JOEEP

e-ISSN: 2651-5318
Journal Homepage: http://dergipark.org.tr/joeep

Araştırma Makalesi ● Research Article

Stock Market Index Prediction Using Machine Learning Techniques: Application of BIST Indices

Makine Öğrenme Tekniklerini Kullanarak Borsa Endeksi Tahmini: BIST Endeksleri Uygulaması

Hayrettin Uzunoğlu[a,*], Sevgi Sümerli Sarıgül[b] & Ramazan Aldemir[c]

[a] Dr.Öğr. Üyesi, Kayseri University
ORCID: 0000-0003-1550-5549
[b] Doç. Dr., Kayseri University
ORCID: 0000-0002-3820-6288
[c] Dr.Öğr. Üyesi, Kayseri University
ORCID: 0000-0002-4279-3649

M A K A L E B İ L G İ S İ

Makale Geçmişi:
Başvuru tarihi: 02 Nisan 2024
Düzeltme tarihi: 10 Ekim 2024
Kabul tarihi: 22 Ekim 2024

Anahtar Kelimeler:
BIST
Yapay Zekâ
Makine Öğrenme

ÖZ

Hisse senedi piyasası, ülke ekonomilerinin önemli göstergelerinden biri olup, bu piyasanın bileşenleri arasındaki ilişkiler birçok çalışmada araştırılmıştır. Hisse senedi piyasasında geleceğe yönelik tahmin çalışmaları hem firma sahipleri hem de yatırımcılar için oldukça önemlidir. Bu yüzden hisse senetlerinin gelecek fiyatını tahmin etmeye yönelik çok sayıda model geliştirilmiştir. Özellikle yapay zekânın önem kazanmaya başladığı günümüzde geleceğe yönelik tahmin modellerinde artık makine öğrenmesi modelleri popüler hale gelmiştir. Bu kapsamda çalışmamızda, çeşitli makine öğrenme algoritmaları kullanılarak Borsa İstanbul sektör endeksleri arasında yer alan; Sınai Endeksi (XUSIN), Hizmetler Endeksi (XUHIZ) ve Mali Endeks (XUMAL) firmalarının 2019-2024 yılı verileri analiz edilmiştir.

A R T I C L E I N F O

Article history:
Received: April 02, 2024
Received in revised form: Oct 10, 2024
Accepted: Oct 22, 2024

Keywords:
BIST
Artificial Intelligence
Machine Learning

A B S T R A C T

The stock market is one of the important indicators of national economies and the relationships between the components of this market have been investigated in many studies. Forecasting in the stock market is very important for both firm owners and investors. Therefore, many models have been developed to predict the future price of stocks. Especially in today's world where artificial intelligence is gaining importance, machine learning models have become popular in future forecasting models. In this context, in our study, the 2019-2022 data of the Industrial Index (XUSIN), Services Index (XUHIZ) and Financial Index (XUMAL) companies, which are among the Borsa Istanbul sector indices, were analysed using various machine learning algorithms.

1. Introduction

The stock market is one of the important indicators of the economies of the countries, and the relationships between the components of this market have been investigated in many studies. Forecasting for the future in the stock market is very important for both company owners and investors. For this reason, many models have been developed to predict the future price of stocks. Especially today, when artificial intelligence has started to gain importance, machine learning models have become popular in prediction models for the future (Örnek,2023).

Artificial intelligence (AI) and Machine learning (MO) are

* *Sorumlu yazar/Corresponding author*.
e-posta: *hayrettinuzunoglu@kayseri.edu.tr*

Atıf/Cite as: Uzunoğlu, H., Sümerli Sarıgül, S. & Aldemir, R. (2024). Stock Market Index Prediction Using Machine Learning Techniques: Application of BIST Indices. Journal of Emerging Economies and Policy, 9(2), 96-106.

making huge impacts in many industries today. Especially in the financial sector, it is used in complex and uncertain topics such as stock price prediction. MO algorithms attempt to predict future market movements by analyzing large amounts of financial data. These forecasts are critical in the strategic decisions of investors and companies.

Artificial intelligence (AI) is an overarching term for the different strategies and techniques used to make machines more human-like. Machine learning (ML) is one of the many branches of artificial intelligence. ML is the science of developing algorithms and statistical models that computer systems use to perform complex tasks without explicit instructions. Systems rely on patterns and inference rather than explicit instructions. Computer systems use ML algorithms to process large amounts of historical data and identify data patterns. While machine learning is artificial intelligence, not all artificial intelligence activities are machine learning.

Artificial intelligence (AI) is a discipline that aims to enable computer systems to have human-like intelligence and learning capabilities. It aims to enable computers to behave like humans by using algorithms and techniques to perform tasks such as problem solving, decision making, natural language processing and visual perception.

Artificial intelligence was born from the idea of mathematically modelling the working principle of the human brain. This idea is a simulation of the situation in which "the brain at the centre of the nervous system continuously receives and processes information with biological neural networks and makes appropriate decisions according to the results of the neural networks (Haykın, 2009, 36). The main purpose of AI is to develop computer systems that can think and learn like humans. These systems can analyse large data sets, interpret the results and make better decisions using the results. However, artificial intelligence still cannot imitate all of human intelligence and can only achieve limited success in certain areas.

The term "big data" is used when data sets are large, complex and diverse. This term refers to situations where traditional data processing methods are insufficient due to the size, speed and variety of data sets. Big data and artificial intelligence are two complementary technologies. Artificial intelligence is used to analyse, interpret and learn from large data sets. Big data, on the other hand, provides the data necessary for artificial intelligence to learn and make decisions.

