



Twitter Data Predicting Stock Price Using Data Mining Techniques



Umang Patel

Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT 06604, USA

Abstract

In this project, we apply sentiment analysis and data mining techniques to discover the correlation between “public sentiment” and “market sentiment”. We use twitter data to predict public mood and use the predicted mood and previous days’ NASDAQ values to predict the stock market movements. We foremost look for a correlation between twitter sentiment and stock prices. Secondly, we determine which words in tweets correlate with changes in stock prices by doing a post analysis of price change and tweets. Also, we discover the relationship between tweets of the vital Twitter user related to the stock and the corresponding one stock price behavior. Lastly, we will try to analyze trending mood on twitter of Top Gainers and Top losers. We achieved this by mining tweets using Twitter’s search API and subsequently processing them for analysis using Sentimental Analysis. For the task of determining sentiment, we test the effectiveness of three data mining techniques: Naive Bayes classification, Decision Tree, and Text Mining.

Introduction

Prediction of stock has been the most active area of research in this era. Whereas, prediction of stock price is an extremely complex and very challenging task because there are too many factors involved such as

- ❖ Economic circumstances,
- ❖ Political events, and
- ❖ Other environmental factors

This factors which may impact the stock price. Historically, stock market price and movements have been highly unpredictable due to lack of technologies and availability of data. In this era of data, due to the availability of data and techniques to use the data has been evolved rapidly. Twitter data is the solution for the problem which exists in predicting stock. The factors discussed above like economic circumstances, political events, and other environmental factors which may impact the stock price are all available in twitter data. Twitter sentimental analysis can be extremely helpful for predicting emotions or opinion of certain product. So examining Twitter’s predictive potential of consumer purchasing and opinion by observing the relationship between societal Twitter trends in the technology sector and hourly stock prices of the top gainers and top losers of companies in the technology sector. Using Data Mining Techniques like: Naive Bayes classification, Decision Tree, and Text Mining, we can hypothesize that the trending mood in Twitter about the top gainers in the stock will be positive, while the trending mood about top losers will be significantly more negative compared to a baseline measurement of the trending mood in the stock market

