Stock Market Prediction and Analysis: A Comprehensive Review and Development of an Intelligent Stock Market Guide

Ayush Srivastava   Ankush Gupta   Harshvardhan Singh   Aditya Dahiya

Shantanu Prakash

1. Updated Problem Statement

Predicting stock market trends has always been a challenge for professionals in the financial sector, including statisticians, economists, and other experts. Stocks play a vital role in driving the economy, being bought, sold, and traded on various stock exchanges. However, the market is influenced not only by economic factors but also by psychological and human elements, making it inherently unpredictable. Traditional methods of stock prediction, such as technical and fundamental analysis, rely on historical data and company fundamentals to forecast future stock prices. In our research initiative, we endeavor to craft an intelligent system employing natural language processing and machine learning techniques such as LSTM and Random Forest. This system will provide users, both novice investors and seasoned traders, with easily interpretable visual as well as textual insights to mitigate risk and make informed investment decisions. By summarizing past financial activities, integrating with the latest news, and offering personalized advice, the system seeks to enhance user engagement and facilitate better financial outcomes.

2. Literature Review

1. **Sentiment analysis of financial news using unsupervised approach**

   This paper explores Sentiment analysis and aims to determine the sentiment strength from a textual source for good decision making. This work focuses on application of sentiment analysis in financial news. The semantic orientation of documents is first calculated by tuning the existing technique for financial domain. The existing technique is found to have limitations in identifying representative phrases that effectively capture the sentiment of the text. Two alternative techniques - one using Noun-verb combinations and the other a hybrid one, are evaluated. Noun-verb approach yields best results in the experiment conducted.

2. **Stock market prediction using machine learning classifiers and social media news**

   This research paper explores the use of social media and financial news data to predict stock market trends. The authors collected data from Twitter, Yahoo Finance, and Business Insider for two years, and used sentiment analysis to assess the overall sentiment of the data. They then trained ML classifiers on the combined data set to predict future stock market trends. The results showed that the classifiers were able to predict stock market trends with some accuracy, but that the performance varied depending on the classifier and the stock market being predicted. The authors also found that social media data was more useful for predicting short-term trends, while financial news data was more useful for predicting long-term trends.

3. **Stock Market Prediction using Financial News Articles**

   This paper discusses extracting data from trusted news sources, cleaning the text using natural language processing, and applying sentiment analysis techniques to determine sentiment polarity. The review highlights the use of machine learning algorithms, such as Linear Regression, to predict stock prices based on historical patterns and news sentiment. It also mentions the integration of sentiment analysis with financial news articles, giving equal weightage to the model to enhance predictive models. The methodology emphasizes feature selection methods and classification techniques to improve the accuracy of stock price predictions.

4. **News Sensitive Stock Trend Prediction**

   The paper introduces a novel methodology for stock trend prediction that integrates incremental K-means clustering, new weighting schemes, and market simulation. They explored time series segmentation techniques and document clustering. It addresses the limitations of traditional approaches by incorporating incremental K-means clustering to filter news articles

and align them with stock trends. They introduced a new weighting scheme that enhances feature importance identification within article collections, improving prediction accuracy. Their methodology also leverages agglomerative hierarchical clustering based on slopes and coefficients of determination to cluster interesting trends and align news articles accordingly.

5. **Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis**

   Previous studies have focused on analyzing the correlation between economic news sentiment and stock market behavior, utilizing techniques like SVM and KNN classifiers. These studies have shown accuracies ranging from 72.73% to 86.21% in predicting stock trends. In the research paper, researchers highlights the use of a Java-based machine learning toolkit for natural language processing to analyze textual data related to the stock market. They combine sentiment analysis of news articles with historical stock price data to increase accuracy. This paper emphasizes the importance of preprocessing techniques, feature weighting methods like TF-Idf, and the integration of sentiment analysis with technical analysis for robust stock market prediction models.

6. **NEU-Stock: Stock market prediction based on financial news**

   The paper addresses the enduring interest in forecasting stock price movements, particularly focusing on the influence of financial news on FPT Group's stock. Employing PhoBERT, a language model, the authors achieve a 93% accuracy in classifying the impact of financial article titles on stock prices. Subsequently, they introduce the NEU-Stock model, utilizing LSTM-Attention architecture to forecast the next day's stock price, integrating past prices and news impact. The model exhibits strong performance, as evidenced by high R2 coefficients and notable RMSE, underscoring its efficacy in stock price prediction.

3. Baseline Implementation

3.1. Stock prediction

3.1.1 Data Collection and Preparation:

Utilized Yahoo Finance API via the 'yfinance' library to retrieve historical stock price data for selected companies in the Dow Jones Industrial Average.

Specified a start date of March 1, 2015, and an end date of March 1, 2024.

Extracted stock price data for companies.

3.1.2 Data Preprocessing

Applied Min-Max scaling to normalize the data within a range of 0 to 1, ensuring consistency in the scale of features.

3.1.3 Model Building

Employed Long Short-Term Memory (LSTM) neural network architecture for time-series forecasting, implemented using TensorFlow and Keras.

Constructed a sequential model consisting of multiple LSTM layers with dropout regularization to mitigate overfitting.

Defined the network architecture with varying LSTM units to effectively capture temporal dependencies.

Compiled the model using the Adam optimizer and mean squared error loss function.

3.1.4 Training

Trained the LSTM model on each selected company's training data from 2015 to 2021.

Configured the model to predict the stock price for the following day based on past closing prices and news impact.

3.1.5 Testing and Evaluation

Evaluated the model's performance using test data from 2022 onwards.

Predicted the stock prices for the test period and compared them with actual values.

Calculated metrics such as coefficient of determination (R2) and root mean squared error (RMSE) to assess the model's accuracy and reliability.

3.1.6 Result Analysis

Analyzed the model's performance based on the visualizations and evaluation metrics, highlighting its effectiveness in forecasting stock prices.

3.2. News Retrieval

We retrieved news articles from the internet using NewsApi's JSON Api for 30 companies. The API was used to fetch news headline, description, content(truncated to 200 chars), url, author, date published and company tags for news articles related to any of the 30 companies from 8th February 2024 to 6th March 2024. These were then put into a csv file in a structured manner for sentiment analysis.
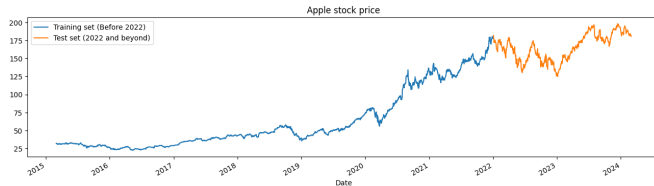
Figure 1. Extracting Stock Prices using API and splitting it into train and test data

3.3. News Sentiment Analysis

To perform sentiment analysis on the financial news obtained, we used the FinBert Model which is built by fine tuning a Bert Model on a large financial corpus. For the financial news data fetched for the past one month, we obtained the sentiment scores(positive, negative and neutral) using softmax activation on the output of the FinBert model which are further utilized by the random regressor model.

3.4. Merging News Sentiments and Stock Predictions

3.4.1 Merging the datasets

For all dates of month February 2024, we took the average sentiments (positive. negative, neutral) of the news for the specific company, and added them in the stock predictions dataset.

Combined the dataset of all the companies to train a random forest regressor.

3.4.2 Training and Testing

Created a 75-25 split training set to train the random forest.

Made predictions on the test set to check the accuracy of the model.

4. Results

We extracted the stock prices and split the dataset into train and test dataset as shown in Figure[1]. Then we used LSTM to predict the values of stock price based on the trend, the variation of stock price of 'Apple' is shown in Figure[2]. Then we fetched the news articles using API and after performing pre-processing on fetched data, we performed sentiment analysis and the result is shown in Figure[3]. Then the random forest predicts the stock prices of company based upon the LSTM and sentiment analysis as shown in Figure[4]. We have got R2 loss ¿ 0.9.

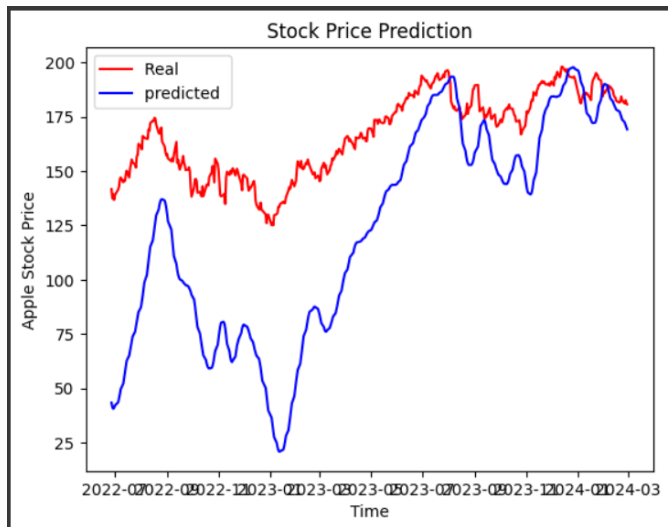We have connected our backend code to a basic Frontend, the front end is similar to Figure[5].
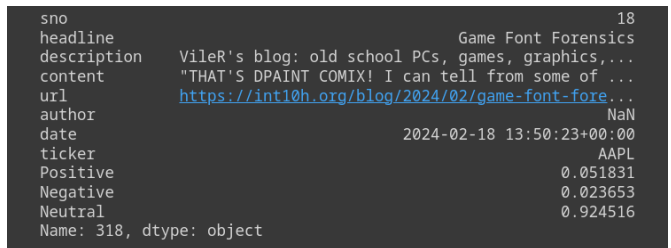

