Received: 5 July 2023 | Revised: 18 August 2023 | Accepted: 30 August 2023 | Published online: 31 August 2023

RESEARCH ARTICLE

Archives of Advanced Engineering Science
2024, Vol. 2(3) 123–133
DOI: 10.47852/bonviewAAES32021325

BON VIEW PUBLISHING

The Intraday High-Frequency Trading with Different Data Ranges: A Comparative Study with Artificial Neural Network and Vector Autoregressive Models

Ayben Koy[1] and Andaç Batur Çolak[2,*]

[1]*Faculty of Business, Finance and Banking Department, Istanbul Ticaret University, Türkiye*

[2]*Information Technologies Application and Research Center, Istanbul Ticaret University, Türkiye*

Abstract: With the high-frequency trading process, which is a subclass of algorithmic trading transactions, intraday information has increasing importance. Traditional statistical methods often fall short in capturing the intricate patterns and volatility inherent in such high-frequency data. In contrast, artificial neural network (ANN) models demonstrate remarkable capability in handling these challenges, and vector autoregressive (VAR) models provide insights into short-term relationships among variables. This study highlights the importance of using both ANN and VAR models for processing these short time intervals. BIST100 index, which is the main index of Borsa Istanbul, is predicted with two different models in different data ranges with ANN models and VAR models. Both generated ANN models successfully complete the training stages, with extremely high precision, and exhibit exceptionally low error values in their predictions. Although both models are effective, the evidence favors the model evaluated using 5-min data for both the training and prediction phases of ANN models. However, the relative importance of 15-min data in explaining the variation of BIST100 is higher. Moreover, the VAR model results indicate that the short-term relationship between variables can be influenced by the range of data and the 15-min interval data of the variables play a more significant role in explaining the BIST100 index over the longer term.

Keywords: high-frequency trading, technical indicators, artificial intelligence, BIST100, simulation

1. Introduction

Automatic buying and selling that occurs in microseconds in accordance with pre-programmed investment strategies is known as high-frequency trading (HFT), a subset of algorithmic trading operations. HFT is employed in the stock, foreign currency, and futures markets via structured exchanges and computerized trading platforms, all of which rely on high technology connection infrastructures. The spread of algorithmic trading and HFTs in these markets increased the importance of academic studies on these transactions. There are two main sources that HFTs make a profit from. Short-term purchase and sale transactions and high transaction volumes and earning returns from small price differences are the first of these resources. Information regarding securities traded on a single market can be released through arbitrage, or the price discrepancies of securities traded on many markets can be calculated. The liquidity provision function is the secondary revenue driver for HFTs. Continuous bilateral quotations are given to gain profit from the difference between buying and selling prices or gain cost advantage, thanks to the market maker function. Although HFT's income-seeking activities have positive contributions such as increasing transaction volumes and liquidity in the markets, improving price efficiency with their bids and prices incorporate information more efficiently [1–3], as a result of selling pressure with high order cancellations in markets where prices fall, they increase the severity of fluctuations in the markets. In times of market stability, heightened levels of HFT engagement correlate with a reduction in the volatility of stock prices. Nevertheless, when the market experiences sudden intraday crashes, the swift interplays among HFT algorithms trigger elevated instances of order cancellations. This, in turn, leads to a simultaneous withdrawal of high-frequency traders from the limit order book [4]. On the other side, if the divide algorithmic traders into two groups as the HFT's and the others, there is evidence of the contrary behavior [5]. This contrarian behavior supports the market stabilizing effects of algorithmic trading. But also the HFTs have also crowding out effect on the classical low-frequency traders [6]. It is very difficult for an individual investor to compete with HFTs. On the other

*__Corresponding author:__ Andaç Batur Çolak, Information Technologies Application and Research Center, Istanbul Ticaret University, Türkiye. Email: abcolak@ticaret.edu.tr.

hand, there is evidence on bidirectional connectedness among HFT and price efficiency. HFTs respond dynamically to engagements in price efficiency, while increasing intense of HFT is connected with more efficient prices [7]. Several investigations have also yielded findings suggesting that heightened HFT activity enhances market liquidity in certain models, while, in others, the converse outcome is observed. This uncertainty is not solely contingent upon the stock market or the timeframe of data collection. It is also intricately linked to the specific metric employed to quantify HFT: augmenting high-frequency orders results in diminished market liquidity, whereas an escalation in high-frequency trades bolsters liquidity [8].

However, the random walk hypothesis states that stock prices cannot be projected [9], with the advantages of present technology, artificial intelligence supported algorithmic transactions and HFTs continue to take place in the financial markets by taking advantage of stock price predictions. HFTs are taking an increasing share in algorithmic trading. It is important to decide at which frequencies the HFTs, which also make artificial intelligence-supported decisions, should act with the data.

There is a dearth of research investigating intraday price changes in emerging markets using artificial neural network (ANN) models. De Oliveira et al. [10] foresee the route of the stock prices in Brazil Bolsa Balcao with 26 variables consist of technical indicators and macroeconomic variables for different window sizes with ANN model. The best ANN model they found has got window size of three. Using 24 technical indicators in their ANN model, Chang et al. [11] practice evolving partially connected neural networks (EPCNNs) to foresee stock price in SP500 and Taiwan stock exchange (TSE) index in three different experiments. In their last part, the prediction power of EPCNN is found advantageous for TSE. In Naik and Mohan [12]'s study a five-layer deep neural network model performed better when compared with three-layer ANN model for the National Stock Exchange in India. The candlestick data and technical indicators are used in the models to identify the up and down movements of the stocks. In a new research for Chinese stock markets [13], a series of novel hybrid models are introduced that combine wavelet analysis with deep learning. The findings of the empirical analysis from the Chinese future market's CSI 300, SSE 50, and CSI 500 futures indices highlight the considerable performance of these proposed hybrid models in anticipating the intraday trends exhibited by stock index futures.

The studies on the intraday price dynamics on Borsa Istanbul are limited. Moreover, there are a few studies doing prediction using ANN models with intraday price dynamics. Using hourly prices of 100 stocks, Gunduz et al. [14] use convolutional neural network (CNN). In their study using deep learning methodology, technical indicators temporal features are used. In 54 out of 100 equities, the CNN-Corr model with the restructured topographies based on clustered feature correlation had the highest macro-averaged (MA) F-Measure metric scores. Another deep learning study predicts the BIST 30 Index over the course of 27 months utilizing technical indicators like the Relative Strength Index (RSI), Bollinger Bands, Stochastic Oscillator, and Moving Average Convergence Divergence (MACD) [15]. In most of the studies on predicting stock prices and stock indices for Borsa Istanbul, daily observation is used. Egeli et al. [16] studied the day of the week effect adding 5 dummy variables to their models using index value, TL/USD exchange rate, and overnight interest rate and compare with moving averages (MAs). Their ANN models performed better than MA models. In the following study, Kutlu and Badur [17] predicted the stock market index using nearly the same variables with Egeli et al. [16] for the period 2001–2006. Eight foreign stock markets' indices are added to index closing price, TL/USD exchange rate, overnight interest rate, and 5 dummy variables' dataset. In the study, it has been seen that the stock market index can be successfully modeled with feed-forward (FF) ANN. Telli and Coşkun [18] forecast a number of variables connected to BIST100, such as currency rates, stock indices of developed markets (DAX, Dow Jones, FTSE), and an economic calendar in which news days related to Türkiye are marked with a 1, while all other days are marked with a 0. One of the primary findings is that the economic datebook is a helpful expounding inconstant for forecasting BIST100 between July 29 and November 15, 2015. Research by Bildirici and Ersin [19] demonstrates the usefulness of combining ANN and generalized autoregressive conditional heteroskedasticity models on a daily dataset covering the period October 23, 1987–February 22, 2008. Based on their analysis of several hybrid models' ability to predict the stock market index, they conclude that conditional variance models supplemented with ANNs capture instability well. Technical indicators such as closing price, moving averages, and momentum price were used in the hybrid model study by Göçken et al. [20], and based on a heuristic optimization methodology (Harmony Search (HS) or GA) the HS-ANN was found to be significantly superior to others.

