

Stock Market Prediction Using Machine Learning and Twitter Sentiment Analysis: A Survey

Likitha Thripuranthakam^{1*}, Shreya B. Gohad², H. N. Sujay Singh³, G. Charan⁴, Rashmi Amaradeep⁵

^{1,2,3,4}B.E. Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, India

⁵Associate Professor, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, India

Abstract: Sentiment analysis has received a variety of interest in the ultimate decade in particular on the availability of statistics from social media web sites due to this quite a few researchers are showing greater interest on this subject. Many researchers believe that the mood of people or sentiment expressed by using them on social media has an impact on economic markets actions. This paper plans to notice the effect of opinion communicated by means of Twitter on the financial exchange and afterward anticipate the moves of the financial exchange for the next days. For extracting tweets data from Twitter, we will be using Tweepy and for the verification of extracted information, we will use Yahoo Finance. To predict we intend to apply three machine algorithms Auto Regressive integrated moving average, long short-term memory, and Linear Regression. The expectation is to apply authentic stock information in relationship with opinion examination of information features and Twitter tweets information, to anticipate the future pace of a stock of interest.

Keywords: Sentiment analysis, Stock market prediction, Machine Learning, Twitter.

1. Introduction

Prediction of stock market movements may be a very challenging task because of non-linear and dynamic nature of stock markets. Now a days the employment of social media has reached another level. The information about public emotions has become considerable on social media. Social media is rebuilding into sort of an optimal stage to share public sentiments about any subject and comprises of huge impact on typical assessment. Techniques which use machine learning will give more accurate, specific and easy to predict stock market movements. This case makes Twitter kind of a corpus with valuable data for researchers. each tweet is of 140 characters long and speaks popular opinion on a subject concisely. The data exploited from tweets are very beneficial for generating predictions. Sentiment analysis of twitter data and sentiment classification is that the task of judging opinion in a very piece of textual content as positive, negative or neutral. During this task a method for predicting stock movements is developed using Twitter tweets about various enterprise. Sentiment analysis of the accumulated tweets is employed for prediction model for locating and analyzing correlation among contents of news articles and stock prices and so making predictions for future prices are developed by the use of

machine learning.

2. Literature Survey

There are a lot different data sets, techniques, machine learning models that have been utilized by many distinct researchers to predict the stock market movement by using sentiment analysis. In this section there some of the work is defined.

Saloni Mohan et.al, [1] have gathered two different datasets for this exploration. The everyday stock cost dataset incorporates shutting stock costs of 500 organizations from February 2013 to March 2017. He additionally collected news stories for the S&P 500 organizations from February 2013 to March 2017 from global day to day paper sites. A lot of information is being created for different organizations and, on the off chance that this information was to be handled physically, it very well may be hard to accomplish a calling time. in this manner, the creator utilized profound learning models to handle the information being produced. The creator has proposed a web-based scrubber to obtain the information, literary realities and stock costs, from the on the web. This information is taken care of to an information pipeline which processes the information and sends them to the AI motor, which recognizes feelings inside the gave messages and their impact on stock costs or anticipating the more extended term cost of the stock. the creator has referenced the models they prepared, their improvement and boundary tuning. in their investigations, they separated the dataset into 90% train, 5% approval, 5% test. They utilized autoregressive (AR), coordinated (I), and moving normal (MA) models anticipated transient estimating. another model they have utilized is Facebook Prophet on this test, they coordinate the model with the information comprising of stock costs of each organization. The creator has utilized RNN LSTM engineering that is gainful for demonstrating inconsistent times of data. The writer presumed that RNN models achieved well when contrasted with ordinary models like ARIMA and Facebook Prophet. The RNN model performed well for stable stocks and had the option to agree with their qualities, yet performed ineffectively for low cost and high unpredictable stocks. The RNN-pp model gave the incredible outcomes across all tests. This model done well

*Corresponding author: likitha.tv@gmail.com

for organizations for which we had more printed information. RNN models are best appropriate at anticipating stock costs contrasted with traditional ARIMA and Facebook. Kesavan M, Karthiraman *et. al.*, [2] analyzed four deep neural networks explicitly RNN, Multi-Layer Perceptron (MLP), CNN and LSTM units and closes CNN did well among various models. The expectation can be moreover exceptional by utilizing a half breed network that joins different models. Zhang K.*et al* proposed Generative Adversarial Networks (GAN) based way to deal with foresee close cost of the stock wherein the generator is made utilizing LSTM and discriminator is utilized MLP. The work introduced by Shi. L, *et. al.*, completed preparing a profound brain network for message-based forecast, picturing designs and assess with genuine ways of life situation and it is improved by the translation of other web-based entertainment messages along with monetary news occasions. A forecast approach through consolidating autoencoder, profound learning model and confined Boltzmann machine was added by Chen. L.A profound learning based multi feed brain network was proposed by Long W. Lu. to remove highlights from multivariate chronicled financial time-series information. The fig 1 features the itemized design outline of the indicator model. A half breed structure upheld AI methods for marker value forecast is presented by Chen Y. *et. al.*, inside which loads were appointed to highlights over preparing stage utilizing Support Vector machine (SVM) and through testing stage highlights with loads were used by K-closest neighbor (KNN) calculation.

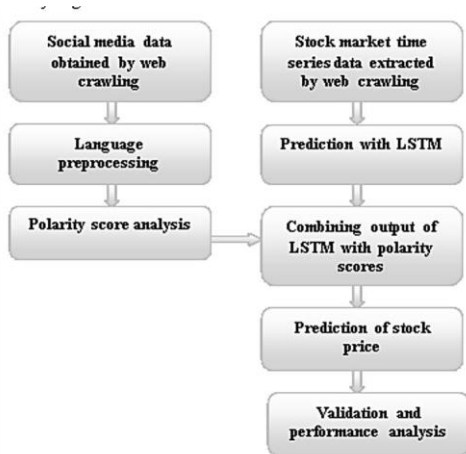


Fig. 1. The detailed architecture of the predictor model

device mastering models, SVM, ANN, naïve-Bayes and random forest were employed by Patel J. *et. al.*, Sentiment analysis technique analyses the tweeter tweets and classifies the sentiment of the tweets into three categories specifically positive, negative or neutral. A simple algorithm has been proposed to associate the feelings and time-series data prediction which incorporates four essential activities:

- A. Information Extraction and preprocessing
- B. Sentimental analysis of social media data
- C. Prediction with LSTM on statistic data input
- D. Combining sentiments with time series prediction.

The overall performance of the proposed stock price prediction technique can be analyzed supported the parameters like error percentage, accuracy, and precision. The paper provides an explicit approach for exchange prediction within the perspective of Indian economic system. It makes use of sentiment analysis to extract polarity scores from the news and social media content and consists of the extracted sentiments collectively with historic stock data to forecast the stock price.

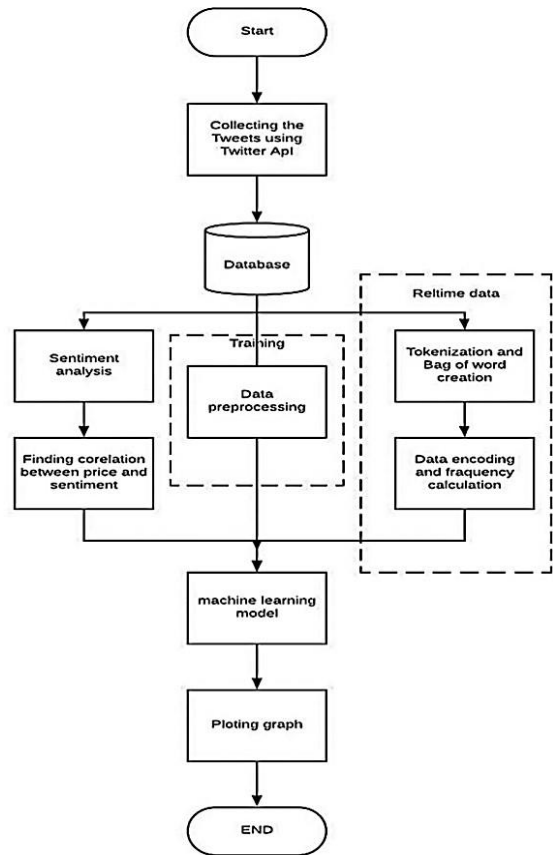


Fig. 2. Framework for Predicting stock price using sentimental analysis through twitter data

