Stock market price prediction using Long Short-Term Memory (LSTM) in combination with Random Forest algorithm

A.AJAY , DR.DINESH KUMAR

A.Ajay

Research Scholar,

Department of computer science,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pincode-602105.

[192124147.sse@saveetha.com](mailto:192124147.sse@saveetha.com)

DR. Dinesh kumar

Project Guide, Corresponding Author,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

KEYWORDS:

machine learning, stock market prediction, literature review, research taxonomy, artificial neural network, support vector machine, genetic algorithm, investment decision

**ABSTRACT**

AIM: The aim of the research on market price prediction using Long Short-Term Memory (LSTM) in combination with Random Forest algorithm is to enhance the accuracy and efficiency of stock market forecasting by leveraging advanced deep learning techniques. MATERIALS AND METHODS: The introduction to this research on market price prediction using Long Short-Term Memory (LSTM) in conjunction with Random Forest algorithm delves deeper into the intricacies of stock market forecasting[(Jabed 2024)](https://paperpile.com/c/53rYV5/xNC3). The volatile and unpredictable nature of financial markets poses a significant challenge for investors and analysts seeking to make informed decisions. Traditional methods of stock price prediction often fall short in capturing the complex patterns and trends inherent in market data. In response to these challenges, this study explores the application of advanced deep learning techniques, specifically LSTM, renowned for its ability to model long-term dependencies in sequential data. By combining LSTM with Random Forest algorithm, a powerful machine learning algorithm known for its effectiveness in classification and regression tasks, the research aims to enhance the accuracy and reliability of stock price predictions.

This research delves into the intricate comparison of Random Forest algorithm and Long Short-Term Memory (LSTM) deep learning algorithms for the prediction of stock market prices. The study focuses on analyzing and contrasting the performance of Random Forest algorithm and LSTM models using historical data from the Dow Jones Industrial Average (DJIA) stock prices, while also integrating external factors such as crude oil and gold prices as additional input parameters[(Bosco and Khan 2018)](https://paperpile.com/c/53rYV5/3AhH). The experimental results reveal that LSTM consistently outperforms Random Forest algorithm across various scenarios due to its inherent ability to capture and process complex temporal dependencies within the data. By leveraging LSTM's capability to retain information over extended periods, the model demonstrates superior predictive accuracy compared to Random Forest algorithm. Furthermore, the study explores the impact of incorporating moving averages into both Random Forest algorithm and LSTM models when trained on combined datasets. The results indicate that integrating moving averages enhances the predictive capabilities of both algorithms. Notably, the LSTM model utilizing moving averages over the combined dataset exhibits the highest level of accuracy in forecasting future stock prices. In conclusion, this research underscores the efficacy of LSTM, particularly when coupled with moving averages, in predicting stock market prices. The findings suggest that the advanced LSTM model holds significant promise for broader applications beyond stock market forecasting, showcasing its potential as a robust tool for accurate and reliable price predictions in various financial domains.

**INTRODUCTION**

The introduction to this research on market price prediction using Long Short-Term Memory (LSTM) in conjunction with Random Forest algorithm delves deeper into the intricacies of stock market forecasting. The volatile and unpredictable nature of financial markets poses a significant challenge for investors and analysts seeking to make informed decisions.[(Brownlee 2018)](https://paperpile.com/c/53rYV5/7B1a) Traditional methods of stock price prediction often fall short in capturing the complex patterns and trends inherent in market data. In response to these challenges, this study explores the application of advanced deep learning techniques, specifically LSTM, renowned for its ability to model long-term dependencies in sequential data. By combining LSTM with Random Forest algorithm, a powerful machine learning algorithm known for its effectiveness in classification and regression tasks, the research aims to enhance the accuracy and reliability of stock price predictions.

Drawing on historical data from the Dow Jones Industrial Average (DJIA) and incorporating external factors such as crude oil and gold prices, this study seeks to create robust predictive models that can adapt to the dynamic nature of financial markets[(Brownlee 2017)](https://paperpile.com/c/53rYV5/xN5b). The introduction underscores the importance of feature engineering and data preprocessing techniques in extracting meaningful insights from raw data, thereby improving the performance of predictive models. Additionally, the integration of technical indicators and moving averages further enriches the predictive capabilities of the LSTM-Random Forest hybrid model.

By bridging the gap between financial analysis and machine learning, this research not only aims to advance the field of stock market forecasting but also to provide valuable tools for investors, traders, and financial institutions seeking to navigate the complexities of modern financial markets. Through a comprehensive analysis of LSTM and Random Forest algorithm models in predicting market prices, this study sets out to contribute to the development of more accurate and efficient forecasting methodologies that can empower stakeholders with actionable insights for better decision-making in the realm of finance.

