iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68455

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Using Financial News Headlines to Predict Stock Data and Conduct Sentiment Analysis

Dr. Madhav M.Bokare[1], Mr.Amol V.Suryawanshi[2]
Department of computer Science, SRTMU Nanded

Abstract: This study offers a web application for stock prediction that uses machine learning algorithms to forecast market performance and analyse the sentiment of financial news. LSTM, lexicon-based analysis, and vector-based machine learning approaches are employed. for stock forecasting and sentiment analysis. The sentiment analysis system classifies news headlines as either positive or negative with an accuracy rate of 86%. The mean absolute error of the stock prediction model's stock performance estimation is 3.4 percent. The total accuracy of both models in predicting stock performance is 83%. The results indicate the system's potential for use in the stock market by offering crucial insights into machine learning algorithms for sentiment analysis using financial news headlines and stock data prediction.
Keywords: machine learning, stock-prediction, sentiment-analysis, fake news detection

I. INTRODUCTION

The stock market is one of the most intricate and dynamic systems, influenced by a variety of factors such as global events, business performance, and economic indicators. Predicting future stock performance is one of the challenges of stock market investing. Investors can learn important information from financial news headlines, but it can be challenging and time-consuming to analyse a large number of news headlines. In recent years, machine learning algorithms have been used extensively to estimate stock values and assess financial news. This research presents a sentiment analysis-based stock data prediction method employing news headlines from the financial sector.

A. Background and motivation

Because they include information on market trends, company performance, and industry changes, financial news headlines are a valuable information resource for investors. However, it can be challenging and time-consuming for investors to sift through a large number of news headlines and extract insightful information. By classifying the language as neutral, negative, or positive, a method known as sentiment analysis can be used to identify the sentiment expressed in textual data, including news headlines. By examining the emotion of financial news articles, investors can gain insight into the general mood of the market and find possibilities for investment. While machine learning algorithms assist users identify precise favourable stock results, traditional methods of financial news research take a lot of time.Using sentiment analysis, machine learning algorithms give users more insight into which stocks can be predicted. The accuracy and effectiveness of the system are increased when machine learning and sentiment analysis are used together. Stock data prediction, which projects future stock values based on historical data as well as additional variables like market trends and economic indicators, is another crucial strategy. Combining sentiment analysis with stock data prediction can help investors make better informed investment decisions and have a more comprehensive understanding of the market.

B. Objectives

The main objective of the aforementioned research study is to outline different approaches for predicting stock data and then conducting sentiment analysis using news headlines taken from the financial industry. In particular, the paper aims to: Create a machine learning-based system that can accurately classify financial news headlines as neutral, bad, or positive. Build a stock data prediction model that uses historical data together with additional variables like market trends and economic indicators to forecast future stock performance. To provide investors with a comprehensive analysis of financial news and estimate stock performance, combine sentiment research with stock data prediction algorithms.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

C. Scope and limitations

The development of a system for stock data prediction and sentiment analysis based on financial news headlines is the primary goal of this study. The algorithm will be trained and tested using historical stock data and a collection of financial news headlines. The system's performance will be evaluated based on how well it predicts stock prices and sentiment analysis. The accuracy of the stock data prediction model, which depends on a number of factors such market movements, economic indicators, and company performance, is one of the system's drawbacks. Moreover, linguistic complexities and cultural differences in comprehension.Users' ability to make sound financial decisions will be positively impacted by this system. This method will deliver positive as well as negative attitudes to consumers based on current top financial headlines. This will assist the user in stock prediction, which will help him expand his firm. Additionally, this system cuts down on the amount of time users spend analysing each news item separately.

II. LITERATURE SURVEY

Li et al. used MKSVR, a method that quantitatively evaluates intraday financial market news and integrates it with stock tick price [1]. Experiments were conducted using tick data and market news from the Hong Kong stock exchange during a one-year period. The findings demonstrate that the MKSVR can leverage hidden facts from news stories more effectively. In contrast to algorithms that solely rely on news artefacts for stock price estimation, past stock price data is utilised. The authors further demonstrate that the MKSVR approach outperforms models that just use one information source.

For profit analysis, Li et al. [2] described a system they built called eMAQT. An earlier assumption regarding public finance is supported by this system. It makes clear that investors may give a variety of reasons for public information events. After the day of the news announcement, this result offers skilled investors great trading opportunities. Stated differently, stock markets react to public knowledge in the era of social media.

The contribution of Nguyen et al.'s work is summed up as follows [3]. First, this work develops a method for predicting stock market data in terms of subject sentiment, even if previous research examined broad sentiments in documents. Second, we introduced two methods for identifying sentiment relationships between topics. The first is a JST-based strategy that uses the current topic model names, while the second is a method that employs a defined way to identify topics and emotions. This approach is referred to as an aspect-based sentiment analysis. Lastly, by examining large test data, this is the first study to demonstrate the value of sentiment analysis. The recommended method may predict a change in stock price with more than 60% accuracy, despite the average accuracy being only 54.41 percent. Market activity can be correctly predicted by sentiment indicators, as demonstrated by Crone and Koeppel [4]. The non-linear descriptive model for continuous returns has a real-world rate of 60.26% and a directional accuracy of 75.64% on the validation set. The reported results demonstrate that non-linear models outperform linear regressions and other benchmark models. The bivariate analysis also showed that following market movements, there is a stronger correlation between the exchange rate and other moods. The study demonstrated that emotion indicators can be used to define financial time series.

