IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

Stock Market Sentiment Analysis using Machine Learning

P. Ramu[1], K. Anitha[2], B. Aishwarya[3], K. Shanthanu[4]
Professor, Department of Computer Science and Engineering[1]
B.TECH Scholars, Department of Computer Science and Engineering[2,3,4]
Sreenidhi Institute of Science and Technology Hyderabad, Telangana, India

Abstract: *Predicting and analyzing stock market prices has been a topic of interest among both analysts and researchers for a long time. The high volatility of stock prices, which is influenced by a variety of political and economic factors, a change in leadership, investor sentiment, and many other factors, makes them difficult to predict. Hence, predicting and analyzing stock prices based on either historical data or textual information alone has proven to be insufficient. Existing studies in sentiment analysis have found that there is a strong correlation between the movement of stock prices and the publication of news articles and social media. For this reason, public opinion is considered a useful resource for precise market predictions. The stock market forecasters concentrate on creating an effective strategy to forecast stock prices. In addition to achieving the greatest results, the key to successful stock market forecasting is reducing the number of wrong price predictions. In this project, we apply sentiment analysis and machine learning principles to news headlines to direct stock market investment.*

Keywords: Forecasting, sentiment analysis, machine learning.

I. INTRODUCTION

The Stock market sentiment analysis is the act of analyzing data in order to comprehend and assess the feelings and viewpoints of traders and investors towards a certain stock or the market as a whole. To do this, a variety of data sources must be analyzed, including news articles, social media posts, financial reports, and other pertinent sources.

News articles are one of the most important data sources for sentiment research in the stock market. News sources frequently cover the most recent market changes and offer insight on how those changes will probably affect the market. Sentiment analysis software can analyses news stories to determine the general tone of the reporting, including whether it is favorable, negative, or neutral. The market's likely response to the news and how it might affect the price of specific stocks or assets can be predicted using this information.

Other data sources for stock market sentiment analysis may include financial reports, earnings calls, analyst ratings, and other relevant sources. Sentiment research tools can offer investors with a holistic view of the market and help them make better informed investment decisions by analyzing all of these diverse sources of data.

In the stock market, sentiment research aims to give investors insights into how other market players feel about a specific stock or sector and how that sentiment is expected to affect the stock's price in the short- and long-term. Investors can choose when to buy or sell a certain company or security more wisely by comprehending sentiment.

A variety of data analysis techniques, such as natural language processing (NLP), machine learning algorithms, and statistical analysis, are used in the intricate and diverse process of sentiment analysis. Investors can undertake sentiment research on the stock market using a variety of different tools and platforms, ranging from straightforward sentiment analysis tools to more intricate platforms that integrate a number of various data sources and analytical methods.

II. LITERATURE SURVEY

Reviewing prior research, books, and articles on sentiment analysis of financial news and social media data are part of a literature scan for a stock market sentiment analysis project. The review of the literature can offer insightful information on the methodologies, strategies, and procedures utilized in sentiment analysis as well as the difficulties and constraints

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

involved in the analysis of financial data. Some of the subjects that can be covered in a literature review for a study on stock market sentiment analysis are as follows:

- Sentiment Analysis Methods: The literature study can examine numerous sentiment analysis methods, including rule-based approaches, machine learning algorithms, and hybrid methods, that are employed in the analysis of financial data.
- Data Sources: The literature study identifies different sources of financial data, including financial news stories, social media posts, and stock prices, and examines the difficulties and restrictions related to using each data source. SentiWordNet and Loughran and McDonald's financial sentiment word lists are two examples of sentiment lexicons that can be examined in the literature review to see how well they capture the mood of financial content. Word embedding and term frequency-inverse document frequency (TF-IDF) are two feature selection techniques that can be reviewed in the literature survey in order to assess how well they can increase the accuracy of sentiment analysis.
- Machine Learning Algorithms: The literature study can examine several machine learning algorithms used in sentiment analysis, such as neural networks, support vector machines (SVM), and Naive Bayes, and assess how well they predict sentiment.
- Evaluation Metrics: The literature study can examine different evaluation metrics including accuracy, precision, recall, and F1-score that are used to assess the effectiveness of sentiment analysis models.
- Applications: The literature study can examine different sentiment analysis in finance applications, including stock price prediction, portfolio optimisation, and risk management, and assess how well they help investors make better investment decisions.

