

Stock Prediction Using Sentiment Analysis: A Comprehensive Review of Techniques And Challenges

Heli Patel¹, Dr. Tejas V. shah², Dr. Deepali H. shah³

¹Dept of IC

^{2,3}Associate Professor, Dept of IC

^{1,2}L.D.college of Engineering

³Govt. Engineering college, sector 28, Gandhinagar

Abstract- Sentiment analysis, also known as opinion mining, is a process of using natural language processing and machine learning techniques to identify and extract subjective information from text data. Stock prediction using sentiment analysis involves analyzing public sentiment towards a company or its stock to predict its future performance in the stock market. Sentiment analysis involves using natural language processing and machine learning techniques to analyze large amounts of text data from sources such as news articles, social media posts, and financial reports to determine the sentiment towards a particular company or stock.

Keywords- Sentiment Analysis, opinion, emotions, Machine Learning, Accuracy

I. INTRODUCTION

As social creatures, humans are naturally wired to be influenced by the beliefs, attitudes, and opinions of others around them. This can impact our decision-making process as well as our overall behavior. Our abstract sentiments and beliefs play a significant role in shaping our perception of the world around us. They are often influenced by a range of factors such as our upbringing, cultural background, social interactions, education, and personal experiences. These sentiments and beliefs can influence our emotions, thoughts, and behavior in different ways

Nowadays, social communication channels like Twitter, Facebook, and YouTube have obtained so much popularity. Opinion mining is the other name of Sentiment Analysis [2]. Sentiment analysis involves analyzing large volumes of text data, such as social media posts, product reviews, and customer feedback, to determine whether the overall sentiment is positive, negative, or neutral. This can be done using various methods such as rule-based systems, machine learning algorithms, and deep learning models. The amount of information being recorded in digital forms is increasing at an unprecedented rate. This is due to the

increasing use of digital devices, the internet, and social media platforms. Every day, we generate massive amounts of data through our online activities such as browsing, social media interactions, and online shopping. This data is recorded and stored in databases and servers, and it can be analyzed to gain valuable insights and information.

Furthermore, social media platforms such as forums, blogs, customer reviews, Twitter, and social network sites provide an enormous amount of information and data on individuals and their behaviors. This data can be used to analyze trends, track sentiment, and understand consumer preferences and behaviors.

Applications of sentiment analysis include brand reputation management, customer service, market research, and political analysis. By analyzing the sentiment of social media posts and customer reviews, companies can gain insights into how their products or services are perceived by the public and make necessary changes to improve their offerings. In politics, sentiment analysis can be used to analyze public opinion towards candidates and issues, and to predict election outcomes. [3] (e.g., a movie review, a person, a political party, or a policy or product feature review .

The staying of this paper is organized as the followings:

The next segment portray explain about the introduction of the sentiment mining and opinion mining. From that point onwards, different stages of opinion Analysis are displayed. The portrayal of the previous work that is in Section IV was done on sentiment analysis techniques is available. Introduction about the sentiment analysis sources/resources are in Section V. Part VI explained regarding challenges of sentiment analysis. Finally, the Discussion and conclusion is stated in section VII.

➤ *Sentiment:*

Sentiments can be expressed in different ways. In the context of natural language processing and machine learning, sentiment refers to the overall emotional tone or attitude expressed in a piece of text. It is often analyzed by computer programs using various techniques to determine whether a given text expresses a positive, negative, or neutral sentiment.

Sentiment Analysis:

Sentiment analysis or opinion mining is the process to identifying and recognize or categorized the users emotion or opinion for any services like movies, product issues ,events or any attribute is positive ,negative or neutral. The sources for this analysis is social communication channels i.e Web site which include reviews, forum discussions, blogs, microblogs, Twitter etc. This research field is very popular nowadays because of its opinioned data where user can find reviews for any services which is useful for their daily lives. Sentiment analysis can be applied to a wide range of texts, such as product reviews, social media posts, news articles, and customer feedback. It can be used to gather insights about public opinion, identify trends and patterns, and support decision-making processes in various fields, including marketing, finance, and politics[5].