In order to emotionally examine and classify various economic news items and evaluate their impact on various stock market value movements—even in the absence of context—Nemes and Kiss [5] used a number of sentiment analysis algorithms. Three types of emotions were distinguished: neutral, negative, and positive. Recurrent neural networks (RNNs) lacked neutral categories, in contrast to Text Blob and NLTK-VADER Lexicon tools. The BERT result served as a standard against which the results of the other sentiment investigations were evaluated. Zuo and Kita [6] described a Bayesian network-based P/E ratio prediction method. They clustered the P/E ratio frequency distribution using the Ward procedure, also known as uniform clustering, after digitising the P/E ratio data. A Bayesian network for the interdependence among previous P/E ratio distributions is created using the digitalised P/E ratio data. The authors contrasted methods of traditional time-series forecasting with the correlation coefficient and accuracy of predicting the actual stock price. The ARMA, MA, ARCH, and AR models are these algorithms. The stock market influence of the emotion identified by the dictionary of polarity was confirmed by Katayama and Tsuda [7]. The writers developed three hypotheses and tested each one. According to Hypothesis 1, if positive news is reported, the company's stock price will rise and the result will be as expected. Many investors believe they are evaluating the substance when they read the news and make investment decisions. According to the second hypothesis, front-page items will boost the effect of hypothesis 1. The validity of this hypothesis was proven by the regression-based study. Sidogi et al. [8] investigated the influence of mood in financial news headlines on stock price prediction using Long-Term Short-Term Memory (LSTM) networks. Intra-day data with exact lag times between the published article's headlines and actual stock values are used for the analysis. In order to objectively assess the value of adding financial news emotions as model inputs, they systematically compared the performance of LSTM models for stock market forecasting under the same conditions. Sheta [9] used the Takagi-Sugeno (TS) method to create fuzzy models for two nonlinear processes.

They were the stock market forecast for the SP 500 the following week and the estimated amount of work required to develop a NASA software project. There are two possible ways to create the TS fuzzy model. 1) Find the membership functions in the rule antecedents using the model input data. 2) Calculate the parameters of the consequences. They used least-squares estimation to estimate these parameters. The results were positive.

An intraday financial trading system powered by a novel brain-inspired evolving Mamdani Takagi-Sugeno Neural-Fuzzy Inference System (eMTSFIS) was proposed by Ho et al. [10]. The eMTSFIS prediction model included synaptic mechanisms and information processing capabilities of the human hippocampus, making it more robust and flexible than existing econometric and neural-fuzzy forecasting approaches. The proposed system's trading strategy was founded on the moving-averages convergence/divergence (MACD) theory in order to provide buy-sell trading signals. By incorporating forecasting abilities into the computation of the MACD trend signals, the lagging component of the MACD trading rule may be addressed. Based on the SP500 Index, the experimental results showed that eMTSFIS could produce extremely accurate forecasts and that the resulting system could identify profitable trading opportunities while reducing inefficient trading transactions. Better multiplicative returns for investors were made possible by these features of the eMTSFIS-based trading system. In order to create a news corpus, Nagar and Hahsler [11] offered a novel method for combining news from several sources using an automated text mining technique. After the corpus was filtered to exclude relevant sentences, it was analysed using natural language processing (NLP) techniques.News Sentiment, a sentiment metric based on the number of words with positive and negative polarities, was introduced. resources for news collecting and compilation as well as a sentiment analysis tool were built using open-source resources. The researchers found a strong relationship between the real-time changes in stock prices and the temporal variations of news sentiment. A five-layer self-organised neuro-fuzzy model was used by Su et al. [12]. This methodology models the dynamics of the stock market using technical indicators. A dataset with four indicators—Volume Adjusted Moving Average (VAMA), Stochastic Oscillator, and Ease of Movement (EMV)—basically taken from TAIEX was used to verify the model's forecasting and prediction skills. A suggested modification of the moving average technique was presented to improve stock price prediction by creating the input set for the neuro-fuzzy model. The results of the simulation showed how well the model predicted and how accurate it was. The modified moving average method's input mistakes were greatly reduced by the neuro-fuzzy model, improving prediction results.

An expert system for stock price analysis based on type 2 fuzzy rules was put into place by Zarandi et al. [13]. Each feature membership value was represented as an interval in a fuzzy logic system of interval type 2 to allow for modelling rule uncertainties. Technical and fundamental indices were included as input variables in the proposed type 2 fuzzy model, which was tested for its ability to forecast the stock price of an Asian automaker.The model's effectiveness in precisely forecasting price changes across a variety of stock sectors was demonstrated by extensive experimental testing. The results showed great promise and were included into a real-time trading system to predict stock values during periods of active trading.

Sim et al. [14] introduced 3D subspace clustering, a novel technique for creating algorithms for choosing cheap stocks. When it comes to managing high-dimensional financial data, 3D subspace clustering is incredibly effective and adaptable to new datasets. This method produces outcomes that are simple to understand and is immune to human prejudices and emotions. XII found that utilising criteria based on 3D subspace clustering algorithms, including CAT Seeker and MIC, produced earnings 60 percent higher than those obtained using Graham's principles alone during a 28-year trial period in the stock market (1980 to 2007).

Chowdhury et al. [15] developed a prediction-based methodology to analyse the sentiment of each investor or corporation after the market closes. The initial stock price movement and the effective positive index curve generated by the predictive news mining model described in this article showed a significant correlation throughout a 4-week sentiment tracking period. The analysis revealed a strong correlation—roughly 67%—between the first stock price curve and news sentiment. The efficient market theory is highly supported by this association, which shows that changes in stock prices clearly reflect sentiment. The significance of social media factors in the context of sentiment analysis for stock prediction and the e-commerce industry was demonstrated by Sarma et al. [16] and Nalabala et al. [17]. A comparison of the use of different machine learning algorithms in stock prediction is given by Wahyuningsih et al. [18]. Tudor [19] tests the Romaninan stock market to determine the significance of sentiment analysis. Several metrics utilised for performance analysis in prediction are explained by Lenten et al. [20]. The focus of Venturini [21] is on the several risk factors that are involved in prediction. Nalabala and Nirupamabhat expound on the significance of correlation in prediction [22]. The significance of sentiment analysis is explained by Gupta et al. [23].