Big data is the mass of data kept in different and relational databases such as social media posts, network logs, blogs, photos, videos, log files, microblogs, information from climate sensors and similar sensors, call records obtained from GSM operators, etc., which are outside the structured data. Big data is quite large, complex and dynamic, and it is too large to be managed by a standard software such as Microsoft and is measured in units such as terabytes and petabytes (Vassakis et al., 2018; Bilik & Aydın, 2018). As a result, when big data and artificial intelligence are used together, they form a powerful combination for making more accurate and effective decisions. Big data is essential for the training and development of artificial intelligence models. Artificial intelligence also provides the technologies required for big data analysis and is used to better understand patterns and relationships in large data sets.

The issue of the predictability of stock returns is one of the most researched topics in the finance literature due to the lack of any method that can accurately determine stock price behaviour. The stock market is highly dynamic in nature, meaning that stock prices can change within minutes or even seconds. Due to the high uncertainty and volatility that make it difficult to predict stock price behaviour, stock investments carry more risk than other investment instruments. In this context, machine learning approach is a relatively new, active and promising field in the prediction of stock price behaviour.

Machine learning methods are popularly applied to many financial problems such as stock market index prediction, bankruptcy prediction or corporate bond classification. In this study, different machine learning methods will be applied to the future stock price prediction of the companies in the Borsa Istanbul sector indices and the results will be compared.

Review of Literature

With the rapid development of technology in recent years, many techniques in the field of artificial intelligence have found their application areas. One of the most preferred among these areas is the finance sector, where future prediction is extremely difficult and important. Although there are many statistical methods used to make predictions in finance, one of the most preferred methods in recent years is artificial neural networks.

The use of artificial neural networks in finance dates back to about 30 years ago. Stock index prediction studies with artificial neural networks have been carried out for about 18 years. One of the first studies was conducted by Kimoto et al. (1990) on the Tokyo Stock Exchange. In the study, a forecasting system that recommends the timing of when to buy and sell stocks was put forward. They stated that they made accurate forecasts with this forecasting method and obtained a high degree of profit return as a result of the forecast.

Bengoechea, Ureta, Saavedra and Medina (1996) used two different artificial neural network models to predict the general stock index in Santiago de Chile Stock Exchange. According to the results of the study, they found that the combined artificial neural network model produced more accurate forecasting results than the simple architectural artificial neural network model.

Mizuno, Kosaka, Yajima and Komoda (1998) proposed a neural network model for the technical analysis of the stock market and a learning method to improve the forecasting ability. In this method, they tried to predict the buy and sell signals of the Tokyo Stock Exchange using artificial neural networks and obtained the prediction result with an accuracy of 63%.

In his study, Rast (1999) compared fuzzy neural networks and classical neural networks approach on the DAX Index, which includes the thirty largest stocks in Germany, using 1998 and 1999 data and concluded that the forecasting performance of the fuzzy neural networks model is better. In a similar study, Jandaghi et al. (2010) compared the forecasting performance of fuzzy neural networks and classical neural networks approach using Tehran Stock Exchange data and concluded that the forecasting performance of the fuzzy neural networks model is higher in the long run.

Yao, Li and Tan (2000) reported that the neural network model outperforms the traditional Black-Scholes option pricing model for volatile markets. They suggested that those who prefer lower risk and return may prefer the traditional Black-Scholes model, while those who prefer higher risk and return should prefer the neural network model.

Phua et al. (2001) used a prediction model that uses neural networks with genetic algorithms to predict stock price movements on the Singapore Stock Exchange. The prediction accuracy of the model was found to be 81% in the test data set.

Zorin and Borisov (2002) compared the forecasting performance of artificial neural networks and BOX-Jenkins models using Riga Stock Exchange data and reported that the artificial neural network model gave more accurate results.

Phua et al (2003) used the data of five major stock market indices (DAX, DJIA, FTSE-100, HSI and NASDAQ) to test the forecasting performance of different artificial neural network models. According to the calculation results obtained from five different financial markets, it was observed that the confidence region based neural network model gave better results compared to the results obtained with other neural network models. Using the trust region based neural network model, they predicted the index increases with a success rate of more than 60% in five stock markets.

Altay and Satman (2005) used daily, weekly and monthly data of the ISE for the period 1997-2005 to compare the forecasting performance of artificial neural networks and linear regression models. According to the results of the study, the forecasting performance of the artificial neural network model is higher. In a similar study (Karaatlı et al., 2005), monthly data of the ISE 100 index between 1990 and 2002 were used. In the study, it was concluded that the artificial neural network model showed higher forecasting performance than the regression method.

Fernandez and Gomez (2007) compared Hopfield type artificial neural network model, annealing simulation algorithm (SA), tabu search (TS) and genetic algorithm using stock market data from different countries (Hong Kong, Germany, UK, USA and Japan) and concluded that artificial neural networks outperformed other models in portfolio selection.

Vaisla and Bhatt (2010) used artificial neural networks and statistical techniques to model and predict the daily stock prices of stocks traded on the Indian National Stock Exchange and then compared the results of these two models. They evaluated the predictive ability of these two models using MAPE, MSE and RMSE. They concluded that artificial neural networks can predict stock market prices very well when trained with sufficient data, correct inputs and appropriate architecture.

Kara, Boyacıoğlu and Baykan (2011) compared artificial neural networks and support vector machine (SVM) models in terms of forecasting performance using ISE National 100 Index data. According to the results of the study, the prediction success rate of the artificial neural network model is 75.74%, while the success rate of the support vector machine is 71.52%.

Cao, Parry and Leggio (2011) used different models such as CAPM, Fama and French's three-factor model to predict stock returns in China. In addition, they compared the predictive ability of each of these models with an artificial neural network (ANN) model with the same variables. They found no statistical difference in the forecasting accuracy of the CAPM and the three-factor model, reflecting the emerging nature of Chinese equity markets. They suggested that the ANN model performed better and could be a useful tool for stock price forecasting in emerging markets.

Khansa and Liginlal (2011) studied the prediction of stock returns of 88 enterprises between 1996 and 2008. In the analysis in which both vector autoregression (VAR) and artificial neural network models were used, the prediction success of the artificial neural network model in the stock price prediction of the firms was 95%, while the prediction success of the VAR model was 85%.

