EQUITY MARKET SENTIMENT ANALYSIS USING NLP TECHNIQUES

Rekha A

Assistant Professor, Department of Information Technology, Panimalar Engineering College
rekhapecit2023@gmail.com

Aruna S

Department of Information Technology, Panimalar Engineering College
shanmuganathanaruna2002@gmail.com

Harini T

Department of Information Technology, Panimalar Engineering College
harinithangamani14@gmail.com

Lavanya K

Department of Information Technology, Panimalar Engineering College,
lavanyakumar0529@gmail.com

Abstract—Financial decision-making requires an understanding of equity market sentiment since investor opinion frequently affects stock prices and market trends. The proliferation of social media and online financial forums has resulted in a wealth of textual data that offers insightful information on the current state of the market. This presentation presents an analysis of the sentiment of the equity market using Natural Language Processing (NLP) techniques. The aim of this project is to develop a robust framework that leverages textual data from several sources, including investor forums, social media posts, and financial news items, in order to forecast and assess sentiment in the equities market using Natural Language Processing (NLP) techniques. The stages of the proposed approach are as follows: gathering data, pre- processing it, extracting features, sentiment analysis, and prediction.

Keywords—Equity market, sentiment analysis, Natural Language Processing (NLP), data pre-processing, feature extraction, machine learning, deep learning, sentiment prediction, market trends, stock prices.

I.   INTRODUCTION

Making wise investing decisions in the dynamic financial markets requires a solid understanding of investor attitude. Together, market participants' attitudes and viewpoints have a big influence on the equity market, which is esteemed for its static volatility and swings. In order to achieve deeper insights, traditional market analysis strategies frequently struggle to capture the subtleties

of investor emotion. Thus, cutting-edge technologies are being integrated, such as natural language processing (NLP) and machine learning.

Market participants may now extract valuable information from significant amounts of unstructured textual data, such as financial news, social media communications, and analyst reports, through NLP techniques. Through the application of language patterns and sentiment analysis, investors are able to ascertain the current state of the market, recognize new trends and predict future moves in the market. The integration of technology and finance improves decision-making and provides new perspectives on the behavioural characteristics of the equity markets.

In order to understand the complex interaction between textual data and market dynamics, this paper explores the field of equity market sentiment analysis using Natural Language Processing (NLP) approaches. By delving into the subtleties of investor emotion, we hope to add to the expanding body of knowledge and offer insightful commentary on the changing financial markets environment.

II.   LITERATURE REVIEW

Aastha Saxena,Sparsh Singla, Arpit Jain, Amrita Ticku and Prateek Sharma, [2] discussed the sentiment analysis techniques on news headlines published as "Sentiment Analysis of Stocks Based on News Headlines Using NLP". Today's society has everyone, from kids to adults, researching stocks and looking for ways to increase their income by understanding market trends. The stock market is an intricately linked network of different investors. Because of its rapid fluctuations, it is difficult to forecast what lies ahead. To boost earnings, businesses all across the world rely on these projections and conjectures. The financial news of the day has a major effect on determining whether stock prices rise or fall. The investor improves significantly from the market news' abundance of information when deciding how much to bid. For the purpose of this study, we take relevant information out of a company's news headlines and examine how it affects the company's stock price immediately. Using a text-based dataset, we apply natural language processing (NLP) to compare two methods that employ distinct algorithms. These methods together ascertain the overall sentiment of the news headline, in comparison to the company's stock prices, and whether it is positive or negative.

Om mane and Sarvanakumar Kandasamy

[3]   performed a survey on stock market prediction published as "Stock market prediction using Natural Language Processing -A survey". The network that serves as a platform for practically all significant economic activities is the stock market. Purchasing individual stocks may not be a wise move, even though investing in the stock market is, especially for novice investors.Thorough research and a great deal of effort are required for prudent stock selection. There are countless possibilities of potential for arbitrage profit when predicting this stock value. The allure of discovering a solution has driven researchers to devise strategies to overcome problems including volatility, seasonality, and temporal dependence. This paper looks at recent studies on natural

language processing and machine learning techniques for stock market prediction. This paper's major contributions are the detailed classifications of numerous recent papers and the depiction of current developments in research in the field of stock market prediction and adjacent fields.

