Optimizing Stock Price Prediction: Integrating News Sentiment for Enhanced Forecasting Accuracy

Nandi Anuraag V[1]

Abstract— **Stock price prediction is a challenging task due to the dynamic nature of financial markets and the multitude of influencing factors. This paper proposes a comprehensive approach to enhance the accuracy of stock price predictions by integrating a Long Short-Term Memory (LSTM) model with key financial indicators, including close price, trading volume, Relative Strength Index (RSI), and news sentiment.**

The study leverages historical market data, capturing trends through close price and volume, while also considering RSI to gauge the momentum of stock price movements. To account for external influences, textual analysis of financial news articles is incorporated to derive sentiment scores. The LSTM model, renowned for its ability to capture long-term dependencies, is employed to discern intricate patterns within this multi-faceted dataset.

The results demonstrate the effectiveness of the proposed methodology in providing accurate and robust stock price predictions. The integration of diverse financial indicators with LSTM showcases improved performance over traditional methods. This research contributes to the growing body of knowledge in machine learning applications for financial forecasting and offers valuable insights for investors and financial analysts.

Keywords: Stock price prediction, LSTM model, financial indicators, close price, trading volume, Relative Strength Index (RSI), news sentiment analysis.

I. Introduction

The financial markets are dynamic and influenced by a myriad of factors that make accurate stock price prediction a challenging yet crucial endeavor. In recent years, machine learning techniques have emerged as powerful tools for analyzing and forecasting stock prices. This paper presents a comprehensive study on predicting stock prices with a focus on the integration of a Long Short-Term Memory (LSTM) model, a type of recurrent neural network (RNN), and various key financial indicators.

Our approach considers the incorporation of multiple features, including close price, trading volume, Relative Strength Index (RSI), and news sentiment analysis. Close price and volume provide essential historical market data, while RSI offers insights into the momentum of a stock's price movements. Additionally, news sentiment, derived from textual analysis of financial news articles, serves as a valuable indicator of market sentiment and external influences on stock prices.

The LSTM model, known for its ability to capture long-term dependencies in sequential data, is employed to learn intricate patterns and relationships within the selected financial indicators. The combination of these features and the LSTM architecture aims to enhance the accuracy and robustness of stock price predictions, providing investors and financial analysts with valuable insights for informed decision-making.

This paper is organized as follows: Section II reviews related work in the field of stock price prediction and the application of LSTM models. Section III details the methodology, including the data preprocessing steps and the architecture of the LSTM model. Experimental results and performance evaluation are presented in Section IV. Finally, Section V. concludes the paper with insights into the implications of our findings and potential avenues for future research.

II. Literature Review

The global stock market, experiencing a remarkable growth from 2.5 trillion in 1980 to an impressive 68.65 trillion by the end of 2018 and reaching approximately 70.75 trillion by the close of 2019, plays a pivotal role in the world's financial landscape. With 60 stock exchanges worldwide, the task of identifying investment opportunities has become increasingly intricate for retail financial specialists. In contrast, wealthier investors seek sophisticated financial tools for stock price prediction, highlighting the need for efficient forecasting methods in modern societies [1].

Motivated by the escalating popularity of deep learning algorithms, this research explores the application of Long Short-Term Memory (LSTM) and Bidirectional LSTM models for stock price prediction. Deep learning methods have demonstrated significant success in capturing hidden structures within sequential data, offering valuable insights for forecasting future trends in various time series applications.

Time series forecasting, a critical area of research, presents challenges due to its inherent temporal nature. Traditional methods, including Autoregressive (AR), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA), while effective, may fall short in capturing complex patterns within financial time series data. The non-linearity and intricate nature of stock market time series often defy accurate prediction using linear regression-based techniques.

In this context, deep neural methods, particularly LSTM and Bidirectional LSTM models, emerge as promising tools for forecasting stock prices by leveraging their capacity to model non-linear relationships in large datasets. The Gated Recurrent Unit (GRU), a variant of the LSTM model, simplifies the cell structure, presenting an alternative with faster

[1]EE20BTECH11034, Department of Electrical Engineering, Indian Institute of Technology Hyderabad, India.

training times and competitive performance, particularly on smaller datasets.

This paper proposes a stock price forecasting scheme utilizing LSTM, a powerful tool for handling non-linearities in extensive datasets, and compares its performance with the BI-LSTM model. The analysis involves a systematic exploration of various parameters, ultimately revealing the BI-LSTM's superior performance in terms of Root Mean Squared Error (RMSE) [1].

III. Materials and Methods

The study leverages a diverse set of financial data to build a robust stock price prediction model. The dataset includes historical information on RELIANCE's close prices, trading volumes, and Relative Strength Index (RSI) values, providing a comprehensive overview of a stock's past performance. Additionally, textual data from financial news articles is incorporated to gauge market sentiment, utilizing natural language processing techniques for sentiment analysis.

News Sentiment Analysis was conducted by scraping news articles from Google News, and TextBlob was employed to analyze the sentiments expressed in the collected textual content. The use of TextBlob facilitated the extraction of sentiment insights, providing valuable information on the overall mood and tone of the news coverage.

