Revue d'Intelligence Artificielle
Vol. 38, No. 3, June, 2024, pp. 999-1008
Journal homepage: http://iieta.org/journals/ria

Sentiment Analysis and Stock Data Prediction Using Financial News Headlines Approach

Shraddha R. Khonde[1*], Shyamal S. Virnodkar[2], Sangita B. Nemade[3], Manisha A. Dudhedia[4], Bhavana Kanawade[5], Shravan H. Gawande[6]

[1] Department of Computer Engineering, M.E.S. Wadia College of Engineering, S. P. Pune University, Pune 411001, India
[2] Department of Computer Engineering, K. J. Somaiya Institute of Technology, Mumbai 400022, India
[3] Department Information Technology, Government College of Engineering, Chh. Sambhajinagar 431005, India
[4] Department of Electronics and Telecommunication, Marathwada Mitra Mandal's College of Engineering, Pune 411052, India
[5] Department of Information Technology, International Institute of Information Technology, S. P. Pune University, Pune 411057, India
[6] Department of Mechanical Engineering, M. E. S. College of Engineering, S. P. Pune University, Pune 411001, India

Corresponding Author Email: shraddha.khonde@mescoepune.org

https://doi.org/10.18280/ria.380325

Received: 27 November 2023
Revised: 2 January 2024
Accepted: 29 March 2024
Available online: 21 June 2024

Keywords:
stock-prediction, sentiment-analysis, machine learning, fake news detection

ABSTRACT

This research provides a web application for stock prediction that analyze financial news sentiment and predicts market performance using machine learning techniques. Machine learning techniques used are vector based, lexicon based analysis and LSTM. For sentiment analysis and stock prediction. With 86% accuracy, the sentiment analysis algorithm categorizes news headlines as positive or negative. The stock prediction model estimates stock performance with a mean absolute inaccuracy of 3.4 percent. When both models are combined, they forecast stock performance with an accuracy of 83%. The system's potential for usage in the stock market is demonstrated by the results, which provide vital insights into machine learning algorithms for stock data prediction along with sentiment analysis utilizing headlines from financial news.

1. INTRODUCTION

One of the most complex and dynamic system is a stock market which is driven by a wide range of elements which includes economic indicators, corporate performance, and world events. One of the difficulties in the stock market investment is anticipating stock performance in the future. Financial news headlines provide essential information to investors, but analyzing a big number of news headlines may be time-consuming and difficult. Machine learning algorithms have been widely employed in recent years to analyses financial news and forecast stock values. In is paper a method for prediction of stock data using sentiment analysis is presented using headlines from financial sector news.

1.1 Background and motivation

Financial news headlines are an important source of information for investors since they offer insight into market trends, corporate performance, and industry changes. However, for investors, analyzing a big number of news headlines and extracting valuable insights may be time-consuming and difficult. A technique called sentiment analysis is used for determining the sentiment represented in textual data, such as news headlines, by categorizing the language as positive, negative, or neutral. Investors can acquire insights into the overall sentiment of the market and uncover investment opportunities by analyzing the sentiment of financial news items. Financial news analysis takes plenty amount of time for the users who uses traditional methods whereas machine learning algorithms helps to find exact positive results of stock. Machine learning algorithms provides better insight using sentiment analysis to users depend on which stock predictions can be done. Use of machine learning along with sentiment analysis increases the accuracy and efficacy of system. Another important approach is stock data prediction, which forecasts future stock values based on previous data and other factors such as market patterns and economic indicators. Investors may gain a more thorough insight of the market and make more educated investment decisions by combining sentiment analysis with stock data prediction.

1.2 Objectives

The primary goal behind mentioned research study is to describe various methods for stock data prediction followed by sentiment analysis based on news headlines extracted from financial sector. The paper specifically seeks to: Develop a system using machine learning techniques that can reliably categorize financial news headlines as good, negative, or neutral. Create a stock data prediction model that can estimate stock performance in the future based on previous data and other variables such as market movements and economic indicators. Combine sentiment research with stock data prediction models to give investors with a thorough study of

financial news and forecast stock performance.

1.3 Scope and limitations

This research mainly focuses on the creation of a system for predicting data from stock and analyzing sentiment based on headlines of financial news. A collection of financial news headlines and historical stock data will be used to train and test the system. The accuracy of the system in forecasting stock prices and sentiment analysis will be used to assess its performance. The system's limitations include the stock data prediction model's accuracy, which is impacted by a variety of elements such as market movements, economic indicators, and corporate performance. Furthermore, the intricacies of language and cultural variances in understanding. This system will have good impact over user's financial decision-making abilities. This system will provide positive as well as negative sentiments to users based on current top financial headlines. This will help user in predicting stock which will be beneficial for him to grow in the business. This system also reduces the time user take in analysis each news individually.

2. LITERATURE SURVEY

MKSVR, a system that quantitatively analyses news of financial market in intra-day and combines with stock tick price, was employed by Li et al. [1]. Over the course of a year, experiments were carried out utilizing market news and tick data from the Hong Kong stock exchange. The results show that the MKSVR can make greater use of hidden facts obtained from news articles. Information from historical stock prices are used than models which just make use of news artefact for stock price estimation. Authors also proves that the MKSVR method surmount models that employ just single source of information.

Li et al. [2] elaborated a developed system names eMAQT for profit analysis. This system supports an assumption mentioned previously about finance related to public. It clarifies that public information events are vulnerable to multiple explanations by investors. This outcome provides excellent trading possibilities for knowledgeable investors to benefit after the news release day. In other words, in the age of social media, stock markets are responsive to public information.

The following summarizes the contribution of the work undertaken by Nguyen et al. [3]. First, although earlier research analyzed general feelings in documents, this study elaborates a strategy for prediction of market data related to stock in terms of subject sentiment. Second, we presented two approaches for capturing topic sentiment connections. The first is technique which makes use of current topic model names as JST-based method, and the other is a method which uses defined way to recognize topics and feelings. This method is called as an Aspect-based sentiment method. Finally, this is the initial study to exhibit the usefulness of using sentiment analysis by investigating big scale test data. Although the average accuracy is only 54.41 percent, the suggested technique may foresee change in stock price with more than 60. Crone and Koeppel [4] showed that sentiment indicators may accurately predict market action. In terms of the non-linear descriptive model for continuous returns it has a validation set directional accuracy of 75.64% and a generalization set real-world rate of 60.26%. The observed findings proves that non-linear models performs way ahead than linear regressions and another benchmark models naive technique. The bivariate study also revealed that the relationship between the exchange rate and other sentiments strengthens after market moves. The investigation revealed the capability of employing sentiment indicators to define financial time series.

