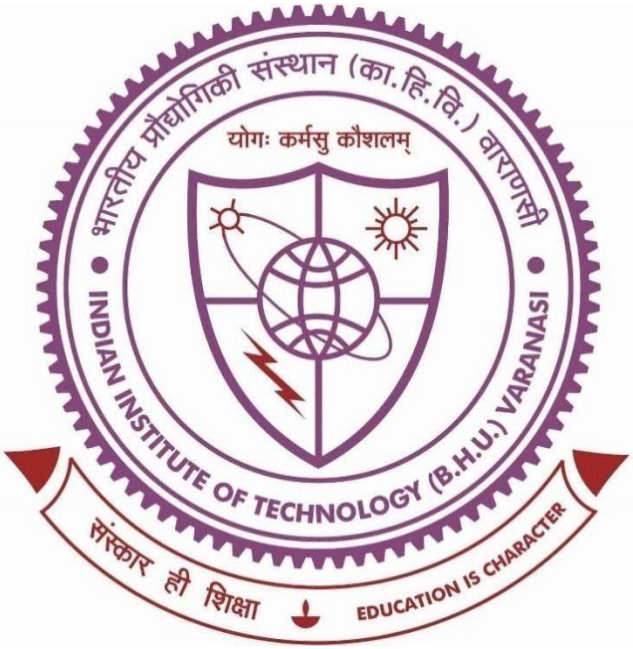


Algorithmic trading using sentiment analysis on news headlines

Members:

Ravi Pratap Singh (17045083)

Malay Shukla (17095038)

Roshan Pramod Chouhan

(17095080)

INTRODUCTION

In the finance field, stock market and its trends are extremely volatile in nature It attracts researchers to capture the volatility and predicting its next moves. Investors and market analysts study the market behaviour and plan their buy or sell strategies accordingly. As stock market produces large amount of data every day, it is very difficult for an individual to consider all the current and past information for predicting future trend of a stock. Mainly there are two methods for forecasting market trends. One is Technical analysis and other is Fundamental analysis. Technical analysis considers past price and volume to predict the future trend where as

Fundamental analysis On the other hand, Fundamental analysis of a business involves analyzing its financial data to get some insights. The efficacy of both technical and fundamental analysis is disputed by the efficient-market hypothesis which states that stock market prices are essentially unpredictable.This research follows the Fundamental analysis technique to discover future trend of a stock by considering news articles about a company as prime information and tries to classify news as good (positive) and bad (negative). If the news sentiment is positive, there are more chances that the stock price will go up and if the news sentiment is negative, then stock price may go down.This research is an attempt to build a model that predicts news polarity which may affect changes in stock trends. In other words,check the impact of news articles on stock prices.We are using VADER library for classification and business times website to gather dataset for news analysis.We are using a Crawler to gather news on a specific company i.e. facebook and gathering news article with date and time stamps. Then applying classifidcation to get sentiment of the news and predicting thus predicting the trend of stock for facebook. By training this model we are then comparing our predicted trends with the real market and also running an algorithm on our prediction model to invest in the market by giving an initial virtual sum on $100000 and an investment allowance of $10 per prediction , if the prediction is positive then shares are invested and if its negative then shares are sold.

METHODOLOGY

.

News Collection: We collected facebook Company's data for past 7 days. This data includes major key events news articles of the company and also daily stock prices of AAPL for the same time period. Daily stock prices contain six values as Open, High, Low, Close, Adjusted Close, and Volume. For integrity throughout the project, we considered Adjusted Close price as everyday stock price. We used a crawler on business times website to get this data.

