Stock Prediction using Twitter

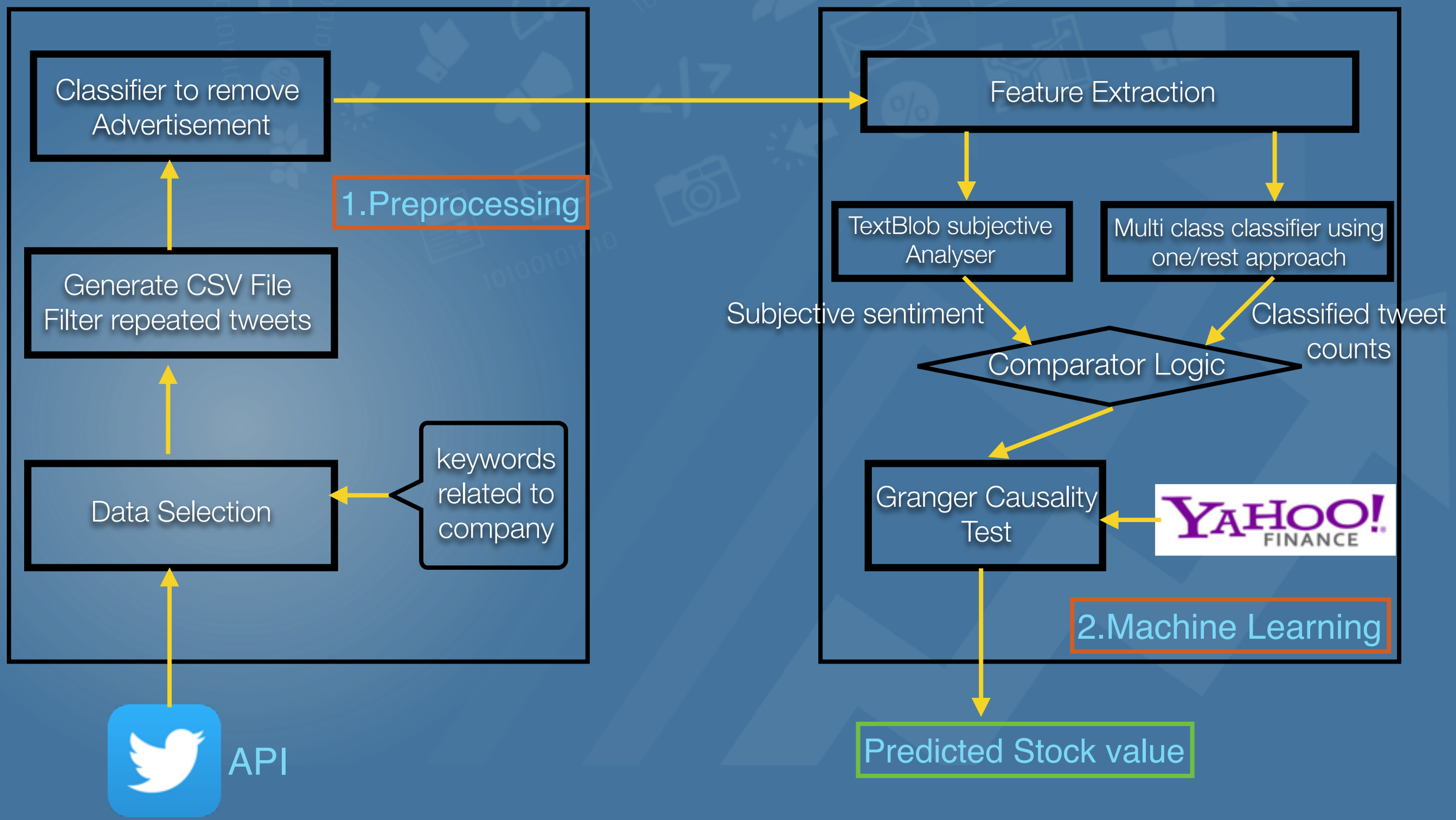
Aishwarya Anand (A20331867)
Vivek Vijaykumar Bajpai (A20361204)

Problem

A system to understand the twitter sentiment of a company and predict its stock value

Case study:
We have taken Apple Inc. for case study

Approach


Classifier to remove Advertisement
1.Preprocessing
Generate CSV File Filter repeated tweets
Data Selection
keywords related to company
API
Feature Extraction
TextBlob subjective Analyser
Multi class classifier using one/rest approach
Subjective sentiment
Classified tweet counts
Comparator Logic
Granger Causality Test
YAHOO! FINANCE
2.Machine Learning
Predicted Stock value

Data

Dataset

Category	Number of Tweets
Filtered Tweets	17K
Tweets without advertisement	9K

Labeled Dataset

Labeled For	Number of Tweets
Advertisement Classifier	700
Sentiment Classifier	800
Total	1.5k

Twitter Field Used : Text, creation timestamp,cashtag,screen_name

Yahoo Data : Closing stock value, opening stock value,

Result

Advertisement Classifier Accuracy = 76%
Sentiment Classifier Accuracy = 70%

Although prediction of stock values is very hard and it depends on numerous factors

During 10 days of Apple Inc. casestudy, we are able to predict the direction of stock values for 4 days

What went wrong:

- Stock values are depended on very large number of variables and we are considering very few of them. (eg.: companies finance, health of the sector etc.)
- Noise in Twitter data (advertisements and promotions, people asking questions to apple about TV shows, music etc.)
- Consumers (twitter users) may find some news bad, which may be considered as good news by investors (eg.:All iPhones booked for next two months)
- Limitation on amount of twitter data available.
- Very less number of negative tweets.

Conclusion

- With a data set of 9.7 millions tweets, the most famous paper on same topic could predict with an accuracy of 87% for 16 days, but the same system predicted with an accuracy of 3.6% for next 20 days.
- We may get the sentiment of the crowd with a good accuracy using social media, but predicting the exact stock value is still out of reach.
- The system does work really well on special days, when the company is in news for certain reason. (eg.: new product launch, Scam uncovered in management etc.), but on normal days, it is not reliable.

References:

1. J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. Journal of Computational Science, 2(1):1–8, 2011.

2. Sang Chung & Sandy Liu. Predicting Stock Market Fluctuations from Twitter. Dec 12, 2011

3. Social Media in Financial Markets: The Coming Of Age…

4. http://www.theatlantic.com/technology/archive/2010/10/predicting-stock-market-changes-using-twitter/64897/