MATERIAL REQURIED

The materials required for this research on market price prediction using Long Short-Term Memory (LSTM) in combination with Random Forest algorithm include historical stock price data from the Dow Jones Industrial Average (DJIA), external factors like crude oil and gold prices, and technical indicators. [(Brownlee 2017)](https://paperpile.com/c/53rYV5/xN5b)Additionally, feature engineering techniques will be employed to extract relevant patterns from the data, and moving averages will be integrated into both LSTM and Random Forest algorithm models to enhance predictive performance. The study will also utilize various performance evaluation metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) to assess model accuracy. By combining these materials and methodologies, the research aims to develop robust predictive models for stock market price forecasting.

**MATERIAL AND METHOD**

The introduction to this research on market price prediction using Long Short-Term Memory (LSTM) in conjunction with Random Forest algorithm delves deeper into the intricacies of stock market forecasting. The volatile and unpredictable nature of financial markets poses a significant challenge for investors and analysts seeking to make informed decisions. Traditional methods of stock price prediction often fall short in capturing the complex patterns and trends inherent in market data. In response to these challenges, this study explores the application of advanced deep learning techniques, specifically LSTM, renowned for its ability to model long-term dependencies in sequential data. By combining LSTM with Random Forest algorithm, a powerful machine learning algorithm known for its effectiveness in classification and regression tasks, the research aims to enhance the accuracy and reliability of stock price predictions.

Drawing on historical data from the Dow Jones Industrial Average (DJIA) and incorporating external factors such as crude oil and gold prices, this study seeks to create robust predictive models that can adapt to the dynamic nature of financial markets. The introduction underscores the importance of feature engineering and data preprocessing techniques in extracting meaningful insights from raw data, thereby improving the performance of predictive models. Additionally, the integration of technical indicators and moving averages further enriches the predictive capabilities of the LSTM-Random Forest hybrid model.

By bridging the gap between financial analysis and machine learning, this research not only aims to advance the field of stock market forecasting but also to provide valuable tools for investors, traders, and financial institutions seeking to navigate the complexities of modern financial markets. Through a comprehensive analysis of LSTM and Random Forest algorithm models in predicting market prices, this study sets out to contribute to the development of more accurate and efficient forecasting methodologies that can empower stakeholders with actionable insights for better decision-making in the realm of finance.

**RELATED WORK**

The study published in the Journal of Big Data explores the implementation of Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) on grouped time-series data to predict stock prices. The research introduces eight new architectural models for stock price forecasting by identifying joint movement patterns in the stock market.[(Thampi et al. 2021)](https://paperpile.com/c/53rYV5/OocF) By combining LSTM and GRU models with four neural network block architectures, the proposed models are evaluated using accuracy measures like Mean Absolute Percentage Error (MAPE), Root Mean Squared Percentage Error (RMSPE), and Rooted Mean Dimensional Percentage Error (RMDPE). The study emphasizes the importance of accurate data modeling for forecasting stock prices with low error rates, highlighting the potential of deep learning algorithms like LSTM and GRU in predicting stock price movements effectively. Another research article focuses on predicting stock market indices using Long Short-Term Memory (LSTM) neural network architecture.[(Tripathi and Upadhyaya 2021)](https://paperpile.com/c/53rYV5/SxbH) The study utilizes LSTM to forecast the next-day closing price of the S&P 500 index, showcasing the application of advanced deep learning techniques in financial forecasting. By leveraging LSTM, a specific neural network architecture known for its ability to capture long-term dependencies in sequential data, the research aims to enhance predictive accuracy in predicting stock market indices.

**LSTM Algorithm**

Long Short-Term Memory (LSTM) is a specialized type of recurrent neural network (RNN) designed to address the vanishing gradient problem commonly encountered in traditional RNNs.[(Thach, Kreinovich, and Trung 2020)](https://paperpile.com/c/53rYV5/Y0YN) LSTMs excel in capturing long-term dependencies in sequential data, making them ideal for tasks like language translation, speech recognition, and time series forecasting.

**Key Points:**

Memory Cell: LSTMs introduce a memory cell that can retain information over extended periods, allowing the network to learn and remember crucial patterns in data sequences.

Gating Mechanism: LSTMs incorporate three gates - the input gate, the forget gate, and the output gate - which regulate the flow of information within the network. These gates control what information is added to, removed from, and output from the memory cell.

Bidirectional LSTM: Bidirectional LSTMs process sequential data in both forward and backward directions, enabling them to capture longer-range dependencies compared to traditional LSTMs.

