

FINANCIAL MODELING OF THE EQUITY MARKET

David Duemig
10.02.2020





— Question

- › “Starting with 27 indicators, can we determine the direction of the stock market?”

Data

1 Raw Data

- › 15,117 observations from 1960-01-04 to 2020-01-27
- › 27 Indicators and 1 Target variable
- › Unbalanced dataset
(53% Ups and 47% Downs)

2 Processed Data

- › Consideration of a balanced dataset

3 Model Data

- › Train/Val/Test Split

1. Iteration



... last (8.) Iteration



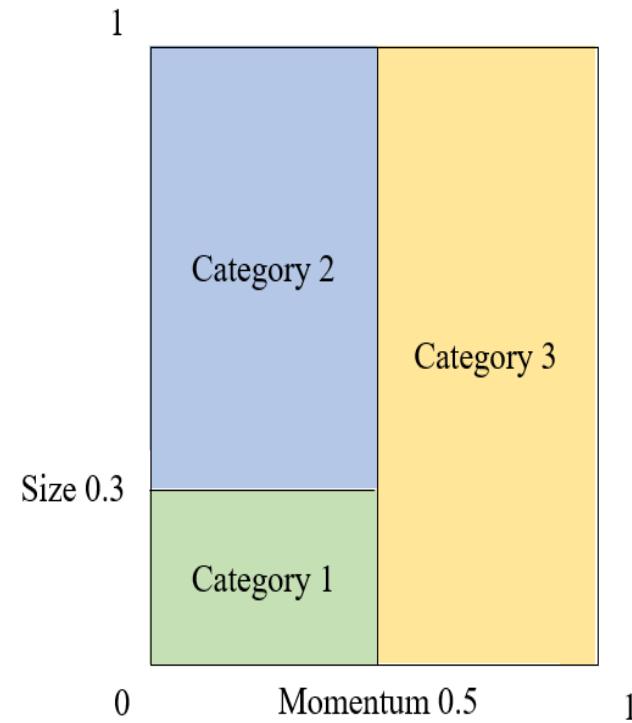
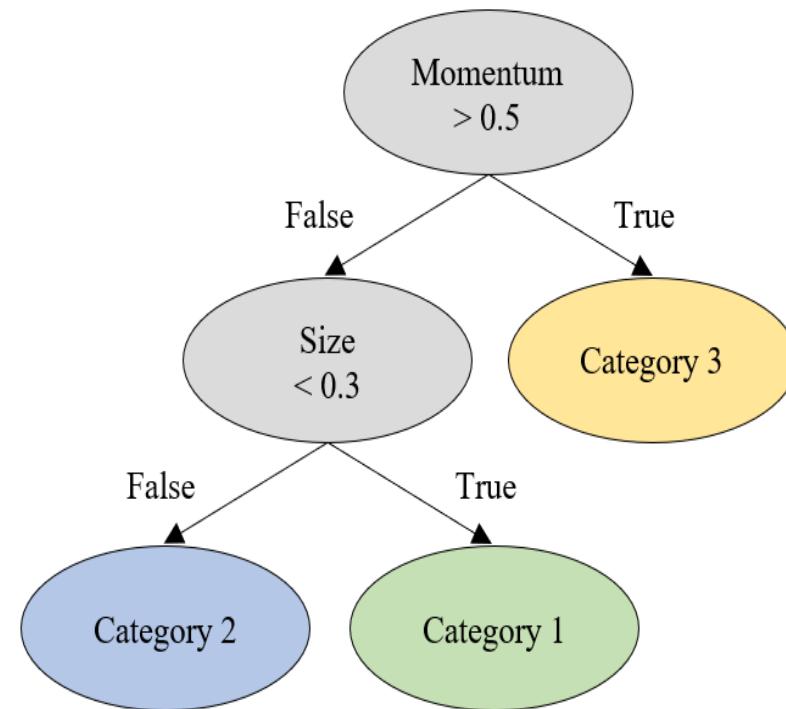
Refit (using the best parameters found)