In sentiment analysis, classifying sentences as either subjective or objective is an important step in identifying the overall sentiment of a text[9]. Subjective sentences express opinions, emotions, attitudes, and personal feelings, whereas objective sentences provide factual information without expressing personal opinions or feelings. By identifying subjective sentences, sentiment analysis models can then focus on extracting the sentiment or emotional tone of those sentences to determine the overall sentiment of the text[10]. This can be particularly useful in various fields like politics, business, marketing, and advertising, as you mentioned, where understanding the sentiment of customers, voters, or audiences can be critical in making informed decisions.

Recent or existing research is using both supervised and unsupervised learning technique to provide diferent techniques for several purpose of sentiment analysis.

The main purpose is to divide the entire documents or sentences into one of the two classes; objective or subjective [11]. the choice of algorithm depends on the problem domain, the size and nature of the dataset, and the performance requirements. Sometimes, a combination of multiple algorithms is used to improve the performance or handle different aspects of the problem. In Opinion mining, the sentences in the text are labeled as either

Classification Level

There are different type of classification levels:

- 1) Document-level
- 2) the Sentence level
- 3) Attribute/Aspect Level

1) *Document Level:* Document level sentiment analysis refers to the process of determining the overall sentiment or emotional tone of a piece of text, such as a document or an article. This is typically done using natural language processing (NLP) techniques to analyze the text and identify specific words, phrases, or patterns that indicate positive, negative, or neutral sentiment[4]. The goal of document level sentiment analysis is to provide an understanding of the writer's overall opinion or attitude towards a particular topic or subject. This can be valuable in a variety of contexts, such as market research, social media monitoring, or customer feedback analysis[5].

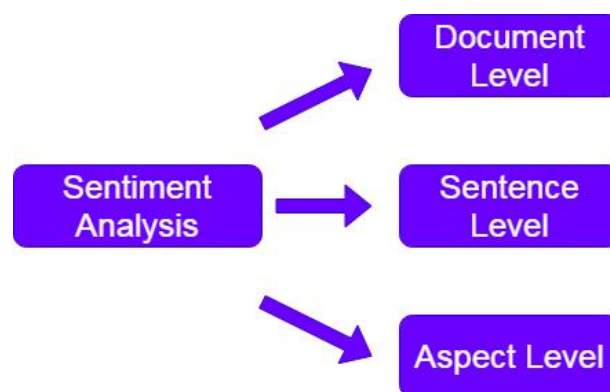


Fig.1 Categories of Sentiment Analysis

2) *Sentence Level:* Sentence-level sentiment analysis is a more granular approach to sentiment analysis, where each sentence in a document is analyzed to determine its polarity (positive, negative, or neutral). This approach can provide more detailed insights into the sentiment expressed within a document, particularly when there is a mixture of positive and negative sentiments. Subjective sentences are those that express opinions, views, or perspectives, while neutral sentences do not imply any sentiment or opinion[8]. By analyzing the sentiment of each sentence in a document, we can gain a better understanding of the overall sentiment expressed within the text.

3) *Aspect/Feature Level:* This type of sentiment analysis is focused on identifying the sentiment towards specific aspects or entities mentioned in a text. It involves identifying and extracting the relevant entities and their associated sentiments.

II. APPLICATIONS

➤ *Decision making support:*

Building a website that could perform decision making is a very crucial part. Analysis has its own advantage like, it can lead to different ideas which can help us to make decision in day to day life such as choosing a good restaurant to go for dinner, or buying a new car or selecting a good movie to watch etc.

➤ *Business related application:*

Because of everyday changing market, the competition has increased a lot in co-operative world. Every wants to create an innovative and newest product which can fully satisfy their customers. To achieve more valuation of their product, organization can assemble all the needs of their users and enhance the efficiency of product from feedback collected from their customers.

