

# Report of Stock Movement Prediction and Portfolio Optimization

By Pong Kaming, Lee Adwin Richie,  
Mamik Prabhsharan, Tsang Dylan

## Introduction

The stock market has long been an area where people attempted to implement machine learning to forecast the return of stocks. This project explores the possibility of combining stock movement prediction by machine learning and portfolio optimization with prediction results through the following process. Firstly, a machine learning model will be trained by historic data to predict the return of each stock inside a large portfolio, like S&P 500. Then, an optimization model will take these predictions as inputs and generate a portfolio. To assess the performance of the model, the return of the generated portfolio will be compared against the return of the original large portfolio with the metric Information Ratio(IR). In this report, we will discuss the methods used to construct the model, as well as the results generated by the model.

## Methods

### Data Retrieval and Preprocessing

Due to time constraints, this project only explores the implementation of stocks in the S&P 500 index. Historical daily close prices of these stocks, from 10/11/2016 to 29/4/2022, were retrieved from Yahoo Finance through the python package “yfinance”.

Daily close prices of these stocks are then used to calculate features like 30-day volatility of the stock return, 30-day beta, which describes the expected movement of a stock relative to the market, and return of the stock in previous days. Stock code is also kept as a categorical feature for the machine learning model.

### Machine Learning

CatboostRegressor, which is a variation of gradient boost, was used for the project due to its convenience in handling categorical data and ability to utilise GPU for faster training. For each trading(test) day, the past 100 days of historic data was used to train the model, where cross validation on the number of iterations was performed to prevent overfitting. The model is then used to predict the return of all stocks on that trading(test) day; and the predicted returns are used for optimising the portfolio.

### Optimization

#### Black-Litterman Allocation

The Black-Litterman (BL) model takes a Bayesian approach to asset allocation. Specifically, it combines a prior estimate of returns (for example, the market-implied returns or market capitalization) with views on certain assets, to produce a posterior estimate of expected returns. The advantages of this are:

- You can provide views or predictions on only a subset of assets and BL will meaningfully propagate it, taking into account the covariance with other assets.
- You can provide confidence/accuracy of your views.
- Using Black-Litterman posterior returns results in much more stable portfolios than using mean-historical returns.

The BL model outputs posterior estimates of the returns and covariance matrix. These returns can then be used to compute the implied weights and assets can be calculated accordingly.

## Results

The simulation of trading based on the above model last year yielded a 30.9% return, which is significantly higher than the S&P return of -2.5% in the same period. The annualised IR is also at a satisfactory level of 96%. However, in terms of IR, the strategy only yielded a score of 6.5%. This means that our strategy performs better than S&P, but not very consistently.

We believe the strategy of combining machine learning and portfolio optimization has its potential, as demonstrated by the impressive 30.9% annual return. However, modification needs to be made such that it can outperform the market on a more consistent basis. Some possible methods could be incorporating more sophisticated machine learning algorithms like neural networks, or including derivatives like options as part of the portfolio optimization.