Random Forest



Tree-Based Methods





Model Assessment



Accuracy

$$\frac{\text{\# of correct predictions}}{\text{\# of predictions}}$$



Confusion matrices

True Negatives (Predicted = 0, Actual = 0)	False Positives (Predicted = 1, Actual = 0)
False Negatives (Predicted = 0, Actual = 1)	True Positives (Predicted = 1, Actual = 1)

$$\text{Precision (0)} = \frac{TN}{TN + FN}$$

$$\text{Recall (1)} = \frac{TP}{TP + FP}$$

$$\text{Recall (0)} = \frac{TN}{TN + FP}$$

$$\text{Recall (1)} = \frac{TP}{TP + FN}$$



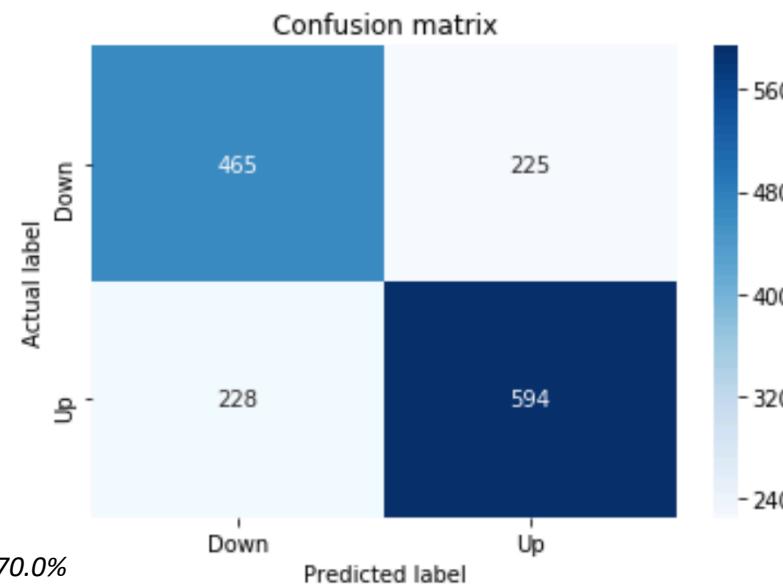