Boyacioglu and Avci [21] use monthly data and the adaptive network-based fuzzy inference system to forecast BIST100. The index is anticipated with a 98.3 percent rate of accuracy in the study based on the use of DJI, DAX, and BOVESPA indices, as well as monthly returns of BIST100. While the studies in the literature mainly try to make predictions with the information on historical price and historical price-related variables, Ustali et al. [22] estimated stock prices using firm-specific financial ratios. Their results indicate that XGBoost and Random Forest algorithms performed better than ANN models.

In today's global financial markets, especially large funds trade within microseconds with the latest data, it is striking that there is not enough study in this field in the literature. Especially in a stock exchange such as Borsa Istanbul, where foreign investors have a significant share, it is important to make analyses that combine technology with intraday data. In this context, the information that this study, which estimates the bist100 index with 5 and 15 mins' data, will provide to the literature and the finance sector is important. On the other hand, a study that elaborates on the relationships between variables through time series analysis will make a different contribution to the literature alongside ANN models.

Given the rise of HFT within the realm of algorithmic trading, the importance of intraday information has become increasingly pronounced. Traditional statistical methods struggle to capture the intricate patterns and volatility present in high-frequency data. In contrast, ANN models exhibit exceptional proficiency in handling these challenges, while vector autoregressive (VAR) models offer insights into short-term relationships among variables. This study underscores the significance of employing both ANN and VAR models to process short time intervals, particularly in the context of predicting the BIST100 index, the main index of Borsa Istanbul. In Borsa Istanbul where foreign investors show great interest, it is possible that many transactions are made with HFTs from all over the world. Foreign investors will continue to make up more than half of Borsa Istanbul's total investor base till at least 2021. This paper attempts to investigate, in this context, which data range should be successful in the intraday prediction of the main index of Borsa Istanbul. Two different ANN models are estimated with different data ranges as 5 and 15 mins. The utilization of ANN models in different data ranges yields highly precise predictions with minimal prediction errors, underlining their efficacy in capturing market dynamics. Moreover, in the realm of high-frequency financial data, VAR models provide a complementary approach by focusing on the short-term relationships among multiple variables. These models capture the interdependencies and responses between variables, shedding light on immediate reactions to market shocks and changes in key economic indicators. This paper is organized as follows: Section 2 reviews the data, and methodology described, Section 3 with empirical results, and the study is concluded in Section 4.

2. Methodology

In this section, comprehensive information regarding the data employed in the study and the methodologies utilized is presented. The study incorporates both linear econometrics and followed by the presentation of the ANN model.

2.1. Data collection

Table 1 displays all of the study's variables, all of which were taken from the Matriks IQ trading platform. Matriks Bilgi Dağıtım A.Ş. is a Turkish fintech firm that was founded in 2003 and had its stock market debut in 2021. Matriks IQ is the company's database. The study includes four basic indicators utilized in technical analysis, as well as the opening, closing, lowest, and highest prices of the BIST100 index, the primary index of Borsa Istanbul. Matriks IQ's downloadable technical indicators are pre-calculated and ready for analysis. The field of finance employs a wide variety of technical analysis instruments. Moving averages calculated from past prices are of general use in technical analysis. In addition to moving averages, there are many technical indicators calculated on the moving averages and past prices.

The frequently used technical indicators are determined by interviewing 21 professionals. After much discussion, it was decided that, in addition to looking at historical prices and moving averages, the RSI and the MACD index are employed most frequently. The RSI index, which is thought to be between 0 and 100, signals an overbought zone above 70 and an oversold zone below 30 is a momentum index. The other momentum index is MACD, which shows the relationship between two different moving averages of prices. If the MACD indicator is under the signal line, it creates a bearish signal and may indicate that the

Table 1
Variables used in the study

Variables		Source
BIST100 CLOSE	BIST100	MatrixIQ database
BIST100 OPEN	OPEN	MatrixIQ database
BIST100 HIGH	HIGH	MatrixIQ database
BIST100 LOW	LOW	MatrixIQ database
MOV15	MOV15	MatrixIQ database
MOV60	MOV60	MatrixIQ database
RSI	RSI	MatrixIQ database
MACD	MACD	MatrixIQ database
TRIGGER	TRIGGER	MatrixIQ database

Figure 1
The basic configuration topology of an MLP network model

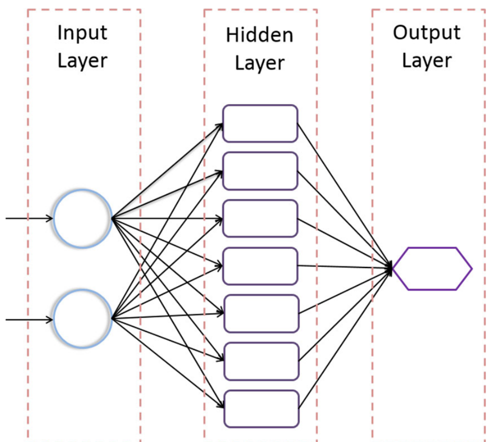
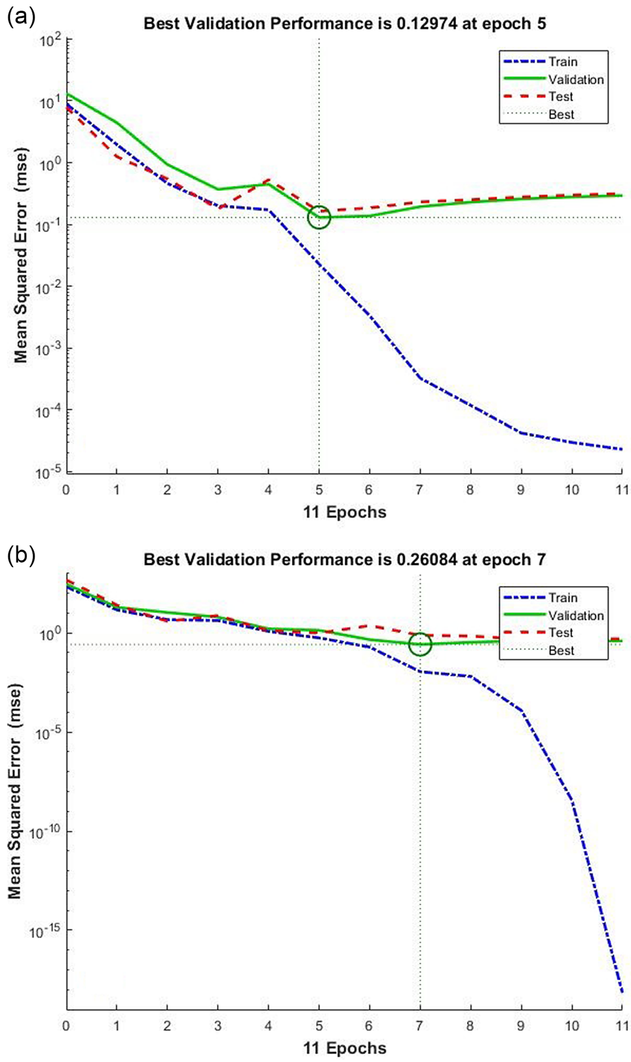


Figure 2
Validation performances of ANN models (a) model with 5-min and (b) model with 15-min



related asset should be sold. If the MACD indicator rises above the signal line, this signal may indicate that buying levels came.