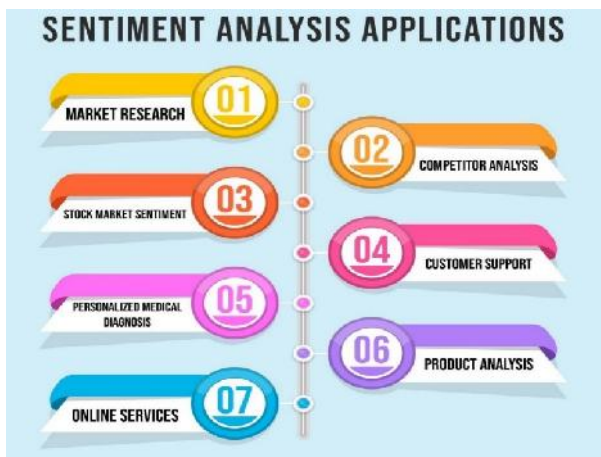


Fig .2 sentiment analysis application

➤ *Predictions and trend analysis:*

Tracking views of public by sentiment scrutiny which enable any person to predict the market scenario which help any person for trading and polls market. By using this all opinions user can predict the market trends.

III. METHOD AND APPROACHES

“There are various ways and techniques available for opinion mining, there are majorly two groups used. 1) Uses lexicon methods and 2) machine learning method which resolves the problems of SA.

1) *Lexicon based approach:* In this current approach, when using the available lexicon techniques for a text which is

given, will separate the words. In general it is performed by aggregation of scores: for example subjective words scores as positive, negative and neutral etc are summed up separately for same. It assigns a score to each word. At least four scores are generated. The one which gets the maximum score gives the overall split of the text [10]. It is mainly divided into two parts: a) Dictionary-based b) corpus-based

a) *Dictionary-based approach:* In this system, the user collected a set of sentiment words and seed list is prepared by them. After that, the user starts searching for phrasebooks and lexicon to find synonyms and antonyms of particular text. Once this is done, the newly created substitutes are added into the seed list. Until there are no new words are found to users this process continues..

Disadvantage: There has to have struggle in finding context or domain-oriented emotion words.

b) *Corpus-based approach:* Corpus is basically a term which is a cluster of writing like group of some writing which is often on a very precise matter. In this, users use the help of corpus text to draw out the seed list which is in organized situation [9].

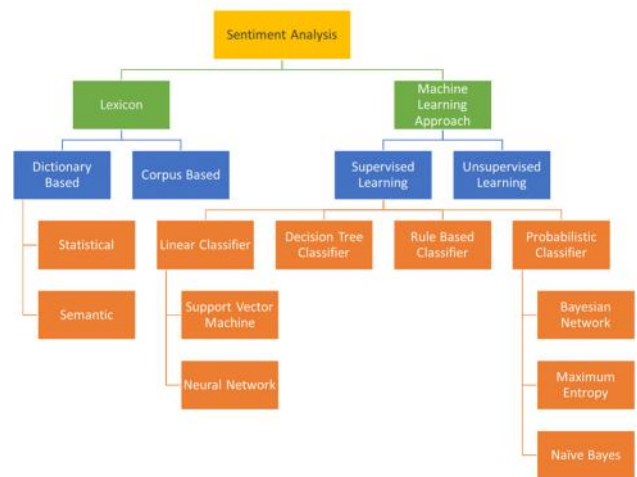


Fig. 3 Various Approaches of Sentiment Analysis

2) *Learning approach:* In this approach, initially classification is performed by taking two different assemblies of the document. Trained data and test data are part of these. This is termed as involuntary classification. Further text is extracted from the features and categorized into I) supervised and II) unsupervised.

a) *Supervised System:* Among various kind of datasets, Labeled training dataset is one of them which is used in supervised

system. Each type of class has its own property and advantages and has its label related to it which can be used for this system. Each word, upon arriving is categorized under a label depending on its type and characteristics related with it.

1. Probabilistic classifier: Predicts or anticipates probability function related to input records among different modules.

Naïve Bayes: In this, to generate possibilities of a group to provide prediction that group of properties belongs to one particular label with help of Bayes theorem using merely a text document as an input. BOW- Bag of Words is a way to extract a text with using machine learning methodologies which is simple and easy to implement. This existing model conducts that these all the features which are given are autonomous. [20]

$$P(\text{label}/\text{features}) = P(\text{label}) * P(\text{features}/\text{label}) / P(\text{features})$$

Bayesian network: It is used to manifest relationships among different features. It can be compared with acyclic graph in which nodes represent random variable and edges represent dependencies. This model is very pricey and hence it's hardly used.

Maximum Entropy: By doing encoding, the labeled feature sets are converted into vectors by using classifiers. This vector which is converted is utilized to decide the weights of these features which can be able to use to suppose and predict the label for each of their feature set. ure set.

2. Linear classifier: The characteristics of the linear classification can be implemented by using this classifier which is used to show predictor as result and can be divided into two classes:

A) *Support Vector Machine (SVM)*: This learning model is under supervision to be utilized for classification. The most important purpose of this particular model is to assure that this is the best linear separator for classification. This will make a model that results in new information into one or two classes using SVM training.

B) *Neural Network (NN)*: It is a neural structure of the brain having electronic networks of neurons. In this network, Neuron is the basic component. Neurons are categorized into three parts- input, hidden and output.

3. Decision tree classifier: To make division of the data, there is a condition which is used. One class consists of those data which mollify the condition and other class consists of the remaining of the data. This technique is called a recursive

technique which has two parts: single attribute split and multi attribute split.

4. Rule based classifier- It is a condition based classifier which makes use of condition or rule like IF, THEN. It can be written as IF condition THEN decision"

We can produce the rules based on our requirements at the time of training phase [2].

IV. RESOURCES OF SENTIMENT ANALYSIS

"To collect data is the main purpose of Sentiment analysis where social communication channels like Twitter, Facebook or any pre-existing resources.

A) *Blogs & Forums*: It is a source of opinions and emotions where we get information for research purpose and that all information can be used by researchers via Web forums and blogs. "Generally, for only single subject forums are designed; thus, by using the forums we can ensure the sentiment mining in single domain. As well, it's the trend that bloggers updated their blogs and reviews everyday after activities in and around their areas, countries and around the world.

B) *Reviews*: There are many available studies which dedicated only on reviews because of their usability with the opinions and sentiment. During any research, Movie and product reviews were mostly studied by researchers where the main purpose is to get the feedback from the sentiment and opinions.

C) *News Articles*: News articles, such as financial articles and political reviews are a popular source of sentiment analysis [51]". The main format of News articles texts are structured and formal.

D) *Social Networks*: Many social networks sites are available from which we can take the opinions and reviews for sentiment analysis like Twitter, Facebook, etc.

- Twitter:

Tweets are the messages posted by different users, having a restriction of 140 characters. Users can read messages (called Tweets) of one another. The micro-blogging service which provides this facility is known as Twitter. By using this tweets which can work as opinions and reviews for future patterns where we can generate the poll results.

- Facebook:

The provision of posting personal profile, photos, videos and other related information are provided by most famous social networking facility called Facebook which got famous right after it got launched in 2004.

Hence, these much ample amount of information available in form of users message, computer technology which is dependant on sentiment behind these message is introduced known as sentiment analysis.

• Related work

A lot of studies have been done by scholars to analyze emotions or opinions. There are many methods used to extract the data.

Comparison of various Approaches and Methods

Many algorithms of data mining have been proposed to predict stock price. The opinion summary is explored based on the opinion sentences. This paper summarizes only factual information, and is not useful for the opinion-based summary. In order to avoid the complex space modelling of various sources a tensor based information framework was introduced to predict stock [1,2].

Daily text content in Twitter is analyzed by mood tracking tools and investigated for predicting the changes of DJIA closing values, which was based on a self-organizing fuzzy neural network [3]. This approach is similar to our approach of mood tracking.

The frequency of blog posts can be used to predict spikes in the actual consumer purchase quantity at online retailers [4]. Moreover, it was shown by Tong [6] that references to movies in newsgroups were correlated with their sales. Sentiment analysis of weblog data was used to predict movies' financial success.

The literature review reveals two main text mining approaches are adopted in the analysis, prediction or mining of stock market features: (i) machine learning such as Support Vector Machine, decision rules/trees, regression algorithms, naïve Bayes, and (ii) natural language processing algorithms [5].

In our preliminary work [6] we used the volume and sentiment polarity of Apple financial tweets to identify important events, as a step towards the prediction of future movements of Apple stock prices. They explain that the application of data mining to the analysis of stock market data using current approaches may not be sufficient to model and

justify random behavior of the market based only on quantitative data such as the values of stocks and historical market prices.

Machine learning algorithms give computers the ability to learn without being explicitly programmed by involving a set of data to train the algorithm and using another set of data to test the generated predictions. The natural language processing approach involves lexical, syntactic, semantic and pragmatic analysis of unstructured texts [8,9].

Table 1.0 COMPARATIVE STUDY OF TECHNIQUES OF SENTIMENT ANALYSIS

Sr. No	Year	Paper Title	author	Methodology Used/Findings
1	(2021)	Stock price prediction based on multiple data sources and sentiment analysis[1]	Shengtingwu,Yuiling, Lu,Ziranzou,Tian-Hsiung Weng	<ul style="list-style-type: none"> S_I_LSTM The stock price prediction method that incorporates multiple data sources and the investor sentiment, which can be called S_I_LSTM.
2	(2021)	Harvesting social media sentiment analysis to enhance stock market prediction using deep learning[2]	Pooja Shah,SharniPandya,Ketan Kotecha	<ul style="list-style-type: none"> ML Techniques ,Customized LSTM The algorithm considers public sentiment, opinions, news and historical stock prices to forecast future stock price.
3	(2022)	HISA-SMFM: HISTORICAL AND SENTIMENT ANALYSIS BASED STOCK MARKET FORECASTING MODEL[3]	Ishu Gupta,Sukhman Singh,Tarun Kumar Madan,	<ul style="list-style-type: none"> SVM, Random Forest,LSTM The use of historical as well as sentiment data to efficiently predict stock prices by applying LSTM.
4	(2019)	A Deep Learning Model for Predicting Buy and Sell Recommendations in Stock Exchange of Thailand using Long Short-Term Memory[4]	Ying-Lai Lin, Chi-Ju Lai and Ping-Fang Pai	<ul style="list-style-type: none"> SVM, Naive Bayes, Deep Learning This research adopts deep learning techniques for predicting buy and sell recommendations in Stock Exchange of Thailand using Long Short-Term Memory. The proposed model can capture long-term dependencies in stock price data in order to enhance the prediction accuracy.
5	(2022)	Using Deep Learning Techniques in Forecasting Stock Markets by Hybrid Data with Multilingual Sentiment Analysis[5]	Yang Li, Yi Pan	<ul style="list-style-type: none"> Naive Bayes LSTM LSTMGA model with hybrid multilingual sentiment analysis is a feasible and promising way of forecasting the stock market.
6	(2019)	Deep learning based sentiment analysis and text summarization in social networks[6]	Dogan, E. & Kaya, B	<ul style="list-style-type: none"> Deep learning Based techniques Analysis and summary of the text were made with data from a hashtag on Twitter. The methods used in the analysis of emotions were compared to the methods of word embedding and a success rate of 93% was obtained.
7	(2020)	Sentiment Analysis for Stock Price Prediction[7]	James Briggs	<ul style="list-style-type: none"> Naive Bayes, SVM This research is focus to give nearly full image of SA techniques and the related fields with brief details. The main research finding of this paper include the sophisticated categorizations of a large number of recent articles and the illustration of the recent trend of research in the sentiment analysis and its related areas.
8	(2021)	Explainable stock prices prediction from financial news articles using sentiment analysis[8]	Shikha Ghite,H Khataavkar,K Kotecha,N Pandey	<ul style="list-style-type: none"> LSTM , Explainable AI (XAI) Hence, finding is one more dimension of sentiments along with technical analysis should improve the prediction accuracy; the LSTM model uses historical stock data along with sentiments from news items to create a better predictive model.
9	(2018)	Stock Movement Prediction from Tweets and Historical Price [9]	Yunmo Xu and Shay B. Cohen	<ul style="list-style-type: none"> ARIMA, Random Forest The research analysis is hybrid objective with temporal auxiliary to flexibly capture predictive dependencies. We demonstrate the state-of-the-art performance of our proposed model on a new stock movement prediction dataset which we collected.
10	(2020)	Stock Trend Prediction Using News Sentiment Analysis[10]	Kalyani Joshi, Prof. Bharathi H. N., Prof. Jyothi Rao	<ul style="list-style-type: none"> SVM, Random Forest and Naive Bayes classification Observations and analysis show that RF and SVM perform well in all types of testing. Naive Bayes gives good result but not compared to the other two. Experiments are conducted to evaluate various aspects of the proposed model and encouraging results are obtained in all of the experiments.
11	(2018)	Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers[11]	GhaithAbdulshatarAJ abbar,Alkubaisyil, Siti Sakira KamaruddinI &Husniza HusniI	<ul style="list-style-type: none"> Hybrid Naive Bayes Classifiers[11] This research suggested Hybrid Naive Bayes Classifiers (HNBCs) as a machine learning method for stock market classification. The outcome is instrumental for investors, companies, and researchers whereby it will enable them to formulate their plans according to the sentiments of people.

VI. CONCLUSION & FUTURE SCOPE

Stock trends are particularly important and unpredictable in the world of finance. Daily headlines are just one of the many variables that influence stock trends. The aim focused in this publication is to predict direction of movement for specified stock. We have validated relation between predicted news sentiment and future stock prices. This is quite important on the advent of stock markets which is useful to people for providing multiple avenues to make their investment grow. It is a potential study path to investigate different algorithms to determine whether they are strong enough to forecast for the longer term. The work presented here may be optimized further by hybrid method of statistical and machine learning.

REFERENCES

- [1] Shengting Wu, Yuling Liu, Ziran Zou & Tien-Hsiung Weng, S_I_LSTM: stock price prediction based on multiplied data sources and sentiment analysis, Taylor & Francis, 2021.
- [2] P Shah, S Pandya, K Kotecha, Harvesting social media sentiment analysis to enhance stock market prediction using deep learning, PeerJ Computer Science, 2021.
- [3] Igupta, T Madan, S Singh, A Singh, HISA-SMFM: HISTORICAL AND SENTIMENT ANALYSIS BASED STOCK MARKET FORECASTING MODEL, Semantic Scholar, 2022.
- [4] T. Sanboon, K. Keatruangkamala, and S. Jaiyen, "A Deep Learning Model for Predicting Buy and Sell Recommendations in Stock Exchange of Thailand using Long Short-Term Memory," in 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), 2019, pp. 757–760.
- [5] Ying-Lei Lin, Chi-Ju Lai and Ping-Feng Pai, Using Deep Learning Techniques in Forecasting Stock Markets by Hybrid Data with Multilingual Sentiment Analysis, MDPI, 2022.
- [6] P Shah, S Pandya, A-Review-On-Sentiment-Analysis-Methodologies-Practices-And-Applications, IJSTR 2020.
- [7] Yang Li, Yi Pan; A novel ensemble deep learning model for stock prediction based on stock prices and news, International Journal of Data Science & Analytics, 2021.
- [8] Dogan, E., & Kaya, B. Deep learning based sentiment analysis and text summarization in social networks. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE. (2019).
- [9] James Briggs. Sentiment Analysis for Stock Price Prediction. <https://towardsdatascience.com/sentiment-analysis-for-stock-price-prediction-in-python-bed40c65d178>, last accessed 2020. Information Technology (IIT), 2015 11th International Conference on IEEE, 2015.
- [10] S Ghite, H Khatavkar, K Kotecha, N Pandey, Explainable stock prices prediction from financial news articles using sentiment analysis, PeerJ Computer Science, 2021.
- [11] Yumo Xu and Shay B. Cohen, Stock Movement Prediction from Tweets and Historical Price Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), pages 1970–1979, Melbourne, Australia, July 15 - 20, 2018. 2018 Association for Computational Linguistics.
- [12] Kalyani Joshi, Prof. Bharathi H. N., Prof. Jyothi Rao, Stock Trend Prediction Using News Sentiment Analysis, 2016.
- [13] Ghaith Abdulsattar A. Jabbar Alkubaisi¹, Siti Sakira Kamaruddin¹ & Husniza Husni¹, Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers, Canadian Center of Science and Education, Vol. 11, No. 1; 2018, ISSN 1913-8989 E-ISSN 1913-8997
- [14] Jigar Patel, Sahil Shah, Priyank Thakkar, K. Kotecha, Predicting stock market index using fusion of machine learning techniques, Expert Syst. Appl. 42(4), 2162–2172 (2015)