International Journal of Futuristic Innovation in Engineering, Science and Technology

Vol.**03**, Issue 3, pp. 54-60, Dec-**2024**
ISSN:2583-6234(Online)
Available online at: IJFIEST

Research Paper

A Comprehensive Survey on Sentiment-Based Approaches in Stock Market Analysis

Mr. Digvijay Singh [*1], Dr. Abhishek Guru [*2], Dr. Asha Ambhaikar [*2]

[1] Mats University, Raipur, Chhattisgarh, India
[2] Associate Professor, Mats University, Raipur, Chhattisgarh, India
[3] Director at MSEIT, Mats University, Raipur, Chhattisgarh, India

[1] digvijayvp2414@gmail.com, [2] abhishekguru0703@gmail.com, [3] drasha@matsuniversity.ac.in

**Corresponding Author: digvijayvp2414@gmail.com*

***Received**: 15/Sep/ 2024*
***Revised:** 22/Oct/2024*
***Accepted**: 28/Nov/2024*
***Published**:27/Dec/2024.*

In recent years, sentiment analysis has emerged as a critical tool for enhancing stock market prediction by capturing investor mood and market psychology from textual data sources such as news articles, financial reports, and social media platforms. This paper presents a comprehensive survey of sentiment-based approaches in stock market analysis, highlighting the evolution, methodologies, and applications of sentiment-driven models. The study examines various techniques including traditional machine learning, deep learning, and natural language processing (NLP) methods employed to extract and quantify sentiment. It also explores the integration of sentiment signals into predictive financial models and their impact on forecasting stock price movements, volatility, and market trends. A comparative analysis of datasets, sentiment lexicons, model performance, and evaluation metrics is provided to identify existing challenges and research gaps. Furthermore, the paper discusses recent advancements such as the use of transformer-based architectures and multimodal sentiment fusion. This survey aims to serve as a foundational resource for researchers and practitioners seeking to leverage sentiment analysis in financial markets, and proposes future directions for developing more robust, explainable, and real-time sentiment-aware forecasting systems.

Keywords: Sentiment Analysis, Stock Market Prediction, Natural Language Processing, Machine Learning, Financial Forecasting.

1. INTRODUCTION:

As more and more individuals have access to the internet, they are turning to social media to disseminate information, read reviews, etc. Numerous applications rely on this data, opinion mining being one of them. The goal of sentiment analysis is to determine whether the information is good or negative. Recognising the positive or negative tone of a text is an essential part of sentiment analysis. In order to classify emotions, advanced polarity classification considers states such as joyful, sad, furious, etc. [1]: Yes. Classical approaches of surveying public opinion were time-consuming and inaccurate. Therefore, opinions expressed on social media are far more genuine than data gathered from interviewers, who are usually reluctant and lack any genuine interest in the subject matter [2]. It is also much simpler to analyse data from an automated system than from a human survey. Consumers' purchasing habits may be impacted by the diverse viewpoints expressed on social media [3]. To improve their goods, companies may also use it [4]. Sentiment analysis, depending on the application and dataset, is often carried out using a variety of approaches and algorithms. [5] Forecasts stock values using social media platforms, particularly Twitter. Here, the simplified programming model was used for computation, and a machine learning method was employed to categorise data and predict future stock values. In contrast,[6] analyses the tone of story comments using a lexicon-based method. Here, we use a Lexicon-based method to determine polarity, and then we feed that data into two machine learning algorithms, SVM and K-nearest neighbour. Additionally, sentiment analysis is known as opinion mining as it is a subfield of that larger field. In sentiment analysis, as described by Pang and Lee [7], a number of terminologies are used.

Due to the incorporation of tongue processing, the phrase sentiment analysis might also be employed on occasion. A phrase or sentiment's polarity may indicate its positive, negative, or neutral tone. Determining whether a piece of writing is objective or subjective is an example of subjectivity. While sentimental statements tend to be subjective, factual sentences tend to be objective. "Suppose he did lie beside Lenin, wouldn't it be permanent?" is an example of a remark that might be seen as subjective. Explicitly expressed sentiments, as "the movie was fantastic," are often intrinsic to the statement.

In order to make educated judgements about whether or not to participate in a firm, sentiment analysis of the stock market is useful. Examining the trading activity of a business or organisation is known as stock analysis. The results of the study demonstrate that the mood on the internet may be used to forecast the market's next move. When people feel good about a company, its stock value rises, but when an unfavourable comment damages it. The stock price is very sensitive to fresh information. A variety of sources, such as publications, news, and social media opinions, provide the information..

2. Challenges and Applications

Sentiment analysis or opinion mining can have various applications like movie reviewing, stock exchange prediction, product feature reviews etc. the varied challenges [8] are:

- Finding the right dictionary: it's difficult to seek out the foremost accurate dictionary that contains all required words, and typically we will overcome this problem by creating our own dictionary supported the need .

- Detection of sarcasm in statements: it's difficult to detect sarcasm and supply it an appropriate polarity value.

- Detection of faux reviews: internet contains spam content also . Effective Sentiment classification requires this spam content to be eliminated before processing. this will be done by identifying duplicates, by detecting outliers. [10]
- Use of orthographic words: orthographic words like too, veeery, etc. are difficult to polarize.
- Use of abbreviations: short forms like u for; you', 4 for 'for' are difficult to supply polarity for.

3. Methodology

Sentiment analysis are often administered at three levels, namely, document level, sentiment level, and aspect level [11]. Sentiment analysis includes three main steps, identification, classification, and aggregation [12].

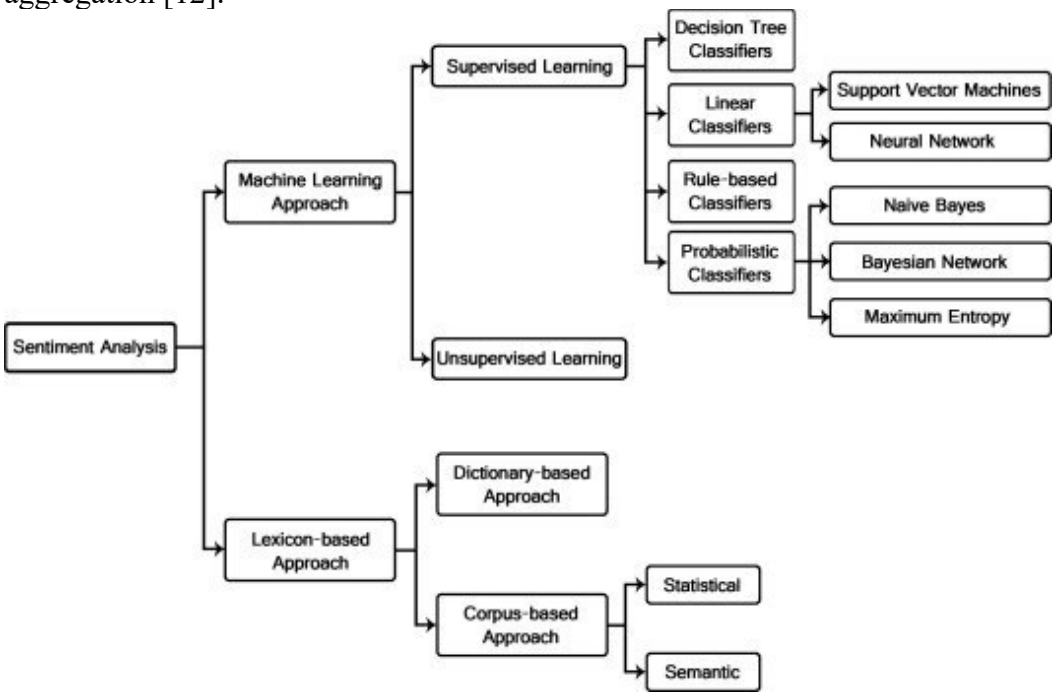

Figure 3.1. Sentiment classification techniques

Dictionary Based

It is common practice to use a sentiment dictionary to determine the polarity of a word in a text. In this approach, a predetermined dictionary is first constructed by hand using a collection of seed words [9]. Using these established dictionaries, text may be tagged as positive, negative, or neutral in terms of polarity. Word count and frequency of recurrence are used by this strategy. And other approaches to add polarity to existing data [13]. Here is an explanation of how this method is often used for sentiment analysis: seed words are built by hand using polarity values that have already been determined. The next step is to use an algorithm that looks for additional terms that are similar by searching dictionaries like wordnet. After that, you may add them to the list, and you can keep doing this until you run out of words.

Supervised Learning

New data is given polarity using supervised learning on top of a training dataset. Input files and output variables make up the training data. Among the many techniques used in supervised machine learning, support vector machines (SVMs) stand out. Using the idea of a decision plane, the support vector machine's algorithm may identify patterns in the provided data and classify individuals into groups with shared characteristics [15]. Accuracy levels of up to 80% are achievable using SVM on the correct dataset [16].