In his study, Ticknor (2013) compared Bayesian artificial neural networks and integrated autoregressive moving averages (ARIMA) models for the prediction of stock returns and concluded that the predictive success of the artificial neural networks model was higher.

In their study, Çalışkan and Deniz (2015) analyzed the daily prices and price change direction of 30 stocks in Borsa Istanbul (BIST) using artificial neural networks technique. While the price direction prediction success in the study was 58%; The mean absolute percentage error (MAPE) was 1.8%.

Malakooti and AghaSharif (2015) stated that using artificial neural networks for the learning and curve fitting process, genetic algorithms for the optimisation process and using support vector machines in the forecasting phase will give accurate results. With the historical data of the last thirty days and support vector machines, NASDAQ, DJIA and S&P 500 indices were predicted by 74.4%, 77.6% and 76%, respectively.

Özçalıcı (2016) predicted the stock prices one day ahead, two days ahead and twenty days ahead with artificial neural networks technique. The price and volume information of the stocks in the BIST 30 Index between January 2010 and November 2015 were included in the model as input variables and it was reported that the prices of stocks twenty days in advance could be predicted by 72.88%.

Özer et al (2017) used weekly closing data of China (Shanghai), India (Nifty 50), Mexico (IPC-Mexico), Istanbul (BIST 100), USA (Nasdaq), UK (FTSE-100), Germany (DAX) and France (CAC-40) indices between 2012 and 2016. In the related study, fuzzy logic techniques and artificial neural network models were tried to be compared and it was found that successfully applying various artificial intelligence models gave promising results. In a different study (Manurung et al., 2018), using Bank Central Asia (BCA) data for 2013-2018 for stock price forecasting, they performed a forecasting study with long short-term memory (LSTM), a type of recurrent artificial neural networks for important parameters in the data (opening, high, low, closing). The results of the analysis show that the most accurate prediction in LSTM is obtained by using less than 1 year of short-term data instead of using 3 years or 5 years of training data, and that LSTM is superior to the traditional forecasting method of autoregressive moving average (ARIMA) with an accuracy of 56% compared to 94% for short-term data.

Pabuçcu (2019) tried to predict the negative and positive movements of the BIST 100 stock market index using artificial neural network, support vector machine and naive Bayes algorithm. As a result of the study, it was concluded that support vector machines are the best classifiers.

Li, Li, Liu, and Wang (2019) consider both latent states and capital asset pricing model based on integrated ARIMA. They used medium-long short-term memory (medium-LSTM) for medium-term stock forecasting and designed an intermediate LSTM-based deep neural network consisting of three components: LSTM, hidden Markov model and linear regression networks. In their experiment on S&P 500 stocks, they showed that the proposed intermediate-LSTM provides a 2-4% improvement in prediction accuracy.

Kim and Kang (2019) used various deep learning models to predict the trends of the Korean Stock Exchange (KOSPI) 200 index. They stated that expanding the input data to include more days in the forecasting study will increase the forecasting accuracy and that the forecasting power of models with long short-term memory (LSTM) is better than others. In Taş, Gülüm, and Tulum (2021), a future price forecasting study was carried out with the help of deep learning and shallow learning methods using the data of the S&P 500 index obtained from Yahoo Finance. In this context, daily price data between 12.08.2000 and 13.8.2020 were separated as training for the first 19 years (95% of the data set) and testing for the last 1 year (5% of the data set), and the prediction performances of long short-term memory (LSTM) and multilayer perceptron (MLP) methods were compared. The study concludes that the training and test errors obtained in both methods are close to each other and that these methods are suitable options for forecasting studies.

In his study, Şerbetçi (2022) investigated whether the stock exchanges of the Turkish stock exchange and the stock exchanges of BRICS, MIST and other countries within the scope of the Fragile Five, which are among the country leagues that emerged as a result of the international determinism in the economic situation and course of the countries in the global system, affect each other.

In their studies, Ilgın and Sarı 2022 aimed to predict the BIST 100 index movements with the benchmark indices of the BRICS countries. In the models they developed for the implementation of their studies, the monthly closing prices of the BIST 100 index were used as the dependent variable and the monthly closing prices of the benchmark indices of the BRICS countries were used as the independent variable. According to the findings, it has been observed that the models established to predict the BIST 100 index with the Artificial Neural Networks (ANN) method give successful results for India, South Africa and Russia, respectively, from the BRICS countries.

Özgür and Sarıkovanlık (2022) examines the forecasting performance of traditional (GARCH and ARMA-GARCH) and novel hybrid machine learning (ML) algorithms for predicting daily returns of BIST100 and NASDAQ indices. The ML algorithms evaluated include Random Forest, XGBoost, and Artificial Neural Networks (ANN). Hybrid models integrating (ARMA-)GARCH forecasts with ML predictions are developed and assessed. Results indicate that the developed hybrid models show promising performance compared to traditional (ARMA-)GARCH forecasts and other tested models.

Ayyıldız and İskenderoğlu (2023) conducted a study to predict the movement of the BIST 100 index using various machine learning methods. They compared decision trees, random forests, k-nearest neighbor, naive Bayes, logistic regression, support vector machines, and artificial neural networks. Data from January 1, 2012, to December 31, 2021, including technical indicators like moving averages, MACD, CCI, RSI, Stochastic %K, %D, William’s %R, and Momentum, were used. The results showed that artificial neural networks performed the best in predicting the index's movement direction, with logistic regression and support vector machines also performing well.

In Akyol Özcan's 2023 study, machine learning algorithms that predict the increases and decreases of stock market indices were compared. An in-depth analysis was made on the accuracy and success criteria of the methods used in the study. A particular point of emphasis is that linear methods produce more successful forecasting results. The findings that simple models, such as linear regression, can be more effective than complex machine learning algorithms under certain conditions.

Data, Methodology and Findings

In the literature, there are many studies on predicting the return of stocks in the stock market with machine learning. However, previous studies have generally focused on the entire stock market or only a single stock. Although there are also studies focusing on the sectors in the stock market, only one sector has been taken into account in these studies. In addition, studies have generally focused on a single machine learning method. In this context, the motivation of our study is to determine which method is more appropriate for more than one sector index by using different machine learning methods.

In this study, the 2019-2024 data of BIST Industrial Index (XUSIN), Services Index (XUHIZ) and Financial Index (XUMAL) firms are used. The number of firms is taken into account in the selection of the relevant sector indices. Since the number of firms in other sector indices in BIST is small, other sectors are not included in the study. Within the scope of the research, the opening price, highest price and lowest price values for each business day between 01.01.2019 and 31.05.2024 for 143 companies whose data can be accessed in the Industrial Index, 65 companies whose data can be accessed in the Service Index and 98 companies whose data can be accessed in the Financial Index were used. The data used for each firm included in the study were obtained from the investing website.

In the selection of the methods used within the scope of the study, studies on similar topics in the literature were taken into consideration (Choudhry & Garg, 2008; Shen, Jiang, & Zhang, 2012; Yoo, Kim, & Jan, 2005; Patel, Shah, Thakkar, & Kotecha, 2015). In this context, the machine learning methods used for testing in our study are Linear Model (LM), Artificial Neural Networks (ANN), stepwise linear regression (SLR), Gaussian Process regression (GSR), Support Vector Regression (SVR), Decision Tree technique.

Linear model (LM), also called linear regression, is a statistical approach typically used to model the relationship between two variables and also for time series forecasting (Harrell, 2015). In this study, single-output regression was selected by modelling different dependent variables from the same input dataset and analysed using Matlab Software (Version 9.13.0, Mathworks, USA).

Artificial neural networks (ANN) are models used in classification and regression calculations. Artificial neural networks were developed inspired by the human brain (Vapnik, 1995). In this study, a feed-forward ANN consisting of an input layer, a hidden layer and an output layer is used.

Stepwise linear regression (SLR) is a method of regressing multiple variables while simultaneously removing those that are not important and can be applied as an exploratory analysis (Harrell, 2015). Gaussian process regression (GPR) is a non-parametric, Bayesian regression approach that has made waves in the field of machine learning (Rasmussen and Williams, 2006). Support Vector Regression (SVR) is a machine learning algorithm for regression analysis. SVR can be used for both linear and non-linear classification problems using various kernel functions (Vapnik, 1995). There are various algorithms used in the Decision Tree technique. Commonly used algorithms are ID3 (Iterative Dichotoiser 3), C4.5, CART (Classification and Regression Trees) and CHAID (Chi-squared Automatic Interaction Detection). The CART algorithm of the decision tree technique, which can work with categorical and continuous values, has the feature of performing two functions for both classification and regression tree. (Akpınar,2017) Machine learning methods use multiple learning algorithms to achieve better predictive performance than can be obtained from any of the constituent learning algorithms alone. In contrast to a statistical ensemble, which is usually infinite in statistical analysis, a machine learning ensemble consists only of a concrete finite set of alternative models, but allows a much more flexible structure to exist between these alternatives (Freund & Schapire, 1997; Hastie, Tibshirani, & Friedman, 2008; Opitz & Maclin, 1999; Polikar, 2006; Rokach, 2010).

Statistical Error Metrics

Mean Square Error

Mean square error is a frequently used performance metric, especially in regression models. MSE is the mean square of the differences between actual and predicted values. Mean square error is a measure where higher errors have more weight. Therefore, it is used in situations where large errors will further degrade model performance. As the mean square error value approaches zero, the model's predictions become closer to the true values and the model's performance increases. However, in order to interpret the mean square error, it is necessary to evaluate the range of actual values and the characteristics of the data set.

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y_t})^2 \quad (1)$$

Root Mean Square Error

It is obtained by taking the square root of the mean square error and gives the average of the magnitudes of the differences between the actual values and the predicted values. RMSE, unlike Mean Square Error, ensures that large errors are penalized more. This feature makes it a more appropriate metric, especially in cases where large errors are significant. Additionally, the RMSE value indicates how close the predictions are to the actual values and can be used to compare the performance of different models.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - y_t^{-})^2} \quad (2)$$

Mean Absolute Error

Mean absolute error (MAE) is a performance metric frequently used in regression and time series problems and calculates the average of the absolute differences between actual and predicted values. Mean absolute error is a measure by which each data point gives equal weight to model performance. Therefore, it is used in cases where small mistakes are as important as big mistakes. As the mean absolute error value approaches zero, the model's predictions become closer to the true values and the model's performance increases. Mean absolute error is especially preferred in industrial and commercial applications because it is an easily interpretable metric.

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}|y_t - y_t^{-}| \quad (3)$$

Adjusted R Square (R2)

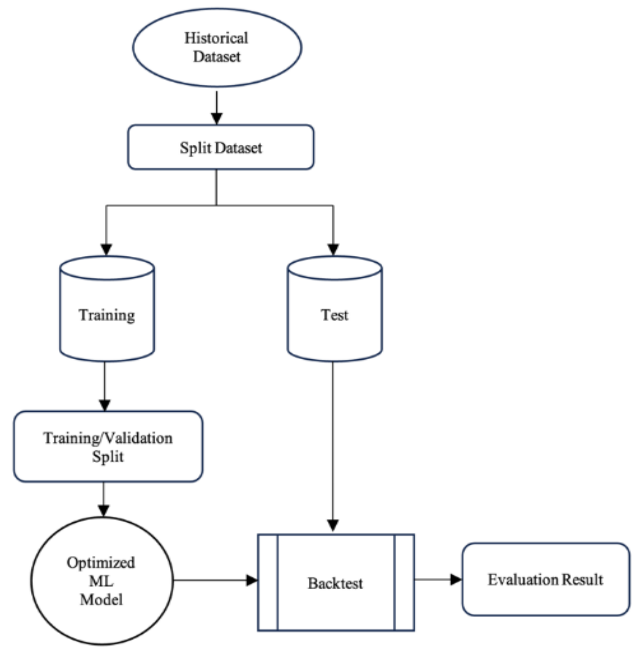
The adjusted R squared error metric determines the accuracy of the model by measuring the rate of variation of the independent variables in the dependent variable. The R2 error metric can increase the likelihood of the model failing on test data because it ignores the overfitting issue. To avoid this problem, the corrected R squared error metric reduces the problem of overfitting by penalizing additional independent variables added to the model. The R2 value indicates how well the experimental data are fitted to a linear curve, and it is preferred to be close to 1.