Receiver Operating Characteristic Curve

Plots true positive rate against false positive rate

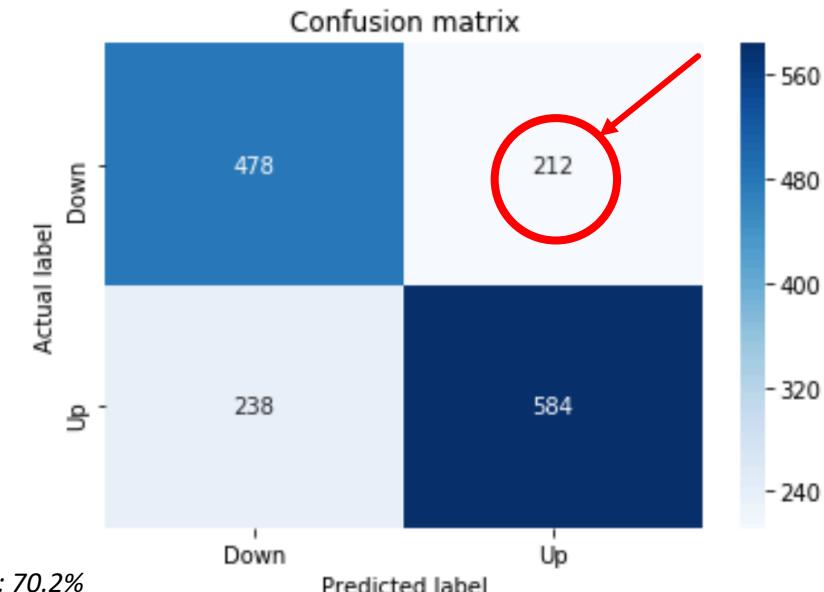
Model Assessment



Moderate strategy Unbalanced data



Conservative strategy Balanced data

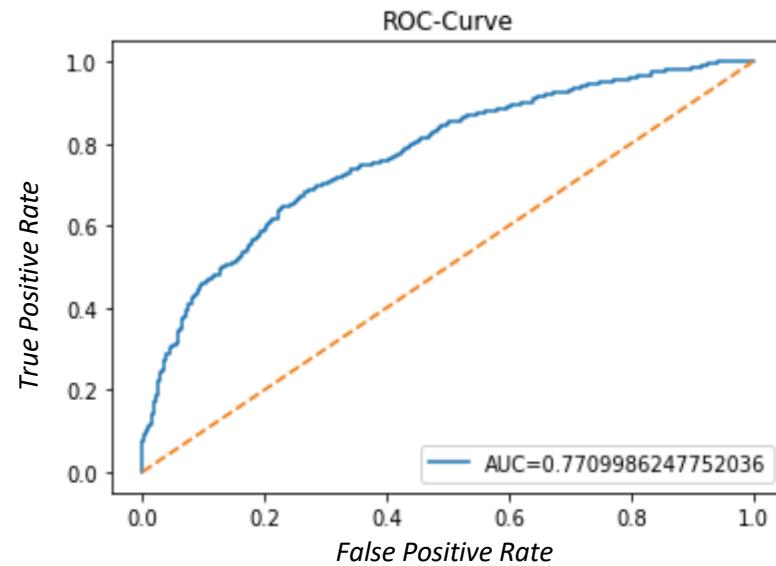


Model Assessment



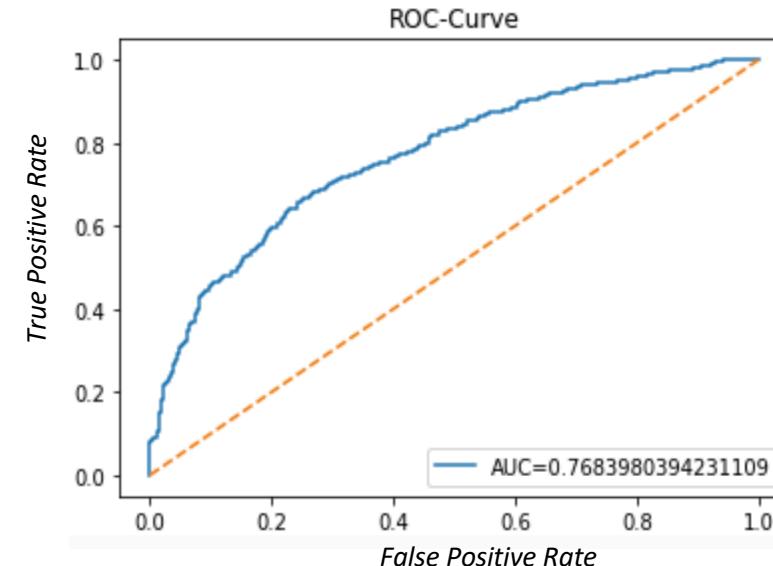
Moderate strategy

Unbalanced data



Conservative strategy

Balanced data



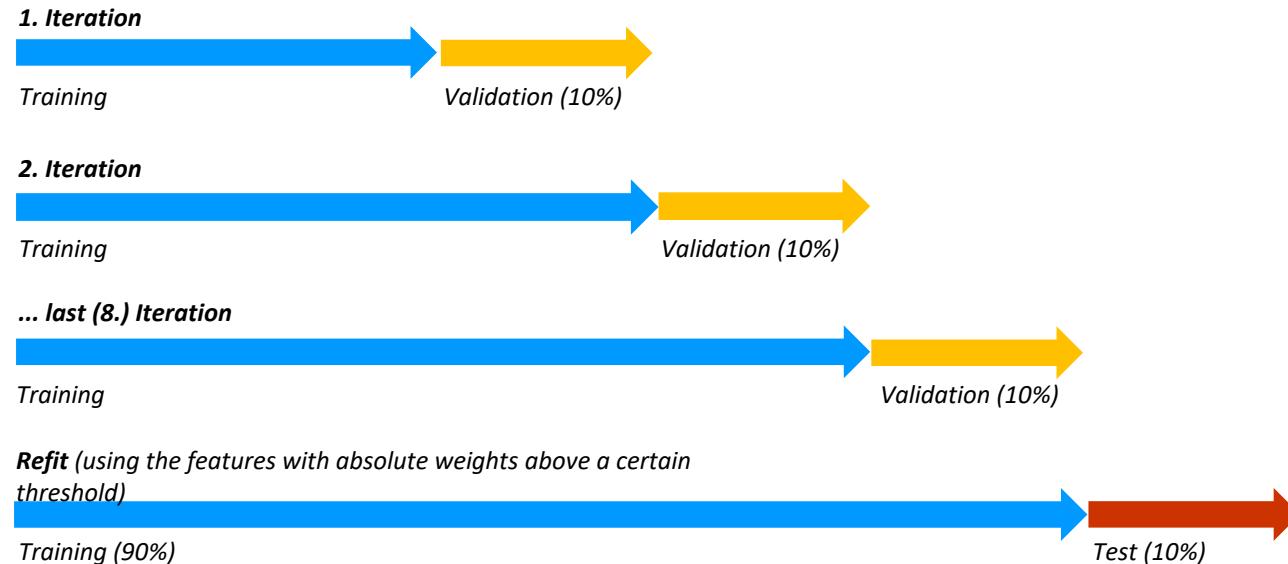
TP=Of all the observations that went up, how many did we predict to go up? ➡ We want this number to be high