Applications: LSTMs find applications in various fields such as language simulation, voice recognition, sentiment analysis, time series prediction, video analysis, and handwriting recognition.

**Structure of LSTM:**

An LSTM network comprises a series of LSTM cells, each equipped with input, output, and forget gates that manage information flow. The gates selectively retain or discard information from previous time steps to maintain long-term dependencies in the input data. The memory cell within each LSTM cell stores past information influencing the cell's output and subsequent processing of sequential data.

Applications of LSTM:

Language Simulation: Used for tasks like machine translation, language modeling, and text summarization.

Voice Recognition: Applied in speech-to-text transcription and command recognition.

Sentiment Analysis: Utilized for classifying text sentiment as positive, negative, or neutral.

Time Series Prediction: Employed to forecast future values based on historical data patterns.

Video Analysis: Used for tasks like object detection, activity recognition, and action classification.

Handwriting Recognition: Applied in recognizing handwritten characters or text.

LSTMs stand out as powerful tools in deep learning due to their ability to capture long-term dependencies effectively across various applications. Their structured architecture and gating mechanisms make them versatile for processing sequential data with intricate patterns.

**RANDOM FOREST ALGORITHM :**

Random Forest is an ensemble learning method used for classification and regression tasks. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Key Points:

Ensemble Learning: Random Forest operates by constructing multiple decision trees during training and outputting the mode (classification) or mean prediction (regression) of the individual trees.

Decision Trees: Each decision tree in a Random Forest is trained on a subset of the training data and a random subset of features, reducing overfitting and improving generalization.

Bootstrapping: Random Forest uses bootstrapping to generate multiple training datasets from the original data, ensuring diversity among the trees in the ensemble.

Feature Importance: Random Forest calculates feature importance by measuring the decrease in impurity (e.g., Gini impurity) when a particular feature is used for splitting data points.

Applications: Random Forest is widely used in various domains, including finance, healthcare, marketing, and bioinformatics, for tasks like credit scoring, disease diagnosis, customer segmentation, and gene expression analysis.

**Structure of Random Forest:**

A Random Forest comprises multiple decision trees, each trained on a subset of the training data and a random subset of features. During prediction, the class (classification) or mean prediction (regression) of the individual trees is aggregated to obtain the final output.

**Applications of Random Forest:**

Finance: Used for credit scoring, risk management, and stock market prediction.

Healthcare: Applied in disease diagnosis, patient outcome prediction, and drug discovery.

Marketing: Utilized for customer segmentation, churn prediction, and campaign optimization.

Bioinformatics: Used in gene expression analysis, protein structure prediction, and drug-target interaction prediction.

Random Forest algorithms are powerful and versatile machine learning techniques capable of handling both classification and regression tasks across diverse domains. Their ability to reduce overfitting, handle high-dimensional data, and provide robust predictions makes them valuable tools for various applications.

**STATISTICAL ANALYSIS:**

The statistical analysis for stock price prediction using Long Short-Term Memory (LSTM) in conjunction with Random Forest algorithm involves a meticulous process encompassing data collection, preprocessing, feature engineering, model development, training, testing, performance evaluation, and result interpretation. Initially, historical stock price data from the Dow Jones Industrial Average (DJIA) is collected along with external factors like crude oil and gold prices to enrich the dataset.[(Venkataraman et al. 2023)](https://paperpile.com/c/53rYV5/eFik) Subsequently, feature engineering techniques are applied to extract meaningful patterns from the data, followed by the creation of LSTM and Random Forest algorithm models tailored to capture temporal dependencies and optimize predictive accuracy. Through rigorous training and testing procedures, the models are evaluated based on performance metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to gauge their predictive capabilities. The analysis culminates in a comprehensive discussion of the results obtained from comparing the LSTM and Random Forest algorithm models in predicting stock prices, shedding light on their respective strengths and areas for improvement[(Venkataraman et al. 2023; Patterson and Gibson 2017)](https://paperpile.com/c/53rYV5/eFik+mePC). By leveraging these statistical methodologies, researchers can derive valuable insights into the efficacy of LSTM and Random Forest algorithm for stock market forecasting, paving the way for enhanced predictive modeling techniques in financial analysis.

**RESULT**

The research on market price prediction using Long Short-Term Memory (LSTM) in conjunction with Random Forest algorithm has culminated in significant findings and outcomes. Through the application of LSTM models, the study has successfully captured long-term dependencies in sequential data, enhancing the accuracy and efficiency of stock price forecasts. Concurrently, Random Forest algorithm has proven instrumental in predicting stock market indices and analyzing complex stock price movements, showcasing its efficacy in handling high-dimensional data and nonlinear classification challenges. The research findings have underscored the effectiveness of LSTM and Random Forest algorithm models in capturing intricate patterns and relationships within financial data, highlighting their strengths in forecasting stock market trends with precision.