4. Results and Discussions

Paper	Summary	Performance
Sentiment Analysis on News Articles for Stocks [17]	Sentiment analysis based on data culled from news items is the focus of this work. After retrieving the news links from the Bing API, the articles were analysed using a sentiment dictionary. We tested two separate machine learning algorithms and compared their anticipated outcomes to the real market fluctuations in stock prices. Future enhancement opportunities were found by mapping the findings to current market trends.	Accuracy for normal equation- 53.2%, and for gradient decent-59.5%.
Stock Trend Prediction Relying on Text Mining and Sentiment Analysis with Tweets [18]	The feature sparse issue that arises from sentiment analysis utilising tweets is addressed in this study. A text-sentiment-based stock trend prediction technique was used to circumvent this. This model employs a support vector machine classifier. The model's accuracy was enhanced by including the hybrid feature selection approach. Here, a SentiWordNet is used to give certain traits more importance. After comparing several learning methods, this article finds that support vector machine (SVM) provides the best results.	An accuracy of 90.34%.
Twitter mood predicts the stock market [19]	If there is a correlation between people's emotions and the value of the Dow Jones Industrial Average (DJIA), then this study will show it. This is accomplished by utilising two monitoring tools: opinion finder, which contrasts positive and negative mood, and Google-Profile of Mood States (GPOMS), which assesses mood along six dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) in order to gauge reaction to the presidential election and Thanksgiving. Additionally, this study provides evidence supporting the concept that the closing values of the DJIA may be predicted by people's emotions as assessed by opinion finder and GPOMS.	An accuracy of 87.6% was obtained and Mean Average Error was reduced by more than 6%.

Collective Sentiment Mining of Micro blogs in 24-hour Stock Price Movement Prediction [20]	The purpose of this study is to assess and forecast the change in stock prices for the following day using collective sentiment analysis. A two-stage technique using a natural language processing (NLP) approach and a statistical analysis approach is included. The SVM method forms its foundation. The expertise in NLP and machine learning is put to use. The Granger Causality test is used to evaluate the prediction's outcome.	An accuracy of 71.84% for positive sentiments and 74.3% for negative
Stock Price Prediction using Linear Regression based on Sentiment Analysis [21]	This paper stresses on the fact that according to efficient market hypothesis (EMH) stock prices depend on a number of factors, one of them being peoples' opinion or sentiment. This paper surveys the Indonesian stock market using sentiment analysis. Two algorithms are used for analysing the tweets, namely, Naïve Bayes and Random Forest algorithms. As future work, the authors suggest increasing the data collection period, using non-linear prediction models and including features such as POS tagging, word weighting, etc. in classification.	The Random model algorithm yields an accuracy of 60.39% and the Naïve Bayes algorithm gives 56.50% accuracy.
Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis [22]	This paper tests the assumption of increasing accuracy of stock market prediction by analysing the psychological moods of twitter users. Eight different emotions can be analyzed lexicon-based approach to classify peoples'psychological states. Support vector machine algorithms are used to predict the DJIA indicators. Focusing on the increase of training period and efficiency of algorithms was saved as future work.	The accuracy rate of 64.10% was achieved using Support Vector Machine algorithm to predict DJIA indicator.
A Hybrid Approach to Sentiment Analysis of News Comments [23]	This research uses two main algorithms, that is, Support Vector Machine (SVM) and K-Nearest Neighbour (KNN), to perform sentiment analysis of news comments. For this, a hybrid approach is adopted, wherein, a sentiment lexicon is used to classify the comments and calculate polarity. The results of this are then used to train the machine learning algorithms.	The accuracy of 60.96 for SVM
Prediction Relying on Text mining and Sentiment Analysis with Tweets [24]	This paper focuses mainly on the twitter feature sparse problem and also the unreliability of using average sentiment score. This is overcome using a hybrid approach consisting of entiwordnet to give additional weightage. SVM cross validation is used.	Accuracy without Feature selection (FS) was 53.62%, whereas with FS, an accuracy of up to 84.06% was obtained.