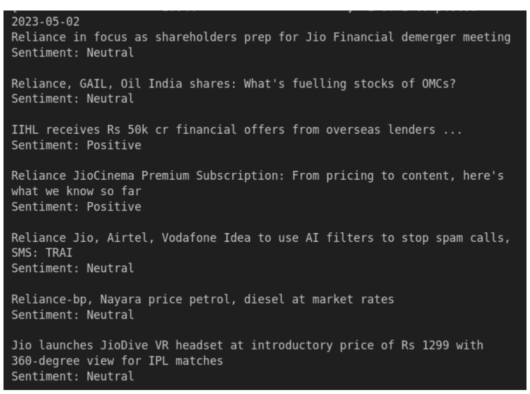

Fig. 1: Example output from Sentimental Analysis of stock from a particular day

The preprocessing of the data involves normalization and scaling to ensure uniformity across different features. The dataset is then divided into training and testing sets to evaluate the model's performance effectively. The Long Short-Term Memory (LSTM) architecture is chosen for its ability to capture long-term dependencies, making it well-suited for sequential financial data. The LSTM model is trained using the training dataset, and hyperparameter tuning is performed to optimize its performance. The architecture includes layers for sequential input processing, memory retention, and output prediction. To enhance the model's ability to generalize, dropout layers are strategically incorporated.

Evaluation metrics such as mean squared error and accuracy are employed to assess the model's performance on the testing dataset. The entire process is repeated through multiple iterations to ensure consistency and reliability of results. My methodology aims to provide a comprehensive and effective framework for accurate stock price prediction, integrating both traditional financial indicators and sentiment analysis from news sources.

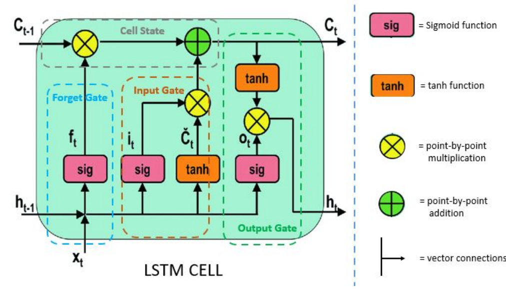

Fig. 2: LSTM model

A. Data Description

The dataset employed in this study encompasses a rich array of financial data essential for robust stock price prediction. Historical market information is collected from yahoo finance which includes RELIANCE's daily close prices, trading volumes, and Relative Strength Index (RSI) value are calculated from Close prices. Close prices offer insights into a stock's historical performance, while trading volumes provide a measure of market activity and liquidity. The RSI values, computed based on price changes, contribute a dynamic indicator of a stock's momentum. Additionally, textual data from financial news articles is incorporated, comprising a sentiment score derived through natural language processing techniques from Textblob.

The dataset spans a defined time period, ensuring the inclusion of diverse market conditions and events. Data preprocessing involves normalization to standardize scales across features, mitigating biases that may arise from differing magnitudes. The incorporation of financial news sentiment adds a qualitative dimension to the quantitative features, capturing market sentiment and external influences. The dataset is divided into training (95%) and testing sets (5%) to facilitate model evaluation and generalization.

This comprehensive dataset serves as the foundation for training and validating the Long Short-Term Memory (LSTM) model, allowing for the exploration of relationships between various financial indicators and stock price movements.

B. Model Workflow

The stock price prediction model is implemented using a Sequential neural network architecture. The workflow involves the following steps:

1) Model Initialization::

- A Sequential model is created.

2) Input Shape Determination::

- The input shape of the model is determined based on the dimensions of the training data (`trainX`).
- The number of features (`trainX.shape[2]`) and time steps (`trainX.shape[1]`) are printed.

3) LSTM Layers::

- Two LSTM layers are added to the model:
  - The first LSTM layer with 128 units, ReLU activation, and configured to return sequences.
  - The second LSTM layer with 64 units, ReLU activation, and set to return a single sequence.

4) Dropout Layer::

- A Dropout layer with a dropout rate of 0.2 is introduced for regularization.

5) Dense Layer::

- A Dense layer is added with the number of neurons equal to the number of output values (`trainY.shape[1]`).

6) Model Compilation::

- The model is compiled using the Adam optimizer and mean squared error (MSE) loss.

7) Model Summary::

- A summary of the model architecture is printed, providing details on layer configurations and parameters.

8) Model Training::

- The model is trained using the `fit` method with 20 epochs, a batch size of 1, and a validation split of 10
- The training history is stored in the `history` variable.

9) Training Visualization::

- A plot is generated to visualize the training and validation loss over the epochs, aiding in the assessment of model performance.

