International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Stock Market Prediction Using LSTM and ARIMA

S. Phani Harshitha

B. Tech Student, Department of IT, GMR Institute of Technology, Rajam-532127, Andhra Pradesh, India
Email: 22345A1208@gmrit.edu.in

ABSTRACT

The stock market is a platform where the buying and selling of shares (also known as stocks) takes place. It provides a marketplace where investors can trade shares of publicly listed companies. The stock exchange acts as a mediator and facilitator for these transaction. The Bombay Stock Exchange (BSE), founded in 1875, holds the distinction of being Asia's first stock exchange. It is located in Mumbai (formerly known as Bombay), India The LSTM model, a type of recurrent neural network, excels at capturing complex temporal dependencies and patterns in historical stock price data. In contrast, the ARIMA model is well-suited for modeling short-term trends and seasonality in financial time series. By integrating these two models, we aim to harness the complementary capabilities of each approach. The methodology begins with preprocessing the historical stock price data, including data sorting, feature engineering, and autocorrelation checks. Subsequently, we employ both LSTM and ARIMA models independently to forecast stock prices. These individual forecasts are then combined using a weighted ensemble approach, allowing us to leverage the strengths of both models.

Keywords*: Support Vector Machine, Random Forest, Long Short-Term Memory, Bombay Stock Exchange, National Stock Exchange, ARIMA*

INTRODUCTION

Predicting stock market movements involves employing sophisticated methodologies, with two prominent approaches being AutoRegressive Integrated Moving Average (ARIMA) models and Long Short-Term Memory (LSTM) networks. ARIMA, a time series forecasting model denoted as ARIMA(p, d, q), utilizes autoregression, differencing, and moving averages to capture linear relationships in historical data. It excels at short-term predictions, making it effective for immediate trends in stock prices.

In contrast, LSTM, a type of recurrent neural network, is adept at capturing intricate patterns and long-term dependencies in time series data. Equipped with memory cells and gates, LSTMs selectively remember or forget information over extended sequences, making them well-suited for handling non-linear relationships and irregular patterns in stock market data. Combining ARIMA and LSTM models has become a prevalent strategy. ARIMA is typically employed for short-term predictions, while LSTM captures long-term trends. The predictions from both models are intelligently combined, often with assigned weights to enhance overall forecasting accuracy.

However, caution is paramount in stock market predictions. Past performance doesn't guarantee future results, and financial markets are influenced by unpredictable factors. Metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) evaluate the combined model's performance. Additionally, risk management strategies are crucial to mitigate potential losses given the inherent uncertainty. Investors should view predictions as tools for informed decision-making rather than certainties.

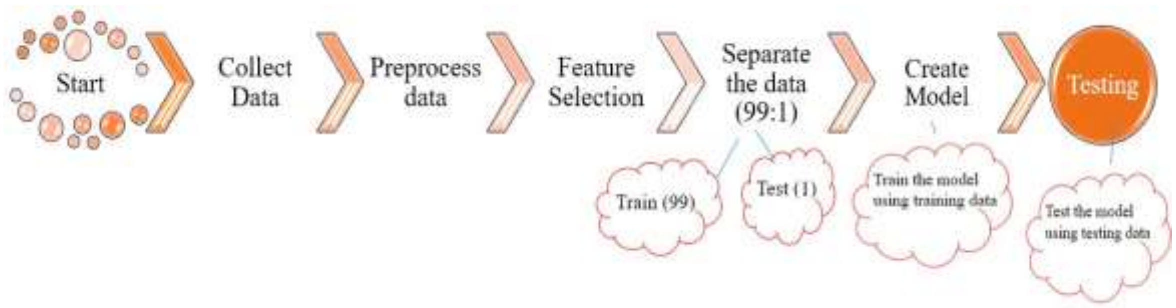
RESEARCH APPROACH

Conducting research on stock market prediction through the amalgamation of AutoRegressive Integrated Moving Average (ARIMA) models and Long Short-Term Memory (LSTM) networks necessitates a systematic and well-defined approach. Firstly, researchers should articulate clear and specific objectives, outlining whether the aim is to enhance prediction accuracy, evaluate the effectiveness of the combined model, or explore the impact of different weighting strategies. This initial step provides a roadmap for the subsequent phases of the research.