FP=Of all the observations that went down, how many did we predict to go up? ➡ We want this number to be low

Indicator Selection

Feature Importance

- >Selects indicators with weights above a certain threshold



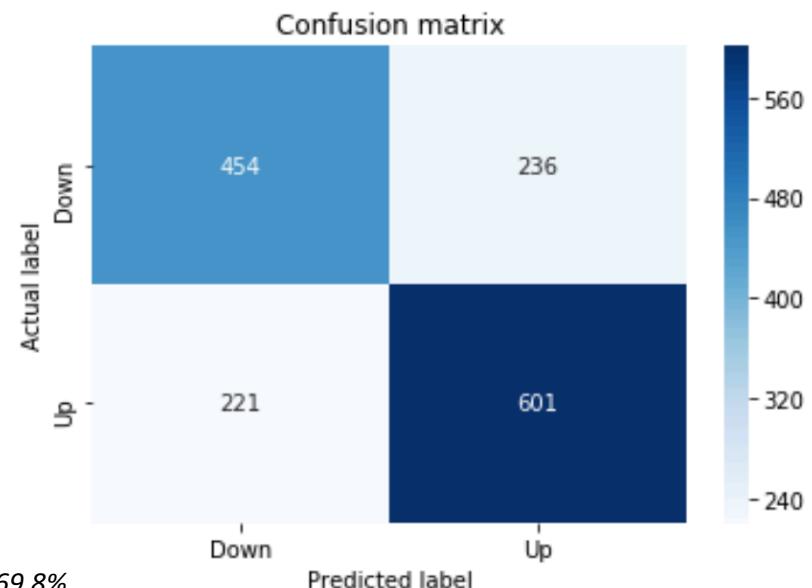
- In our model, **indicator 11** (and to a lesser extent **indicator 22**) are the most important