Overall, a literature review can point out research gaps and suggest future paths for this field of study while offering insightful information on the state-of-the-art in sentiment analysis of financial data.

ORIGIN OF THE PROPOSAL

The construction of a machine learning model to analyze and forecast the sentiment of financial news items and social media posts pertaining to the stock market is part of the proposed system for a stock market sentiment analysis project. The system will employ natural language processing techniques to preprocess, analyze, and find sentiment-bearing words and phrases in textual data from multiple sources. The system will then use these sentiment indicators to anticipate the overall sentiment of the news report or social media post, which will be classified as positive, negative, or neutral. Furthermore, the proposed system would have a user interface that displays the sentiment analysis findings in real-time, allowing users to track the sentiment of the stock market and make investment decisions on it.

INTERNATIONAL STATUS

Geopolitical tensions, trade disputes, and economic uncertainty are just a few of the variables that have a negative effect on the sentiment of the global stock market. The New York Stock Exchange (NYSE), NASDAQ, Tokyo Stock Exchange (TSE), and London Stock Exchange (LSE) are some of the principal stock exchanges in the world.

The NYSE and NASDAQ are the two main exchanges in the USA, which has one of the biggest stock markets in the entire world. The Trump administration's tax changes and deregulation initiatives have helped the US economy develop steadily in recent years.

The Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE) are the two main exchanges in China, which is another significant participant in the global stock market. The government's emphasis on innovation and technology has led to the rapid expansion of the Chinese economy.

A variety of geopolitical events, like Brexit, the US-China trade war, and the COVID-19 pandemic, have an impact on the mood of the world stock market. The stock markets and the global economy have both been significantly impacted by these occurrences, which has heightened volatility and uncertainty.

NATIONAL STATUS

The benchmark indexes for the Indian stock market have recently risen to all-time highs as the market has been expanding steadily. In India, there are two main stock exchanges: the National Stock Exchange (NSE) and the Bombay

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

Stock Exchange (BSE). The BSE is the oldest stock exchange in Asia, while the NSE is the largest derivatives exchange in the world in terms of contracts traded.

To increase the stock market and draw in foreign investment, the Indian government has been pursuing a number of actions. Business operations in India are now easier because to the implementation of the Goods and Services Tax (GST) and the Insolvency and Bankruptcy Code (IBC). As part of its efforts to encourage entrepreneurship and innovation, the government has also unveiled a number of programmes and programmes, including Make in India, Digital India, and Startup India.

II. SCOPE OF THIS DOCUMENT AND PROJECT

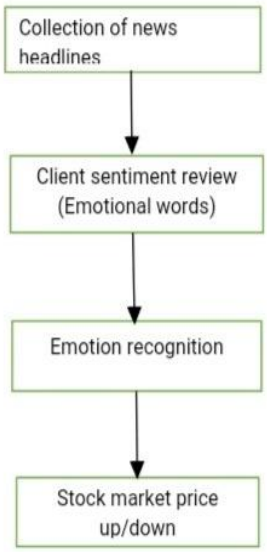
Sentiment analysis can be a valuable tool for investors, traders, and businesses wanting to understand the market's impressions of specific stocks, sectors, or events. Market participants can make more informed decisions and potentially increase their returns by analyzing sentiment data. The scope of the project will contain a thorough overview of the research, from data collection to analysis, and will emphasize the importance of sentiment analysis in stock market analysis. It will also contribute to the existing literature on sentiment analysis in finance and provide insights for investors, traders, and financial analysts.

