IMPACT OF PUBLIC SENTIMENTS ON STOCK PERFORMANCE

Amit S. Bardhan[1], Prof. Sandip Kumar S. Chauhan[2], Dr. Dinesh B. Vaghela[3]

[1]Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat
[2]Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat
[3]Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat

Abstract —Stock market is considered to be a sector where the people, other companies and mutual funds organizations heavily invest in order to gain maximum profit. As compared to the other investment sector stock market gives more returns if one is able to perform proper analysis in order to rectify the factors on which the momentum of the stock price depend. Stock prices in the stock market are governed by various factors like company profit and loss ratios, projects company has acquired, scandals, economic outlook, inflation, political shocks , investor sentiments etc. Out of all these one of the major factor is public sentiment. We can analyze public sentiment and can classify investors views are positive or negative for the particular stock or company. For collecting public sentiments we can use various API of python and for classifications we can use various machine learning algorithms. Sentiment classification using machine learning algorithms provide higher accuracy than others.

Keywords- Market analysis, Sentiment analysis, Machine learning, NLP

I. INTRODUCTION

 Investments in stock prices are always considered to be one of the best choices for the investors who believe in making more profit as the analytics figures prove that as compared to other sectors share market gives better profit ratios. But on the same side without proper analysis of the company investing in shares may lead to huge loss. The stock price of a particular share in a share market depends upon many factors like inflation, economic outlook, scandals, profit and loss ratios of the company as per quarter results, projects the company as finished and acquired, public sentiments etc.

Out of all these factors we consider of over here public sentiments that are spread by various individual investors, financial organizations that heavily impact the price of a particular stock price. A public sentiment can be classified into either positive, negative or neutral. People use online digital platforms like Twitter, Facebook, LinkedIn, Flicks etc. to share their personal opinions or sentiments [5]. Apart from that companies that are listed in the stock market also have their own social media profile where there share information regarding various new ventures or actions that the company is going to conduct or also ready achieved some goals.

The reason behind considering public opinions or sentiments for analyzing the stock price momentum is that a huge number of people are sharing a lot amount of digital content online which greatly drive the market. Everyone is using smart phones to transmit their views and opinions using social media platforms and blogs. Data available from these sources are having large amount of priceless information that has to be filtered and mined for predicting the stock momentum [7].

Apart from public sentiments here we also gather data from various news portals using web crawlers and a huge amount of sentiment corpus can be generated for data analytics. One of the major problems while dealing with such corpus is that the data here is in fully unstructured format [9]. And dealing with such a unstructured format of data needs high amount of data preprocessing in order to perform data cleansing operation before actually carrying out the sentiment classification and classify the public opinions in terms of positive and negative sentiments [11].

Once the data cleansing operation is performed we will use various supervised machine learning algorithms like Naïve Bayes, SVM, Neural Networks , Decision Tree etc into order to classify the words in terms of there their actual sentiments and can assign the positive +1, negative -1 and neutral 0 to the analyzed opinions [12]. We are supervised ML algorithms as these algorithms are considered to be very efficient as compared to other available algorithms. Also various data mining open source tools like Rattle, Weka, KNIME etc. are available in which experimental evaluations can be easily carried out and a working model can be created for testing and prediction the future stock momentum for prediction with higher accuracy.

II. RELATED WORKS

Focusing on the stock market prediction Antonios and his colleagues have worked on Facebook's Gross Happiness Index (FGNHI) has been developed by Facebook data team and offers daily sentiments. FGNHI is calculated on basics of millions of Facebook participants where the sentiment index can also be classified in terms of positive and negative words. So they examine Facebook's daily sentiment index and scrutinize its relation to the stock market [1]. The internet expressed sentiment are potentially very helpful as many people spend considerable amount of time on reading digital content online which comprise of valuable insights, reaction to other sources of new etc as stated by Colm Kearney and Sha Liu [2].

Detection of Negation in news is considered to be very important. Nicolas and his colleagues uses two approaches one is rule based another is machine learning in order to perform the classification and to detect the negative sentiments [3]. In another research done by Nuno Oliveria and his colleagues forms dictionary of lexicon for which they are using Term Frequency-inverse document frequency (TF-IDF), Information gain (IG) and Pointwise Mutual Information (PMI) to compute the sentiment score [4]. Collecting the sentiments from the twitter feeds for some selected companies and using Granger causality test proving the sentiments in the stock related tweets can be used for analyzing the momentum if the particular stock [8]. Building a dictionary named SenticNet3 that contains both common and common sense knowledge in order to improve sentiment classifications has been stated by Erik Cambria and his colleagues [14]. Deep convolutional neural networks can also be applied to the corpus and sentiment prediction can be carried out efficiently [15].

III. PROPOSED METHOD

In our research Sentiment Analysis is been performed using natural language processing in order to methodically recognize and find out the actual information. It is used to define the actual feelings of the people by handling all the negations in the given text. We are using here NLTK (natural language toolkit) which is a free, open source and can be used with different platforms like R and python for working with human language data. It does the task of classification, tokenization, stemming, tagging, parsing and semantic reasoning. After performing all the data preprocessing on the given corpus we implement different machine learning to the prepared dataset or the corpus and finally the model is trained accordingly.

Machine Learning is a field of computer science which utilizes statistical approaches to the given computer system and provides an ability to gain knowledge from the given model [16]. It is broadly classified into supervised learning and unsupervised learning. In supervised learning we give the computer some example inputs and predefined desired outputs. But in case of unsupervised learning no particular targets are given, and the computer needs to learn by its own. List of famous ML algorithms are Decision Tree, Association rule, Artificial neural networks, Deep learning, SVM, Clustering, Naïve Bayes etc [2].

In our proposed method we fetch the data from tweets, Facebook posts and News articles. There after we combine create a corpus of sentiments and implement data preprocessing, which includes data cleansing, removing top words , negation handling and finally classification. After classification we relate the sentiment score with the price of the stock and try to predict the next day stock momentum.
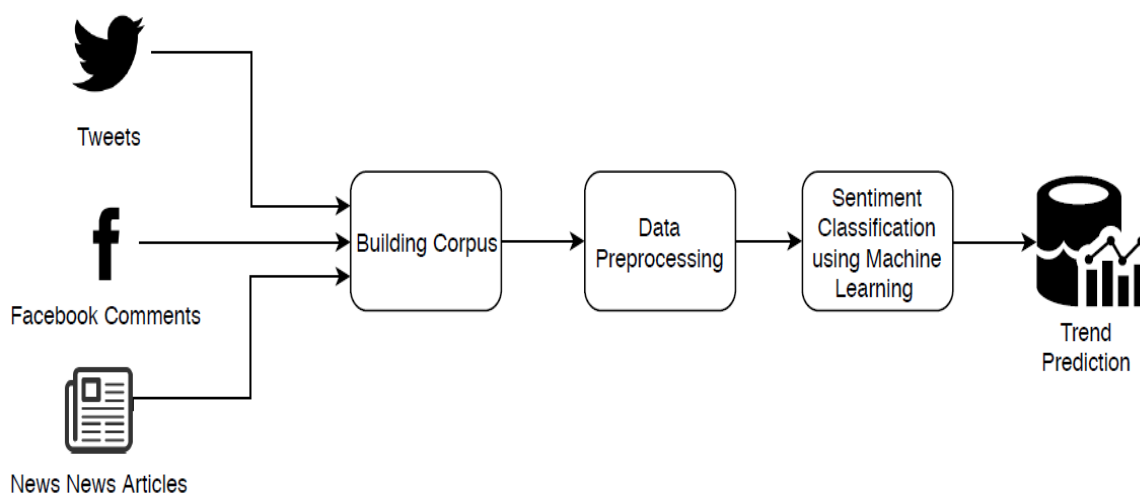

Fig. 1. Proposed Algorithm

IV.   EXPERIMENTAL EVALUATION

Experiment Environment and momentum prediction:
 For fetching tweets we are using a python client tweepy which is an official Twitter API, in which we are passing consumer key, consumer secret, access token and access token secret which will provided when you create twitter application. All these four parameters provide an authentication to twitter to fetch the tweets. Similarly we use Facebook Graph API that allows fetching data from Facebook. And lastly we use the web scrapping API BeautifulSoup for web data mining. We use Stocker class in python in order to fetch the stock prices of a particular stock for a specific period of time interval.

After fetching data we create a single corpus and assign the sentiment polarity to the given text of data as 1 for positive -1 for negative and 0 for neutral. And depending upon the sentiments score we try to predict the momentum of the particular share for the next day.
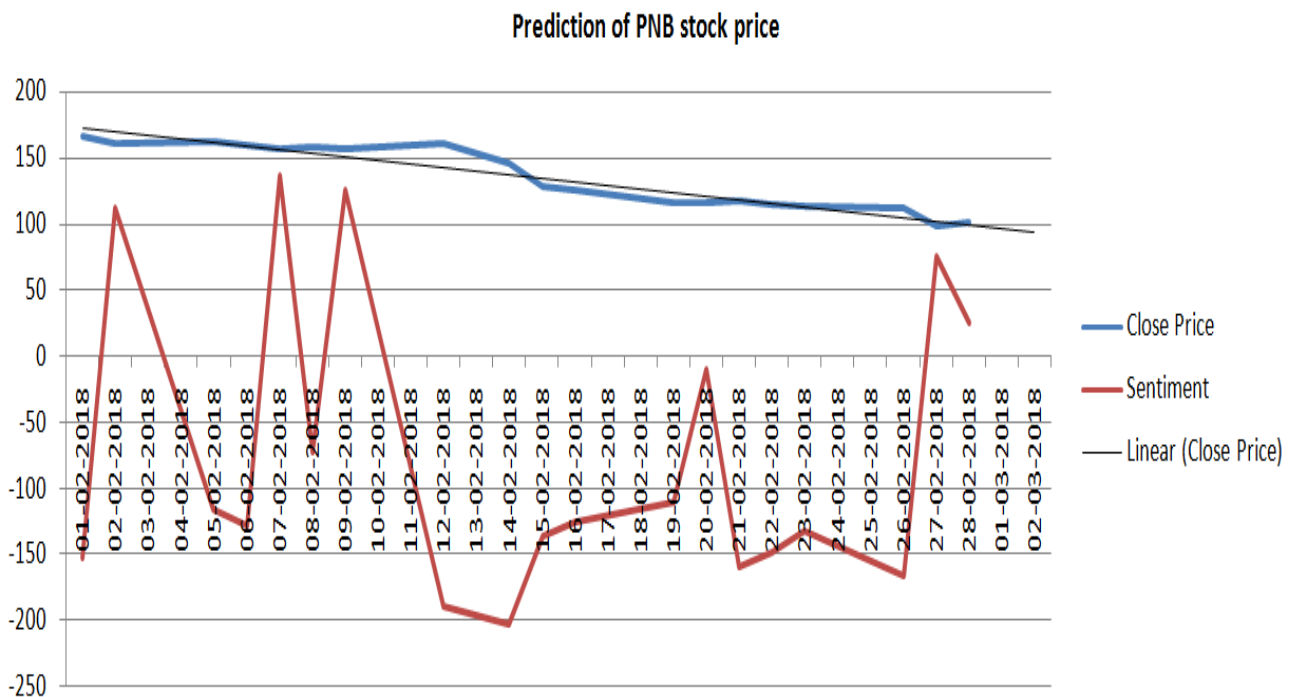

Fig. 2. Prediction output

The Fig. 2 shows the prediction of stock for the next two days. Here we have taken the stock price of PNB for the month of Feb 2018 and we have also calculated the sentiment score of that particular share on that day and tried to predict the next day stock momentum. Here the we are predicting the stock price of PNB for the dates 1st and 2nd of March and going to be negative. For predicting the price of next two days we are using Decision Tree algorithm of machine learning using an on open source data mining tool rattle.

V.   CONCLUSION.

Public sentiment analysis is of great use when it comes to opinion mining of or taking a survey of reactions of the people. As more and more people are using digital media for expressing views and opinion a huge amount of data across web is available that are used to get public sentiments. But dealing with public sentiments is very typical as we need to deal with unstructured data, so a very efficient data preprocessing has to be implement for the given collected corpus and then data analytics can be performed. Here we have used public sentiments to predict the market mood for a given particular stock and using ML algorithm we have tried to predict the momentum.

VI.   FUTURE SCOPE

To increase the accuracy of prediction other machine learning algorithms can be implemented. Also more factors can be taken into consideration instead of only working with the close price and sentiments. Dealing with live public sentiments can also be implemented for intraday trading predictions for a given stock.

REFERENCES

[1] A. Siganos, E. Vagenas-Nanos, and P. Verwijmeren, "Facebook's daily sentiment and international stock markets," *J. Econ. Behav. Organ.*, vol. 107, no. PB, pp. 730–743, 2014.

[2] C. Kearney and S. Liu, "Textual sentiment in finance: A survey of methods and models," *Int. Rev. Financ. Anal.*, vol. 33, no. Cc, pp. 171–185, 2014.

[3] N. Pröllochs, S. Feuerriegel, and D. Neumann, "Enhancing sentiment analysis of financial news by detecting negation scopes," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 2015–March, pp. 959–968, 2015.

[4] N. Oliveira, P. Cortez, and N. Areal, "Stock market sentiment lexicon acquisition using microblogging data and statistical measures," *Decis. Support Syst.*, vol. 85, pp. 62–73, 2016.

[5] S. P. Trends, S. G. Chowdhury, S. Routh, and S. Chakrabarti, "News Analytics and Sentiment Analysis to Predict," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 3, pp. 3595–3604, 2014.

[6] M. Skuza and A. Romanowski, "Sentiment Analysis of Twitter Data within Big Data Distributed Environment for Stock Prediction," vol. 5, pp. 1349–1354, 2015.

[7] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 306–324, 2015.

[8] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Stream-based active learning for sentiment analysis in the financial domain," *Inf. Sci. (Ny).*, vol. 285, no. 1, pp. 181–203, 2014.

[9] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining for market prediction: A systematic review," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7653–7670, 2014.

[10] D. D. Wu, L. Zheng, and D. L. Olson, "A decision support approach for online stock forum sentiment analysis," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 44, no. 8, pp. 1077–1087, 2014.

[11] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9603–9611, 2015.

[12] T. H. Nguyen and K. Shirai, "Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction," *Proc. 53rd Annu. Meet. Assoc. Comput. Linguist. 7th Int. Jt. Conf. Nat. Lang. Process. (Volume 1 Long Pap.*, pp. 1354–1364, 2015.

[13] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4: Sentiment Analysis in Twitter," *Proc. 11th Int. Work. Semant. Eval.*, pp. 502–518, 2017.

[14] E. Cambria, D. Olsher, and D. Rajagopal, "SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis," *Twenty-eighth AAAI Conf. ...*, pp. 1515–1521, 2014.

[15] C. N. dos Santos and M. Gatti, "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts," *Coling-2014*, pp. 69–78, 2014.

[16] D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," *J. King Saud Univ. - Eng. Sci.*, 2016.

[17] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Comput. Sci.*, vol. 17, pp. 26–32, 2013.

[18] M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," *Procedia Comput. Sci.*, vol. 87, pp. 44–49, 2016.

[19] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.

[20] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowledge-Based Syst.*, vol. 69, no. 1, pp. 14–23, 2014.