A thorough literature review constitutes the second critical stage, wherein researchers delve into existing literature on stock market prediction, ARIMA models, LSTM networks, and the integration of multiple models. This review not only informs the research design but also identifies gaps and opportunities for contribution to the field. Subsequently, the data collection process involves gathering historical stock market data, ensuring its cleanliness, accuracy, and relevance to the research objectives. Financial databases, APIs, and reliable sources play a pivotal role in this phase. The subsequent stage involves defining evaluation metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and accuracy, to gauge the performance of the combined model. Comparative analysis against individual ARIMA and LSTM models provides insights into the efficacy of the integrated approach. Results are then interpreted, and conclusions are drawn based on the performance metrics, with a discussion of the strengths and limitations of the combined model. Researchers should also explore implications for investors, financial analysts, and potential applications in practical contexts.

In the conclusion, the research findings are summarized, and avenues for future work are proposed, addressing any limitations and suggesting areas that warrant further investigation. This structured approach ensures a rigorous and comprehensive exploration of the integration of ARIMA and LSTM models for stock market prediction, contributing to the ongoing evolution of predictive analytics in financial markets.

METHODOLOGY:



The provided image illustrates the Data Modeling Methodology, which is a systematic approach to creating data models that accurately represent real-world concepts and relationships. It involves several steps that ensure the model's effectiveness and consistency with the underlying data.

1. Define the Purpose of the Model

Before embarking on the modeling process, it's crucial to clearly define the purpose of the model. What problem does it aim to solve? What information should it capture? Who will use the model and for what purposes? Understanding the model's intended use will guide the subsequent steps.

2. Collect and Assess the Data

The quality and relevance of the data are paramount to building an accurate model. Gather the necessary data from various sources, ensure its completeness and consistency, and assess its suitability for the intended model. Identify any data quality issues that need to be addressed before proceeding.

3. Data Preprocessing

Data preprocessing involves preparing the collected data for modeling by cleaning, transforming, and normalizing it. This may involve handling missing values, outliers, and inconsistencies, transforming data types, and scaling numerical values. The goal is to ensure the data is in a suitable format for modeling.

4. Feature Selection

Feature selection involves identifying the most relevant and informative features from the dataset to be incorporated into the model. This helps reduce dimensionality and improve model performance. Various techniques, such as statistical measures, correlation analysis, and machine learning algorithms, can be employed for feature selection.

5. Model Selection and Training

Choose an appropriate machine learning algorithm based on the nature of the problem and the characteristics of the data. Train the chosen model on the prepared and preprocessed data, optimizing its parameters to minimize error and maximize accuracy.

6. Model Evaluation

Evaluate the trained model's performance on unseen test data to assess itsgeneralizability and predictive power. Common evaluation metrics include accuracy, precision, recall, and F1 score. Evaluate the model's performance against different performance metrics to gain a comprehensive understanding of its strengths and weaknesses.

7. Model Improvement and Refinement

Based on the evaluation results, identify areas for improvement and refine the model. This may involve adjusting hyperparameters, modifying feature selection, or exploring different modeling techniques. Continuously evaluate and refine the model to optimize its performance and ensure its effectiveness.