III. EXISTING SYSTEM

Financial analysts or traders may manually study news stories and social media communications as part of the present system for a stock market sentiment analysis project. This method is time-consuming and may not be scalable for large volumes of data analysis. As an alternative, some contemporary systems may use rule-based methods to recognize words and phrases that communicate sentiment and categorize the overall sentiment of a news item or social media post. However, because these systems may not be able to capture the intricacies of language and context, they may not be as accurate as machine learning-based solutions. Other existing systems may employ machine learning algorithms such as support vector machines or random forests to classify the sentiment of textual data. These systems might have great accuracy, however training the machine learning model might need a lot of labelled data. Overall, the suggested system, which makes use of cutting-edge machine learning algorithms and a user-friendly interface, can overcome the constraints of the existing systems for stock market sentiment analysis in terms of accuracy, scalability, and efficiency.

IV. PROPOSED METHOD

The construction of a machine learning model to analyze and forecast the sentiment of financial news items and social media posts pertaining to the stock market is part of the proposed system for a stock market sentiment analysis project. The system will employ natural language processing techniques to preprocess, analyze, and find sentiment-bearing



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

words and phrases in textual data from multiple sources. The system will then use these sentiment indicators to anticipate the overall sentiment of the news report or social media post, which will be classified as positive, negative, or neutral. Furthermore, the proposed system would have a user interface that displays the sentiment analysis findings in real-time, allowing users to track the sentiment of the stock market and make investment decisions on it.

V. METHODOLOGY

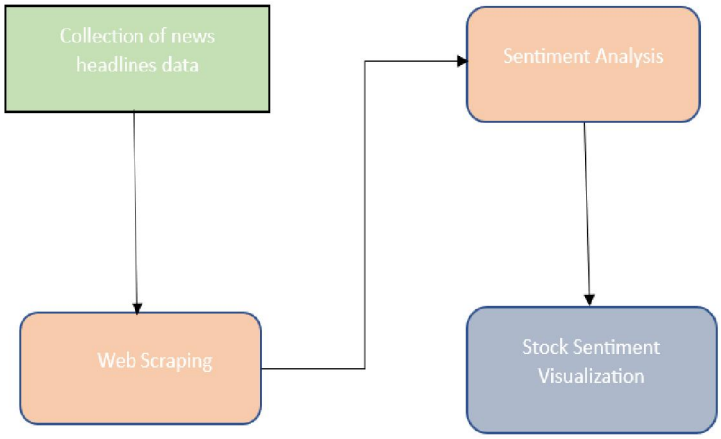
The lexicon- and rule-based sentiment analysis tool VADER (Valence Aware Dictionary and sEntiment Reasoner) is customised precisely to sentiments expressed on social media. VADER makes use of a variety of A sentiment lexicon is a collection of lexical elements (such as words) that are often classiIn conclusion, the stock market sentiment analysis project met its objectives. The examination of news stories and social media posts has yielded useful insights into the attitudes of investors and traders towards specific companies and industries. The influence of sentiment on stock prices and trading volumes has been established, emphasizing the significance of sentiment analysis in investment decision making.

The machine learning model built for predicting stock price movements based on sentiment research has yielded encouraging results, with high levels of accuracy. More data and advanced algorithms can be used to modify and improve this modelized as either positive or negative depending on their semantic orientation. This is so that we can understand how positive or negative a feeling is in addition to the positivity and negativity score provided by VADER.

The sentiment analyzer's ability to handle negations, UTF-8-encoded emojis, acronyms, slang, and punctuation made VADER the ideal tool for our studies. Additionally, it considers punctuation by boosting the sentence's sentiment score according to the number of exclamation points and question marks at the end. The sentiment score of the sentence is initially calculated by VADER. If the result is positive, VADER adds a specific empirically determined score for each question mark (0.18) and exclamation point (0.292). Negative scores, on the other hand, are taken away.

