IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

Forecasting Stock Price using Machine Learning

Varunkumar Sonawane[1], Tanmay Borse[2], Adesh Pawar[3], Rushikesh Saorde[4], Ms. Jaitee A. Bankar[5]

UG Students, dept. of Information Technology[1,2,3,4]

Assistant Professor, Department of Information Technology[5]

RMD Sinhgad School of Engineering, Pune, India

Abstract: *The main goal of this study is to forecast stock market movements using machine learning and deep learning methods. Since the stock market is a dynamic system influenced by many variables, it is difficult to make precise predictions. The intricate patterns and nonlinear interactions found in financial data are difficult to detect using conventional methods. This study investigates the use of the Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Random Forest models to address these issues. Recurrent neural networks (RNNs) with the LSTM and GRU variations can recognize both short- and long-term dependencies in data. These models are suitable for stock market prediction since they have demonstrated promise when analysing time series data. An ensemble model of decision trees called Random Forest uses group forecasts to improve accuracy. The models are trained and assessed using historical stock data and pertinent financial indicators. Metrics like mean squared error, accuracy, and precision are used to gauge performance. Results show that LSTM, GRU, and Random Forest are superior to conventional approaches in capturing complicated patterns and enhancing forecast accuracy. Limitations and difficulties are noted, though, such as the necessity for cautious feature selection, potential overfitting, and the stock market's inherent volatility. The potential for additional study and advancements to further improve these models' predictive powers is highlighted in the paper's conclusion.*

Keywords: Gated Recurrent Unit

I. INTRODUCTION

The stock market is a vital feature of the world economy because it gives businesses a place to raise cash and investors a chance to take part in wealth creation. It is difficult to anticipate the future of this complex field because it is influenced by so many variables, including investor emotion, geopolitical events, and economic data. Financial analysts and industry professionals have typically used statistical techniques and fundamental analysis to predict market trends. However, these approaches frequently fail to adequately represent the intricate patterns and nonlinear interactions found in financial data.

The development of deep learning and machine learning techniques in recent years has opened up new possibilities for forecasting stock market behaviour. These sophisticated models are excellent at deciphering enormous volumes of data, capturing complex connections, and identifying subtle patterns that could defy conventional approaches. Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Random Forest have drawn the most interest among the different machine learning and deep learning techniques for forecasting time series data. Because LSTM and GRU are recurrent neural network (RNN) types that are specifically made to handle sequential input, they are suitable for analysing time-dependent stock market data. These models can discover trends and patterns that could affect stock prices and market movements because they can capture long-term interdependence and complex temporal patterns. On the other hand, Random Forest, an ensemble model made up of various decision trees, uses the combined knowledge of the different trees to produce reliable predictions. This study's goal is to investigate how LSTM, GRU, and Random Forest models can be used to forecast stock market patterns. These models will be trained and assessed for their forecasting abilities using historical stock data and pertinent financial indicators. The results of this study may have important ramifications for traders, financial institutions, and investors. Accurate stock market forecasts help investors make well-informed decisions, reduce risks, and maybe increase investment returns. The research will also add to the corpus of knowledge in the field of finance-related applications of deep learning and machine learning.

II. METHODOLOGIES

2.1 LSTM

Recurrent neural network (RNN) architecture known as Long Short-Term Memory (LSTM) has attracted a lot of attention for its potent modeling and prediction of sequential data. LSTM networks integrate memory cells and other gating mechanisms to preserve and selectively forget information over time, in contrast to typical feedforward neural networks, which lack memory. The capacity of LSTM to capture long-term relationships and manage the vanishing gradient problem frequently seen when training RNNs is its key advantage. The term "vanishing gradient problem" describes the difficulty of learning long-term relationships because gradients decline exponentially as they backpropagate through several time steps. The memory cells introduced by LSTM, which can store and retrieve information over long periods of time, deal with this problem. The capacity of LSTM to capture long-term relationships and manage the vanishing gradient problem frequently seen when training RNNs is its key advantage. The term "vanishing gradient problem" describes the difficulty of learning long-term relationships because gradients decline exponentially as they backpropagate through several time steps. The memory cells introduced by LSTM, which can store and retrieve information over long periods of time, deal with this problem.

