Analysis of Financial News Sentiment

Sandeep Yadav[1]

[1]*Assistant Professor, Department of Computer Application,*

Buddha Institute of Management, GIDA, Gorakhpur, Uttar Pradesh, India

sandeepyadav1616@bimgkp.ac.in

Abstract—

Sentiment analysis is a branch of computational semantics and data mining. It alludes to the understanding of data from sources with a lot of emotion, including news, social media, reviews, and so on. Nowadays, with data growing in volume and being essential to every organization, manual data analysis is no longer practical in our fast-paced environment. As a result, data mining and artificial intelligence methods must be used. Among a number of factors, a company's profits or losses are a major one that influences how much its stock price fluctuates. Since the majority of traders rely on news sources for information, news is a key determinant of the stock market movement.

This study illustrates how changes in stock market prices are impacted by the classification of emotions.By applying sentiment analysis to some of the most liquid equities utilizing the VADER (Valence Aware Dictionary and Sentiment Reasoner) technology, it produces investing insights.

Keywords—

Financial news stories, sentiment analysis, VADER, data mining, the stock market, and liquid stocks.

INTRODUCTION

One important aspect of a country's economy is its stock market. It plays a vital part in a nation's business and trade expansion. Consequently, it significantly affects the national economy and, with the rise of globalization, now has an international impact. As a result, everyone involved in the stock market, whether directly or indirectly, keeps a careful eye on it.

Forecasting changes in the market can be aided by sentiment analysis of news, which is unstructured data [1,2]. Indicators of market sentiment are vital in figuring out the impact of fairness in the foreseeable upcoming. For instance, the price of a stock will either keep rising or stay the same if there is a favorable sentiment regarding it. However, declining mood may be interpreted as a sign that prices are about to decline.

If numbers are to be acknowledged, we must handle massive amounts of data today, and innovation is growing at an exponential rate. According to a current projection, the world's total data would grow at a compound annual growth rate (CAGR) of 61% from 33 zettabytesin 2018 to 175 zettabytes by 2025. [3] Higher news volumes signal an occurrence that could cause the price of assets to significantly increase or decrease.

The primary factor influencing news headline selection is the fact that 70–80% of organizational data is unstructured. There is some uncertainty that the measure of unstructured data will continue to evolve regardless of the true pace. Many associations now find that it is difficult to coordinate such massive amounts of data with huge business frameworks using poorly designed data management devices due to the volume of information testing. Additionally, it takes into consideration risks that cannot be attributed to the market.

Consequently, there is still a significant volume of unstructured data unexplored.

In this project, For news headlines and descriptions, Finwiz is utilized. Finwiz indexes content from a number of well-known international sources. It has been a crucial component in providing its users with timely and frequent news. The NLTK (Natural Language Toolkit) package's Word-based sentiment analysis instrument will make use of The Sentiment Reasoner and Valence Aware Dictionary, or VADER library. This tool is used since it is quick and doesn't require any training data; therefore it may even be used with online streaming data.

I. RELATED WORKS

The effectiveness of market sentiments in predicting stock trends has been confirmed by numerous researchers [4-5], and this also holds true for the Bitcoin exchange market [6]. Market sentiment has been analyzed using a variety of methodologies. They fall into two general categories: Lexicon-based methods and machine learning [7]. There are two methods: dictionary-based and corpus-based subcategories of lexicon-based methods. Learning by machine can be further divided into supervised and unsupervised methods, such as decision tree, rule-based, and probabilistic classification.

Here is a discussion of a few models created in this manner.

According to Dev Shah et al. [8], a dictionary-based approach is recommended. It converts text corpora to numerical vectors using a Python package named pattern. This library calculates the frequency of both positive and negative phrases to produce a sentiment score aggregates them according to frequency. Other industries can also use the strategy. The problem with the model is that the emotion each token's score is produced without being weighted, and a straightforward accumulation of scores can result in notable deviations from the sentiment of the market.

According to M.S. Usha et al. [9], unsupervised learning can be used to generate market sentiment insight. Their suggested model can simultaneously identify topic and sentiment and is based on the Gibbs sampling approach. Multiple data sets, or information from different kinds of product reviews that were crawled from an online store, are used in the article. As a result, the data contains a range of sentiment depending on the topic. Because it uses an unsupervised learning approach, it is very portable. The inability of the suggested paradigm to identify neutral viewpoints is a disadvantage.

To generate sentimental values that are be utilized to obtain understanding of the shift in the market trends, D.K. Kirange et al. [10] suggest a forecasting model derived from the emotion news classification. Naïve Bayes, SVM, and KNN are among the techniques used; KNN has the highest accuracy—this model collects data from a variety of well-known news sources over an extended a while, and then uses emotion categorization to determine sentiment polarity. The writers next try to identify a correlation by comparing their newly acquired knowledge with previous stock market data. The results show that the SVM model seems to have the best precision in predicting change.

In order to create a forecasting model that can forecast stock prices with an accuracy of between 65% and 92%, SnehKalra et al. [11] suggest using Sentiment analysis using the Naïve Bayes classifier in combination with past information of nearby news dates. The model makes use of the pre-processed stock news data collection, after which naïve Bayes classification is used to analyze each article's sentiment. By analyzing the news and numerical data using a variety of machine learning techniques, this is combined with the stock variance around neighboring dates that is retrieved from Yahoo Finance to produce investing insights. Regarding the identical data set, KNN prediction produced the highest level of accuracy. This model's limitation is the restricted supply of data it uses, which may lead to variations from the actual outcomes.

A alternative tack has been adopted by Xiadong Li et al. [12]. A layered deep learning model was used to generate market sentiment in a suggested stock prediction system. The learned sentiment was then used to a neural network that is fully coupled to produce market forecasts. Their algorithm generates an output that is subsequently fed into the neural network by analyzing sentiment using dictionaries and technical stock indicators.

Other models that have been suggested include the following: Various time series analysis models are used [18], Ngram and Naïve Bayes algorithms are used for sentiment analysis [14], sentiment analysis using dictionaries [15], classifying moods and calculating daily scores are used to map sentiment scores [16–17], and particular sub-modules based on needs are designed depending on the natural language processing module's current modules to analyze sentiment [1,13].

II. METHODOLOGY ADOPTED

Sentiment analysis is performed on Finwiz data scrapes. Real-time headlines and stock updates from respectable publications, including the Bloomberg, Yahoo Finance, the Wall Street Journal, the Financial Times, and others, are available on Finwiz. The choice of data source is crucial to avoiding the inclusion of inaccurate data because the veracity and relevance of the news are so important. An inclusion of inconspicuous but truthful headline may result in the emotion analyzer producing an incorrect sentiment score. The sentiment analyzer may produce an unrelated emotion score determined by the tokens "Tesla" and "crash" in response to a potential headline such as "Tesla's CEO Elon Musk crashes a party," which could result in erroneous predictions.

After classifying the data according to the proportion of the sentence that is positive, negative, or neutral, a compound score is produced using the sentence's normalized aggregate. In order to assess market sentiment toward the stock, the sentiment score of each headline is tallied and condensed.
The Python utility VADER determines a sentence's sentiment values using a lexicon-based method. In order to enable the analyst to interpret these terms in their pecuniary sense, this is employed in conjunction with sentiment values that are specifically allocated to keywords that are frequently found in news headlines that refer to stocks, such as Falls, Crushes, Plunges, etc.

III. IMPLEMENTATION

1. Using HTML Files to get data

The user adds HTML files of pertinent stocks to the dataset they have gathered. To make sure the servers are not overloaded, this is done by hand.

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

Figure 1: Data Extracted for Apple Stock (AAPL)

2. Data extraction from webpages

The Beautiful Soup library is used to extract pertinent data from the stored webpage files, which is then tabulated under ticker, date, time, and headline. To get ready for preprocessing, data for Apple (AAPL) stock as of November 1, 2019, is retrieved and tabulated in Figure 1.

3. Cleaning Data

After weekend data and duplicates are eliminated, the obtained data is text-processed to increase accuracy.

4. Using VADER Library and Sentiment Assignment

The Python utility VADER determines a sentence's sentiment values using a lexicon-based method. In addition to using VADER to obtain a general sentiment analysis, specific keywords such as crushes, beats, misses, difficulty, and falls are also updated as lexicons with their corresponding sentiment values so that the analyst can comprehend these terms in their financial context. To enable the tool to comprehend new terms in their fiscal sense, Figure 2 displays the inclusion of fresh vocabulary to the lexicon along with their pertinent mood.

```
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. Data Visualization and Summarization

The data is categorized according to the proportion of the sentence that is positive, negative, or neutral, as was previously mentioned. A compound score is then produced using the sentence's normalized aggregate. To assess the sentiment of the market toward the stock, the sentiment score of each headline is tallied and condensed.

IV. Results

Sentiment analysis of news headlines yields data that can be categorized as either positive, neutral, or negative. The percentage of the language that is positive, negative, or neutral is used to generate this. For every statement, the sum of these three values equals one. Lexicon ratings are normalized to provide the compound sentiment score.

Figure 3 shows the sentiment values for Apple (AAPL) stock for a single day. Figure 4 shows the sentiments yellow for neutral, red for negative, and green for positive feelings. It is evident that the total of the positive (pos), neutral (neu), and negative (neg) values is 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 pertinent news stories [stock of Apple (AAPL)]

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

Figure 5 summarizes and presents the daily compound score so that the user can assess the stock's market sentiment. Apple is shown in blue on the graph, and Tesla is shown in red. A negative feeling about the stock is shown by values less than zero, which also suggests that stock prices, may decline.

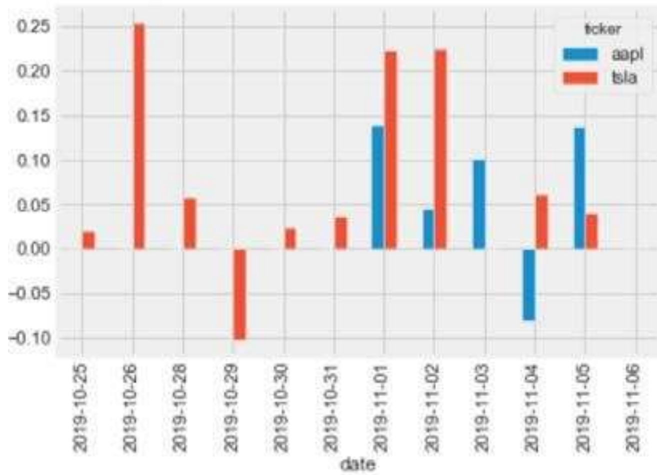

Figure 5: Headline News The sentiment compound score for the stocks of Tesla (TSLA) and Apple (AALP)

Figure 6 shows a graphic representation of the Tesla (TSLA) stock prices for the same time period that were obtained from Yahoo Finance. Changes in sentiment are found to be strongly correlated with changes in stock prices. Specifically, an increase in the stock price is associated with a positive change, and a decrease in prices is associated with a negative change. For example, we looked at the Tesla pricing change case. Prices increased on Monday, October 28, 2019, the first day of trading for the week, reflecting the favorable shift in attitude on October 26, 2019. Comparably, a decline in stock values on October 29, 2019, which is followed by a stabilization on October 30, 2019, with a more optimistic market attitude, is reflected in a negative shift in sentiment on that date. An effective association between market mood and the equity's unit price was found in a case study of stock price fluctuations over the same time period. It is crucial to keep in mind that there are additional influencing factors that can occasionally lead to a divergence from the relationship between stock prices and market sentiment, including Overvaluation correction and the bandwagon effect, the impact of related markets, the economic climate of the nation in which the company is incorporated, the global economic climate, and so forth.

Figure 6: Stock Price graph for tesla

V. CONCLUSION

A mechanism for interpreting market mood is developed in this study. Using VADER and user-defined Lexicons, which can be used to determine change, the relationships between the volume, polarity, and subjectivity of news pertaining to stocks are examined in order to get results for market sentiment. The frequency of sentiment shifts is another indicator of equity volatility. Using such a tool in conjunction with need-based vocabulary expansion has the advantage of enabling quick and flexible data analysis, which makes it applicable to livestream data. The same technique can be applied to live-streamed data, although it is not used in this paper to save needless strain on the website server. Additionally, with a few tweaks to the terminology, this methodology can be used to analyze a particular market sector, such as e-commerce, healthcare, pharmacy, etc., based on product, or on capital, such as large-, mid-, or small-cap companies, to predict the growth or decline of their stock value.

VI. REFFERENCES

[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/ourstory/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.