Nemes and Kiss [5] employed several sentiment analysis algorithms to emotionally inspect and categorize different news headlines from economics and assess their influence on distinct stock market value movements even when the context was missing. Emotions were divided into three categories: good, negative, and neutral. Text Blob and NLTK-VADER Lexicon tools have neutral categories, however Recurrent Neural Network (RNN) did not. The outcomes of the different sentiment studies were compared to the BERT result as a benchmark. Zuo and Kita [6] reported the P/E ratio prediction technique employing a Bayesian network. They digitized the P/E ratio data by using the uniform clustering or Ward technique to cluster the P/E ratio frequency distribution. The digitized P/E ratio data are used to generate a Bayesian network for the interdependence among prior P/E ratio distributions. Authors compared the correlation coefficient and accuracy of prediction of the actual stock price with algorithms of classic time-series forecast. These algorithms are ARMA, MA, ARCH and AR models.

Katayama and Tsuda [7] verified that the emotion determined by the dictionary of polarity had a stock market effect. Three hypotheses were developed and tested each one by authors. Hypothesis 1 states that if favorable news is released, increase in the stock price of company will be observed, and the outcome will be as predicted. Many investors feel that by reading the news and making investing decisions, they are assessing the substance. The second hypothesis states that hypothesis 1 effect will increases for front-page articles. The analysis based on regression confirmed that this hypothesis was valid. Sidogi et al. [8] used Long-Term Short-Term Memory (LSTM) networks to study the im- pact of financial news headline sentiment on the prediction of stock prices. The analysis is carried out on intra-day data with precise lag durations between headlines from the published article and realized stock values. They conducted a systematic comparison of the performance of LSTM models for stock market forecasting under identical settings, but with an objective evaluation of the relevance of including financial news emotions as model inputs. Sheta [9] developed fuzzy models for two nonlinear processes using the Takagi-Sugeno (TS) approach. They were the development effort estimate for a NASA software project and the stock market projection for the following week SP 500. The creation of the TS fuzzy model may be accomplished in two parts. 1) Using model input data, determine the membership functions in the rule antecedents. 2) estimate the consequence parameters. They estimated these parameters using least-squares estimation. The findings were encouraging.

Ho et al. [10] suggested an intraday financial trading system driven by a unique brain-inspired evolving Mamdani Takagi-Sugeno Neural-Fuzzy Inference System (eMTSFIS). When compared to current econometric and neural-fuzzy forecasting methodologies, the eMTSFIS predictive model contained synaptic mechanisms and in- formation processing capabilities of the human hippocampus, resulting in a more resilient and adaptable forecasting model. To create buy-sell trading signals, the suggested system's trading approach was based on the

moving-averages convergence/divergence (MACD) concept. The lagging aspect of the MACD trading rule may be addressed by including forecasting skills into the calculation of the MACD trend signals. The experimental findings based on the SP500 Index demonstrated that eMTSFIS could deliver highly accurate forecasts and that the resulting system could discover timely trading opportunities while minimizing wasteful trading trans- actions. These characteristics allowed the eMTSFIS-based trading system to provide investors with better multiplicative returns. Nagar and Hahsler [11] presented a unique approach to aggregate news from various sources using automated text mining approach to build a corpus for news. The corpus underwent filtration to isolate pertinent sentences, subjected to analysis through natural language processing (NLP) methods. A sentiment gauge, denoted as News Sentiment, was introduced, relying on the count of words displaying positive and negative polarities. Open-source tools were employed to construct tools for both news compilation and aggregation, along with a tool for sentiment assessment. The researchers observed a robust correlation between the temporal fluctuations of News Sentiment and the real-time shifts in stock prices. Su et al. [12] implemented a self-organized neuro-fuzzy model with five layers. This system uses technical indicators for modeling stock market dynamics. The model's predictive and forecasting capabilities were validated using a dataset comprising four indicators: Volume Adjusted Moving Average (VAMA), Stochastic Oscillator and Ease of Movement (EMV) which is basically sourced from TAIEX. To enhance stock price prediction, a proposed adaptation of the moving average method was introduced for generating the input set for the neuro-fuzzy model. Simulation outcomes demonstrated the effectiveness of the model in prediction and its accuracy in forecasting. The neuro-fuzzy model significantly mitigated input errors stemming from the modified moving average method, leading to improved prediction outcomes.

Zarandi et al. [13] implemented an expert system for stock price analysis based on type 2 fuzzy rules. A fuzzy logic system of interval type 2 was employed to account for uncertainties in modeling rules, with each feature membership value represented as an interval. The suggested type 2 fuzzy model incorporated technical and fundamental indices as input variables and underwent testing for predicting the stock price of an Asian automobile manufacturer. Rigorous experimental testing revealed the model's efficacy in accurately predicting price fluctuations across diverse sectors of stocks. The outcomes were highly promising and were integrated into a real-time trading system for the anticipation of stock prices during active trading periods.

Sim et al. [14] proposed a new method for generation of rules for selection of undervalued stock named as 3D subspace clustering. 3D subspace clustering proves to be a highly efficient method for handling high-dimensional financial data, demonstrating adaptability to new datasets. This approach remains impervious to human biases and emotions, providing easily interpretable results. Over an extensive 28-year experimentation period in the stock market (1980 to 2007), XII observed that applying rules derived from 3D subspace clustering algorithms, specifically CAT Seeker and MIC, yielded profits 60 percent higher than those achieved using Graham's rules in isolation.

