SENTIMENT ANALYSIS OF TWITTER DATA FOR PREDICTION OF STOCK PRICES

National University of Singapore

Kaustubh Jagtap

Marcus Chia Shan Wei

Ola Irgens Kylling

Rahul Rajesh

Executive Summary

The growth of digital technologies and the Internet have fueled newer ways of sharing information and interacting. An increasing number of people have started to rely on social media and information content platforms for valuable and timely information, and investors are no exception. In today's financial markets, many traders rely on Sentiment Analysis to evaluate overall market sentiment of stocks, before employing machine learning techniques such as a Modified Bag of Words Model to guide trade decisions. In this paper, we propose a machine learning model that can process subjective information effectively in order to guide investment decisions.

Methodology

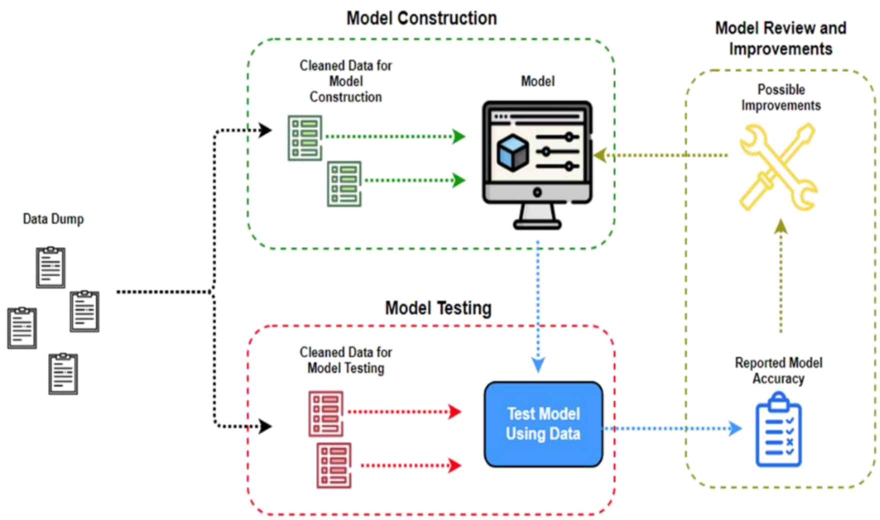
The Bag-of-Words Model is one of the most basic ways of extracting information from subjective text. As its name suggests, “Bag of Words” implies that any information regarding the order or context of the words is discarded. The only feature that the model is concerned with is the occurrence of each individual word. However, when using the Bag-of-Words Model, we can immediately notice that with the addition of a few simple rules, we can drastically improve the accuracy of the model. For instance, we can take note of stop words, intensifier words and negation words which change the meaning or the magnitude of the following word. As such, we have decided to add three additional rules (carried out in the following order) to the Bag-of-Words Model:

1. **Filter out all stop words (as specified by the Python NLTK) which do not carry meaning**
   It is a good practice to remove stop words to lower the dimensional space since stop words do not affect the overall sentiment of statements.

2. **When there is an occurrence of an intensifier word, increase the goodness/badness score of the following positive/negative word by 1**
   By amplifying the score of the word following an intensifier word, we extract the amplitude of the modifier along with the sentiment of the word.

3. **When there is a Negation word, reverse the goodness and badness scores of the following word**
   A negation word turns parts of the statement into its opposite. By reversing the score of the word following a negation word, we can capture the true meaning of the word that is the focus of negation.

Methodology



As with other Natural Language Processing models, the most fundamental steps when implementing the Bag-of-Words model are:

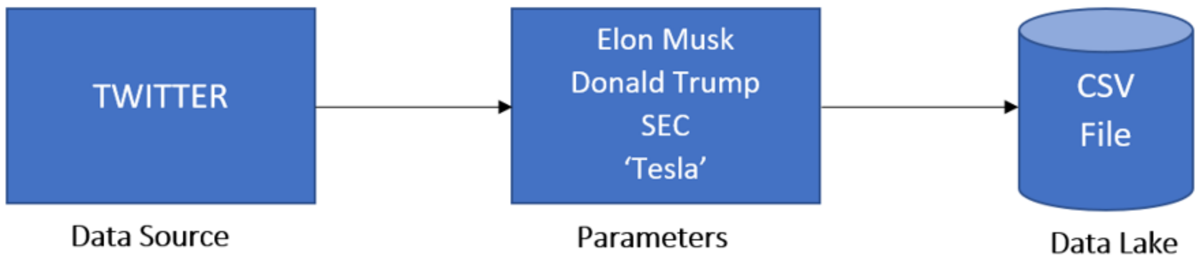
1. **Forming your training set of data**

1000 articles from the Thomson-Reuters News Archive database were used to form our training sample. Only news articles which are related to one specific company are used. In addition, to have a balanced sample, we randomly select 500 positive news articles and 500 negative news articles as our training sample.

2. Scraping Twitter data

With the use of Twitter API, tweets regarding the stock 'Tesla' are scraped into CSV files. Important information that will be collected are the user, the time of the tweet, the number of likes and retweets of the tweet and the contents of the tweet.

Data Collection - Features



Data Lake Schema - Features

Field Name	Data Type
username	STRING
timestamp	DATETIME
num_likes	INT
num_retweets	INT
tweet_content	STRING

3. Word Tokenization

Word Tokenization is carried out on the contents of the tweet, using the Python Natural Language Toolkit, a tokenizer that divides a string of text into individual words (tokens).

4. Stemming

Stemming is the process of reducing derived words to their base word. Stemming is carried out using the Porter stemming algorithm, which comes with the Python Natural Language Toolkit.

5. Filtering

After stemming, we filter out intensifier words, negation words and stop words and apply the rules as discussed earlier.

6. Usage of Model for Sentiment Analysis of Tweets

The model will be used to analyze the sentiments of the most recent social media mentions of selected stock counters. We find the frequency that each word in the tweet appears in positive and negative articles in our training data. Sentiment scores that range from 1 to -1 will be calculated for each word. The overall sentiment for the tweet will be the sum of the sentiment scores of all the words (after applying rules regarding intensifier, negation and stop words).

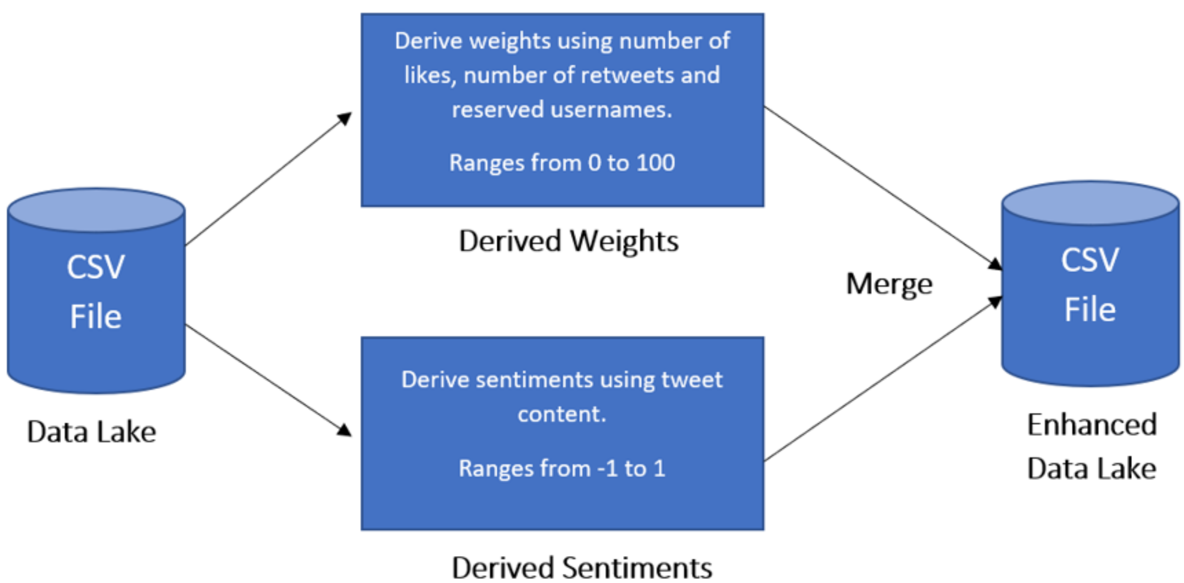
7. Verification and Validation

Using the sentiment scores obtained, we plot a linear regression of sentiment score against percentage change in stock price to see our sentiment scores can explain much of the variation in stock prices.

Words			Goodness	Badness
Fantastic	15	3	0.833	0.167
video	200	200	0.5	0.5
and	100	100	0.5	0.5
surprisingly	70	30	0.7	0.3
excellent	20	1	0.952	0.047
images	100	100	0.5	0.5
		Total	3.99	2.01

Example of calculating Sentiment Score of a tweet is shown in the diagram on the left. Overall sentiment score of the tweet is calculating using the difference between the goodness and badness score of the tweet.

Data Processing - Features



Results

For the ease of explanation, our full results will not be displayed in this report. A random sample (5% of the total tweets) is selected for manual inspection to determine the accuracy of the sentiment scores. Furthermore, only selected parts of the random sample will be used in the explanation of the results. The table below serves as a preface of our results.

USER	TWEET	TIME	FOLLOWERS	SENTIMENT SCORE	PRICE CHANGE
Stephen Ellison	#BREAKING: Tesla Model S that caught fire in Los Gat breaking, catch, fire	0800	979	-1.32	0.0044%
KPIX 5	UPDATE: The @Tesla Model S that caught on fire earl catch, fire, extinguish	1500	85100	-0.84	0.0044%
Amanda Del Castilo	#NEW @Tesla owner shares this incredible video of h incredible, catch, fire	1500	337	-0.29	0.0044%
Autotestdrivers.com	Tesla Model S catches on fire after being towed to a g catch, fire, investigate, burn, loss, unknown	1800	3141	-3.97	-0.1468%
Price Action	A Tesla catching on fire is nothing to be worried abou catch, fire, nothing, worry, suppress, burn, suffocate	2100	32	-4.81	-0.0032%

In general, the following observations can be made from our results:

1. **<u>Sentiment scores are accurate</u>**

   When we manually inspected a random sample of the tweets, we found that the sentiment scores were generally coherent with the sentiment score, meaning to say that our modified bag of words model was accurate in gauging the positivity and negativity of a tweet. However, this could have been by chance as we noted some irregularities in the results. As seen in the table above, the sentiment "Tesla owner shares this incredible video of his car catching fire" is in no way a positive sentiment. However, since the word "incredible" tends to appear in more positive sentiments, it has a large positive sentiment score and could have very well made the overall sentiment a positive one.

2. <u>Sentiment scores and changes in prices are not positively correlated</u>

When a simple linear regression was conducted to see how sentiment scores affected changes in price, we found sentiment scores explained very little of the variation in the stock price. We noted however, that the R-squared value was higher during the NYSE opening times.

Assumptions

In our research, there were several simplifying assumptions we made. However, during implementation, investors should account for and mitigate these assumptions:

1. We assumed that the delays in investor response to tweets is the same for all investors at all timings. This delay was taken to be 1 hour, independent of the actual user itself. However, from our results, we see that the sentiment scores tended to be more positively correlated to the percentage change in the price of the stock, proving our assumption invalid.

2. While we noted that the tweets of certain influential people might have a more significant effect on stock price than others, we assumed the 'weight' of their tweets to be negligible in this study. Further studies, however, should account for the weight and aggregate the sentiment of the overall market by taking into consideration every single person who tweets anything about Tesla.

Limitations

There are certain limitations and areas of improvement for our research.

1. Firstly, we only focused our entire research just based on 1 stock - Tesla. The twitter accounts we tracked as well as the nuances of the NLP sentiments we assigned were very specific to this stock. Therefore, our research cannot be generalized to other stocks before carrying out further studies.

2. In addition, we only managed to scrape the past seven days of tweets as the complementary version of the Twitter API only allows the collection of the past week's tweets. This makes it difficult to assume that our sentiment scores will be accurate beyond the previous week.

3. Furthermore, the methodology we used to assign the scores is largely heuristic. This means that the formula is 'made up' by us as a nominal, or ordered, variable. It provides a decent *relative* indication of sentiment, but taken in absolute sense, does not provide a meaningful measure. However, heuristic measures by nature are meant for this purpose. It could be improved using computational methods that focus on improving candidate language identification techniques during each step.

4. Lastly, the output from a model is a simple up or down classification. Although this gives a good prediction of the direction of movement of the security, it does not indicate a target price. Due to this shortcoming, it is not easy to place a trade with a properly defined target price, stop loss and take profit.