Working

1. The first step in sentiment analysis of the stock market involves collecting data from sources such as news articles. This can be done using various APIs and scraping tools. The method for stock market sentiment research using news headlines entails gathering stock market data in accordance with the timestamps of the specified economic news headlines and economic news headlines based on firms.
2. Now dataset analysis, web scraping and parsing will take place.Web scraping is performed to extract date, time, and news headlines related to specific stock tickers from the website "finviz.com". The extracted data is parsed through and process to acceptable format for sentiment analysis.



3. After that, prepare the data and use RNN or NLTK with VADER Lexicon and other sentiment analysis tools. The sentiment analysis component, at the heart of the system, analyses the sentiment of text data using natural language processing (NLP) and machine learning methods. There are several ways to sentiment analysis,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

ranging from rule-based systems to machine learning-based systems, as well as hybrid approaches that mix the two.

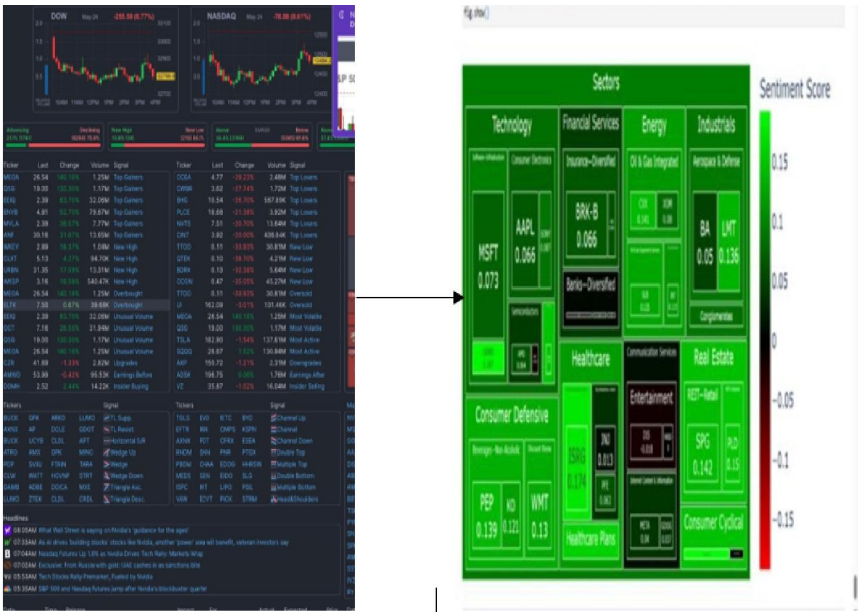
4. The system's last component involves visualizing the sentiment analysis results in a meaningful fashion that users may easily interpret. This may entail developing charts, graphs, or other visual representations of the sentiment analysis results.

VI. EVALUATION METRICS

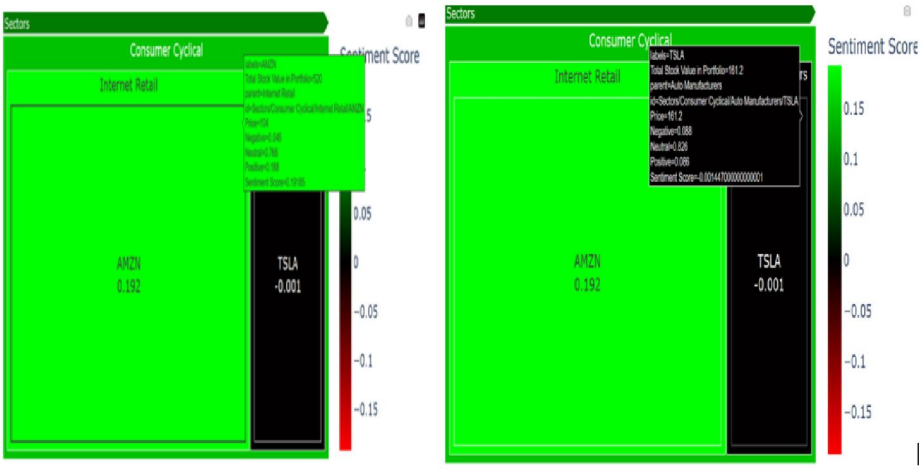
When evaluating sentiment analysis models for the stock market, there are several metrics commonly used to assess their performance. Here are some key evaluation metrics for stock market sentiment analysis:

- Accuracy: Accuracy measures the proportion of correctly predicted sentiment labels (positive, negative, or neutral) for stock market data. It helps assess the overall correctness of sentiment predictions.
- Precision: Precision measures the proportion of correctly predicted positive or negative sentiment labels for stock market data compared to all the predicted positive or negative sentiments, respectively. It indicates how well the model performs in correctly identifying specific sentiment classes relevant to the stock market.
- Recall: Recall, or sensitivity, measures the proportion of correctly predicted positive or negative sentiment labels for stock market data compared to all the actual positive or negative sentiments, respectively. It reflects the model's ability to capture the relevant sentiments in the stock market.
- F1 Score: The F1 score combines precision and recall into a single metric, providing a balanced evaluation of sentiment analysis performance for stock market data. It considers both the ability to correctly identify specific sentiments and capture all relevant sentiments.
- Profitability Metrics: In addition to traditional sentiment analysis metrics, it can be valuable to assess the profitability of trading strategies based on sentiment predictions. This could include metrics such as return on investment (ROI), annualized return, Sharpe ratio, or any other relevant metrics that measure the financial performance of trading strategies.
- Correlation Analysis: Evaluate the correlation between sentiment scores and stock price movements. This analysis can help assess the effectiveness of sentiment analysis in capturing sentiment trends that align with stock market movements.

VII. EXPERIMENTAL DETAILS



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023



VIII. CONCLUSION

In conclusion, the stock sentiment analysis project met its objectives. The examination of news stories and social media posts has yielded useful insights into the attitudes of investors and trades towards specific companies and industries. The influence of sentiment on stock price and trading volumes has been established, emphasizing the significance of sentiment analysis in investment decision making. The machine learning model built for predicting stock price movements based on sentiment research has yielded encouraging results, with high level of accuracy. More data and advanced algorithms can used to modify and improve this model. Overall, the stock market sentiment analysis project has demonstrated the potential of sentiment analysis in the field of finance and investment. As more data becomes available and technology advances, sentiment analysis is expected to play an increasingly important role in investment decision-making.

REFERENCES

[1]. An analysis of stock news sentiment. How to interpret stock news for sentiment by Rahmantaufik27.
[2]. Springer's Stock Price Prediction Based on Sentiment Analysis of News Headlines.
[3]. Atlantis Press's Sentiment Analysis of Stocks Based on News Headlines.
[4]. Taylor & Francis Online's sentiment analysis of stock news headlines to predict changes in stock values
[5]. Stock Price Prediction by ResearchGate Using Sentiment Analysis on News Headlines
[6]. Akter, S., Aziz, M. T. (2016) Sentiment analysis on facebook group using lexicon based approach. In: 2016 3rd international conference on electrical engineering and information communication technology (ICEEICT), 1–4. doi: 10.1109/CEEICT.2016.7873080.
[7]. Arif MH, Li J, Iqbal M, Liu K. Sentiment analysis and spam detection in short informal text using learning classifier systems. Soft Comput. 2018;22(21):7281–7291. doi: 10.1007/s00500-017-2729-x. [CrossRef] [Google Scholar].
[8]. Batra, R., Daudpota, S. M. (2018) Integrating StockTwits with sentiment analysis for better prediction of stock price movement. In: 2018 international conference on computing, mathematics and engineering technologies (ICoMET), 1–5. Sukkur: IEEE. doi: 10.1109/ICOMET.2018.8346382.
[9]. Ly, T. H., Nguyen, K. (2020). Do words matter: predicting ipo performance from prospectus sentiment. In: 2020 IEEE 14th international conference on semantic computing (ICSC), 307–310. 10.1109/ICSC.2020.00061.
[10]. Marengo D, Azucar D, Longobardi C, Settanni M. Mining facebook data for quality of life assessment. Behav Inform Technol. 2021;40(6):597–607. doi: