12th International Conference on Computational Intelligence and Communication Networks

Sentiment Analysis of Financial News

Arul Agarwal
Department of Computer Science and Engineering, SCOPE
Vellore Institute of Technology, Vellore, India
arulagarwal@gmail.com

Abstract— **Sentiment analysis is a subdiscipline covered under data mining and computational semantics. It refers to the comprehension of gathered data that is procured from sentiment rich sources like news, social media sites, reviews, and so forth. In the current era where data is becoming increasingly voluminous and yet crucial to all businesses, manual analysis of data does not remain viable in this fast-moving world. Thus, it is necessary to make use of artificial intelligence and datamining techniques. Amongst several variables, a key determinant that result in the fluctuation in stock prices is the gains or losses incurred by a company. As most traders get their information from news, it makes news as a core influential factor to forecast change in the stock market. This study focuses on sentiment classification and shows its effect on the change in stock market prices. It generates investing insight by applying sentiment analysis using VADER (Valence Aware Dictionary and Sentiment Reasoner) tool on some of the most liquid stocks.**

Keywords—Stock market, Data analysis, Data mining, Financial News articles, Sentiment analysis, VADER, Liquid stocks.

I. Introduction

The stock market is a significant component of the economy of a nation. It performs a crucial role in the growth of a business and trade of a country. Thus, it greatly influences the economy of the nation and with the advent of globalization now has a global effect. Therefore, all parties linked directly or indirectly to the stock market maintain a close watch on it. Sentiment analysis of news which is unstructured data can help in predicting market changes [1-2].

Market Sentiment indicators are essential in determining how the equity will bear in the upcoming future. For example, if there is a positive sentiment about a stock, the price for it will continue to increase or remain constant. On the other hand, deteriorating sentiments can be seen as harbingers to the drop in prices.

"Innovation is scaling at an exponential rate, and today we are handling loads of data if numbers are to be accepted. An ongoing report uncovers that the all-out data existing in the world will develop at a CAGR of 61% to 175 zettabytes by 2025 from 33 zettabytes in 2018." [3] Larger volumes of news indicate an event that may lead to a significant rise or drop in the price of assets.

The key reason behind choosing news headlines is that 70 – 80% of data in organizations is unstructured. Regardless of the genuine rate, there is a little uncertainty that the measure of unstructured data keeps on developing. Today numerous associations discover the volume of information testing and coordinating such humongous data with big business frameworks with the assistance of not well-prepared data management devices is a challenging task. Also, it accounts for kinds of risk not attributable to the market. As a result, a substantial amount of unstructured data remains untapped.

In this study, Finwiz is being used for news headlines and descriptions. Finwiz indexes articles from various prominent worldwide resources. It has been an integral element that offers regular and timely news to its users. We will use the VADER (Valence Aware Dictionary and Sentiment Reasoner) library in NLTK (Natural Language Toolkit) package for its lexicon-based sentiment analysis tool. This tool is used for its fast nature that can be used without any training data and thus, can be utilized on online streaming data as well.

II. Related Works

Many researchers have validated the efficacy of market sentiments in stock trend predictions [4-5], this extends to the Bitcoin exchange market as well [6]. Various approaches have been adopted to analyse market sentiment. They can be broadly classified into the lexicon-based approach and machine learning approach [7]. Lexicon based approach can be subdivided into dictionary-based and corpus-based. Machine Learning is further classified into unsupervised and supervised techniques, which include probabilistic classification, rule-based, decision tree classification, etc. Some of the models designed along these lines are discussed here.

Dev Shah et al [8] have suggested a dictionary-based model. It utilizes a python library called pattern to transform text corpus to numerical vectors. This library tallies the frequency of positive and negative words and summarizes them on the basis of frequency to generate sentiment score. The model can be utilized in other sectors as well. The drawback of the model is that the sentiment score for each token that is generated is not weighted and a simple summation of scores can lead to significant variations from the actual market sentiment.

M.S. Usha et al [9] have suggested generating market sentiment insight using unsupervised learning. Their proposed model is based on the Gibbs sampling algorithm and can detect sentiment and topic in a simultaneous manner. The database used in the paper is Multiple data sets i.e. it contains data from distinct kinds of merchandise reviews crawled from an e-commerce website and thus, contains a variety of sentiment depending on the topic. The use of unsupervised learning method makes it highly portable. The drawback of

978-1-7281-9393-9/20/$31.00 ©2020 IEEE
DOI: 10.1109/CICN.2020.57

the proposed model is that it is unable to detect neutral opinions.

D.K. Kirange et al [10] propose a prediction model that is based on emotion classification of news to generate sentiment values that can be used to gain insight into the change in market trends. This model gathers news from various prominent news sources over a large period of time and then applies emotion classification to obtain sentiment polarity using techniques like naïve Bayes, SVM, KNN with KNN having the highest accuracy. Then the authors proceed to correlate their gained insights with stock market historical data and attempt to find a correlation. From the results, the accuracy of the SVM model for the accuracy of predicting change appears to be the highest.

Sneh Kalra et al [11] propose the use of Naïve Bayes classifier used for sentiment analysis used in conjunction with historical data of adjacent dates of the news to develop a prediction model that can predict stock prices with as accuracy between 65% to 92%. The model utilizes the stock news data set which is pre-processed and then the sentiment of each article is analysed using naïve Bayes classification. This is used in conjunction with stock variance around adjacent dates which is obtained from yahoo finance to generate investing insight by processing both the news data and the numeric data through various machine learning algorithms. Prediction using KNN resulted in the highest degree of accuracy for the same set of data. The drawback of this model is that the source of information that it utilizes is limited and can cause variation from the actual results.

Xiadong Li et al [12] have taken a different approach. A stock prediction system was proposed that utilizes a layered deep learning model for generating market sentiment which then applies the gained insight to a fully connected neural network to generate market forecasts. Their model makes sentiment analysis using dictionaries and uses it along with technical stock indicators to generate an output which is then passed through the neural network.

Other models proposed have covered methods like designing specific need-based sub-modules based on existing modules within the natural language processing module to analyse sentiment [1,13], Sentiment analysis using N-gram and Naïve Bayes Algorithm [14], Dictionary Based sentiment analysis [15], Sentiment analysis using mood classification and daily score computation to map sentiment scores [16-17], use of different time series analysis models [18]

III. METHODOLOGY ADOPTED

Data scraped from Finwiz is passed to sentiment analysis. Finwiz provides real-time stock updates and headlines from well-established newspapers including but not limited to Financial Times, Wall Street Journal, Bloomberg, Yahoo Finance, and so on. As the factuality and pertinence of news sourced are extremely vital, the choice of source of data is critical to prevent the addition of erroneous data. An addition of impertinent but factual headline can cause the sentiment analyser to generate a wrong sentiment score. For example, a possible headline like "Tesla's CEO Elon Musk crashes a party" can cause the sentiment analyser to generate an irrelevant sentiment score based on tokens "Tesla" and "crash" that may create inaccurate predictions.

The data is classified on the basis of what percentage of the sentence has a positive, negative, or neutral sentiment, and then generates a compound score on the basis of its normalized aggregate for the sentence. The sentiment score of each headline is tabulated and summarized so that it can be utilized to evaluate market sentiment regarding the stock.

Python tool VADER uses a lexicon-based approach for determining sentiment values of a sentence. This is used along with sentiment values assigned explicitly to keywords commonly found amongst news headlines, referring to stocks, such as Falls, Crushes, Plunges, etc so to allow the analyser to understand these words in their pecuniary sense.

IV. IMPLEMENTATION

1. Using HTML Files to get data

   HTML files of relevant stocks are downloaded and added to the dataset collected by the user. This is done manually to ensure no unnecessary load is put on the servers.

date		
2019-11-01	09:33PM	The Best Stock Market Ever Heads Into Stocks B...
2019-11-01	07:50PM	Garmin Flies Under the Radar. Its Stock Has St...
2019-11-01	06:52PM	Time to Buy Facebook (FB) Stock After Earnings...
2019-11-01	05:18PM	Nvidia's Jensen Huang named top-performing CEO...
2019-11-01	04:53PM	Googles Fitbit Acquisition Gets Instant Antitr...
2019-11-01	04:24PM	S&P 500, Nasdaq Hit Record Highs: Apple, Faceb...
2019-11-01	04:22PM	Google To Buy Fitbit, Takes On Apple In Data P...
2019-11-01	04:15PM	How Many Streaming Video Choices Are Too Many ...
2019-11-01	03:52PM	What Are Warren Buffett's Top 10 Stock Holdings?
2019-11-01	03:27PM	'Forging Their Own Path': The Top Stocks Owned...
2019-11-01	03:16PM	The Best Fidelity Funds for 401(k) Retirement ...
2019-11-01	03:14PM	UPDATE 1-Apple asks U.S. to waive tariffs on C...
2019-11-01	02:54PM	Buy Apple Stock at New Highs Heading into the ...
2019-11-01	02:36PM	Google Invades Apple's Wearables Territory Wit...
2019-11-01	02:31PM	Earnings Reports for the Week of Nov. 4-8 (PTO...
2019-11-01	02:31PM	PowerSchool searching for more office space as...
2019-11-01	02:25PM	Sources: Apple scouting for a larger office in...
2019-11-01	02:25PM	Apple asks U.S. to waive tariffs on Chinese-ma...

Figure 1: Data Extracted for AAPL (Apple) stock

2. Extracting data from Webpages

   Relevant data is extracted from the saved webpages files using Beautiful Soup library and tabulated under ticker, date, time, headline. In Figure 1, data for Apple (AAPL) stock for the date 1st November 2019 is extracted and tabulated so to be prepared for pre-processing.

3. Cleaning Data

   Data acquired is cleaned by removing duplicates and weekend data followed by text processing to get improve accuracy.

4. Use of VADER Library and assigning Sentiment Values

   Python tool VADER uses a lexicon-based approach to determining sentiment values of a sentence. Use of VADER is done to get a general sentiment analysis in addition to it, certain keywords like crushes, beats, misses, trouble, falls with their respective sentiment values are also updated as lexicons so to allow the analyser to understand these words in their financial sense. In Figure 2, the addition of new words with their relevant sentiment to the lexicon is shown so to allow the tool to understand these words in their financial sense.

```
new_words = {
    'crushes': 10,
    'beats': 5,
    'misses': -5,
    'trouble': -10,
    'falls': -100,
}
vader = SentimentIntensityAnalyzer()
vader.lexicon.update(new_words)
```

Figure 2: Addition of New Words to the Lexicon

5. Summarizing and Visualizing Data
   As discussed, the data is classified on the basis of what percentage of the sentence has a positive, negative, and neutral sentiment, and then generates a compound score on the basis of its normalized aggregate for the sentence. The sentiment score of each headline is tabulated and summarized so that it can be utilized to evaluate market sentiment regarding the stock.

V. Results

Data generated from sentiment analysis of news headlines is classified under negative, neutral, and positive sentiment. This is generated on the basis of the percentage of the sentence is positive, negative, or neutral. These three values sum up to one for each sentence. The compound sentiment score is obtained by normalizing lexicon ratings.

Sentiment values are tabulated for Apple (AAPL) stock for a single day in Figure 3, which has also been visualized in figure 4 for positive(green), negative (red), and neutral (yellow) sentiments. It can be observed that the negative values (neg), neutral (neu), and positive (pos) sum up to one.

date	compound	neg	neu	pos
2019-11-01	0.8555	0.000	0.567	0.433
2019-11-01	0.0000	0.000	1.000	0.000
2019-11-01	0.3252	0.000	0.811	0.189
2019-11-01	0.0000	0.000	1.000	0.000
2019-11-01	0.0000	0.000	1.000	0.000
2019-11-01	-0.2732	0.123	0.877	0.000
2019-11-01	0.0000	0.000	1.000	0.000
2019-11-01	0.1779	0.000	0.892	0.108
2019-11-01	0.2023	0.000	0.795	0.205
2019-11-01	0.2023	0.000	0.833	0.167
2019-11-01	0.6369	0.000	0.625	0.375
2019-11-01	0.0000	0.000	1.000	0.000
2019-11-01	0.4019	0.000	0.803	0.197
2019-11-01	0.0000	0.000	1.000	0.000

Figure 3: Sentiment Analysis for relevant News Headlines [Apple (AAPL) Stock]

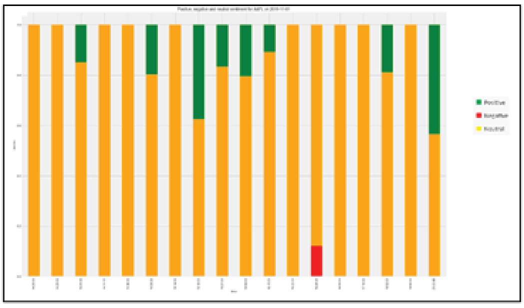

Figure 4: Visualization of Sentiment Score of Headlines for AAPL on 2019-11-01

In Figure 5, the compound score for each day is summarised and displayed for the user to evaluate market sentiment for the stock. In the graph, red is for tesla and blue is for apple. Values below 0 indicate a negative sentiment regarding the stock and is an indicator indicating a possible fall in stock prices

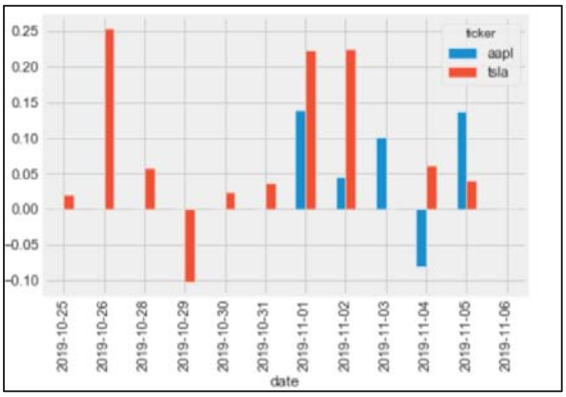

Figure 5: News Headline Sentiment Compound Score for TSLA (Tesla) and AALP (Apple) Stocks

The stock prices for Tesla (TSLA) was sourced from yahoo finance for the same time period and depicted graphically in Figure 6. A strong correlation is observed between change in sentiment which correlates to a change in stock prices. In particular, a positive change correlates to a rise in stock price and a negative change correlates to a fall in prices. For instance, we studied the case for price changes in tesla. The positive change of sentiment on 26-10-2019 is reflected with an increase in prices on 28-10-2019 i.e. Monday which is the first day of trading in the week. Similarly, a negative change in sentiment on 29-10-2019 reflects a drop of stock values on 29-10-2019 which later stabilizes on 30-10-2019 with a more positive market sentiment. A case study of variation of stock prices for the same duration was conducted and an efficacious correlation was observed between market sentiment and unit price for the equity. It is important to remember here that there are other influencing factors such as the bandwagon effect, overvaluation correction, effect of related markets, economic situation of the country where the company is incorporated in, global economic situation and so on, that may as occasionally cause deviation from the correlation between market sentiment and stock prices.

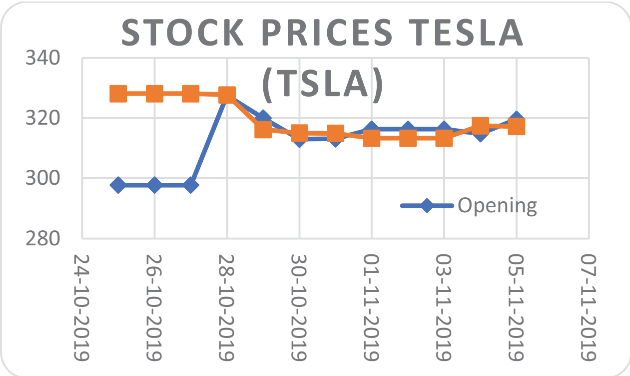

Figure 6: Stock Price graph for tesla

VI. CONCLUSION

In this study, a methodology to interpret market sentiment is developed. The relationships between the volume of news, polarity, and subjectivity of news referring to stock to allow generation for market sentiment results are analysed by using VADER along with user-defined Lexicons, which can be utilized to ascertain change. The volatility of equity can also be observed by the frequency of sentiment change. The benefit of using such a tool alongside need-based addition to the lexicon is that it allows fast and versatile analysis of data thus allowing the use of it on livestream data. This paper does not use live-streamed data so to prevent unnecessary load on the website server, but the same method can also be utilized with live-streamed data. Also, with minor additions and modifications to the lexicon this methodology can be adopted for the analysis of specific market sector say on basis of product viz pharmacy, healthcare, e-commerce, etc. or on basis of capital like large-cap, mid-cap or small-cap to forecast the rise or decline of its stock value.

REFERENCES

[1] O. C. Sert, S. D. Şahin, T. Özyer, and R. Alhajj, "Analysis and prediction in sparse and high dimensional text data: The case of Dow Jones stock market," *Phys. A Stat. Mech. its Appl.*, vol. 545, p. 123752, 2020, doi: 10.1016/j.physa.2019.123752.

[2] J. Griffith, M. Najand, and J. Shen, "Emotions in the Stock Market," *J. Behav. Financ.*, vol. 21, no. 1, pp. 42–56, Jan. 2020, doi: 10.1080/15427560.2019.1588275.

[3] D. Reinsel, J. Gantz, and J. Rydning, "The Digitization of the World - From Edge to Core," *IDC White Pap.*, no. November, p. US44413318, 2018, [Online]. Available: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf.

[4] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011, doi: https://doi.org/10.1016/j.jocs.2010.12.007.

[5] Q. Li, T. Wang, P. Li, L. Liu, Q. Gong, and Y. Chen, "The effect of news and public mood on stock movements," *Inf. Sci. (Ny).*, vol. 278, pp. 826–840, 2014, doi: 10.1016/j.ins.2014.03.096.

[6] J. H. Yu, J. Kang, and S. Park, "Information availability and return volatility in the bitcoin Market: Analyzing differences of user opinion and interest," *Inf. Process. Manag.*, vol. 56, no. 3, pp. 721–732, 2019, doi: https://doi.org/10.1016/j.ipm.2018.12.002.

[7] A. Bhardwaj, Y. Narayan, Vanraj, Pawan, and M. Dutta, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty," *Procedia Comput. Sci.*, vol. 70, pp. 85–91, 2015, doi: 10.1016/j.procs.2015.10.043.

[8] D. Shah, H. Isah, and F. Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market," *Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018*, pp. 4705–4708, 2019, doi: 10.1109/BigData.2018.8621884.

[9] M. S. Usha and M. Indra Devi, "Analysis of sentiments using unsupervised learning techniques," *2013 Int. Conf. Inf. Commun. Embed. Syst. ICICES 2013*, pp. 241–245, 2013, doi: 10.1109/ICICES.2013.6508203.

[10] D. K. Kirange and R. R. Deshmukh, "Sentiment Analysis of News Headlines for Stock Price Prediction," vol. 5, no. Iii, pp. 1–5, 2016, doi: 10.13140/RG.2.1.4606.3765.

[11] S. Kalra and J. S. Prasad, "Efficacy of News Sentiment for Stock Market Prediction," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 491–496, 2019, doi: 10.1109/COMITCon.2019.8862265.

[12] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Inf. Process. Manag.*, vol. 57, no. 5, p. 102212, 2020, doi: 10.1016/j.ipm.2020.102212.

[13] 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, doi: 10.1016/j.eswa.2015.07.052.

[14] A. E. Khedr, S. E. Salama, and N. Yaseen, "Predicting stock market behavior using data mining technique and news sentiment analysis," *Int. J. Intell. Syst. Appl.*, vol. 9, no. 7, pp. 22–30, 2017, doi: 10.5815/ijisa.2017.07.03.

[15] V. Kalyanaraman, S. Kazi, R. Tondulkar, and S. Oswal, "Sentiment analysis on news articles for stocks," *Proc. - Asia Model. Symp. 2014 8th Asia Int. Conf. Math. Model. Comput. Simulation, AMS 2014*, pp. 10–15, 2014, doi: 10.1109/AMS.2014.14.

[16] A. Mittal and A. Goel, "Stock prediction using twitter sentiment analysis. Working paper," *Standford Univ. CS229*, no. June, 2012, doi: 10.1109/WI-IAT.2013.48.

[17] F. Audrino, F. Sigrist, and D. Ballinari, "The impact of sentiment and attention measures on stock market volatility," *Int. J. Forecast.*, vol. 36, no. 2, pp. 334–357, 2020, doi: 10.1016/j.ijforecast.2019.05.010.

[18] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," *Proc. - 5th IEEE Int. Conf. Big Data Serv. Appl. Big Data Service*, pp. 205–208, 2019, doi: 10.1109/BigDataService.2019.00035.