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum_i(\hat{y}_i - y_i^{-})^2}{\sum_i(y_i - y_i^{-})^2} \quad (4)$$

The machine learning process begins by splitting the dataset into “training data” and “test data.” Although there is no specific rule in the literature, generally 70% of the data is randomly selected and assigned as training data. In this article, the data is divided into 70% training data and 30% testing data. The purpose of using training data is to determine the optimum values of control parameters for various machine learning algorithms (learning process). When selecting optimum parameter values, emphasis is placed on minimizing the total error value, which can be defined as the difference between the actual observed value and the calculated result of the model (prediction output). Parameter values that give the smallest error are selected. The parameters of the model were optimized using the Bayesian Search method. The flowchart of the proposed algorithm for data analysis is given in Figure 1.

Bayesian optimization algorithm: It is the most accurate algorithm in terms of accuracy and time saving. Instead of searching for parameters randomly or sequentially, it detects hyperparameters using the probability function created using past values. In this way, a Instead of choosing the next set of parameters randomly, the algorithm optimizes its choice and detects the best set of parameters in the shortest time. Since the machine learning algorithms used in this study do not have too many hyperparameters, this algorithm was preferred to achieve the best results.

Figure 1: General scheme for data analysis and testing.



Summary of the index data is shown in Table 1 and Hyperparameter values of the method selected according to the performance criteria are shown in Table 2.

Table 1: Summary of the indices data

Symbol	Start	End	#observations
XUSIN	2019-01-02	2024-05-31	1356
XUHIZ	2019-01-02	2024-05-31	1356
XUMAL	2019-01-02	2024-05-31	1356

Table 2: Artificial Neural Networks Hyperparameters

ANN	#It.	=	1000	Maximum number of iterations
	Size	=	2	Number of fully connected layers
	First layer Size	=	10	-
	Second layer Size	=	10	-
	Activation	=	ReLU	-

The performance evaluation criteria of machine learning algorithms are given in Table 3. According to this table; the machine learning methods shown in Table 3 were used on XUSIN, XUHIZ and XUMAL data and it was shown that the best performing method among these methods was the Artificial Neural Networks for all three indices. This method is followed by Stepwise Linear Regression method in the second place. These results indicate that the generalization ability of the model is good. The findings from Table 3 clearly illustrate the superior performance of Neural Networks across all three sectors. In the first sector, Neural Networks achieved the lowest RMSE (1.4059) and MSE (1.9766), indicating high accuracy, while SVM lagged significantly with the highest RMSE (2.3602) and MSE (5.5703). In the second sector, both Neural Networks (0.0379) and Linear Regression (0.0390) demonstrated exceptional performance with low RMSE values, surpassing Decision Trees (0.0444).

In the financial sector, Neural Networks continued to lead with an RMSE of 1.7525, followed closely by Linear Regression at 1.8340. SVM, however, performed the worst in this sector as well, with a notable RMSE of 3.4798. Overall, Neural Networks consistently deliver strong results, while SVM struggled significantly, particularly in the validation phase, suggesting it may not be well-suited for the regression tasks evaluated.

Table 3: The performance evaluation criteria of machine learning algorithm

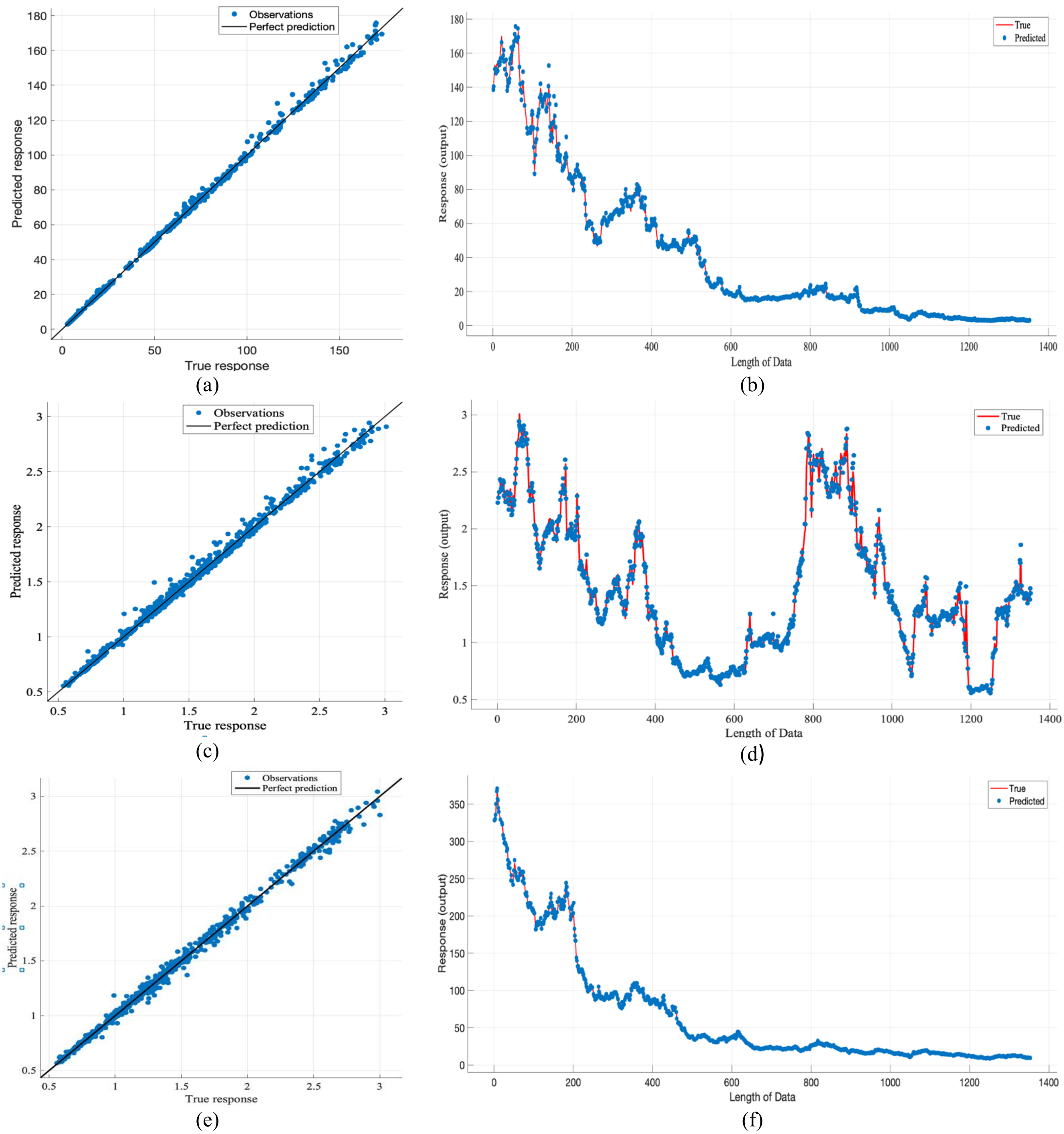
	Model Type	RMSE	MSE	RSquared	MAE
Industry	Neural Network	1,4059	1,9766	0,9990	0,7151
	Linear Regression	1,4086	1,9841	0,9990	0,7250
	Stepwise Linear Regression	1,4124	1,9949	0,9990	0,7275
	Gaussian Process Regression	1,4675	2,1536	0,9989	0,7260
	RUSBoosted Trees	1,5548	2,4175	0,9987	0,7744
	Decision Tree	1,7288	2,9889	0,9985	0,8798
	SVM	2,3602	5,5703	0,9971	2,1424
Services	Neural Network	0,0379	0,0014	0,9960	0,0248
	Linear Regression	0,0390	0,0015	0,9958	0,0256
	Stepwise Linear Regression	0,0391	0,0015	0,9958	0,0256
	Gaussian Process Regression	0,0391	0,0015	0,9958	0,0253
	RUSBoosted Trees	0,0401	0,0016	0,9955	0,0267
	Decision Tree	0,0444	0,0020	0,9945	0,0304
	SVM	0,0485	0,0023	0,9935	0,0396
Financial	Neural Network	1,7525	3,0713	0,9995	0,8884
	Linear Regression	1,8340	3,3636	0,9995	0,9458
	Stepwise Linear Regression	1,8428	3,3958	0,9995	0,9464
	Gaussian Process Regression	1,8435	3,3985	0,9995	0,9508
	RUSBoosted Trees	1,8460	3,4051	0,9989	0,9606
	Decision Tree	1,8465	3,4078	0,9965	0,9686
	SVM	2,2798	5,1494	0,9951	2,0600

The Artificial Neural Networks (ANN) model exhibits strong performance, evidenced by its low RMSE, MSE, and MAE values, alongside high R^2 scores. These metrics reflect the model's robustness in learning from the data and its capability to generalize effectively to new datasets.

In Figure 2, the effectiveness of the ANN method in predicting outcomes across industrial, service, and financial sectors is prominently displayed. Specifically, Figures 2(a) and 2(b) showcase a close alignment between the observed values (in red) and the predicted values (in blue) for industrial data, further underscoring the accuracy and reliability of the ANN approach. This close correlation not only highlights the model's predictive power but also reinforces its suitability for various applications in different sectors.

Similarly, Figures 2(c) and 2(d) depict the successful modeling of service data, where the agreement between observed and predicted values underscores the capability of artificial neural networks. Furthermore, Figures 2(e) and 2(f) demonstrate the accuracy of predictions in financial data, further supporting the robustness of the Artificial Neural Network method in this domain. Overall, the data presented in these graphs validate the effectiveness and reliability of artificial neural networks in predicting diverse datasets, emphasizing their applicability across industrial, service, and financial sectors.

Figure 2: Artificial neural networks method Industrial data (a) Observed and best forecast values (b) Actual values and forecast values Services data (c) Observed and best forecast values (d) Actual values and forecast values Financial data (e) Observed and best forecast values (f) Actual values and forecast values.



Conclusion

This study can be extended by using data with more year intervals and variables. When forecasting the stock market, data sets covering a wider time interval can be used in addition to historical data. In this study, Machine learning methods are used on BIST Industrial Index, Services Index and Financial Index data. It is shown that the best performing method among all the machine learning techniques tested is the Artificial Neural Networks for all three indices. The good performance of this method is well known in stock market forecasting and therefore it has been widely used in previous studies. The good performance of the Artificial Neural Networks may indicate that for a smaller period of time the linear model is suitable for forecasting purposes. Despite the good results obtained, different data should be analysed with different technical analysis methods other than the proposed methods and the results should be re-evaluated. In addition, various economic indicators and factors can also be taken into account to improve forecasting performance. These can be economic indicators such as inflation rates, exchange rates, interest rates, trade data, unemployment rates. However, it may also be useful to categorise the data into smaller time periods. For example, it may be possible to make more dynamic and near-term forecasts by using hourly, daily or monthly data. This can capture economic fluctuations more precisely and enable us to obtain more up-to-date forecasts.

