International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Prediction of Stock Values using Sentiment Analysis on Twitter Data

Aishwarya .V[1], Dr. Vani Priya[2]

[1]Computer Science and Engineering, Sir M Visveswarya Institute of Technology Bangalore, India

[2]Infromation Science and Engineering, Sir M Visveswarya Institute of Technology Bangalore, India

Abstract: ***Today, public opinion on social media is used as a primary key source to take informed decisions like to buy a product, to invest in a stock a company or to predict whether a movie would be block buster at the box office and so on. Manual digestion of such a huge data is very tedious and time consuming. Sentiment analysis is a technique which can be used to categorize emotions/feelings within people's opinions expressed within an online mention., automatically without manual intervention.Among all the applications of the sentiment analysis, the stock value prediction is quiet a difficult task since it is dependent on the demand of the stock and new information significantly. Also, it has been an active area of research for a long time. In our experiment we show how to connect to twitter and perform sentiment analysis.We used TextBlob method to build the prediction model. The main objective of this study is to determine the accuracy of a machine learning technique (TextBlob) with respect to providing a positive, negative and a neutral classification using Sentiment Analysis for stock related tweets.Our experiment shows that the percentage for positive sentiment of tweets are significantly higher.***

Keywords: Twitter Sentiment Analysis, stock values, Predictive Analysis, twitter, TextBlob

1. Introduction

Sentiment Analysis also known as opinion mining is used to analyze the emotional tone or underlying sentiment behind words to categorize emotions or feelings within people's opinion expressed within an online mention. Sentiment Analysis is basically the process of determining the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral.

Sentiment Analysis has got many applications in different domains. One of the many examples is in businessesto providing feedbacks about a company's stock by the reviews on social media. In this paper we mainly focus on Twitter Sentiment Analysis.

Stock market or Stock Exchange is the act of forecasting the future of a company. It is a collection of markets and exchanges where regular activities of buying, selling and issuance of shares of publicly-held companies take place. The Stock market allows companies to issue and sell their shares to the common public. It essentially means that a company divides itself into a number of shares and sells a part of those shares to the common public [8]. Many large companies have their stocks listed on Stock exchange [1]

The Stock value prediction is quite a task as it fluctuates rapidly and poses many challenges as new information is a significant factor which affects change in the stock price.[2] People usually tend to buy a company's stock if the company's reputation is good [3]. Social media and few networking sites like Facebook, Twitter, YouTube can influence people's decision whether or not they will buy the company's stock [4]. One of the many social media that is commonly used by a company is Twitter as it is open for public consumption, has well defined API and tweets are posted at the 'speed of thought' [5] .

Twitter is used as a real time microblogging service platform that allows people to share their views, express their curiosity, thoughts or opinions as to how they feel about any topic and communicate with short messages/limited length of text say which comprises of 140 characters or less [6]. Twitter is an excellent data source to be considered for sentiment analysis for predicting stockvalues.

In our research, we have applied Sentiment Analysis on tweets extracted from Twitter and classification of tweets into positive, negative and neutral is carried out. In phrase level, the Sentiment Analysis system recognizes the polarity of the phrase.

There can be two approaches to Sentiment Analysis:
- Lexicon-based-method.
- Using Machine LearningTechniques.

A Lexicon-based method uses a dictionary of words incorporating part-of-speech tagging with assigned to them a semantic scores to predict the final polarity of the tweets.

The second approach, using machine learning techniques which include Unsupervised and Supervised learning algorithms that the problemacts as a text classification.

The two methods of Sentiment Analysis lexicon-based method and machine learning based technique both rely on the bag-of-words.

We have applied the Machine Learning Techniques.

The various areas where we can apply sentiment analysis are:
a) Business: The company can use sentiment analysis to analyses the product reviews so that they make strategies to improvise the business plan.
b) Politics: The public opinion matters a lot as far as the

social media is concerned, the political parties, now a days are engaging the technical experts who analyze the opinion of people on social media.

c) Security concerns: Sentiment analysis can also be used to find the dangers that can be posed by criminal or terrorists activities

In this experiment we focus mainly on the textual information of the message.

From the stand of text classification, some of the challenges include limited length of text, sarcasm and irony, language style on circumstances and consequently what is considered positive to one person may be considered as negative to another person [7].

To do this, we present a step-by-step methodology on twitter sentiment analysis

Twitter Sentiment Analysis

This technical paper records the implementation of Twitter Sentiment Analysis by using the APIs provided by twitter.
The approach to extracting tweets from twitter is as follows:

1) Importing the necessary packages and libraries.
2) Defining parameters that will establish connections with the twitter API.
3) Clean the tweets by removing the stop words like “the”, ”as”, ”a”, “in”.
4) Tokenizing each word from the dataset and feeding into the program.
5) For each word/sentence compare it with positive or negative.
6) Finally, determine the accuracy for positive, negative and neutral tweets.

The below figure shows the twitter sentiment analysis algorithm:

```
procedure TWITTER-CONNECTION
consumer key = "xyxyxyxyxxy"
consumer secret = "xyxyxyxyxxy"
access token = "xyxyxyxyxxy"
access_token_secret = "xyxyxyxyxxy"
auth = tweepy.OAuthHandler(consumer key, consumer
secret)
auth.set_access_token(access token, access_token_secret)
end procedure

procedure TWEET-CLEANING
tweet= Lemoore-stop words
Return tweet
end procedure

procedure TWEET-CLASSIFCATION(t)
for tweet in cursor. Items(3000):
   analysis =tweet. Text
   polarity += analysis.sentiment.polarity
end procedure
procedure GET-TWEETS(count)
cursor = tweepy.Cursor(
   OpenSearch, q=query + "-fetch: retweets", Lang="end",
result type="09-05-2016"
)
end procedure

procedure MAIN()
for tweet in cursor. Items(3000):
   analysis =tweet. Text
if polarity == 0:
   print("neutral")
elf polarity < 0:
   print("negative")
elf polarity > 0:
    print("positive")
end procedure
```

1) Implementation

In our experiment we used python as our platform for implementing sentiment analysis and TextBlob as a prediction model.
Some packages have been utilized including tweepy, NumPy and TextBlob.The required libraries can be installed by following commands:

- Pip install tweedy
- Pip install NumPy
- Pip install text blob

To download the dictionary NLTK corpora we run the following command Python-m textblob.download_corpora. Tweepy is the python client for the official Twitter API. Text blob is used as a python library for text processing. Corpora is a large and structured set of texts that are stored and processed for analyzing tweets.

2) Connecting to Twitter APIs

In order to fetch tweets from Twitter API, one needs to register an App through their twitter account and define an application. The following steps are to be followed: Weneed to get to https://developer.twitter.com/en/apps and generate the APIs.

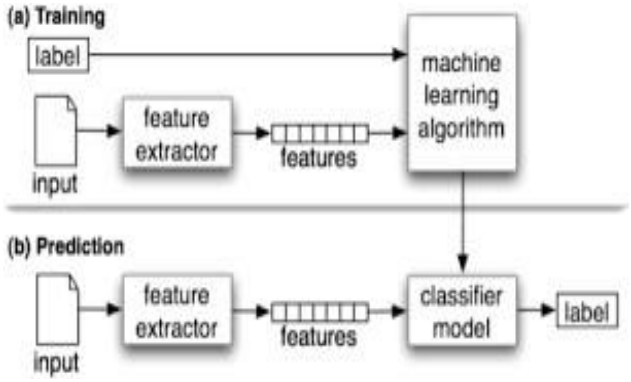
2. Methodology

In our experiment we have used TextBlob which is built upon Natural Language Toolkit. *TextBlob* is used for text data processing which is a Python library. It comes with an API, which provide solutions NLP tasks like POS(Part of Speech) tagging ,sentiment classification It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. We have used Python as our platform for implementing the data. Stock price values of a particular company ismentioned.Then sentiment analysis has been employed , the company being researched to perform sentiment analysis, to predict the Stock price value of that company, and as to how people are reacting about itis Google.

Here, a simple python library known as TextBlob,

TextBlob

a machine learning method is used as a dataset for processing textual data and it provides a simple API to access it's methods and for diving into common natural language processing tasks like extraction of live streams from twitter, Sentiment Analysis, Classification, Tokenization (splitting of text into words and sentences), Part-Of-Speech-Tagging, and more. Finally, sentiment score for tweets have been computed.

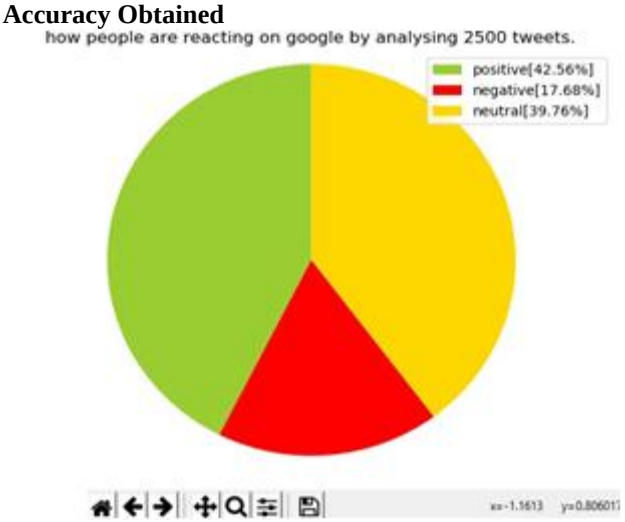


Data can be downloaded from the internet/extracted from twitter. It was done by live streaming on twitter using TextBlob as our model. The sentiment function of TextBlob returns two properties, Polarity and Subjectivity.

Polarity is a float value which lies in the range [-1,1], where 1 means positive sentiment and -1 is negative sentiment. Subjectivity is also a float value which lies in the range[0,1]. Subjective sentence expresses some feelings, views, opinions, allegations etc. [6].

TextBlob works as a framework for all the necessary tasks we need in the basic, NLP. The TextBlob analyzes the tweets for the keyword we want to perform sentiment analysis on, trains the model to classify the tweets and it will give us the tweet status as to how many people are feeling positive, negative or sayneutral about the company and the accuracy isknown.

Accuracy Obtained



As we observe from the results obtained that the percentage of positive tweets are significantly high. It is also important to mention that depending on the data obtained in the experiment we tend to get different results as people's opinions might change.

3. Conclusion

From the result of our experiment, there are a few things which can be concluded: Sentiment Analysis in our research mainly focuses on to predict stock values and people's opinion through tweets i.e., through the use of context, tone etc. We have mined data from twitter and trained a model TextBlob to process the live tweets and classify it into positive, negative and neutral. This trained model thenpredicts the rates of the stock markets. We realized that the percentage for positive tweets are significantly high. We havenoticed that the results would be more accurate if we studied people's sentiment in real.

4. Acknowledgement

I am extremely thankful to my mentor Dr. Vani Priya who guided me throughout my entire project work and motivated me for more.

References

[1] E. F. Fama, "The behavior of stock-market prices," Journal of Business, pp. 34-105, 1965.
[2] J. Bollen and H. Mao. Twitter mood as a stock marketPredictor. IEEE Computer, 44(10):91–94.
[3] Google Finance. Available from:https://www.google.com/finance. [cited 27th july 2017]
[4] Web scraping. Available from: https://www.webharvy.com. [cited 27th july 2017]
[5] M. A. Russell, Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. O'Reilly Media, Inc, 2013.
[6] Zheng Chen.Study of Stock Prediction Based on Social network, International conference of social Computing, 2013, 913-916.
[7] L. Bing, K. C. Chan and C. Ou, "Public sentiment

analysis in twitter data for prediction of a company's stock price movements," in EBusiness Engineering (ICEBE), 2014 IEEE 11th International Conference on, 2014, pp. 232-239.

[8] Bracey, L. (n.d.). The Importance of Business Reputation. Retrieved Juli 2015, from Business in Focus:http://www.businessinfocusmagazine.com/2012/10 /the-importance- ofbusiness-reputation