



# DSC Boston

**Mar 2019**

Quarterly Stock Performance Prediction

BV Nithish A  
Shreyas C  
Ravi C  
Jatin K  
Ileshan V



# Solution Approach

- Data Exploration
- Data Preparation
  - Cleansing
  - Feature Engineering
    - Removal of blank/invalid, highly correlated features
    - Handling missing values
    - Skew handling
    - Standardization/Normalization per Period (Quarter)
    - Feature Extraction: 'Quarter'
- Modeling
  - Shortlisting of feasible/best possible model for task at hand and evaluation against each
  - Ranking code for conversion of regression score into quantiles by quarter



# Feature Importances

	Feature	Importance
105	10	22.6191
104	07	21.9214
103	04	17.4153
102	01	17.3335
101	Gr2Ind_Var11	17.0959
100	Gr2Ind_Var10	12.4031
99	Gr2Ind_Var9	11.1703
98	Gr2Ind_Var6	10.9819
97	Gr2Ind_Var5	8.13196
96	Gr2Ind_Var3	7.885
95	Gr2Ind_Var1	7.65166
94	Gr1Ind_Var104	7.61696
93	Gr1Ind_Var103	7.533
92	Gr1Ind_Var102	6.74913
91	Gr1Ind_Var101	6.14073
90	Gr1Ind_Var100	5.54866



# Validation Approach

- Validation dataset chosen based on the dates closest to the test dataset.

## Regression

- Grouped the data by date, sorted the prediction to divide the data into 20 groups. Calculated correlation of the predicted group with ground truth.

## Classification

- Correlation done based on predicted class and ground truth.



## Results Comparison - on validation set

Model	Correlation
Lasso Regression	0.036
SVM	0.062
Random Forests	0.055
MLP (Regression)	-0.009
MLP (Classification)	0.033
LightGBM	0.02



# Appendix

Code: <https://github.com/abvnithish/Data-Science-Challenge-Boston-Mar-2019>