Research results are consistent with the studies of Zorin and Borisov (2002), Altay and Satman (2005), Vaisla and Bhatt (2010), Kara, Boyacıoğlu and Baykan (2011), Khansa and Liginlal (2011), and Ticknor (2013), as mentioned in the literature review.

In future studies, we aim to optimise the data by adding more identifying features of the data and validate it with new real data from other stock market indices and foreign exchange market data. We also intend to include more technical analysis methods and existing index data in the metric study.

To guide future research efforts effectively, it is crucial to conduct comparative studies between advanced machine learning techniques and traditional linear models. Additionally, exploring how these methods perform under diverse market conditions will provide valuable insights. Furthermore, expanding research to incorporate new data sources and advanced technical indicators can significantly enhance the accuracy of forecasting models and augment decision-making capabilities within financial markets. The integration of AI and machine learning in financial forecasting offers substantial benefits, particularly in areas such as risk management, investment strategies, and market analysis. Nonetheless, continuous research and development efforts are essential to further refine these technologies, enhance their predictive accuracy, and broaden their scope of applications.

References

Altay, E., Satman, M.H. (2005), Stock Market Forecasting: Artifical Neural Network And Linear Regression Comparison In An Emerging Market, *Journal Of Financial Management And Analysis*, 18(2), 18-33.

Akpınar, H. (2017). Data, Veri Madenciliği Veri Analizi. 2.Baskı. Papatya YayıncılıkEğitim, İstanbul

Akyol Özcan K. (2023), Borsa endeksi yönünün makine öğrenmesi yöntemleri ile tahmini: bist 100 örneği. Gümüşhane Üniversitesi Sosyal Bilimler Dergisi, 14(3), 1001-1018.

Ayyıldız, N., İskenderoğlu, Ö. (2023), Prediction of Stock İndex Movement Using Machine Learning Methods: an Application on BIST 100 İndex, II. Eurasian Conference on Economics, Finance and Entrepreneurship, 20-21 May, İstanbul, 101-113.

Bengoechea, A.G., Ureta, C.O., Saavedra, M.M., Medina, N.O. (1996), *Stock Market Indexes In Santiago De Chile: Forecasting Using Neural Networks*, In Proceedings of International Conference on Neural Networks (ICNN'96) 4, 2172-2175.

Bilik, M., Aydın, Ü. (2018), *Finansal Hizmetlerde Dijital Dönüşüm ve Etkileri*. In book of Proceedings, 3rd. International Congress on Economics, Finance, and Energy, ISBN: 978-601-7805-32-6.

Cao, Q., Parry, M.E., Leggio, K.B. (2011), The Three-factor Model and Artificial Neural Networks: Predicting Stock Price Movement in China, *Annals of Operations Research*, 185(1), 25-44.

Choudhry, R., Garg, K. (2008), A hybrid machine learning system for stock market forecasting, *World Academy of Science, Engineering and Technology*, 39(3), 315-318.

Çalışkan, M.M.T., Deniz, D. (2015), Yapay Sinir Ağlarıyla Hisse Senedi Fiyatları ve Yönlerinin Tahmini, *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 31, 177-194.

Fernandez, A., Gomez, S. (2007), Portfolio Selection Using Neural Networks, *Computers & Operations Research*, 34(4), 1177-1191.

Freund, Y., Schapire, R.E. (1997), A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *Journal of Computer and System Sciences*, 55, 119-139.

Harrell, F. (2015), *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*, New York, NY: Springer.

Hastie, T., Tibshirani, R., Friedman, J. (2008), *The Elements of Statistical Learning*, New York: Springer.

Haykın, S. (2009), *Neural Networks and Learning Machines*. Boston: Pearson International Edition.

Jandaghı, G., Tehranı, R., Hossınpour, D., Gholıpour, R., Shadkam, S.A.S. (2010), Application Of Fuzzyneural Networks In Multi-Ahead Forecast Of Stock Price, *African Journal Of Business Management*, 4(6), 903-914.

Kara, Y., Boyacıoğlu, M.A., Baykan, Ö.K. (2011), Predicting Direction Of Stock Price Index Movement Using Artificial Neural Networks And Support Vector

Machines: The Sample Of The Istanbul Stock Exchange, *Expert Systems With Applications*, 38, 5311–5319.

Karaatlı, M., Güngör, İ., Demir, Y., Kalaycı Ş. (2005), Hisse Senedi Fiyat Hareketlerinin Yapay Sinir Ağları Yöntemi İle Tahmin Edilmesi, *Yönetim Ve Ekonomi Araştırmaları Dergisi*, 3, 22-48.

Khansa, L., Liginlal, D. (2011), Predicting Stock Market Returns From Malicious Attacks: A Comparative Analysis of Vector Autoregression And TimeDelayed Neural Networks, *Decision Support Systems*, 51(4). 745-759.

Kim, S., Kang, M. (2019), *Financial series prediction using attention LSTM*, arXiv preprint arXiv:1902.10877.

Kimoto, T., Asakawa, K., Yoda, M., Takeoka, M. (1990), Stock Market Prediction System With Modular Neural Network, *Proceedings Of The International Joint Conference On Neural Networks*, 1-6.

Li, X., Li, Y., Liu, X.Y., Wang, C.D. (2019*)*, *Risk management via anomaly circumvent: mnemonic deep learning for midterm stock prediction*, in Proc. 2nd KDD Workshop on Anomaly Detection in Finance, Anchorage, AK, USA.

Malakooti, M.V., AghaSharif, A. (2015), *Prediction of Stock Market Index based on Neural Networks, Genetic Algorithms and Data Mining Using SVD*, The Prooceedings of the International Conference on Digital Information Processing, Data Mining, and Wireless Communications, Dubai, UAE.

