VIVA Institute of Technology

10th National Conference on Role of Engineers in Nation Building - 2022 (NCRENB-2022)



# VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION

ANNUAL RESEARCH JOURNAL ISSN(ONLINE): 2581-7280

# "SENTIMENT ANALYSIS OF STOCK NEWS USING NLTK"

Hardik Raut<sup>1</sup>, Rohini Hodge<sup>2</sup>, Harshal Bhoir<sup>3</sup>, Reema Yedge<sup>4</sup>

(Electronics & Telecommunication, Viva Institute of Technology, India)

**Abstract:** The popular theory about stock prediction is Efficient Market Hypothesis. Due to its failure many research has been done in area of stocks prediction. In this research non quantifiable data such as news about financial activities of a company is taken to predict its future stock trend with the sentiment of the news. As news have a strong impact on stock values, this study aims at relationship of company news with stock trend. For this, we created a NLTK model which depicts polarities of news article as negative or positive. To evaluate various aspects of the proposed model experiments were conducted. The NLTK model is accurate and this tool is capable of determining the emotional values without neutral sections. By comparing the results of experiments with the movement of stock market values in the same time periods, we can establish the relation between change occurred in the stock values with sentiment analysis of economic news headlines. The model has prediction accuracy of more than 80% and in comparison, with 50% of accuracy with news random labelling; thereby increasing the model accuracy by 30%.

Keywords - NLTK, VADER LEXICON, BERT, Sentiment analysis, Stocks

# **1** Introduction

The goal is to develop and use a model to sentiment prediction by making connections between news articles and marking them with negative or positive sentiments. There square measure several opportunities currently to perform sentiment analyses, for instance external services that square measure nearly fully able to use it during a given context wherever it's required like TextBlob. additionally, there square measure choices that enable us to make our own models, train them supported our own knowledge. Sentiment analysis with BERT is one among the foremost powerful tools that we are able to use, however we are able to conjointly produce a repeated Neural Network (RNN) furthermore or use the VADER Lexicon with NLTK tool and SentimentIntensityAnalyzer[1].

The share/stock market is one among the foremost vital economic participants. many folks try and interpret and outline the various share market movements in some ways. during this article, we have a tendency to use totally different tool to the sentiments analysis, particularly absorption on the economic or financial news, however in terms of news, absorption solely on the headline of economic news. In today's communication and news utilization, the headlines of varied articles play an excellent additional vital role than before. Now, we have a tendency to use sentiment analysis on the headlines of a selected company or corporations to see the consequences of the headlines to the stock market. The question arises what proportion result has the economic headline while not the economic news whole context, if it's any measurable result in any respect. we've got found that it extremely has. Thus, we have a tendency to outline the various impacts and their proper significance with an awfully specific and distinctive new approach.

#### VIVA Institute of Technology

#### 10th National Conference on Role of Engineers in Nation Building – 2022 (NCRENB-2022)

Data is a crucial pillar of research. Primarily the headlines of economic news square measure required, what we have a tendency to use for sentiment analysis. Secondary, totally different stock market knowledge are required supported corporations. There square measure several potentialities for knowledge assortment and analysis from 'conventional' dictionary-based performed by humans to 'more serious' neural network that verify the sign of the headlines of every economic news and label with applicable emotional polarity. within the case of stock market knowledge, various tools square measure on the market to get stock market knowledge which might be even company-specific that is vital to us. In each case, we have a tendency to work with the foremost up-to-date knowledge as doable, supported the data provided by the businesses. Both, the headlines of the economic news and stock price knowledge square measure associated with the fundamental measure that specified by the news. So, the result of the given sentiment analysis and therefore the vary of stock market knowledge are applicable.

The analysis is often separated to consecutive sections, first collecting headlines of financial economic news supported corporations and collect stock market knowledge as per the timestamps of the given economic news headlines. Then prepare this knowledge and apply sentiment analysis tool NLTK and VADER Lexicon. Manage this knowledge and compare the stock market knowledge and emotional knowledge with image and rationalization. gift however the headlines of economic news will have an effect on totally different stock market changes and therefore the public[9].