According to a review of the literature, stock predictions in real-time settings can be made using sentiment analysis of financial news. Many real-time applications may be limited by the standard datasets used by the majority of the academics. Sentiment analysis employs a variety of methods, including natural language processing, fuzzy logic, clustering, and classification. Because of its effective performance, natural language processing is employed prudently, according to the survey.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

The most effective and precise techniques for stock prediction based on financial news are AIFFN and TF-IDF. These techniques get over the drawbacks of vector-based prediction techniques and data storage. In comparison to typical datasets and analysis used for stock forecasts, it can be stated that real-time financial news retrieve and prediction will yield better prediction and good sentiment analysis for stock data.
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.

III. OVERVIEW OF THE SYSTEM

The system architecture used to implement the research is depicted in Figure 1 along with the system overview. The financial industry news is stored in the backend. The news is retrieved from a number of popular websites. The webpages save data in a database and are runtime. Financial feeds are processed using natural language processing.

A. Data flow diagram

The data flow diagram for the system used in this study is shown in Figure 2. The data flow between the project and the many websites is depicted in this diagram. Data from different websites is gathered in real time via web scrapers, preprocessed, and stored in a database for future use. The final prediction is found via natural language processing. This will assist with stock prediction sentiment analysis.

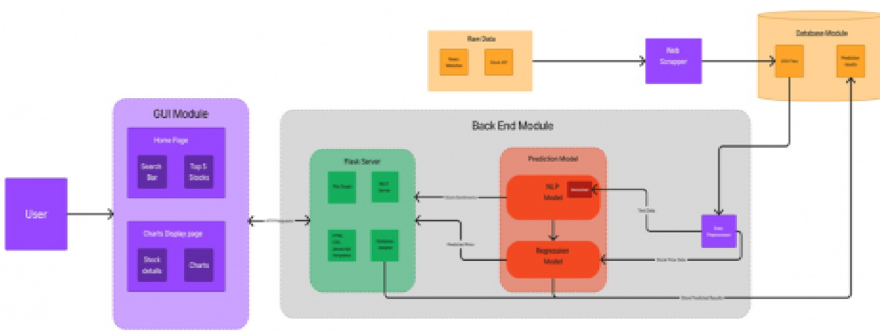

Figure 1. System architecture

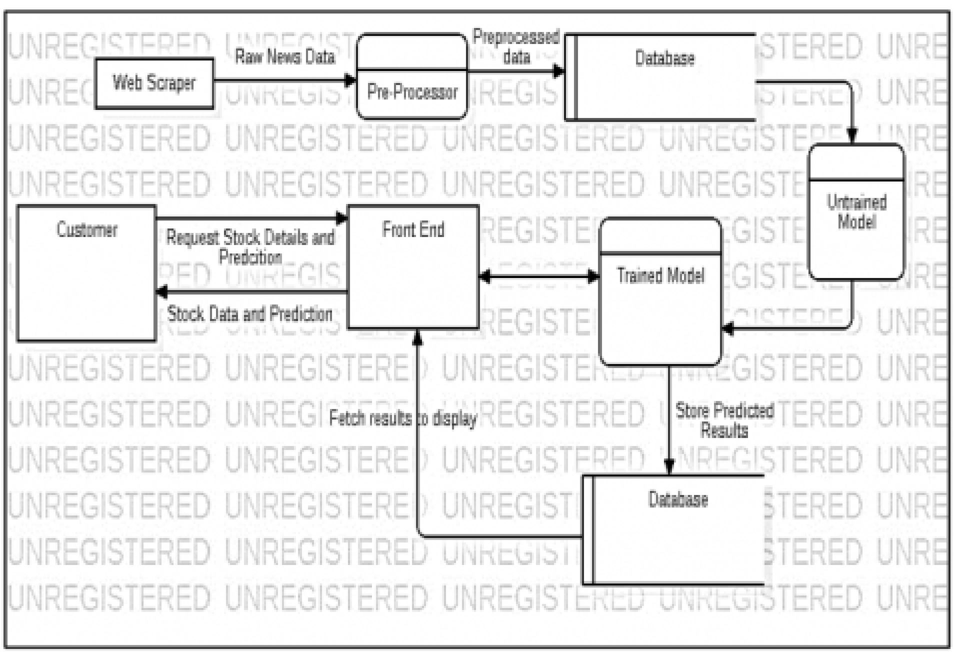

Figure 2. Data flow diagram

B. Component design and interaction

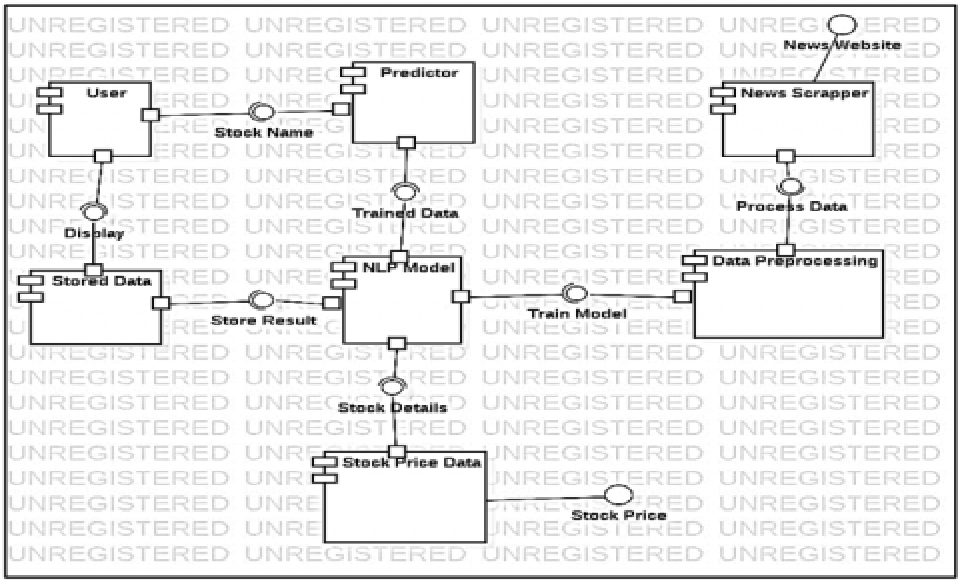

Figure 3. Component diagram

The interaction diagram throughout the implementation of the different components in this system during the research is shown in Figure 3. To obtain the final sentiment analysis and prediction, all of the modules are merged. The interaction diagram between the several system modules from this study is shown in Figure 4. In order to develop final predictions, the modules interacted with one another by sending input and intermediate outcomes. The input is created in real time based on news headlines from different websites. Sentiment analysis is performed via natural language processing.

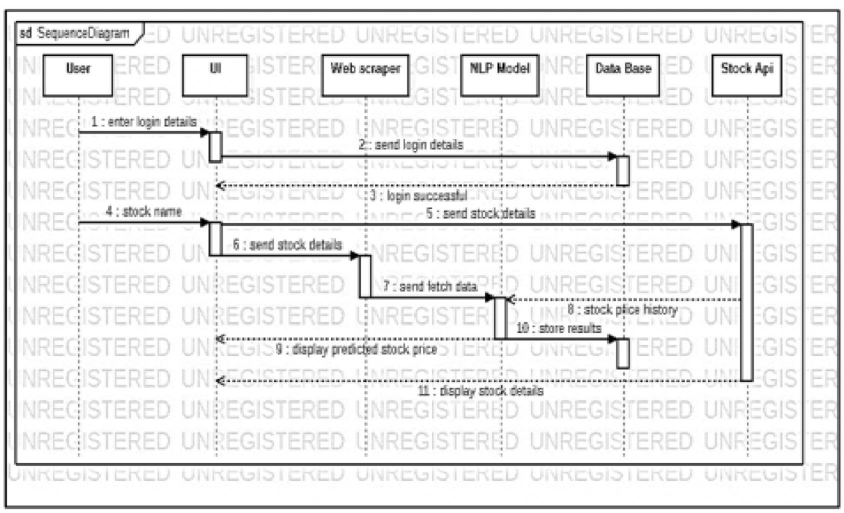

Figure 4. Interaction diagram

C. Tools, technologies and libraries

The majority of the techniques in this study are employed in accordance with the requirements when putting a stock prediction system into practice. Together, these technologies are able to gather data from multiple websites and store it on the database server. Information about the stock price is dynamically gathered using the threshold value. Information is gathered using a scraper and then subjected to additional evaluation and analysis in Python. The final forecasts are obtained using a variety of libraries. Table 1 provides a summary of all the tools and libraries that were utilised, as shown below. The tools and technology used for various functionalities are represented in Table 1.

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

