A SUPERVISED MACHINE LEARNING IN FINANCIAL FORECASTING: IDENTIFYING EFFECTIVE MODELS FOR THE BIST100 INDEX

Cansu Ergenç[1][✉], Rafet Aktaş[2]

[1] *Ankara Yıldırım Beyazıt University, Türkiye*

ABSTRACT

The purpose of this study is to identify the most effective supervised machine learning models for predicting the financial performance of companies listed on the BIST100 index. In the rapidly evolving field of financial forecasting, machine learning techniques offer robust predictive capabilities. This research evaluates a range of supervised models, including Tree-Based Models (Decision Trees, Bagging, Random Forests, Adaboost, Gradient Boosting Machine (GBM), LightGBM, XGBoost, CatBoost), Neural Network-based Models (Artificial Neural Networks (ANN), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTM)), and Instance-based Learning Models (K-Nearest Neighbors (KNN) and Support Vector Machines (SVM)). The models' performance is assessed using comprehensive error metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and relative Root Mean Squared Error (rRMSE). The findings reveal that no single machine learning model consistently outperforms others across all companies in the BIST100 index. However, models like XGBoost and Random Forests demonstrate strong and consistent performance, making them particularly effective for financial performance forecasting. Furthermore, deep learning models such as CNNs, RNNs, and LSTMs show promising results, especially for certain firms. The research highlights key insights for investors and financial analysts seeking to leverage machine learning for data-driven decision-making in the Turkish stock market. This study offers a unique contribution to the field by applying and comparing advanced machine learning techniques in the context of the BIST100 index. It provides actionable insights for improving financial prediction accuracy and offers a foundation for further research in other stock market contexts.

KEY WORDS

tree-based models, neural network-based models, instance-based learning models

JEL CODES

C45, C53, G17, G32

1   INTRODUCTION

Accurate financial performance forecasting is integral to a company's strategic planning and financial management, playing a critical role in shaping both short- and long-term business decisions. By predicting future financial performance, companies can make well-informed decisions about resource allocation, risk management, and investment opportunities (Roberts and Dowling, 2002; Koller et al., 2010). As the financial landscape becomes increasingly complex, with companies operating in volatile and unpredictable markets, the demand for more accurate forecasting methods has grown significantly. Recent advancements in computational techniques, coupled with the exponential growth of financial data, have prompted a surge in the application of machine learning (ML) models to financial performance prediction (Jordan et al., 2015; Nosratabadi et al., 2020; Iqbal et al., 2020). These models offer robust capabilities for capturing complex, non-linear relationships within financial data, addressing many of the limitations of traditional forecasting approaches.

Despite the potential of machine learning, gaps remain in our understanding of which models perform best in specific financial contexts, particularly in emerging markets such as Turkey. The unique economic structure and market dynamics of Turkey, as represented by the companies listed on the Istanbul Stock Exchange (BIST100), offer an ideal environment for testing and comparing machine learning models. The BIST100 index, composed of the 100 largest publicly traded companies in Turkey, reflects the diverse economic landscape of the country and plays a significant role in its financial stability and growth. Accurate predictions of financial performance within this index are not only essential for companies themselves but also for investors, analysts, and policymakers who rely on such forecasts for decision-making.

Traditional financial forecasting methods, while useful in many cases, often fall short in capturing the multi-dimensional and non-linear interactions inherent in financial data (Refenes et al., 1997; Clements et al., 2004; Gradojevic and Yang, 2006; Adegbite et al., 2019). Financial data is affected by a wide range of factors, including macroeconomic indicators, company-specific financial ratios, and market sentiment. These factors interact in complex ways that can be difficult to model using conventional statistical approaches. In contrast, machine learning models, which are designed to handle large and complex datasets, offer significant advantages for financial forecasting. Specifically, they can learn from historical data and identify intricate patterns and relationships that may not be apparent through traditional methods.

This study aims to address the research gap by conducting a comprehensive evaluation of several supervised machine learning models for predicting the financial performance of companies in the BIST100 index. The models selected for this analysis include Tree-Based Models such as Decision Trees, Random Forests, Bagging, and Boosting methods (XGBoost, LightGBM, CatBoost), Neural Network-based models like Artificial Neural Networks (ANN), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTM), and Instance-based Learning Models such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). These models were chosen for their proven capabilities in handling high-dimensional, non-linear financial data (Jones, 2017; Chow, 2018; Jan, 2021; Reel et al., 2021).

The dataset for this study is drawn from the BIST100 index, covering a ten-year period from 2014 to 2024. It includes daily stock prices, trading volumes, and various technical indicators such as Moving Averages (MA_50, MA_200), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Bollinger Bands (BB_Mid, BB_Upper, BB_Lower). Financial ratios, derived from companies' balance sheets and income statements – such as the debt-to-equity ratio, current ratio, return on capital employed, and net profit margin – serve as key inputs for the machine learning models (Höbarth, 2006; Chen, 2011; Xiuguo and Shengyong, 2022). By assessing

model performance through error metrics such as Mean Square Error (MSE), Root Mean Square Error (RMSE), Normalized Root Mean Square Error (rRMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE), this study provides a comprehensive evaluation of which algorithms deliver the most accurate and reliable predictions in this financial context.

In addition to determining which models provide the best overall accuracy, this research seeks to explore the conditions under which certain models may outperform others. For instance, tree-based models have often demonstrated strong performance in financial forecasting due to their ability to handle non-linear relationships, while neural networks, especially deep learning methods like CNNs and LSTMs, may excel in capturing more subtle patterns and temporal dependencies in financial time series data (Jordan et al., 2015; Jan, 2021). Understanding these nuances is essential for improving the practical application of machine learning in financial forecasting.

The practical implications of this study are far-reaching. Investors, financial analysts, and corporate decision-makers can leverage the insights from this research to make more informed investment strategies, optimize financial performance, and mitigate risks. By comparing the strengths and weaknesses of various machine learning models, this research provides a valuable resource for selecting the most appropriate tools for financial forecasting in diverse economic contexts.

This paper is structured as follows: Section 2 presents a comprehensive literature review of machine learning applications in financial forecasting. Section 3 describes the dataset and methodology used in the study. Section 4 discusses the results of the model evaluations, while Section 5 elaborates on the practical and theoretical implications of the findings. Finally, Section 6 concludes with key insights and recommendations for future research in this rapidly evolving field.

2   LITERATURE REVIEW

The financial forecasting landscape has witnessed a significant transformation with the advent of machine learning (ML) techniques, particularly in the context of stock market predictions. The BIST100 index, comprising the top 100 companies listed on Borsa Istanbul, is a critical benchmark for Turkish equities and has increasingly attracted the interest of researchers aiming to leverage advanced ML models for financial prediction. A substantial body of research has explored the use of supervised machine learning in financial forecasting, especially for developed markets. Tree-based models such as Random Forests, Gradient Boosting Machines (GBM), and their derivatives have become popular due to their ability to model complex, non-linear relationships and effectively reduce overfitting through ensemble strategies (Hoque and Aljamaan, 2021; Ryll and Seidens, 2019; Chen and Guestrin, 2016). Neural network-based methods – including Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs) – have also gained traction for their capability to capture sequential dependencies and hidden patterns in financial time series data (Shen et al., 2021; Bai et al., 2023; Kim and Kim, 2019). Evaluation of these models often involves robust error metrics such as MSE, RMSE, MAE, and MAPE, as well as sensitivity analyses to assess their reliability and robustness under varying market conditions (Kambale et al., 2024; Yaprakdal and Bal, 2022).

Despite these advancements, a critical review of studies published between 2022 and 2024 reveals several important limitations. First, the majority of recent comparative studies remain largely focused on developed financial markets – such as the United States, Western Europe, or East Asia – while emerging markets like Turkey remain underrepresented (Alizadegan et al., 2024; Bhuiyan et al., 2023; Liu et al., 2023; Alshehri, 2023). Second, existing studies often restrict themselves to a small set of algorithms – most frequently Random Forest, XGBoost, or LSTM – without evaluating a broader methodological landscape that includes both instance-based (e.g.,

KNN, SVM) and hybrid models. Furthermore, many analyses rely on short time frames, limited variable sets, or lack statistical significance testing, reducing the robustness and generalizability of their findings (Margarat et al., 2023; Benchikh et al., 2024). Studies on the Turkish market are particularly scarce and typically focus on a single model or a limited set of predictors, which constrains the understanding of model transferability in volatile, emerging market conditions (Beyaz and Efe, 2019).

This study directly addresses these gaps by providing a comprehensive head-to-head comparison of a diverse set of supervised machine learning models – including ensemble, deep learning, and instance-based approaches – using a large-scale, up-to-date BIST100 dataset spanning a decade. By benchmarking model performance using multiple error metrics and conducting statistical tests, this research not only advances the methodological rigor of comparative ML forecasting but also provides actionable insights for both researchers and practitioners seeking effective and adaptable tools for financial forecasting in emerging markets. Thus, the present study extends the frontier of machine learning research by systematically evaluating which models offer the greatest accuracy and robustness for BIST100 companies, thereby supporting more data-driven decision-making in the Turkish stock market and beyond.

3 THEORETICAL BACKGROUND

Ensemble models, particularly Random Forest and XGBoost, have demonstrated consistently strong performance in financial forecasting due to their robust theoretical underpinnings. One primary reason for their effectiveness is their ability to reduce model variance and mitigate overfitting – a common issue in volatile and high-dimensional financial datasets (Breiman, 2001; Chen and Guestrin, 2016). Ensemble learning combines the predictions of multiple weak learners (typically decision trees) to form a more stable and accurate model, capturing complex, non-linear relationships within the data that single models may miss (Dietterich, 2000). In financial time series, where market dynamics can shift rapidly and involve intricate interdependencies, ensemble methods provide both flexibility and resilience. Random Forest achieves this by aggregating numerous uncorrelated decision trees, each built on randomly sampled subsets of features, which enhances generalizability and reduces sensitivity to noise (Liaw and Wiener, 2002; Davis and Nielsen, 2020). Similarly, XGBoost employs gradient boosting to sequentially optimize model performance, focusing learning on difficult-to-predict cases and delivering high accuracy even on challenging, noisy datasets (Chen and Guestrin, 2016; Alshehri, 2023).

Another crucial advantage of tree-based ensemble models is their robustness to multicollinearity and missing values, as well as their ability to provide feature importance measures – essential tools for financial analysts working with large, multifaceted datasets (Lundberg et al., 2019; Wang et al., 2022). These properties make ensemble methods especially suitable for financial applications where predictors may be numerous, highly correlated, or partially missing. In contrast, deep learning models such as Long Short-Term Memory networks (LSTM) and Convolutional Neural Networks (CNN) excel in capturing temporal dependencies and extracting meaningful features from sequential data (Hochreiter and Schmidhuber, 1997; Kim and Kim, 2019; Liu et al., 2022). LSTM networks are specifically designed to overcome the limitations of traditional recurrent neural networks by addressing the vanishing gradient problem, allowing them to model long-term dependencies in time series – an essential requirement for financial forecasting where past trends and shocks can impact future prices (Bai et al., 2023; Fischer and Krauss, 2018). This theoretical advantage is supported by recent studies showing LSTM's superior ability to model non-stationary, sequential financial data, especially in volatile markets (Wang, 2024; Diqi et al., 2024).

CNNs, although originally developed for image processing, have been successfully adapted for financial time series analysis due to their ability to extract and hierarchically combine local patterns within the data (Cao and Wang, 2019; Qi et al., 2022). When used alone or in hybrid architectures (e.g., CNN-LSTM), these networks can learn both short-term local features and long-term dependencies, providing a comprehensive approach to modeling market behaviors (Widiputra et al., 2021; Kim and Kim, 2019). Moreover, deep learning models are data-driven and require minimal feature engineering, enabling them to automatically uncover complex, non-linear, and hidden relationships in large, high-frequency financial datasets (Jan, 2021; Liu et al., 2022). However, while ensemble and deep learning models offer significant advantages, their success also depends on careful hyperparameter tuning, appropriate data preprocessing, and understanding their limitations regarding interpretability and computational cost (Henrique et al., 2023; Beyaz and Efe, 2019).

