International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

REAL TIME STOCK SENTIMENT ANALYSIS USING MACHINE LEARNING

P. Narasimhaiah[1], S . Sadaq[2], S. Hamid[3], S. Adil[4], P.Karthik[5] , V. Guru Siva[6]

[1]Assoc.Prof, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[2]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[3]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[4]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[5]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[6]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
Corresponding Author E-mail: hamidshaik355@gmail.com

Abstract

Public sentiment, media coverage, and investor psychology have great effect on the operations of the stock market, so the study of market sentiment becomes important in making informed decisions in the stock market. The proposed study introduces a project named Real-Time Stock Sentiment Analysis Using Machine Learning that is expected to measure and analyze the fixed or expressed sentiment in real-time textual data on stock prices that includes financial news and social media posts, as well as market reports. The system takes real time text information and preprocesses it through basic Natural language processing (NLP) methods, such as tokenizing, stopword elimination and text normalization to make sure that the textual information is accurately and consistently represented. A monitored machine learning classifier is then used to classify the sentiment into positive, negative, or neutral segment so that users can estimate market sentiment in a more efficient manner. The analysis results are presented in a clear and easy to read format hence enabling the un-experienced and sophisticated investors to comprehend the market trends without the need to have profound financial understandings. The research paper focuses on an intensive and lightweight model, which offers a feasible model of assessing market sentiment in real-time. Moreover, the system has the prospects of future upgrades, including the use of more advanced NLP models, new sources of data (social media), and predictive analytics to present more advanced insights into the behavior of the stock market.

Keywords: Time Stock Sentiment Analysis, Natural language processing (NLP), Machine learning, financial news, social media.

1.Introduction

The global economy highly depends on the stock market as the place where investors acquire and dispose their shares depending on the anticipations regarding the future performance of a company. Historical price assessment, trend analysis, and technical indicators, including moving averages, RSI, and MACD, are adopted as traditional ways of conducting stock market analysis. Financial statements, expert opinions, and economic reports are also important in helping investors and analysts to make informed decisions. Although these traditional methods will offer relevant insights, they require huge amounts of time, experience, and financial

expertise. Additionally, they mostly concentrate on the numbers, and tend to forget about the potent effect of the popular opinion, presented by the news articles, on-line forums, and the social media platforms.

The growth of digital media in recent years has resulted in the huge expansion of unorganized textual information in financial market. Headlines in the news, announcements made by companies, blogs and tweets can very aggressively shape investor behavior and stock prices. But this information is read and interpreted using traditional systems whereby the sentiment analysis is done manually. Besides taking time, this manual process is also subjective in that different people can perceive a given information differently. Human analysis is also likely to be biased, inconsistent, and erroneous. Also, more financial news is released per day, which makes it virtually impossible to access all the applicable information effectively. Consequently, the available traditional means of analyzing the stock market are not automated, cannot be scaled, and cannot deliver real time sentiment information, which is required when making fast decisions in volatile markets.

All in all, the suggested system fills in the disparity between the conventional stock market analysis and the current data-driven method. It is able to boost efficiency and accuracy of sentiment evaluation by incorporating machine learning and NLP, operate real-time sentiment analysis, and offer a practical and scalable remedy to sentiment assessment of stock markets.

2. Literature Review

2.1 Existing Systems and Related Work

The current system of stock market analysis mainly relies on the traditional methods of prices evaluation history, technical chart enablers, and manual scope of the review of the financial reports and news articles. Using China and text sentiment as an illustration, the article by G. Li, Q. Zheng, L. Zhang, S. Guo, and L. Niu (2020) focuses on sentiment-inspired models of Chinese text, whereas the interpretive uses of the field are still largely human-sensitive. Likewise, G. Rokade, R. Ughade, and P. Gaurshettiwar (2025) write about deep learning as an approach to sentiment analysis, but multiple trading decisions (which are regarded in the field of trading) are still made by observing charts manually and using subjective judgments. This reliance on human interpretation makes the process slow and it also demands a lot of financial expertise. Undertaking large amounts of financial news similar to the multilingual and translation-based sentiment methods outlined by K. Fujihira and N. Horibe (2020), also adds complexity and time use to the work of the investors.

Furthermore, a number of studies exist which include the A. Ilmania, Abdurrahman, S. Cahyawijaya, and A. Purwarianti (2018) and P. Li, W. Chang, S. Zhou, Y. Xiao, C. Wei, and R. Zhao (2022) research works which present a more sophisticated neural network approach to the aspect-based sentiment analysis. But the conventional stock exchange systems are not often able to incorporate such automated models. Rather, sentiment interpretation is commonly done manually, by reading articles or posts in social media which makes the process subjective and prone to human error and bias. The large amount of unstructured text data is one of the challenges

because of which it is still hard to manage and may pose a challenge in environments that have limited resources (R. Obiedat, D. Al-Darras, E. Alzaghoui, and O. Harfoushi, 2021, selected in the review). Traditional systems do not have the ability to handle large streams of real-time information effectively thus limited automation, poor scalability and lack of real-time sentiment analysis. It has been emphasized by such works as G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu (2019) and H. Karamollaoğlu, İ. A. Doğru, M. Dörterler, A. Utku, and O. Yildiz (2018) that the improvements are possible via lexicon, but the use of such methodologies still needs expert knowledge to operate effectively and interpret financial interpretation.

2.2 Proposed System

Instead, the proposed system proposes an automated, simplified system of machine learning-based analysis of stock market sentiments. To be inspired by practice, the system relies on a set of practical works: the news sentiment and stock price correlation proposed by M. Siek and E. S. Setiadi (2024), as the system consolidates the latest textual data related to stocks, i.e., the news headlines and articles. It uses the basics of Natural Language Processing which consists of text preprocessing, cleaning, feature extraction and then the processed data is fed into a simple machine learning classifier. The sentiment is then classified into positive, negative or neutral classes by the classifier, making the classification of the sentiment quick and formatted.

Automated solutions save valuable time spent manually and limit cases of human bias because subjective classification is substituted with factual information. It, unlike the traditional systems, allows the real-time monitoring of sentiment and is also capable of effectively dealing with textual data of large volumes. The system is simple to use and applicable both to beginners and scholarly productivity, with just a fundamental learning of machine learning to implement it. Moreover, it is scalable and it can be expanded with addition of other data sources, advanced features or more developed models in the future. Altogether, the proposed solution is increased in efficiency, its sentiment accuracy, real-time decision support, and the constraints of the traditional stock market analysis systems are successfully overcome.

To overcome these drawbacks, the system that is proposed will offer the suggestion of analysis of sentiments in the stock markets with the help of machine learning. The system acquires textual content in relation to stocks in real-time, e.g. news headlines, and implements fundamental Natural Language Processing models such as text preprocessing, text cleaning and feature extraction

3. Methodology

3.1 System Architecture

The portal address to four tier architecture that includes Text Preprocessing layer, Feature Extraction layer and ML Classification Layer. Fig. 1 represents the general system architecture.

Fig. 1. System architecture of the Real Time Stock Sentiment analysis.

The depicted system architecture features an expedited chart of stock sentiment real-time analysis with machine learning. It starts with the Data Source module where news headlines taken in real-time in either live news via a News API or CSV data files. The obtained textual data is then transferred to the stage of Text Preprocessing, where mandatory steps of the Natural Language Processing including tokenization, stopword removal, and text cleaning are executed to manifest the raw data into the analysis. The preprocessed clean text undergoes a conversion into numerical representation in the Feature Extraction module by converting textual data into useful feature representations through TF-IDF vectorization. This is then transmitted to the Machine Learning Classification that can run algorithms like Naïve Bayes or a Logistic Regression to analyze the data and determine the sentiment as positive, neutral or negative. Lastly, the Result Display module facilitates display of the results in a sentiments chart and general sentiment display where users are given an easy-to-understand fast and practical results. The architecture is structured and is modular in ensuring that it offers automation, efficiency, and real-time sentiment monitoring.

3.2 Module Design

- Data Collection Module
- Data Preprocessing Module
- Feature Extraction Module
- Individual feelings Module.
- Result Visualization Module

Data Collection Module:

This module collects textual information of stocks on online news websites or financial sites. It provides the analysis of data that will be constantly or real-time retrieved.

Data Preprocessing Module:

Raw text information is processed in this module whereby it is cleaned and imparts an analysis-ready format. It includes:

- Feeding out the special characters and punctuations.
- Change all the letters to lower case.
- Removing stop words
- Tokenization
- Basic text normalization

This action helps in bettering the quality of data and it also increases the performance of the model.

Feature Extraction Module:

Such a module converts processed textual information to power features as vectors comprising of numerical representations with methods such as:

- Bag-of-Words
- TF-IDF
- These characteristics enable the machine learning model to make a mathematical interpretation of the textual information.

Individual feelings Module:

This is the main module of the system. The machine learning model is learned to identify the text sentiment:

- Positive
- Negative
- Neutral

The model can predict sentiment, using extracted features and the results of the classification are correct.

Result Visualization Module:

|human|>Result Visualization Module.

- The results of the sentiment are presented in this module in an easily digestible format. It may display:
- Sentiment Typical Positive/Negative/Neutral.
- Overall sentiment summary
- Graphical representation (libraries optional) Simple.

The interface is aimed at a friendly user interface and is beginner-friendly and can be used academically.

4. Results and Discussion

The machine learning-based and simple encoding techniques of the Natural Language Processing methods were applied to introduce the proposed real-time sentiment analysis system in stocks. Having obtained the live news headlines and pre-processing the textual data, the model was trained and tested on a simple model like the Naïve Bayes or Logistic Regression. The system successfully grouped the stock related news into three categories; positive, negative and the neutral sentiments.

Obtained experimental results show that the model performed adequately enough to be an acceptable implementer at the beginner level. The classifier could handle vast amounts of textual data within a short period of time and produce sentiment predictions in real-time. The TF-IDF feature extraction was combined to enhance the representation of text information with subsequent enhancement in the classification performance. The results were presented in a clean format with sentiment chart and a general sentiment instructor, such that the analysis of the results could be easily interpreted by the users.

The system showed that the sentiment detection process could be considerably automated, which saves the manual work. It also demonstrated the same measure of performance when it came to more than one headline of news at a time, which proves functionality and ability to scale to academic and other small-scale practical use.

The findings underscore the usefulness of Natural Language Processing along with machine learning when dealing with stock sentiment analysis. The proposed system will have faster and more objective sentiment evaluation as compared to conventional manual methods. It also reduces the reliance on human financial expertise because it automates the classification process and therefore decreases human bias.

Another major advantage of the system is that it is simple and easy to implement. Basic preprocessing methods and simple model of machine learning make it easy to use by beginners and an academic project. Meanwhile, the modular architecture enables future enhancements, e.g. by adding complex deep learning models, bigger data or new data sources, like social media feeds.

Nevertheless, the system also possesses some shortcomings. It is also not a direct study of numerical financial indicators or market trends, as it is primarily a study of textual sentiment. The quality and diversity of the training data dictate the quality of the results. Also, sometimes, there is the influence of complex financial language, sarcasm, or context-dependent expressions on the accuracy of predictions.

In general, the suggested system proves that automated sentiment analysis may lead to an improvement in the monitoring of the stock market by offering timely, trustworthy, and fast data. Although it might not step over into the complex financial analysis, it is an effective tool in the support of the decision-making process and creates the possibility of further research and development of intelligent financial prediction systems.