Figure 2. LSTM output for "Apple"

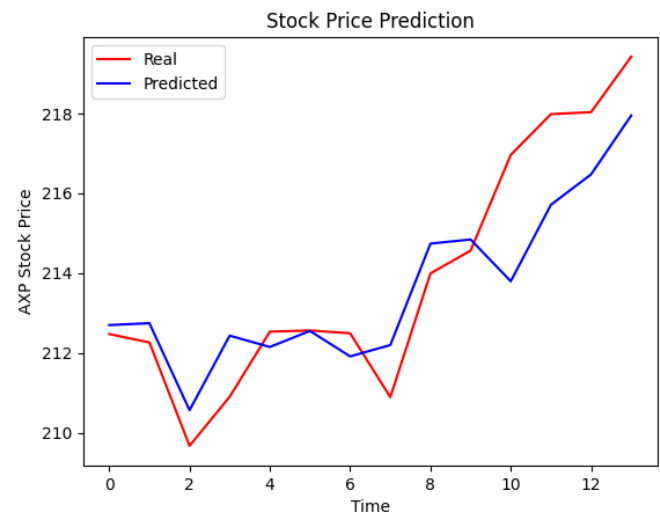Figure 3. Sentiment Analysis output

Figure 4. Random Forest output

5. Improvements in future deadline

- We will expand our news dataset by using web scrapping to collect dataset from URLs, and other multimodal inputs.
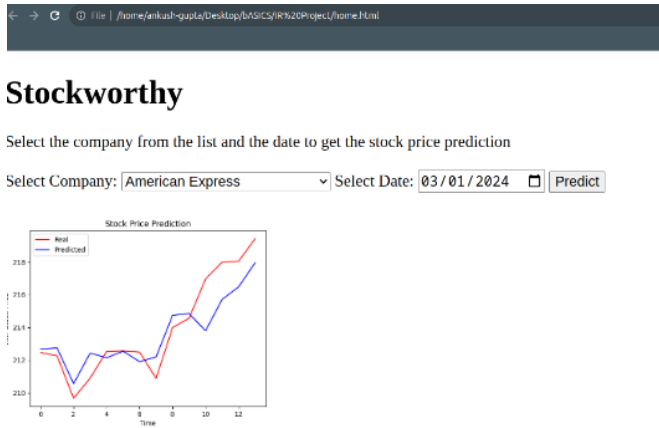
Figure 5. Basic Frontend

- To incorporate the vast dataset indexing will be performed to rank them according to the proximity to the commodity and company

- To expand the use case of our product, we will try to include chatbot, for easier breakdown of stocks, previous trends and factors affecting their price ranges.

- We will modify the UI-UX to make it more user-friendly with various options such as chatbot, graphical representation, detailed analysis of stocks etc. to explore so that the user can choose the service they want.

- To fine-tune the working of the random forest regressor, we will utilise other macro-attributes such as the INR-USD currency exchange values, gold and brent prices, to get a better picture of the market.

6. References

- Chen MS, ed. Advances in Knowledge Discovery and Data Mining: 6th Pacific-Asia Conference; Proceedings. Springer; 2002.

- Faculty of Computers and Information Technology, Future University in Egypt, Khedr AE, S.E.Salama, Yaseen N. Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis. IJISA. 2017;9(7):22-30. doi:10.5815/ijisa.2017.07.03

- M S. Stock Market Prediction using Financial News Articles. https://www.irjet.net/archives/V6/i12/IRJET-V6I12229.pdf

- Khan W, Ghazanfar MA, Azam MA, Karami A, Alyoubi KH, Alfakeeh AS. Stock market prediction using machine learning classifiers and social media, news. J Ambient Intell Human Comput. 2022;13(7):3433-3456. doi:10.1007/s12652-020-01839-w

- Maqbool J, Aggarwal P, Kaur R, Mittal A, Ganaie IA. Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach. Procedia Computer Science. 2023;218:1067-1078. doi:10.1016/j.procs.2023.01.086

- Kalyani J, Bharathi ProfHN, Jyothi ProfR. Stock trend prediction using news sentiment analysis. Published online 2016. doi:10.48550/ARXIV.1607.01958

- NEU-Stock: Stock market prediction based on financial news https://ceur-ws.org/Vol-3026/paper24.pdf

- Sentiment analysis of financial news using unsupervised approach https://www.sciencedirect.com/science/article/pii/S1877050920307912 $RR - 2rr = 86247f833c3fecac$