4   PREDICTION MODELS

In this study, Tree-Based Models (Decision Trees, Bagging, Random Forests, Adaboost, Gradient Boosting Machine (GBM), LightGBM, XGBoost, CatBoost), Neural Network-based Models (Artificial Neural Networks (ANN), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTM)) and Instance-based Learning Models (K-Nearest Neighbors (KNN) and Support Vector Machines (SVM)) were used. Fig. 1 shows the models used in this study. The selection of machine learning models in this study is guided by both empirical evidence from recent literature and practical considerations relevant to financial time series forecasting. Ensemble methods – notably Random Forest, LightGBM, and XGBoost – were included due to their demonstrated ability to handle large, high-dimensional datasets with complex, non-linear relationships, and their robust performance in diverse financial prediction tasks (Henrique et al., 2023; Chen and Guestrin, 2016; Davis and Nielsen, 2020). These models effectively reduce overfitting and variance through aggregation and boosting techniques, making them highly suitable for volatile and multi-factor financial environments such as BIST100. Neural network-based models, including LSTM, CNN, and ANN, were selected for their theoretical and empirical strengths in modeling temporal dependencies and uncovering hidden structures in sequential data (Hochreiter and Schmidhuber, 1997; Kim and Kim, 2019; Liu et al., 2022). LSTM networks, in particular, excel at learning long-term dependencies, a critical feature for financial time series characterized by delayed effects and memory. CNNs have also shown promise in extracting localized features and patterns, especially when integrated into hybrid architectures (Cao and Wang, 2019; Widiputra et al., 2021). Their flexibility and scalability make them strong candidates for financial forecasting with large and complex datasets. Instance-based methods such as SVM and KNN are widely used in the literature for classification and regression in financial applications, providing valuable benchmarks and offering interpretability advantages (Cao and Tay, 2001; Islam et al., 2021; Bellotti et al., 2011). SVMs are particularly robust to outliers and noise, and are effective for high-dimensional spaces – a common characteristic in financial data. Alternative statistical methods such as ARIMA, GARCH, or Facebook Prophet were intentionally not included in this study. While these approaches have been widely used in earlier financial forecasting research, they typically rely on strong linearity or stationarity assumptions and have limited capacity to model the multi-dimensional and non-linear dependencies prevalent in large-scale financial datasets (Gradojevic and Yang, 2006; Jan, 2021). Recent comparative studies also show that advanced machine learning models routinely outperform traditional time-series techniques, especially in contexts with complex interactions and numerous predictive features (Bhuiyan et al., 2023; Alizadegan et al., 2024). Finally, model selection also considered computational efficiency and interpretability. While deep learning models

can require significant computational resources, the models chosen represent a balance between predictive power, scalability, and practical implementation in real-world financial analysis.

4.1 Tree-Based Models

4.1.1 Decision Trees

Decision trees are widely used in financial forecasting for their robust decision-making capabilities. In stock price forecasting, they analyze historical market data to predict trends (Li et al., 2023). They are also valuable in risk management, integrated into multi-agent systems for financial distress prediction (Yan et al., 2015). In real option valuation, decision trees model uncertainties and complex payout structures flexibly (Hahn et al., 2007). Beyond finance, they are effective in load forecasting for energy (Faber and Finkenrath, 2021), credit behavior prediction (Anand et al., 2022), traffic flow prediction (Xia and Chen, 2017), financial statement fraud detection (Chen et al., 2014), and asset investment discharge prediction (Jannink and Bos, 2005). Their adaptability and ability to handle non-linear data make them invaluable in forecasting and financial analysis.

4.1.2 Bagging

Bagging, or bootstrap aggregation, is a powerful ensemble technique in financial forecasting, known for addressing parameter sensitivity and model uncertainty by averaging estimates across bootstrap samples to reduce parameter instability (Hao et al., 2023; Jordan et al., 2017). Widely applied in economic and financial forecasting, bagging is effective in reducing forecast mean square error, as seen in inflation forecasting (Lee et al., 2019). It also excels in stock price forecasting, outperforming other methods in financial crisis prediction and stock return forecasting (Tsai et al., 2011). Bagging improves forecast accuracy for crude oil prices (Aydoğmuş et al., 2015) and enhances financial risk prediction by creating diverse training datasets for classifiers (Sun, 2012). Additionally, in stock premium predictability, bagging uses moving block bootstrapping, demonstrating its versatility in financial forecasting (Gupta et al., 2016).

4.1.3 Random Forests

Random Forests play a critical role in enhancing predictive accuracy and addressing complex financial challenges via ensemble learning. Introduced by Breiman, Random Forests combine multiple random decision trees to deliver reliable predictions for classification and regression tasks (Davis and Nielsen, 2020). This approach improves accuracy by leveraging an ensemble of trees and injecting randomness during training (Rawnaq et al., 2024). Random Forests have been widely applied in financial forecasting, including stock price prediction, fraud detection, bankruptcy prediction, and option pricing (Qin et al., 2013; Li et al., 2015; Karminsky and Burekhin, 2019; Chen et al., 2021b). Studies confirm their strength in modeling nonlinear relationships, outperforming traditional linear models (Qin et al., 2013). They have also effectively predicted abnormal trading behavior linked to internet rumor spread, highlighting their adaptability to complex market dynamics (Cheng et al., 2023).

4.1.4 AdaBoost

AdaBoost, or Adaptive Boosting, combines multiple weak learners to form a strong learner, effectively capturing patterns in financial data (Li, 2024a). It has been applied successfully in credit assessment, bankruptcy prediction, stock index forecasting, and economic time series prediction, often outperforming traditional models like logistic regression in complex financial scenarios (Bluwstein et al., 2023; Zhou and Lai, 2016; Park et al., 2019; Heo and Yang, 2014). AdaBoost's accuracy improves further when integrated with models such as LSTM and SVM (Nabi et al., 2020; Wu and Gao, 2018). Its ability to handle nonstationary financial data highlights its value in trading and decision-making contexts (Chang et al., 2019). By focusing on misclassified instances, AdaBoost effectively reduces bias and strengthens predictive accuracy (Busari et al., 2021).

4.1.5 Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM), pioneered by Friedman, has become essential for regression tasks across financial applications (Pedchenko et al., 2018). GBM is widely used for forecasting in domains such as natural gas spot prices (Li et al., 2023), house prices (Yan et al., 2015), startup performance (Hahn et al., 2007), agricultural prices (Faber and Finkenrath, 2021), rent prices (Anand et al., 2022), electricity prices, and GDP growth (Xia and Chen, 2017). It often outperforms traditional models in predicting financial crises, bankruptcy, stock indices, and economic time series, showcasing its strength in complex scenarios (Pedchenko et al., 2018; Hahn et al., 2007; Faber and Finkenrath, 2021; Anand et al., 2022; Xia and Chen, 2017). GBM is also effectively combined with other algorithms like XGBoost, LightGBM, and Random Forest to improve forecasting accuracy (Yan et al., 2015; Chen et al., 2014). These hybrid models have shown promise in capturing nuanced patterns in financial data (Yan et al., 2015; Chen et al., 2014). Additionally, GBM is well-suited for nonstationary financial data analysis, supporting robust trading and decision-making processes (Jannink and Bos, 2005). By iteratively refining challenging cases, GBM ensures high accuracy, making it a preferred choice in financial forecasting (Hao et al., 2023).

4.1.6 LightGBM

LightGBM has gained prominence in financial forecasting due to its efficient data handling, low memory usage, and fast processing, especially in large-scale applications (Liu et al., 2023). It has been widely applied to forecast financial parameters like stock returns, water quality (dissolved oxygen levels), and stock prices (Zhou et al., 2023; Cao and Sun, 2024; Liu et al., 2023). Its capability to handle high-dimensional datasets effectively makes it ideal for financial forecasting tasks (Chen et al., 2021a). LightGBM is also combined with algorithms like LSTM, XGBoost, and Random Forest to enhance forecasting accuracy in diverse applications, yielding strong predictive results (Cao and Sun, 2024; Alizadegan et al., 2024). It has further been used in predicting stock volatility, global stock prices during market stress, and Bitcoin prices, underscoring its adaptability in financial forecasting (Gong et al., 2022; Bhuiyan et al., 2023; Liu et al., 2022). LightGBM's efficiency with large datasets and rapid processing capabilities make it a valuable tool for financial measurement and market analysis (Zhang, 2022).

4.1.7 XGBoost

XGBoost has become popular across fields like finance, energy, traffic, and medicine due to its optimization strengths (Zhu et al., 2023). In finance, it is frequently used for predicting price movements (Alshehri, 2023; Choi et al., 2022) and assessing credit debt default risks through variable importance analysis (Wang et al., 2022). Recognized for its powerful optimization, XGBoost is widely adopted for complex research challenges (Bačanin et al., 2022) and has been applied in trading to predict apple futures prices (Deng et al., 2021). Outside of finance, XGBoost also sees applications in healthcare and natural language processing (Du et al., 2024).

4.1.8 CatBoost

CatBoost is highly regarded for its ability to manage categorical features effectively across various applications, including financial forecasting (Prokhorenkova et al., 2017; Dorogush et al., 2018). It has proven successful in predicting credit debt defaults, corporate failure, hotel cancellations, and liquidated damages in construction projects, showcasing strong predictive performance in complex scenarios (Yao and Yang, 2024; Wang et al., 2022; Alshboul et al., 2022). CatBoost also excels in fault diagnosis, anomaly detection, quality prediction, and stock forecasting, illustrating its versatility in financial and industrial domains (Fadzil et al., 2024; Alfarhood et al., 2024). Known for efficiently handling categorical data, it is a preferred tool for precise decision-making. Additionally, CatBoost has been combined with models like XGBoost, LightGBM, and neural

networks to boost accuracy and address unique challenges in financial forecasting, with promising improvements in capturing complex patterns (Alshboul et al., 2022; Sun and Tian, 2023).

4.2 Neural Network-Based Models

4.2.1 Artificial Neural Networks (ANN)

ANNs are extensively applied in financial forecasting due to their capacity to model complex financial data relationships. Researchers have used ANN for tasks such as predicting stock movements, credit risk, bankruptcy risk, and stock index forecasting. Casas (2012) emphasized ANN's importance in series prediction, particularly for financial instruments, by enabling parallelized training. Alamsyah et al. (2021) developed a financial distress early warning model using ANN with backpropagation on financial indicators. Johari et al. (2018) demonstrated ANN's efficacy in financial time series prediction, while Verma et al. (2017) compared ANN algorithms for time series forecasting, underscoring their popularity in finance. Selmi et al. (2015) applied ANN for stock return forecasting, and Cao and He (2019) used it to predict financial crises by linking commodity and stock markets, showcasing ANN's adaptability.

4.2.2 Convolutional Neural Networks (CNNs)

CNNs have gained traction in financial forecasting due to their ability to identify complex patterns in financial time series. They have been applied to various tasks, including stock price forecasting, credit risk assessment, and market trend analysis. For instance, Cao and Wang (2019) developed a CNN-based stock price model that effectively predicted stock indices, while Lu et al. (2023) further validated CNNs' capabilities in stock index forecasting. Mode and Hoque (2020) examined CNN-based model robustness by generating adversarial examples for deep learning regression. Widiputra et al. (2021) introduced a Multivariate CNN-LSTM model for predicting parallel financial time series, highlighting CNN-LSTM's versatility across indices. Qi et al. (2022) proposed a CNN-GRU-attention model for stock prediction in Chinese markets, emphasizing CNNs' feature extraction efficacy and forecasting accuracy in financial data.

