**Project Outline**

**Problem motivation:**

In the stock market, the stock price of each publicly listed company is ever changing, it is difficult for people to predict where the stock market is heading because there are many complex factors that could influence the trend. In this project, I am planning to use information extraction techniques on articles related to the targeted company, perform sentiment analysis on the data acquired, to predict its share prices.

**Possible Solutions:**

**Using Machine Learning algorithm to learn and predict stock prices**

**Feature gathering:**

1. Develop an NLP system that analyse financial articles related to the targeted company. Extract meaningful words which holds a sentimental value, or words that can be used as an indicator as to whether the company is performance well or worse. These words will make up a large bulk of features.
2. The past stock prices may be used as output values in a training set. Timeline of the data gathered is important because they must correspond to the feature values in the respective training examples.

**Construction of Training sets:**

* 1. Each training example will have the features (x(n)) ­­gathered in step 1) from above, and their respective stock prices (y) as the output value.
  2. Perhaps training examples may be distinguished by date, so forming a training set (in matrix form) where each row represents a training example of a particular date. Each column represents a particular feature, the output value y will occupy the rightmost column.

**Supervised Learning:**

1. Training set may be fed into a supervised learning algorithm to predict certain outcomes:
   1. Linear Regression: Predict stock prices (difficult to produce a precise figure that has high accuracy)
   2. Neural Network (Classification problem): Whether the stock price is likely to Rise or Fall (a general prediction of rise or fall in stock price may be more reliable)

**Research (to do):**

1. Understand the mechanisms of the stock market
   1. Examine the current popular methods of predicting stock prices, what are the main variables other researchers like to consider?
2. Identify the suitable timelines between articles and the date of the stock prices, for example will today’s new articles provide an insight to today’s, or tomorrow’s, or the day after tomorrow’s stock trends? “Expiration date” of an article’s usefulness.
3. Find different sources of articles
4. Mathematical model used to calculate the company’s value
5. Are there stock market API’s that would allow fast and easy extraction of data.

**Tools (software/programming language):**

NLTK

Octave/Matlab

Python