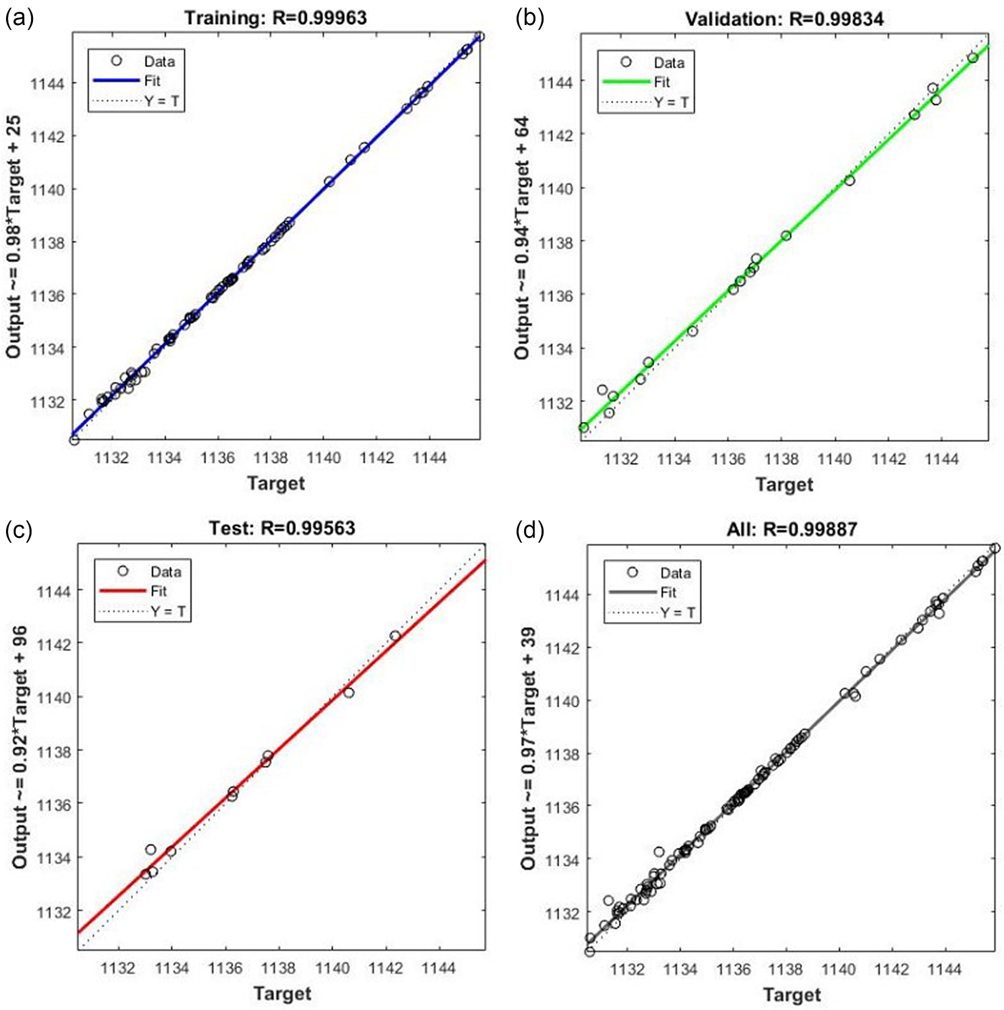
The final three days of September 2020 are chosen because they represent a time when the large price fluctuations resulting from the Covid-19 epidemic had somewhat subsided. Two different data ranges are used as 5 and 15 mins. There are 98 observations for the 5-min dataset for the last day of September, and there are 99 observations for the 15-min dataset of the last 3 days of September 2020.

2.2. Development of the ANN model

Two dissimilar ANN models have been developed for Borsa İstanbul to analyze the effect of data ranges on intraday HFTs. In both ANN models, the FF back-propagation (BP) multi-layer perceptron (MLP) network model, which is often preferred, is used [23]. In MLP network models, there is an input layer, minimum one hidden layer, and an output layer where the prediction results are obtained. The hidden layer encloses a computational element called a neuron, and each layer is connected to the next layer [24]. The basic configuration topology of an MLP network model is shown in Figure 1.

In ANN models developed using 5 and 15-min data, Open, High, Low, MOV60, MOV15, RSI, MACD, and TRIGGER values were defined as input parameters and BIST100 values were predicted in the output layer. There is no model or correlation used to determine the number of neurons in the hidden layer of MLP models [25]. For this reason, the performance of ANN models developed with different neuron numbers was analyzed, as was done in the literature, and the model with 5 neurons in the hidden layer was used. Optimization of the data to be used in training ANN models is an important parameter for the prediction performance of the model [26]. For this reason, the dataset was optimized and the grouping with the highest predictive performance was used. In the ANN model developed with 98 data, 70% of the data used for training, 20% for validation, and 10% for testing phase. As the training algorithm, the Levenberg–Marquardt algorithm, which has high performance and is frequently used in MLP networks, was preferred [27]. The Tan-Sig function is used as the transfer function in the hidden layer of the ANN model and the Purelin function is used in the output layer. The transfer functions used are given below [28].

Figure 3
Training performances of ANN model with 5 mins



$$f(x) = \frac{1}{1 + exp(-x)} \quad (1)$$

$$purelin(x) = x \quad (2)$$

Mean squared error (MSE), coefficient of determination (R), and margin of deviation (MoD) values were used to analyze the prediction performance of the developed ANN model. The equations used in the calculation of the parameters are as follows [29]:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(X_{exp(\text{i})} - X_{ANN(i)}\right)^2 \quad (3)$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^{N}\left(X_{\text{exp}(i)} - X_{ANN(i)}\right)^2}{\sum_{i=1}^{N}\left(X_{\text{exp}(i)}\right)^2}} \quad (4)$$

$$\text{MoD } (\%) = \left[\frac{X_{exp} - X_{ANN}}{X_{exp}}\right] \times\ 100 \quad (5)$$

2.3. VAR model

Within the scope of the study, where prognostications were realized through the application of ANN models, VAR models were harnessed to achieve a deeper comprehension of the intricate interconnections inherent within the dataset.

VAR models were introduced by Sims (1980) [30] for macroeconometric analysis. A VAR model is used to show the dynamic structure of the selected variables. In the VAR models, the variables do not have to be specified as endogenous or exogenous (Brooks), and the current values of variables are partly explained by past values of the variables involved. With the help of functions as impulse response and variance decomposition, relations of the variables are captured.

Traditional VAR models set up with stationarity series. In the present study, the variables are analyzed by augmented Dickey–Fuller