**DISCUSSION**

The discussion section will likely delve into the findings and implications of the study, analyzing the performance of LSTM and Random Forest algorithm models in predicting stock prices and comparing their accuracy using statistical metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Additionally, the discussion may explore the impact of incorporating external factors and moving averages on predictive outcomes. By engaging in a detailed discussion, researchers can elucidate the strengths and limitations of the models, provide insights into the effectiveness of deep learning algorithms in financial forecasting, and suggest avenues for future research to enhance predictive capabilities further.

**CONCLUSION**

In the conclusion of the research on market price prediction using Long Short-Term Memory (LSTM) in conjunction with Random Forest algorithm, a comprehensive summary of the findings and implications of the study will be presented. The conclusion will likely highlight the performance of LSTM and Random Forest algorithm models in forecasting stock prices, emphasizing the superiority of LSTM in capturing long-term dependencies and providing more accurate predictions compared to Random Forest algorithm. It may discuss the significance of incorporating external factors such as crude oil and gold prices, as well as the impact of integrating moving averages on predictive accuracy. The conclusion is expected to underscore the potential applications of advanced deep learning techniques in financial forecasting and offer recommendations for further research to enhance model performance and address any limitations identified during the study. By encapsulating key insights and implications, the conclusion aims to provide a clear and concise summary of the research outcomes and their implications for future developments in stock market prediction methodologies.

**DECLARTION**

The declaration for market price prediction using Long Short-Term Memory (LSTM) in combination with Random Forest algorithm involves a commitment to leveraging advanced deep learning techniques to enhance the accuracy and efficiency of stock market forecasting. This declaration signifies a dedication to harnessing the power of LSTM's ability to capture long-term dependencies in data and Random Forest algorithm's robust classification capabilities to develop predictive models that can navigate the complexities of financial markets. It underscores a commitment to incorporating external factors, employing feature engineering techniques, and utilizing performance evaluation metrics to optimize predictive accuracy. This declaration signifies a pursuit of innovation in financial analysis, aiming to provide stakeholders with reliable tools for making informed decisions in the dynamic realm of stock market investments.

**Conflicts of Interest:**

The authors of this paper declare no conflict of interest.

**Author Contributions:**

The development and implementation of the Random Forest algorithm, as well as data collection and data analysis, were all performed by author GRK. The conceptualization of the study, oversight of the research process, and a critical assessment of the manuscript were all carried out by author VGK.

**ACKNOWLEDGMENT :**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

REFERENCE

[Bosco, Joish, and Fateh Khan. 2018. *Stock Market Prediction and Efficiency Analysis Using Recurrent Neural Network*. GRIN Verlag.](http://paperpile.com/b/53rYV5/3AhH)

[Brownlee, Jason. 2017. *Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future*. Machine Learning Mastery.](http://paperpile.com/b/53rYV5/xN5b)

[———. 2018. *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.](http://paperpile.com/b/53rYV5/7B1a)

[Jabed, Mahfuz Islam Khan. 2024. *Stock Market Price Prediction Using Machine Learning Techniques*. Ocleno.](http://paperpile.com/b/53rYV5/xNC3)

[Patterson, Josh, and Adam Gibson. 2017. *Deep Learning: A Practitioner’s Approach*. “O’Reilly Media, Inc.”](http://paperpile.com/b/53rYV5/mePC)

[Thach, Nguyen Ngoc, Vladik Kreinovich, and Nguyen Duc Trung. 2020. *Data Science for Financial Econometrics*. Springer Nature.](http://paperpile.com/b/53rYV5/Y0YN)

[Thampi, Sabu M., Jaime Lloret Mauri, Xavier Fernando, Rajendra Boppana, S. Geetha, and Axel Sikora. 2021. *Applied Soft Computing and Communication Networks: Proceedings of ACN 2020*. Springer Nature.](http://paperpile.com/b/53rYV5/OocF)

[Tripathi, Meenakshi, and Sushant Upadhyaya. 2021. *Conference Proceedings of ICDLAIR2019*. Springer Nature.](http://paperpile.com/b/53rYV5/SxbH)

[Venkataraman, Neelanarayanan, Lipo Wang, Xavier Fernando, and Ahmed F. Zobaa. 2023. *Big Data and Cloud Computing: Select Proceedings of ICBCC 2022*. Springer Nature.](http://paperpile.com/b/53rYV5/eFik)

**Table and Figures:**

**SPSS STATISTICAL ANALYSIS**