Stock Trend Forecasting Method Based on Sentiment Analysis and System Similarity Model [25]	This paper proposes a system wherein the Bayesian classifier is used based on the system similar model to predict stock movement. The system is tested using inter cross and turnover test.	An accuracy of up to 78.5% is obtained
Stock Market Forecast Using Sentiment Analysis [26]	In this paper they have used twitter and BSE to calculate public moods and used granger causality to predict the results, they also implemented SOFNN (Self organised fuzzy neural network) for decoding non-linear time series.	An accuracy of 75.56% was obtained

References

[1] Manisha Shinde-Pawar,"Formation of Smart Sentiment Analysis Technique for Big Data",International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 12, December 2014

[2] Kim Schouten and Flavius Frasincar, "Survey on aspect-level sentiment analysis", IEEE Transactions on Knowledge and Data Engineering, Volume: 28, Issue: 3, March 1 2016

[3] E.Van Kleef, H. C. M. Van Trijp, and P. Luning, "Consumer research in the early stages of new product development: a critical review of methods and techniques," food quality preference, Vol. 16, no. 3, pp. 181–201, 2005.

[4] Y. Chen and J. Xie, "Online Consumer review: word-of-mouth as a new element of Marketing Communication mix," manage. sci., Vol. 54, no. 3, pp. 477–491, 2008.

[5] Michał Skuza, Andrzej Romanowski, "Sentiment Analysis of Twitter Data within Big Data Distributed Environment for Stock Prediction", Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1349–1354 ACSIS, Vol. 5.

[6] B. Pang and L. Lee, "Opinion mining and sentiment analysis," found. trends inf. retrieval, vol. 2, no. 1-2, pp. 1–135, 2008.

[7] Asmita Dhokrat, Sunil Khillare, C. Namrata Mahender, "Review on techniques and tools used for opinion mining", international journal of computer applications technology and research volume 4– issue 6, 419 - 424, 2015

[8] Laura Cruz, Jose Ochoa, Mathieu Roche, Pascal Poncelet, "Dictionary-based Sentiment Analysis applied to specific domain using a Web Mining Approach"

[10] Indurkhya, N. & Damerau, F.J. [eds.] Liu, B., 2010 "Sentiment Analysis and subjectivity," appeared in handbook of natural language processing. Bing l., 2012 "Sentiment analysis and opinion mining," Morgan & Claypool publishers.

[11] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," data mining know. discovery, vol. 24, no. 3, pp. 478–514, 2012.

[12] Douglas R, Rice Christopher, "Corpus-based dictionaries for Sentiment Analysis of specialized
Vocabularies"version 0.1 September 19, 2013

[13] Reshma Bhonde, Binita Bhagwat , Sayali Ingulkar, Apeksha Pande, "Sentiment Analysis based on dictionary approach", International Journal of Emerging Engineering Research and

Technology Volume 3, issue 1, January 2015

[15] Pravesh Kumar Singh, Mohd Shahid Husain," Methodological Study of Opinion Mining and Sentiment Analysis Techniques", International Journal on Soft Computing (ijsc) vol. 5, no. 1, February 2014

[16] Phayung Meesad and Jiajia Li, King Mongkut's, "Stock Trend Prediction Relying on Text Mining and Sentiment Analysis with Tweets"

[17] VaanchithaKalyanaraman, Sarah Kazi, Rohan Tondulkar,"Sentiment Analysis on News Articles for Stocks".

[18] Phayung Meesad and Jiajia Li," Stock Trend Prediction Relying on Text Mining and Sentiment Analysis withTweets"

[19] Johan Bollen,Huina Mao,Xiao-Jun Zeng," Twitter mood predicts the stock market"

[20] Feifei Xu and Vlado Keˇselj,"Collective Sentiment Mining of Microblogs in 24-hour Stock Price
MovementPrediction", 2014 IEEE 16th Conference on Business Informatics

[21] Yahya Eru Cakra,Bayu Distiawan Trisedya," Stock price prediction using linear regression based on sentiment analysis", Advanced Computer Science and Information Systems (ICACSIS), 2015 International Conference.

[22] Alexander Porshnev, Ilya Redkin, Alexey Shevchenko,"Machine learning in prediction of stock market indicators based on historical dataand data from Twitter sentiment analysis", 2013 IEEE 13th International Conference on Data Mining Workshops

[23] Addlight Mukwazvure,K.P Supreethi,"A Hybrid Approach to Sentiment Analysis of NewsComments"

[24] Phayung Meesad, Jiajia Li, "Stock trend prediction relying on text mining and sentiment analysis with tweets",2014 4th World Congress on Information and Communication Technologies (WICT 2014)

[25] Kaihui Zhang, Lei Li, Peng Li,"Stock trend forecasting method based on sentiment analysis and system similarity model",Strategic Technology (IFOST), 2011 6th International Forum.

1. Rajat Ahuja, Harshil Rastogi,Arpita Choudhuri," Stock market forecast using sentiment analysis", Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference