NU

NIIT UNIVERSITY

THE UNIVERSITY OF THE FUTURE

Industry-linked I Technology-based I Research-driven I Seamless

Group Project on

Natural Language Processing (CS 491)

Academic Year- 2020-21

On

EXTRACT STOCK SENTIMENTS FROM NEWS HEADLINES

Submitted by

Group No. 8

1. Astha Kumar	(BT18GCS016)	CSE	C4
2. Shreya Chauhan	(BT18GCS154)	CSE	C4
3. Muskan Goel	(BT18GCS188)	CSE	C3

Table of Contents

Introduction

Gone are the days when it used to take a lot of time for the financial news to spread via newspapers, radio or any oral communication. Nowadays, in this era of the internet, it just takes seconds. And we don't realise the impact of this everyday news on the stock market. Stock market is the bone of this rapidly emerging economy and our country's growth is somewhere bound to the fact of stock market prediction. Considering the lack of awareness and knowledge among the people, stock market prediction techniques will play a major role in choosing the right path, bringing them to the market and also retain the existing investors.

The arrival of news at every point of time changes the sentiment towards the company. And these days due to the bliss of the internet, the traders and investors have constant access to the trending news which keeps them updated and shapes their sentiments towards the particular piece of news or the trading company helping them decide to invest in that company.

News has always taken the lead in providing a clear crystal perception about a market investment. With the growing time, news has been increasing voluminously, which in result makes it impossible for an investor or a group of investors to decide or find out the best relevant part of news from the large chunk of news available. But we know that it's important to make the right choice about an investment timely in order to extract maximum profit out of the investing plan. And for that, Computation comes into place where we can extract the main news headlines, analyze them and then calculate the real time sentiments regarding that piece of news to know whether the market feels good or bad about a stock.

Problem Statement

News articles have always been on priority to know about a company and it's stocks as prime information, considering which they classify the news as positive or negative. If there is a positive news sentiment, then there are more chances that the price of a particular stock will rise, and we can buy the stock. On the other hand, if the news sentiment is negative,then the price of the stock may go down, and we can sell it. So this is the key for the people to trade stocks profitably.
In this project, we have generated an investing insight by applying sentiment analysis on financial news headlines from Finviz website which gives the real time data regarding the stock. Using this natural language processing technique, we have tried to analyze the emotion behind the headlines and predict whether to buy or sell a stock.

Literature Survey

Stock price trend prediction is an active research area, as more accurate predictions are directly related to more returns in stocks. It uncovers the future market behavior which always helps the investors to understand when and what stocks can be purchased for the growth of their investment .Therefore significant efforts have been put into developing models that can predict future trends of a specific stock or overall market. Some of the researchers showed that there is a strong relationship between news articles about a company and its stock prices fluctuations.

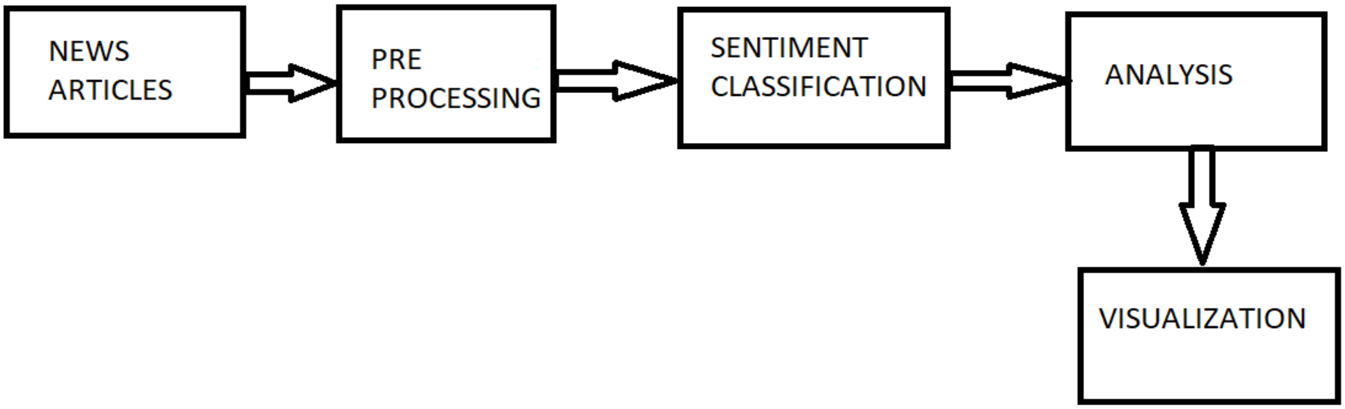
Ref. [4] inspected the ability to use sentiment polarity (positive and negative) and sentiment emotions selected from financial news or tweets to predict the market movements or its life cycle . For this, they have collected a large dataset of the top 25 historical financial news headlines in addition to a large set of financial tweets collected from Twitter. For analysis, they used the Granger causality test [5] that is a statistical test technique majorly used to reveal causality in time series data and explore if one-time series data can predict the other. For sentiment analysis, the authors examined two machine learning methods SVM (Support vector machine)and LSTM(Long short term memory) . The experiment result illustrated that in some cases sentiment emotions contribute to Granger-cause stock price fluctuation, but the finding was not inclusive and must be examined for each case. Also, it has been revealed that for some stocks, adding sentiment emotions to the machine learning market prediction model will increase the prediction accuracy. Comparing the two machine learning methods, SVM achieved better and more balanced results, and that's because the size of the dataset is quite small to be sufficiently used with SVM.

Nagar and Hahsler in their research [6] presented an automated text mining based approach to aggregate news stories from various sources and create a News Corpus which is filtered down to relevant sentences and analyzed using Natural Language Processing (NLP) techniques. A sentiment metric, called NewsSentiment, utilizing the count of positive and negative polarity words is proposed as a measure of the sentiment of the overall news corpus. They also assert that the time variation of NewsSentiment shows a very strong correlation with the actual stock price movement.

Yu et al [7] presents a text mining based approach to determine the sentiment of news articles and depict its impact on energy demand. News sentiment is measured and presented as a time series , compared with variations in energy demand and prices.

J. Bean [8] involves keyword tagging on Twitter feeds about airlines satisfaction and scores them for polarity and sentiment. This provides a quick idea of the sentiments p about airlines and their customer satisfaction ratings.

Proposed Methodology



❑ Workflow

New Articles /Headlines

We have collected real time news headlines of various companies from a reputed website for financial visualization known as “finviz.com”. It is a browser-based stock market research platform which easily provides the headlines of any company you desire.

Pre-Processing

Now the headlines provided to us by this website have to be preprocessed so that we can segregate headlines of each company according to their ticker . If we have to process them they have to be stored in a certain form to give us the desired result

Sentiment Classification

We will be using a lexicon based approach for sentiment analysis , Lexicon-based approaches are based on the use of a sentiment lexicon, i.e., a list of words each mapped to a sentiment score, to rate the sentiment of a text chunk,Our model for this is the VADER model .VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.

Analysis

After the scores of each headline is provided ,Stock sentiment analysis can be used to determine investors’ opinions of a specific stock or asset.

Visualization

For better understanding and analysis of sentiment of each sentiment we have also plotted the results.

❑ Technology

NLTK

The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP).It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning.We will be using it for sentiment analysis using the vader model.

VADER MODEL

VADER (Valence Aware Dictionary for Sentiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion.VADER sentiment analysis relies on a dictionary that maps lexical features to emotion intensities known as sentiment scores. The sentiment score/Compound score of a text can be obtained by summing up the intensity of each word(Valence score) in the text.Compound scores are normalized between -1 (most negative sentiment) and +1 (most positive sentiment).

$$\textit{Compound Score}=\frac{x}{\sqrt{x^2+\alpha}}$$

Here x is the valence score of each word , α is a constant which usually equal to 15.Vader follows certain rules to calculate valence scores of each word :

1.**Punctuation**, namely the exclamation point (!), increases the magnitude of the intensity without modifying the semantic orientation.For eg: ”This is disgusting!!!” Is more negative than “This is disgusting”.

2.**Capitalization**, specifically using ALL-CAPS to emphasize a sentiment-relevant word, increases the magnitude of the sentiment intensity without affecting the semantic orientation. For example :”This is DISGUSTING “conveys more intensity than ,”This is disgusting”.

3.**Degree modifiers** sentiment intensity increases or decreases with degree modifiers. For example: “This is extremely disgusting” is more intense than “This is disgusting”, whereas “This is slightly disgusting.” reduces the intensity.

4.**Polarity shift due to Conjunctions**, The conjunction “but” signals a shift in sentiment polarity, For example: “This is disgusting,but not that much.” has mixed sentiment, with the latter half dictating the overall rating.

5.**Catching Polarity Negation**, By analysis of the continuous sequence of preceding words a sentiment-laden lexical feature, we catch cases where negation flips the polarity of the text. For example a negated sentence would be “This isn't that disgusting.”.

The result obtained is as follows : compound scores, and new “pos/neg” labels derived from the compound score ,where positive sentiment :

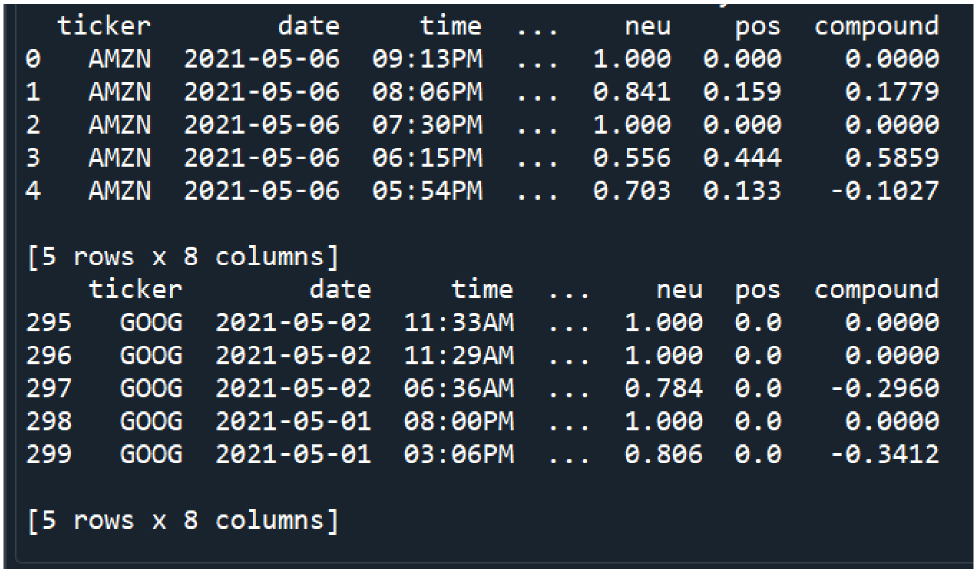
(compound score >= 0.05),neutral sentiment : (compound score > -0.05) and (compound score < 0.05)negative sentiment : (compound score <= -0.05)

Result and Analysis

We have used VADER Sentiment Analysis to predict stock sentiment as there is no need to train the model and therefore, the entire process becomes faster and more efficient.

Here, we can see the bar graph with tickers Amazon (AMZN-blue), Google (GOOG-orange) and Tesla (TSLA-green). Tickers are abbreviations that are used to uniquely identify or denote a particular company’s publicly traded stocks. On the bar graph, the x-axis represents the date the headline was published on and the y-axis represents the compound scores.

If we consider 05 May, 2021; we can see that AMZN has the highest compound score as compared to the others. However, looking at TSLA, we can see that the compound score tends towards -1.

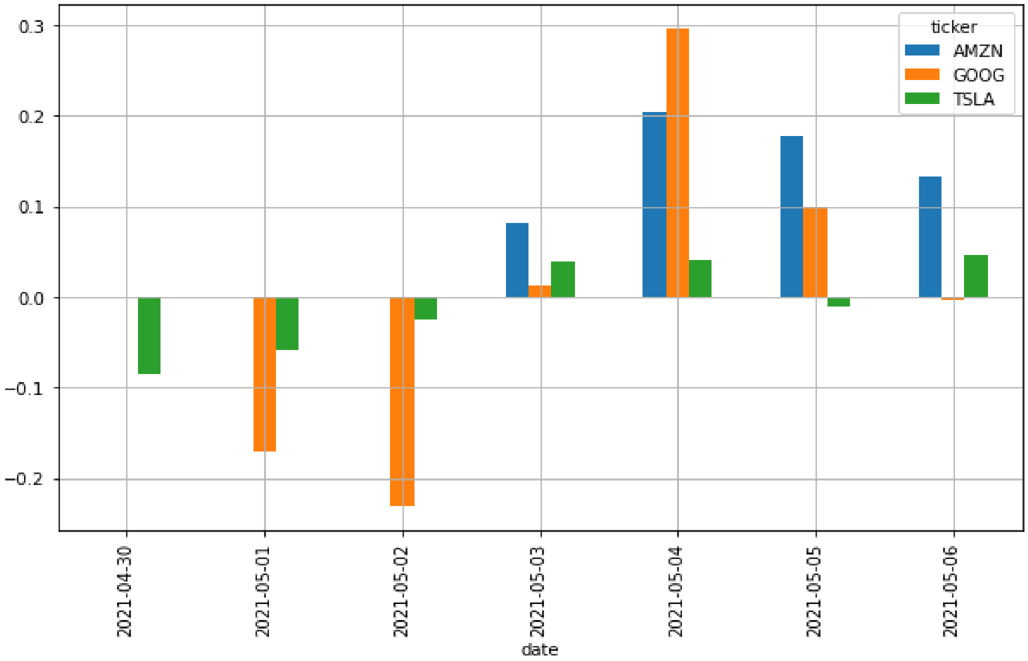


```
  ticker        date     time  ...    neu    pos  compound
0   AMZN  2021-05-06  09:13PM  ...  1.000  0.000    0.0000
1   AMZN  2021-05-06  08:06PM  ...  0.841  0.159    0.1779
2   AMZN  2021-05-06  07:30PM  ...  1.000  0.000    0.0000
3   AMZN  2021-05-06  06:15PM  ...  0.556  0.444    0.5859
4   AMZN  2021-05-06  05:54PM  ...  0.703  0.133   -0.1027

[5 rows x 8 columns]
    ticker        date     time  ...    neu  pos  compound
295   GOOG  2021-05-02  11:33AM  ...  1.000  0.0    0.0000
296   GOOG  2021-05-02  11:29AM  ...  1.000  0.0    0.0000
297   GOOG  2021-05-02  06:36AM  ...  0.784  0.0   -0.2960
298   GOOG  2021-05-01  08:00PM  ...  1.000  0.0    0.0000
299   GOOG  2021-05-01  03:06PM  ...  0.806  0.0   -0.3412

[5 rows x 8 columns]
```

Given above are the negative, neutral, positive and compound scores corresponding to the tickers and their headlines. We can see that when the sentiment is completely neutral, the compound score becomes 0. Whereas, when there’s a negative emotion displayed, the compound score tends to -1 and similarly, when the sentiment is positive, the compound score tends towards +1.



We can calculate the compound scores for other companies as well using VADER Sentiment Analysis.

Concluding Remarks

In this project, we used different sentiment analysis tools to emotionally analyze and classify different stock news headlines. Emotions were classified into the usual positive, negative and neutral categories. Neutral categories appeared for TextBlob and NLTK-VADER Lexicon tools. We used a financial website called FinViz.com that works in real time and prints headlines based on recent time and date.

In the analysis of sentiment results and stock market changes, we compared the last seven days' results of compound values. We obtained an appropriate bar diagram for the reading of the results and the stock market movements, but we could detect differences according to the effect of the neutral values for the headlines. Overall, financial news headlines have an impression on stock exchange values even without their textual context, and significant differences are often observed between different sentiment analytical tools.

But the stock exchange impact also depends on how the data within the current study period was affected.

Of course, there are several challenges that we face while predicting and analyzing stock sentiments. Firstly, only headlines written in the English language can be parsed through the code and analysed. Secondly, sarcasm detection still acts as an issue. Many times we can see that certain news headlines tend to be sarcastic, however, at this stage, our code cannot detect sarcasm.

Future work could include further expansion of the analysis, possible additions of features that would help combat the above mentioned challenges. In addition, comparison between other tools to analyze which gives the best prediction and analysis of the stock market could be done.

References

[1] Oshi Gupta, “Sentiment Analysis of News Headlines For Stock Trend Prediction”, volume 5, issue 12, 2020,: 2456-3315

[2] Spandan Ghose Chowdhury, Soham Routh, Et al. “News Analytics and Sentiment Analysis to Predict Stock Price Trends”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3595-3604

[3] Robert P. Schumaker, Yulei Zhang, Et al. “Sentiment Analysis of Financial News Articles”

[4] Mudinas A, Zhang D, Levene M. Market trend prediction using sentiment analysis: Lessons learned and paths forward. 2019.

[5] Granger CW. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society.

[6] 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

[7] 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

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