8. Model Deployment and Monitoring

Deploy the final model into a production environment, integrating it into the application or system where it will be used. Monitor the model's performance over time, identifying any

RESULTS

Algorithm	**MAE**	**MSE**	**RMSE**	**R Squared**
Linear Regression	2.6312	11.7550	3.4286	0.6573
Polynomial Regression	2.0037	7.9467	2.8190	0.7134
Random Forest	2.3053	9.5669	3.0930	0.6705
Gradient Boosting	**1.9642**	**7.2011**	**2.6835**	**0.7485**
SVM	2.4373	10.6333	3.2609	0.3458
Ridge Regression	2.6323	11.7500	3.4278	0.4971
Lasso Regression	3.5850	20.1185	4.4854	−2.9327
Elastic Net Regression	3.6595	20.9698	4.5793	−4.0050

CONCLUSION

In summary, this study introduces a novel hybrid approach for stock price prediction, harnessing the strengths of both Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) models. The focus on the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE) involves thorough preprocessing of historical stock prices and financial data, including sorting, feature engineering, and autocorrelation checks. LSTM, known for capturing intricate temporal dependencies, and ARIMA, well-suited for short-term trend modeling, are utilized independently for stock price forecasting. The integration of these models is accomplished through a weighted ensemble method, leveraging the complementary abilities of LSTM and ARIMA. The ultimate goal of this hybrid model is to improve prediction accuracy and robustness in the dynamic landscape of financial markets. It is essential, however, to recognize the inherent complexities and uncertainties associated with stock price prediction. While the model provides valuable insights for investors and analysts, its application should be approached with caution given the unpredictable nature of market dynamics. Future research directions may involve refining the hybrid model, incorporating additional features, and exploring alternative machine learning algorithms. Continuous validation against real-world market data remains crucial for evaluating the model's performance across various market conditions.

References

1. Lakshminarayanan, S. K., & McCrae, J. P. (2019, December). A Comparative Study of SVM and LSTM Deep Learning Algorithms for Stock Market Prediction. In *AICS* (pp. 446-457).
2. Suhaime, S. (2021). Stock Price Prediction Analysis Dashboard using Machine Learning with LSTM Neural Network.
3. Hasan, M. M., Roy, P., Sarkar, S., & Khan, M. M. (2021, January). Stock market predictionweb service using deep learning by lstm. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0180-0183). IEEE.
4. Zou, Z., & Qu, Z. (2020). Using lstm in stock prediction and quantitative trading. *CS230: Deep Learning, Winter*, 1-6.
5. Jamil, H. (2022). *Inflation forecasting using hybrid ARIMA-LSTM model* (Doctoral dissertation, Laurentian University of Sudbury).
6. Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, *9*, 100320.
7. Long, B., Tan, F., & Newman, M. (2023). Forecasting the Monkeypox Outbreak Using ARIMA, Prophet, NeuralProphet, and LSTM Models in the United States. *Forecasting*, *5*(1), 127-137.
8. Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016, August). Stock market prediction using machine learning techniques. In *2016 3rd international conference on computer and information sciences (ICCOINS)* (pp. 322-327). IEEE.
9. Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018, December). Stock market prediction using machine learning. In *2018 first international conference on secure cyber computing and communication (ICSCCC)* (pp. 574-576). IEEE.
10. Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, *167*, 599-606.
11. Misra, M., Yadav, A. P., & Kaur, H. (2018, July). Stock market prediction using machine learning algorithms: a classification study. In *2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering*
12. Kompella, S., & Chakravarthy Chilukuri, K. C. C. (2020). Stock market prediction using machine learning methods. *International Journal of Computer Engineering and Technology*, *10*(3), 2019.
13. Pahwa, K., & Agarwal, N. (2019, February). Stock market analysis using supervised machine learning. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 197-200). IEEE.

14. Strader, T. J., Rozycki, J. J., Root, T. H., & Huang, Y. H. J. (2020). Machine learning stock market prediction studies: review and research directions. *Journal of International Technology and Information Management*, *28*(4), 63-83.
15. Kohli, P. P. S., Zargar, S., Arora, S., & Gupta, P. (2019). Stock prediction using machine learning algorithms. In *Applications of Artificial Intelligence Techniques in Engineering: SIGMA 2018, Volume 1* (pp. 405-414). Springer Singapore.
16. Subasi, A., Amir, F., Bagedo, K., Shams, A., & Sarirete, A. (2021). Stock market prediction using machine learning. *Procedia Computer Science*, *194*, 173-179.