Using news analytics to predict stock market movements

Ivo Tadic

Hood College / CS-522 Data Mining

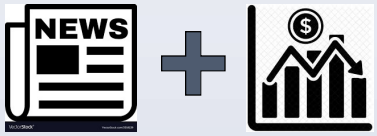
Introduction

Do News have an effect on Stock Market movement?

Two Sigma has released a Kaggle competition for Data Scientist to try to find the answer to that question. Based on Datasets from Thomson Reuters (News Dataset) and Intrinio (Market Dataset). [1]

Objective

Predict Market movement based on news and market datasets.



News Data [3]

Market Data [2]

PREDICT

Market Movement

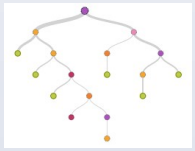
Model

The solution has some processing restrictions (CPU) and some runtime memory restrictions (RAM). The Algorithm selected should be able to process the data within the required boundaries.

6 hours of runtime
2 CPU cores

16 GB RAM Limit

Amongst tree classifiers the LGBM Classifier provides the speed and memory usage to allow us to process the data within the boundaries.



LGBM Classifier

Decision Tree Boosting Algorithm
Fast and low memory usage!

RESULTS

Our results show that news data provide value to the accuracy but the market data itself accounts for the most important features in the decision tree generated with the train data.

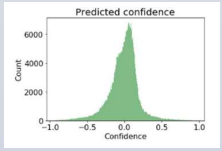
The confusion matrix shows that there is a big number of true negatives and false positives.

Accuracy / AUC Scores

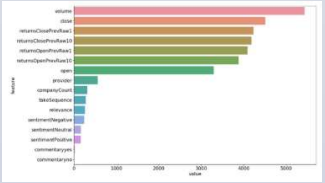
Accuracy score	0.654023
AUC score	0.726820

Confusion Matrix

Prediction Confidence



Feature importance



Conclusions

The LGBM Classifier provides a good starting point to create a prediction that includes both News and Market data and the results show better accuracy compared to other models that only use Market data.

Future Work

Now that we have a solid baseline the next step would be to try a neural network with limited amount of data.

Understanding why the true negatives and false positives is that big would be something to review further and to understand the data behind that specific use case.

REFERENCES

[1] Kaggle (2018), Two Sigma: Using News to Predict Stock Movements, https://www.kaggle.com/c/two-sigma-financial-news.

[2] Intrino (2018), Intrino's Main Site, https://intrinio.com/.

[3] Thomson Reuters (2018), Main Site , https://www.thomsonreuters.com/en/products-services/financial.html/.

Contact Information

Ivo Tadic
Hood College Graduate School
it3@hood.edu