Model Assessment



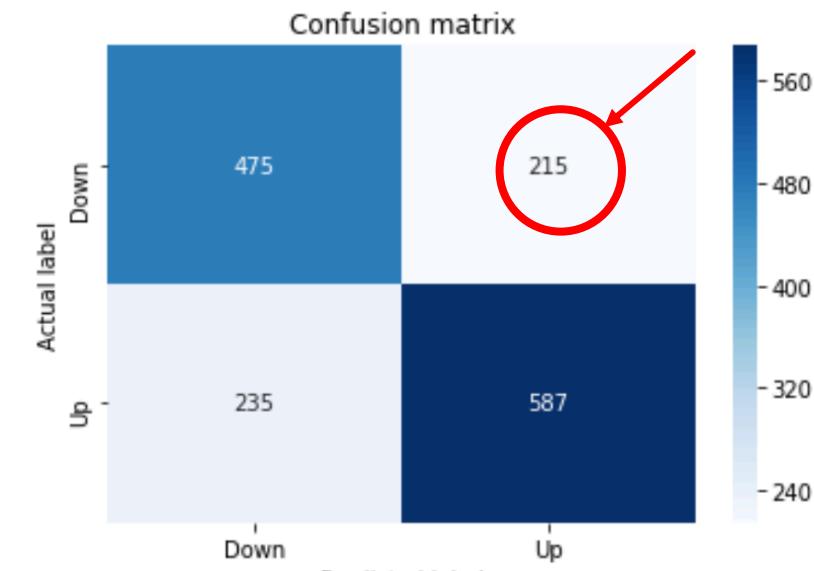
Moderate strategy

Unbalanced data



Conservative strategy

Balanced data

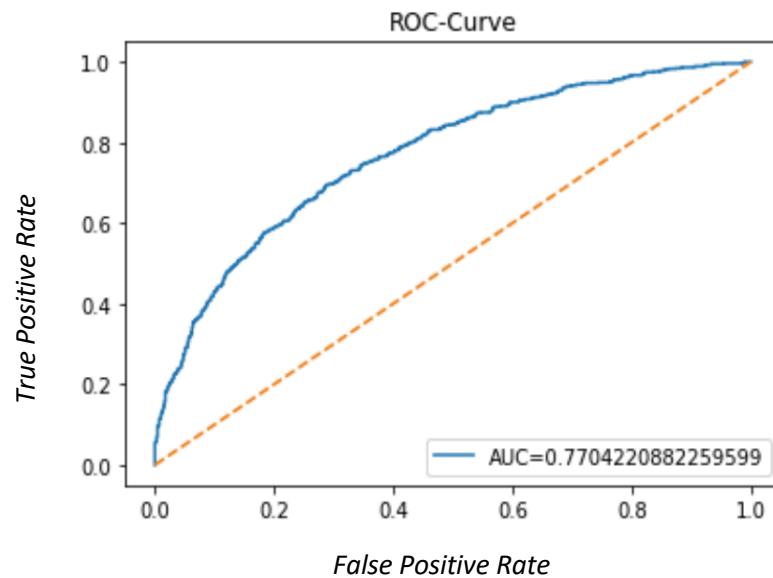


Model Assessment



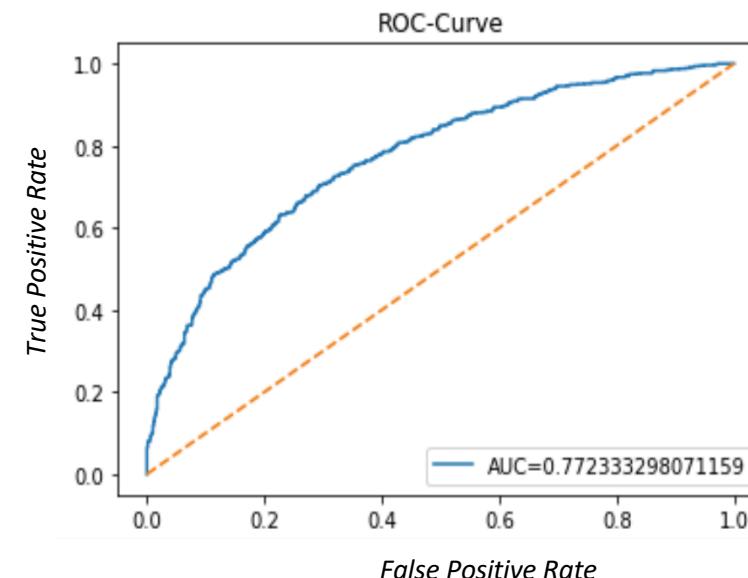
Moderate strategy

Unbalanced data



Conservative strategy

Balanced data



TP=Of all the observations that went up, how many did we predict to go up? ➡ We want this number to be high

FP=Of all the observations that went down, how many did we predict to go up? ➡ We want this number to be low

Advantages and Disadvantages

+ *Advantages*

- › Ability to reveal complex nonlinear patterns and complex non-linear interactions
- › Robust against:
 - outliers
 - non-scaled data
 - multi-collinearity
- › Simplicity
- › Interpretability
- › Feature importance

- *Disadvantages*

- › More complex optimization compared to linear setting
- › Potential to overfit



Thank you