Results

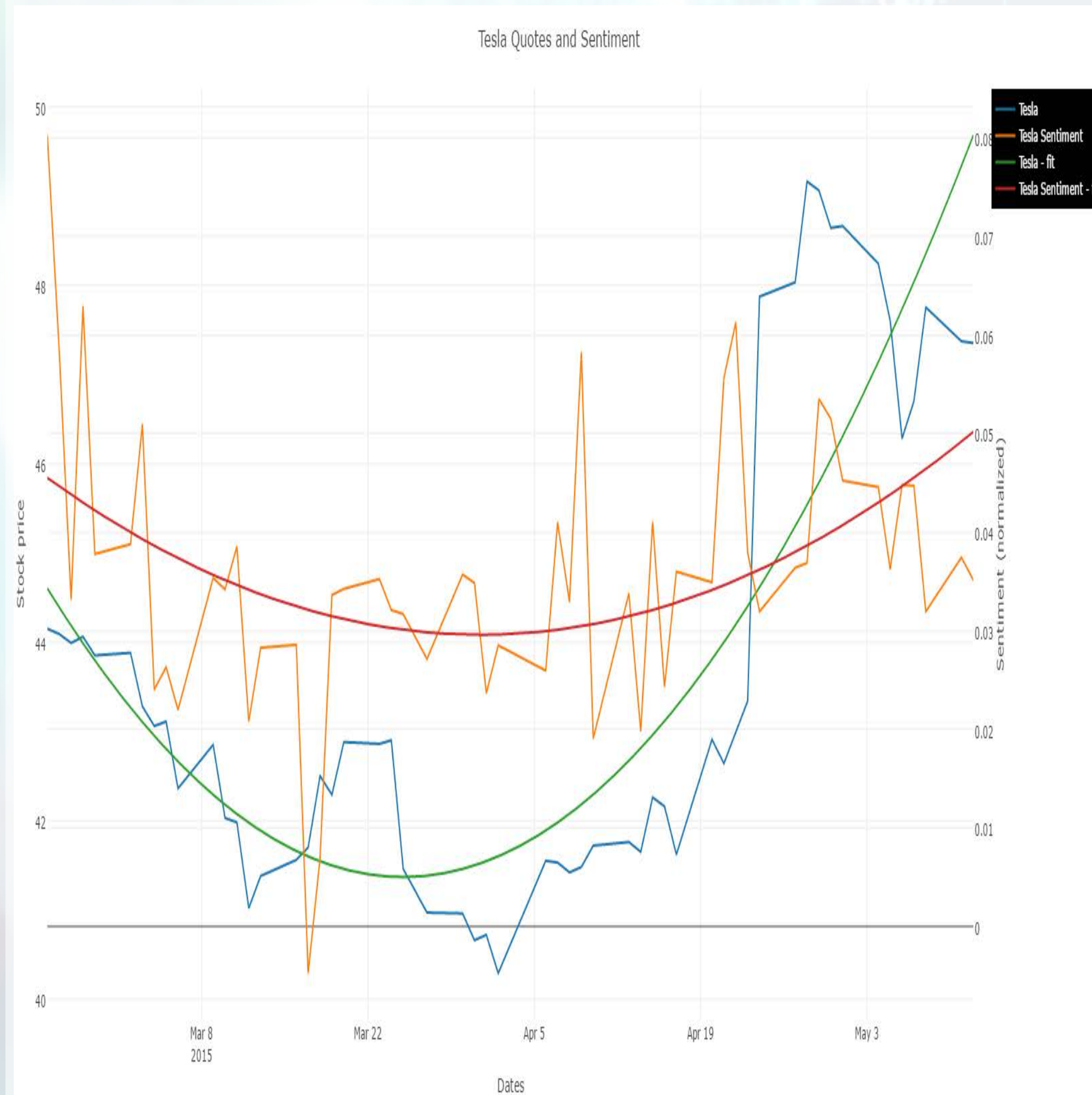


Fig. 1 Output

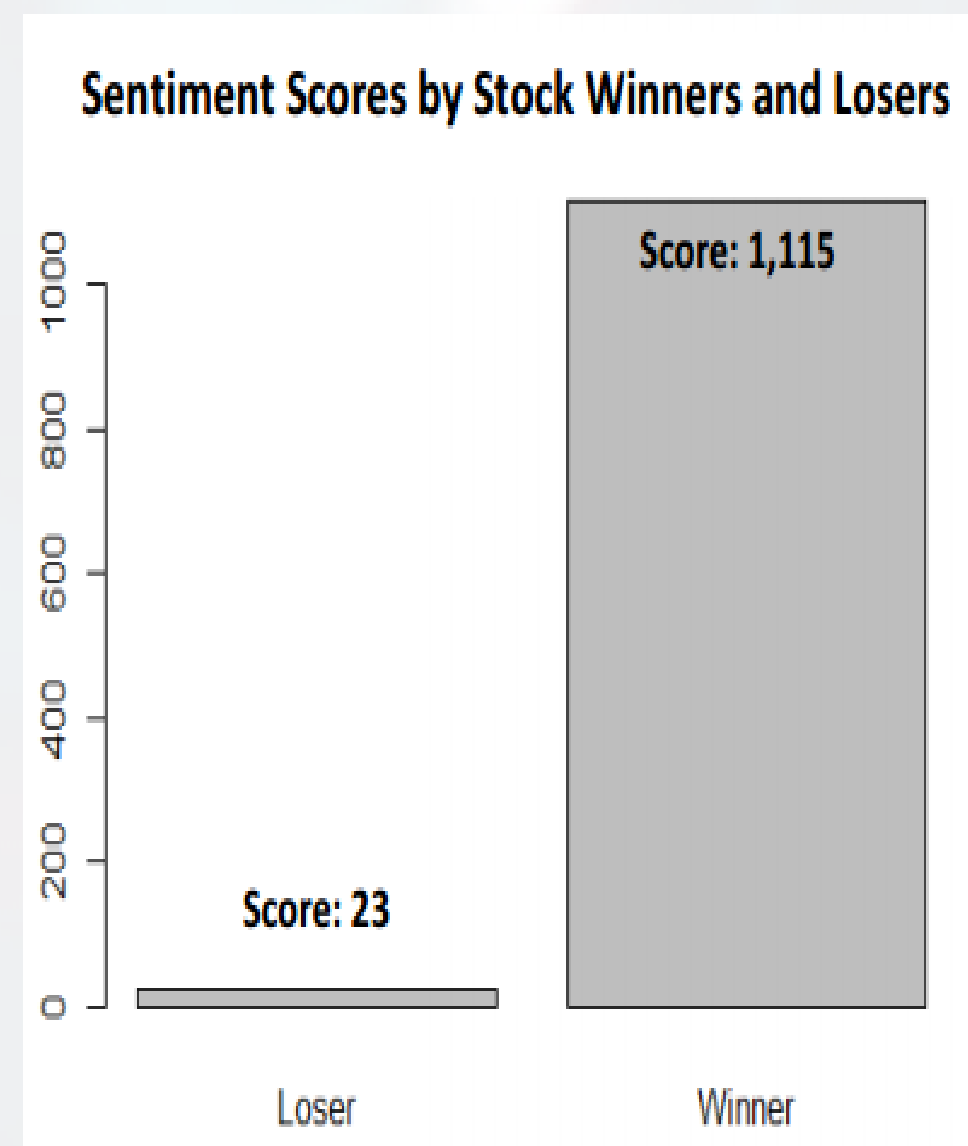


Fig. 2 Trending mood for Gainer and Loser



Fig. 3 Word Cloud – Image Courtesy: <https://www.jasondavies.com/wordcloud/>

Conclusion

We conclude that with the help of this project, the twitter sentiment score can be predict the movement of stocks if the sentiment is trending positive, not negative. However, stock price movements are more strongly predictive of twitter sentiment movements. There is no significant predictive power of trending negative sentiment scores on stocks relating to the subject. In this project, we will try to evaluate all the problem definition defined. Primarily, the goal is to find out the correlation between the Twitter Sentiments and Market Sentiments. Secondary, the importance of the Vital Twitter user and his influence on the stock. This analysis will be helpful, if successful to various investors. The result may not be perfect due to political events, general economic conditions, and investors’ expectations influence stock market.

Data Mining Techniques

- 1. Text Mining:** Text mining, also referred to as text data mining, roughly equivalent to text analytics, refers to the process of deriving high-quality information from text. Sentimental Analysis, one of the typical text mining tasks will be use for twitter data analysis.
- 2. Naive Bayes classification:** A Naive Bayes classifier is probabilistic classifier based on Bayes Rule, and the simplest form of a Bayesian network. The classifier is an application of Bayes Rule:

$$P(c|F) = \frac{P(F|c)P(c)}{P(F)}$$
- 3. Decision Tree:** Decision tree (DT) algorithm is a data mining induction technique which recursively partitions a dataset of records using depth-first greedy approach or breadth-first approach until all the data items belong to a particular class.

Problem Statement

❖ Problem Definition I: Analysis on Movement of Stock based on Twitter sentiments

In this problem definition, we will try to evaluate the twitter sentiments and correlate it with the market sentiments. We will primarily focus on the positive, negative or neutral sentiments based on tweets of the particular day for a particular stock.

❖ Problem Definition II: Analysis on Trending mood of Top Gainer and Top Loser

In this analysis, we try to analyze the trending mood for the Top gainer and Top Loser. Analyzing trending mood in terms of positive and negative, ideally Trending mood for the Top gainer should be positive on twitter whereas opposite for the trending mood for Top Loser i.e. the negative mood on twitter.

❖ Problem Definition III: Analysis on Vital Twitter User related to stocks

In this analysis, we will find out the relation between the vital twitter user related to stocks. Vital user means the important twitter user which is related to the stock.

❖ Problem Definition IV: Word cloud for Stock change

We will try to create the word cloud which help us to easily analyze and help us to get the words which was most used during the top gains or top loss.