Chowdhury et al. [15] elaborated a model based on prediction which happens after market close to analyze each investor or company sentiment. Over a 4-week sentiment tracking period, a notable association emerged between the initial stock price trend and the effective positive index curve produced by the predictive news mining model presented in this article. The investigation highlighted a robust correlation, approximately 67%, between news sentiment and the original stock price curve. This correlation strongly supports the efficient market hypothesis, indicating that sentiment is distinctly mirrored in the movements of stock prices. Sarma et al. [16] and Nalabala et al. [17] revealed the importance of social media aspects in consideration of sentiment analysis for stock prediction along with e-commerce field. Wahyuningsih et al. [18] provides the comparision of using various machine learning algorithms in stock prediction. An experiment is conducted by Tudor [19] on Romaninan stock market for understanding sentiment analysis importance. Lenten et al. [20] elaborates various parameters used for performance analysis in prediction. Venturini [21] focuses on various risk factors involved during prediction. Importance of correlation in prediction is elaborated by Nalabala and Nirupamabhat [22]. Gupta et al. [23] explain the importance of sentiment analysis.

From literature survey, it is observed that sentiment analysis for financial news can be used for stock prediction in real time environment. Most of the researchers used standard dataset which can be the limitation for many real time applications. Various techniques are used for sentiment analysis like fuzzy logic, clustering, classification and natural language processing. According to survey natural language processing is used wisely due to its efficient performance. AIFFN and TF-IDF methods are the most efficient and accurate methods which can be used in the stock prediction for financial news. These methods overcome limitations provided by data storage and vector-based methods used in prediction. It is can concluded that real time financial news fetch and prediction will provide better prediction and good sentiment analysis for stock data as compared to standard datasets and analysis used for stock predictions.

3. SYSTEM ARCHITECTURE

This section elaborates system architecture of the research implementation. System focuses on the detection of fake news and stock prediction based on news feed runtime. News will be taken from various known stock markets predictions which will be further used for sentimental analysis.

3.1 Overview of the system

System overview is shown in Figure 1 that is system architecture of the implementation of the research. The backend is used to store the news related to financial sector. The news is fetched from the various well-known websites. The websites are executed runtime and stores data in the database. Natural language processing is used do financial feeds.

3.2 Data flow diagram

Figure 2 represents the data flow diagram for the system implemented in this research. This diagram is used to show the data flow between the various websites and the project. Web scraper is used to collect data from various websites runtime which is then preprocessed and save in the database for

prediction. Natural language processing is used to find the final prediction. This will help in doing sentiment analysis of stock prediction.

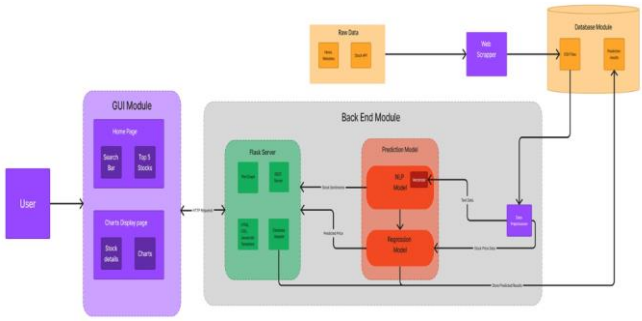

Figure 1. System architecture

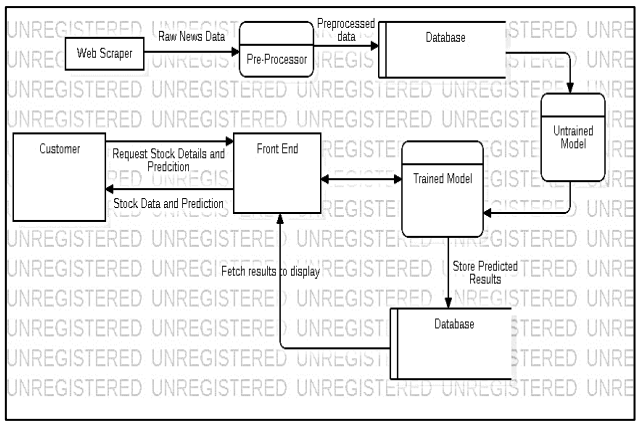

Figure 2. Data flow diagram

3.3 Component design and interaction

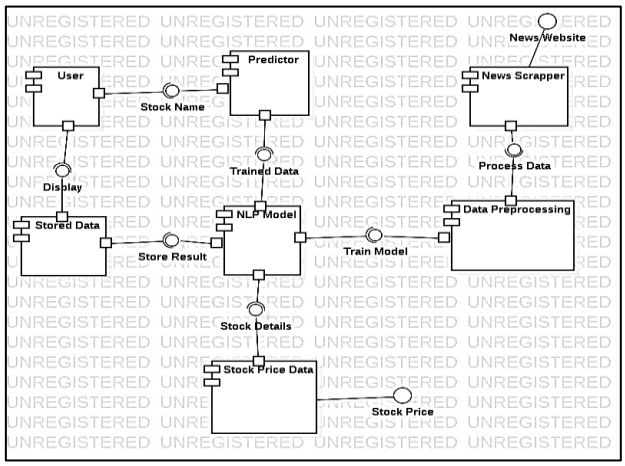

Figure 3. Component diagram

Figure 3 represents the interaction diagram while implementation of various components in this system during research. All the modules are integrated to get the final prediction and sentiment analysis. Figure 4 represents interaction diagram between various modules of the system from this research. The modules used to interact with each other to send input and intermediate results so that final predictions can be generated. The input is generated in real time environment according to news headlines in various websites. Natural language processing is used to do sentiment analysis.

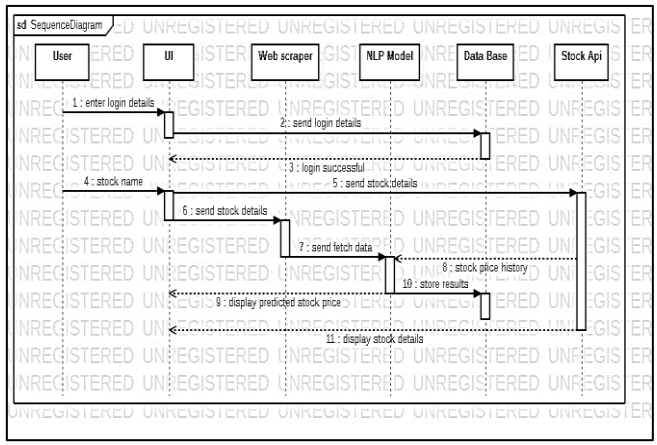

Figure 4. Interaction diagram

3.4 Tools, technologies and libraries

In this research while implementing a stock prediction system most of the tools are used as per the requirements. These technologies are working together to collaborate input together from various websites and save the records in the database server. The threshold value is used to gather the information related to stock price dynamically. Scrapper is used to collect information which will be further evaluated and analyzed using python language. Various libraries are used to get the final predictions. All the tools and libraries used are summarized in Table 1 as given below. Table 1 represents various tools and technologies used for different functionalities.

