

CS 483 Machine Learning

Team: Swag

Group Project 1

Group Members: Wei Lei, Adam Pipes, Sravani Ravula

- **Model Chosen:**

- Linear Regression: Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between the dependent and independent variables, they are considering and the number of independent variables being used.
- Random Forest Regression: To perform the prediction using the trained random forest algorithm we need to pass the test features through the rules of each randomly created tree. Suppose let's say we formed 100 random decision trees from the random forest. Each random forest will predict different targets (outcomes) for the same test feature.
- Lasso Regression: Lasso regression are some of the simple techniques to reduce model complexity and prevent over-fitting which may result from simple linear regression.

- **Time Frame:**

After we have looked at the data and did a visual analysis, some stocks date were chosen on specific dates to eliminate outliers or more closely follow current stock trends for all the ten stocks we picked. So below are the stocks with there time frames.

**Stocks - Date Frame**

APPLE - 2009-03-02  
AMAZON - 2014-10-01  
GOOGLE - 2004-08-19 (MAX Range)  
FACEBOOK - (MAX Range)  
MICROSOFT - 2012-12-01  
TESLA - 2013-04-01  
TWITTER - 2013-07-11 (MAX Range)  
GE - 2016-12-01  
AT&T - 1983-11-21 (MAX Range)  
INTEL - 2001-03-01

- **Source Code:**
  - .ipynb file - includes a summary of date ranges chosen for stock companies. Essentially stocks with aggressive changes in their trends were limited to the most recent long-term trend.
  - .pdf file of the notebook.
  - The program reads CSV files based on their default names and outputs CSV files for all with predictions for our date range.
  - Graphs and the trend lines for historical data also outputted.
- **Historical Data:**
  - The data from Yahoo Finance.
- **Historical Graphs:**
  - Graph and prediction lines for actual data.
- **Prediction Data:**
  - Output prediction data from the program
  - RandomForest Prediction Folder
  - Lasso Prediction Folder
- **Team Job :**

Project Overview is that, as a team, we have tried three different models to come with the best-predicted model. We used the same data set from the 10 different stocks with the same time frame and making sure everything else is contact expects the prediction models to continue with the best model for the next three-week prediction.

Adam: Linear Regression

Sravani: Random Forest Regression

Wei: Random Forest Regression/ Lasso Regression

- **Conclusion:**

After all three approaches on the data with the companies and the time frame being constant, we discovered drawbacks of using different regression models on the stocks data set and got a deeper understanding of these models in practice. Some of the issues we observed are

Linear modeling: using the full historical data is that it throws the linear regression off so it is not very accurate. There is a single row in the Apple dataset which had NULL values. I manually removed from the file just to simplify things. Apple and Amazon both had sudden growth patterns that emerged after being public for years. The values (low/high/open/close) tend to be very close as well so at the range of all history the graphs look very similar. Comparatively, Lasso Regression and Random Forest Regression prediction values looked better and we would like to continue the project and observe the next 3 weeks and decide which module has a better result.

