TWEET-BASED SENTIMENT ANALYSIS OF APPLE

BUSI4392 ANALYTICS SPECIALIZATIONS & APPLICATIONS, 2019-2020

Shiqi BAI (20219140)

CONTENT

EXECUTIVE SUMMARY

The stock exchange is a subject highly influenced by economic, social and political factors. There are several factors, such as external or internal factors that may affect and move the stock market. Due to changes in supply and demand, stock prices rise and fall every second. Solving this problem usually involves various data mining techniques. But using machine learning techniques will provide more accurate, precise, and simple methods for solving problems related to stocks and market prices. Stock related analysis is one of the most important topics in academic and financial research.

Various data mining techniques are often involved in research. solve this problem. But using machine learning/deep learning techniques will provide a more accurate, precise, and simple method for solving problems related to stocks and market prices. In this report, we applied the sentiment analysis for the tweet content and implement a model by using Logistic Regression, to differentiate the engagement score of a tweet.

With the datasets provided by tweeter and collected by Kaggle, we completed the sentiment analysis and topic modeling with the help of LDA algorithm after a series of data cleaning processes. Then we found that all topics except of 'IWatch' related are highly relevant with the customers engagement according to the regression results. Based on the findings, we recommended that Apple company can consider predicting stock volume and price using Tweets data in the future research.

The rest of the report is organized as follows. In the second part, approaches breakdown introduces the content of our report. The third part shows the method and process of data collection in tweets. The fouth part explains the sentiment analysis result and important findings in this field. Section five illustrates the potential research ideas for the evaluation and tweets and the Apple Stock Index. Final section is about the conclusion and reflection of this survey.

APPROACH BREAKDOWN

EXPLORATION DATA ANALYSIS: There are limit number of innovative tweets from the dataset and many individuals preferred retweeting the posts from some users (i.e. KOLs) directly. Since that we cannot guarantee those individuals motivations, we selected the samples that posted by some Key Opinion Leader (KOLs) for our analysis.

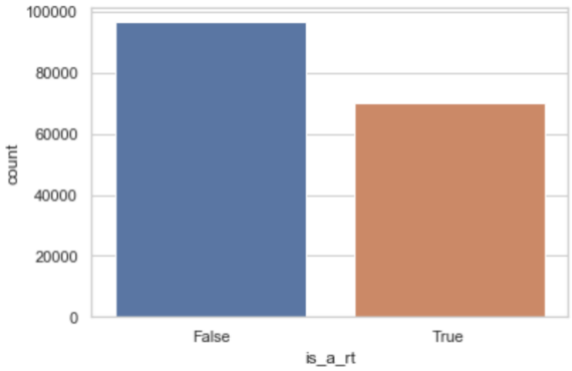

Figure 1: The ratio of tweets whether retweet or not

SENTIMENT ANALYSIS: Sentiment analysis has become the core of social media research. In the fields of e-commerce and trade, management and politicians, many studies have been conducted to obtain user opinions. Social media has recently become a rich resource for tapping user emotions. Using sentiment analysis to analyze public opinion, some studies have shown that sentiment analysis of news, documents, quarterly reports, and blogs can be used as part of a trading strategy. In this report, Twitter was chosen as the platform for mining opinions when trading strategies with the Apple stock market to execute and explain the relationship between tweets and the Apple Market Index. Furthermore, we applied a Logistic Regression model to identify the important topics to affect the customer engagement.

DATA COLLECTION

This section will outline data collection methods, including samples selection and sentiment preparation methods.

TWEETS DATA COLLECTION:

We explored that many individuals retweeted the 'KOLs' tweets, which we regarded their posts as outliers. Therefore, we filtered Top 10 KOLs who posted the tweets more frequently as our target observations as shown in Figure 2.

username	
Peripheral News	6971
Computer News	6920
Electronic News	6372
trader whodont trade	4703
Tim Spencer	1636
Sam Miller	1619
MacHash	1447
The Stock Professor	1233
John Lee	1162
David Moadel	1029
Stakepool	983
NASDAQStocks	977
AAPL Stock Alerts	892
Stock News Herald	845
Capital Market Labs	645

Figure 2: List of TOP 10 KOLs and their post amount

APPLE STOCK DATA COLLECTION:

In addition, we collected past stock data from 2016-04-02 to 2016-05-15 through the Kaggle dataset [1], so that for the preparation in digging the relationship between tweets information and stock index.

	Date	Close/Last	Volume	Open	High	Low
0	02/28/2020	$273.36	106721200	$257.26	$278.41	$256.37
1	02/27/2020	$273.52	80151380	$281.1	$286	$272.96
2	02/26/2020	$292.65	49678430	$286.53	$297.88	$286.5
3	02/25/2020	$288.08	57668360	$300.95	$302.53	$286.13
4	02/24/2020	$298.18	55548830	$297.26	$304.18	$289.23

Figure 3: Samples of Apple stock data we collected

DATA PREPROCESSING: After the data tagging is complete, the data is stored in our system in a standard format as described in the framework shown in Table 1.

Choose relevant tweet observations and stock data from 2016-04-02 to 2016-05-15

No duplicates or retweets in KOLs' tweets
Remove URLs, mention symbol (@), hashtag symbol (#) in tweet content
Convert the original data types into proper ones

Table 1: The general tweet content cleaning processes

Since then, four preprocessing steps have been carried out for sentiment analysis.

- Tokenization: Divide each tweet into multiple token-based space characters.
- Process of deleting stop words: This operation is performed to delete stop words in Arabic.
- Light root: remove the suffix and prefix and return the word to its root.
- Filter tokens by length: delete useless words and set them to three.

ANALYSIS SECTION

THE RELATIONSHIP BETWEEN POSITIVE EVENTS AND MARKET PERFORMANCE: Although it is generally believed that stock market prices are mainly driven by new information and follow a random pattern, many studies have tried to use external stimuli to predict stock market behavior based on behavioral economics, which emphasizes the important role of emotions in decision-making . Overall, we believe that because it is difficult to quantify events changes, there is limited evidence that there is a direct link between social events and market performance. However, we applied sentiment analysis to calculate emotional polarity for each tweet and obtained the statistical information of positive tweets and stock volume. We have found a clear relationship between sentiment of tweets and Apple stock index.

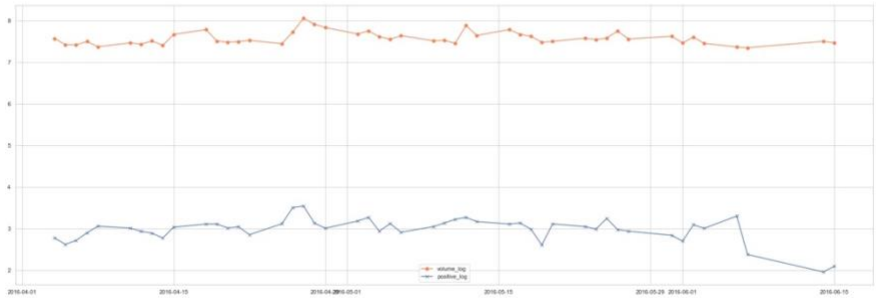

Figure 4: Correlation between sentiment analysis and Apple stock market (Volume)

SENTIMENT ANALYSIS AND TOPIC MODELING BY LDA METHOD:

	tweet_cnt			perc_negative	perc_positive	perc_neutral
sentiment_bucket	negative	neutral	positive			
User Name						
AAPL Stock Alerts	124	519	249	13.901345	27.914798	58.183857
Capital Market Labs	106	345	194	16.434109	30.077519	53.488372
Computer News	56	199	6665	0.809249	96.315029	2.875723
David Moadel	175	532	322	17.006803	31.292517	51.700680
Electronic News	56	171	6145	0.878845	96.437539	2.683616
John Lee	123	790	249	10.585198	21.428571	67.986231
MacHash	158	770	519	10.919143	35.867312	53.213545
NASDAQStocks	82	530	365	8.393040	37.359263	54.247697
Peripheral News	60	202	6709	0.860709	96.241572	2.897719
Sam Miller	167	917	535	10.315009	33.045090	56.639901
Stakepool	72	502	409	7.324517	41.607325	51.068159
Stock News Herald	2	4	839	0.236686	99.289941	0.473373
The Stock Professor	234	657	342	18.978102	27.737226	53.284672
Tim Spencer	190	820	626	11.613692	38.264059	50.122249
trader whodont trade	451	2157	2095	9.589624	44.546034	45.864342

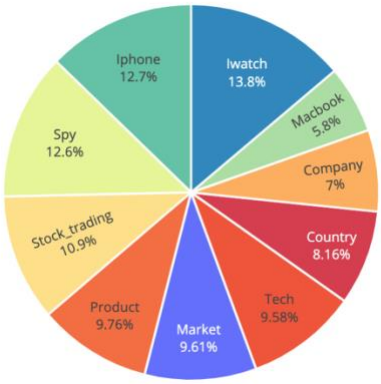

Figure 5: Sentiment classification (Left) and topic distribution of KOLs (Right)

"Using Twitter sentiment analysis for stock price prediction" can be developed as a method for predicting stock prices. Changes in the Apple's stock price are related to public opinion expressed in tweets about this company. Understanding the author's point of view from a passage is the purpose of sentiment analysis. Positive news and tweets about Apple on social media will definitely encourage people to invest in the company's stock, and as a result, Apple's stock price will rise.

ENGAGEMENT ANALYSIS BY LOGISTIC REGRESSION: the results of regression are shown as Figure 6, which indicates that only the topic of 'IWatch' has a weak relation with the customer's engagement. For the marketing consideration, the Apple company should focus on the more relative topics to develop a marketing strategy.

	coef	std err	t	P>|t|	[0.025	0.975]
Repeats	60.0000	6.07e-17	9.88e+17	0.000	60.000	60.000
Favorates	40.0000	2.28e-15	1.75e+16	0.000	40.000	40.000
sentiment_score	-9.77e-15	3.24e-15	-3.013	0.003	-1.61e-14	-3.41e-15
Country	1.36e-14	3.06e-15	4.444	0.000	7.6e-15	1.96e-14
Product	-5.773e-15	2.86e-15	-2.022	0.043	-1.14e-14	-1.76e-16
Company	9.714e-14	3.26e-15	29.773	0.000	9.07e-14	1.04e-13
Stock_trading	2.764e-14	2.65e-15	10.420	0.000	2.24e-14	3.28e-14
Spy	-4.063e-14	2.45e-15	-16.601	0.000	-4.54e-14	-3.58e-14
Macbook	-2.537e-14	3.59e-15	-7.059	0.000	-3.24e-14	-1.83e-14
Iphone	6.883e-15	2.53e-15	2.723	0.006	1.93e-15	1.18e-14
Iwatch	8.882e-16	2.47e-15	0.359	0.719	-3.96e-15	5.73e-15
Market	-6.062e-14	2.78e-15	-21.795	0.000	-6.61e-14	-5.52e-14
Tech	2.598e-14	2.84e-15	9.162	0.000	2.04e-14	3.15e-14

Figure 6: The degree of influence of independent variables on dependent variables

FURTHER ANALYSIS RECOMMENDATION

As we discussed in the last section, we recommend the company can develop a prediction model by using machine learning, which is used to find and analyze the correlation between tweet content and stock prices, and then predict future prices. Vocabulary-based sentiment analysis allows the classification of data sentiment to measure public sentiment related to Apple products and itself, and collects stock market indicators such as trading volume, market closing price, and average daily price changes to track market trends roughly. A series of tests based on correlation and regression determine the presence or absence of the relationship between variables in the context and enable the study to explore the impact of public opinion on investment decisions.

This early study of the relationship between public tweet discussions and market performance shows that it is expected to use sentiment analysis on Twitter data to potentially predict market trends. The results highlight a group of variables with stronger causality, which should become the focus of similar studies in the future. More research may be needed to use a wider sample of tweets or longer observation times to explore the relationship between Twitter-based discussions and stock market trends, even predict the close prize could be a possible research in the future research of Apple.

CONCLUSION

We can determine some of the main limitations of this survey: the first question of the sample size is whether it can well represent the views of the public Twitter population. In this survey, the final tweet sample contained more than 160,000 tweets within a 70-day period. Considering that Twitter is updated hundreds of millions of times a day, this seems to be a trivial sample that can well represent demographic sentiment. Having said that, the sample in this study should be designed to well represent the discussions on Twitter by the KOLs, and because the total number is unknown, it is difficult to say whether the sample of tweets fairly represents the overall sentiment of the public.

The second and more harmful problem of this study is to specify the size of the sampling period. Although the purpose of the study was to select only 70 days to measure the existence of short-term relationships based on the study of events, the stock performance is a long-term trend. Since the trend of the relationship is not obvious in the smaller sample, the size of the sample is not sufficient to classify this relationship as being significant with certainty. However, just because a relationship is found to be statistically unimportant does not mean that the relationship does not exist.

For future research, we recommend finding the pattern between Apple's stock index and public opinion on Twitter by adding emotional characteristics to Apple's stock closing price and Twitter. This may enable our model to predict the opening price of the Apple stock market.

REFERENCE

1. https://www.kaggle.com/tarunpaparaju/apple-aapl-historical-stock-data