4.2.3 Recurrent Neural Networks (RNNs)

RNNs have become central to financial forecasting, valued for their ability to capture temporal dynamics and model nonlinear relationships in stock market data. Studies demonstrate RNNs' use in predicting stock prices, volatility, trading patterns, and portfolio strategies (An, 2023; Dey et al., 2021; Alotaibi et al., 2018; Singh et al., 2019; Yadav and Vasuja, 2019). Integrating RNNs with advanced techniques like LSTM further improves forecast accuracy by capturing market complexity (Raza et al., 2023; Upadhyay et al., 2023). RNNs have also been applied in sentiment analysis to integrate sentimental information with market data, aiding in volatility prediction (Liu et al., 2017). Techniques such as Denoising Autoencoders and Transformers enhance RNN-based stock trend forecasting (Chen, 2024). Combining RNNs with time series algorithms has shown reliable results in historical data forecasting (Yadav and Vasuja, 2019). Recent findings highlight LSTM's superiority over traditional RNNs in precision for stock price prediction (Upadhyay et al., 2023), with effectiveness demonstrated across various companies, showcasing RNN architectures' adaptability to different market scenarios (Gohil and Shah, 2022; Firouzjaee and Khaliliyan, 2024).

4.2.4 Long Short-Term Memory networks (LSTM)

LSTM networks are pivotal in financial forecasting, especially for stock prices and market trends, due to their ability to capture long-term dependencies in financial time series data (Zahrah et al., 2021; Li, 2024b; Zhang, 2022; Reddy et al., 2020; Gajamannage and Park, 2022; Deshpande, 2023; Bhalke et al., 2022; Raut and Shrivas, 2024). Studies consistently demonstrate LSTM's precision in managing both short and long-distance sequences in stock price prediction (Bai et

al., 2023; Qiu et al., 2020; Firouzjaee and Khalilian, 2024; Diqi et al., 2023; Luo, 2018). Techniques like wavelet denoising, whale optimization, and stacked architectures enhance LSTM's predictive accuracy (Wang and Wu, 2023; Kim and Kim, 2019; Fischer and Krauss, 2018). Combining LSTM with CNNs further improves forecasting by leveraging varied data representations (Kim and Kim, 2019; Li et al., 2020). LSTM surpasses models like ARIMA and GRU in stock market prediction (Das, 2024; Fatima and Rahimi, 2024) and is also effective in volatility forecasting, bankruptcy prediction, and financial time series analysis (Ho et al., 2021; Vochozka et al., 2020).

4.3   Instance-based Learning Models

4.3.1   K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm has been applied in financial forecasting, particularly in predicting stock prices. Tanuwijaya and Hansun (2019) demonstrated KNN's potential in predicting the LQ45 stock index. Chen (2024) also employed KNN alongside other machine learning methods for stock price forecasting, highlighting its relevance in financial analysis. Additionally, Islam et al. (2021) emphasized the popularity of KNN and support vector regression in predicting stock prices on the Dhaka Stock Exchange. These studies collectively indicate the practical application and effectiveness of KNN in financial prediction tasks.

4.3.2   Support Vector Machines (SVM)

Support Vector Machines (SVM) have been widely utilized in financial forecasting due to their effectiveness in handling classification and regression tasks. Cao and Tay (2001) introduced the application of SVMs in financial forecasting, highlighting their significance in computational science and engineering. Hossain et al. (2020) successfully used SVMs to predict stock prices, specifically forecasting daily closing prices on the Dhaka Stock Exchange. Bellotti et al. (2011) compared SVMs with ordered choice models in predicting international bank ratings, demonstrating SVMs' capability in time series forecasting. Zhang et al. (2016) enhanced financial time series forecasting accuracy by integrating SVMs with autoregressive integrated moving average (ARIMA). The robustness of SVMs in handling noisy data and generalization challenges has been emphasized by Lin (2012). Overall, SVMs have proven to be a valuable tool in financial forecasting, offering accurate predictions and efficient handling of complex financial data.

5   EXPERIMENTAL SETTING

5.1   Data Description

The dataset utilized in this study encompasses financial data from companies listed on the BIST100 index, which includes the 100 largest companies traded on Borsa Istanbul. The data spans from 2014 to 2024 and includes various financial indicators and technical metrics. Key variables in the dataset include the date of observation, opening price of the stock on the given day, highest price of the stock on the given day, lowest price of the stock on the given day, and closing price of the stock on the given day. Additional variables include the Simple Moving Average (SMA), which is the average of closing prices over a specified period, and the Weighted Moving Average (WMA), which is the average of closing prices over a specified period, weighted by recent prices. Momentum (MOM) is the rate of change of the closing price over a specified period. Stochastic %K (STCK) represents the current closing price relative to the range over a recent period, while Stochastic %D (STCD) is the moving average of the Stochastic %K. The Relative Strength Index (RSI) measures the magnitude of recent price changes to evaluate overbought or

Tab. 1: Characteristics of the dataset

Variable	Description
Date	Date of observation
Open	Opening price on the given day
High	Highest price on the given day
Low	Lowest price on the given day
Close	Closing price on the given day
Simple Moving Average (SMA)	Average of closing prices over a specified period
Weighted Moving Average (WMA)	Average of closing prices over a specified period, weighted by recent prices
Momentum (MOM)	Rate of change of the closing price over a specified period
Stochastic %K (STCK)	Current closing price relative to the range over a recent period
Stochastic %D (STCD)	Moving average of the Stochastic %K
Relative Strength Index (RSI)	Measures the magnitude of recent price changes to evaluate overbought or oversold conditions
Moving Average Convergence Divergence (MACD)	Difference between the 12-day and 26-day exponential moving averages (EMAs)
Larry William's %R (LWR)	Momentum indicator measuring overbought and oversold levels
Accumulation/Distribution Oscillator (ADO)	Indicator combining price and volume to show how much of a stock is being accumulated or distributed
Commodity Channel Index (CCI)	Measures the deviation of the price from its average over a specified period, indicating overbought or oversold conditions

oversold conditions. Moving Average Convergence Divergence (MACD) is the difference between the 12-day and 26-day exponential moving averages (EMAs). Larry William's %R (LWR) is a momentum indicator measuring overbought and oversold levels. The Accumulation/Distribution Oscillator (ADO) combines price and volume to show how much of a stock is being accumulated or distributed. Finally, the Commodity Channel Index (CCI) measures the deviation of the price from its average over a specified period, indicating overbought or oversold conditions.

A comprehensive set of financial indicators has been used to estimate the performance of companies listed in the BIST100 index. These indicators have been selected after a in depth review of the existing literature. The financial indicators have been selected based on their relevance and effectiveness in financial analysis and stock market forecasting and form the basis of machine learning models. Tab. 1 includes detailed explanations of the variables.

5.2 Data Preprocessing

Data preprocessing is a crucial step in preparing the dataset for machine learning models. The following steps outline the preprocessing techniques applied to the BIST100 dataset: First, the dataset was thoroughly cleaned to address missing values, outliers, and inconsistencies. Missing values were handled using interpolation methods, while outliers were identified and treated using statistical techniques such as z-score and IQR methods. Normalization was then applied to ensure that all features contribute equally to the analysis. Numerical variables, including stock prices and technical indicators, were normalized using Min-Max scaling, rescaling the values to a range between 0 and 1. Feature engineering was conducted to create additional features that capture more information from the raw data. New indicators such as moving averages (MA_50, MA_200), Bollinger Bands (BB_Mid, BB_Upper, BB_Lower), and other technical indicators were calculated. Since the dataset consists of time series data, it was necessary to structure the data into a format
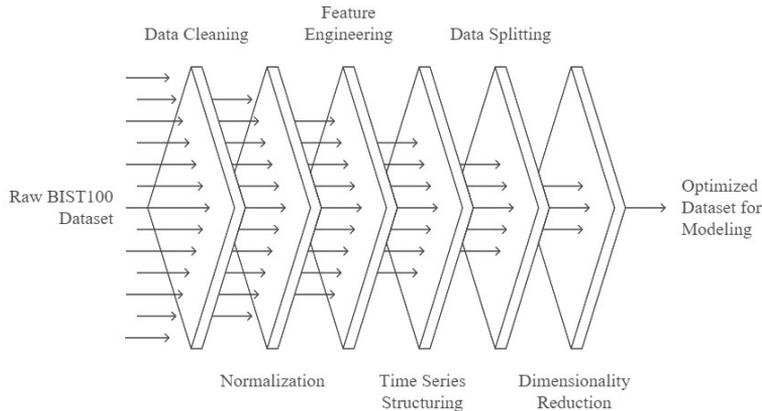
Fig. 1: Data Preprocessing for Machine Learning

suitable for time series analysis. Lag features were created to include past values of stock prices and technical indicators as inputs for the models. The dataset was divided into training and testing sets using a time-based split to ensure that the training data precedes the testing data, maintaining the chronological order of financial data. Dimensionality reduction techniques, such as Principal Component Analysis (PCA), were applied where necessary to reduce computational complexity and enhance model performance. In cases where the target variable exhibited class imbalance, techniques such as oversampling, undersampling, or synthetic data generation (e.g., SMOTE) were employed to balance the dataset. By implementing these preprocessing steps, the dataset was prepared to maximize the performance and accuracy of the machine learning models in predicting the financial performance of companies in the BIST100 index. The data preprocessing workflow – comprising data cleaning, normalization, feature engineering, time series structuring, dimensionality reduction, and data splitting – is summarized in Fig. 1. These steps are critical to ensuring that the raw BIST100 dataset is transformed into an optimized format suitable for machine learning modeling.

5.3   Experimental Design

The experimental design of this study aims to evaluate the effectiveness of various supervised machine learning models in predicting the financial performance of companies listed on the BIST100 index. The process involves several stages, including model selection, training, validation, and evaluation. For model selection, the study includes Tree-Based Models such as Decision Trees, Bagging, Random Forests, AdaBoost, Gradient Boosting Machine (GBM), LightGBM, XGBoost, and CatBoost. Neural Network-Based Models, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory networks (LSTM), are also considered. Additionally, Instance-Based Learning Models such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) are included. Data preparation involves preprocessing the dataset, which comprises financial indicators and technical metrics. This includes data cleaning, normalization, feature engineering, categorical encoding, and creating lag features for time series analysis. The dataset is then split into training and testing sets using a time-based split to maintain the chronological order of the financial data. During model training, each selected model is trained on the training set. Hyperparameter tuning is performed using grid search or random search techniques to optimize the performance of each model. Cross-validation is employed to ensure the robustness of the models, with the training set
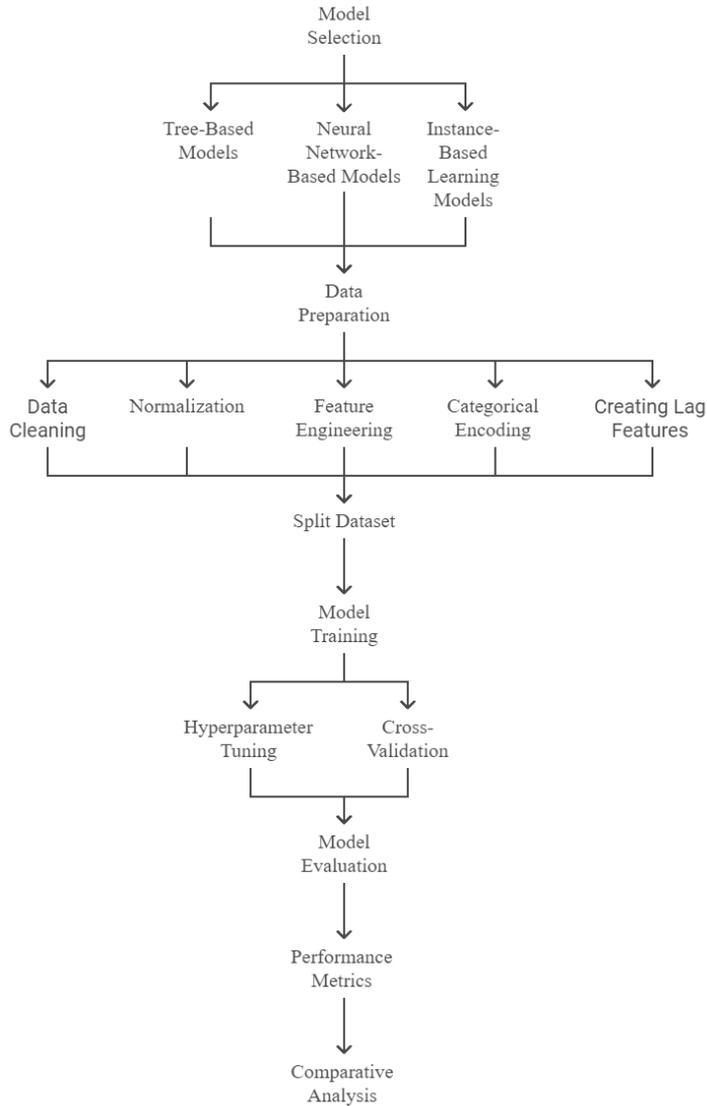
Fig. 2: Experimental design of study

further divided into k-folds (e.g., 5-fold or 10-fold cross-validation) to validate the models during the training phase. Model evaluation involves assessing the trained models on the testing set using various performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and relative Root Mean Squared Error (rRMSE). Comparative analysis is conducted to assess the performance of different models, and statistical significance tests (e.g., paired t-tests) are used to determine if differences in performance are significant. The overall experimental workflow, from model selection and data preparation to model evaluation and comparative analysis, is illustrated in Fig. 2.

5.4   Evaluation Metrics

The evaluation of the machine learning models in this study is based on several key performance metrics that quantify the accuracy and robustness of the predictions. The metrics used are as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2 \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2} \tag{2}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| Y_i - \widehat{Y}_i \right| \tag{3}$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \widehat{y}_i}{y_i} \right| \tag{4}$$

$$\text{rRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}}{\overline{y}} \tag{5}$$

6   RESULTS

The results of this study highlight the effectiveness of various supervised machine learning models in predicting the financial performance of companies listed on the BIST100 index. The evaluation of these models was conducted using several key performance metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and relative Root Mean Squared Error (rRMSE). Tab. 2 presents the comparative error metrics – MSE, RMSE, MAE, MAPE, and rRMSE – for each evaluated model, with the best-performing values highlighted in bold. Fig. 3 to 5 illustrate the model fitting results for the Tree-Based Models, Neural Network-Based Models, and Instance-Based Learning Models, respectively, providing a visual comparison of their predictive accuracy on the BIST100 dataset.

Performance of Tree-Based Models

Tree-based models, including Decision Trees, Bagging, Random Forests, AdaBoost, Gradient Boosting Machine (GBM), LightGBM, XGBoost, and CatBoost, demonstrated strong performance in predicting financial outcomes. Among these, XGBoost and Random Forests exhibited particularly robust and consistent performance across all evaluation metrics. XGBoost, known for its scalability and flexibility, provided high accuracy and low error rates, making it one of the best performers in this category. Random Forests, with its ensemble approach, effectively handled nonlinear relationships and reduced overfitting, resulting in reliable predictions.

Performance of Neural Network-Based Models

Neural Network-based models, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory networks
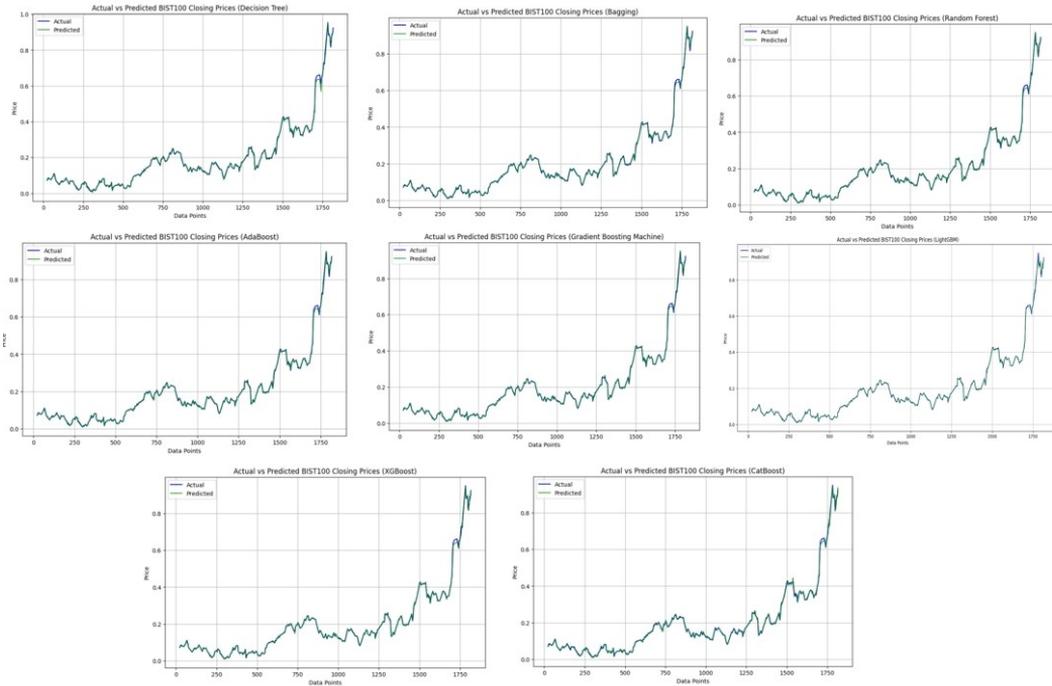
Fig. 3: Tree-based models fitting results. Each subplot shows the predicted vs. actual closing prices for a specific Tree-Based Model (Decision Trees, Bagging, Random Forests, AdaBoost, GBM, LightGBM, XGBoost, CatBoost) applied to the BIST100 dataset. The blue line represents the actual closing prices, while the orange line represents the model's predicted values. Closer alignment between the lines indicates higher predictive accuracy.

Tab. 2: MSE, RMSE, MAE, MAPE and rRMSE values

	MSE	RMSE	MAE	MAPE	rRMSE
Decision Trees	3.258e−05	0.0057	0.0032	2.4591	0.0283
Bagging	1.671e−05	0.0041	0.0026	2.1876	0.0202
Random Forests	1.635e−05	0.0041	0.0024	1.9541	0.0201
Adaboost	2.272e−05	0.0047	0.0025	2.0166	0.0236
GBM	2.035e−05	0.0045	0.0023	1.6585	0.0223
LightGBM	2.739e−05	0.0052	0.0026	1.7510	0.0259
XGBoost	2.662e−05	0.0051	0.0025	1.7688	0.0256
CatBoost	3.129e−05	0.0055	0.0037	3.0035	0.0277
ANN	0.00042	0.0205	0.0151	13.6640	0.1019
CNNs	2.886e−05	0.0053	0.0039	4.1893	0.0266
RNNs	6.565e−05	0.0081	0.0060	5.6011	0.0402
LSTMs	0.0001	0.0106	0.0078	7.3147	0.0528
KNN	0.0001	0.0101	0.0071	7.0558	0.0501
SVM	1.592e−05	0.0039	0.0031	3.5148	0.0198

Note: MSE = Mean Squared Error, RMSE = Root Mean Squared Error, MAE = Mean Absolute Error, MAPE = Mean Absolute Percentage Error, rRMSE = relative Root Mean Squared Error.
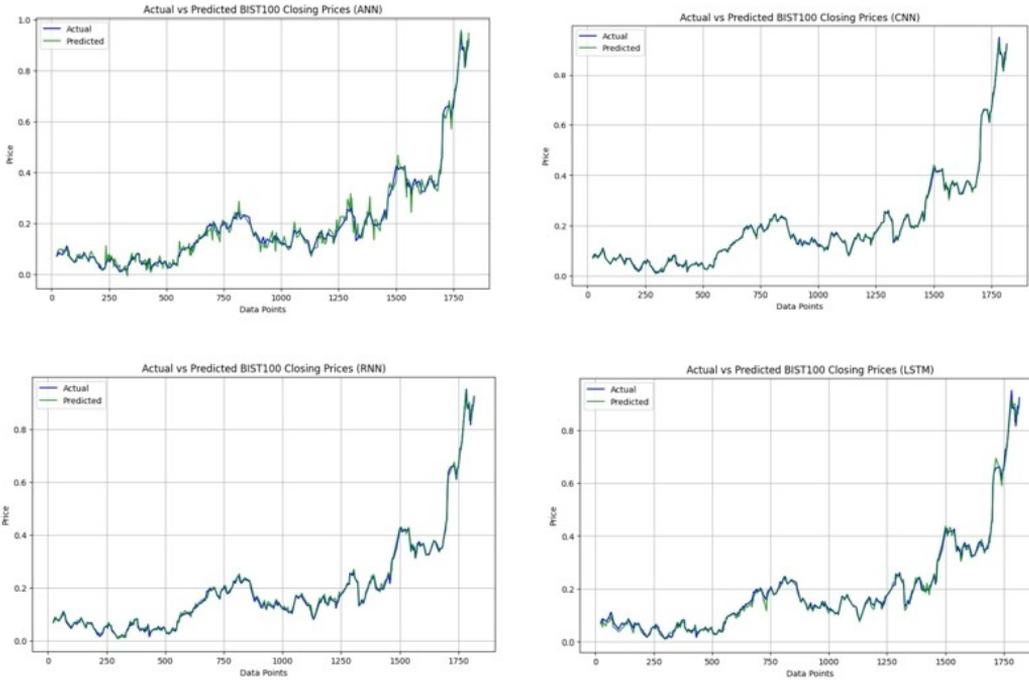
Fig. 4: Neural Network-Based Models fitting results. This figure illustrates the predicted vs. actual financial performance for Neural Network-Based Models (ANN, CNN, RNN, LSTM) using the BIST100 dataset. Lower deviation from the diagonal line indicates better predictive accuracy.
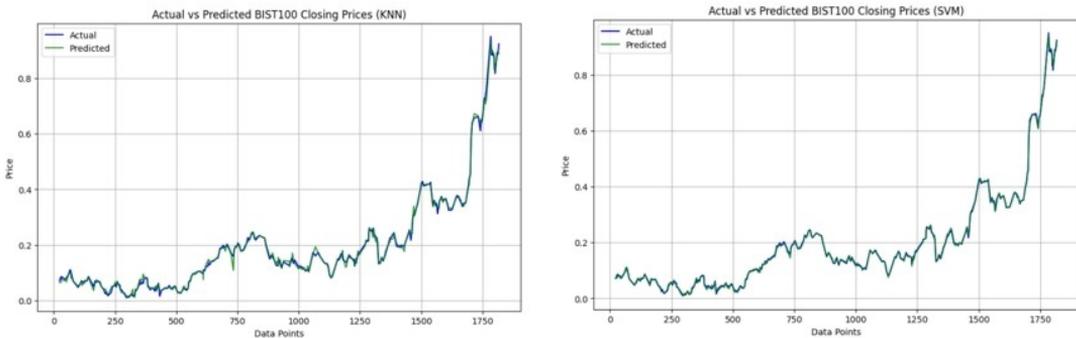

Fig. 5: Instance-based Learning Models fitting results. This figure compares the predicted vs. actual financial performance for Instance-Based Learning Models (KNN, SVM) on the BIST100 dataset. Performance differences are reflected in the proximity of predictions to the actual values.

(LSTM), also showed significant promise. LSTM networks, designed to handle long-term dependencies in time series data, achieved high accuracy in predicting stock prices and market trends. The integration of LSTM with techniques such as wavelet denoising and optimization algorithms further enhanced its predictive capabilities. CNNs were effective in capturing intricate patterns in financial data, and their performance was further improved when combined with RNNs in hybrid models. ANNs and RNNs also provided competitive results, demonstrating their versatility in financial forecasting tasks.

Performance of Instance-Based Learning Models

Instance-based learning models, specifically K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), were evaluated for their predictive accuracy. SVMs showed strong performance in handling classification and regression tasks, effectively predicting daily closing prices and financial risk assessments. The robustness of SVMs in managing noisy data and complex financial patterns was evident from the results. KNN demonstrated its potential in stock index prediction, although its performance was slightly lower compared to SVMs and other advanced models.

Comparative Analysis

The comparative analysis revealed that no single model consistently outperformed others across all companies in the BIST100 index. However, certain models, such as XGBoost and Random Forests, stood out due to their strong and consistent performance. Deep learning algorithms, including CNNs, RNNs, and LSTMs, also showed significant accuracy in predicting the financial outcomes of specific companies. The integration of various techniques and hybrid models often resulted in improved prediction performance, highlighting the importance of model selection and customization based on the specific financial forecasting task. Paired t-tests, were conducted to compare the performance of different models. The results indicated that the differences in performance between the top-performing models were statistically significant, reinforcing the robustness of models like XGBoost, Random Forests, and LSTM networks.

7   DISCUSSION

This study offers valuable insights into the predictive capabilities of supervised machine learning models applied to the financial forecasting of companies listed on the BIST100 index. Our findings reveal significant differences in model performance, with ensemble methods such as Random Forest and XGBoost emerging as the most effective for this context. These models consistently produced lower error metrics across MSE, RMSE, MAE, and MAPE, highlighting their robustness and adaptability to the complexities inherent in financial data.

The superior performance of Random Forest and XGBoost can be attributed to their ensemble learning structures, which mitigate overfitting and better capture non-linear relationships within financial datasets. These characteristics make them highly suitable for environments with high variability, such as the Turkish stock market, where economic and market dynamics are influenced by a range of macroeconomic and company-specific factors. This aligns with Viswanathan and Stephen's (2021) findings, where ensemble approaches, including cross-validation within Random Forest, significantly enhanced the accuracy of stock market index predictions, confirming the effectiveness of ensemble methods in improving error metrics. Similarly, Yang (2023) observed that Random Forest produced high accuracy and precision in predicting stock price directional movements, further supporting the robustness and adaptability of ensemble models in complex financial environments. In line with our results, Shende et al. (2022) highlighted the practical utility of ensemble methods for handling non-linear financial dynamics, demonstrating the suitability of Random Forest and similar models for financial forecasting tasks based on financial ratios.

Additionally, models like LightGBM and Gradient Boosting Machine (GBM) offered a balanced trade-off between accuracy and computational efficiency, positioning them as practical alternatives for similar financial applications where efficiency is prioritized alongside accuracy. Henrique et al. (2023) provide a balanced view on the limitations of these models, noting that while ensemble models like Random Forest are highly effective for general financial forecasting, they may

face challenges in certain short-term prediction tasks. This finding aligns with our observations and reinforces the importance of context-specific adjustments when applying ensemble methods in financial forecasting. Together, these references and our findings underscore the adaptability and accuracy of ensemble models like Random Forest and XGBoost, particularly for long-term forecasting in emerging markets. This study thus provides a solid foundation for leveraging ensemble models in financial markets and supports further research into optimizing these methods for even more nuanced applications.

In contrast, neural network models, including LSTMs and CNNs, while effective in identifying complex temporal patterns in financial data, demonstrated higher error rates in our tests. This suggests that while deep learning techniques hold promise for sequential and time-series analysis, they may require additional tuning and integration with feature engineering techniques to enhance their performance in financial forecasting for the BIST100 index. Beyaz and Efe (2019) similarly observed that neural network ensembles could optimize feature selection and improve forecast accuracy, although single networks displayed greater sensitivity to parameter tuning. This finding reinforces our results, suggesting that an ensemble approach or hybrid models might further optimize neural networks' forecasting accuracy. Furthermore, Cao and Wang (2019) highlighted the utility of CNNs in identifying temporal dependencies, yet found that SVMs demonstrated resilience in managing diverse data patterns, underscoring the balance between deep learning and traditional models. The strengths of neural networks in time-series analysis are further supported by Liu et al. (2022), who applied a bidirectional LSTM model with an attention mechanism, achieving superior results in capturing intricate temporal dependencies compared to SVM and KNN in stock price prediction.

Interestingly, our evaluation of Support Vector Machines (SVM) showed its robustness in handling complex financial patterns, particularly in classification and regression tasks. This aligns with existing literature, such as that by Patel et al. (2023), which demonstrates SVM's effectiveness in high-dimensional, noisy financial data environments. However, K-Nearest Neighbors (KNN) exhibited limited performance, likely due to its sensitivity to the local structure of data points, which may not generalize effectively within the BIST100's diverse economic landscape. Reddy et al. (2024) similarly found that while LSTMs excel in stock market prediction, SVM offers competitive performance in classification tasks involving high-dimensional data, underscoring the versatility of SVM in complex financial contexts. These findings suggest that future studies could explore hybrid models that combine the interpretability and stability of tree-based methods with the temporal insights provided by neural networks, leveraging the strengths of each model type. Such hybrid approaches might bridge the gap between deep learning's advanced pattern recognition and traditional models' robustness, offering improved predictive capabilities in financial forecasting. From a practical standpoint, our findings offer actionable insights for investors, financial analysts, leaders in financial institutions, market authorities, and supervisory professionals operating in emerging market contexts. The consistent superior performance of ensemble models such as XGBoost and Random Forest indicates that practitioners seeking robust, scalable stock prediction solutions should prioritize these algorithms in their forecasting workflows (Henrique et al., 2023; Du et al., 2022). Their adaptability to non-linearities, ability to handle large and complex datasets, and resistance to overfitting make them especially well-suited to volatile markets like Turkey. For institutional investors and portfolio managers, integrating advanced ensemble models into their decision-making processes can enhance return forecasting, risk assessment, and portfolio optimization. For example, using XGBoost-based predictive signals alongside traditional financial analysis can improve the identification of under- or overvalued assets and help to time market entries more effectively (Alshehri, 2023; Moon and Kim, 2019). Financial analysts can benefit by combining model-driven insights with fundamental or technical analysis. Ensemble models' feature importance scores can highlight which macroeconomic

indicators or company ratios most influence stock prices, enabling more targeted research and investment recommendations (Lundberg et al., 2019; Wang et al., 2022). Leaders in financial institutions, market authorities, and supervisory professionals are encouraged to leverage the predictive capabilities of these models for developing AI-driven, real-time decision-support tools in market supervision, systemic risk assessment, and compliance monitoring. As machine learning models can rapidly adapt to shifting market patterns, their integration into industry technology solutions may help in identifying early warning signals of instability or market manipulation, thereby supporting financial system resilience (Du et al., 2022; Henrique et al., 2023).

8 CONCLUSION

This study provides a comprehensive assessment of supervised machine learning models for forecasting the financial performance of companies listed on Turkey's BIST100 index. By conducting a rigorous comparative analysis of tree-based models (Random Forest, XGBoost, LightGBM), neural network models (LSTM, CNN), and instance-based models (SVM, KNN), we have identified ensemble approaches – particularly XGBoost and Random Forest – as consistently delivering superior predictive accuracy and error reduction across multiple performance metrics. Our results indicate that ensemble models are especially adept at capturing the complex, non-linear dynamics characteristic of emerging financial markets, while also maintaining robustness to noise and high-dimensionality. This highlights their practical value for investors, financial analysts, and decision-makers aiming to deploy data-driven strategies in rapidly changing environments. Although neural network models like LSTM and CNN demonstrated strengths in sequential data analysis, their higher error rates point to the need for further refinement – potentially through hybrid architectures or enhanced feature engineering. SVM also exhibited strong predictive capabilities, especially in classification tasks, while KNN's performance was less favorable, likely due to its sensitivity to the local structure of financial data. In terms of actionable recommendations, our findings suggest that financial institutions, analysts, and market authorities in emerging economies should prioritize ensemble models such as XGBoost and Random Forest for reliable, interpretable, and scalable financial forecasting. Moreover, the integration of explainable AI tools can further enhance the transparency and acceptance of these models in professional practice. For future research, we recommend several promising directions: Exploring hybrid models that combine the interpretability and robustness of ensemble methods with the advanced pattern recognition capabilities of deep learning. Incorporating alternative data sources, such as sentiment analysis from social media, macroeconomic indicators, or ESG metrics, to enrich model inputs and capture a broader spectrum of market signals. Applying advanced feature selection and engineering techniques to further improve model accuracy and generalizability. Benchmarking model performance across different emerging markets to evaluate the transferability and contextual robustness of leading algorithms.

9 REFERENCES

ADEGBITE, E., GUNEY, Y., KWABI, F. and TAHIR, S. 2019. Financial and Corporate Social Performance in the UK Listed Firms: The Relevance of Non-Linearity and Lag Effects. *Review of Quantitative Finance and Accounting*, 52, 105–158. DOI: 10.1007/s11156-018-0705-x.

ALAMSYAH, A., KRISTANTI, N. and KRISTANTI, F. T. 2021. Early Warning Model for Financial Distress Using Artificial Neural Network. *IOP Conference Series: Materials Science and Engineering*, 1098 (5), 052103. DOI: 10.1088/1757-899X/1098/5/052103.

Alfarhood, M., Alotaibi, R., Abdulrahim, B., Einieh, A., Almousa, M. and Alkhanifer, A. 2024. Predicting Flight Delays with Machine Learning: A Case Study from Saudi Arabian Airlines. *International Journal of Aerospace Engineering*, 2024, 3385463. DOI: 10.1155/2024/3385463.

Alizadegan, H., Radmehr, A. and Ilani, M. A. 2024. *Forecasting Bitcoin Prices: A Comparative Study of Machine Learning and Deep Learning Algorithms.* Preprint. DOI: 10.21203/rs.3.rs-4390390/v1.

Alotaibi, T., Nazir, A., Alroobaea, R., Alotibi, M., Alsubeai, F., Alghamdi, A. and Alsulimani, T. 2018. Saudi Arabia Stock Market Prediction Using Neural Network. *International Journal on Computer Science and Engineering*, 10 (2), 62–70. DOI: 10.21817/ijcse/2018/v10i2/181002024.

Alshboul, O., Shehadeh, A., Al Mamlook, R. E., Almasabha, G., Almuflih, A. S. and Alghamdi, S. Y. 2022. Prediction Liquidated Damages via Ensemble Machine Learning Model: Towards Sustainable Highway Construction Projects. *Sustainability*, 14 (15), 9303. DOI: 10.3390/su14159303.

Alshehri, A. S. 2023. Predicting Cryptocurrency Returns Using Classification and Regression Machine Learning Models. *Journal of Electrical Systems*, 20 (4s), 539–553. DOI: 10.52783/jes.2065.

An, Z., Jiang, K. and Zheng, J. R. 2023. Features of Realized Volatility Analysis and Return Predicting Based on LGBM and RNN Model. *Applied and Computational Engineering*, 27, 38–48. DOI: 10.54254/2755-2721/27/20230133.

Anand, M., Velu, A. and Whig, P. 2022. Prediction of Loan Behaviour with Machine Learning Models for Secure Banking. *Journal of Computer Science and Engineering*, 3 (1), 1–13. DOI: 10.36596/jcse.v3i1.237.

Aydoğmuş, H. Y., Ekinci, A., Erdal, H. İ. and Erdal, H. 2015. Optimizing the Monthly Crude Oil Price Forecasting Accuracy via Bagging Ensemble Models. *Journal of Economics and International Finance*, 7 (5), 127–136. DOI: 10.5897/JEIF2014.0629.

Bačanin, N., Živković, M., Stoean, C., Antonijević, M., Janičijević, S., Šarac, M. and Štrumberger, I. 2022. Application of Natural Language Processing and Machine Learning Boosted with Swarm Intelligence for Spam Email Filtering. *Mathematics*, 10 (22), 4173. DOI: 10.3390/math10224173.

Bai, Y., Bai, P., Zhang, X. and Li, H. 2023. Financial Asset Volatility Forecasting using LSTM with Intraday High-Low Price Information. In Vilas, G., Shvets, Y. and Mallick, H. (eds.). *MSEA 2023: Proceedings of the 2nd International Conference on Mathematical Statistics and Economic Analysis*. DOI: 10.4108/eai.26-5-2023.2334349.

Bellotti, T., Matousek, R. and Stewart, C. 2011. A Note Comparing Support Vector Machines and Ordered Choice Models' Predictions of International Banks' Ratings. *Decision Support Systems*, 51 (3), 682–687. DOI: 10.1016/j.dss.2011.03.008.

Benchikh, S., Tarik, J., Roa, L. and Elmehdi, N. 2024. Impact of Feature Selection on the Prediction of Global Horizontal Irradiation under Ouarzazate City Climate. *Data & Metadata*, 3, 363. DOI: 10.56294/dm2024363.

Beyaz, K. and Efe, M. Ö. 2019. Forecasting BIST100 Index with Neural Network Ensembles. In *2019 11th International Conference on Electrical and Electronics Engineering (ELECO)*, 940–944. DOI: 10.23919/ELECO47770.2019.8990659.

Bhalke, D. G., Bhingarde, D., Deshmukh, S. and Dhere, D. 2022. Stock Price Prediction Using Long Short Term Memory. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, 14 (2s), 271–273. DOI: 10.18090/samriddhi.v14spli02.12.

Bhuiyan, M. A. M., Zurita-Valdebenito, C., Alam, M. S. and Sarmin, N. 2023. Short and Long-term Forecasting of Emerging Market Data using ARIMA-based and Boosting Machine Learning Algorithms. In *Proceedings of the 5th International Conference on Statistics: Theory and Applications (ICSTA'23)*. DOI: 10.11159/icsta23.158.

Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S. and Şimşek, Ö. 2023. Credit Growth, the Yield Curve and Financial Crisis Prediction: Evidence from a Machine Learning Approach. *Journal of International Economics*, 145, 103773. DOI: 10.1016/j.jinteco.2023.103773.

Breiman, L. 2001. Random Forests. *Machine Learning*, 45, 5–32. DOI: 10.1023/A:1010933404324.

Busari, G. A., Kwak, N. W. and Lim, D. H. 2021. An Application of AdaBoost-GRU Ensemble Model to Economic Time Series Prediction. *Indian Journal of Science and Technology*, 14 (31), 2557–2566. DOI: 10.17485/IJST/v14i31.1204.

Cao, J. and Sun, X. 2024. Analysis of the Difference in Stock Price Between A-shares and American Stocks in Machine Learning. *SHS Web of Conferences*, 181, 02011. DOI: 10.1051/shsconf/202418102011.

Cao, J. and Wang, J. 2019. Stock Price Forecasting Model Based on Modified Convolution Neural Network and Financial Time Series Analysis. *International Journal of Communication Systems*, 32 (12), e3987. DOI: 10.1002/dac.3987.

Cao, L. and Tay, F. E. H. 2001. Financial Forecasting Using Support Vector Machines. *Neural Computing and Applications*, 10 (2), 184–192. DOI: 10.1007/s005210170010.

Cao, W. and He, T. 2019. Predictability of Financial Crisis via Pair Coupling of Commodity Market and Stock Market. *Journal of Finance and Accounting*, 7 (1), 9–16. DOI: 10.11648/j.jfa.20190701.12.

Casas, C. A. 2012. Parallelization of Artificial Neural Network Training Algorithms: A Financial Forecasting Application. In *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics*, 337–342. DOI: 10.1109/CIFEr.2012.6327811.

Chang, V., Li, T. and Zeng, Z. 2019. Towards an Improved Adaboost Algorithmic Method for Computational Financial Analysis. *Journal of Parallel and Distributed Computing*, 134, 219–232. DOI: 10.1016/j.jpdc.2019.07.014.

Chen, H., Didisheim, A. and Scheidegger, S. 2021a. *Deep Structural Estimation: With an Application to Option Pricing*. ArXiv:2102.09209. DOI: 10.48550/arXiv.2102.09209.

Chen, J. M., Zovko, M., Šimurina, N. and Zovko, V. 2021b. Fear in a Handful of Dust: The Epidemiological, Environmental, and Economic Drivers of Death by $PM_{2.5}$ Pollution. *International Journal of Environmental Research and Public Health*, 18 (16), 8688. DOI: 10.3390/ijerph18168688.

Chen, M.-Y. 2011. Bankruptcy Prediction in Firms with Statistical and Intelligent Techniques and a Comparison of Evolutionary Computation Approaches. *Computers & Mathematics with Applications*, 62 (12), 4514–4524. DOI: 10.1016/j.camwa.2011.10.030.

Chen, S., Goo, Y.-J. J. and Shen, Z.-D. 2014. A Hybrid Approach of Stepwise Regression, Logistic Regression, Support Vector Machine, and Decision Tree for Forecasting Fraudulent Financial Statements. *The Scientific World Journal*, 2014 (1), 968712. DOI: 10.1155/2014/968712.

Chen, T. and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. DOI: 10.1145/2939672.2939785.

Chen, Z. 2024. Stock Price Prediction with Denoising Autoencoder and Transformers. *Highlights in Science, Engineering and Technology*, 85, 803–810. DOI: 10.54097/1skct023.

Cheng, L.-C., Lu, W.-T. and Yeo, B. 2023. Predicting Abnormal Trading Behavior from Internet Rumor Propagation: A Machine Learning Approach. *Financial Innovation*, 9, 3. DOI: 10.1186/s40854-022-00423-9.

Choi, I., Yun, W. and Kim, W. C. 2022. Improving Data Efficiency for Analyzing Global Exchange Rate Fluctuations Based on Nonlinear Causal Network-Based Clustering. *Annals of Operations Research*. DOI: 10.1007/s10479-022-05101-8.

Chow, J. C. K. 2018. *Analysis of Financial Credit Risk Using Machine Learning*. ArXiv:1802.05326. DOI: 10.48550/arXiv.1802.05326.

Clements, M. P., Franses, P. H. and Swanson, N. R. 2004. Forecasting Economic and Financial Time-Series with Non-Linear Models. *International Journal of Forecasting*, 20 (2), 169–183. DOI: 10.1016/j.ijforecast.2003.10.004.

Das, J. D., Thulasiram, R. K., Henry, C. and Thavaneswaran, A. 2024. Encoder–Decoder Based LSTM and GRU Architectures for Stocks and Cryptocurrency Prediction. *Journal of Risk and Financial Management*, 17 (5), 200. DOI: 10.3390/jrfm17050200.

Davis, R. A. and Nielsen, M. S. 2020. *Modeling of Time Series using Random Forests: Theoretical Developments*. ArXiv:2008.02479. DOI: 10.48550/arXiv.2008.02479.

Deng, S., Huang, X., Wang, J., Qin, Z., Fu, Z., Wang, A. and Yang, T. 2021. A Decision Support System for Trading in Apple Futures Market Using Predictions Fusion. *IEEE Access*, 9, 1271–1285. DOI: 10.1109/access.2020.3047138.

Deshpande, V. 2023. Implementation of Long Short-Term Memory (LSTM) Networks for Stock Price Prediction. *Research Journal of Computer Systems and Engineering*, 4 (2), 60–72. DOI: 10.52710/rjcse.74.

Dey, P., Hossain, E., Hossain, M. I., Chowdhury, M. A., Alam, M. S., Hossain, M. S. and Andersson, K. 2021. Comparative Analysis of Recurrent Neural Networks in Stock Price Prediction for Different Frequency Domains. *Algorithms*, 14 (8), 251. DOI: 10.3390/a14080251.

Dietterich, T. G. 2000. Ensemble Methods in Machine Learning. In *Proceedings of the First International Workshop on Multiple Classifier Systems*. Lecture Notes in Computer Science, 1857, 1–15. DOI: 10.1007/3-540-45014-9_1.

Diqi, M., Ordiyasa, I. W. and Hamzah, H. 2024. Enhancing Stock Price Prediction Using Stacked Long Short-Term Memory. *IT Journal Research & Development*, 8 (2), 164–174. DOI: 10.25299/itjrd.2023.13486.

Dorogush, A. V., Ershov, V. and Gulin, A. 2018. *CatBoost: Gradient Boosting with Categorical Features Support*. ArXiv:1810.11363. DOI: 10.48550/arXiv.1810.11363.

Du, H., Lv, L., Wang, H. and Guo, A. 2024. A Novel Method for Detecting Credit Card Fraud Problems. *PLOS One*, 19 (3), e0294537. DOI: 10.1371/journal.pone.0294537.

Du, S., Hao, D. and Li, X. 2022. Research on Stock Forecasting Based on Random Forest. In *2022 IEEE 2nd International Conference on Data Science and Computer Application*, 301–305. DOI: 10.1109/ICDSCA56264.2022.9987903.

Faber, T. and Finkenrath, M. 2021. Load Forecasting in District Heating Systems Using Stacked Ensembles of Machine Learning Algorithms. In *Proceedings of the 14th International Renewable Energy Storage Conference 2020*. DOI: 10.2991/ahe.k.210202.001.

Fadzil, M. A. M., Razali, A. A., Zabiri, H. and Hussin, A. H. C. 2024. Investigative Analysis of Automatic Mode Detection for a Lubricant Base Oil Production Plant Using PCA and Machine-Learning Models. *ACS Omega*, 9 (3), 3525–3540. DOI: 10.1021/acsomega.3c07331.

Fatima, S. S. W. and Rahimi, A. 2024. A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems. *Machines*, 12 (6), 380. DOI: 10.3390/machines12060380.

Firouzjaee, J. and Khalilian, P. 2024. The Interpretability of LSTM Models for Predicting Oil Company Stocks: Impact of Correlated Features. *International Journal of Energy Research*, 1–18. DOI: 10.1155/2024/5526692.

Fischer, T. and Krauss, C. 2018. Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. *European Journal of Operational Research*, 270 (2), 654–669. DOI: 10.1016/j.ejor.2017.11.054.

Gajamannage, K. and Park, Y. 2022. *Real-time Forecasting of Time Series in Financial Markets Using Sequentially Trained Many-to-one LSTMs*. ArXiv:2205.04678. DOI: 10.48550/arxiv.2205.04678.

Gohil, J. and Shah, M. 2022. *Application of Big Data in Petroleum Streams*. 1st ed. CRC Press. DOI: 10.1201/9781003185710.

Gong, X., Guan, K. and Chen, Q. 2022. The Role of Textual Analysis in Oil Futures Price Forecasting Based on Machine Learning Approach. *Journal of Futures Markets*, 42 (10), 1987–2017. DOI: 10.1002/fut.22367.

Gradojevic, N. and Yang, J. 2006. Non-linear, Non-parametric, Non-fundamental Exchange Rate Forecasting. *Journal of Forecasting*, 25 (4), 227–245. DOI: 10.1002/for.986.

