Data 620 - Final Project Proposal

Team1: Ramnivas Singh, Tage Singh, Deepak Sharma

July 11, 2021

Proposal: Stock Market Prediction using Social Media Analysis

Methodology:

- To explore the positive and negative definition of investment directions based on the data about latest news from Stock Market in the United States.
- Use natural language processing function, like tokenize, NaiveBayesClassifier to provide efficient ways to analyze the scraping news data from website of Stock Market.
- It shows empirical results by using machine learning method to represent information from Stock Market's New, and simple naive Bayes classifiers to predict the direction of movements.

Motivation:

1. Use Social network analysis concepts to mine and generate graph dataset
2. Correlate centrality measures with the stock price movements
3. To use island method to create network graph for analysis
4. To use Naive Bayses classifier to predict the company in positive or negative direction by scraping News in Website of Stock Market
5. To see whether news sentiment is associated with stock price movement by analyzing the time series data

Data source:

- The first data for this application contains total 46 cases of positive and negative definition from the web http://positivewordsresearch.com/sentiment-analysis-resources/#content
- Second data contains total 3389 cases of positive and negative definition from Twitter
- The dataset is from scraping the latest news for the specific required stock shares or index. https://www.nasdaq.com
- The dataset of stock close price is from scraping the website of Nasdaq.

Deliverables:

1. Scrap the latest news data for comparing the definition data of positive and negative.
2. Analyze the definition data of positive and negative from twitter.
3. Obtain the positive or negative trend from the latest news of specific index or stock shares by natural language processing function - classify.
4. Test the accuracy by different features and choose the best model.
5. Calculate Score of document = count (pos.matches) – count (neg.matches). If the Score is 0 or more, we consider the document is positive or else, negative. Then compare the movement sentiment score and stock price.

Work Plan

- **Proposal:** Ramnivas Singh, Tage Singh, Deepak Sharma
- **Data Acquisition & Preparation:** Tage Singh, Ramnivas Singh
- **Data Exploration**: Deepak Sharma, Tage Singh
- **Network Design & Analysis:** Deepak Sharma, Ramnivas Singh
- **Sentiment Analysis:** Ramnivas Singh, Tage Singh, Deepak Sharma
- **Conclusion:** Ramnivas Singh, Tage Singh, Deepak Sharma
- **Final Presentation:** Ramnivas Singh, Tage Singh, Deepak Sharma

Concerns

1. Tweets to analyze are from the past and are in limited numbers. So generated predictions can may be slightly different as compare to actual data
2. Social network analysis need high computation hardware. We may have to reduce large dataset to small manageable dataset