Analyzing Stock Market Price Based on Historical data and twitter sentiment Analysis

DSI Program
Capstone Project
By

Dereje Workneh

06.09.2020 GA-DC-DSI-CC-11

6/12/20

Outline

- INTRODUCTION
- PROBLEM STATEMENT
- Work flow
- METHODOLOGY
- RESULTS
- CONCLUSION AND RECOMMENDATION

Introduction

- Stock market prediction
- Stock market predication has always been an interesting topic among researchers.
- market data with public sentiment to predict market movement

Problem Statement

- Utilize social media to access market sentiment and predict the behavior of the apple

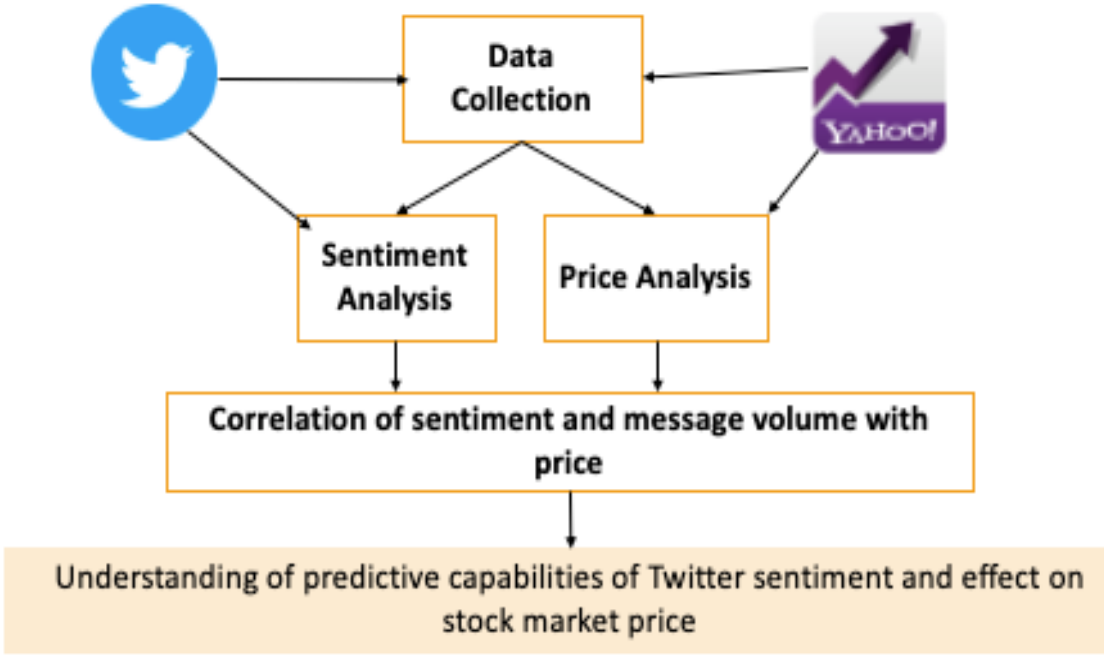
- Classify polarity of given text at document, sentence or feature level and determine whether opinion in text is positive, negative or neutral.

- Using ML algorithm to predict sentiment and find correlation between sentiment and stock price

Data Collection

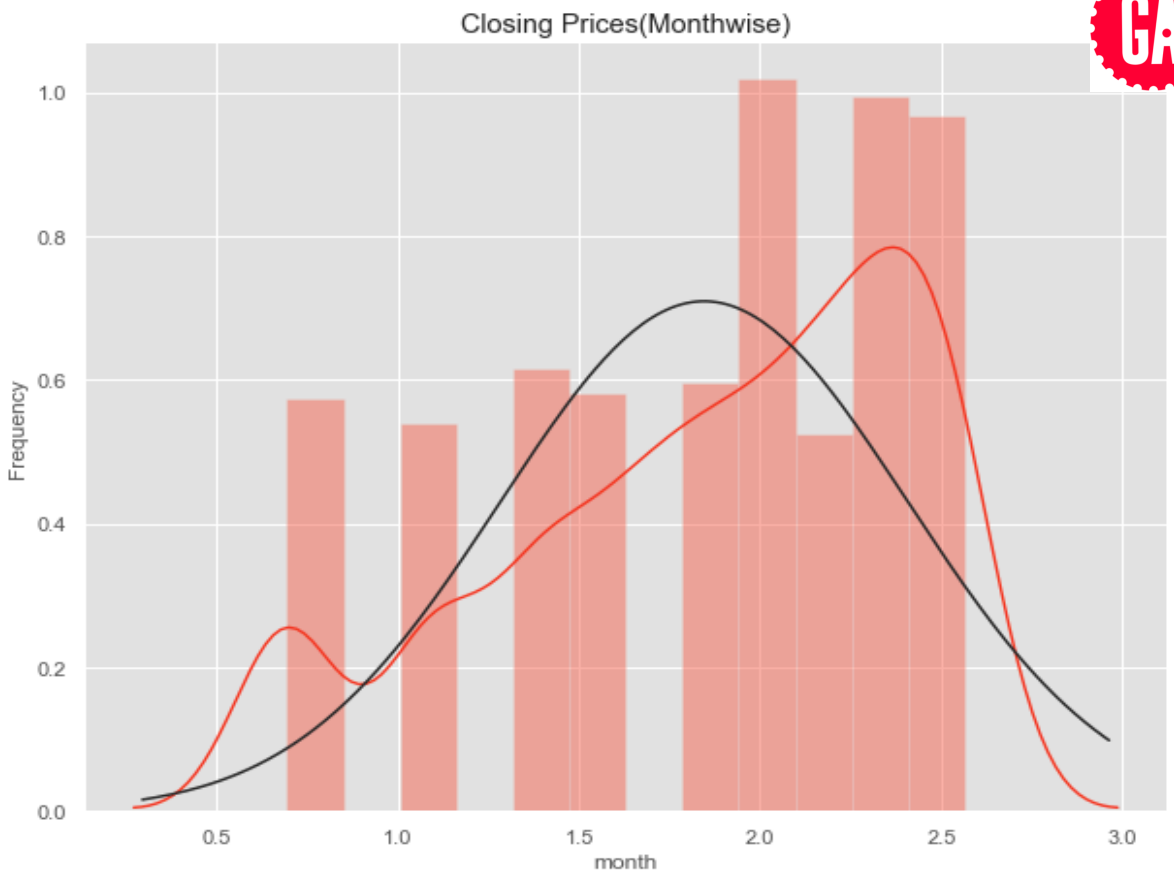
- Twitter is popular microblogging website
- Each tweet contains different characters in length
- Tweets varies in time (hours)
- Only text message (English)
- 10 tweets/day
- 5yrs
- Yahoo Finance data