Pre Processing: Text data is unstructured data. So, we cannot provide raw test data to classifier as an input. Firstly, we need to tokenize the document into words to operate on word level. Text data contains more noisy words which are not contributing towards classification. So, we need to drop those words. In addition, text data may contain numbers, more white spaces, tabs, punctuation characters, stop words etc. We also need to clean data by removing all those words. For this purpose, we created own stop-word list which specifically contains stop words related to finance world and also general English stop words. This stop words list contains general words including Generic, names, Date and numbers, Geographic, Currencies.Also, to ignore words that appear in only one or two documents, we are considering minimum document frequency which considers words that appear in minimum three documents.Stemming is also important to reduce redundancy in words. Using stemming process, all the words are replaced by its original version of word. For example, the words 'developed', 'development', 'developing' are reduced to its stem word 'develop'. Some of the pre-processing is done before applying polarity detection algorithm. And some of them are applied after applying polarity detection algorithm.

SentimentDetection Algorithm: For automatic sentiment detection of news articles, we are following Dictionary based approach which uses Bag of Word technique for text mining. This method is based on the research of J. Bean in his implementation of Twitter sentiment analysis for airline companies. To build the polarity dictionary, we need two types of words collection; i.e. positive words and negative words. Then we can match the article's words against both these words list and count numbers of words appears in both the dictionaries and calculate the score of that document. We created the polarity words dictionary using general words with positive and negative polarity. Also addition to this, we used Finance specific words with its polarity using McDonald's research . In this dictionary, we collected 2360 positive words and 7383 negative words.For the news article, we are considering the string

which contains headline and news body, both. The algorithm to calculate sentiment score of a document is given below.Algorithm:
1. Tokenize the document into word vector.
2. Prepare the dictionary which contains words with its polarity (positive or negative)
3. Check against each word weather it matches with one of the word from positive word dictionary or negative words dictionary.
4. Count number of words belongs to positive and negative polarity.
5. Calculate Score of document = count(pos.matches) –count(neg.matches)
6. If the Score is 0 or more, we consider the document is positive or else, negative.
Here, we are considering one assumption as if the score of the document is 0, then we label it as positive as we are considering two class problem for this implementation.As a result, we get news collection with its sentiment score and polarity as positive or negative.

Document Representation:

In order to reduce the complexity of text documents and make them easier to work with, the documents has to be transformed from the full text version to a document vector which describes the contents of the document. To represent text documents, we are using TF-IDF scheme. The higher tf-idf value a term gets, the more important it is. A high value is reached when the term frequency in the given document is high and when there are few other documents in the collection containing the given term/feature. This term weighting method tends therefore to filter out common terms by giving them a very low value.

VADER Sentiment Analysis:

VADER (Valence Aware Dictionary for sEntiment Reasoning) is a pre-built sentiment analysis model included in the **NLTK** package. It can give both positive/negative (polarity) as well as the strength of the emotion (intensity) of a text. It is rule-based and relies heavily on humans rating texts via Amazon Mechanical Turk — a crowd-sourcing e-platform which utilizes human intelligence to perform tasks that computers are currently unable to do. This literally means that other people have already done the dirty work of building a sentiment lexicon for us. These are *words or any textual form of communication generally labelled according to their semantic orientation as either positive or negative)* for us.

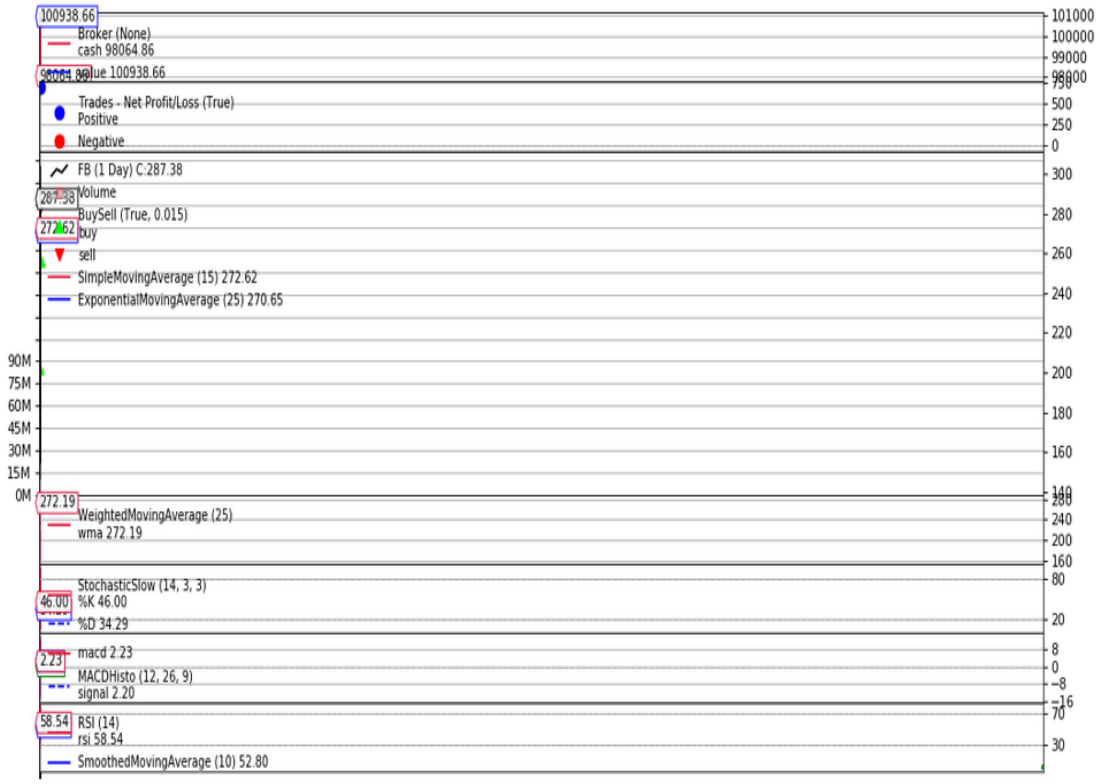
The sentiment score of a text can be obtained by summing up the intensity of each word in the text and then normalizing it. The human raters of Vader used 5 heuristics to analyze the sentiment:

1. **Punctuation** — I love pizza vs I love pizza!!

2. **Capitalization** — I'm hungry!! vs I'M HUNGRY!!
3. **Degree modifiers (use of intensifiers)**— I WANT TO EAT!! VS I REALLY WANT TO EAT!!
4. **Conjunctions (shift in sentiment polarity, with later dictating polarity)** — I love pizza, but I really hate Pizza Hut (bad review)
5. **Preceding Tri-gram** (identifying reverse polarity by examining the tri-gram before the lexical feature— Canadian Pizza **is not really** *all that great.*

Plotting the values: After classification of unknown data, we plotted the news score chart and compared with historical price chart.
Given below is the process of crawling business today website to gather news on facebook.

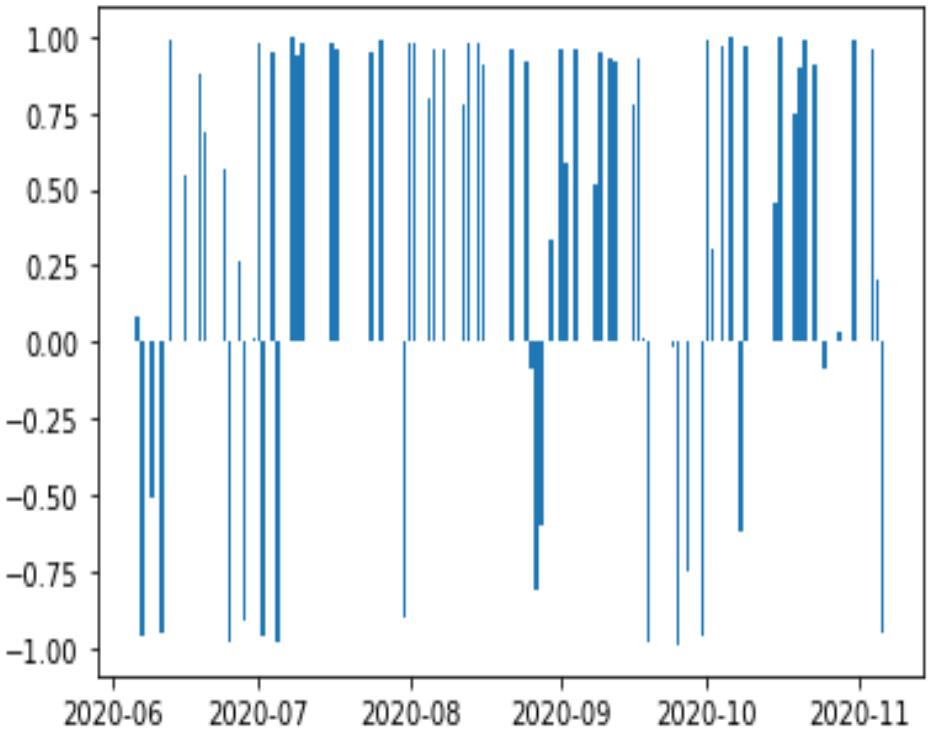
```
05 Nov 2020 https://www.businesstimes.com.sg/technology/twitter-facebook-suspend-some-accounts-as-us-election-misinformation-spreads-
online
04 Nov 2020 https://www.businesstimes.com.sg/technology/twitter-facebook-push-back-on-trumps-election-posts
03 Nov 2020 https://www.businesstimes.com.sg/technology/major-leak-sees-one-million-swedes-data-shared-with-facebook-google
30 Oct 2020 https://www.businesstimes.com.sg/technology/facebook-anticipates-tougher-2021-even-as-pandemic-boosts-ad-revenue
27 Oct 2020 https://www.businesstimes.com.sg/technology/with-new-tools-facebook-aims-to-avoid-election-fiasco-repeat
27 Oct 2020 https://www.businesstimes.com.sg/technology/facebook-takes-mobile-games-into-the-cloud
24 Oct 2020 https://www.businesstimes.com.sg/technology/facebook-twitter-ceos-to-testify-post-election-us-senate-panel
24 Oct 2020 https://www.businesstimes.com.sg/technology/us-may-file-antitrust-charges-against-facebook-as-soon-as-november-washington
-post
22 Oct 2020 https://www.businesstimes.com.sg/technology/facebook-launches-dating-service-in-europe
20 Oct 2020 https://www.businesstimes.com.sg/technology/facebook-unveils-machine-learning-translator-for-100-languages
19 Oct 2020 https://www.businesstimes.com.sg/technology/irish-regulator-probes-facebooks-handling-of-childrens-data-on-instagram
18 Oct 2020 https://www.businesstimes.com.sg/technology/22m-facebook-and-instagram-ads-rejected-ahead-of-us-vote
15 Oct 2020 https://www.businesstimes.com.sg/technology/zoom-opens-platform-for-paid-events-following-facebook
14 Oct 2020 https://www.businesstimes.com.sg/technology/facebook-bans-ads-discouraging-vaccines
08 Oct 2020 https://www.businesstimes.com.sg/government-economy/facebook-to-pause-political-ads-as-us-election-day-ends
07 Oct 2020 https://www.businesstimes.com.sg/government-economy/facebook-pulls-trump-post-for-minimising-covid-19-danger
05 Oct 2020 https://www.businesstimes.com.sg/energy-commodities/sunseap-to-supply-facebook-with-solar-energy-for-singapore-operations
03 Oct 2020 https://www.businesstimes.com.sg/technology/facebook-twitter-ceos-to-testify-before-us-senate-committee
01 Oct 2020 https://www.businesstimes.com.sg/garage/news/india-startups-join-effort-to-break-google-facebook-dominance
01 Oct 2020 https://www.businesstimes.com.sg/government-economy/facebook-bans-us-ads-that-call-voting-fraud-widespread-or-election-in
valid
30 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-overhauls-instagram-messaging-enabling-cross-app-chats-with-messenge
r
29 Sep 2020 https://www.businesstimes.com.sg/technology/philippines-accuses-facebook-of-censoring-pro-government-content
26 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-gets-reprieve-from-apple-on-live-events-cut
24 Sep 2020 https://www.businesstimes.com.sg/government-economy/thailand-takes-first-legal-action-against-facebook-twitter-over-conte
nt
23 Sep 2020 https://www.businesstimes.com.sg/government-economy/thailand-to-start-legal-action-against-facebook-google-twitter-over-c
ontent
23 Sep 2020 https://www.businesstimes.com.sg/government-economy/facebook-says-fake-accounts-from-china-aimed-at-us-politics
18 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-to-curb-private-groups-spreading-hate-misinformation
17 Sep 2020 https://www.businesstimes.com.sg/technology/australian-regulator-dares-facebook-to-block-news-content
17 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-plans-ray-ban-smart-glasses-as-it-eyes-ar
16 Sep 2020 https://www.businesstimes.com.sg/sme/facebook-singapore-to-give-s475m-in-grants-to-small-businesses-hit-by-covid-19
15 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-launches-climate-science-info-centre-amid-fake-news-criticism
11 Sep 2020 https://www.businesstimes.com.sg/technology/eu-urges-facebook-google-twitter-to-do-more-against-fake-news
10 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-google-twitter-urged-by-eu-to-do-more-against-fake-news
08 Sep 2020 https://www.businesstimes.com.sg/real-estate/facebook-campus-hotel-a-bet-on-comeback-for-business-travel
07 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-plans-to-label-posts-more-aggressively-clegg-says
03 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-bans-indian-ruling-party-politician-for-policy-violation
01 Sep 2020 https://www.businesstimes.com.sg/technology/facebook-threatens-to-stop-publishers-in-australia-share-local-news-if-regula
tion-becomes
31 Aug 2020 https://www.businesstimes.com.sg/technology/facebook-says-eu-data-demands-included-risks-to-staffs-families
29 Aug 2020 https://www.businesstimes.com.sg/technology/google-facebook-dump-hong-kong-cable-after-us-security-alarm
29 Aug 2020 https://www.businesstimes.com.sg/technology/facebook-fights-uk-merger-regulator-over-giphy-acquisition
29 Aug 2020 https://www.businesstimes.com.sg/technology/facebook-ceo-says-kenosha-guard-page-was-left-up-by-mistake
27 Aug 2020 https://www.businesstimes.com.sg/technology/facebook-says-apple-mobile-software-will-cut-ad-revenue
26 Aug 2020 https://www.businesstimes.com.sg/government-economy/thai-minister-says-clampdown-wont-stop-as-facebook-plans-to-fight-ord
er
25 Aug 2020 https://www.businesstimes.com.sg/technology/after-block-new-facebook-group-criticising-thai-king-gains-500000-members
```



CONCLUSION

Finding future trend for a stock is a crucial task because stock trends depend on number of factors. We assumed that news articles and stock price are related to each other. And, news may have capacity to fluctuate stock trend. So, we thoroughly studied this relationship and concluded that stock trend can

be predicted using news articles and previous price history. As news articles capture sentiment about the current market, we automate this sentiment detection and based on the words in the news articles, we can get an overall news polarity. If the news is positive, then we can state that this news impact is good in the market, so more chances of stock price go high. And if the news is negative, then it may impact the stock price to go down in trend. We used polarity detection algorithm for initially labelling news and making the train set. For this algorithm, dictionary based approach was used. The dictionaries for positive and negative words are created using general and finance specific sentiment carrying words. Then pre-processing of text data was also a challenging task. We used VADER dictionary for stop words removal which also includes finance specific stop words. Based on this data, we allocated a sum of money $100000 to the model to see its performance in the real market and allocated $10 for each transaction when the model predicted that stock price would go up by a cerain value then our model would make an investment of allocated amount and whe the news was negative and the predicted price was to drop by a certain value it would sell , based on these activities we studied the model on a week's data and finally concluded that our model was in profit of $75. The above given graph was generated to compare the predicted price with actual price variation and the image given below depicts the intensity of predicted price variation for buy and sell operation.



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 MeetupPresentation, 2011