STOCK PREDICTION

FNCE 2431

ANTHONY WANG
DYLAN OTTAWAY
SCOTT BEVAN

AGENDA

- Problem Domain
- Dataset Description
- Pipeline Summary
- Approach
  - Initial Results Summary
  - Hyperparameter Tuning
- Results (PnL)
- Risks / Recommendations
- References

PROBLEM

How to achieve consistent excess returns on trading public equities based on a machine learning model or model variants that can predict 1 day ahead with acceptable accuracy that a stock will go up or down.

Key questions:

- Best or performant model?
  - Appropriate feature selection?
    - Tuning analysis and results evaluation?

DATASET DESCRIPTION

Source: import yfinance as yf

Stock Data:

- Top 50 US Technology (by MKTCAP)

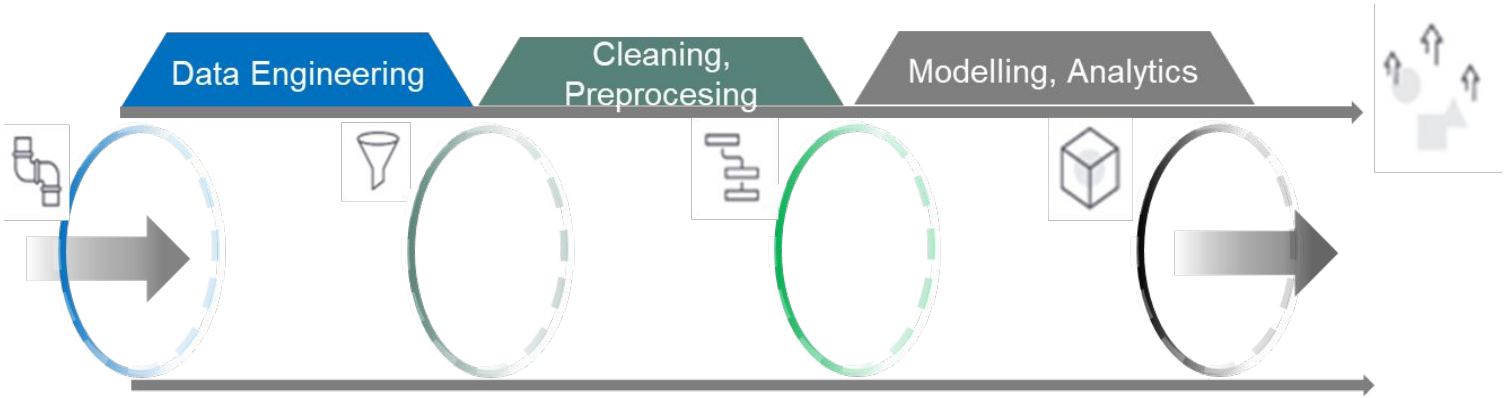
Type: Time Series Price and Volume data
Period: 2016-01-01 until 2022-01-01
Total rows 75,550

Feature Selection Augmentation

- **19 self selected features,** variations of RSI, SMA, MFI, EMA, and Lag periods
- **91 TA package feature** augmentation

Total Feature number: 110

PIPELINE SUMMARY AND APPROACH



- Data ingest
- Data grouping

- Feature selection
- Normalization
- Refinement

- Classifier selection (initial)
- Target: % change returns 1 day ahead
- Target Class (Prediction)

2 Class Model

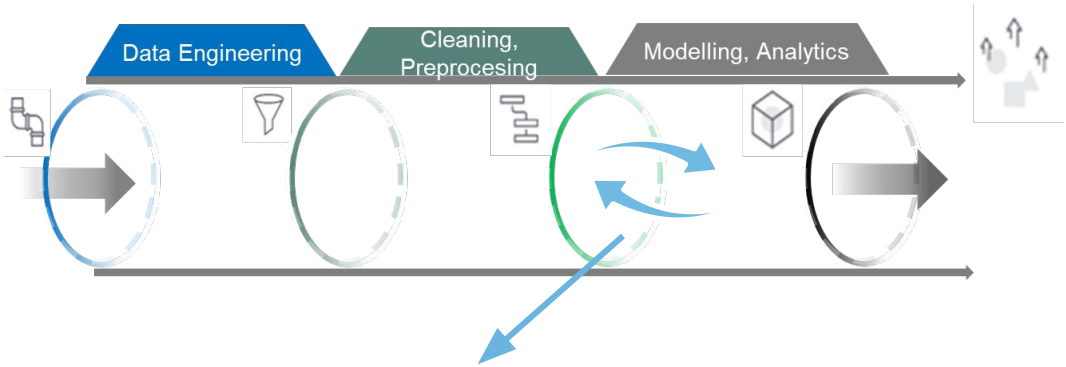
Up {1}, Down {-1}

3 Class Model

Up {1}, Flat {0}, Down {-1}

- Fit and score
- Visualize and Evaluate
- Parameter Tuning
- Business Representation (P&L)

APPROACH



Iterative Model, Evaluate, Tune

- **Gradient Boosting Classifier**
- **Decision Tree Classifier**
- Logistic Regression
- Gaussian NB
- Ensemble Method (Gaussian NB, Logistic Regression, Bernoulli NB, SGDC)

Iterative Selection and Tuning

Model Selection

- Best initial model selection
- Check for over/underfitting, accuracy
- Hyperparameter tuning

Tuning

- Manual parameter tuning
  - Range of learning rates
  - Estimators (# of sequential trees)
  - Random subsampling
- Hyperparameter tuning using RandomizedSearchCV

SCORING RESULTS

[0.75153323 0.76440687 0.78584452 0.78341345]
Avg cross-validation: 0.771 (K-Fold, 4 folds)

In Sample

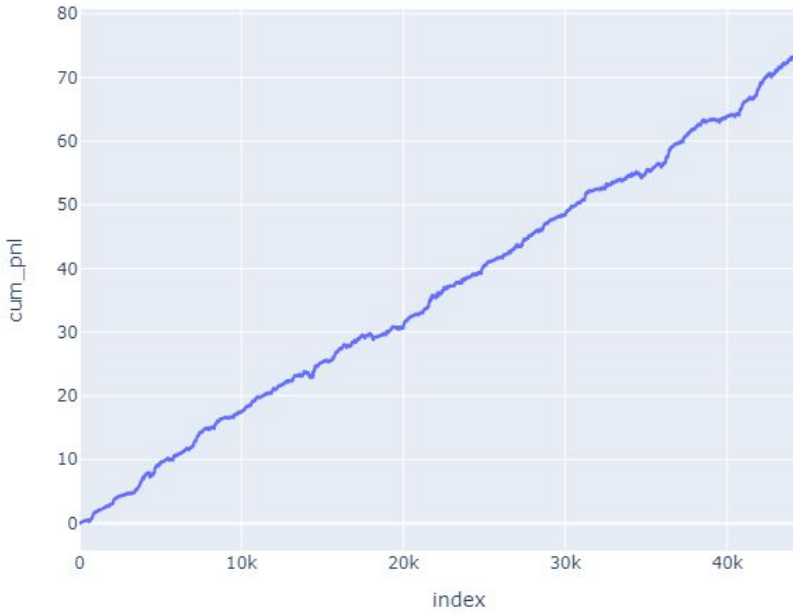
Classifier	Accuracy	Target Class	Precision	Recall
Decision Tree Classifier	(training):1 (validation): 0.518	Down {-1}	0.47	0.48
		Up (1}	0.56	0.55
GradientBoosting	(training): 0.66 (validation): 0.54	Down {-1}	0.49	0.51
		Up (1}	0.58	0.56

Out of Sample (on unseen data)