D. Data sources

A range of sources are used to collect the information for this project. For the purpose of training, Kaggle supplied the news headline data, particularly the "Indian Financial News Articles (2003-2020)" dataset by hkapoor. We collected news headline data from the websites of Money Control, Economic Times, and Business Standard in addition to the Kaggle dataset. Using web scraping techniques, we were able to obtain the news headlines from these websites. Fyers, which provides an API for acquiring historical stock data, provided us with stock data. We collected data on a range of stocks that are traded on Indian stock exchanges, such as the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE).

E. Data cleaning and formatting

To ensure the highest quality and accuracy of the news headline data used in this study, many steps were performed. In order to minimise word diversity and improve consistency, the data was first lemmatised, which reduced words to their most basic form. The frequent terms that did not significantly add meaning to the text were then removed using stop word removal. In order to eliminate unnecessary characters that could impede the analytical process, punctuation was also eliminated. Lastly, the term frequency-inverse document frequency (TF-IDF) approach was used to vectorise the data, converting the text into numerical values for easier analysis. Finding the right feed and cutting down on processing and overhead time are two benefits of this method.

F. Data storage and retrieval

The financial news headlines were saved in both CSV and JSON formats to encourage usage and compatibility with various computer languages. The pd.read function of the Pandas library was used to retrieve the data, which made it easy and rapid to put the data into memory for training.

IV. FEATURE SELECTION AND EXTRACTION

A. Feature selection techniques

To find the most pertinent and instructive aspects for the models that were found, a variety of feature selection strategies were used in this system on the prediction of sentiment analysis and stock data using financial news headlines.

Correlation-based feature selection was one of the methods that were used. We were able to choose characteristics that significantly influenced sentiment analysis and stock performance prediction by using this strategy to examine the link between each feature and the goal variable. aims to improve the precision and effectiveness of our prediction and analysis models by locating and choosing the most important features.

B. Feature extraction methods

It used feature extraction techniques to glean representative and significant information from the financial news headlines. TF-IDF (Term Frequency-Inverse Document Frequency) vectorisation was one of the main approaches employed. A popular method in natural language processing, TF-IDF assigns weight to words according to their prominence and frequency in a corpus of documents. The system will be able to determine the significance of each word in the headlines by converting the textual data into numerical feature vectors using TF-IDF. This will enable more precise sentiment analysis and prediction. The inability of TF-IDF to assist in determining a word's semantic meaning is one of its limitations. Semantic meaning, however, is crucial in sentiment analysis in order to produce either positive or negative feelings. The user might decide on its financial stock forecasts based on these feelings.

