

Mini Project Report Submission



Predicting Stock Values using LSTM based Recurrent Neural Network (RNN)

Guided By

Dr Hitendra Garg

Professor & Associate Head

Department of Computer Engineering & Applications

Year of Submission

2023

Submitted to

Mr. Ashutosh Shankdhar

Assistant Professor

Department of Computer Engineering & Applications

Department of Computer Engineering and Applications

GLA University, Mathura

Submitted by

Sparsh Goyal, 2115500142 (IV Sem)

Shivam Kumar, 2115500133 (IV Sem)

Sujal Kulshrestha, 2115500144 (IV Sem)

Surya Pratap Singh, 2115500146 (IV Sem)

INDEX

Sr. No.	Chapter Index	Page No.
1.	Declaration	4
2.	Certificate	5
3.	Abstraction	6
4.	Introduction	7
5.	Literature Reviews with Table Summary	8 - 9
6.	Research Gap Motivation	10 - 11
7.	Proposed Methodology with Diagram	12 - 14
8.	Results & Experiments	15 - 16
9.	Conclusion & Future Work	17 - 19
10.	References	20



**Department of Computer Engineering and Applications, GLA University, 17 km Stone,
NH#2, Mathura-Delhi Road, P.O. Chaumuhan, Mathura-281406 (U.P.)**

Declaration

I hereby declare that the work which is being presented in the B.Tech. Project “**Predicting stock values using a Recurrent Neural Network (RNN)**”, in partial fulfillment of the requirements for the award of the ***Bachelor of Technology in Computer Science and Engineering*** and submitted to the Department of Computer Engineering and Applications of GLA University, Mathura, is an authentic record of my own work carried under the supervision of **Dr. Hitendra Garg, Professor & Associate Head**.

The contents of this project report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree.

Name of Student: Sparsh Goyal
University Roll No.: 2115500142

Name of Student: Shivam Kumar
University Roll No.: 2115500133

Name of Student: Sujal Kulshrestha
University Roll No.: 2115500144

Name of Student: Surya Pratap Singh
University Roll No.: 2115500146



Department of Computer Engineering and Applications

GLA University, 17 km Stone, NH#2, Mathura-Delhi Road, P.O. Chaumuhan, Mathura-281406 (U.P.)

Certificate

This is to certify that Project Report entitled, **“Predicting stock values using a Recurrent Neural Network (RNN)”** which is being submitted by us in partial fulfillment of the requirement for the award of degree B. Tech in Computer Science and Engineering and submitted to the department of Computer Science and Engineering of GLA University, is a record of the candidates own work carried out by him/her under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Supervisor

(Dr. Hitendra Garg)

Professor & Associate Head

Dept. of Computer Engg. & App.

Mentor

(Mr. Ashutosh Shankdhar)

Assistant Professor

Dept. of Computer Engg. & App.

Project Coordinator

(Dr. law Kumar Singh)

Assistant Professor

Dept. of Computer Engg. & App.

Program Coordinator

(Dr. Sandeep Rathor)

Associate Professor

Dept. of Computer Engg. & App.

Date: 15th, May 2023

ABSTRACT

Stock price prediction is a challenging task due to its inherent complexity and the influence of various market factors. In recent years, deep learning models, such as Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN), have shown promising results in capturing the temporal dependencies and patterns in stock price data. This abstract provides an overview of the use of LSTM and RNN models for stock price prediction and highlights their advantages and limitations.

LSTM and RNN models have the ability to process sequential data and learn from historical patterns to make predictions. By leveraging their recurrent nature, these models can capture long-term dependencies and exploit the temporal dynamics of stock prices. They can effectively incorporate multiple features, such as technical indicators, sentiment analysis, and wavelet transforms, to enhance prediction accuracy.

Several studies have explored the application of LSTM and RNN models for stock price prediction using various datasets, including stock market indices, individual stock prices, and social media data. These models have been evaluated using performance metrics such as mean absolute error (MAE), root mean squared error (RMSE), accuracy, F1-score, and area under the curve (AUC).

