 contributions to statistics* (s. 115-127). Paper Presented at International Work-Conference on Time Series Analysis, Cham: Springer. https://doi.org/10.1007/978-3-319-96944-2_8.

Liu, B., Hui, H., and Li, J. (2013). "The Research of Commodity E-Commerce and Logistics Collaborative System Based on the Electronic Warehouse Receipts", *Proceedings of 3rd International Conference on Logistics, Informatics and Service Science* (s. 659-666). Reading: Springer.

Liu, N., and Yu, J. (2019). "Raw Grain Price Forecasting with Regression Analysis", *2019 International Conference on Modeling, Simulation and Big Data Analysis (MSBDA 2019)* (s. 372-378.). Dordrecht: Atlantis Press.

Maabreh, M. and Almasabha, G. (2023). "Machine Learning Regression Algorithms for Shear Strength Prediction of SFRC-DBs: Performance Evaluation and Comparisons", *Arabian Journal for Science and Engineering*.

Matloff, N. (2017). *Statistical Regression and Classification-From Linear Models to Machine Learning.* Boca Raton: CRC Press.

Maulud, D.H., and Abdulazez, A. M. (2020). "A Review on Linear Regression Comprehensive on Machine Learning", *Journal of Applied Sciences and Technology Trends*, 1(4), 140-147.

Melichar, M., and Atems, B. (2019). "Global Crude Oil Market Shocks and Global Commodity Prices", *OPEC Energy Review*, 43(1), 92-105.

Miller, J. D. (2017). *Statistics for Data Science.* Birmingham: Packt Publishing.

Mollaahmetoğlu, E., and Yaşar Akçali, B. (2022). "Elektronik Ürün Senedi (ELÜS) Endeksleri ile ABD Dolar Endeksi ve Dolar Kuru Arasındaki İlişkinin Simetrik ve Asimetrik Nedensellik Testleri ile Analizi", *Ekonomi Politika ve Finans Araştırmaları Dergisi*, 7(IERFM Özel Sayısı), 45-60.

Murphy, K. P. (2012). *Machine Learning - A Probabilistic Perspective.* Cambridge: The MIT Press.

Nagarajan, S., Feyisa, T., and Degefa, T. (2021). "Forecasting Retail Price of Consumer Food Product's: In the Case of Oromia Region, Ethiopia", *Journal of Tianjin University Science and Technology*, 54(7), 629-646.

Nazlioglu, S., Erdem, C., and Soytas, U. (2013). "Volatility Spillover Between Oil and Agricultural Commodity Markets", *Energy Economics*, 36, 658-665.

Nyarko, Y., and Pellegrina, H. S. (2022). "From Bilateral Trade to Centralized Markets: A Search Model for Commodity Exchanges in Africa", *Journal of Development Economics*, 157, 102867.

Özçelik, A., and Özer, O. O. (2006). "Koyck Modeliyle Türkiye'de Buğday Üretimi ve Fiyatı İlişkisinin Analizi", *Journal of Agricultural Sciences*, 12(04), 333-339.

Özsoy Çalış, N., Babuşçu, Ş., and Hazar, A. (2022). "Bir Tarımsal Kredi Olarak ELÜS Rehni Karşılığında Kredi Kullandırımı", *Gümrük ve Ticaret Dergisi*, 9(27), 25-41.

Pajankar, A., and Joshi, A. (2022). *Hands-on Machine Learning with Python: Implement Neural Network Solutions with Scikit-learn and PyTorch.* New York: Apress.

Paper, D. (2020). *Hands-on Scikit-Learn for Machine Learning Applications - Data Science Fundamentals with Python.* Logan: Apress.

Paris, A. (2018). "On the Link Between Oil and Agricultural Commodity Prices: Do Biofuels Matter?", *International Economics*, 155, 48-60.

Polat, K. (2022). Durum ve Tahmin: Buğday 2021/2022, Ankara.

Powell, N., Ji, X., Ravash, R., Edlington, J., and Dolferus, R. (2012). "Yield Stability for Cereals in a Changing Climate", *Functional Plant Biology*, 39(7), 539-552.

Pozez, A.M. (2016). "A Roadmap to Better Understanding the Issuance and Transfer of Negotiable Electronic Warehouse Receipts in the American Cotton Trade", *Arizona Journal of International and Comparative Law*, 33: 205-218.

Raschka, S., Liu, Y., and Mirjalili, V. (2022). *Machine Learning with Pytorch and Scikit-learn - Develop Machine Learning and Deep Learning Models with Python.* Birmingham: Packt Publishing Ltd..

Rashid, M., Bari, B.S., Yusup, Y., Kamaruddin, M.A., and Khan, N. (2021). "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction", *IEEE Access*, 9, 63406-63439.

Rezitis, A.N. (2015). "The Relationship Between Agricultural Commodity Prices, Crude Oilprices and US Dollar Exchange Rates: A Panel VAR Approach and Causality Analysis", *International Review of Applied Economics*, 29(3), 403-434.

