Stock Market Trend Analysis Using Twitter

Husam Chaudhary, Abdullah Sheikh & Prof. M Lutful Islam
Department of Computer Engineering
M. H. Saboo Siddik College of Engineering
Mumbai, India

Abstract: Technology is developing at a rapid rate. With the advancement in technology huge amount of data is being generated every day. This data comes from mainly social media sites such as Twitter, Facebook, etc. In this social media platforms people share their views or opinions about various range of subjects. This data can be mined and used for various purposes. In this work, sentiment analysis of stocks through data mining will be performed.

Keywords—Twitter, Stock Market, Big Data, Hadoop, Sentimental Analysis,

1. Introduction

Internet technology such as social media (e.g. Facebook, Twitter, Snapchat, WhatsApp, etc.) has been gaining huge popularity than ever before. Nowadays people of all age groups spend more time on social media than anywhere else. Among various social media platforms, Twitter has seen a massive rise in popularity. Twitter is a platform where people from all over the world share their views on various subjects such as trending news, politics, sports, stocks, gadgets, movies, etc. “Tweet” is the heart of twitter, every post posted on twitter is called a tweet. The tweet has a character’s restriction on it i.e. tweet should not exceed 140 characters. This format forces the user to be more concise about the topic they are posting. Hence tweet is a short, condensed information about a particular topic. This makes tweets easy to read and understand, easy to data mine and find various trends in various subjects.

Over the past few years Twitter has seen steep exponential growth. It has up to 1.3 billion registered users, Twitter sees 100 million users a day and about 500 million visitors a month. Approximately 60 million tweets are generated by Twitter users every day. The huge popularity of Twitter has led many celebrities, sportsmen, politicians, etc. to join Twitter and connect with their fans and increase their popularity. With millions of people tweeting about their various opinions about various subjects ranging from sim cards to smartphones, Twitter is a very rich source of real-time concise information regarding current trends and opinions.

Study of Behavioral economics tell us that people are not logical consumers and individual decision making is greatly affected by the opinions of others. So, if each tweet is considered as a brief summary of a person’s mood or opinion about a particular subject, then the aggregate of tweets about that subject should tell us the collective mood. Although market news (Quarterly results, performance, etc.) mostly influences stock prices public sentiments also plays an important role. Psychological researches have proven that emotions and information, play a important role in human decision-making. Behavioral finance has further proven that financial decisions are largely driven by emotions and mood. It is therefore safe to assume that the public mood and sentiment can drive stock market prices as much as news.

So, this project focuses on short-term prediction of stock price trend, and takes the approach of analyzing the time based indicators as features to classify trend (Rise and Fall). The user can give a stock name as an input and the system will give prediction of the input stock trend in near future.

We use twitter data to predict public mood and use the predicted mood and previous day’s stock values to predict the stock market movements. In this project, we will apply sentiment analysis to find the correlation between public sentiment and market sentiment.

2. Literature Survey

In reference [1] the authors attempt to examine Twitter’s predictive potential of consumers by observing the relationship between Twitter trends in a particular sector and stock prices of the top gainers and top losers of top companies in the that sector.

In reference [2] the authors presented the first indications that there can be a correlation between Twitter sentiment and the stock market. In their work, a sentiment score is correlated with the Dow Jones Industrial Average and then fed into a neural network to predict market trends. The authors use a mood tracking tool named Opinion Finder to measure mood in 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). Then, they correlate the mood time series with DOW JONES INDUSTRIAL AVERAGE closing values by using a Self Organizing Fuzzy Neural Network. Using their techniques, they measured an improvement on DOW
JONES INDUSTRIAL AVERAGE prediction accuracy.

In reference [3] the authors divided selected tweets data and stock price data from the same period into different time slots by ‘unit time’, and used SVM and Naïve Bayes algorithms. Their work included four main steps. 1. Data collection from Twitter and NASDAQ website. 2. Keywords selection. 3. Optimization of the number of keywords from 2, and unit time length 4. Checking prediction using both learning algorithms.

In reference [4] the authors apply sentiment analysis and machine learning principles to find the correlation between” public sentiment” and” market sentiment”. They use twitter data to predict public mood and use the predicted mood and previous days’ DOW JONES INDUSTRIAL AVERAGE values to predict the stock market movements.

3. Proposed System

In our system, we will extract stock market data from Twitter API’s by applying keywords and store it into HDFS (Hadoop filesystem) using Flume or Hadoop commands. We will then perform processing of extracted data and remove redundant data and finally perform sentiment analysis on the data.

The Tweets with higher no of likes or retweets will have more weightage in determining the overall score of the sentiment analysis. As this will give us opinion of more number of people (the one who only likes or retweets a tweet).

![Fig.1.Data Storage and Analysis](image1)

The workflow of the project is as shown in the fig below.

A. Working of the system

B. Algorithm

Input:
D: Dataset of Stock Market
Output:
Buy or Sell Recommendation
Processing:
1. Read Dataset D.
   a. Enter the keyword for extracting live tweets.
   b. Store the extracted tweets on HDFS.
2. Perform Pre-processing
   a. Removal of raw data such as @, #, http removal.
   b. Remove all slang words.
   c. Remove stop words.
3. Classification:
   a. Classify words based on positive and negative dictionary.
   b. Calculate the frequency of each word in the document.
   c. Classify the document and label them as positive, negative, and neutral based on the frequency of word present in each line of the document (i.e. more positive or more negative).
   d. Give more weightage to tweets with more likes and retweets.
   e. Calculate the sentiment score for each document.
   f. Calculate the sentiment score for the entire dataset.

4. Display the output of the sentiment analysis.

C. Expected Results

After getting the sentiment results we can predict what stocks to buy and what to sell.

4. Conclusion

Many individuals develop their own methodologies to increase the probability of making a profit in their stock investment. The overall success rate of these methodologies are generally too low to be practical for real-world application. Using our system users can predict short-term trend of stock and has potential application for personal investment. Opinions play a central role in almost all human activities, because it influences our behaviors and decisions. Therefore, whenever we make a decision, we want to know other's opinions. Using our system users can enter stock name and get overall sentiments of different users about the stock. Tracking the public's sentiment by gathering online information about the market and stocks can be valuable in creating various trading strategies.

5. Future Scope

The limitation of our project is the amount of data available about stocks on twitter which is quite low. To get more accurate results, we need huge data and that is a drawback of our system. These remain as area of future research.

6. Acknowledgement

We wish to express our sincere thanks to our Principal Mr. Mohiuddin Ahmed, M.H. Saboo Siddik College of Engineering for providing us all the facilities, support and environment to meet our project requirements. We would also take the opportunity to express our humble gratitude to our Head of Department of Computer Engineering Mr. Z.A. Usmani for supporting us in all aspects. We are highly thankful to our internal project guide Er. Lutful Islam whose invaluable guidance helped us understand the project better and helped us to complete the project successfully.

7. References

[1] Predicting Stock Market Fluctuations from Twitter. Sang Chung & Sandy Liu [Berkeley, California]