Manurung, A.H., Budiharto, W., Prabowo, H. (2018), Algorithm and modeling of stock prices forecasting based on long short-term memory (LSTM), *International Journal of Innovative Computing Information and Control*, 12, 1277-1283.

Mızuno, H., Kosaka, M., Yajıma, H., Komoda N. (1998), Application Of Neural Network To Technical Analysis Of Stock Market Prediction, *Studies In Informatic And Control*, 7(3), 111-120.

Opitz, D., Maclin, R. (1999), Popüler Topluluk Yöntemleri: Ampirik Bir Çalışma, *Yapay Zeka Araştırmaları Dergisi*, 11, 169-198.

Örnek, A. (2023). Yapay Zeka ve Makine Öğrenmesinin Hisse Senedi Fiyat Tahmini Üzerindeki Etkileri. Finans Araştırmaları Dergisi, 10(2), 45-58.

Özçalıcı, M. (2016), Yapay Sinir Ağları ile Çok Aşamalı Fiyat Tahmini: Bist 30 Senetleri Üzerine Bir Araştırma, *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 31(2), 209-227.

Özer, A., Sarı, S.S., Başakın, E.E. (2017), Bulanık Mantık ve Yapay Sinir Ağları ile Borsa Endeks Tahmini: Gelişmiş ve Gelişmekte Olan Ülkeler Örneği, *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(1), 99-123.

C. Özgür, V. Sarıkovanlık (2022), Forecasting Bist100 and Nasdaq Indices with Single and Hybrid Machine Learning Algorithms, Economic Computation and Economic Cybernetics Studies and Research, 56(3), 235-250.

Pabuçcu, H. (2019), Borsa Endeksi Hareketlerinin Tahmini: Trend Belirleyici Veri, *Selçuk Üniversitesi Sosyal Bilimler Yüksekokulu Dergisi*, 22(1), 246-256.

Patel, J., Shah, S., Thakkar, P., Kotecha, K. (2015), Predicting stock market index using fusion of machine learning techniques, *Expert Systems with Applications*, 42(4), 2162-2172.

Phua, P.K.H., Mıng, D., Lın, W. (2001), Neural Network With Genetic Algorithms For Stocks Prediction, *Asia Pacific Journal of Operational Research*, 18(1), 103-107.

Phua, P.K.H., Zhu, X., Koh, C.H. (2003), *Forecasting Stock Index Increments Using Neural Networks With Trust Region Methods*, Proceedings Of The International Joint Conference, 1, 260-265.

Polikar, R. (2006). Karar vermede topluluk tabanlı sistemler, *IEEE Devreler ve Sistemler Dergisi*, 6(3), 21–45.

Rasmussen, C.E., Williams, C.K.I. (2006), *Gaussian Processes for Machine Learning*, Massachusetts: MIT Press. Cambridge.

Rast, M. (1999), *Forecasting With Fuzzy Neural Networks: A Case Study in Stock Market Crash Situations. Fuzzy Information Processing Society, Nafıps*, 18th International Conference Of The North American, 418-420.

Rokach, L. (2010), Ensemble tabanlı sınıflandırıcılar, *Yapay Zeka İncelemesi*, 33(1), 1-39.

Saka Ilgın, K., & Sercan Sarı, S. (2022). BIST-100 Endeks Hareketlerinin BRICS Endeksleri Aracılığıyla Tahmin Edilmesi: Yapay Sinir Ağları Uygulaması. Abant Sosyal Bilimler Dergisi, 22(1), 350-366.

Shen, S., Jiang, H., Zhang, T. (2012), *Stock Market Forecasting Using Machine Learning Algorithms*, Department of Electrical Engineering, Stanford University, Stanford, CA, 1–5.

Şerbetçi, A. (2022). Brıcs, Mıst, Kırılgan Beşli Kapsamında Yer Alan Ülkelerin Menkul Kıymetler Borsaları Arasındaki Kısa Ve Uzun Dönemli Etkileşime İlişkin Amprik Bir Araştırma. Premium E-Journal Of Social Sciences (Pejoss), 6(19), 76–97.

Taş, A.İ., Gülüm, P., Tulum, G. (2021), Finansal Piyasalarda Hisse Fiyatlarının Derin Öğrenme ve Yapay Sinir Ağı Yöntemleri ile Tahmin Edilmesi; S&P 500 Endeksi Örneği, *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, Additional Issue, 446-460.

Ticknor J.L. (2013), A Bayesian Regularized Artificial Neural Network for Stock Market Forecasting, *Expert Systems with Applications*, 40(14), 5501-5506.

Usul, H., Küçüksille, E., Karaoğlan, D. S. (2017), Güven Endekslerindeki Değişimlerin Hisse Senedi Piyasalarına Etkileri: Borsa İstanbul Örneği, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22(3), 685-695.

Vaisla, K.S., Bhatt, A.K. (2010) An Analysis of The Performance of Artificial Neural Network Technique For Stock Market Forecasting, I*nternational Journal On Computer Science And Engineering*, 2(6), 2104-2109.

Vapnik, V. (1995), *The Nature of Statistical Learning Theory*. New York: Springer.

Vassakis, K., Petrakis, E., Kopanakis, I. (2018), Big Data Analytics: Applications, Prospects and Challenges, Mobil Big Data, *Lecture Notes on Data Engineering and Communications Technologies*, 10, 3-20.

Yao, J., Li, Y., Tan, C.L. (2000), Option price forecasting using neural networks, *Omega*, 28(4), 455-466, ISSN 0305-0483, https://doi.org/10.1016/S0305-0483(99)00066-3.

Yoo, P.D., Kim, M.H., Jan, T. (2005), *Machine learning techniques and use of event information for stock market prediction: A survey and evaluation*, in: International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, CIMCA-IAWTIC'06, 835–841.

Zorın, A., Borısov, A. (2002), *Modelling Riga Stock Exchange Index using neural networ*ks, Proceedings of the International Conference Traditions and Innovations in Sustainable Development of Society. Rezekne, Latvia.