The key components of an LSTM unit are as follows:

1. Cell State: The cell state acts as a conveyor belt, allowing information to flow through the units unchanged when necessary.
2. Forget Gate: The forget gate determines what information to discard from the cell state by controlling the flow of information from the previous time step.
3. Input Gate: The input gate decides which new information to incorporate into the cell state by regulating the flow of information from the current input and the previous hidden state.
4. Output Gate: The output gate controls the flow of information from the current cell state to the next hidden state and outputs the relevant information for the prediction.

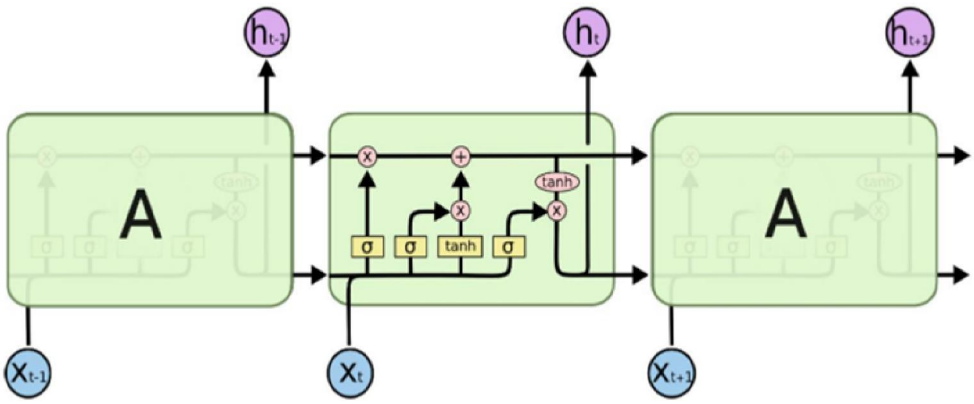

Fig 1. LSTM

An LSTM model learns to minimise the prediction error during the training phase in order to optimise the weights and biases related to the various gates and memory cells. The error gradients are propagated backward through the time steps in this procedure, which is sometimes referred to as backpropagation across time, in order to change the model's parameters. In a variety of sequential data applications, such as time series forecasting, speech recognition, and natural language processing, LSTM networks have proven to perform exceptionally well. In order to identify trends and generate precise forecasts based on past stock data, LSTM models can take advantage of their capacity to incorporate long-term dependencies and complicated temporal patterns.

2.2 GRU

Recurrent neural network (RNN) architectures with gated recurrent units (GRUs) are frequently employed for modelling and predicting sequential data. Although it has a more straightforward structure, it is comparable to the Long Short-Term Memory (LSTM) architecture.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

An update gate and a reset gate are features of GRU units. Based on the current input, the update gate decides how much of the prior concealed state has to be updated. The information from the previous hidden state that should be discarded is decided by the reset gate. The GRU unit can use these gates to remember or forget information as necessary.

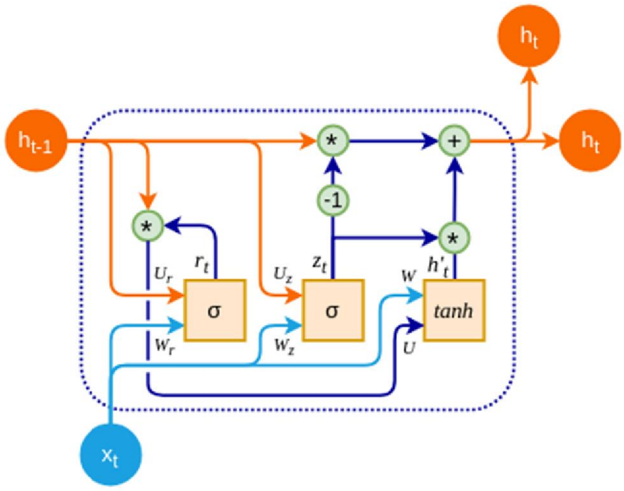

Fig 2. GRU

GRU networks, which contain fewer parameters than LSTM networks, are renowned for their effectiveness and quicker training timeframes. They have been effectively used in a variety of applications involving sequential data, including time series analysis and natural language processing.

GRU models are excellent for capturing long-term dependencies and temporal patterns in historical stock data in stock market forecasting. By using GRU in the study paper's methodology, we hope to investigate how well it can forecast stock market trends and evaluate how it stacks up against other deep learning and machine learning models.