Niveditha N. Reddy *et. al.*, [3] proposed procedures at expectation of costs. Likewise proposed not many specialized and major strategies. while the exploratory methodology is utilized for outer data like cost and interest. The method utilized is SVM and opinion analyzer have been ended up being the most simple and productive model for the forecast of varieties in stock cost, basically founded on the inclination of feelings of the tweets. The specialized techniques utilize the notable information comprising of: instability, volume of exchanging, past preparation and numerous others. To obtain exact outcomes concerning expectation of the stock cost straight classifiers are utilized [2] or a similar methodology can be accomplished via utilizing the S&P500 record [7]. The strategy utilized is LSTM. Fig. 2 demonstrates the proposed structure for anticipating stock cost utilizing examination through twitter information. This approach confirms the adequacy of an assortments of Recurrent on the utilization of AAPL ticker from NASDAQ trade. The precision of this model is almost half.

Bollen makes sense of about the overall population temperament there are three assortments of public states of mind they're fulfilled, quiet, tension. The methodology utilized is the wistful examination calculation wherein the feeling of each tweet and each sentence is determined. The nostalgic investigation is utilized on the tweets which can be aggregated from twitter, we obtain tweets with the assistance of Twitter API. The client must have the API key verification, which is expected by the API. The python library is alluded to as Tweepy. The literary substance information in a tweet will contain parcel of additional information to not be thought about for the investigation of feelings. The assembled information will be tokenized, and a sack of words is shaped. next ascertain the recurrence and feeling of each tweet which can be utilized as the highlights for our model. For this intention, the tokenization by utilizing sack of words creation, information encoding and recurrence estimation. After the model is prepared, the relationship is seen among the cost and feeling. The task proposed an AI model purposes Random Forest for stock cost forecast the utilization of Twitter suppositions. essentially founded on the studies and correlations completed with all the different AI models at the stock cost expectation the utilization of twitter, to explore and anticipate in view of the public mind-set utilizing nostalgic investigation, to find out about the public mind-set the tweets are grouped into positive, unbiased and negative.

Rakhi Batra *et. al.*, [4] done feeling investigation on tweets related with Apple organization items. They used StockTwits to extract the company-related tweet data through pipeline API. It offers options for APIs to gather various assortments of information for example search API and stream API. Authors used seek API as they needed historical data of tweets. The result of this technique comes as JSON objects of tweets and these articles incorporate tweet distinguishing proof, client id, time, tweet text-based content, retweets, the feeling of clients on that tweet (bullish and negative), from there, the sky is the limit. The JSON data is then transformed to CSV file format. The flowchart of all of the proposed procedure is displayed in fig three the accompanying advance is to bring securities exchange information of Apple from Yahoo Finance from 2010 to 2017 in succeed design and the qualities were close cost, opening value, low and excessive cost, volume, and changed close. They processed the extracted stock records through using various natural Language Processing strategies (which includes text processing, Tokenization) to remove emoticon, punctuation, URLs, stop words, numbers, etc. for exact forecast from printed information. To reject the prevent words from the rundown creators have utilized a characteristic language tool compartment (NLTK) which consolidates a word reference of stop words. each word in the rundown of words is contrasted with word reference words, on the off chance that it coordinates, the comparing word is prohibited from the rundown. To filter out the symbols from the tweet data they used R proگرامing "sub" function. The stock information separated from Yahoo Finance is then preprocessed and used to conclude whether on a particular day the stock cost expanded or diminished. To settle on this decision, the end cost of today was deducted from the

end cost of yesterday, assuming the final product is more noteworthy than zero methods security (portion of an organization) cost is expanded and an individual can offer the security to acquire benefit. On the other hand, on the off chance that the thing that matters is significantly less than zero methods the security cost is diminished and an individual can buy the security or hold on the off chance that he/she has any. at last, the buy and sell determination was determined for the entire days and two segments for example the date and stock buy choice were presented in the information record. To compute the missing qualities from end of the week information they utilized a trademark that requires the day preceding today and resulting day's cost to find this day's value ($Y = (\text{PreviousDay} + \text{NextDay})/2$). The creators utilized the SVM model to anticipate the opinion of each tweet. They involved SVM as it is sensibly powerful to overfitting, it can manage huge element spaces and it is memory productive. To begin with, every one of the information was parted into preparing information and test sets. They utilized the R "createDataFunction" approach with 80% preparation set likelihood to make the preparation set and test set. Then they utilized preparing sets to prepare the SVM model and both preparation and test exactnesses were determined to assess the model. Fig 3 demonstrates the flowchart for proposed strategy. The ensuing advance as indicated by the flowchart given in [7] is to combine the feeling information and the stock cost information separated from Yahoo Finance. Then, at that point, they again utilized the SVM model to foresee securities exchange development and to give suggestions on whether an individual ought to trade an offer.

