M6 Competition Project

M6 Competition

The M6 Financial Forecasting Competition is held to gather empirical evidence that sheds light on the causes and implications of the Efficient Market Hypothesis (EMH) paradox. According to this hypothesis, financial share prices reflect all available information, making investment outperformance relative to the market impossible.

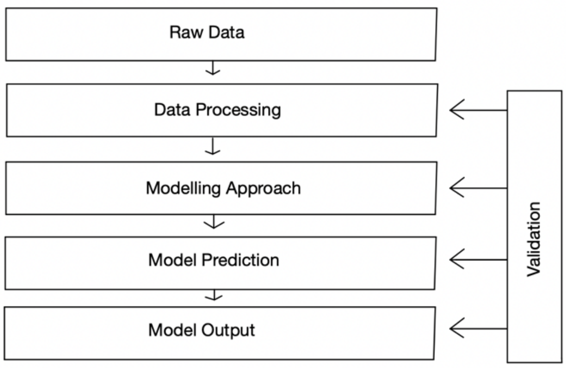
However, there have been instances where the explanation of EMH appears paradoxical, such as when the market observes both the poor performance of active funds and the exceptional performance of individuals such as Warren Buffet.

There are a total of 100 equities provided in this competition. Teams are expected to predict the bucket of rankings where rank 1 is the lowest forecasted percentage return and rank 5 is the highest forecasted percentage return for the next 9 weeks. Teams are also expected to provide investment decisions where they decide the amount and whether to long or short each of the 100 equities.

Workflow

Please refer to "SampleData.R" where the script provides a sample workflow covering the whole workflow. Do note that you are not required to follow the exact workflow provided.

A workflow is crucial in transforming the raw data through a series of steps to valuable data or insights. A workflow also ensures that the detailed steps are consistent and repeatable, where other users can recreate the workflow and arrive at a similar result. Implementing a workflow is also beneficial for the team to incorporate improvements or make changes without disrupting the code downstream.



Data Source

There are four data sources provided in this project, namely the equity details, ETF prices, stock prices, and sentiment on the sector level from 2019 onwards. The team has the freedom to utilize any of the data sources provided or scout for any alternative data source available. Scouting for additional features in alternative data sources may provide some insights for the model during the prediction stage.

Data Processing

In the real-world environment, data rarely come processed and ready to be churned into predictive models. The primary step of completing the challenge is through the acquisition of valuable data through the mountain of data available.

Even though the data provided in the M6 competition went through a series of processing, further improvements can be made to the raw data before we train our predictive models with it. Some examples of other improvements include, but are not limited to, addressing the unequal periods between each time-series data and accounting for special holidays and seasons.

Feature Engineering

Frequently, feature engineering provides the most improvements to the predictive performance of our models. Using the data sources available and the teams' domain knowledge, teams can create relevant metadata that is fed into the predictive models. The decision to include a feature is dependent on the type of application. Having too many unrelated features may deteriorate the models' performance, referred to as 'noise.' Hence, the teams should always verify if the set of features created is relevant to the cases.

For example, there is some sentiment data provided in sector-level detail by analyzing the news published using NLP. The team may decide to utilize the sentiment data as part of the feature for prediction, or data engineer it for more valuable insights using domain knowledge.

Modeling Approach

With the data ready to be predicted, the teams can plan how to carry out the prediction. An example approach is first to perform the train-test split, then train a model on all the available training data. Next, the teams would perform hyperparameter tuning on the model to improve the prediction capabilities. Lastly, the teams would assess the performance of the model based on the test data. The teams are free to use any approaches which they deem fit. The guidelines given below are just as the phrase suggested, as guidelines. Teams are not expected to strictly follow the steps below.

Model Selection

During this stage, the teams select the model(s) that execute the teams' modeling approach. To compare the performance of each model, the teams can use log loss (If performing categorical prediction) as the comparison metrics. Some teams may also choose to conduct model calibration to assess the model performance.

Hyperparameters

After selecting the best model to carry out the prediction task, teams can optimize the model by going through hyperparameter tuning. Sometimes, hyperparameters may provide better predictive performance. However, teams should be aware of the risk of overfitting while employing hyperparameter tuning. In the space of financial forecasting, it may be wise to perform boot-strapping rather than hyperparameter tuning to prevent overfitting.

Model Performance on Test Dataset

Teams are advised to compare the performance of their models on the test dataset, which is the dataset that the model has not encountered. If the model's performance deteriorates significantly, then it suggests that the model is overfitted on the training dataset.

Model Prediction

With the predictive model that the teams have prepared, teams are expected to predict the bucket of rankings where rank 1 is the lowest forecasted percentage return and rank 5 is the highest forecasted percentage return for the next 9 weeks. [Eg: On Sept 16th, 2022, your model should predict the ranking buckets for Nov 18th,2022]

Teams are also expected to provide investment decisions where they decide the amount and whether to long or short each of the 100 equities for the next 9 weeks.

Please refer to "sample_submission.csv" for an example of the prediction result.

Validation

Every step in the workflow detailed above should be subjected to validation. In forecasting, a common problem is data leak, where the model has access to features that are not made available at the time of prediction. Data leaks are common if teams are doing feature engineering without an adequate understanding of the created features. Another common issue is infeasible predictions. Hence, teams should perform validation and sanity checks at every step of the workflow.

Evaluation

Teams are assessed based on a mixture of interpretability, accuracy, and creativity. **Please submit the predicted ranking buckets for Nov 18th.**

On interpretability, I will need to be able to understand your code and be able to run it on my local machine. Do specify any non-native packages used so I can install the dependencies on my local machine.

On accuracy, I will be using the actual stock returns to evaluate the accuracy of the prediction output as well as the investment decision output.

On creativity, teams are encouraged to venture with materials outside of the curriculum.

There are two types of files to be submitted:

1. Your code that includes the workflow (Your code will need to be able to produce the model(s) and the output CSV)
2. Output CSV (refer to the sample output provided) Before submission, compress all the required files into a single zip file and label the file as M6Forecasting_TEAMNAME.

FAQs

1. Must we have exactly 20 equities in a single rank bucket?
    a. No, some equities are subjected to high volatility. Hence, it may be prudent to have more/less than 20 equities in a single rank bucket to capture the performance. Note that the results are evaluated in terms of probabilities in each ranking bucket.
2. Can we choose not to invest in certain equities?
    a. Yes, you are not required to invest in every single equity. However, you do need to make sure that the sum of the absolute value for the investment decision should not be greater than 1.