Roy, D., and Bhar, R. (2020). "Trend of Commodity Prices and Exchange Rate in Australian Economy: Time Varying Parameter Model Approach", *Asia-Pacific Financial Markets*, 27, 427-437.

Shalev-Shwartz, S., and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms.* New York: Cambridge University Press.

Song, L., Tian, G., and Jiang, Y. (2022). "Connectedness of Commodity, Exchange Rate and Categorical Economic Policy Uncertainties — Evidence from China", *North American Journal of Economics and Finance* , 60, 101656.

Song, Q.C., Tang, C., and Wee, S. (2021). "Making Sense of Model Generalizability: A Tutorial on Cross-Validation in R and Shiny", *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920947067.

Sun, Y., Guo, J., Shan, S., and Khan, Y.A. (2021). "Wheat Futures Prices Prediction in China: A Hybrid Approach", *Discrete Dynamics in Nature and Society*, 1-9.

Tiwari, A.K., Khalfaoui, R., Solarin, S.A., and Shahbaz, M. (2018). "Analyzing the Time-Frequency Lead–Lag Relationship Between Oil and Agricultural Commodities", *Energy Economics*, 76, 470-494.

TMEX. (2022). “Turkish Mercantile Exchange”, https://www.turib.com.tr/en/indices/ adresinden alındı

UNCTAD. (2009). Development Impacts of Commodity Exchanges in Emerging Markets, New York.

Vancsura, L., Tatay, T., and Bareith, T. (2023). "Evaluating the Effectiveness of Modern Forecasting Models in Predicting Commodity Futures Prices in Volatile Economic Times", *Risks*, 11(2), 27-43.

Wang, Q., and Hu, Y. (2015). "Cross-correlation Between Interest Rates and Commodity Prices", *Physica A: Statistical Mechanics and its Applications*, 428, 80-89.

Wang, Y., Liu, L., and Wu, C. (2020). "Forecasting Commodity Prices Out-Of-Sample: Can Technical Indicators Help?", *International Journal of Forecasting*, 36(2), 666-683.

Willmott, C.J., and Matsuura, K. (2005). “Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance”, *Climate Research*, 30(1), 79-82.

Worku, M.A., Ejigu, A., and Gebresilasie, G. (2016). "The Contribution of Ethiopia Commodity Exchange for Promoting Exports of Agricultural Products", *Journal of Economics and Sustainable Development*, 7(9), 81-90.

Xu, M., Watanachaturaporn, P., Varshney, P.K. and Arora, M.K. (2005). "Decision Tree Regression for Soft Classification of Remote Sensing Data", *Remote Sensing of Environment*, 97, 322-336.

Xu, X., and Zhang, Y. (2022). "Commodity Price Forecasting Via Neural Networks for Coffee, Corn, Cotton, Oats, Soybeans, Soybean Oil, Sugar, and Wheat", *Intelligent Systems in Accounting, Finance and Management*, 29(3), 169-181.

EXTENDED ABSTRACT

GENİŞLETİLMİŞ ÖZET

MAKİNE ÖĞRENMESİNDE REGRESYON MODELLERİNİN TAHMİN PERFORMANSLARININ KARŞILAŞTIRILMASI: TÜRKİYE ÜRÜN İHTİSAS BORSASI BUĞDAY ENDEKSİ ÜZERİNE BİR UYGULAMA

Giriş ve Çalışmanın Amacı: Tarımsal emtia fiyatları, ülkelerin ekonomileri üzerinde önemli bir etki yapmakta, mali ve para politikalarında düzenlemelere neden olmaktadır. Ancak son yirmi yılda tarımsal emtia fiyatları, birçok faktör nedeniyle önemli dalgalanmalar göstermiştir. Bu bağlamda, emtia fiyatlarının tahmin edilmesi, hem politika yapıcıları hem de piyasa katılımcıları için büyük önem taşımaktadır. Emtia piyasalarına artan ilgi, riskleri yönetme, fiyat istikrarını sağlama ve işlem maliyetlerini azaltma ihtiyacını doğurmuş ve bu da emtia borsalarının kurulmasına yol açmıştır. Ayrıca söz konusu piyasalarda lisanslı depolama alanları ve Elektronik Ürün Senetleri (ELÜS) sunulmaktadır. Türkiye gelişmekte olan bir ülke olarak, Türkiye Ürün İhtisas Borsası (TÜRİB) kuruluşu aracılığıyla lisanslı depolama sistemini uygulamakta ve ELÜS ticaretini denetlemektedir. TÜRİB, çeşitli tarımsal emtialar için ulusal bir platform olarak faaliyet göstermekte ve ELÜS'ler aracılığıyla bir araya getirilen bu emtialara ait senetlerle yapılan işlemlerden elde edilen fiyatlar kullanılarak endeksler oluşturulmaktadır. Bu çalışmanın temel amacı, Polinom Regresyonu (PR), Destek Vektör Regresyonu (DVR), Karar Ağacı Regresyonu (KAR) ve Rastgele Orman Regresyonu (ROR) makine öğrenmesi algoritmalarının tahmin performansını emtia borsaları bağlamında ve TÜRİB'deki en önemli endekslerden biri olan buğday endeksi (TRBBGD) özelinde değerlendirmek olarak ifade edilebilir.