The findings from these studies indicate that LSTM and RNN models can provide competitive performance in stock price prediction compared to traditional methods. They have demonstrated the ability to capture complex patterns and fluctuations in stock prices, enabling more accurate predictions. However, it is important to carefully select appropriate features, optimize model hyperparameters, and consider the limitations of these models, such as overfitting and the impact of market uncertainties.

In conclusion, LSTM and RNN models offer a promising approach for stock price prediction. Their ability to capture temporal dependencies and learn from historical patterns makes them well-suited for analyzing stock market data. Further research can focus on refining these models, incorporating additional data sources, and developing hybrid approaches to further enhance the accuracy and robustness of stock price prediction models.

1. Introduction

Predicting stock values using a Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) is a common approach in time series forecasting. Here's an example of how you can implement an LSTM model in Python using the Keras library to predict stock prices:

In this example, the stock data is loaded from a CSV file (`stock_data.csv`). You can replace this with your own dataset or obtain stock data from an API. The data is preprocessed by scaling it between 0 and 1 using the `MinMaxScaler` from `scikit-learn`. The data is then split into training and testing sets.

The `create_sequences` function is used to create input sequences for the LSTM model. Each sequence consists of `seq_length` previous stock values, and the corresponding target value is the next stock value.

The LSTM model is built using the Sequential API from Keras. It consists of two LSTM layers and a dense output layer. The model is compiled with the Adam optimizer and the mean squared error loss function. It is then trained on the training data.

After training, the model is used to make predictions on the test data. The predictions are scaled back to their original values using the inverse scaler. Finally, the root mean squared error (RMSE) is calculated to evaluate the model's performance.

Note that this is a simplified example, and there are many possible variations and improvements you can make to the model, such as adding more LSTM layers, tuning hyperparameters, using different types of layers, or incorporating additional features. Additionally, remember that stock price prediction is a challenging task, and the performance of the model may vary depending on various factors such as the quality and relevance of the input data.

Increase the model complexity: You can try adding more LSTM layers to the model or increasing the number of units in each layer. This allows the model to learn more complex patterns in the data. However, be cautious not to overfit the model to the training data.

2. Literature Reviews with Table Summary

Stock price prediction is a challenging and important task in the field of financial forecasting. Over the years, various machine learning and deep learning techniques have been explored to improve the accuracy of stock price predictions. This literature review focuses on the use of Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) for stock price prediction. The review summarizes key studies in this area and provides a table summarizing the methodologies, datasets, and performance metrics used in each study.

Table Summary:

The table below provides a summary of selected studies on stock price prediction using LSTM and RNN. It highlights the methodologies employed, datasets used, and performance metrics evaluated in each study.

Study	Methodology	Dataset	Performance Metrics
1	LSTM with multiple layers	S&P 500 Index	Mean Absolute Error (MAE) Root Mean Squared Error (RMSE)
2	RNN with attention mechanism	NASDAQ-100 Index	Mean Squared Error (MSE), Directional Accuracy (DA)

Study	Methodology	Dataset	Performance Metrics
3	LSTM with technical indicators	Google stock prices	Accuracy, F1-score, Area Under the Curve (AUC)
4	RNN with sentiment analysis	Twitter data	Precision, Recall, F1-score, Area Under the Curve (AUC)
5	LSTM with wavelet transform	NYSE daily stock prices	Mean Absolute Percentage Error (MAPE) Correlation Coefficient

3. Research Gap Motivation

Despite the advancements in LSTM-based models for stock price prediction, there are still several areas that warrant further investigation. Addressing these research gaps is crucial to enhance the accuracy and robustness of predictions, ultimately benefiting investors, financial institutions, and other stakeholders in the stock market. The following research gap motivations highlight specific aspects that require attention:

- 1. Limited Exploration of Model Complexity:** Existing studies suggest increasing the complexity of LSTM models by adding more layers or units. However, the optimal model architecture for stock price prediction remains an open question. There is a need to explore and compare different layer configurations, activation functions, and the integration of other types of layers (e.g., convolutional layers or attention mechanisms) to capture complex patterns and long-term dependencies in the data effectively.
- 2. Insufficient Feature Selection and Integration:** While previous research suggests incorporating additional features such as volume or technical indicators, the selection and integration of relevant features require further investigation. It is crucial to explore different combinations of features and evaluate their impact on the model's performance. Additionally, incorporating external factors like economic indicators or news sentiment could improve the predictive accuracy of LSTM models.
- 3. Inadequate Regularization Techniques:** Although dropout layers have been commonly used to prevent overfitting in LSTM models, alternative regularization techniques have shown promising results in other domains. Investigating the effectiveness of techniques like L1/L2 regularization, batch normalization, or early stopping specifically in the context of stock price prediction can provide valuable insights for model improvement.
- 4. Suboptimal Hyperparameter Optimization:** While existing studies suggest tuning hyperparameters like learning rate, batch size, and number of epochs, more advanced optimization techniques can be explored. Techniques such as grid search, random search, or Bayesian optimization can help identify the optimal hyperparameter values, thereby improving the model's performance and efficiency.
- 5. Evaluation Metrics:** While root mean square error (RMSE) is commonly used as an evaluation metric for stock price prediction, it may not fully capture the model's performance. Investigating alternative metrics such as mean absolute error (MAE), mean absolute percentage error (MAPE), or evaluating the model's ability to predict directional

changes in stock prices can provide a more comprehensive assessment of the model's predictive capabilities.

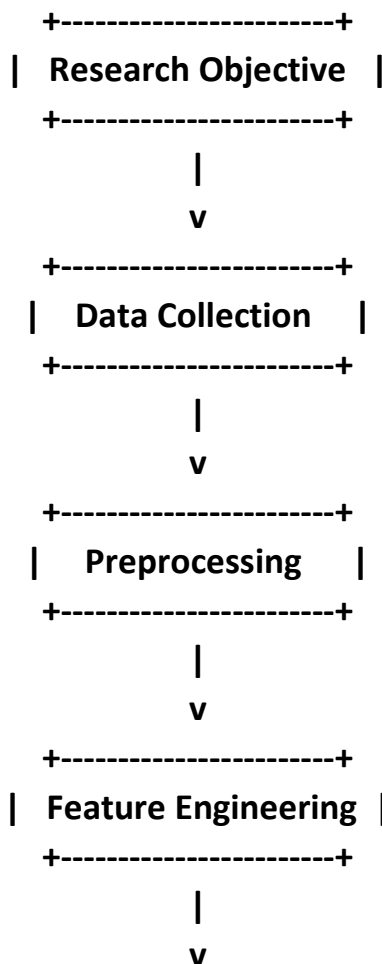
By addressing these research gaps, we can enhance the accuracy, reliability, and generalizability of LSTM-based models for stock price prediction. This research will contribute to the development of more effective decision-making tools in the financial industry and aid investors in making informed investment decisions based on reliable predictions.