Ayman E. Khedr, S.E. Salama, Nagwa Yaseen.et.al.,[5] "Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis" Because stock market prediction is beneficial and plays a big part in the economy, it has become a major subject of study. In order to reduce the dangers involved in investing, one must be informed about future stock market trends.A wealth of information for shareholders is thought to be produced by the stock market's massive volume of data. The research's objective is to create a

predictive model with a low error ratio that can predict future stock market movements and improve prediction accuracy. Based on sentiment analysis of past stock prices and financial news, this prediction model was created. Because it takes into account a variety of news sources relevant to the market and the firm with historical stock prices, Compared to earlier research, this model produces outcomes that are more accurate. Employed is a dataset with three firms' stock prices. The sentiment analysis of the news must be done using the naïve Bayes algorithm before it can establish the text polarity. Between 72.73% and 86.21% of the predictions were made correctly in this step. To predict stock prices in the future, the second level integrates news polarities and past stock prices. Predictive accuracy increased to 89.80% as a result.

Kalyani joshi, Prof. Bharathi H. N , Prof. Jyothi rao[1] published the paper related to sentiment analysis named "Stock Trend Prediction using News Sentiment Analysis". The widely accepted hypothesis on stock prediction is called the Efficient Market Hypothesis. Because stock prediction has been so unsuccessful, a lot of research has been done in this area. The goal of this research is to use news sentiment categorization to forecast future stock trends using non-quantifiable data, such as financial news stories about a company. Assuming that news items have an impact on the stock market, this is an attempt to investigate the relationship between news and stock trends. We created three different classification models to show whether news items are polarized positively or negatively in order to highlight this. Results indicate that RF and SVM function well across all testing scenarios. Although not as good as the other

two, Naïve Bayes produces good results. Research studies are carried out to assess several facets of the suggested model, and positive outcomes are found in each test. The prediction model's precision exceeds 80%, and it has improved accuracy by 30% when contrasted with news random labels, which has an accuracy of only 50%.

Arpit Goel and Anshul Mittal [4], "Stock Prediction Using Twitter Sentiment Analysis": To ascertain the correlation between "market sentiment" and "public sentiment," they employ machine learning and sentiment analysis techniques.." We utilize DJIA values from prior days and public

mood estimations from Twitter to forecast stock market moves. We propose a novel cross-validation method for financial data to validate our results, and we use Self- Organizing Fuzzy Neural Networks (SOFNN) to achieve 75.56% accuracy on the DJIA values and Twitter feeds between June and December 2009. On the basis of our anticipated values, we also put into practice a naive portfolio management technique. The renowned work by Bollen et al., which made an 87% accurate prediction, served as the foundation for our effort.

III.   EXISTING SYSTEM

Because of the volatility of financial markets, Stock Movement Prediction (SMP) attempts to forecast the future price movement of stocks of listed businesses, a difficult undertaking. Current research in finance indicates that the momentum spill over effect contributes significantly to stock volatility. To represent the intricate relationships of listed firms in the actual financial market, prior research has, however, generally only learned the basic connection information among connected organizations. We first create a more thorough Market Knowledge Graph (MKG) to solve this problem. This MKG comprises hybrid relationships, such as explicit and implicit linkages, as well as bi- typed entities, such as listed firms and the executives connected to them. Next, we suggest using a new Dual Attention Network called DANSMP to identify momentum spillover signals using the built-in MKG for stock prediction.

DEMERITS:
- They did not implement the deployment process.
- They did not compared more than an algorithms to getting better accuracy level.
- Accuracy was low.

IV.   PROPOSED SYSTEM

The initial step in the recommended approach is to collect significant textual data from several sources, such as financial news articles, social media platforms, and investment forums. Sentiment Analysis: In this phase, machine learning and deep learning models are implemented to classify the sentiment of the extracted features. Among the methods that can be deployed are support vector machine algorithms (SVM), Naive Bayes, Recurrent Neural Networks (RNN), and Transformers like BERT or GPT.

To discover patterns and correlations between the textual properties and sentiment categories (positive, negative), these models are trained using labelled sentiment datasets. Sentiment Prediction and Trend Analysis: In order to predict sentiment and analyze how it affects equity markets and shares, the suggested system uses time-series analysis techniques. Sentiment values can be predicted using techniques such as LSTM models or autoregressive integrated moving average (ARIMA).