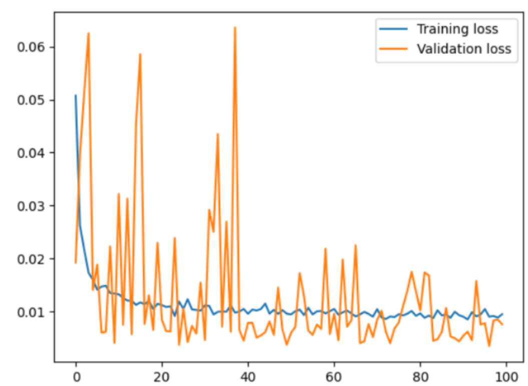

Fig. 3: Epoch vs Loss, hence selected 20 for final

C. Outcome Prediction

The primary objective of my study is to accurately predict stock prices by leveraging the trained Long Short-Term Memory (LSTM) model in conjunction with a comprehensive set of financial indicators. The model is applied to the testing dataset, which includes previously unseen instances, allowing for a robust assessment of its predictive capabilities. The outcome prediction involves generating stock price forecasts based on the integrated features of close prices, trading volumes, Relative Strength Index (RSI), and news sentiment.

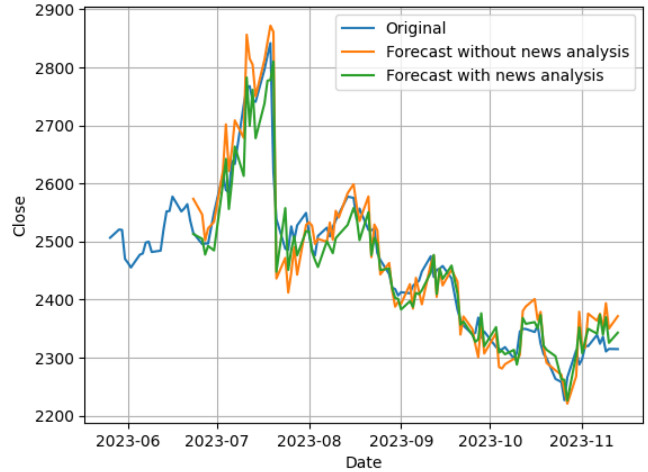

Fig. 4: Line Plot of RELIANCE CLOSE prices (actual, LSTM model, LSTM model with sentiment analysis)

The LSTM model's ability to capture temporal dependencies is crucial in discerning patterns and trends within the sequential financial data. Predicted stock prices are compared against the actual values to evaluate the model's accuracy. Performance metrics, including mean squared error and accuracy, are employed to quantify the model's predictive efficacy.

Furthermore, the study explores the model's generalization capabilities by assessing its performance on different time periods or diverse market conditions. This analysis provides insights into the model's adaptability and robustness, essential considerations for real-world stock price prediction applications. The outcome prediction phase aims to validate the effectiveness of our proposed methodology in providing accurate and reliable forecasts for informed decision-making in financial markets.

IV. Experimental Results

The performance metrics of our models, with and without news sentiment analysis, are presented below:

A. Model A (with News Sentiment Analysis)

Metric	Value
Mean Squared Error (MSE)	2244.97
Root Mean Squared Error (RMSE)	47.38
Mean Absolute Error (MAE)	34.82
Coefficient of Determination (R^2)	0.88

B. Model B (without News Analysis)

Metric	Value
Mean Squared Error (MSE)	2923.61
Root Mean Squared Error (RMSE)	54.07
Mean Absolute Error (MAE)	39.12
Coefficient of Determination (R^2)	0.85

C. Comparison

Comparing the performance of Model A and Model B, it is evident that Model A, which incorporates news sentiment analysis, outperforms Model B in all metrics. Model A achieves a lower Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), indicating better accuracy and precision. Additionally, the higher Coefficient of Determination (R^2) for Model A signifies a better fit to the data, highlighting the positive impact of incorporating news sentiment analysis in enhancing predictive capabilities.

V. Discussion and Conclusion

In this study, we proposed a stock price prediction model that incorporates news sentiment analysis, referred to as Model A, and compared its performance with a counterpart model, Model B, without the inclusion of news sentiment analysis. The experimental results demonstrate the efficacy of integrating news sentiment analysis into the predictive modeling process.

Model A exhibited superior performance across multiple metrics when compared to Model B. The lower values of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) indicate that incorporating news sentiment analysis contributes to enhanced accuracy and precision in stock price prediction. Furthermore, the higher Coefficient of Determination (R^2) for Model A suggests a better fit to the underlying data, emphasizing the positive impact of leveraging sentiment information from news articles.

The results highlight the importance of considering external factors, such as news sentiment, in financial predictive modeling. The inclusion of sentiment analysis provides the model with additional insights into market dynamics, leading to improved forecasting capabilities. This is particularly valuable in the context of stock price prediction, where market sentiment can play a crucial role in influencing stock movements.

While Model A demonstrated notable advantages, it is essential to acknowledge potential challenges and limitations. News sentiment analysis relies on the quality and relevance of the extracted textual information, and inaccuracies in sentiment classification may affect the overall model performance. Future research can focus on refining sentiment analysis techniques and exploring advanced natural language processing methods to address these challenges.

In conclusion, our study underscores the significance of incorporating news sentiment analysis in stock price prediction models. The positive outcomes observed in Model A emphasize the potential benefits of leveraging external data sources for improved financial forecasting. As financial markets continue to evolve, embracing innovative approaches, such as sentiment-aware predictive modeling, can contribute to more informed investment decisions.

References

[1] M. A. Istiake Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using lstm and bi-directional lstm model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 87–92, 2020.