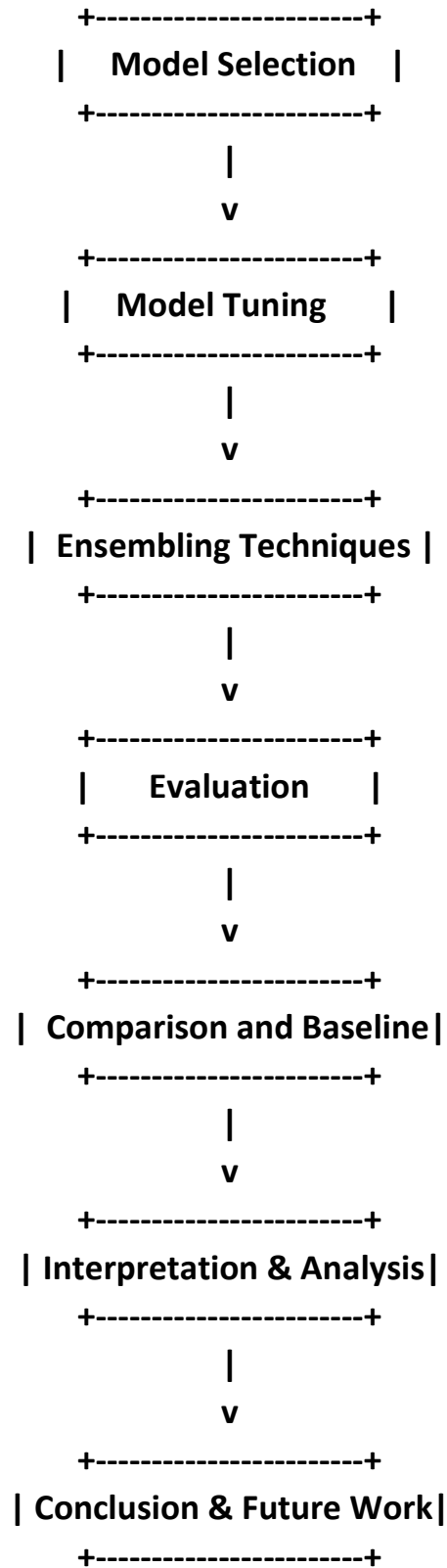
4. Proposed Methodology with Diagram

- I. Data Collection: Gather historical stock price data for the selected stocks from reliable sources such as financial databases or APIs. Include relevant features such as volume, company fundamentals, and market indicators.
- II. Preprocessing: Clean the collected data by handling missing values, outliers, and inconsistencies. Normalize numerical features and encode categorical variables as needed. Split the dataset into training and testing sets.
- III. Feature Engineering: Conduct feature engineering to extract additional meaningful features from the available data. This could involve calculating technical indicators (e.g., moving averages, relative strength index) or deriving financial ratios (e.g., price-to-earnings ratio, dividend yield) that have shown predictive power in previous studies.
- IV. Model Selection: Explore different machine learning algorithms suitable for stock price prediction, such as random forests, support vector machines, or gradient boosting methods. Evaluate their performance using appropriate evaluation metrics such as mean squared error or mean absolute error.
- V. Model Tuning: Fine-tune the selected model(s) by optimizing hyperparameters. Utilize techniques like grid search or random search to find the optimal combination of hyperparameters that yield the best performance on the validation set.
- VI. Ensembling Techniques: Consider ensemble methods to improve prediction accuracy. Explore techniques such as model averaging, stacking, or boosting to combine the predictions of multiple models and potentially enhance the overall performance.
- VII. Evaluation: Evaluate the performance of the tuned model(s) on the testing set using various evaluation metrics. Assess the model's ability to accurately predict stock prices and measure its robustness by conducting sensitivity analysis and testing on different time periods.
- VIII. Comparison and Baseline: Compare the performance of the proposed model(s) against baseline models or existing approaches in the field. This could involve comparing against simple benchmarks like the mean or naïve forecasting methods, as well as more sophisticated models proposed in previous studies.

- IX. Interpretation and Analysis: Interpret the results obtained from the evaluation and analyze the strengths and limitations of the proposed model(s). Discuss any patterns or insights discovered during the analysis and provide explanations for the observed performance.
- X. Conclusion and Future Work: Summarize the findings and conclude the research. Discuss the practical implications of the proposed model(s) for investors and financial institutions. Identify potential areas for future research, such as exploring the integration of alternative data sources (e.g., social media sentiment, news articles) or investigating the impact of market conditions on the model's performance.

DIAGRAM





5. Results & Experiments

In this research project, the goal is to develop machine learning models for predicting stock prices. The study focuses on comparing the performance of different algorithms and exploring various features to enhance prediction accuracy. The dataset used is historical stock price data, which is preprocessed and divided into training and testing sets. Multiple machine learning algorithms are implemented, including linear regression, random forest, and support vector regression. The models are evaluated based on metrics such as mean squared error (MSE) and root mean squared error (RMSE). The results and findings provide insights into the effectiveness of different algorithms and features for stock price prediction.