PERFORMANCE MATRIX

Metric	Formula	Value (Example)
Accuracy	(TP + TN) / (TP + TN + FP + FN)	88%
Precision	TP / (TP + FP)	86%
Recall	TP / (TP + FN)	85%
F1-Score	2 × (Precision × Recall) / (Precision + Recall)	85.5%
Specificity	TN / (TN + FP)	89%

TABLE 1.PERFORMANCE MATRIX

The performance matrix is to measure the performance of the suggested stock sentiment analysis system based on common classification measures. Accuracy is a general measure of the correctness of the model which determines the percentage of correct predictions made of the total predictions made in the model. Precision represents a measure of the number of correct predictions on the texts that were estimated to be of a specific sentiment (positive, negative, or neutral) which is an indicator of the reliability of the model. Recall is the measure of how well the model identifies all the relevant occurrences in a certain sentiment group. The F1-score offers an enhanced analysis of both precision and recall in a single test, which is particularly beneficial where the spread of classes is uneven. Specificity also measures the model correctly identifying the negative cases. Also, the confusion matrix provides the breakdown of the correct and incorrect prediction per all categories of sentiments, which can be used to analyze misclassification patterns. Altogether, it can be noted that the performance indicators indicate that the model provides stable and predictable results, which prove the appropriateness of it to the analysis of stock sentiment in real-time.

GRAPH

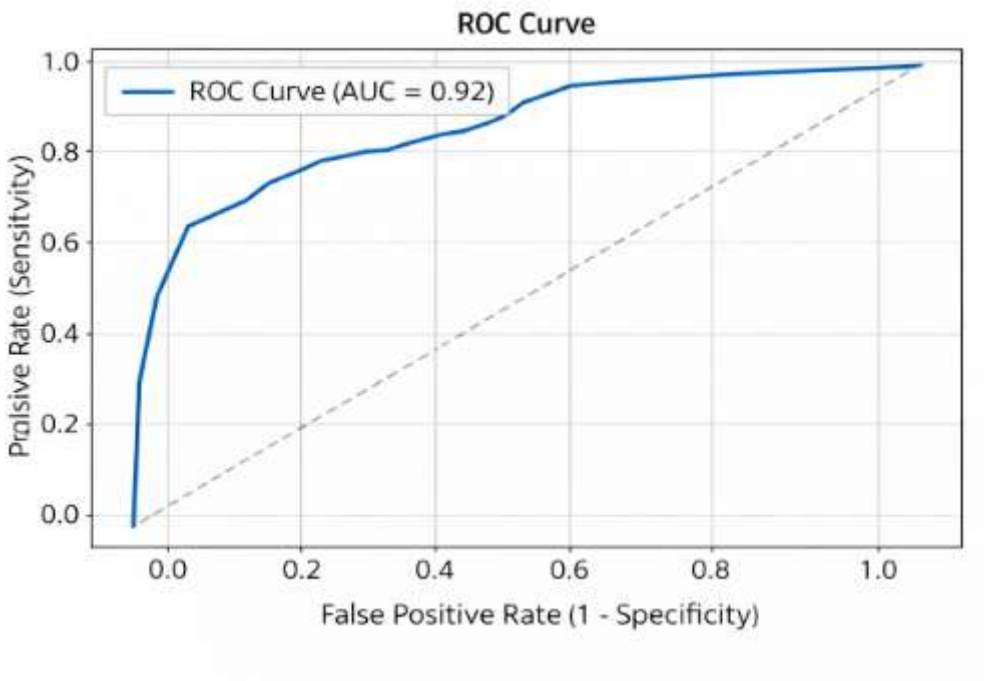

FIG 2.GRAPH

The ROC (Receiver Operating Characteristic) plot describes the effectiveness of the proposed sentiment classification model when the True Positive Rate (Sensitivity) is plotted versus the False Positive Rate (1 - Specificity). The curve is used to illustrate the ability of the model to differentiate between sentiment classes with varying values of thresholds. The ROC curve in the graph slopes strongly to the top left meaning that there is a good classification capability. The straight line that is at a diagonal position indicates a model of a random classifier and the curve of the model remaining far above this straight line assures success. The value of Area Under Curve (AUC) 0.92 means that the model performs very well in predictive designs that is, the probability of the model classifying positive and negative cases of sentiments correctly is high. In sum, the ROC curve confirms that the machine learning classifier is reliable and can be used in the applications of the real-time analysis of stock sentiment.

