



Predicting the Stock Market

Taylor Rohrich (trr2as) Alex Stern
(acs4wq) Chirag Kulkarni (ck3fz)



Executive Summary

Problem

Predict the short-term direction of a stock price using both technical and fundamental analysis

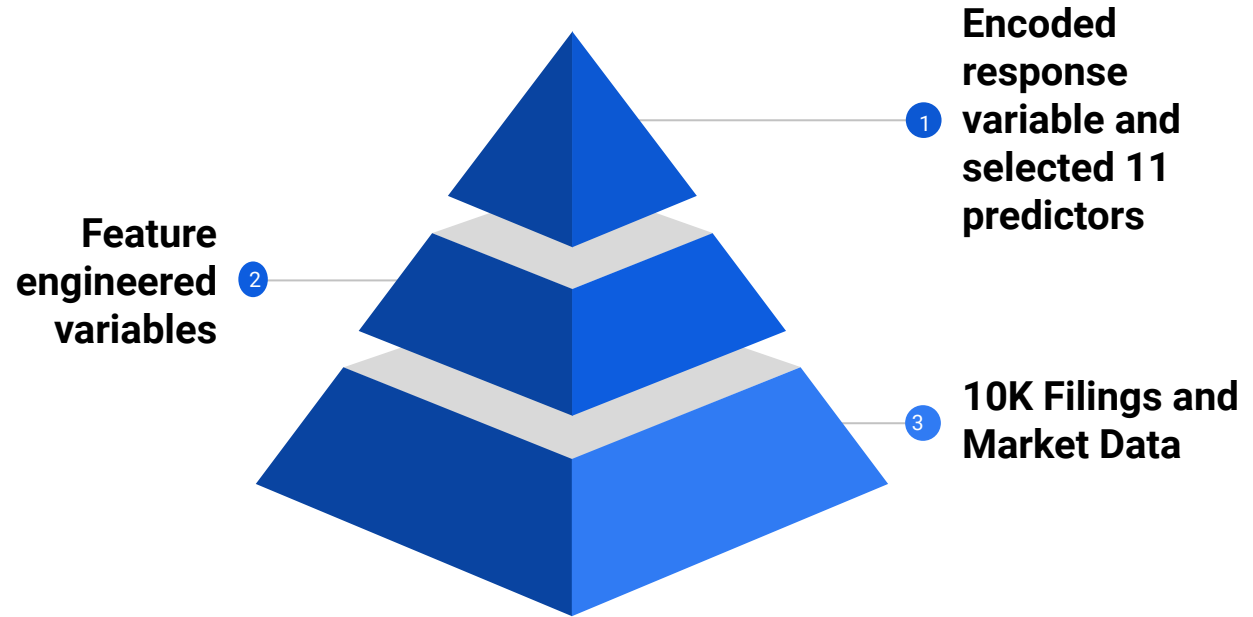
Model

Created three models to compare performance: neural network, logistic regression and gradient boosted tree

Results

Gradient Boosted Tree model led to final accuracy of 77.9%

Data Summary



Transformations & Preprocessing

- 70-30 Train-Test Split
- Use of Several Custom Transformers
 - `changeColumnType`: Modify data types, `responseVariable`: create response variable
- Pre-built Transformers
 - `Imputer` - fills NA's
 - `Indexer & Encoder` - vectorize and one hot encode categorical features
 - `getFeatures` - utilize `VectorAssembler` and create features column

```
pipeline = Pipeline(stages=[changeColumnType,imputer,indexer,responseVariable,encoder,getFeatures,gbt])
```

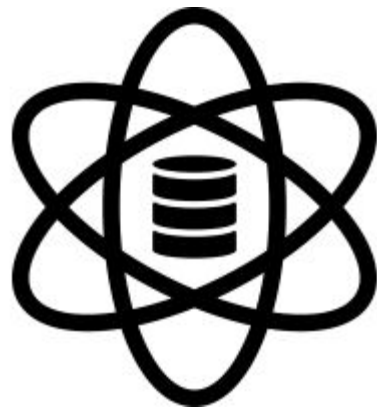
Models Constructed

Models

- Logistic Regression
- Multilayer Perceptron (Neural Network)
- Gradient Boosted Trees

Fine-Tuning Procedure

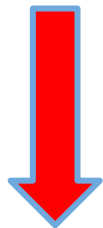
- Logistic Regression: max iterations, elastic net value
- MLP: max iterations, block size (batch size)
- GBT: max bins, max depth



Model Performance

	predicted 0	predicted 1
actual 0	0.0	104067.0
actual 1	0.0	140817.0

LR + NN



	predicted 0	predicted 1
actual 0	65571.0	38496.0
actual 1	15588.0	125229.0

GBT



- Confusion Matrix
- Champion Model - GBT
 - 77.9% accurate
- Sensitivity Analysis

Changing feature by one standard deviation

	LR	DL	GBT
gross_profit	0.576032	0.576032	0.781389
long-term_debt	0.576032	0.576032	0.781728
open	0.576032	0.576032	0.781058
close	0.576032	0.576032	0.781585

Conclusions & Future Research



Time Series



Cross Validation



Text Analysis



Industry Specialization

Questions