Figure 4
Training performances of ANN model with 15 mins

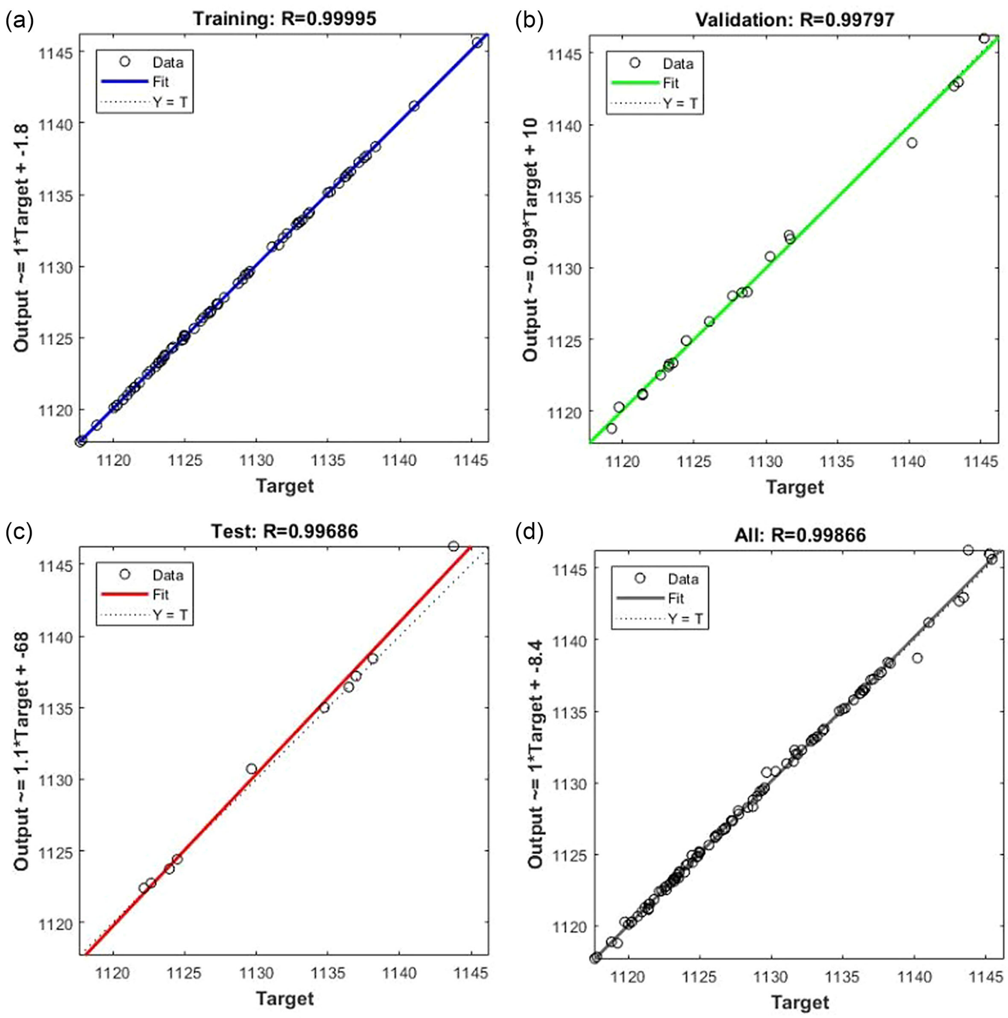


Figure 5
Error histogram of ANN models (a) model with 5 mins and (b) model with 15 mins

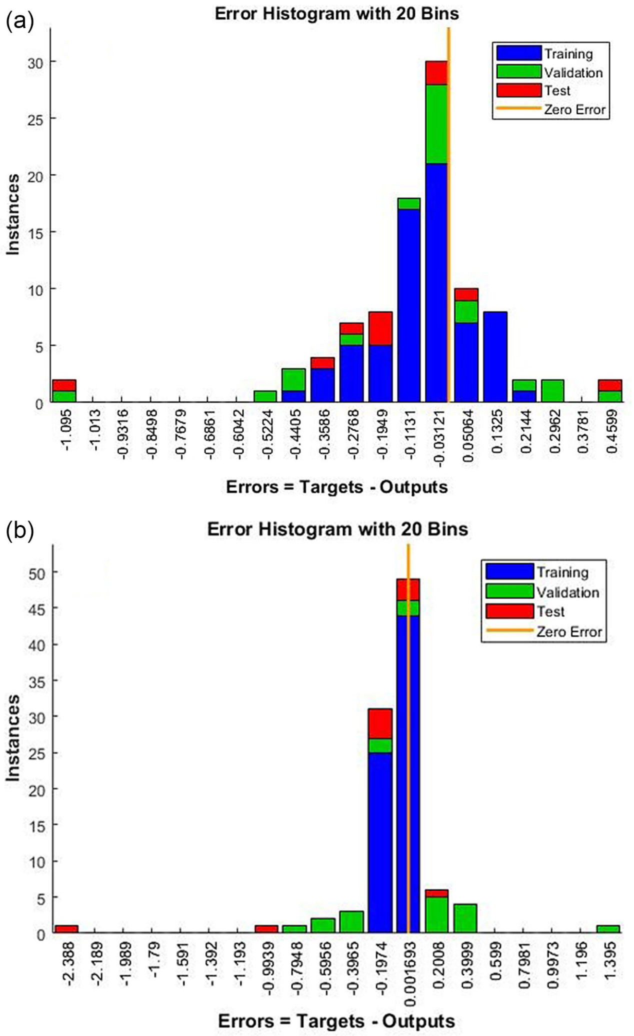
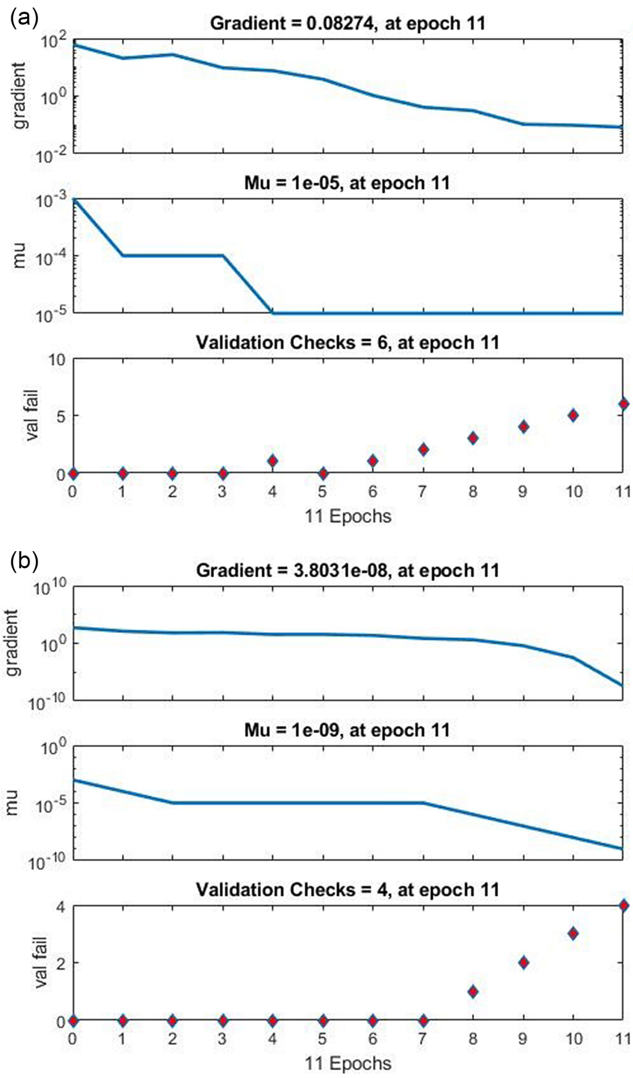


Figure 6
Training state of ANN models (a) model with 5 mins and (b) model with 15 mins



and Phillips–Perron unit root tests. According to the unit root tests, the logarithmic difference series used are stationary.[1]

3. Results and Discussion

This section provides an exhaustive exposition of the study's outcomes. Primarily, an in-depth analysis of the results derived from the ANN models is conducted. Subsequently, the outcomes of the VAR model, which delineates short-term relationships among the variables, are elucidated and deliberated upon.

In Figure 2, the validation performances of the training phase of the developed ANN models are shown. It is seen in the graphs that the MSE values, which are high at the beginning of the education phase, decrease with the advancing epochs. The closeness of MSE values to zero is an indicator of the low error rate of the training phase of the ANN model. The training process of the models has been completed at the point where the MSE values obtained from all three datasets reached the most ideal value.

Figures 3 and 4 show the data obtained from the training phase of ANN models. While the target data are on the x-axis of the graphs, the predicted values obtained from the ANN model are located on the y-axes. There are three different values in the graphs. Values expressed by dots represent data points, dotted lines represent equality lines, and solid lines represent fitted line. When the graphs are examined, the data points are seen located on the equality line. However, it is also seen that the equality lines are in perfect agreement with the fitted line. It should also be noted that the R values obtained using equation 4 are close to 1. These findings show that the training stages of the developed ANN models have been strongly completed with very high accuracy. However, the fact that the R value of the ANN model developed using 5-min values is higher than that of the model developed using 15-min values also indicates that the 5-min model has relatively higher prediction performance.

Error histograms of ANN models are given in Figure 5. In the error histograms, the errors belonging to the training processes of the models can be analyzed by showing the differences between the data obtained from the ANN models and the target data. In the graphs examined, it is seen that the error points are concentrated around the zero error line. However, it should be noted that the numerical

[1]The results of unit root tests can be shared if requested.

values of the error values are also very low. This information obtained from error histogram graphs shows that the training processes of the developed ANN models have been completed with low errors.