Kavramsal/kuramsal çerçeve: Tarımsal emtia fiyatları, politika yapıcıları, firmalar, tüketiciler ve komisyoncular dahil olmak üzere piyasa katılımcıları için kritik bir önem taşır. Dolayısıyla, emtia fiyatlarının tahmin edilebileceği etkili yöntemlerin araştırılması ilgili literatürde üzerinde durulan bir konu olagelmiştir. Emtia fiyatlarını tahmin etmek için kullanılan çeşitli yaklaşımlar göz önüne alınarak, bu çalışmada makine öğrenmesi regresyon tekniklerinin emtia piyasası endekslerini tahmin etmedeki başarısı değerlendirilmiş, elde edilen sonuçların gerçek değerlerle karşılaştırması ve çeşitli performans değerlendirme kriterlerinden faydalanılması yoluyla tahmin başarılarının incelenmesi amaçlanmıştır.

Yöntem ve Bulgular (Methodology and Findings): Çalışma, ABD Doları-Türk Lirası döviz kuru (USD/TRY), Brent ham petrol fiyatları ve gecelik faiz oranları değişkenlerini içreren bir model dahilinde TÜRİB'in en önemli endekslerinden biri olan buğday endeksi TRBBGD için Polinom Regresyon (PR), Destek Vektör Regresyon (DVR), Karar Ağacı Regresyon (KAR) ve Rastgele Orman Regresyon (ROR) yöntemleri odaklı bir tahminleme uygulamasını içermektedir. 01/04/2021-20/02/2023 dönemi günlük verilerinin kullanıldığı çalışmada ilgili yöntemlerin potansiyel uygulanabilirliğinin değerlendirilmesi ve böylece gelecekte bu alanda yapılacak araştırmalara alternatif bir yaklaşım sunulması amaçlanmıştır.

Sonuç ve Öneriler: Elde edilen bulgulara göre, tüm modeller gerçek değerlere yakın tahminler üretmiş olup, özellikle, DVR modeli daha fazla doğru sonuç üretmiştir. Ancak diğer modellerle karşılaştırıldığında, KAR ve ROR modellerinin gerçek değerlere daha da yakın sonuçları içeren örüntüler sergilediği görülmüştür. Son değerlendirmede, model performansları ortalama mutlak yüzde hata (OMYH) değerleri bağlamında ölçülmüş olup tüm modellerin %10 altında OMYH değerleri ile yüksek doğruluk seviyesine ulaştığı görülmüştür. Bununla birlikte, genel performansta ağaç tabanlı modellerin (KAR ve ROR), diğer makine öğrenmesi yöntemlerini geride bıraktığı ifade edilebilmektedir. Literatüre katkı bağlamında bu çalışma, ELÜS'ler temel alınarak TÜRİB endekslerini tahmin etmek için makine öğrenmesi tekniklerini kullanan ilk çalışma olup; elde edilen doğru sonuçların performans değerlendirme kriterleri tarafından da onaylanması nedeniyle benzer tahminleme amaçları için uygulanabilir bir yaklaşım önermektedir. Bu çalışmanın bazı önemli sınırlamaları vardır: İlk olarak, birçok TÜRİB endeksi arasından sadece buğday endeksi seçilmiştir. Ayrıca, emtia fiyatları üzerinde etkisi olan birçok faktör bulunmasına rağmen, sadece üç tanesi modele dahil edilmiştir. Diğer bir sınırlama ise, algoritmaların eğitimi ve tahmin çalışmalarının ilgili kurum veri tabanındaki mevcut veri miktarı nedeniyle sınırlı sayıda veri seti üzerinde gerçekleştirilmiş olmasıdır. Çalışmanın diğer bir sınırlılığı ise birçok makine öğrenmesi algoritmasından dört tanesinin kullanılmış olmasıdır.

KATKI ORANI BEYANI VE ÇIKAR ÇATIŞMASI BİLDİRİMİ

Sorumlu Yazar *Responsible/Corresponding Author*	Hasan Arda BURHAN
Makalenin Başlığı *Title of Manuscript*	COMPARISON OF PREDICTION PERFORMANCES OF REGRESSION MODELS IN MACHINE LEARNING: AN APPLICATION ON THE TURKISH MERCANTILE EXCHANGE WHEAT INDEX
Tarih *Date*	17.10.2023
Makalenin türü (Araştırma makalesi, Derleme vb.) Manuscript Type (Research Article, Review etc.)	Research Article

Yazarların Listesi / *List of Authors*

Sıra No	**Adı-Soyadı** *Name - Surname*	**Katkı Oranı** *Author Contributions*	**Çıkar Çatışması** *Conflicts of Interest*	**Destek ve Teşekkür (Varsa)** *Support and Acknowledgment*
1	Hasan Arda BURHAN	%100	-	-