Table 1. Tools and technologies

Name	Version	Use
HTML5, CSS, Javascript	-	Used for Frontend
Python	3.11.2	For backend
Flask	2.3	For backend server
mySQL	8.0	For database
fyers api	2	To get stock price data
requests	2.25.1	To fetch news headlines
beautifulsoup4	4.11.1	To create news scrapper
scikit-learn	1.0.2	Python library to implement machine learning models
Jinja2	3.1.2	Used to create webpages
Jupyter	6.5.4	To test ml models in python
lxml	4.9.2	To parse the xml data
matplotlib	3.7.0	Used to make graphs
nltk	3.8.1	Used for sentiment analysis
numpy	1.24.3	Used for mathematical calculations
pandas	2.0.1	Used for data processing
pickle	5	Object serialization and Deserialization

3.5 Data sources

Information for this project is gathered from a variety of sources. Kaggle provided the news headline data for training, especially the “Indian Financial News Articles (2003-2020)” dataset by hkapoor. In addition to the Kaggle dataset, we gathered news headline data from Money Control, Economic Times, and Business Standard websites. We extracted the

news headlines from these websites using web scraping methods. We received stock data from Fyers, which offers an API for obtaining historical stock data. We gathered information on a variety of equities traded on the Indian stock exchanges, including the National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE).

3.6 Data cleaning and formatting

Multiple procedures were taken to assure the greatest quality and accuracy of the news headline data utilized in this research. The data was first lemmatized to reduce words to their base form, minimizing word variety and enhancing consistency. Following that, stop word removal was used to eliminate frequent terms that did not contribute substantial sense to the text. Punctuation removal was also used to remove extraneous characters that might interfere with the analytical process. Finally, the data was vectorized to turn the text into numerical values for simpler analysis using the term frequency-inverse document frequency (TF-IDF) approach. This approach helps in finding the appropriate feed and reducing the processing and overhead time.

3.7 Data storage and retrieval

To promote usage and interoperability with other computer languages, the financial news headlines were saved in both JSON and CSV formats. The data was obtained using the Pandas library's pd.read function, which made loading the data into memory for training purposes simple and quick.

4. FEATURE SELECTION AND EXTRACTION

4.1 Feature selection techniques

In this system on Prediction of Sentiment Analysis and stock data using Financial News Headlines, the various feature selection techniques were employed to identify the most relevant and informative features for identified models. One of the techniques systems utilized was correlation-based feature selection. This method enabled us to analyze the correlation between each feature and the objective variable, enabling us to select features that had a strong impact on predicting stock performance and sentiment analysis. By identifying and selecting the most significant features, intends to increase the accuracy and efficacy of our prediction and analysis models.

4.2 Feature extraction methods

To extract meaningful and representative information from the financial news headlines, employed feature extraction methods. One of the primary methodologies used was TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. TF-IDF is a widely used technique in natural language processing that allocates weight to words based on their frequency and prominence in a document corpus. By transforming the textual data into numerical feature vectors using TF-IDF, system will able to capture the relevance of each word in the headlines, thus facilitating more accurate prediction and sentiment analysis. Limitation of TF-IDF is it does not help in getting semantic meaning of words. However in sentiment analysis, semantic meaning is very important to generate positive or negative sentiments. On these sentiments user can take decision about its financial stock predictions.

4.3 Feature engineering approaches

In this system various feature engineering methods were used in addition to feature selection and extraction to improve the predictive capability of our models. Feature engineering is basically used to extract complex patterns and features from the data. This follows the process for creating new features or modifying existing feature. System use many feature engineering approaches specialized to financial news and market data in research. Sentiment word embedding was one of these technologies, which involves encoding sentiment-related information into numerical vectors to capture the sentiment polarity and intensity of news headlines. Due to this system hoped to increase the performance and resilience of stock prediction and sentiment analysis system by including such manufactured characteristics.

Overall, in this system various techniques were attempted to optimize the representation and relevance of features in Stock Data Prediction and Sentiment Analysis system by utilizing feature selection techniques, feature extraction methods such as TF-IDF vectorization, and feature engineering approaches such as sentiment word embedding. In word embedding polarity of sentiments is consider as positive or negative using vector based matrix. System was able to extract significant information from financial news headlines, increase the accuracy of system models, and get a better grasp of stock market dynamics and sentiment patterns using these approaches.

5. SENTIMENT ANALYSIS

5.1 Sentiment analysis methods

In this system both vector-based and lexicon-based sentiment analysis algorithms were analyzed. System used term frequency-inverse document frequency (TF-IDF) to turn each news headline into a numerical vector in the vector-based technique. After then used logistic regression to categorize each vector as positive, negative, or neutral. Then utilized a pre-defined sentiment lexicon to give scores to each word in the news headline and then summed these values to produce an overall sentiment score in the lexicon-based technique.

5.2 Sentiment lexicon and dictionary based approaches

In this system testing with numerous sentiment lexicons in the sentiment lexicon and dictionary-based techniques, including the AFINN lexicon and the SentiWordNet lexicon. System also investigated the usage of custom-built emotion dictionaries for certain businesses and financial and stock market subjects. Table 2 represents the performance of system using AIFFN Lexicon method. Performance is evaluated using various parameters as precision, recall and F1-score. These parameters are used to check performance of system in terms of stability and robustness for stock predictions. Performance is evaluated using two classes as negative and positive. From table we can say that this method provides better performance in both the classes as positive and negative.

Table 2. Performance of AIFFN lexicon

Class	Precision	Recall	F1-Score
Negative	0.59	0.78	0.67
Positive	0.95	0.88	0.91

Figure 5 represents the graphical representation of system performance for AIFFN Lexicon method. Graph represents performance for both classes as negative and positive.

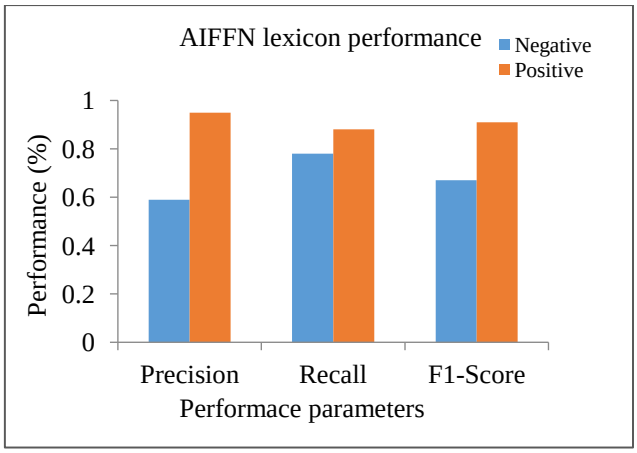

Figure 5. Graphical representation of AIFFN lexicon

5.3 Vector based approaches

In system the TF-IDF technique is employed for the vector-based methods, followed by logistic regression to predict sentiment labels. To improve the performance of our models, system was tested with various hyperparameters such as regularization strength and maximum number of iterations. Table 3 represents the performance of TF-IDF method using vector based methods. Performance is calculated using various parameters as precision, recall and f1-score. From figures mentioned in table this method provides good performance for both the classes. Figure 6 shows the graphical representation of the TF-IDF method.

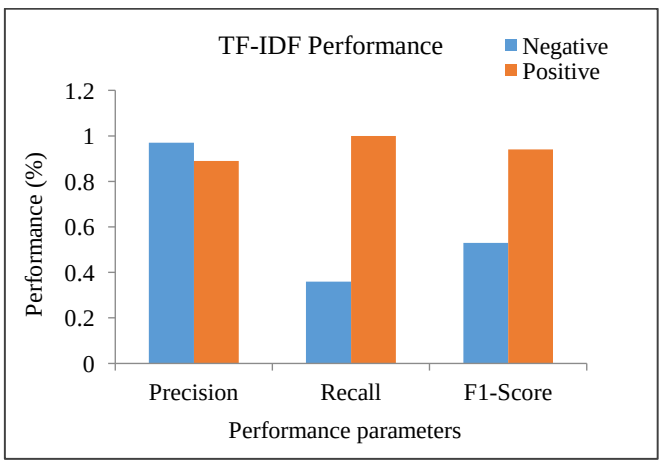

Figure 6. Representation of TF-IDF

Table 3. Performance of TF-IDF

Class	Precision	Recall	F1-Score
Negative	0.97	0.36	0.53
Positive	0.89	1.00	0.94

5.4 Evaluation metrics

In terms of assessment measures, system computes the accuracy of sentiment analysis models using the model. Score function. Performance is also assessed using accuracy, recall, and F1 score to evaluate models' ability in recognizing positive and negative thoughts. According to system performance it is discovered that TF-IDF beat AFINN in terms of accuracy, recall, and F1-score after analyzing the statistical data presented. When TF-IDF was compared to AFINN, its precision, recall, and F1-score for the positive class were much greater, suggesting its better accuracy in properly categorizing positive thoughts. Although TF-IDF had a poorer recall for the negative class, the total F1-score was higher, indicating a better balance of accuracy and recall. Furthermore, TF-IDF evaluates the relevance of terms in a text by taking their frequency and rarity in the corpus into consideration. This allows TF-IDF to capture the distinctive features of financial news headlines and their influence on sentiment analysis. AFINN, on the other hand, is based on a pre-defined sentiment vocabulary, which may not capture the subtleties and intricacies of financial language. Taking these variables into account, the selected TF-IDF as the sentiment analysis approach for system because of its better performance in properly identifying feelings and its ability to capture the importance of words in financial news headlines.

6. STOCK DATA PREDICTION

6.1 Prediction model

The system examined many models, including regression-based models, time-series analysis-based models, and machine learning-based models, and selected the best model for stock price prediction.

6.2 Regression based models

Linear Regression was one of the models we picked after researching numerous regression models for stock data prediction. Linear Regression was chosen because of its ease of use, readability, and ability to capture linear connections between variables. It is hoped to find underlying patterns and trends in the stock market by fitting a linear equation to the data. The R-squared statistic was used to assess the effectiveness of the Linear Regression model, which represents the percentage of the variation in stock price explained by the model. The Linear Regression model scored an impressive 96% accuracy, suggesting that it accurately captured the link between the given characteristics and the stock price. It is crucial to note, however, that the accuracy may vary depending on market circumstances and datasets, and more validation and testing are necessary to evaluate its effectiveness in various situations. Nonetheless, the findings underscore Linear Regression's promise as a trustworthy technique for stock data prediction. Data captured online is nonlinear in manner so it need to be converted into the specific vector as mention in previous section for sentiment analysis. Also real time data capture is dynamic which also requires feature extraction for sentiment analysis. Though the data is dynamic and nonlinear the regression models provides more efficacy on stock data prediction. In real time data capture each feature is important for sentiment analysis whereas the

features having more polarity towards positive sentiments are used to generate a linear based stock prediction model.

6.3 Time series analysis based models

Further in this system another approach was investigated as time series analysis-based models for stock data prediction in system, and one of the models we developed was the Long Short-Term Memory (LSTM) recurrent neural network (RNN). LSTM was selected because of its ability to capture temporal relationships and successfully manage data sequences. We hoped to use LSTM to utilize past stock data and capture patterns and trends over time. The LSTM model's correctness was assessed using relevant metrics such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE). The LSTM model has an accuracy of 71%, demonstrating that it can forecast stock values to some degree. It is vital to note that the accuracy will vary dependent on the dataset, hyperparameter tweaking, and LSTM model architecture. Further investigation and testing are necessary to optimize the performance and examine the appropriateness of the LSTM model for various stock market circumstances and datasets.

6.4 Evaluation metrics

To analyze the accuracy and performance of our prediction models, system employed assessment measures such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These metrics compute the average squared difference between anticipated and actual stock prices, giving a quantitative assessment of the model's ability to forecast stock prices reliably. In addition, system looked at measures like Mean Absolute Error (MAE) and R-squared (R2) to assess the prediction models' overall performance and capacity to incorporate volatility in stock price data. Sentiment analysis plays a vital role in stock data prediction when both the models are used in an ensemble approach. Both models help in compromising the limitations of each other to improve upon system performance. These assessment measures assisted us in assessing the efficacy and dependability of our prediction models and comparing their performance.

7. INTEGRATION AND IMPLEMENTATION

7.1 System integration and deployment

In implementation the meticulously merged frontend and backend pieces to provide a strong communication channel between the user interface and the underlying functionality. We used HTML for structural components, CSS for style and layout, and JavaScript for dynamic and interactive features to develop an attractive and engaging frontend. We used the Flask framework on the backend, which offered a lightweight and flexible environment for constructing web apps. The system is successfully organized the backend code by using Flask's modular architecture, allowing for fast routing, request handling, and data processing. We used MySQL as the database management system to store and retrieve essential data, assuring data integrity and simple data editing. We assured that the integrated system worked perfectly via thorough testing and debugging, and that it was ready for deployment in the field. Integrity and consistency of data is maintained with the help of normalization where each dataset used is properly normalize and follows all the properties of transaction. Data preprocessing is done to clean the data and maintain integrity of real time data. Figure 7 shows the backend table structure used in the implementation of this system. Backend databases are used to store the news feed received from various websites using real time data capture related to financial news. All the news is converted into the JSON and CSV structure to store a record in the database server.

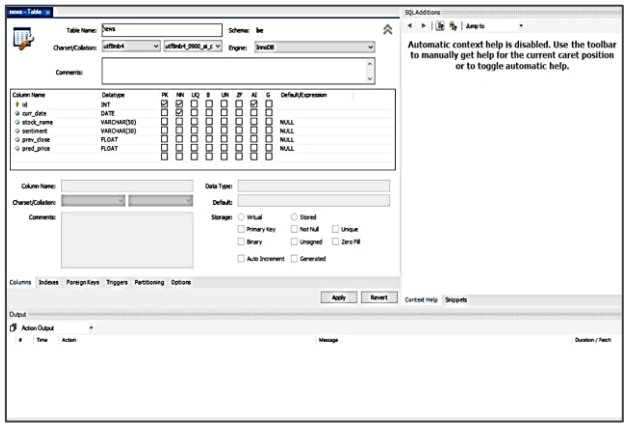

Figure 7. Back-end structure in database

7.2 User interface and functionality

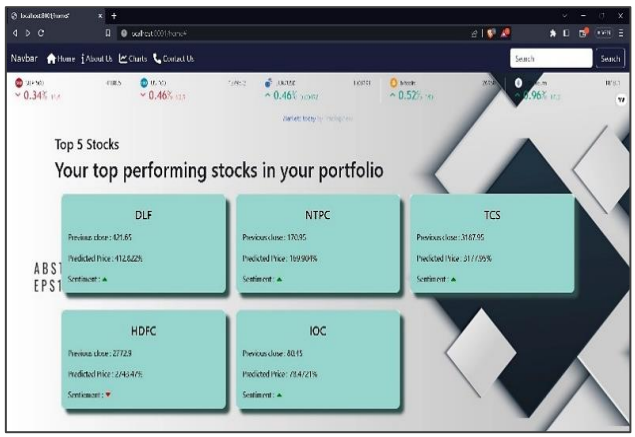

Figure 8. System homepage

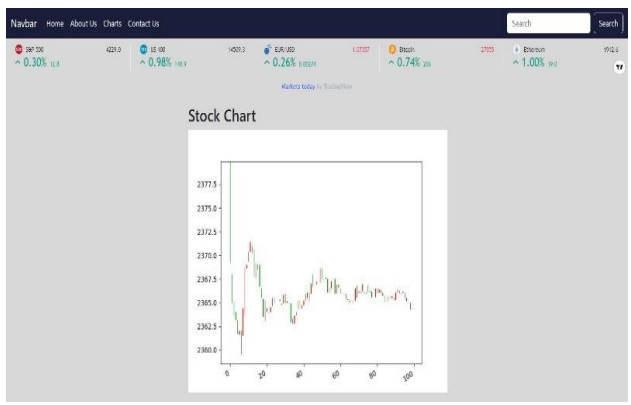

Figure 9. Stock detail page

System's user interface (Figure 8) is intended to deliver a nice and intuitive experience for individuals interested in stock data prediction and sentiment analysis. The interface contains

a navigation bar that provides quick access to various system elements. The homepage, about us, and charts are among the sections. The top five stocks of the day are presented in separate cards on the homepage. This gives consumers a fast summary of the trending stocks and captures their interest. The cards include important information about each stock, such as the symbol, current price, and a short description. The about us page is a knowledge base where people may learn more about our system and its purpose. We describe in detail how the algorithm uses stock data prediction and sentiment analysis based on financial news headlines. Additionally, contact information is given for users to contact us if they have any questions or complaints. The charts area provides visitors with extensive financial charts and information about the chosen company. Users may choose a particular stock of interest, and the system will create interactive charts illustrating historical price patterns, volume, and other pertinent financial information. In addition, the system uses stock data prediction skills to present customers with the expected price of the chosen stock. Overall, the user interface attempts to give consumers with a visually attractive and useful experience. It enables consumers to acquire essential stock data, get insights via charts and projections, and quickly navigate through the system's many areas. Figure 8 represents the top five records and websites selected from online or real time feed as an input. These will be used as highest priority websites for sending news as input to the model for prediction. Figure 9 represents the detail information of particular stock. It represents the fluctuation in the stock prizes from the exact real time data. It helps user to go for prediction and sentiment analysis.

7.3 System performance and scalability

Several optimizations were done in system to boost system speed. One of the most significant advancements was the ability to save forecasts, which lowered the computing time necessary for recalculating predictions. System tries to avoid the need to compute the predictions every time a user requested the data by storing them, resulting in improved response times. Furthermore, for chart visualization, is used as a backend technique in which the charts were created and saved as image files. This method lowered front-end processing time and enabled rapid retrieval and presentation of charts. Furthermore, system optimized the chart rendering process by lowering the quantity of the dataset needed to generate the charts, allowing for quicker rendering times without sacrificing chart quality or accuracy.

Scalability: This system has the potential to be implemented on the cloud in terms of scalability. The cloud has various benefits, including scalability, dependability, and cost- effectiveness. System can manage increasing user traffic and support a rising number of users without experiencing substantial performance deterioration by employing cloud infrastructure.

Furthermore, cloud deployment enables simple scalability by allowing for the dynamic allocation of resources depending on demand. This guarantees that this system can scale up or down as required, providing peak performance and cost-efficiency during low-traffic times. Furthermore, cloud-based services often have built-in redundancy and high availability, guaranteeing that our system remains available even if hardware fails or is disrupted.

In summary, the adoption of optimizations such as storing predictions and creating charts in the backend has improved the speed of our system. These improvements have resulted in quicker reaction times and better overall system performance. Furthermore, system's scalability is strengthened by its cloud deployment capabilities, which allows for effective resource allocation and accommodating rising user demand.

8. RESULTS AND EVALUATION

8.1 Data analysis and visualization

Sentiment Analysis is done as a part of data analysis where word embedding is possible with vectorization. Validation is done with the help of feature extraction for real time data. A rigorous data analysis approach was used in this study effort to get insights from stock data and financial news headlines. The stock data was preprocessed, which included managing missing numbers, outliers, and guaranteeing data quality. To investigate the data's features, exploratory data analysis (EDA) approaches were used. To analyze the data's central trends and dispersion, statistical metrics such as mean, median, and standard deviation were produced. To discover patterns and trends in stock price movements over time, time series analysis methods such as autocorrelation and trend analysis were used. In addition to data analysis, visualization was critical in successfully presenting the results. Various visualization tools and approaches were used to graphically display the data's linkages and patterns. Line charts were used to display previous stock prices and track their patterns. The association between stock prices and sentiment analysis ratings generated from financial news headlines was investigated using scatter plots. Heatmaps were used to visualize sentiment distributions across various time periods or stock categories. These visualizations aided in comprehending the dynamics of stock market data as well as the effect of sentiment analysis on stock performance.

8.2 System performance evaluation

To test its efficiency, the system created for stock data prediction and sentiment analysis underwent a thorough performance review. The system's prediction skills were evaluated using historical stock data and financial news headlines. The data was divided into training and testing sets throughout the assessment process, with the training set used to train the prediction model and the testing set used to evaluate its performance. To assess the accuracy and reliability of the system's predictions, several assessment measures were used. To measure the difference between projected and actual stock prices, mean absolute error (MAE) and root mean square error (RMSE) were determined. These measures revealed the average size of the forecast mistakes. Furthermore, accuracy metrics were used to evaluate the system's capacity to accurately categorize the emotion of financial news headlines. To assess the performance of sentiment classification, precision, recall, and F1- score were calculated.

8.3 Comparison with existing methods

The comparison of suggested approach to current techniques for predicting stock data and analyzing sentiment is done. To prove results a thorough literature analysis is done to discover previously published or implemented relevant research and methodologies. System concentrates on typical

current approaches that used comparable data and addressed similar goals.

The performance of various existing approaches using the same test dataset and assessment criteria as our system was evaluated and tested. The evaluation of their accuracy in prediction, sentiment categorization, and overall system performance to our suggested approach is done. The system is illustrated for the efficacy and superiority of system in terms of predictive skills, sentiment analysis accuracy, and overall performance via this comparison.

8.4 Interpretation of results

System offered an in-depth interpretation and analysis of the data analysis, system performance assessment, and comparison with current methodologies outcomes. Study of all the patterns and insights gleaned through stock data analysis, such as finding repeating trends, seasonality, and association with market occurrences. System also explore about the results of the sentiment analysis, emphasizing any intriguing discoveries or correlations between sentiment and stock performance.

Furthermore, this paper critically assessed system's shortcomings and prospective areas for development. Handling of some issues including data quality, model assumptions, and inherent uncertainties in stock market dynamics might have impacted the system's performance is also considered. System also took into account any possible biases or limits in the sentiment analysis process, such as issues with sentiment categorization accuracy or the effect of news sources.

The goal of the findings interpretation was to offer a thorough knowledge of our system's capabilities, limits, and implications for practical applications in stock prediction and sentiment analysis. The system can be enhanced for stock exchange where dynamic sentiment analysis can be done in real time environment. Collaboration of sentiment analysis in stock prediction in big extends or real time stock market can be considered as future enhancements.

9. DISCUSSION

9.1 Implications of the system

The created system's implications for stock data prediction and sentiment analysis are noteworthy in various ways. To begin, including sentiment analysis of financial news headlines into the stock prediction process gives useful insights into market mood and investor behavior. This enables investors to make better-informed choices and perhaps reduce risks. Second, by proving the feasibility and usefulness of integrating machine learning and natural language processing methods in stock prediction and sentiment analysis, the system adds to the area of financial technology. This brings up new research and development opportunities in the fields of finance and technology integration. Sentiment analysis will have very high impact on the individual user where user can take financial decision based on the sentiment analysis as positive or negative. Positive sentiment reflects upgrowth in the stock prediction and negative reflects downfall. These sentiments are helpful for user for taking his financial decisions.

9.2 Limitations and future work

While the system seems to be promising, it is critical to recognize its limits and identify areas for further development. One restriction is that sentiment analysis is dependent on the accuracy and quality of financial news data. Data noise and biases may have an influence on sentiment categorization accuracy. Furthermore, altering market dynamics and economic considerations may have an impact on the system's performance. Future work might concentrate on improving sentiment analysis approaches, adding new data sources, and refining prediction models to better respond to changing market circumstances. Continuous monitoring and review of the system's performance are required to resolve these constraints and ensure the system's long-term usefulness.

10. CONCLUSION

This research presents a fully working system for predicting stock data and analyzing sentiment using financial news headlines. The system exhibits its potential to aid users in decision-making processes by merging machine learning algorithms, natural language processing, and financial analysis approaches. Sentiment analysis is the key insight presented in this paper as word embedding and vectorization is used for generating sentiment out of real time data feeded to system. This system extract features from various websites from financial news headlines which can be further user in predicting the sentiment as positive or negative. Collaboration of sentiment with stock prediction helps user to take financial decisions about stocks. The research emphasizes the significance of sentiment analysis in comprehending real data and its influence on stock performance. While there are certain drawbacks, the system provides a strong platform for future study and growth in the discipline. Further enhancements can be testing this system for big stock exchange projects and investors in big scale. However, this system developed for financial technology can provide vital insights on the incorporation of data-driven techniques in the stock market.

REFERENCES

[1] Li, X., Huang, X., Deng, X., Zhu, S. (2014). Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information. Neurocomputing, 142: 228-238. https://doi.org/10.1016/j.neucom.2014.04.043

[2] Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., Chen, Y. (2014). The effect of news and public mood on stock movements. Information Sciences, 278: 826-840. https://doi.org/10.1016/j.ins.2014.03.096

[3] Nguyen, T.H., Shirai, K., Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, 42(24): 9603-9611. https://doi.org/10.1016/j.eswa.2015.07.052

[4] Crone, S.F., Koeppel, C. (2014). Predicting exchange rates with sentiment indicators: An empirical evaluation using text mining and multilayer perceptrons. In 2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), London, UK, pp. 114-121. https://doi.org/10.1109/CIFEr.2014.6924062

[5] Nemes, L., Kiss, A. (2021). Prediction of stock values changes using sentiment analysis of stock news headlines. Journal of Information and Telecommunication, 5(3): 375-394.

https://doi.org/10.1080/24751839.2021.1874252
[6] Zuo, Y., Kita, E. (2012). Stock price forecast using Bayesian network. Expert Systems with Applications, 39(8): 6729-6737. https://doi.org/10.1016/j.eswa.2011.12.035
[7] Katayama, D., Tsuda, K. (2018). A method of measurement of the impact of Japanese news on stock market. Procedia Computer Science, 126: 1336-1343. https://doi.org/10.1016/j.procs.2018.08.084
[8] Sidogi, T., Mbuvha, R., Marwala, T. (2021). Stock price prediction using sentiment analysis. In 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, pp. 46-51. https://doi.org/10.1109/SMC52423.2021.9659283
[9] Sheta, A. (2006). Software effort estimation and stock market prediction using takagi-sugeno fuzzy models. In 2006 IEEE International Conference on Fuzzy Systems, Vancouver, BC, Canada, pp. 171-178. https://doi.org/10.1109/FUZZY.2006.1681711
[10] Ho, W.L., Tung, W.L., Quek, C. (2010). Brain-inspired evolving neuro-fuzzy system for financial forecasting and trading of the s&p500 index. In PRICAI 2010: Trends in Artificial Intelligence: 11th Pacific Rim International Conference on Artificial Intelligence, Daegu, Korea, pp. 601-607. https://doi.org/10.1007/978-3-642-15246-7_56
[11] Nagar, A., Hahsler, M. (2012). Using text and data mining techniques to extract stock market sentiment from live news streams. In International Conference on Computer Technology and Science (ICCTS 2012), IACSIT Press, Singapore.
[12] Su, C.L., Chen, C.J., Yang, S.M. (2010). A self-organized neuro-fuzzy system for stock market dynamics modeling and forecasting. In Proceedings of the 14th WSEAS international conference on Computers: part of the 14th WSEAS CSCC multiconference-Volume II, pp. 733-745.
[13] Zarandi, M.F., Rezaee, B., Turksen, I.B., Neshat, E. (2009). A type-2 fuzzy rule-based expert system model for stock price analysis. Expert systems with Applications, 36(1): 139-154. https://doi.org/10.1016/j.eswa.2007.09.034
[14] Sim, K., Gopalkrishnan, V., Phua, C., Cong, G. (2012). 3D subspace clustering for value investing. IEEE Intelligent Systems, 29(2): 52-59. https://doi.org/10.1109/MIS.2012.24
[15] Chowdhury, S.G., Routh, S., Chakrabarti, S. (2014). News analytics and sentiment analysis to predict stock price trends. International Journal of Computer Science and Information Technologies, 5(3): 3595-3604.
[16] Sarma, S.L.V.V.D., Sekhar, D.V., Murali, G. (2023). Neural network and sentimental model for prediction of stock trade value. Revue d'Intelligence Artificielle, 37(2): 315-321. https://doi.org/10.18280/ria.370209
[17] Nalabala, D., Bhat, M.N. (2020). Predicting the E-commerce companies stock with the aid of web advertising via search engine and social media. Revue d'Intelligence Artificielle, 34(1): 89-94. https://doi.org/10.18280/ria.340112
[18] Wahyuningsih, T., Manongga, D., Sembiring, I., Wijono, S. (2024). Comparison of effectiveness of logistic regression, naive bayes, comparison of effectiveness of logistic regression, naive bayes. Procedia Computer Science, 234: 349-356. https://doi.org/10.1016/j.procs.2024.03.014
[19] Tudor, C. (2012). Active portfolio management on the Romanian stock market. 8th International Strategic Management Conference, Procedia - Social and Behavioral Sciences, 58: 543-551. https://doi.org/10.1016/j.sbspro.2012.09.1031
[20] Lenten, L., Crosby P., McKenzie J. (2019). Sentiment and bias in performance evaluation by impartial arbitrators. Economic Modelling, 76: 128-134. https://doi.org/10.1016/j.econmod.2018.07.026.
[21] Venturini, A. (2022). Climate change, risk factors and stock returns: A review of the literature. International Review of Financial Analysis, 79: 101934. https://doi.org/10.1016/j.irfa.2021.101934.
[22] Nalabala, D., Nirupamabhat, M. (2020). An optimized machine learning model for stock trend anticipation. Ingénierie des Systèmes d'Information, 25(6): 783-792. https://doi.org/10.18280/isi.250608
[23] Gupta, K., Jiwani, N., Afreen, N. (2023). A combined approach of sentimental analysis using machine learning techniques. Revue d'Intelligence Artificielle, 37(1): 1-6. https://doi.org/10.18280/ria.370101