Figure 6 shows the training status of ANN models. It can be seen in the graphs that the gradient and mu values decrease with the advancing epochs. However, it is also understood that the training of the models has been terminated after the errors have been repeated 6 times, and therefore the validation check values are equal to 6. Figure 7 shows the values obtained from the ANN model and the target values for each data point. When paying attention to the graphs, it is clearly seen that the values obtained from the ANN model and the target values are in perfect agreement. This result shows that the developed ANN models can predict BIST100 values with high accuracy. While the BIST100 values realized on the x-axis of Figure 8, there are the prediction values obtained from the ANN model on the y-axis. When paying attention to the data points in the graphs, it is seen that they are located very close to the zero error line. This closeness of the data points to the zero error line shows that the error rate between the predicted values obtained from the ANN model and the actual BIST100 values is very low.

Figure 9 shows the MoD values calculated using Equation (5) for each data point. When the graphs are examined, it is seen that the data points are generally located close to the zero error line. The closeness of the data points to the zero error line shows that the error rates between the predicted values obtained from the ANN model and the actual BIST100 values are very low. The results obtained from the graphics clearly show that the developed ANN models can predict BIST100 values with ideal accuracy and very low error rates. However, when both graphs are analyzed, it is seen that the error rates of the ANN model developed with 5 mins of data are relatively lower than the ANN model developed with 15 mins of data. In order to better understand the prediction performance between the ANN model developed with 5 mins of data and the ANN model developed with 15 mins of data, the error rates at each data point are shown in Figure 10 comparatively. When the graph is taken into consideration, it is clearly seen that the error rates of the ANN model, which was developed with 5-min data, are lower than the 15-min model. In Figure 11, the differences between the target values and ANN model outputs for each data point are given and it is aimed to make the error analysis in more detail. When the data obtained are examined, it is seen that the difference values of the ANN model developed with 5-min data are lower than the ANN model developed with 15-min data. The results clearly confirm that the ANN model developed with 5 mins of data has lower error values compared to the ANN model developed with 15 mins of data.