C. Feature engineering approaches

In addition to feature selection and extraction, a variety of feature engineering techniques were applied in this system to enhance our models' predictive power. In essence, feature engineering is the process of identifying intricate patterns and characteristics in data. This comes after the procedure for adding new features or changing ones that already exist. In research, systems employ a variety of feature engineering techniques tailored to market data and financial news. One of these methods was emotion word embedding, which captures the sentiment polarity and intensity of news headlines by encoding sentiment-related information into numerical vectors. It was thought that by adding such manufactured qualities, the stock prediction and sentiment analysis system would perform better and be more resilient.

Overall, this system used a variety of feature selection techniques, feature extraction techniques like TF-IDF vectorization, and feature engineering techniques like sentiment word embedding in an effort to maximize the representation and relevance of features in the Stock Data Prediction and Sentiment Analysis system. Word embedding uses a vector-based matrix to determine if feelings are positive or negative. By employing these strategies, the system was able to improve the accuracy of its models, extract important information from financial news headlines, and gain a better understanding of sentiment patterns and stock market dynamics.

V. SENTIMENT ANALYSIS

A. Sentiment analysis methods

Both lexicon-based and vector-based sentiment analysis methods were examined in this system. In the vector-based technique, the system converted each news headline into a numerical vector using term frequency-inverse document frequency (TF-IDF). Each vector was then classified as positive, negative, or neutral using logistic regression. subsequently assigned scores to each word in the news headline using a pre-established sentiment lexicon, adding together these scores to generate an overall sentiment score using the lexicon-based technique.

B. Sentiment lexicon and dictionary based approaches

This system is being tested using a variety of dictionary-based methods and sentiment lexicons, such as the SentiWordNet lexicon and the AFINN lexicon. The use of specially designed emotion dictionaries for specific industries, as well as financial and stock market topics, was also examined by the system. The system's performance using the AIFFN Lexicon approach is shown in Table 2. A number of metrics, including precision, recall, and F1-score, are used to assess performance. The system's stability and robustness for stock predictions are evaluated using these parameters. Two classes—negative and positive—are used to evaluate performance. Based on the table, we can conclude that this approach improves performance in both positive and negative classes.

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

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

The system performance for the AIFFN Lexicon approach is shown graphically in Figure 5. The graph shows both positive and bad results for both courses.

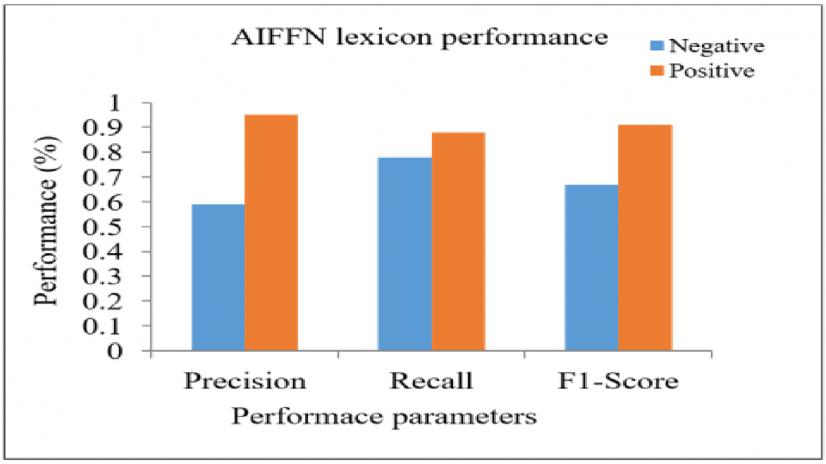

Figure 5. Graphical representation of AIFFN lexicon

C. *Vector based approaches*

The system uses logistic regression to estimate sentiment labels after using the TF-IDF methodology for vector-based methods. The system was evaluated using a variety of hyperparameters, including regularization strength and maximum number of iterations, in order to enhance the performance of our models. The performance of the TF-IDF method using vector-based techniques is shown in Table 3. A number of metrics, including f1-score, precision, and recall, are used to calculate performance. According to the table's numbers, this approach performs well for both groups. The TF-IDF method's graphical depiction is displayed in Figure 6.

