

Stock Market Prediction and Analysis: A Comprehensive Review and Development of the Stockworthy Chatbot

Ayush Srivastava

Ankush Gupta

Harshvardhan Singh

Aditya Dahiya

Shantanu Prakash

1. Introduction/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 this research endeavor, our aim is to develop an intelligent system, such as a chatbot, that leverages natural language processing and machine learning techniques to analyze financial news sentiment and historical market data. This system will provide users, both novice investors and seasoned traders, with insights to mitigate risk and make informed investment decisions. By summarizing past financial activities, integrating with the latest news, and offering personalized advice, the chatbot seeks to enhance user engagement and facilitate better financial outcomes.

2. Importance

Developing a chatbot capable of predicting stock market trends and offering personalized financial advice holds immense importance in the realm of finance and investment. Firstly, it empowers investors, both novice and seasoned, by providing them with valuable insights derived from the analysis of historical market data and current news sentiment. This empowerment enables investors to make more informed decisions regarding their investment portfolios, thereby enhancing their financial well-being. Secondly, such a project contributes to risk management strategies by enabling investors to anticipate market movements and adjust their investment strategies accordingly, thus minimizing potential losses. Furthermore, the accessibility of a chatbot platform ensures that financial advice and market predictions are readily available to a wide range of in-

vestors, promoting financial inclusivity and education. Ultimately, the development of this project not only facilitates individual investors but also contributes to the overall efficiency and stability of the stock market by promoting informed decision-making and reducing market volatility.

3. Literature Review

1. Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach

This research paper aims to investigate the influence of financial news on stock price trends and to develop a predictive model using sentiment analysis of financial news along with historical stock data. The study explores three sentiment scoring methods - VADER, TextBlob, and Flair - along with an MLP Regressor to predict stock price movements and understand the impact of each sentiment analyzer individually and in combination.

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 highlight 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. Stock Trend Prediction using News Sentiment Analysis

They have explored the use of text mining, sentiment analysis, and classification algorithms to extract valuable insights from news articles and predict stock trends. Prior studies have demonstrated the efficacy

of SVM, Random Forest, and Naïve Bayes algorithms in text classification. They draw inspiration from J. Bean's work on sentiment analysis of airline satisfaction using Twitter data, highlighting the relevance of sentiment detection algorithms in capturing consumer sentiment. Furthermore, incorporating a polarity dictionary comprising positive and negative words and finance-specific terms reflects a nuanced approach to sentiment analysis in financial contexts. They emphasize the importance of term weighting in filtering out common terms and capturing document relevance. The research aligns with established practices in model evaluation and system testing by dividing data into training and test sets and evaluating classifier performance through metrics like accuracy and precision.

4. Methodology

Our methodology involves several key steps to develop the **Stockworthy chatbot**. Initially, we train a **sentiment analysis model on financial news datasets**, such as the Ashraq financial news dataset available on HuggingFace. We select a suitable architecture for sentiment analysis, such as **BERT or RoBERTa**, and fine-tune a pre-trained model using the HuggingFace Transformers library. We then perform an extrinsic evaluation of the sentiment analysis model using **finBERT model** as a benchmark, ensuring its efficacy in the domain.

Subsequently, **we merge the sentiment scores obtained from the sentiment analysis model with past news scores and stock datasets**. This integration creates a comprehensive dataset that captures both sentiment trends and stock market movements over time. Utilizing this dataset, we employ a random forest regressor to generate stock price predictions.

Finally, **we integrate the sentiment-enhanced stock predictions into the Bard API**. By providing the API with prompts tweaked according to the inputs by user, giving the summarized history of a desired company and the sentiments of a current news inputted by user (or one can only ask for its past track record), we enable it to act as a financial advisor, offering investment advice based on the combined insights from sentiment analysis, past news study and stock prediction. **This methodology ensures the creation of a robust chatbot capable of providing personalized and informed investment recommendations to users, leveraging sentiment analysis, machine learning, and natural language processing techniques.**

5. Evaluation Criteria

Stock prediction accuracy can be assessed using standard metrics like accuracy, precision, recall, and F1 score. Additionally, consider:

- Confidence scores for evaluating strength/weakness of predictions and flexibility of sentiment analysis.
- Extrinsic evaluation by comparing model predictions with external models (e.g., FinBERT) using Mean Reciprocal Rank.
- Relevance baseline: Set a baseline for article relevance to the search query and measure precision of results.
- Financial regulations and security: Ensure model outputs comply with regulations and user data is handled securely.

6. Novelty

Our chatbot "StockWorthy" stands out with four key features:

Interactive User Interface with Personalized News Feeds: The chatbot offers a user-friendly interface with tailored news feeds focused on stocks of interest, enhancing engagement and informed decision-making.

Integration of Historical and Sentiment Analysis: Combining historical stock price analysis with sentiment analysis on news articles improves prediction accuracy, providing nuanced insights into potential stock movements.

Confidence Score for Predictions: By assigning confidence scores to predictions, users can assess associated risks, empowering them to make informed decisions aligned with their financial goals.

Personalized Investment Advice: Offering customized investment advice tailored to individual portfolios and objectives enhances user experience and facilitates effective portfolio management.

These features collectively establish "StockWorthy" as a comprehensive, user-friendly chatbot, pioneering the integration of historical data and real-time news trends for insightful stock analysis.

7. Potential Contributions

- Aditya Dahiya: Development of chatbot interface and using APIs.
- Ankush Gupta : Performing sentiment analysis of news articles
- Ayush Srivastava : Application of APIs to merge and fine-tune ML/DL models
- Harshvardhan Singh : Designing algo to merge chatbot model with finance.
- Shantanu Prakash : Development of chatbot interface, training and evaluation of models.

The actual contributions may vary from the potential contributions mentioned here.

8. 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