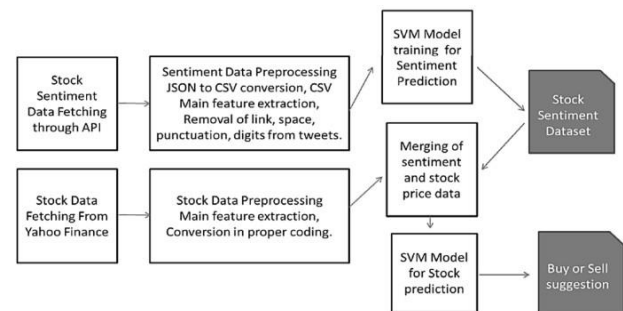


Fig. 3. Flowchart of methodology

They introduced the results as disarray networks. They applied the idea through gathering feeling information and stock cost market information and fabricated SVM models for forecast and in the last, they estimated the expectation exactness. They achieved 75.22% preparation exactness and 76.68% test precision.

Mehar Vijn *et.al.*, [5] amassed the beyond a decade recorded information of few organizations like Nike, Goldman Sachs, Johnson, JP Morgan from Yahoo Finance. The dataset contained records about the stock like high, Low, Open, close, contiguous close and volume. They removed day wise terms of the data. They made six new factors at making stock shutting costs. those factors were acclimated train the model. the new factors are as per the following:

1. Stock high short Low price(H-L)
2. Stock high close Low price(O-C)
3. Stock value's seven days moving normal (7 DAYS MA)
4. Stock value's fourteen days moving normal (14 DAYS MA)
5. Stock value's 21 days moving normal (21 DAYS MA)
6. Stock value's standard deviation for the beyond seven days (7 DAYS STD DEV)

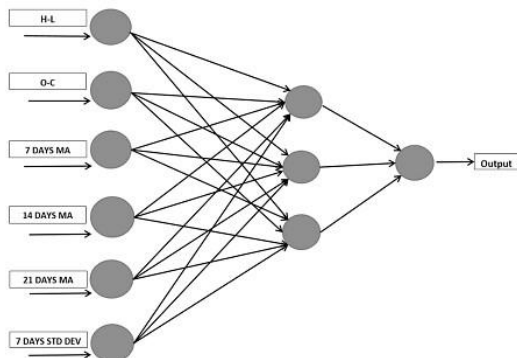


Fig. 4. Detailed architecture of Artificial Neural network (ANN) for stock price prediction

He utilized Artificial Neural Networks (ANN) model to prepared of those factors. The model works with three layers. It incorporates input layer, stowed away layer and furthermore the result layer. The info layer comprises of ongoing factors which are H-L, O-C, and 7 DAYS MA, 14 DAYS MA, 21 DAYS MA, 7 DAYS STD. The loads on each info load is increased and added and shipped off the neurons. The point-by-point engineering of ANN for stock value expectation is given in fig 4. The secret layer or the actuation layer comprises of those neurons. the entire weight is determined and is moved to the 0.33 layer which is that the result layer. The result layer comprises of just 1 neuron that can offer the normal benefit concerning cost of the stock. Irregular Forest (RF) a group machine it was utilized to learn approach. This model is fit for performing both relapse and grouping undertakings. This model goals at limiting gauging mistake by regarding the trade examination as an arrangement issue and upheld preparing factors anticipated the ensuing day cost of the stock for a specific organization.

Arpit Goel *et.al*, [6] utilized fundamental datasets. 1. stock file modern normal (DJIA) values from June to December. The data was gained utilizing Yahoo Finance and incorporated the open, close, high and low qualities for a given day. 2. Freely accessible Twitter information containing very 476 million tweets, for example, more than 17 million clients. the data gained from the previously mentioned sources was to be pre-handled to frame it suitable for trustworthy examination. the creator has preprocessed the DJIA information inside the accompanying way. The procedures are affirmed in fig. 5. The Twitter information was accessible for the entire days existing in the giving span, the DJIA values gained utilizing Yahoo Finance was missing for ends of the week and different occasions when the market is shut. To complete the information, creator has approximated the missing qualities

utilizing an inward capacity. Thus, assuming the DJIA esteem on a given day is x and furthermore the resulting accessible data is y with n days missing in the middle, we inexact the missing information by assessing the essential day after x to be $(y+x)/2$ then, at that point, following the indistinguishable methodology recursively until all holes are filled.

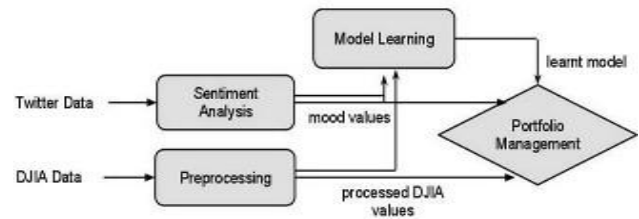


Fig. 5. Techniques

This guess is advocated on the grounds that the stock information by and large follows a curved capacity, until truth be told at oddity points of surprising ascent and fall. Assuming we look at the general development of financial exchanges, it's related for certain startling leaps/falls and a brief length of little variances round the new worth. nonetheless, such leaps/falls are because of a few significant variations and can't be anticipated. besides, as we as a whole acknowledge a great many people memory is incredibly short and even as the market is additionally exchanging at a way more significant level than the previous year, that doesn't infer that serenity are a lot over going before year; public disposition might be an absolutely nearby measurement. different feeling investigation strategies like glossary age, Tweet Filtering, day to day score Computation, score Mapping were proposed.

Nehal Shah *et.al*, [7] gathered a classifier this is ready as "positive" or "negative" or "impartial" named tweet corpus. by using executed classifier, they have arranged future tweets as on the other hand "positive", "negative" or "impartial" basically established on pre-arranged tweets features. Two ordinary classifiers used for text gathering: Naïve Bayes Bernoulli and sponsorship Vector structure models. It is applied to Notwithstanding that Naïve Bayes Classifier. Guileless Bayes Classifier is used for protections trade assumption. stock tweets and stock expenses were connected with expect costs for explicit day. Train classifier was used to expect stock expenses for looming days. Ensuing to exploring tweet assessment, they've expected protections trade expenses of express associations and also try to anticipate region wise all over fall through getting records from Bombay stock exchange site.

Bhavya K *et. al*, [8] used obvious stock records along the edge of feeling examination of stories elements and Twitter posts, to anticipate the future expense of a stockpile of income. The elements were secured by scratching the net site, FinViz, while tweets were taken using Tweepy. both were taken apart using the Vader Sentiment Analyzer. inspect the impact of feeling imparted through Twitter tweets on stock expense gauge. Twitter is that the electronic amusement stage which gives a released stage for every person to unequivocal their considerations unreservedly. They bring the live twitter tweets of the specific association the use of the API. all the stop phrases, exceptional characters are removed from the dataset.

The filtered truths is applied for feeling examination using Naïve bayes classifier. As required, the tweets are portrayed into positive, negative and fair tweets. To expect the stock worth, the stock dataset is gotten from yahoo finance API. The stock data connected with the tweets data are given as commitment to the AI model to activate the result. XGBoost classifier is used as a model to predict the monetary trade cost. The procured conjecture regard is differentiated and the certifiable stock exchange regard. The suitability of the proposed project on stock charge conjecture is checked through tests different associations like Apple, Amazon, Microsoft using live twitter data and everyday stock data. After this, they followed invalidation acknowledgment measures to perceive "fantastic" and "terrible" wherein they endeavored to embrace and do feeling assessment on twitter dataset and orchestrate them into great, critical and unprejudiced approach to acting of people. Expecting the full study includes a positive, charmed disposition o if something is refered to with positive affiliations. Accordingly, it is considered as a magnificent affirmation. Expecting the whole comment joins a terrible, depressed then again if something communicated with negative affiliations. Along these lines, it is contemplated as a negative attestation. Accepting the review gives no private perspective in the comments and researches sends information. in addition, the AI model XGBoost outfits more accurate characteristics in connection with various models. therefore, the utilization of feeling assessment of twitter data and stock data from yahoo finance API, we expect the protections trade esteem that is profitable for anticipating future stock expense. inside the long run, we consider to extra overhaul the add the going with locales. In any case, our assessment is denied to 16 associations. an expansion to more broad game plan of associations or everything Twitter data could yield more encounters into the data, inciting more utilize open expense assumption. second, we use the optional inclination marks given through Twitter clients due to the truth the base truth data for model readiness. As assessed, this data has recently 89.8% accuracy, recuperating arrangement data is expected to help the idea of the assessment analyser. in the long run, this endeavor assesses association at ordinary granularity due to the truth the stock data are only available at the reliably level. It will presumably be captivating to study associations at a superior granularity like hourly. The accuracy accomplished was 89.8%. thusly, online amusement like twitter may in like manner be used as a source to expect the exchange cost with most outrageous precision.

Vikram Kolasani *et.al*, [9] affirmed the adequacy of utilizing Twitter presents on anticipate stock developments. they start by utilizing preparing different models on the Sentiment 140 Twitter records. They saw that help Vector Machines (SVM) accomplished acceptable (0.83 exactness) in the wistful examination, so they utilized it to anticipate the feeling of tweets for each and every day that the market was open. resulting, they utilized the wistful examination of one year's information of tweets that involve the "securities exchange", "stocktwits", "AAPL" catchphrases, fully intent on foreseeing the comparing stock costs of AppleInc. (AAPL) and consequently the United States' Dow-Jones modern normal

(DJIA) list costs. models, Boosted Regression trees and Multilayer Perceptron Neural Networks were acclimated anticipate the worth distinction of AAPL and DJIA costs. They show that brain networks do significantly better compared to customary models at stocks' cost forecast. all through this paper, an opinion labeled Twitter dataset of 1.6 million tweets collected from Sentiment 140 are visiting be utilized for feeling arrangement. Then, the Boosted Regression Tree and Multilayer Perceptron models are utilized for foreseeing the next day's stock development with this current day's tweets containing the "securities exchange", "StockTwits", "AAPL". This paper tried "Are brain networks less difficult at anticipating the stock trade development than regular models?". 5 % of the preparation information from the opinion 140 dataset was acclimated test the prepared models. Likewise, 5 distinct models are prepared on the dataset. especially, the models utilized are Logistic Regression (LR), help Vector Machines (SVM), choice Tree (DT), Boosted Tree (BT), and Random Forests (RF). The 5 models are talented with the preparation dataset. After this, every one of the models are acclimated foresee the feeling an incentive for the test information tweets. Following this progression, the creators contrasted expected feeling values and genuine opinion upsides of the test informational index tweets.

Siti Sakira Kamaruddin *et.al*, [10] proposed an elective structure connected with feeling examination from Twitter posts. Their proposed structure addresses a high-level plan of grouping model that attempts to upgrade the arrangement precision to help leaders inside the area of stock trade. This model beginnings with information assortment part and in second stage filtration is finished on information to encourage just the relevant information. the most fundamental stage is that the naming part inside which extremity of information is set and pessimistic, positive or nonpartisan qualities are allocated to articulations of individuals. The fourth is that the arrangement present which suitable examples of stock trade are recognized by utilizing hybridizing NBCs. The last stage is execution and evaluation. The entire proposed system is displayed in fig 6. This review proposes to a Hybrid Naïve Bayes Classifiers (HNBCs) as an AI approach for stock trade grouping, therefore addresses a helpful report for financial backers, organizations and analysts and can help them to figure out their arrangements as indicated by opinions of people. They separated dataset from Twitter by utilizing Twitter PC program Interface (API) in light of the fact that the streaming can give a constant stream of the information with refreshes. Besides, pre-handling the data set utilizing tongue Pre-handling (NLP), wherein NLP handling start with alterations step and end with the extraction cycle to remove the favored highlights. second stage likewise incorporate Filtering step that is connected with the substance extraction process from the gathered tweets after alterations step. This study centers around separating significant elements, which are spatial and fleeting highlights of tweets. Spatial data about tweets are frequently gotten by utilizing approaches: the essential one is naturally aggregating the precise spatial data accessible on Twitter and furthermore the subsequent one is roughly surmising what is happening of the client from the

client profile. Thirdly, is that the master marking strategy wherein during this stage the dataset is arranged into positive, negative and impartial extremity. Fourthly, is that the order level utilizing HNBCs could be a gauge NB with joining Multinomial NB (MNB) as managing very one component simultaneously, Bernoulli NB (BNB), which is uncommonly simple for additional preparation informational index and Semi-Supervised Naïve Bayes (SSNB), which is reasonable for train NBCs for marked information [10].at last, the order model's general presentation is assessed utilizing review, accuracy, F-measure and precision.

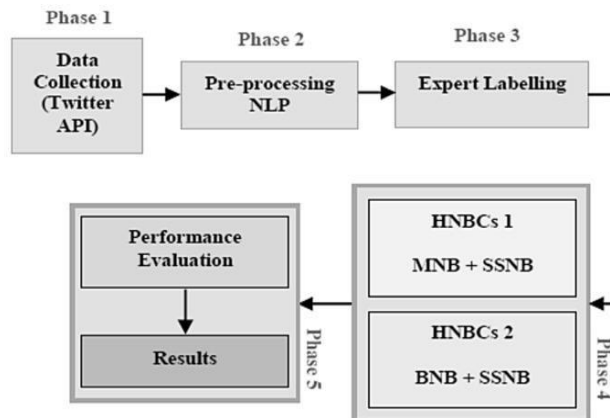


Fig. 6. A conceptual framework

3. Future Scope

Predicting stock market movement through sentimental analysis is a step-by-step process through the literature survey we got to know the basic structure of the whole process. For future work we intend to use of Tweepy to fetch twitter data. To predict we propose to use three algorithms ARIMA, LSTM and Linear Regression. Furthermore, to increase the efficiency more factors such as location of the twitter account, number of retweets and popularity of the twitter user. can be taken into consideration while performing a wistful investigation on the tweets removed from twitter Sentiments from various social media platforms could be incorporated to upgrade the general exhibition of the framework and taking huge informational collections for preparing will assist with expanding the precision of the proposed framework.

4. Conclusion

Predicting movement of Stock Market is an interesting research area. Prior foreseeing Stock Market developments were about arbitrary mathematical expectation in light of the accessible authentic information however with the expansion of ongoing social money individuals' conviction, temperament and response to certain occurrences are additionally taken in thought while anticipating stock development. Similarly, as with the development of web-based entertainment stages and it made accessible online to expand number of convictions. These are different investigations and explores which proposes that feeling examination of public state of mind got from Twitter channels can be utilized to figure developments of individual stock costs. Through the literature survey it's clear that there are few defined processes which was common almost all the research papers i.e., twitter data extraction, processing of the extracted data, performing a sentimental analysis on the extracted data validating the extracted data using Yahoo Finance and then training a machine learning model by using the same data sets. We propose to use Tweepy to fetch twitter data. To predict we intend to use of three algorithms ARIMA, LSTM and Linear Regression. All this data will then be combined with the sentimental analysis of tweets. And finally, it will recommend whether the price will rise or fall.

References

- [1] Saloni Mohan et al" Stock Price Prediction Using News Sentiment Analysis", 2019, IEEE, pp. 205-208, 2019.
- [2] Kesavan M. Karthiraman J, et. al., "Stock Movement Prediction with Historical Time Series Data and Sentimental Analysis of Social Media Data", 2020, IEEE, pp. 477-482, 2020.
- [3] Niveditha N. Reddy et al., "Predicting Stock Price Using Sentimental Analysis Through Twitter Data", 2020, IEEE.
- [4] Rakhi Batra "Integrating StockTwits with Sentiment Analysis for better prediction of Stock Price Movement", 2018, IEEE.
- [5] Mehar Vjijh "Stock Closing Price Prediction using Machine Learning Techniques", 2019, ICCIDS.
- [6] Arpit Goel "Stock Prediction Using Twitter Sentiment Analysis, Neural Networks", 2019, CS229,2:27:1-27:27.
- [7] Nehal Shah "Stock Market Movements Using Twitter Sentiment Analysis" IJSDR, vol. 4, 2019.
- [8] Padmanayana, Varsha, Bhavya K "Stock Market Prediction Using Twitter Sentiment Analysis", IJSRCSEIT, vol. 7, Issue 4, pp. 265-270, 2021.
- [9] Sai Vikram Kolasani, Rida Assaf "Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks", Journal of Data Analysis and Information Processing, vol. 8, pp. 309-319, 2020.
- [10] Siti Sakira Kamaruddin "Conceptual Framework for Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naïve Bayes Classifiers", International Journal of Engineering & Technology, vol 7, pp. 57-61, 2018.