Final Project Proposal

**- Team information**

Yiziyin Yao, 001341398 Sichen Zhao, 001051127 （Team 15）

**- Background and statement**

Nowadays, there are many factors that affect the movement of stock prices, including news. It is not difficult to find that different news will also affect the stock price. Negative news will normally cause people to sell stocks. Positive news will normally cause individuals to buy stocks. Good earnings reports, an announcement of a new product, a corporate acquisition, and positive economic indicators all translate into buying pressure and an increase in stock prices.

By analyzing news data to predict stock prices, we have a unique opportunity to advance the state of research in understanding the predictive power of the news. This power, if harnessed, could help predict financial outcomes and generate significant economic impact all over the world.

In order to find out the relationship between news headlines and the movement of stock prices, we collected a large number of datasets. Therefore, for more efficient data processing, it is necessary to parallelize our code to get a better performance.

**- Datasets:**

1. News Headlines from January 2012 to May 2018

There are near 200 thousand news headlines with categories. We found it at

https://www.kaggle.com/rmisra/news-category-dataset

2. S&P 500 data from January 2012 to May 2018

3. Some news headlines to support our claim

**- Objectives:**

The primary purpose of this project is to figure out whether it have any connection between the daily news and the stock price, of course, we try to deal it parallelly.

First of all, we need to cleaning data, we will connect the news and stock data together (because news headlines exists in everyday, but stock market only has activity in trading days).

After cleaning data, we will separate all headlines into word and vectorized the word, and separate them into training group and test group.

Then we put it into different models such as SVR, SVC, Random Forest, Logic Regression to get the result.