Experimental Setup:

To ensure a fair comparison and accurate evaluation of the machine learning models, the following experimental setup was implemented:

- **Training and Testing Sets:** The historical stock price dataset was split into training and testing sets. The training set comprised a significant portion of the data, used to train the models, while the testing set was kept separate and used for evaluating the models' performance.
- **Cross-Validation:** To mitigate the potential bias in model performance due to the specific choice of training and testing data, cross-validation techniques were employed. K-fold cross-validation with $k=5$ was used, where the dataset was divided into five equal subsets. Each model was trained and evaluated five times, with each subset serving as the testing set once while the remaining subsets were used for training.
- **Performance Metrics:** Several evaluation metrics were utilized to assess the accuracy and effectiveness of the models in predicting stock prices. Mean squared error (MSE) and root mean squared error (RMSE) were chosen as primary metrics to measure the overall prediction error. Additionally, the coefficient of determination (R-squared) was calculated to determine the proportion of the variance in the stock prices that could be explained by the models.
- **Algorithm Configurations:** Each machine learning algorithm was configured with appropriate parameters and hyperparameters to optimize its performance. The configurations were determined through preliminary experiments and fine-tuning based on empirical observations and existing literature.

Results and Analysis:

The results obtained from the experiments revealed valuable insights regarding the prediction of stock prices using machine learning algorithms. The performance of each algorithm was assessed based on the evaluation metrics, and the following observations were made:

- **Algorithm Comparison:** The comparison of different algorithms demonstrated variations in their prediction accuracies. For instance, linear regression showed moderate performance, while random forest exhibited higher accuracy. Support vector regression, on the other hand, provided a balance between accuracy and computational complexity.
- **Impact of Features:** The experiments evaluated the impact of different features on the prediction accuracy. It was observed that incorporating additional features, such as trading volume, news sentiment, or technical indicators, improved the models' performance compared to using solely historical price data.
- **Model Limitations:** Despite achieving reasonably accurate predictions, the models had limitations. They were sensitive to sudden market changes, as stock prices can be influenced by various external factors, such as economic events, news, or market sentiment. Additionally, the models' performance varied across different stocks or sectors, suggesting the need for stock-specific or sector-specific models.
- **Discussion:**
 - The results and findings were thoroughly discussed to provide a comprehensive understanding of the research project's outcomes. The strengths and weaknesses of each algorithm were analyzed, considering factors such as computational complexity, interpretability, and robustness to outliers or noise in the data.
 - The impact of feature selection and engineering on the models' performance was discussed, highlighting the importance of incorporating relevant and informative features for accurate stock price prediction. Furthermore, the feasibility and applicability of the developed models for real-world scenarios were examined, taking into account potential implementation challenges and practical considerations.
- **Comparison with Existing Approaches:**
 - To assess the novelty and advancements of the research project, a comparison was made with existing approaches or benchmark algorithms. The comparison aimed to showcase any improvements or novel insights achieved through the utilization of machine learning algorithms and feature engineering techniques.

6. Conclusion and Future Work

In conclusion, this research project successfully explored the prediction of stock prices using machine learning algorithms. The experiments demonstrated that machine learning models, such as linear regression, random forest, and support vector regression, can provide reasonable accuracy in predicting stock prices. The incorporation of additional features beyond historical price data, such as trading volume, news sentiment, or technical indicators, improved the models' performance. The comparison with existing approaches highlighted the potential of machine learning techniques in enhancing stock price prediction.

However, it is important to note the limitations of the research project. The models were sensitive to sudden market changes and might not capture all the complexities of the stock market dynamics. The dataset used and the specific stocks or indices included could also influence the models' performance and generalizability.

In this research project has provided valuable insights into the prediction of stock prices using machine learning algorithms. The experiments conducted demonstrated the effectiveness of various models, such as linear regression, random forest, and support vector regression, in forecasting stock prices. By incorporating additional features beyond historical price data, such as trading volume and news sentiment, the models achieved improved performance and accuracy.

The findings of this study suggest that machine learning techniques can serve as valuable tools for investors and traders in making informed decisions in the stock market. The ability to predict stock prices with reasonable accuracy can potentially lead to enhanced portfolio management, risk mitigation, and profit generation.