Figure 7
Realized and prediction values according to the data number

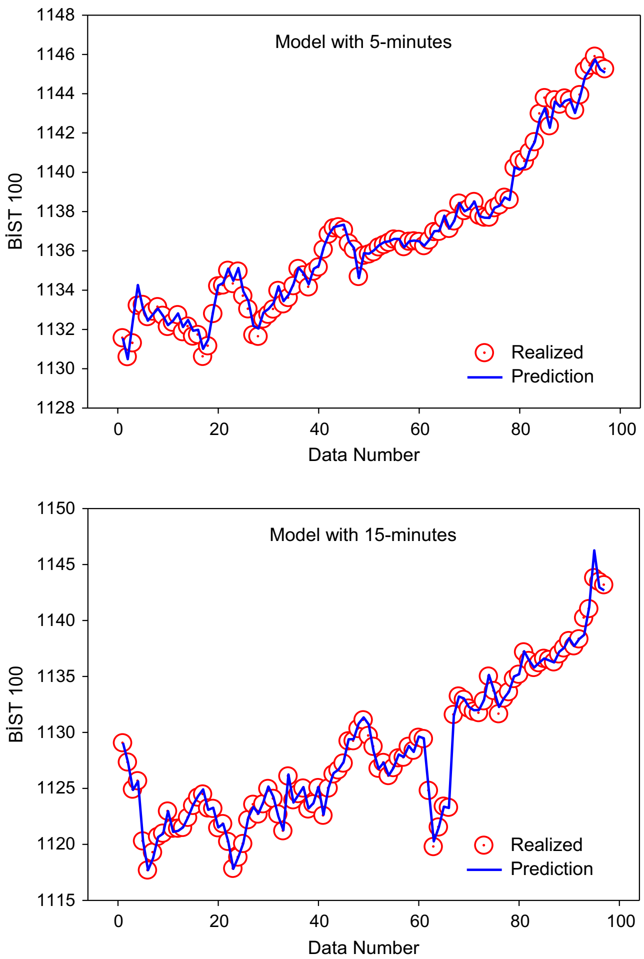


Figure 8
Realized values vs predicted results

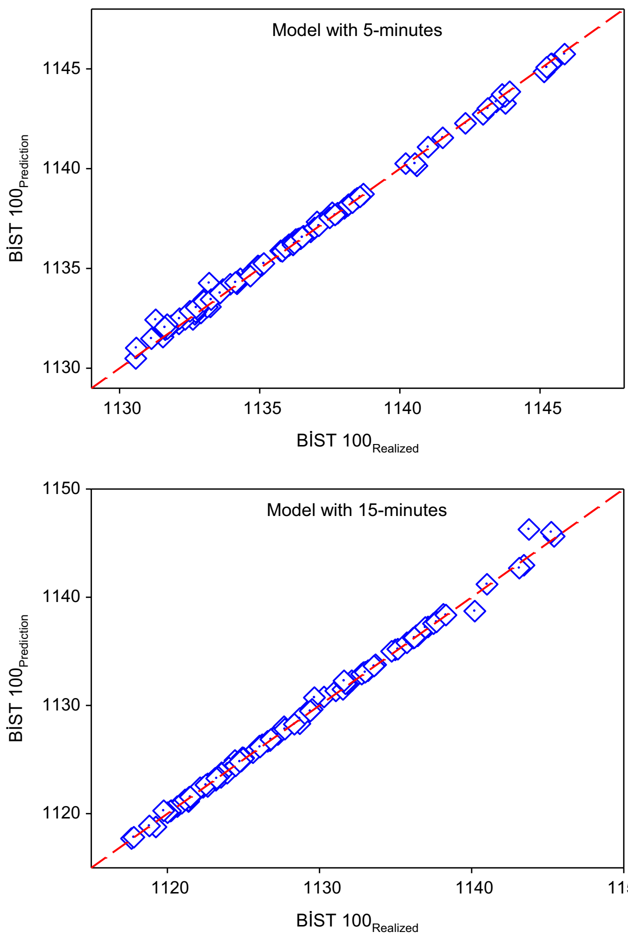


Figure 9
MoD values according to data number

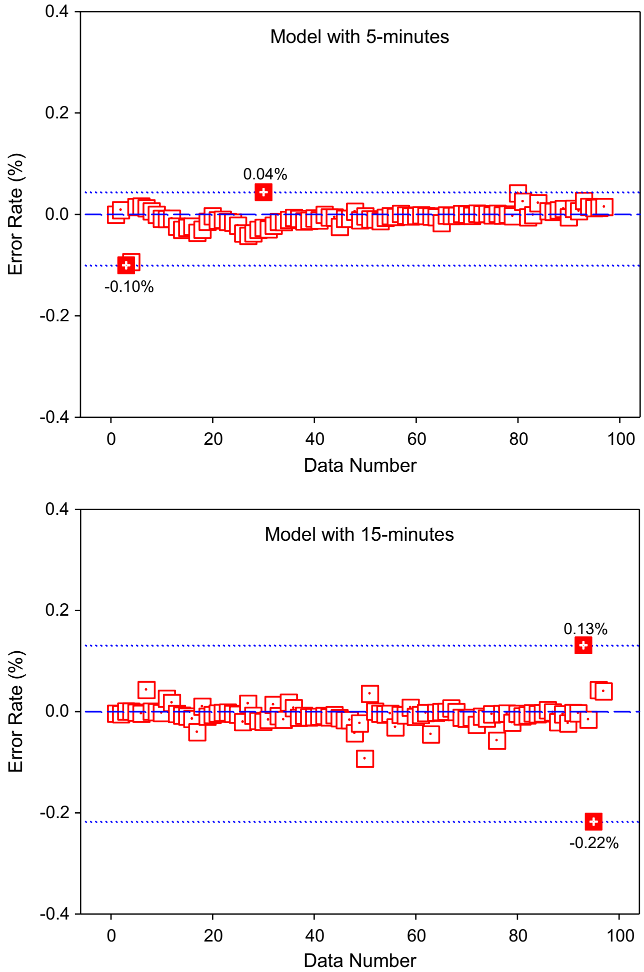
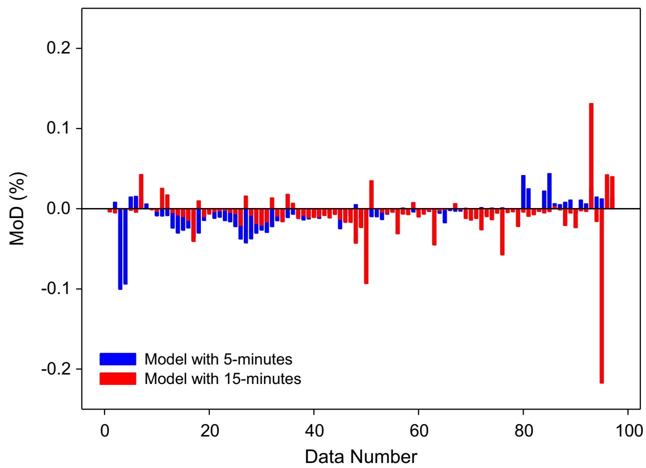
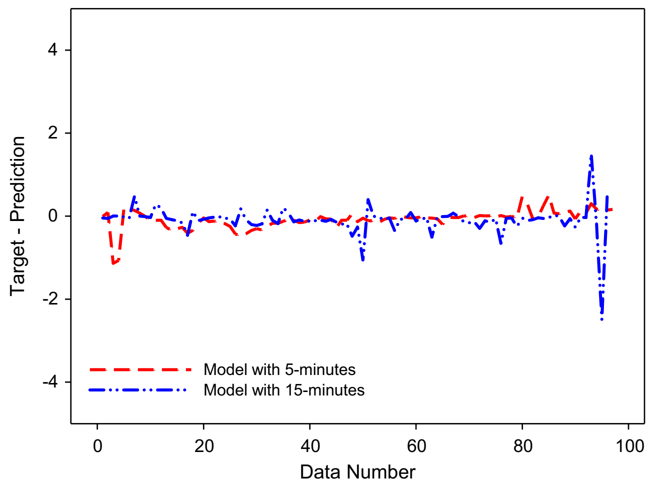


Figure 10
Comparison of the MoD values of ANN models



To analyze short-term relationships among data, this study utilized VAR analysis. Figure 12 illustrates the impulse response analyses of the stationary 5-min variables, while Figure 13

Figure 11
Difference between target and prediction values according to the data number of ANN models



displays those of the 15-min data. In the first row of Figure 12, the response of BIST100 to a one standard deviation shock applied to other variables is depicted. Interestingly, when MACD, MOV60, and RSI receive a one standard deviation shock, BIST100 responds in a decreasing direction for the first 4 periods (20 mins). However, none of the shocks on the variables have a permanent effect on BIST100, as their responses intersect the x-axis after a certain time. The first column of Figure 12 shows the response of other variables to a one standard deviation shock applied to BIST100. With the exception of the opening price (OPEN) of the BIST100 index, all variables respond positively to shocks from BIST100 in the first period (5 mins). Furthermore, the positive responses of MACD and MOV60 to shocks on BIST100 are relatively long compared to other variables.

Figures 14 and 15 include variance decomposition analyses of the VAR models established with, respectively, 5- and 15-min data. Variance decomposition is a statistical technique used to analyze the variance of a variable of interest and to determine the proportion of its variance that can be explained by other variables or factors. In other words, it breaks down the variance of a variable into different components based on the factors that contribute to its variability. This technique is commonly used in econometrics and finance to understand the relationships between variables and to identify the factors that are driving changes in a particular variable. When the results of BIST100 are examined, it can be seen that other variables are not strong enough in explaining the variance of BIST100 with 5-min data, but the importance of other variables increases in the results of 15-min data. In the model constructed with 15-min data, 40% of the variance of BIST100 can be explained by other variables after 45 mins (3 periods). Examining the impulse-response analyses of the VAR model shown in the first row of Figure 13, it becomes apparent that the BIST100 responds negatively to shocks applied to the highest price (HIGH) and MOV15. Additionally, the BIST100 index responds to MACD, OPEN, and RSI shocks in a decreasing direction. In the preceding section, we observed that utilizing different frequencies in econometric analyses of the BIST100 index, which we predicted with ANN models, may lead to outcomes that could alter the relationship direction in the short term. Moreover, upon investigating the response of other variables to a one standard