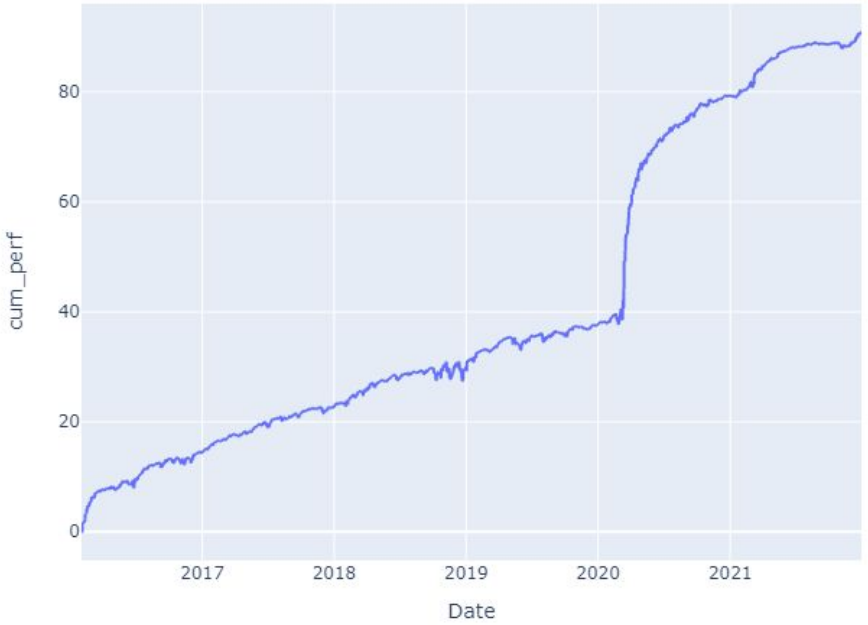
Classifier	Accuracy	Target Class	Precision	Recall
Decision Tree Classifier	(validation): 0.536	Down {-1}	0.55	0.50
		Up (1}	0.52	0.58
GradientBoosting	(validation): 0.555	Down {-1}	0.56	0.64
		Up (1}	0.55	0.46

P&L Analysis

Cumulative P&L Curve (pct. gains)



Cumulative Performance



Risks

-Overfitted

-Poor accuracy on the down

-Bias for our market selection, and regime

-Best models going forward is Gradient Boost

-Additional analysis needed

-Scale tickers and features

-candlesticks, crisis / historical events

- Different data sources

Recommendations

REFERENCES

-https://technical-analysis-library-in-python.readthedocs.io/en/latest/ta.html

-https://github.com/twopirllc/pandas-ta

-https://finance.yahoo.com/

-https://urldefense.com/v3/__https://arxiv.org/pdf/2107.13148.pdf__;!!MLMg-p0Z!EPdjmtu5ROx6-9sHz_oyjhGpvNPDIDWRQd8AyXkF9X9-7pUgZN1QGG-arqc5rEnhzMXcDvX2M0rWrCMhJ157CA$

-https://arxiv.org/pdf/1402.7351.pdf

-http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1049.2078&rep=rep1&type=pdf

-https://www.investopedia.com/terms/r/rsi.asp

-https://www.investopedia.com/terms/s/sma.asp

-https://www.investopedia.com/terms/e/ema.asp

-https://companiesmarketcap.com/tech/largest-tech-companies-by-market-cap/

-https://www.analyticsvidhya.com/blog/2016/02/complete-guide-parameter-tuning-gradient-boosting-gbm-python/

-https://machinelearningmastery.com/nested-cross-validation-for-machine-learning-with-python/

-https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html
-https://farland.faculty.arizona.edu/sites/farland.faculty.arizona.edu/files/2018-11/jama_Norton_2018_gm_OddsRatios.pdf

Contact Information

Anthony Wang - awang12@scu.edu

Dylan Ottaway - dottaway@scu.edu

Scott Bevan - sbevan@scu.edu