#### **2 Literature Review**

Yu et al [2] demonstrated a text mining based mostly framework to work out the sentiment of articles and illustrate its impact on energy demand. News sentiment is quantified as a time series and compared with fluctuations in energy demand and costs.

Khedr and Yaseen[6] aims at having an efficient model to predict stock market further trends with tiny error quantitative relation and improve the prediction accuracy. wherever this prediction model is predicted on sentiment analysis and historical stock market costs, worked with K-NN and naïve mathematician rule to earn the ultimate results. we are able to separate the model for 2 stages. the primary stage is to see the news polarity is positive or negative mistreatment naïve mathematician rule, the second stage incorporates the output of the primary stage as input with the processed historical numeric knowledge to predict the long run stock trend of K-NN rule.

J. Bean [3] used keywords tagging regarding airlines satisfaction on Twitter feeds to attain them for polarity and sentiment. this may offer a fast plan of the sentiment prevailing regarding airlines and their client satisfaction ratings. we've got used the sentiment detection algorithmic program based mostly on this analysis.

Wang et al. [7] introduced a public sentiment analysis throughout the irruption that is ready to provides perceptive info in creating applicable response of public health. They analyse the Sina Weibo standard social media website posts of China, 'wherever the unsupervised BERT model is adopted to classify sentiment classes (positive, neutral, and negative) and TF-IDF (term frequency-inverse document frequency) model is employed to summarize the topics of posts'. Analysing social media posts with negative sentiment may contribute to understanding the experiences and offers examples for different countries. The analysis offer insights on the changing of social sentiment over time and therefore the topic themes connected to negative sentiment on the social media sites. Sizable accuracy was achieved with TF-IDF topic extraction model and BERT classification model.

SmartSA is a lexicon-based sentiment analysis for social media. It integrates methods to capture discourse polarity from 2 ways that, the interaction of terms with their matter neighbourhood and text genre like native and international context. They conjointly introduce associate approach to breed a general lexicon, with genre-specific jargons and sentiment. The results from numerous social media show that these methods of native and international contexts considerably improve sentiment classification, and are complementary together.

Kalyani, Bharathi & Rao[4] in their analysis, used supervised machine learning for classification of headlines and extra text mining techniques to look at news polarity. The news article with its polarity score and text reborn

# VIVA-Tech International Journal for Research and Innovation ISSN(Online):2581-7280

#### VIVA Institute of Technology

#### 10th National Conference on Role of Engineers in Nation Building – 2022 (NCRENB-2022)

to tf-idf vector area area unit fed to the classifier. 3 completely different classification algorithms (Support Vector Machines "SVM", Naïve Bayes and Random Forest) area unit enforced to analyze and enhance classification accuracy. Results of all 3 algorithms area unit compared supported preciseness, recall, accuracy, and different model analysis techniques. once evaluating the results of all classifiers, the SVM classifier performs satisfactorily for unknown information. The Random Forest conjointly showed higher results when put next to the Naïve Bayes algorithmic program. Finally, a graph of link between news articles and stock information is plotted.

Streaming knowledge encourage be an upscale supply of information analysis wherever data square measure collected in Realtime[4]. The most important characteristics of such knowledge being its accessibility and accessibility, facilitate in correct analysis and prediction. Robert et al.[8] show associate analysis that has been created for creating money selections like stock market prediction, to predict the potential costs of a company's stock mistreatment twitter knowledge.

### **3 Methodology**

As earlier mentioned, the most goal within the economic news headlines is to use the foremost upto- date knowledge. All knowledge assortment and management are machine-driven. there's associate choice to the user to mention the portal as a supply to manage the news. we have a tendency to used knowledge from 'finviz.com' for our analyses. Before grouping the info, it's doable to enter the stock market names of the businesses wherever we might wish to collect the info of current economic events for analysis. It is possible to specify over one company by listing as parameter. The task takes care of managing the suitable timestamps (news publication time) and separating the news supported the businesses and build a csv backup file. This freshly compiled knowledge is employed by the appliance for any analysis (as a part of sentiment analysis, comparisons, and different potentialities.) it's vital to say that news timestamps play a task in collection extra stock market knowledge that the analyses come about within the same fundamental measure. Thus, these economic news headlines outline the interval for later stock market knowledge assortment separated for corporations.

#### **4** Sentiment analysis

In the case of sentiment analysis, the headline of the economic news from every company is labelled to what sentiment price it carries, and therefore the polarity price is additionally indicated. With the assistance of those knowledge, we are able to build analyses and predict the result.

	А	В	С	D
1	Ticker	Date	Time	Headline
2	AMZN	Jan-30-22	08:15AM	Sea Limited Hits Triple Digit Growth Rates Despite Inflation
3	AMZN	Jan-30-22	07:20AM	With MercadoLibre Revenue Up 89%, Is Now the Time to Invest?
4	AMZN	Jan-30-22	06:50AM	Buy This Stock in 2022 and You Could Be Rich by 2032
5	AMZN	Jan-30-22	06:30AM	Shopify Stock: Bear vs. Bull
6	TSLA	Jan-30-22	11:18AM	Dow Jones Futures Loom As Market Rally Still Must Take This Step; 5 Key Earnings
7	TSLA	Jan-30-22	09:18AM	How Tesla's Record Earnings Impact Lucid and Rivian
8	TSLA	Jan-30-22	08:00AM	Tesla Allows Customers to Purchase Merch with Dogecoin
9	TSLA	Jan-30-22	06:50AM	Buy This Stock in 2022 and You Could Be Rich by 2032
10	GOOG	Jan-30-22	02:58AM	Wall Street Week Ahead Earnings: Alphabet, PayPal, Exxon Mobil, Meta, Qualcomm and Amazon in Focus
11	GOOG	Jan-29-22	08:49AM	Where Will Alphabet Be in 10 Years?
12	GOOG	Jan-29-22	07:00AM	3 Autonomous Driving Technologies Investors Must Know About in 2022
13	GOOG	Jan-29-22	05:36AM	Got \$3,000? 5 Unstoppable Stocks to Buy as the Market Corrects Lower
14	GOOG	Jan-29-22	02:30AM	Microsoft's Activision Buy Could Take Gaming M&A to Next Level

#### Figure 1. Part from the economic news headlines dataframe.

#### 4.1. NLTK -- VADER lexicon

# VIVA-Tech International Journal for Research and Innovation ISSN(Online):2581-7280

#### VIVA Institute of Technology

### 10th National Conference on Role of Engineers in Nation Building – 2022 (NCRENB-2022)

NLTK is abbreviation for Language Toolkit. This toolkit is one among the foremost powerful information processing libraries that contains packages to create machines perceive human language associated reply to that with an applicable response. Our main focus is to analyse the tendency of sentiments using sentiments Intensity analyzer. The polarity price of the sentence's scales between -1 and one rather like within the TextBlob. The method for labelling info (positive, negative or neutral) and previous tool which was used are sort of similar. we have a tendency to use VADER Lexicon during this section. VADER (Valence Aware wordbook and sEntiment Reasoner) uses the tool specifically attuned to sentiments expressed in social media, and works well on texts from different domains based on lexicon and rule-based sentiment analysis tool that's specifically.

Figure 2 shows that the score of neutral value dominates all the cases of the sentiments result company wise. Significant values were neutral as obtained from the analysis of economic news headlines. Neutral values This level of neutral values has effect on analyses and comparisons of following changes in stock market. The results from NLTK compared to TextBlob, shows the significant reduction in neutral values and so the results are expected to be correct and realistic results with very less neutral values. In 51.50 % of the overall result's neutral additionally to 50 % positive and 17 % negative. From the negative and positive groups, more positive are seen to dominate, but the presence of neutral scores can produce uncertainty in the results.

The Figure 3 shows the outcome graph of scores separated by days interval. The results combined to provides us a normalized score of the negative or positive the news for the company was. The polarity cannot be neutral in total as the news of one polarity tries to move the neutral values some direction depending upon the polarity of news. Thus, the subsequent figure is made, wherever on below of zero suggests that the negative section and top of zero suggests that the positive section.

	А	В	C	D	E	F	G	Н
1	ticker	date	time	headline	neg	neu	pos	compound
2	AMZN	Jan-30-22	08:15AM	Sea Limited Hits Triple Digit Growth Rates Despite Inflation	0.165	0.609	0.226	0.1779
3	AMZN	Jan-30-22	07:20AM	With MercadoLibre Revenue Up 89%, Is Now the Time to Inv	0	1	0	0
4	AMZN	Jan-30-22	06:50AM	Buy This Stock in 2022 and You Could Be Rich by 2032	0	0.753	0.247	0.5574
5	TSLA	Jan-30-22	06:00AM	Are Tesla's Promises of Self-Driving Cars Risky for Investors?	0.158	0.614	0.228	0.2023
6	TSLA	Jan-30-22	05:57AM	Elon Musk Wants to Make Tesla Boring And Be a Car Salesma	0.187	0.813	0	-0.3182
7	TSLA	Jan-29-22	11:45AM	2 Reasons Elon Musk Can't Save Dogecoin	0.345	0.655	0	-0.3875
8	GOOG	Jan-30-22	11:18AM	Dow Jones Futures Loom As Market Rally Still Must Take This	0.128	0.872	0	-0.2263
9	GOOG	Jan-29-22	05:36AM	Got \$3,000? 5 Unstoppable Stocks to Buy as the Market Corr	0.308	0.692	0	-0.4588
10	GOOG	Jan-28-22	04:15PM	How Spotify's Neil Young fight signals platform's 'grown up' t	0.206	0.794	0	-0.3818
11	GOOG	Jan-28-22	04:12PM	YouTube takes down House Republicans campaign ad claimi	0.172	0.828	0	-0.3612

Figure 2. Results of sentiment analysis using NLTK and VADER as % positive, negative or neutral





Figure 3. Result of NLTK and VADER Lexicon Analysis days wise.

## Conclusion

In this study, we classified the news headlines using sentiment analysis tools which is capable of analysing emotions and classify different companies' news headlines to examine the effect of sentiments on the price of stocks in stock exchange. The emotions were segregated into the neutral, negative, positive classes. Neutral classes in NLTK VADER lexicon were very less as compared to other tools used for sentiment analysis as seen in the literature. Neutral values tend to produce variation in results when compared to the actual behaviour of stocks prices. The emotional effect of the news on stock price is significant and so the results of sentiments analysis is comparable to the exchange values for the given script. The positive values indicated that there might be increase in the price of stocks, negative value indicated decrease in stock price, the neutral values have no effect on stock price. The NLTK Vader Lexicon produces results, with very less model size. Since stock price variation is difficult to establish. Future work includes use of more sentiment analysis tool and natural language processing to predict stock behaviour and match the prediction real prices of stock from the stock exchange.

## References

[1] Anurag Nagar, Michael Hahsler, Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams, IPCSIT vol. XX (2012) IACSIT Press, Singapore

[2] W.B. Yu, B.R. Lea, and B. Guruswamy, A Theoretic Framework Integrating Text Mining and Energy Demand Forecasting, International Journal of Electronic Business Management. 2011, 5(3): 211-224

[3] J. Bean, R by example: Mining Twitter for consumer attitudes towards airlines, In Boston Predictive Analytics Meetup Presentation, 2011

#### VIVA Institute of Technology

## 10th National Conference on Role of Engineers in Nation Building – 2022 (NCRENB-2022)

[4] Yauheniya Shynkevich, T.M. McGinnity, Sonya Coleman, Ammar Belatreche, Predicting Stock Price Movements Based on Different Categories of News Articles, 2015 IEEE Symposium Series on Computational Intelligence

[5] P. Hofmarcher, S. Theussl, and K. Hornik, Do Media Sentiments Reflect Economic Indices? Chinese Business Review. 2011, 10(7): 487-492

[6] Khedr, A. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. International Journal of Intelligent Systems and Applications, 9(7), 22. https://doi.org/10.5815/ijisa

[7] Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: Negative sentiment analysis on social media in China via bert model. IEEE Access, 8, 138162–138169. https://doi.org/10.1109/Access.6287639

[8] Robert P. Schumaker, Yulei Zhang, Chun-Neng Huang, Sentiment Analysis of Financial News Articles

[9] Billah, M., Waheed, S., & Hanifa, A (2016, December 8–10). Stock market prediction using an improved training algorithm of neural network. 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh (pp. 1–4). IEEE. https://doi.org/10.1109/ICECTE.2016.7879611