Figure 12
Impulse response – 5 mins

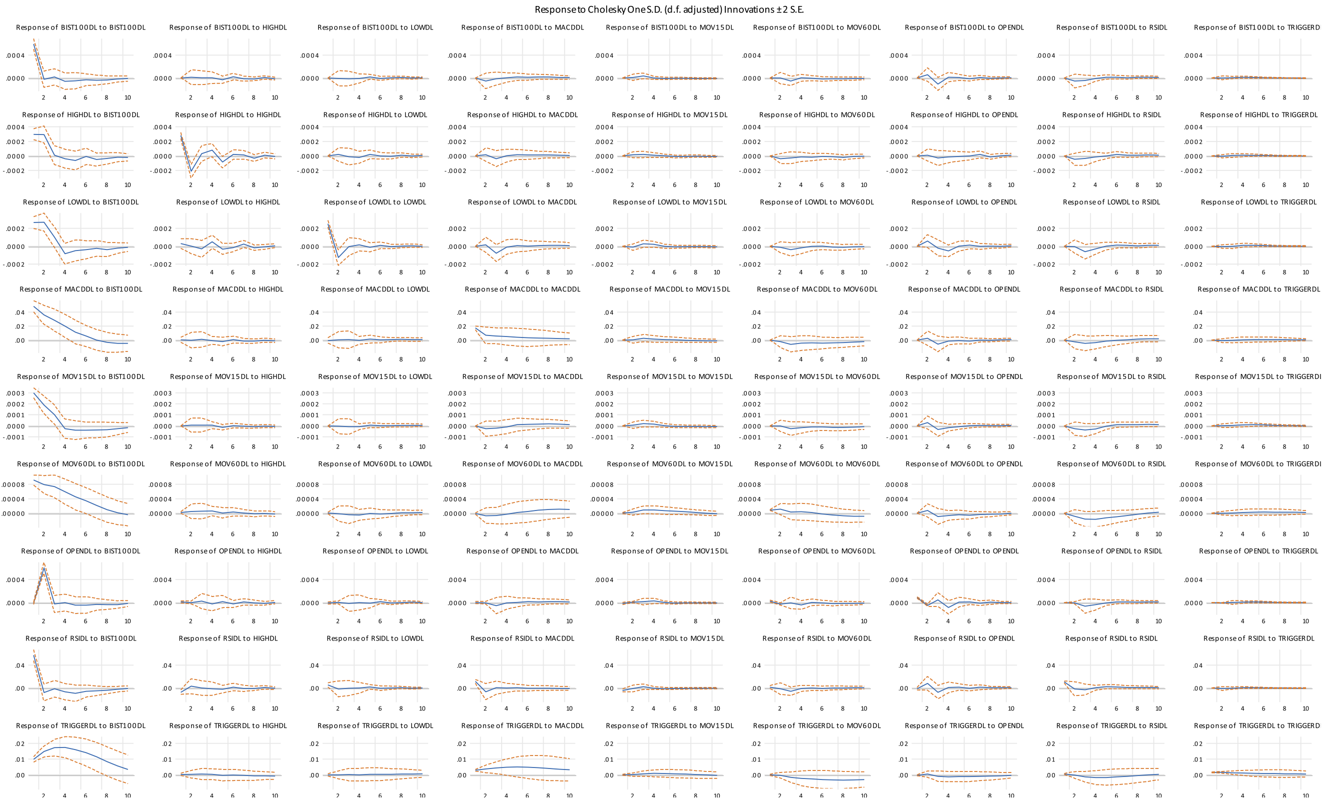


Figure 13
Impulse response – 15 mins

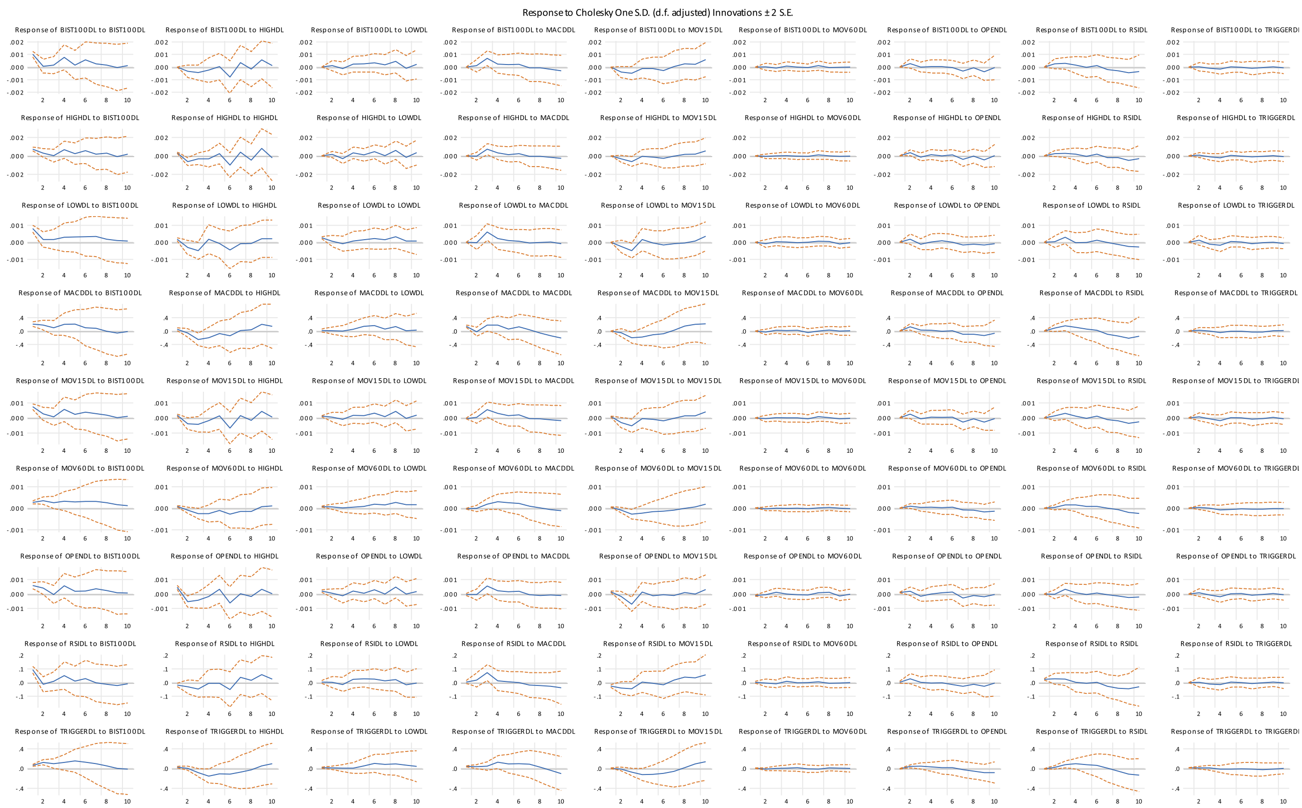


Figure 14
Variance decomposition BIST100 – 5 mins

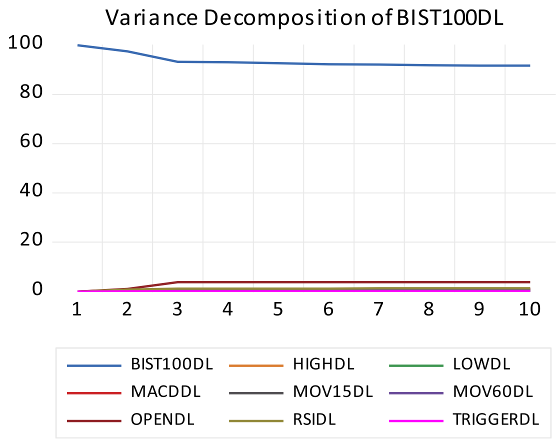
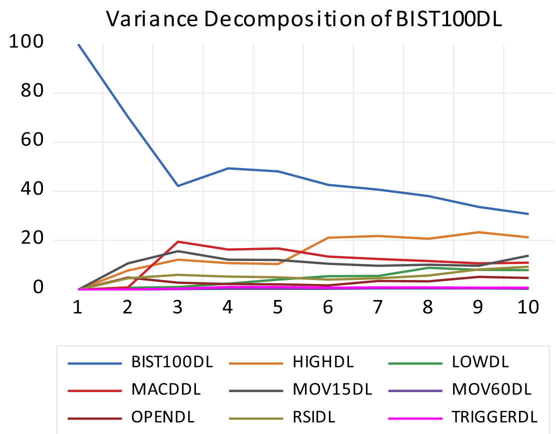


Figure 15
Variance decomposition BIST100 – 15 mins



deviation shock in the BIST100, it was observed that some variables had a weak yet entirely positive response in the first period (15 min).

4. Conclusion

In line with the speed brought by technology, HFTs, which trade in microseconds by using various algorithms in globalized financial markets, evaluate the information coming to the market and price movements with the shortest data intervals. On the other hand, methods such as artificial intelligence, deep learning, and machine learning are included in these fast trading transactions. In this context, this study, in which intraday and different frequency data are analyzed with artificial intelligence models, has been carried out to be a guide for the sector as well as the literature. In the study, the last 3 days of September 2020 were considered as the sampling period, when the high price movements caused by the Covid-19 pandemic decreased relatively. A dataset consisting of 99 observations of 15 mins and 98 observations of 5 mins only for the last day of the month was used. In ANN models, in which BIST100 index's opening price, lowest price, highest price, averages, RSI, and MACD indicators are used as input variables and BIST100 closing price is estimated, feed-forward (FF) back-propagation (BP) multi-layer perceptron (MLP) network model has been used. The findings show that the training stages of the developed ANN models are completed with very high accuracy, but the R value of the ANN model developed using 5-min values is higher than the model developed using 15-min values, which also shows that the 5-min model has a relatively higher prediction performance. When the MSE values and MoD values of the models were examined, it was concluded that both models had very low errors and the predictions of the models were successful. The study provides evidence that predictions made with 5 mins of data are more successful than predictions made with 15 mins of data. On the other hand, the results obtained with the VAR model provide evidence that the direction of short-term relationship between variables can be altered by the data range. The variance decomposition test, on the other hand, provides information that the 15-min data of the variables are more important in explaining the BIST100 index in the longer term.

By analyzing intraday data using artificial intelligence models, this study provides insights that can serve as valuable guidance for developing more effective HFT strategies. Decision-makers in the finance industry can use the findings to refine their trading algorithms and make more informed decisions. They can gain insights into the capabilities of these technologies and their potential impact on trading outcomes. This understanding can inform the adoption and implementation of such technologies in trading systems. Beyond its applicability in the industry, the study contributes to the academic literature by addressing a significant gap in the field of HFT. The comprehensive analysis of intraday data using ANN and VAR models enriches the existing knowledge base and sets a precedent for future research endeavors in understanding short-term market behavior. Moreover, the study's choice of the sampling period during the Covid-19 pandemic allows decision-makers to assess the resilience of the market during times of extreme volatility. The insights gained can aid in devising risk management strategies and preparing for similar situations in the future. The comparison between models developed using 5- and 15-min data offers valuable insights into optimal data frequency for prediction purposes. Decision-makers can use this information to optimize their data collection strategies and improve the accuracy of their predictions. Lastly, the findings regarding the relationship between data range and short-term relationships, as well as the significance of 15-min data in explaining the BIST100 index over the longer term, offer nuanced insights into market behavior. Decision-makers can use this understanding to better anticipate and respond to short-term fluctuations and long-term trends.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

References

[1] Anagnostidis, P., Fontaine, P., & Varsakelis, C. (2020). Are high–frequency traders informed? *Economic Modelling*, *93*, 365–383. https://doi.org/10.1016/j.econmod.2020.08.013

[2] Carrion, A. (2013). Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets*, *16*(4), 680–711. https://doi.org/10.1016/j.finmar.2013.06.005

[3] Hruška, J. (2016). Dynamics of liquidity on German stock market under the influence of HFT. *Procedia-Social and Behavioral Sciences*, *220*, 134–141. https://doi.org/10.1016/j.sbspro.2016.05.477

[4] Ammar, I. B., & Hellara, S. (2022). High-frequency trading, stock volatility, and intraday crashes. *The Quarterly Review of Economics and Finance*, *84,* 337–344. https://doi.org/10.1016/j.qref.2022.03.004

[5] Arumugam, D., & Prasanna, P. K. (2021). Commonality and contrarian trading among algorithmic traders. *Journal of Behavioral and Experimental Finance*, *30*, 100495. https://doi.org/10.1016/j.jbef.2021.100495

[6] Bongaerts, D., & Van Achter, M. (2021). Competition among liquidity providers with access to high-frequency trading technology. *Journal of Financial Economics*, *140*(1), 220–249. https://doi.org/10.1016/j.jfineco.2020.11.002

[7] Ammar, I. B., & Hellara, S. (2021). Intraday interactions between high-frequency trading and price efficiency. *Finance Research Letters*, *4*, 101862. https://doi.org/10.1016/j.frl.2020.101862

[8] Karkowska, R., & Palczewski, A. (2023). Does high-frequency trading actually improve market liquidity? A comparative study for selected models and measures. *Research in International Business and Finance*, *64*, 101872. https://doi.org/10.1016/j.ribaf.2022.101872

[9] Malkiel, B. G. (1973). *A random walk down wall street*. USA: W. W. Norton & Company.

[10] De Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, *40*(18), 7596–7606. https://doi.org/10.1016/j.eswa.2013.06.071

[11] Chang, P. C., Wang, D. D., & Zhou, C. L. (2012). A novel model by evolving partially connected neural network for stock price trend forecasting. *Expert Systems with Applications*, *39*(1), 611–620. https://doi.org/10.1016/j.eswa.2011.07.051

[12] Naik, N., & Mohan, B. R. (2020). Intraday stock prediction based on deep neural network. *National Academy Science Letters*, *43*(3), 241–246. https://doi.org/10.1007/s40009-019-00859-1

[13] Liang, D., Xu, Y., Hu, Y., & Du, Q. (2023). Intraday return forecasts and high-frequency trading of stock index futures: A hybrid wavelet-deep learning approach. *Emerging Markets Finance and Trade*, *59*(7), 2118–2128. https://doi.org/10.1080/1540496X.2023.2177507

[14] Gunduz, H., Yaslan, Y., & Cataltepe, Z. (2017). Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems*, *137*, 138–148. https://doi.org/10.1016/j.knosys.2017.09.023

[15] Halil, R. A. Ş. O., & Demirci, M. (2019). Predicting the Turkish stock market BIST 30 index using deep learning. *International Journal of Engineering Research and Development*, *11*(1), 253–265. https://doi.org/10.29137/umagd.425560

[16] Egeli, B., Ozturan, M., & Badur, B. (2003). Stock market prediction using artificial neural networks. *Decision Support Systems*, *22,* 171–185.

[17] Kutlu, B., & Badur, B. (2009). Stock market index prediction with artificial neural networks. *Management Journal*, *20*(63), 45–40.

[18] Telli, Ş., & Coşkun, M. (2016). Forecasting the BIST 100 Index using artificial neural networks with consideration of the economic calendar. *International Review of Economics and Management*, *4*(3), 26–46. https://dergipark.org.tr/en/download/article-file/266612

[19] Bildirici, M., & Ersin, Ö. Ö. (2009). Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange. *Expert Systems with Applications*, *36*(4), 7355–7362. https://doi.org/10.1016/j.eswa.2008.09.051

[20] Göçken, M., Özçalıcı, M., Boru, A., & Dosdoğru, A. T. (2016). Integrating metaheuristics and artificial neural networks for improved stock price prediction. *Expert Systems with Applications*, *44,* 320–331. https://doi.org/10.1016/j.eswa.2015.09.029

[21] Boyacioglu, M. A., & Avci, D. (2010). An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange. *Expert Systems with Applications*, *37*(12), 7908–7912. https://doi.org/10.1016/j.eswa.2010.04.045

[22] Ustali, N. K., Tosun, N., & Tosun, Ö. (2021). Makine öğrenmesi teknikleri ile hisse senedi fiyat tahmini [Stock price prediction with machine learning techniques]. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *16*(1), 1–16.

[23] Canakci, A., Ozsahin, S., & Varol, T. (2012). Modeling the influence of a process control agent on the properties of metal matrix composite powders using artificial neural networks. *Powder Technology*, *228,* 26–35. https://doi.org/10.17153/oguiibf.636017

[24] Ahmadloo, E., & Azizi, S. (2016). Prediction of thermal conductivity of various nanofluids using artificial neural network. *International Communications in Heat and Mass Transfer*, *74,* 69–75. https://doi.org/10.1016/j.icheatmasstransfer.2016.03.008

[25] Bonakdari, H., & Zaji, A. H. (2016). Open channel junction velocity prediction by using a hybrid self-neuron adjustable artificial neural network. *Flow Measurement and Instrumentation*, *49,* 46–51. https://doi.org/10.1016/j.flowmeasinst.2016.04.003

[26] Çolak, A. B. (2021). An experimental study on the comparative analysis of the effect of the number of data on the error rates of artificial neural networks. *International Journal of Energy Research*, *45*(1), 478–500. https://doi.org/10.1002/er.5680

[27] Ali, A., Abdulrahman, A., Garg, S., Maqsood, K., & Murshid, G. (2019). Application of artificial neural networks (ANN) for vapor-liquid-solid equilibrium prediction for CH4-CO2 binary mixture. *Greenhouse Gases: Science and Technology*, *9*(1), 67–78. https://doi.org/10.1002/ghg.1833

[28] Abdul Kareem, F. A., Shariff, A. M., Ullah, S., Garg, S., Dreisbach, F., Keong, L. K., & Mellon, N. (2017). Experimental and neural network modeling of partial uptake for a carbon dioxide/methane/water ternary mixture on 13X zeolite. *Energy Technology*, *5*(8), 1373–1391. https://doi.org/10.1002/ente.201600688

[29] Çolak, A. B., Yıldız, O., Bayrak, M., & Tezekici, B. S. (2020). Experimental study for predicting the specific heat of water based Cu-Al2O3 hybrid nanofluid using artificial neural network and proposing new correlation. *International Journal of Energy Research*, *44*(9), 7198–7215. https://doi.org/10.1002/er.5417

[30] Lütkepohl, H. (2013). Vector autoregressive models. In N. Hashimzade & M. A. Thornton (Eds.), *Handbook of research methods and applications in empirical macroeconomics* (pp. 139–164). Edward Elgar Publishing.

How to Cite: Koy, A., & Çolak, A. B. (2024). The Intraday High-Frequency Trading with Different Data Ranges: A Comparative Study with Artificial Neural Network and Vector Autoregressive Models. *Archives of Advanced Engineering Science*, *2*(3), 123–133. https://doi.org/10.47852/bonviewAAES32021325