Gupta, R., Modise, M. P. and Uwilingiye, J. 2016. Out-of-Sample Equity Premium Predictability in South Africa: Evidence from a Large Number of Predictors. *Emerging Markets Finance and Trade*, 52 (8), 1935–1955. DOI: 10.1080/1540496x.2015.1058075.

Hahn, W. J., Dyer, J. and Brandão, L. E. 2007. Using Decision Analysis to Solve Real Option Valuation Problems: Building a Generalized Approach. In *Hydrocarbon Economics and Evaluation Symposium*. DOI: 10.2118/108066-MS.

Hao, J., Feng, Q. Q., Li, J. and Sun, X. 2023. A Bi-Level Ensemble Learning Approach to Complex Time Series Forecasting: Taking Exchange Rates as an Example. *Journal of Forecasting*, 42 (6), 1385–1406. DOI: 10.1002/for.2971.

Henrique, B. M., Sobreiro, V. A. and Kimura, H. 2023. Practical Machine Learning: Forecasting Daily Financial Markets Directions. *Expert Systems with Applications*, 233 (2), 120840. DOI: 10.1016/j.eswa.2023.120840.

Heo, J. and Yang, J. J. 2014. Bankruptcy Forecasting Model Using AdaBoost: A Focus on Construction Companies. *Journal of Intelligence and Information Systems*, 20 (1), 35–48. DOI: 10.13088/jiis.2014.20.1.035.

Ho, M. K., Darman, H. and Musa, S. 2021. Stock Price Prediction Using ARIMA, Neural Network and LSTM Models. *Journal of Physics: Conference Series*, 1988 (1), 012041. DOI: 10.1088/1742-6596/1988/1/012041.

Höbarth, L. L. 2006. *Modeling the Relationship Between Financial Indicators and Company Performance. An Empirical Study for US-Listed Companies*. Doctoral thesis. Institute for Statistics and Mathematics. DOI: 10.57938/ f5f31af3-19a7-488e-9497-e8998438e049.

Hochreiter, S. and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural Computation*, 9 (8), 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

Hoque, K. E. and Aljamaan, H. 2021. Impact of Hyperparameter Tuning on Machine Learning Models in Stock Price Forecasting. *IEEE Access*, 9, 163815–163830. DOI: 10.1109/access.2021.3134138.

Hossain, M. F., Islam, S., Chakraborty, P. and Majumder, A. K. 2020. Predicting Daily Closing Prices of Selected Shares of Dhaka Stock Exchange (DSE) Using Support Vector Machines. *Internet of Things and Cloud Computing*, 8 (4), 46–51. DOI: 10.11648/j.iotcc.20200804.12.

Iqbal, R., Doctor, F., More, B., Mahmud, S. and Yousuf, U. 2020. Big Data Analytics: Computational Intelligence Techniques and Application Areas. *Technological Forecasting and Social Change*, 153, 119253. DOI: 10.1016/j.techfore.2018.03.024.

Islam, S., Sikder, M. S., Hossain, M. F. and Chakraborty, P. 2021. Predicting the Daily Closing Price of Selected Shares on the Dhaka Stock Exchange Using Machine Learning Techniques. *SN Business & Economics*, 1 (4), 58. DOI: 10.1007/s43546-021-00065-6.

Jan, C.-L. 2021. Financial Information Asymmetry: Using Deep Learning Algorithms to Predict Financial Distress. *Symmetry*, 13 (3), 443. DOI: 10.3390/sym13030443.

Jannink, J. W. and Bos, C. F. M. 2005. Probabilistic Discharge Forecasting for Improved Asset Investment Decision Support. In *SPE Europec/EAGE Annual Conference.* DOI: 10.2118/94116-MS.

Johari, S. N. M., Farid, F. H. M., Nasrudin, N. A. E. B., Bistamam, N. S. L. and Shuhaili, N. S. S. M. 2018. Predicting Stock Market Index Using Hybrid Intelligence Model. *International Journal of Engineering and Technology*, 7 (3.15), 36–39. DOI: 10.14419/ijet.v7i3.15.17403.

Jones, S. 2017. Corporate Bankruptcy Prediction: A High Dimensional Analysis. *Review of Accounting Studies*, 22 (3), 1366–1422. DOI: 10.1007/s11142-017-9407-1.

Jordan, M. I. and Mitchell, T. M. 2015. Machine Learning: Trends, Perspectives, and Prospects. *Science*, 349 (6245), 255–260. DOI: 10.1126/science.aaa8415.

Jordan, S. J., Vivian, A. and Wohar, M. E. 2017. Forecasting Market Returns: Bagging or Combining? *International Journal of Forecasting*, 33 (1), 102–120. DOI: 10.1016/j.ijforecast.2016.07.003.

Kambale, W. V., Salem, M., Benarbia, T., Al Machot, F. and Kyamakya, K. (2024). Comprehensive Sensitivity Analysis Framework for Transfer Learning Performance Assessment for Time Series Forecasting: Basic Concepts and Selected Case Studies. *Symmetry*, 16 (2), 241. DOI: 10.3390/sym16020241.

Karminsky, A. M. and Burekhin, R. N. 2019. Comparative Analysis of Methods for Forecasting Bankruptcies of Russian Construction Companies. *Business Informatics*, 13 (3), 52–66. DOI: 10.17323/1998-0663.2019.3.52.66.

Kim, T. and Kim, H. Y. 2019. Forecasting Stock Prices with a Feature Fusion LSTM-CNN Model Using Different Representations of the Same Data. *PLOS One*, 14 (2), e0212320. DOI: 10.1371/journal.pone.0212320.

Koller, T., Goedhart, M. and Wessels, D. 2010. *Valuation: Measuring and Managing the Value of Companies.* 5th ed. John Wiley & Sons. ISBN 978-0-470-88996-1.

Lee, T.-H., Ullah, A. and Wang, R. 2019. Bootstrap Aggregating and Random Forest. In Fuleky, P. (ed.). *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*, pp. 389–429. DOI: 10.1007/978-3-030-31150-6_13.

Li, C., Chan, Y., Kazmi, S. H. A. and Fu, H. 2015. Financial Fraud Detection Model: Based on Random Forest. *International Journal of Economics and Finance*, 7 (7), 178–188. DOI: 10.5539/ijef.v7n7p178.

Li, H. 2024a. Optimizing Stock Price Prediction: Exploring LSTM Architectural Parameters in Financial Forecasting. *Highlights in Science Engineering and Technology*, 85, 1095–1100. DOI: 10.54097/40px3f62.

Li, N. 2024b. Literature Review: Machine Learning in Stock Predictions. *Highlights in Business Economics and Management*, 24, 853–859. DOI: 10.54097/81x6z947.

Li, Q., Tan, J., Wang, J. and Chen, H. 2020. A Multimodal Event-Driven LSTM Model for Stock Prediction Using Online News. *IEEE Transactions on Knowledge and Data Engineering*, 33 (10), 3323–3337. DOI: 10.1109/TKDE.2020.2968894.

Li, R., Ma, M. and Tang, N. 2023. Stock Price Prediction Based on Decision Trees, CNN and LSTM. In *Proceedings of the 4th International Conference on Economic Management and Model Engineering.* DOI: 10.4108/eai.18-11-2022.2327160.

Liaw, A. and Wiener, M. 2002. Classification and Regression by randomForest. *R News*, 2/3, 18–22.

Lin, T.-C. 2012. Decision-Based Filter Based on SVM and Evidence Theory for Image Noise Removal. *Neural Computing and Applications*, 21 (4), 695–703. DOI: 10.1007/s00521-011-0648-9.

Liu, W., Liu, S., Hassan, S. G., Cao, Y., Xu, L., Feng, D., Cao, L., Chen, W., Chen, Y., Guo, J., Liu, T. and Zhang, H. 2023. A Novel Hybrid Model to Predict Dissolved Oxygen for Efficient Water Quality in Intensive Aquaculture. *IEEE Access*, 11, 29162–29174. DOI: 10.1109/ACCESS.2023.3260089.

Liu, W., Zhang, Y. and Liu, Y. 2022. Attention-based BiLSTM Model for Stock Price Prediction. In *Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition*, pp. 257–263. DOI: 10.1145/3573942.3574019.

Liu, Y., Qin, Z., Li, P. and Wan, T. 2017. Stock Volatility Prediction Using Recurrent Neural Networks with Sentiment Analysis. In Benferhat, S., Tabia, K. and Ali, M. (eds.). *Advances in Artificial Intelligence: From Theory to Practice*, pp. 192–201. Lecture Notes in Computer Science, 10350. DOI: 10.1007/978-3-319-60042-0_22.

Lu, J., Ding, Y. and Li, Z. 2023. Optimizing Financial Engineering Time Indicator Using Bionics Computation Algorithm and Neural Network. In *Fifth International Conference on Artificial Intelligence and Computer Science.* DOI: 10.1117/12.3009279.

LUNDBERG, S. M., ERION, G. G. and LEE, S.-I. 2019. *Consistent Individualized Feature Attribution for Tree Ensembles.* ArXiv:1802.03888. DOI: 10.48550/arXiv.1802.03888.

LUO, T. 2018. Research on Decision-Making of Complex Venture Capital Based on Financial Big Data Platform. *Complexity*, 1–12. DOI: 10.1155/2018/5170281.

MARGARAT, G. S., KUMAR, C. S., RAJAN, S. and RAJ, M. B. 2023. Forecasting Wind Energy Production Using Machine Learning Techniques. *E3S Web of Conferences*, 387 (2), 01007. DOI: 10.1051/e3sconf/202338701007.

MODE, G. R. and HOQUE, K. A. 2020. *Adversarial Examples in Deep Learning for Multivariate Time Series Regression.* ArXiv:2009.11911. DOI: 10.48550/arxiv.2009.11911.

MOON, K.-S. and KIM, H. 2019. Performance of Deep Learning in Prediction of Stock Market Volatility. *Economic Computation and Economic Cybernetics Studies and Research*, 53 (2), 77–92. DOI: 10.24818/18423264/53.2.19.05.

NABI, R. M., SAEED, S. A. M. and HARRON, H. 2020. A Novel Approach for Stock Price Prediction Using Gradient Boosting Machine with Feature Engineering (GBM-wFE). *Kurdistan Journal of Applied Research*, 5 (1), 28–48. DOI: 10.24017/science.2020.1.3.

NOSRATABADI, S., MOSAVI, A., DUAN, P., GHAMISI, P., FILIP, F., BAND, S. S., REUTER, U., GAMA, J. and GANDOMI, A. H. 2020. Data Science in Economics: Comprehensive Review of Advanced Machine Learning and Deep Learning Methods. *Mathematics*, 8 (10), 1799. DOI: 10.3390/math8101799.

PARK, M., LEE, M. L. and LEE, J. 2019. Predicting Stock Market Indices Using Classification Tools. *Asian Economic and Financial Review*, 9 (2), 243–256. DOI: 10.18488/journal.aefr.2019.92.243.256.

PATEL, V., KUMAR, A. and YADAV, D. 2023. Machine Learning Techniques for Predicting Stock Closing Prices. In *2023 3rd International Conference on Pervasive Computing and Social Networking*, pp. 447–452. DOI: 10.1109/ICPCSN58827.2023.00079.

PEDCHENKO, N., STRILEC, V., KOLISNYK, G. M., DYKHA, M. and FROLOV, S. 2018. Business Angels as an Alternative to Financial Support at the Early Stages of Small Businesses' Life Cycle. *Investment Management and Financial Innovations*, 15 (1), 166–179. DOI: 10.21511/imfi.15(1).2018.15.

PROKHORENKOVA, L., GUSEV, G., VOROBEV, A., DOROGUSH, A. V. and GULIN, A. 2017. *CatBoost: Unbiased Boosting with Categorical Features.* ArXiv:1706.09516. DOI: 10.48550/arxiv.1706.09516.

QI, J., HUANG, S., HU, J., NI, W. and CHEN, H. 2022. Stock Price Prediction in Chinese Stock Markets Based on CNN-GRU-attention Model. In *International Symposium on Artificial Intelligence and Robotics 2022*, 12508, pp. 182–189. DOI: 10.1117/12.2663261.

QIN, Q., WANG, Q.-G., LI, J. and GE, S. S. 2013. Linear and Nonlinear Trading Models with Gradient Boosted Random Forests and Application to Singapore Stock Market. *Journal of Intelligent Learning Systems and Applications*, 5 (1), 1–10. DOI: 10.4236/jilsa.2013.51001.

QIU, J., WANG, B. and ZHOU, C. 2020. Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism. *PLOS One*, 15 (1), e0227222. DOI: 10.1371/journal.pone.0227222.

RAUT, S. and SHRIVAS, A. 2024. Analysis & Stock Price Prediction and Forecasting Using Different LSTM Models. *International Journal of Scientific Research in Engineering and Management*, 8 (4), 1–5. DOI: 10.55041/ijsrem30115.

RAWNAQ, E., ESMATYAR, B., HAMANAKA, A., TASAOKA, T. and SHIMADA, H. 2024. A Comparative Study of Two Tree-Based Models for Predicting Flyrock Velocity at Open Pit Bench Mining. *Open Journal of Applied Sciences*, 14 (2), 267–287. DOI: 10.4236/ojapps.2024.142019.

RAZA, A., JAVED, M., FAYAD, A. and KHAN, A. Y. 2023. Advanced Deep Learning-Based Predictive Modelling for Analyzing Trends and Performance Metrics in Stock Market. *Journal of Accounting and Finance in Emerging Economies*, 9 (3), 277–294. DOI: 10.26710/jafee.v9i3.2739.

REDDY, D. J., DONALD, A. D., RAMANA, K. S., LAKSHMI, K. S. and DIVYA, P. C. S. 2020. Cross Entropy Based Long Short Term Memory Recurrent Neural Network Model for Analyzing the Time Series on Stock Market Price. *International Journal of Intelligent Engineering & Systems*, 13 (2), 259–266. DOI: 10.22266/ijies2020.0430.25.

REDDY, V. M., NAVEEN, D. N., SUNDHAR, D. N. and VICTORIA, K. L. 2024. Deep Insights: Revolutionizing Stock Market Predictions with Machine Learning and Deep Learning Techniques. In *2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence*, pp. 1–6. DOI: 10.1109/RAEEUCCI61380.2024.10547777.

REEL, P. S., REEL, S., PEARSON, E., TRUCCO, E. and JEFFERSON, E. 2021. Using Machine Learning Approaches for Multi-Omics Data Analysis: A Review. *Biotechnology Advances*, 49, 107739. DOI: 10.1016/j.biotechadv.2021.107739.

REFENES, A.-P. N., BURGESS, A. N. and BENTZ, Y. 1997. Neural Networks in Financial Engineering: A Study in Methodology. *IEEE Transactions on Neural Networks*, 8 (6), 1222–1267. DOI: 10.1109/72.641449.

ROBERTS, P. W. and DOWLING, G. R. 2002. Corporate Reputation and Sustained Superior Financial Performance. *Strategic Management Journal*, 23 (12), 1077–1093. DOI: 10.1002/smj.274.

RYLL, L. and SEIDENS, S. 2019. *Evaluating the Performance of Machine Learning Algorithms in Financial Market Forecasting: A Comprehensive Survey.* ArXiv:1906.07786. DOI: 10.48550/arxiv.1906.07786.

SELMI, N., CHAABENE, S. and HACHICHA, N. 2015. Forecasting Returns on a Stock Market Using Artificial Neural Networks and Garch Family Models: Evidence of Stock Market S&P 500. *Decision Science Letters*, 4 (2), 203–210. DOI: 10.5267/j.dsl.2014.12.002.

SHEN, Z., WAN, Q. and LEATHAM, D. J. 2021. Bitcoin Return Volatility Forecasting: A Comparative Study between Garch and RNN. *Journal of Risk and Financial Management*, 14 (7), 337. DOI: 10.3390/jrfm14070337.

SHENDE, S. D., SINGH, A. S., SHAH, S. S., SHINDE, M. M., MORE, S. R. and AINAPURE, B. 2022. Stocks Price Prediction by Fundamental Analysis Using Machine Learning Algorithms. In *2022 5th International Conference on Contemporary Computing and Informatics*, pp. 1515–1522. DOI: 10.1109/IC3I56241.2022.10072563.

SINGH, S., AHMAD, M., BHATTACHARYA, A. and AZHAGIRI, M. 2019. Predicting Stock Market Trends using Hybrid SVM Model and LSTM with Sentiment Determination using Natural Language Processing. *International Journal of Engineering and Advanced Technology*, 9 (1), 2870–2875. DOI: 10.35940/ijeat.a1106.109119.

SUN, J. 2012. Integration of Random Sample Selection, Support Vector Machines and Ensembles for Financial Risk Forecasting with an Empirical Analysis on the Necessity of Feature Selection. *Intelligent Systems in Accounting Finance and Management*, 19 (4), 229–246. DOI: 10.1002/isaf.1331.

SUN, Y. and TIAN, L. 2023. Research on Stock Prediction Based on LSTM and CatBoost Algorithm. In *Proceedings of the 2nd International Conference on Bigdata Blockchain and Economy Management.* DOI: 10.4108/eai.19-5-2023.2334326.

TANUWIJAYA, J. and HANSUN, S. 2019. LQ45 Stock Index Prediction Using *k*-Nearest Neighbors Regression. *International Journal of Recent Technology and Engineering*, 8 (3), 2388–2391. DOI: 10.35940/ijrte.c4663.098319.

TSAI, C.-F., LIN, Y.-C., YEN, D. C. and CHEN, Y.-M. 2011. Predicting Stock Returns by Classifier Ensembles. *Applied Soft Computing*, 11 (2), 2452–2459. DOI: 10.1016/j.asoc.2010.10.001.

UPADHYAY, N. K., SINGH, V., SINGH, S. and KHANNA, P. 2023. Enhancing Stock Market Predictability: A Comparative Analysis of RNN and LSTM Models for Retail Investors. *Journal of Management and Service Science*, 3 (1), 1–9. DOI: 10.54060/jmss.v3i1.42.

VERMA, S. A., THAMPI, G. T. and RAO, M. 2017. Inter-Comparison of Artificial Neural Network Algorithms for Time Series Forecasting: Predicting Indian Financial Markets. *International Journal of Computer Applications*, 162 (2), 1–13. DOI: 10.5120/ijca2017913249.

VISWANATHAN, T. and STEPHEN, M. 2021. Does Machine Learning Algorithms Improve Forecasting Accuracy? Predicting Stock Market Index Using Ensemble Model. In *Advances in Distributed Computing and Machine Learning*, pp. 511–519. Lecture Notes in Networks and Systems, 127. DOI: 10.1007/978-981-15-4218-3_50.

VOCHOZKA, M., VRBKA, J. and ŠULEŘ, P. 2020. Bankruptcy or Success? The Effective Prediction of a Company's Financial Development Using LSTM. *Sustainability*, 12 (18), 7529. DOI: 10.3390/su12187529.

WANG, J., RONG, W., ZHANG, Z. and MEI, D. 2022. Credit Debt Default Risk Assessment Based on the XGBoost Algorithm: An Empirical Study from China. *Wireless Communications and Mobile Computing*, 2022. DOI: 10.1155/2022/8005493.

WANG, W. and WU, Y. 2023. Risk Analysis of the Chinese Financial Market with the Application of a Novel Hybrid Volatility Prediction Model. *Mathematics*, 11 (18), 3937. DOI: 10.3390/math11183937.

WANG, Z. 2024. Stock Price Prediction Using LSTM Neural Networks: Techniques and Applications. *Applied and Computational Engineering*, 86 (1), 275–281. DOI: 10.54254/2755-2721/86/20241605.

WIDIPUTRA, H., MAILANGKAY, A. and GAUTAMA, E. 2021. Multivariate CNN-LSTM Model for Multiple Parallel Financial Time-Series Prediction. *Complexity*, 2021, 1–14. DOI: 10.1155/2021/9903518.

WU, Y. and GAO, J. 2018. AdaBoost-Based Long Short-Term Memory Ensemble Learning Approach for Financial Time Series Forecasting. *Current Science*, 115 (1), 159. DOI: 10.18520/cs/v115/i1/159-165.

XIA, Y. and CHEN, J. 2017. Traffic Flow Forecasting Method Based on Gradient Boosting Decision Tree. In *Proceedings of the 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology.* DOI: 10.2991/fmsmt-17.2017.87.

Xiuguo, W. and Shengyong, D. 2022. An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning. *IEEE Access*, 10, 22516–22532. DOI: 10.1109/ACCESS.2022.3153478.

Yadav, G. and Vasuja, R. 2019. Analysis of Time Series Prediction Using Recurrent Neural Networks. *International Journal of Computer Applications*, 182 (48), 34–40. DOI: 10.5120/ijca2019918732.

Yan, J., Liao, J.-J. and Shih, C.-H. 2015. Multi-Agent Hybrid Mechanism for Financial Risk Management. *Journal of Industrial Engineering and Management*, 8 (2), 435–452. DOI: 10.3926/jiem.1313.

Yang, M. 2023. Predicting the Direction of Stock Price Movement with Machine Learning Algorithms. *Advances in Economics, Management and Political Sciences*, 52 (1), 283–291. DOI: 10.54254/2754-1169/52/20230758.

Yao, X. and Yang, X. 2024. Forecasting Crude Oil Futures Using an Ensemble Model Including Investor Sentiment and Attention. *Kybernetes*, 53 (12), 6114–6138. DOI: 10.1108/K-03-2023-0364.

Yaprakdal, F. and Bal, F. 2022. Comparison of Robust Machine-Learning and Deep-Learning Models for Midterm Electrical Load Forecasting. *European Journal of Technique*, 12 (2), 102–107. DOI: 10.36222/ejt.1201977.

Zahrah, H. H., Sa'adah, S. and Rismala, R. 2021. The Foreign Exchange Rate Prediction Using Long-Short Term Memory. *International Journal on Information and Communication Technology*, 6 (2), 94–105. DOI: 10.21108/ijoict.2020.62.538.

Zhang, X. 2022. A Model Combining LightGBM and Neural Network for High-Frequency Realized Volatility Forecasting. In *Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development*. DOI: 10.2991/aebmr.k.220307.473.

Zhou, L. and Lai, K. K. 2016. AdaBoost Models for Corporate Bankruptcy Prediction with Missing Data. *Computational Economics*, 50 (1), 69–94. DOI: 10.1007/s10614-016-9581-4.

Zhou, Z., Song, Z., Ren, T. and Yu, L. 2023. Two-Stage Portfolio Optimization Integrating Optimal Sharp Ratio Measure and Ensemble Learning. *IEEE Access*, 11, 1654–1670. DOI: 10.1109/access.2022.3232281.

Zhu, C., Beatty, T., Zhao, Q., Si, W. and Chen, Q. 2023. Leveraging Genetic Data for Predicting Consumer Choices of Alcoholic Products. *China Agricultural Economic Review*, 15 (4), 685–707. DOI: 10.1108/caer-09-2022-0214.

AUTHOR'S ADDRESS

Cansu Ergenç, Ankara Yıldırım Beyazıt University, Business School, Ankara, Türkiye, e-mail: cansuergenc@aybu.edu.tr, ORCID: 0000-0002-4722-0911 (corresponding author)

Rafet Aktaş, Ankara Yıldırım Beyazıt University, Business School, Ankara, Türkiye, ORCID: 0009-0008-8033-4604