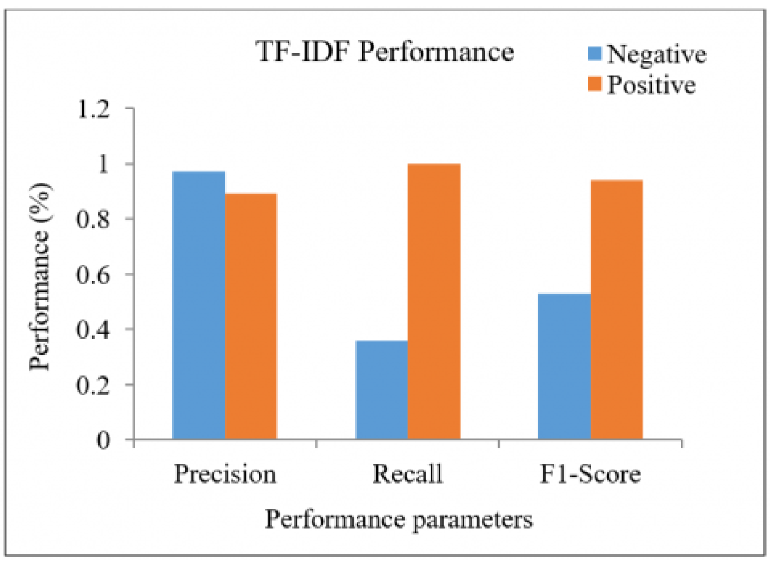

Figure 6. Representation of TF-IDF

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

D. Evaluation metrics

The system uses the model to calculate the accuracy of sentiment analysis models in terms of assessment metrics. score system. Accuracy, recall, and F1 score are also used to evaluate the models' performance in identifying both positive and negative thoughts. After examining the provided statistical data, it was found that TF-IDF outperformed AFINN in terms of accuracy, recall, and F1-score based on system performance. The precision, recall, and F1-score for the positive class of TF-IDF were much higher than those of AFINN, indicating that it is more accurate at correctly classifying positive thoughts. The overall F1-score was greater for TF-IDF, suggesting a better balance between accuracy and recall, even if it had a lower recall for the negative class. Moreover, TF-IDF considers a term's frequency and rarity in the corpus when assessing its significance in a text. As a result, the unique characteristics of financial news headlines and their impact on sentiment analysis can be captured using TF-IDF. Contrarily, AFINN relies on a pre-established sentiment vocabulary, which might not adequately represent the nuances and complexity of financial terminology. Considering these factors, TF-IDF was chosen as the system's sentiment analysis method due to its superior performance in accurately detecting emotions and its capacity to capture the significance of words in financial news headlines.

VI. STOCK DATA PREDICTION

A. Prediction model

The system looked at a variety of models, such as those based on machine learning, time-series analysis, and regression, and chose the most effective model for predicting stock prices.

B. Regression based models

After investigating a variety of regression models for stock data prediction, we decided to use linear regression as one of them. Because of its simplicity, readability, and capacity to identify linear relationships between variables, linear regression was selected. By fitting a linear equation to the data, it is anticipated to uncover underlying patterns and trends in the stock market. The efficiency of the Linear Regression model was evaluated using the R-squared statistic, which shows the proportion of the stock price variance that the model can account for. With an astounding 96% accuracy rate, the Linear Regression model appears to have grasped the relationship between the specified variables and the stock price. It is important to remember, nonetheless, that the accuracy might change based on market conditions and datasets, and further testing and validation are required to assess its efficacy in different scenarios. However, the results highlight the potential of Linear Regression as a reliable method for predicting stock data. Since online data is nonlinear, it must be transformed into the precise vector mentioned in the preceding section in order to be used for sentiment analysis.Additionally, the dynamic nature of real-time data collecting necessitates feature extraction for sentiment analysis. Regression models are more effective at predicting stock data, despite the fact that the data is dynamic and nonlinear. Every characteristic in real-time data capture is crucial for sentiment analysis, while a linear-based stock prediction model is created using the features that are more polarized toward positive feelings.

C. Time series analysis based models

The Long Short-Term Memory (LSTM) recurrent neural network (RNN) was one of the models we built as part of an additional approach to this system that looked into time series analysis-based models for stock data prediction. Because of its capacity to effectively handle data sequences and capture temporal correlations, LSTM was chosen. In order to use historical stock data and identify patterns and trends over time, we intended to employ LSTM. Relevant measures like Mean Squared Error (MSE) or Root Mean Squared Error (RMSE) were used to evaluate the accuracy of the LSTM model. With a 71% accuracy rate, the LSTM model shows some degree of stock value forecasting ability. It is important to remember that the dataset, LSTM model architecture, and hyperparameter adjustments will all affect the accuracy. To maximize performance and assess the suitability of the LSTM model for diverse stock market conditions and datasets, more research and testing are required.

D. Evaluation metrics

The system used evaluation metrics including Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to examine the precision and effectiveness of our prediction models. By calculating the average squared difference between expected and actual stock prices, these metrics provide a quantitative evaluation of the model's accuracy in stock price forecasting. In order to evaluate the prediction models' overall effectiveness and ability to account for volatility in stock price data, the system also examined metrics including Mean Absolute Error (MAE) and R-squared (R2). When both models are utilized in an ensemble manner, sentiment analysis is essential to the prediction of stock data.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

In order to increase system performance, both models assist in compromising each other's shortcomings. These evaluation tools helped us compare the performance of our prediction models and evaluate their dependability and effectiveness.

VII. INTEGRATION AND IMPLEMENTATION

A. System integration and deployment

The frontend and backend components were carefully integrated throughout implementation to create a robust communication route between the underlying functionality and the user experience. To create a visually appealing and captivating frontend, we utilized HTML for structural elements, CSS for style and layout, and JavaScript for dynamic and interactive features. On the backend, we employed the Flask framework, which provided a versatile and lightweight environment for building web applications. Flask's modular architecture is used to successfully structure the system's backend code, enabling quick data processing, request handling, and routing.To ensure data integrity and easy data modification, we utilized MySQL as the database management system to store and retrieve crucial data. Through extensive testing and debugging, we made sure the integrated system operated flawlessly and was prepared for field deployment. By appropriately normalizing each dataset and adhering to all transaction properties, normalization helps to maintain the consistency and integrity of data. The purpose of data preprocessing is to preserve the integrity of real-time data and clean it up. The backend table structure utilized in this system's implementation is depicted in Figure 7.Using real-time data capture for financial news, backend databases are utilized to hold the news feed that is received from multiple websites. All of the news is transformed into CSV and JSON formats so that it may be stored on the database server.

Figure 7. Back-end structure in database

B. User interface and functionality

For those interested in sentiment analysis and stock data prediction, the system's user interface (Figure 8) aims to provide a pleasant and simple experience. A navigation bar on the UI offers easy access to different system components. Among the sections are the charts, about us, and homepage. On the homepage, the top five stocks of the day are displayed in individual cards. This piques customers' attention and provides a brief overview of the trending stocks. Important details about each stock, including its symbol, current price, and a brief description, are included on the cards. A knowledge base regarding our system and its function can be found on the "about us" page. We go into great detail about the algorithm's usage of sentiment analysis and stock data projection based on financial news headlines. We also provide contact details so that users can get in touch with us with any queries or grievances. Visitors can view comprehensive financial charts and details about the selected company in the charts area. The system will generate interactive charts that show past price trends, volume, and other relevant financial data after users select a specific stock of interest.Additionally, the system shows consumers the anticipated price of the selected stock using its stock data prediction capabilities. In general, the user interface aims to provide users with an experience that is both aesthetically pleasing and practical.

