Design Summary

**Abstract**

Good news may raise the stock price and bad news may depress the stock price. Our algorithm focuses on analyzing text patterns and information behind the news to predict the moving of Apple’s stock price.

Our design can be summarized into following sections;

**1.Preprocessing**

To start with, we used 24,363 articles related to Apple from wall street journals and half year stock price as our data source. Since the document term matrix has high sparsity, we used LASSO method and Principal component regression to reduce the dimension and divide it into training and testing set.

**2.Sentimental Test**

Instead of using all words, we constructed a document term matrix of only sentimental words. After preprocessing the data including cleaning and extracting stem words, we applied sentimental test on data. For more details, we recognized and classified our words as positive and negative. Intuitively positive words stimulate the stock price to grow, and negative words make the stock price decrease.

**3.Training and Testing**

During the training procedure, we split our data into validation group and training group using folds validation. We used principal component regression and lasso regression algorithm to train our data with our terms as covariates and stock price change as labels. We tested on the validation test with mean square error.

**Result**

After applied the above-mentioned strategies, we obtained many key words for stock price prediction in headlines, articles and also sentimental words. We also fitted models with relatively low mean square of errors.