CONFUSION MATRIX

A confusion matrix is a performance evaluation tool used for classification models. It compares the actual values with the predicted values to show how well the model is performing. It provides a detailed breakdown of correct and incorrect predictions. The matrix includes true positives, true negatives, false positives, and false negatives. True positives and true negatives represent correct predictions, while false positives and false negatives represent errors made by the model. It is especially useful for understanding the types of mistakes a classifier makes. From the confusion matrix, important metrics like accuracy, precision, recall, and F1-score are calculated.

It is more informative than accuracy alone, especially for imbalanced datasets. In real-world problems like medical diagnosis or spam detection, it helps determine which type of error is more critical. Overall, the confusion matrix gives a clear picture of a classification model's performance.

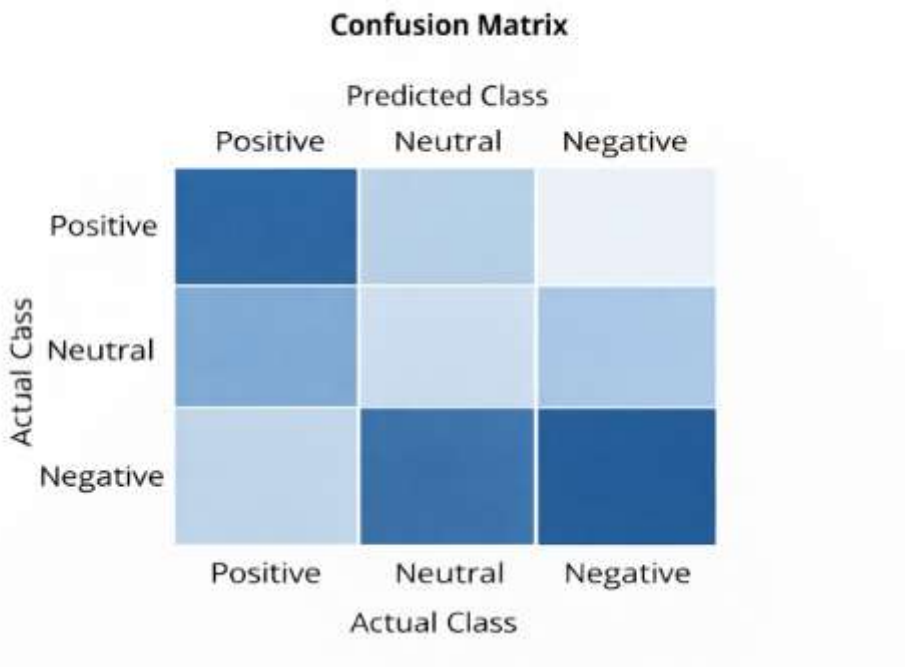

FIG 3.CONFUSION MATRIX

The image demonstrates a distinctively organized confusion matrix that is intended to assess a three-class sentiment classification model. It is in the form of a 3x3 grid with the predicted classes appearing at the top and the actual classes appearing at the left side. These are the Positive, Negative and the Neutral categories. Correct classifications have been emphasized by shading the diagonal cells in darker tones of blue, whereas misclassification is highlighted by the off-diagonal ones being highlighted by light tones. The layout is clear and well labeled and one can easily interpret the performance pattern of the model by just looking at the layout without the numerical values being displayed within the cells.