Customers can swiftly navigate through the system's several sections, obtain crucial stock data, and gain insights through charts and estimates. The top five records and URLs chosen from the online or real-time feed as input are shown in Figure 8. These websites will be given top priority when it comes to delivering news as input for the prediction algorithm.

The specific stock's detailed information is shown in Figure 9. It depicts the change in stock prices based on precise real-time data. It facilitates the user's pursuit of sentiment analysis and prediction.

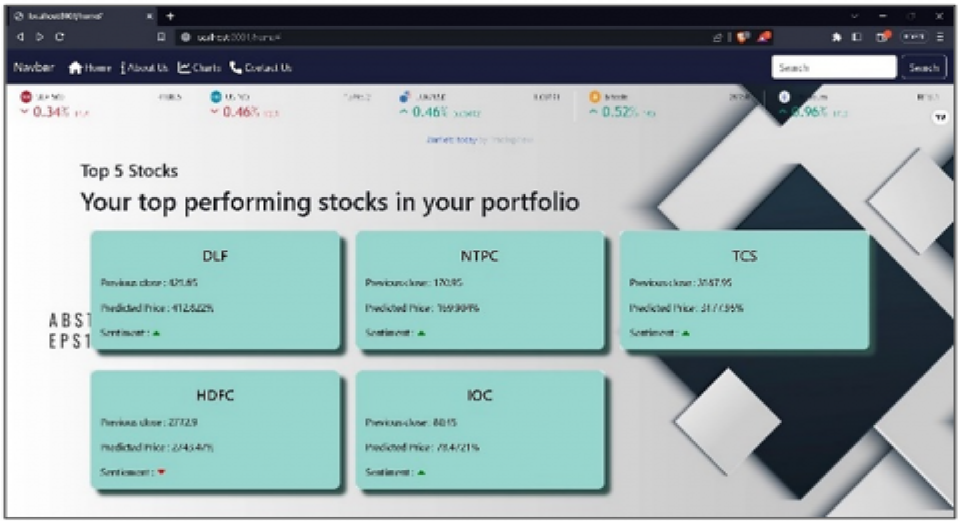

Figure 8. System homepage

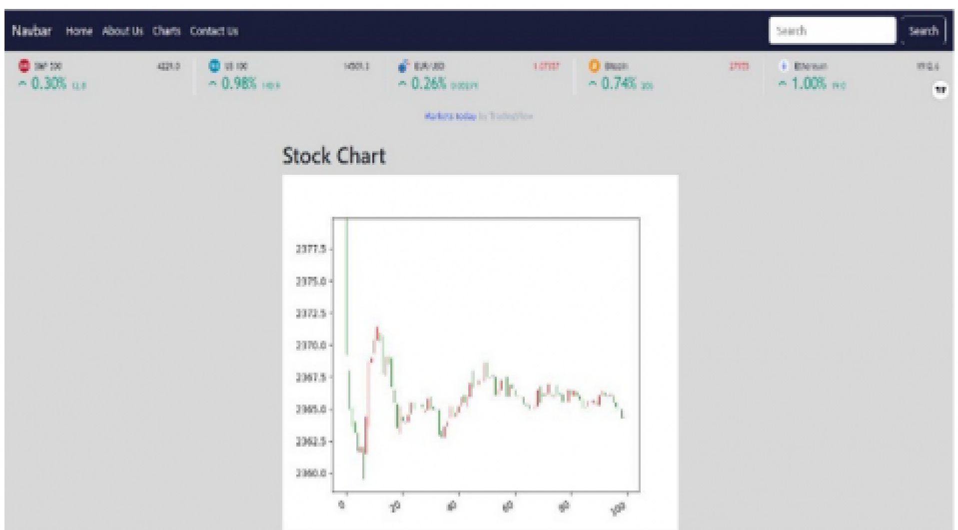

Figure 9. Stock detail page

C. System performance and scalability

To increase system speed, a number of optimizations were made. Predictions could be saved, which reduced the amount of time needed to compute them again. This was one of the biggest innovations. By saving the predictions, the system aims to reduce the need to calculate them each time a user requests the data, which improves response times. Additionally, it is employed as a backend mechanism for chart viewing, where the charts were produced and stored as image files.This technique allowed for quick chart retrieval and presentation while reducing front-end processing time. Additionally, by reducing the amount of dataset required to create the charts, the solution improved the rendering process and enabled faster rendering times without compromising the accuracy or quality of the charts.

Scalability: Because of its scalability, this system might be deployed on the cloud. Among the many advantages of the cloud are its scalability, dependability, and affordability. By using cloud infrastructure, the system can handle growing user traffic and accommodate more users without suffering from a significant decline in performance.

Furthermore, by enabling the dynamic allocation of resources based on demand, cloud deployment makes straightforward scalability possible. This ensures the system can scale up or down as needed, offering optimal performance and cost-effectiveness when traffic is light. Additionally, redundancy and high availability are frequently integrated into cloud-based services, ensuring that our system will continue to function even in the event of hardware failure or interruption.

In conclusion, our system is now faster thanks to the implementation of optimizations like backend chart creation and prediction storage. Better system performance and faster reaction times are the outcomes of these enhancements. Additionally, the system's cloud deployment characteristics, which enable efficient resource allocation and accommodating growing user demand, enhance its scalability.