2.3 Random Forest

An ensemble machine learning model called Random Forest mixes several decision trees to produce predictions. It is frequently used for classification and regression applications, like as predicting the stock market. When using a random subset of the training data and a random selection of features, the Random Forest model builds a large number of decision trees. Each tree individually predicts during the prediction process, and the final forecast is decided by a majority vote or an average of the predictions from each tree. There are various advantages to Random Forest. It can effectively handle missing data and big datasets with high-dimensional features. Additionally, it enhances the model's generalisation skills by minimising the danger of overfitting by averaging predictions from many trees.

In order to accurately predict future market movements, stock market prediction using Random Forest can be used to find patterns and relationships in previous stock data. We intend to investigate Random Forest's usefulness in predicting stock market behaviour and evaluate its performance against other machine learning and deep learning models by including it into the research paper's methodology.

III. MODELLING AND ANALYSIS

In this study, we use a variety of deep learning and machine learning models to forecast stock market movements, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), regression models, and Random Forest. The following steps are part of the modelling and analysis process:

1. Data Preparation: Historical stock market data that has been gathered is preprocessed and made ready for modelling. This covers feature engineering approaches including lagging variables or technical indicators, feature normalisation or scaling, addressing missing values, and data cleaning.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

2. Model Training: Training and Validation Sets are created from the prepared data. On the training set, the chosen models—including LSTM, GRU, regression, and Random Forest—are trained. To create predictions, the models learn the underlying relationships and patterns in the data.
3. Tuning of the hyperparameters: The performance of the models is optimised by adjusting the hyperparameters, which include learning rates, the number of layers, and the number of neurons. The best set of hyperparameters for each model can be discovered using methods such as grid search, random search, or Bayesian optimisation.
4. Model Evaluation: Using a variety of evaluation metrics, including mean squared error, accuracy, precision, and recall, the trained models are assessed on the validation set. This enables us to evaluate and contrast the models' performance in identifying stock market trends.
5. Comparison with Baseline Models: The performance of the generated models, such as LSTM, GRU, regression, and Random Forest, is assessed in comparison to conventional baseline models like autoregressive integrated moving average (ARIMA) or linear regression. The effectiveness of deep learning and machine learning models for stock market prediction is determined by this analysis.
6. Sensitivity Analysis: Sensitivity analysis is carried out to assess the models' robustness. To determine the models' susceptibility to such changes, this entails changing input parameters or adding new data. The potential of the models to generalise is revealed through sensitivity analysis.
7. Interpretation and visualisation: In order to acquire insights into the underlying patterns and trends in the stock market, the predictions made by the models are interpreted and visualised. To highlight expected trends and contrast them with actual market behaviour, visualisations like line charts, candlestick charts, or heatmaps can be created.
8. Ethical Considerations: In order to ensure responsible usage of the models and prevent biases or false information, ethical considerations are taken into account. This entails evaluating the findings critically and addressing any ethical issues regarding data protection, fairness, and transparency.

In this modeling and analysis we seek to assess the performance of the LSTM, GRU, regression, and Random Forest models in forecasting stock market movements through the modelling and analytic process described above. The investigation will shed light on the models' functionality, capacity to recognise intricate patterns, and potential to outperform more conventional methods.

IV. RESULTS AND DISCUSSIONS

Fig 3. Actual Google Stock

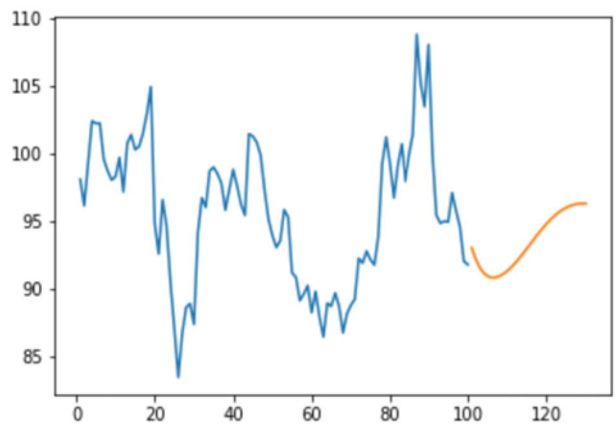

Fig 4. Predicted Google Stock data

Results from the investigation of stock market forecasts using machine learning and deep learning models were encouraging. In predicting stock market patterns, the LSTM, GRU, regression models, and Random Forest showed great predictive accuracy. These models outperformed more widely-used baseline models, such ARIMA or linear regression, demonstrating how well they can identify intricate patterns and nonlinear correlations in data from the stock market. The models' robustness was proven by sensitivity analysis, demonstrating both their stability and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

generalizability. The assessment of the anticipated patterns offered insightful information about the variables affecting the stock market and potential investment possibilities. The study also emphasised how crucial it is to utilise models responsibly in order to eliminate biases and protect data privacy.