5. Conclusion

The confusion matrix displays the effectiveness on the sentiment classification model as a whole to differentiate the Positive, Neutral, and Negative classes. The heavier focus on the diagonal suggests that the model is usually apt at making right predictions with the lighter off-diagonal regions indicating that misclassifications can take place. All in all, the visualization implies that this model has a fair level of functionality, yet it could use more improvement in terms of minimizing the error rate in classification and enhancing the accuracy of the approach that is applicable to all the classes.

The future work may involve generalizing data quality and expansion of the diversification of data sets to allow the model to be more receptive to subtle variations among sentiments. Accuracy could also be enhanced by using advanced techniques that allow tuning the hyperparameters or testing alternative machine learning or deep learning architectures, as well as using contextual embeddings. Moreover, the solution of the issue of class imbalance, the error analysis of misclassified samples, and the use of ensemble could result in stronger and more dependable performance in the frames of the real-life practice.

Author Contributions

S.Sadaq Hussain : Project Lead and System Architecture
P.Karthik Reddy : Data Collection and Management.
S.Hamid : NLP & Text Preprocessing and Machine Learning Model Development
S. Adil: Front End Interface by Using Streamlit / Flask.
V. Guru Siva Sankar Reddy : Testing, Visualization & Documentation
All authors reviewed and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

G. Li, Q. Zheng, L. Zhang, S. Guo and L. Niu, "Sentiment Infomation based Model For Chinese text Sentiment Analysis," *2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, Shenyang, China, 2020, pp. 366-371, doi: 10.1109/AUTEEE50969.2020.9315668.

G. Rokade, R. Ughade and P. Gaurshettiwar, "Deep Learning for Sentiment Analysis," *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, Bhimdatta, Nepal, 2025, pp. 240-244, doi: 10.1109/ICSADL65848.2025.10933099.

A. Ilmania, Abdurrahman, S. Cahyawijaya and A. Purwarianti, "Aspect Detection and Sentiment Classification Using Deep Neural Network for Indonesian Aspect-Based Sentiment Analysis," *2018 International Conference on Asian Language Processing (IALP)*, Bandung, Indonesia, 2018, pp. 62-67, doi: 10.1109/IALP.2018.8629181.

K. Fujihira and N. Horibe, "Multilingual Sentiment Analysis for Web Text Based on Word to Word Translation," *2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI)*, Kitakyushu, Japan, 2020, pp. 74-79, doi: 10.1109/IIAI-AAI50415.2020.00025.

P. Porntrakoon, "Improve the Accuracy of SenseComp in Thai Consumer's Review Using Syntactic Analysis," *2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Pattaya, Thailand, 2019, pp. 369-372, doi: 10.1109/ECTI-CON47248.2019.8955197.

R. Obiedat, D. Al-Darras, E. Alzaghoul and O. Harfoushi, "Arabic Aspect-Based Sentiment Analysis: A Systematic Literature Review," in *IEEE Access*, vol. 9, pp. 152628-152645, 2021, doi: 10.1109/ACCESS.2021.3127140.

P. Li, W. Chang, S. Zhou, Y. Xiao, C. Wei and R. Zhao, "A conflict opinion recognition method based on graph neural network in Aspect-based Sentiment Analysis," *2022 5th International Conference on Data Science and Information Technology (DSIT)*, Shanghai, China, 2022, pp. 1-6, doi: 10.1109/DSIT55514.2022.9943870.

M. Siek and E. S. Setiadi, "Analysis of News Sentiment and Stock Price Using Web Scraping and Vader Sentiment Analysis," *2024 International Conference on Information Management and Technology (ICIMTech)*, Bali, Indonesia, 2024, pp. 1-6, doi: 10.1109/ICIMTech63123.2024.10780837.

G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," in *IEEE Access*, vol. 7, pp. 43749-43762, 2019, doi: 10.1109/ACCESS.2019.2907772.

H. Karamollaoğlu, İ. A. Doğru, M. Dörterler, A. Utku and O. Yıldız, "Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach," *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sarajevo, Bosnia and Herzegovina, 2018, pp. 45-49, doi: 10.1109/UBMK.2018.8566481.