VIII. RESULTS AND EVALUATION

A. Data analysis and visualization

Word embedding is made possible via vectorization in sentiment analysis, a sort of data analysis. Feature extraction is used to help validate real-time data. This study used a comprehensive data analytic approach to glean insights from market data and financial news headlines. Preprocessing the stock data entails managing missing values and outliers while also guaranteeing data quality. The properties of the data were examined using exploratory data analysis (EDA) approaches. The data's central trends and dispersion were examined using statistical metrics such as mean, median, and standard deviation. To identify patterns and trends in the changes in stock prices over time, time series analysis techniques such as trend analysis and autocorrelation were used. Effective presentation of the results required both data analysis and visualization. A range of visualization tools and approaches were used to visually show the relationships and patterns in the data. Line charts were used to display previous stock prices and track their patterns. The association between stock prices and sentiment analysis ratings obtained from financial news headlines was investigated using scatter plots. Heatmaps were used to display sentiment distributions across several time periods or stock categories. These visual aids made it simpler to comprehend the dynamics of stock market data and how sentiment analysis affects stock performance.

B. System performance evaluation

The system designed for sentiment analysis and stock data prediction was put through a rigorous performance review to determine how effective it was. Financial news headlines and historical stock data were used to assess the system's prediction abilities. Throughout the evaluation process, the data was separated into training and testing sets. The prediction model was trained using the training set, and its performance was assessed using the testing set. A number of evaluation metrics were employed in order to evaluate the precision and dependability of the system's forecasts. Mean absolute error (MAE) and root mean square error (RMSE) were calculated to quantify the discrepancy between expected and actual stock values. The average magnitude of the predicted errors was disclosed by these metrics. Additionally, the system's ability to correctly classify the sentiment of financial news headlines was assessed using accuracy measures. Recall, precision, and F1-score were computed to evaluate sentiment categorization performance.

C. Comparison with existing methods

The proposed method is compared to existing methods for sentiment analysis and stock data prediction. A comprehensive literature review is conducted to find previously published or used pertinent research and methodology in order to support the results. The system focuses on common contemporary methods that addressed similar objectives and employed comparable data.

We tested and assessed the performance of several existing methods using the same test dataset and evaluation standards as our system. Their forecast accuracy, sentiment classification, and overall system performance are compared to our recommended method. Through this comparison, the system's effectiveness and superiority in terms of sentiment analysis accuracy, prediction skills, and overall performance are demonstrated.

D. Interpretation of results

The system provided a thorough interpretation and analysis of the data analysis, evaluation of the system's performance, and comparison with the results of existing approaches.

Study of every pattern and insight discovered by analyzing stock data, including seasonality, recurring trends, and correlations with market events. The system also examines the sentiment analysis's findings, highlighting any fascinating findings or connections between sentiment and stock performance.

Additionally, this study evaluated the system's flaws and potential areas for improvement. The performance of the system may have been affected by how certain difficulties, such as data quality, model assumptions, and inherent uncertainties in stock market dynamics, were handled. The system also considered potential biases or limitations in the sentiment analysis process, such as problems with the accuracy of sentiment categorization or the influence of news sources.

Offering a comprehensive understanding of our system's potential, constraints, and implications for real-world stock prediction and sentiment analysis applications was the aim of the findings interpretation. The technique could be improved for stock exchanges, where real-time dynamic sentiment analysis is possible. Future improvements could include the use of sentiment analysis in real-time stock market trading or large-scale stock prediction.

IX. DISCUSSION

A. Implications of the system

The ramifications of the developed approach for sentiment analysis and stock data prediction are significant in a number of ways. First off, incorporating sentiment analysis of financial news headlines into the stock prediction process provides valuable information on investor activity and market sentiment. This makes it possible for investors to make more informed decisions and possibly lower risks. Second, the system advances the field of financial technology by demonstrating the viability and value of combining machine learning and natural language processing techniques in sentiment analysis and stock prediction. This opens up new avenues for research and development in the domains of technology integration and finance. Individual users will be greatly impacted by sentiment analysis, and they may base their financial decisions—whether favorable or unfavorable—on it. A positive sentiment indicates a rise in the stock forecast, while a negative sentiment indicates a decline. The user finds these sentiments useful when making financial judgments.

B. Limitations and future work

Despite the system's apparent promise, it is important to understand its limitations and pinpoint areas that require more work. Sentiment analysis's reliance on the precision and caliber of financial news data is one drawback. Sentiment categorization accuracy may be impacted by biases and data noise. The system's performance may also be impacted by shifting market dynamics and economic factors. In order to effectively adapt to shifting market conditions, future research may focus on enhancing sentiment analysis techniques, incorporating additional data sources, and improving prediction models. To address these limitations and guarantee the system's long-term utility, ongoing performance evaluation and monitoring are necessary.

X. CONCLUSION

In this study, a completely functional system for sentiment analysis and stock data prediction using financial news headlines is presented. By combining techniques from financial analysis, natural language processing, and machine learning algorithms, the system shows promise in supporting users' decision-making. The main takeaway from this study is sentiment analysis, which uses word embedding and vectorization to create sentiment from real-time data that is fed into the system. In order to help users estimate whether a financial news headline will be positive or negative, this algorithm extracts information from a variety of websites. Combining sentiment analysis with stock forecasting aids users in making informed financial decisions on equities. The study highlights how important sentiment analysis is to understanding actual data and how it affects stock performance. Despite these shortcomings, the system offers a solid foundation for further research and advancement in the field. Testing this approach for large stock exchange projects and investors on a large scale could lead to further improvements. Nonetheless, this financial technology system can offer crucial information on integrating data-driven strategies in the stock market.

REFERENCES

[1] 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

[2] 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

[3] 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

[4] 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

[5] 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

[6] 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

[7] 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

[8] 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

[9] 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

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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.

[15] 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.

[16] 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

[17] 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

[18] 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.

[19] 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

[20] 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

[21] 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

[22] 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.

[23] 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