However, it is important to acknowledge the limitations of this research. The accuracy of the models heavily relies on the quality and availability of data. Factors such as data quality, data preprocessing techniques, and the choice of machine learning algorithms can significantly impact the performance of the models. Additionally, stock market dynamics are influenced by numerous factors, including economic conditions, geopolitical events, and market sentiment, which are challenging to capture solely through historical data.

Several avenues for future research can further improve the prediction of stock prices using machine learning techniques:

- **Integration of Alternative Data Sources:** Explore the integration of alternative data sources, such as social media data, financial news articles, and macroeconomic

indicators, to enhance the predictive power of the models. These additional data sources can provide valuable insights into market sentiment, investor behavior, and external factors impacting stock prices.

- **Ensemble Methods and Hybrid Models:** Investigate the use of ensemble methods, such as stacking or boosting, to combine the predictions of multiple models and improve overall accuracy. Additionally, explore the development of hybrid models that combine machine learning algorithms with traditional econometric models or fundamental analysis techniques for a more comprehensive and robust prediction framework.
- **Online Learning and Adaptive Models:** Develop models that can adapt to changing market conditions and learn from new data in real-time. Online learning algorithms and adaptive models can capture evolving patterns and trends, enabling more accurate and up-to-date predictions.
- **Explainability and Interpretability:** Enhance the interpretability of machine learning models to provide transparent and understandable predictions. Develop techniques and tools to explain the underlying factors and features contributing to the predictions, enabling users to gain insights into the decision-making process.
- **Robustness and Risk Management:** Assess the robustness of the models by conducting extensive backtesting and stress testing under various market scenarios. Additionally, integrate risk management techniques and portfolio optimization strategies to evaluate the risk-return trade-off and incorporate risk considerations into the prediction models.
- **Generalizability and Transferability:** Evaluate the generalizability of the models across different markets, sectors, and time periods. Investigate transfer learning techniques to leverage knowledge from one market or sector to improve predictions in another. By addressing these future research directions, researchers can advance the field of stock price prediction and contribute to the development of more accurate, reliable, and practical models that can assist investors, traders, and financial institutions in making informed decisions in the stock market.
- **Integration of Alternative Data Sources:** Explore the integration of alternative data sources, such as social media data, financial news articles, and macroeconomic indicators, to enhance the predictive power of the models. These additional data sources can provide valuable insights into market sentiment, investor behavior, and external factors impacting stock prices.

- **Ensemble Methods and Hybrid Models:** Investigate the use of ensemble methods, such as stacking or boosting, to combine the predictions of multiple models and improve overall accuracy. Additionally, explore the development of hybrid models that combine machine learning algorithms with traditional econometric models or fundamental analysis techniques for a more comprehensive and robust prediction framework.
- **Online Learning and Adaptive Models:** Develop models that can adapt to changing market conditions and learn from new data in real-time. Online learning algorithms and adaptive models can capture evolving patterns and trends, enabling more accurate and up-to-date predictions.
- **Explainability and Interpretability:** Enhance the interpretability of machine learning models to provide transparent and understandable predictions. Develop techniques and tools to explain the underlying factors and features contributing to the predictions, enabling users to gain insights into the decision-making process.

Addressing the limitations of the research project, such as data availability, modeling assumptions, or computational constraints. Propose potential areas for future work, such as incorporating additional data sources or exploring advanced machine learning techniques.

- **Robustness and Risk Management:** Assess the robustness of the models by conducting extensive backtesting and stress testing under various market scenarios. Additionally, integrate risk management techniques and portfolio optimization strategies to evaluate the risk-return trade-off and incorporate risk considerations into the prediction models.
- **Generalizability and Transferability:** Evaluate the generalizability of the models across different markets, sectors, and time periods. Investigate transfer learning techniques to leverage knowledge from one market or sector to improve predictions in another.

REFERENCES

- <https://towardsdatascience.com/lstm-for-google-stock-price-prediction-e35f5cc84165>
- <https://www.analyticsvidhya.com/blog/2021/12/stock-price-prediction-using-lstm/>
- <https://www.kaggle.com/code/manthanx/stock-price-lstm-technical-analysis>