Users can make well-informed investing decisions by identifying patterns and correlations by analysing the relationship between sentiment and market data. To improve accuracy, we compared more than two methods. We accurately determine the worth of the performance and confusion

measurements. Construct an application for the goal of deployment. Performance level and accuracy both increased.

MERITS:
- We compared more than two algorithms to getting better accuracy level.
- We figure out performance and confusion metrics value properly.
- Build an application for deployment purpose.
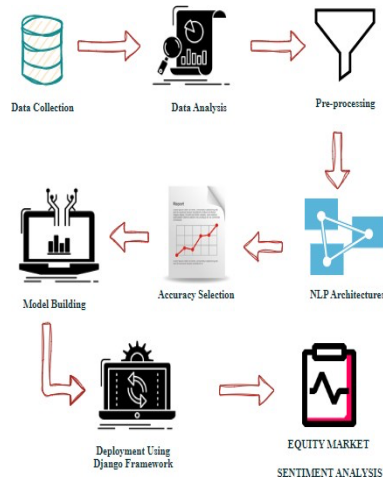- Accuracy & performance level improved.

Fig 1: Architecture diagram

V. METHODOLOGY

This section discusses the model's design that this study proposes. It also includes details on the dataset that was used for the training and testing of the model. It offers comprehensive details on BERT, the dataset, and the preprocessing techniques that were applied before the model was created.

Dataset:

Collect and preprocess monetary news stories and postings on social media concerning the equity market for sentiment analysis using NLP techniques.

News Collection:

The news headlines dataset is being used for this study. The world news and stock price fluctuations that are available on Kaggle are combined to create the dataset.

Preprocessing:

There could be missing values in the obtained data, which could cause inconsistencies. Preprocessing data is necessary to increase the algorithm's efficiency and produce better outcomes. It is necessary to eliminate the outliers and do variable conversion.
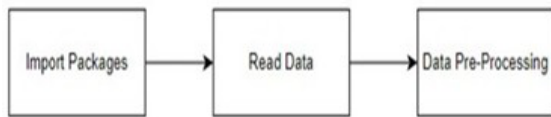
Fig 2: Steps for data pre-processing

Unstructured data is textual data. Therefore, we are unable to feed the classifier with unprocessed test data. To work at the word level, we must first tokenize the document into words. There

are more noisy terms in next data that don't help with classification. Thus, we must discard those terms. Text data may also include stop words, additional white space, tabs, digits, and punctuation to purge the data of all those words. A stop-word list was created for the finance world and general English words, including generic names, dates, numbers, geographic and currencies. Words were considered minimum document frequency and stemming was used to reduce redundancy. Preprocessing was done before and after applying the detection algorithm.

Building the classification model
Sentiment Analysis: In this stage, Artificial intelligence methods include deep learning, the sentiment of the retrieved features is classified using machine learning and natural language processing algorithms.

