

Machine Learning: Python-based model of stock convictions to expected returns within given investment horizons

Student Group Profile: BSc/MSc/MEng level with background in Financial Mathematics, Statistics, Machine Learning, Computer Science, Software Engineering.

Background: As part of its systematic fund, BankingScience Ltd (BSL) has developed a long-only, smart-beta algorithm and is currently integrating it with a Black-Litterman allocation framework.

To enable such integration, BSL needs to develop a model capable of predicting expected returns on the basis of a ranking of scores assigned to all the evaluated stocks on a weekly basis, with investment horizons of: a month, quarter semi-annual and annual.

Key Elements of the Project:

- Functional and non-functional requirements are to be gathered and summarised in a separate document;
- **Initial modelling**
 - pre-process example model logs and extract stock convictions
 - implement a simple regression model to be used as a baseline for further analysis
 - implement a time-series analysis harness that will be used to train, test, cross-validate and evaluate different version of models
 - implement suitable metrics to evaluate predictive power of the models
 - implement a simple, repetitive reporting facility to output results
 - train, test and cross-validate the regression model and show that the model generalizes on the out-of-sample data
 - prepare a presentation of methodology, architecture and results, present to BSL
- **Advanced modelling**
 - develop, train, test and cross-validate three alternative models
 - compare performance of the models against a baseline model and against each other for the various investment horizons
 - perform stability and sensitivity analysis w.r.t baseline model and investment horizons
- Unit/integration/system-tests and documentation is to be provided on GitHub/Confluence
- Final presentation of results (including key findings, advantages, disadvantages and further analysis suggestions)

Key Assumptions:

- All elements of the code will be organized on GitHub and needs to be frequently backed up
- Modelling will be performed in Python, final product needs to work as a standalone script
- List of stocks for analysis and Indices for performance comparison will be provided