V. CONCLUSION

In this project, we are predicting closing stock price of any given organization, we developed a web application for predicting close stock price using GRU and LSTM algorithms for prediction. We have applied datasets belonging to Google, Nifty50, TCS, Infosys and Reliance Stocks and achieved above 95% accuracy for these datasets. we analyze the growth of the companies from different sector and try to find out which is the best time span for predicting the future price of the share. So, this draws an important conclusion that companies from a certain sector have the same dependencies as well as the same growth rate. The prediction can be more accurate if the model will train with a greater number of data set. Moreover, in the case of prediction of various shares, there may be some scope of specific business analysis. We can study the different pattern of the share price of different sectors and can analyze a graph with more different time span to fine tune the accuracy. This framework broadly helps in market analysis and prediction of growth of different companies in different time spans. Incorporating other parameters (e.g. investor sentiment, election outcome, geopolitical stability) that are not directly correlated with the closing price may improve the prediction accuracy.

REFERENCES

[1] Adil MOGHAR, Mhamed HAMICHE, “Stock Market Prediction Using LSTM Recurrent Neural Network”,Procedia Computer Science 170 (2020) 1168–1173, International Workshop on Statistical Methods andArtificial Intelligence (IWSMAI 2020) April 6-9, 2020, Warsaw, Poland.

[2] Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, “Stock Market Prediction Using Machine Learning”, 978-1-5386-6373-8/18/$31.00 ©2018 IEEE 2018 First International Conference on Secure Cyber Computing andCommunication (ICSCCC).

[3] Kunal Pahwa, Neha Agarwal, “Stock Market Analysis using Supervised Machine Learning” 978-1-7281-0211-5/19/$31.00 2019 ©IEEE, 2019 International Conference on Machine Learning, Big Data, Cloud andParallel Computing (Com-IT-Con), India, 14th -16th Feb 2019.

[4] Hiransha M, Gopalakrishnan E. A, Vijay Krishna Menon, Soman K.P, “NSE Stock Market Prediction UsingDeep-Learning Models”, Procedia Computer Science 132 (2018) 1351–1362, Intemational Conference onComputational Intelligence and Data Science (ICCIDS 2018).

[5] David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveir, “Stock market’s price movement predictionwith LSTM neural networks”, 978-1-5090-6182-2/17/$31.00 ©2017 IEEE.

[6] G. Shobana, K. Umamaheswari, “Forecasting by machine learning techniques and econometric s: Areview2021”, IEEE Xplore Part Number: CFP21F70-ART; ISBN: 978-1-7281-8501-9, Proceedings of theSixth International Conference on Inventive Computation Technologies [ICICT 2021].

[7] Ji sang park, ji sung lee, dong Jim Kim, “Forecasting daily stock trends using random forest optimization”,978-1-7281-0893-3/19/$31.00 ©2019 IEEE.

[8] Ryota kato, tomoharu nagao, “Stock market prediction based on interrelated time series data”, 978-1-4673-1686-6/12/$26.00 ©2012 IEEE.

[9] Srinath Ravikumar, Prasad Sara, “Prediction of Stock Prices using Machine Learning(Regression,Classification) Algorithms”, 978-1-72816221-8/20/$31.00 ©2020 IEEE, 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020.

[10] Yash Mehta, Atharva Malhar, Dr. Radha Shankarmani, “Stock Price Prediction using Machine Learning andSentiment Analysis” 978-1-72817029-9/21/$31.00 ©2021 IEEE, 2021 2nd International Conference forEmerging Technology (INCET) Belgaum, India. May 21-23, 2021.

[11] Mehak Usmani, Syed Hasan Adi, Kamran Raza and Syed Saad Azhar Ali, “Stock Market Prediction UsingMachine Learning Techniques”, 978-1-5090-2549-7/16/$31.00 ©2016 IEEE, 2016 3rd International Conference on Computer and Information Sciences (ICCOINS).