Fig 3: Building the classification model

Information Science
Combines information from data with practical insights in a variety of application domains, making it an interdisciplinary field. It achieves this by drawing information and insights from both structured and unstructured data using scientific processes, systems, algorithms, and methods. Applying math, business sense, instruments, algorithms and machine learning techniques is what is meant to be understood as data science to uncover hidden patterns or insights from unprocessed data. These findings can be crucial to important business choices.
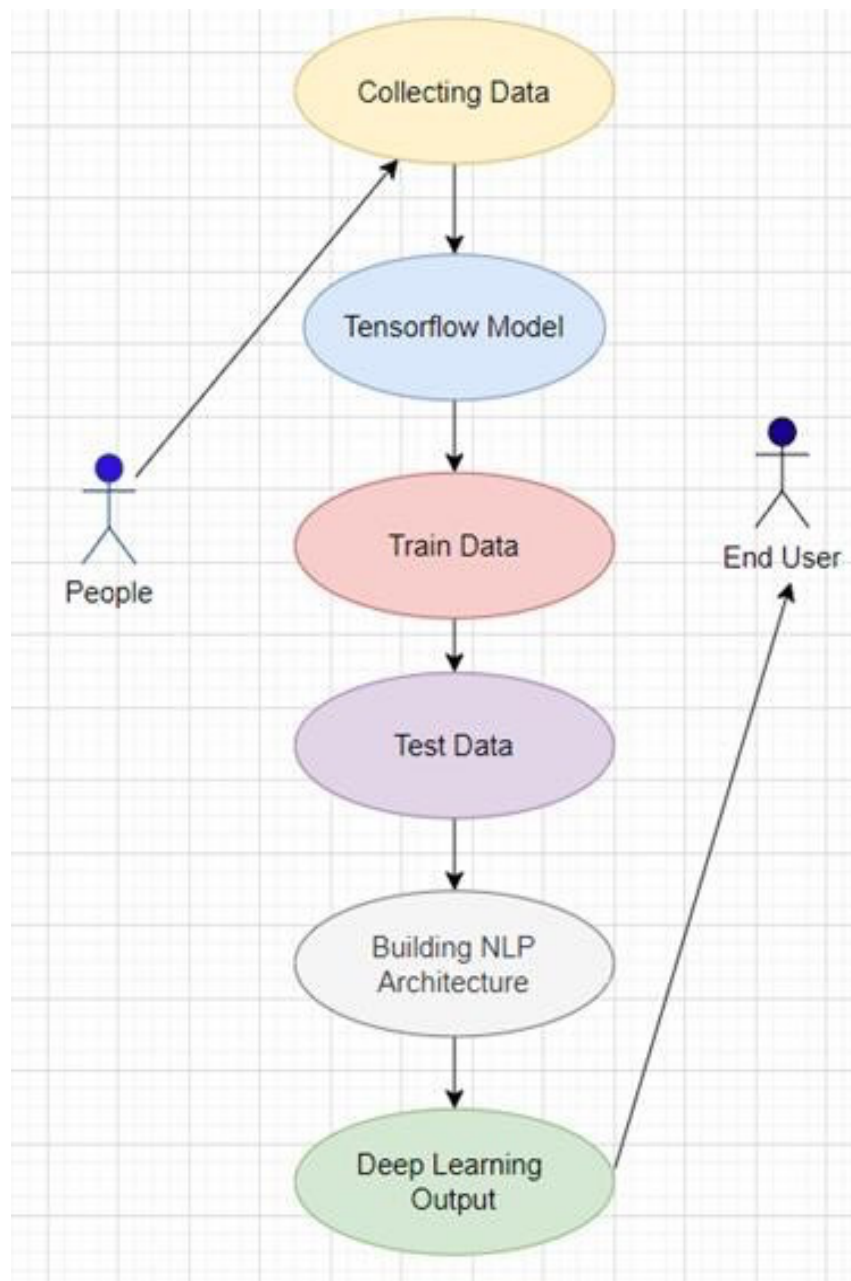
Fig 4: Use case diagram of using NLP model

Artificial Intelligence

Programming for AI emphasises the three cognitive abilities of self-repair, learning and reasoning. Studying procedures involve acquiring information and creating rules for converting it into knowledge that may be used, Selecting the appropriate method for a desired result and self-correction are the main goals of reasoning processes.

Technologies related to deep learning and artificial neural networks are developing quickly due to their ability to swiftly process massive amounts of data and reliably than is humanly possible when making predictions.

Natural Language Processing

NLP is a potent technique that makes it possible for machines to comprehend human language and acquire knowledge from human written sources. It has applications in information retrieval, machine translation, question-answering and text mining. In this research the NLP techniques are used to retrieve accuracy from the textual data of financial news.

Machine Learning

ML is an AI technology that enables computers to learn without explicit programming. Data Scientists use various machine learning techniques to identify trends in Python that provide useful information. Classification is a supervised learning approach that predicts the class of given data points, using techniques like Decision trees, logistic regression, multi- class classification, and support vector machines.