Work flow



Exploratory data of Stock price

- Varies in with months
- Prediction of closing stock prices.
- It is a threshold that indicate the apple will go up or down in the next day



Exploratory data of Stock price

- The trend is highly non-linear
- It is difficult to capture the trend using this information

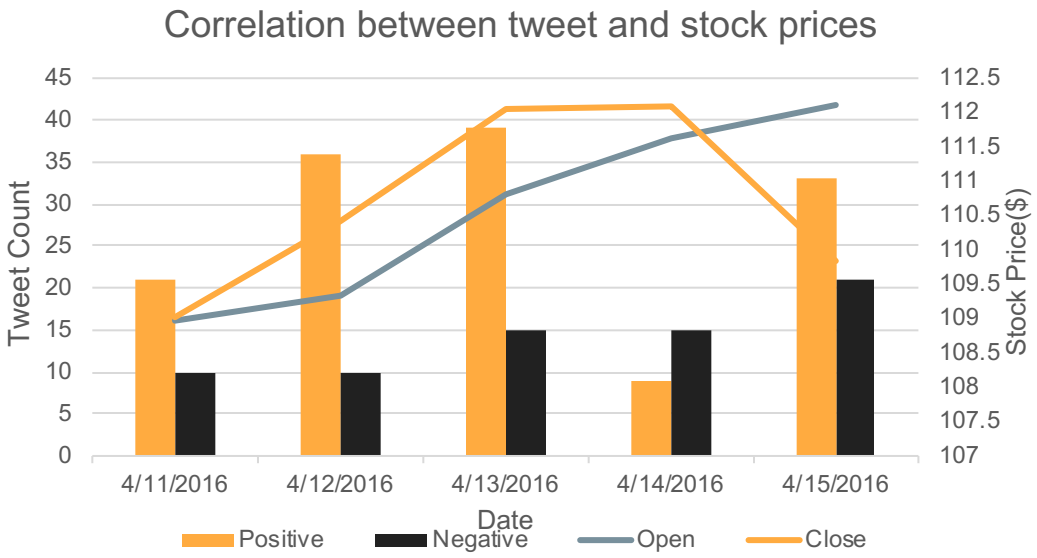
Result

- Tweet sentiment analysis
  - SVM
  - Naïve Bayes Classification
- Stock Price using tweet sentiment analysis
  - SVM

Model	Accuracy	Precision	Recall	F-Measure
Naïve Bayes Bernoulli	0.8039 (80.39 %)	0.4615, 0.45, 0.9710	0.75, 1.0, 0.7882	0.5714, 0.6207, 0.8701
Support Vector Machine	0.8431 (84.31 %)	0.6923, 0.65, 0.9275	0.5625, 1.0, 0.8767	0.6206 0.7878, 0.9014

Result

- Tweet sentiment analysis
  - SVM
  - Naïve Bayes Classification
- Stock Price using tweet sentiment analysis
  - SVM



Accuracy: 0.7876 (78%)

Conclusion and recommendation

- Addition of twitter sentiment analysis in stock prediction results appeared to improve the prediction.
- Supportive Vector Machine better predictor than Naïve Bayes B.
- Social media has an advantage for stock prediction.
- Larger dataset improve the prediction result.

THANK YOU!!