# MLR\_Stock\_Predict

to use multi-linear regression to study each individual factor contributing to the stock performance within 3 months' time.

The code will be written in python language.

For stock data, we downloaded from https://www.zacks.com/screening/stock-screener, as it allows you to export the data to .csv file.

The selection criteria uses in our project is as follows:

Time as of Nov 30/2016

Market Cap >= 100 million

Avg Volume >= 10000

Beta <= 3

P/E(Trailing 12 Months) <= 150

Current Ratio < 10

Current ROE(TTM) >= -100

12 Mo Trailing EPS >= 0

Price as a % of 52 Wk H-L Range > -100%

% Price Change (12 Weeks) > -100%

There are 2718 stocks being screened out through the criteria. And we split the data into train data and test data with about 80% to 20% ratio.

We train the multi-linear regression model with train data and generate the pred data based on test set.

Meanwhile, we are trying to find out which are the most critical factors that would contribute most to the % Price Change (12 Weeks).

The method that we used to optimize our model is backward eliminination (https://en.wikipedia.org/wiki/Stepwise\_regression).

After optimized, the most critical factors are:

market cap (x1)

average volume within one day (x2)

current ratio (x3)

Price as a % of 52 Wk H-L Range (x4)

summary table:

==============================================================================

coef std err t P>|t| [95.0% Conf. Int.]

------------------------------------------------------------------------------

const -20.6866 0.739 -27.995 0.000 -22.136 -19.238

x1 -3.612e-05 7.93e-06 -4.553 0.000 -5.17e-05 -2.06e-05

x2 2.037e-07 6.83e-08 2.982 0.003 6.98e-08 3.38e-07

x3 -0.2819 0.160 -1.758 0.079 -0.596 0.033

x4 0.3669 0.009 42.851 0.000 0.350 0.384

==============================================================================

From the model, we can draw the multi-linear liquation as follows:

% Price Change (12 Weeks) = (-20.6866) + (-3.612e-05)\*(market cap) + (2.037e-07)\*(average volume within one day) + (-0.2819)\*(current ratio) + (0.3669)\*(Price as a % of 52 Wk H-L Range)

Definitely, this model will not be the best one and it is just for study and research purpose.