Supervised Machine Learning is a common practice in machine learning, where an algorithm is used to translate the input variables(X) and output variables(Y) are mapped using an algorithm to predict new data based on the mapping function
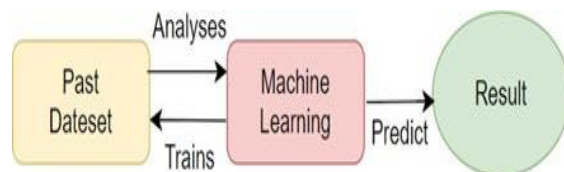
$Y = f(X)$

Fig 5: Process of Machine Learning

Building an Anticipatory Model

Compiling large amounts it takes historical data to do machine learning. Enough is said raw and historical information that can be gathered. There is no direct use for raw data prior to data preparation.

This model has been trained and tested to ensure that it produces little mistakes in its reliable forecasts. To improve the accuracy of the model, it was periodically modified. Below diagram depicts description of the key components and steps typically found in a machine learning workflow diagram.
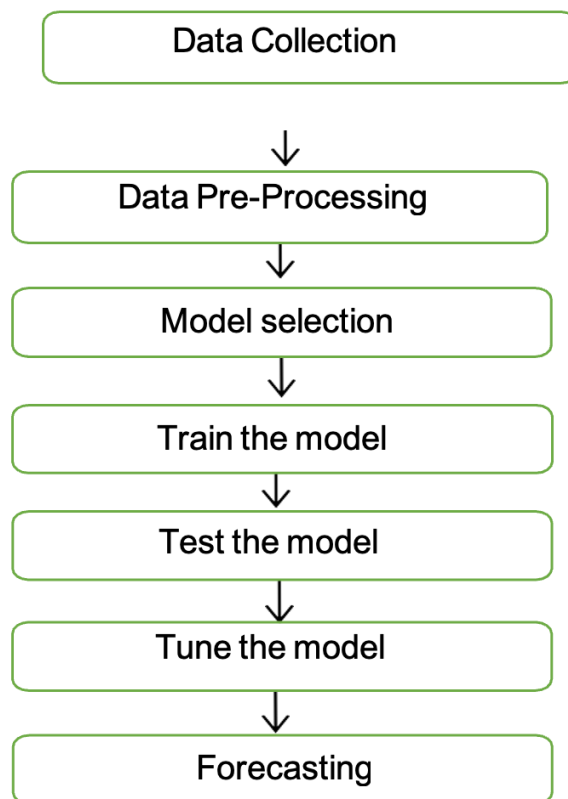
Fig 6: Diagram of workflow process

Multinomial Naive Bayes Algorithm

In natural language processing, this probabilistic classification technique is employed particularly text classification. Our research input is in text format so that this algorithm is used. It determines the likelihood that a document belongs to a particular class using the Bayes theorem.
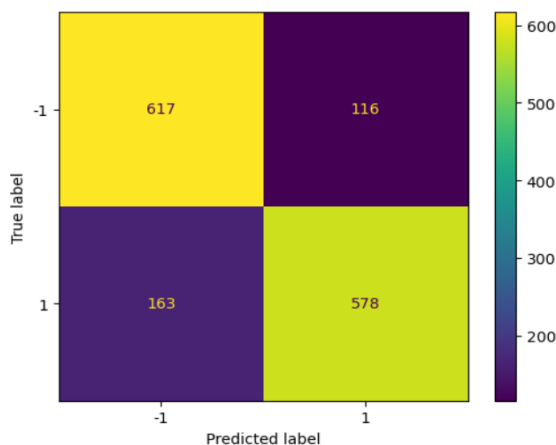
Fig   7:   Confusion   matrix based on Multinomial NB Algorithm

Logistic Regression Algorithm

A statistical technique called logistic regression is employed in machine learning to predict binary classification outcomes based on input features, utilizing the logistic function for interpretability and computational efficiency.

Random Forest Algorithm

A powerful machine learning approach that makes use of decision trees is the random forest classifier to handle complex datasets, improving generalisation and mitigating overfitting, making it widely used in finance, healthcare and image recognition.

Deployment

Deploying the model and forecasting the result in the Django Framework. After being trained, the deep learning model is stored as a file in the hierarchical data format (.h5 file), which is subsequently installed in our django framework to improve user interface and forecast whether the input picture is CKD /Not CKD.

Django

High-level Python web framework Django provides safe, maintainable websites and is built for quick website creation, a thriving community, excellent documentation and various support options for both free and paid users.

VI.   CONCLUSION

In summary, using Natural Language Processing (NLP) tools to analyse sentiment in the equity market yields priceless insights about the attitudes of investors and the state of the market. Investors can make more educated decisions by utilising Natural Language Processing (NLP) to gather and evaluate textual data from social media postings, financial reports, and news articles.

This allows for a more nuanced view of market sentiment. This novel method not only increases the effectiveness of market analysis but also helps create predictive models that are flexible enough to adjust changing market conditions. In the end, there is much space for development in decision making procedures and investment outcome maximisation through the integration of NLP in equities market sentiment analysis. To increase the accuracy and complexity of sentiment predictions for the equities market, we can investigate more sophisticated sentiment analysis models in the future, like BERT or GPT-3. Enhance the timeliness of equities market sentiment Assessments by incorporating social media sentiment research and real-time data sources to capture dynamic market feelings.

VII. REFERENCES

[1]   Kalyani Joshi, Prof. Bharathi H. N. , Prof. Jyothi Rao (2016) "Stock Trend Prediction using News Sentiment Analysis" International Journal of Computer Science and Information Technology 8(3):67-76 DOI:10.5121/ijcsit.2016.8306

[2]   Aastha Saxena, Arpit Jain, Prateek Sharma, Sparsh Singla, and Amrita Ticku (2023) "Sentiment Analysis of Stocks Based on News Headlines Using NLP" V.

E. Balas et al. (Eds.): AITEES 2022, AHIS 3,   pp.   124–135,   2023. https://doi.org/10.2991/978-94-6463-074- 9_12

[3]   Om mane and Sarvanakumar kandasamy (2022) "Stock market prediction using Natural Language Processing -A survey" David C. Wyld et al. (Eds): NLPA, AIS, BDAP, SOENG, IPPR - 2022 pp. 33-48, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121404

[4]   Anshul Mittal and Arpit Goel (2011) "Stock Prediction Using Twitter Sentiment Analysis" Standford University, CS229 (2011   http://cs229.

stanford.edu/proj2011/GoelMittalStockMa rketPredictionUsingTwitterSentimentAnal ysis.pdf)

[5]   Ayman E. Khedr. et.al., S.E.Salama and Nagwa Yaseen.et.al.,(2017) "Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis" I.J. Intelligent Systems and Applications, 2017, 7, 22-30 Published Online July 2017 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijisa.2017.07.03

[6]   Sheikh Mohammad Idrees, M. Afshar Alam, and Parul Agarwal (2019) "A

Prediction Approach for Stock Market Volatility Based on Time Series Data" Digital   Object   Identifier 10.1109/ACCESS.2019.2895252

[7]   S. AL Wadi , Mohammad Almasarweh & Ahmed Atallah Alsaraireh (2018) "Predicting Closed Price Time Series Data Using ARIMA Model" Modern Applied Science; Vol. 12, No. 11; 2018 ISSN 1913-1844E-ISSN 1913-1852 doi:10.5539/mas.v12n11p181   URL: https://doi.org/10.5539/mas.v12n11p181

[8]   Fazlija, B.; Harder, P. (2022) "Using Financial News Sentiment for Stock Price Direction Prediction" Mathematics 2022, 10,   2156.   https:// doi.org/10.3390/math10132156

[9]   Karlo Puh and Marina Bagic Babac (2023) "Predicting stock market using natural language processing" American Journal of Business Vol. 38 No. 2, 2023 pp. 41-61 Emerald Publishing Limited 1935- 5181 DOI 10.1108/AJB-08-2022-0124

[10]   B   L,   S. and B   R,   S. (2023), "Combined deep learning classifiers for stock market prediction: integrating stock price and news sentiments", Kybernetes, Vol. 52   No. 3,   pp.   748-   773.   https://doi.org/10.1108/K-06-2021- 0457

[11] Ji, X., Wang, J. and Yan, Z. (2021), "A stock price prediction method based on deep learning technology", International Journal of Crowd Science, Vol. 5 No. 1, pp. 55-72.

[12]

Kameshwari, S., Kaniskaa, S., Kaushika, S

. and Anuradha, R. (2021),   "Stock trend

prediction using news headlines", 2021 IEEE India Council International Subsections
Conference (INDISCON),pp. 1-5,
doi: 10.1109/INDISCON53343.2021.9582 219.