Case Study: Equity Selection Model

The goal of the Equity Selection Team at Ultramarin is to develop a holistic equity selection model that can extract relevant signals from heterogeneous data powering our investment strategies.

Now imagine you are part of the squad and face the following task: Our data engineers have provided you with a panel data set (see attached data.csv) with [date, identifier, market-cap, sector and index-membership] as meta-information. The dataset contains a target variable ("target") that indicates the total return of the stocks over the next 28 calendar days. In addition, a selection of 10 firm-specific features derived from price and fundamental data is available to help predict the return of individual stocks.

Please build a predictive model to forecast the target variable. It is up to you what model you choose and how you structure the study. There is no right or wrong, we are primarily interested in how you would organise your work.

We'd like you to walk us through your results (including programming code in Python) as part of a technical presentation that should address the following questions:

- Why did you choose this particular model and what are its strengths and weaknesses?
- Which features are important for the chosen model?
- How did you evaluate the performance of your model?
- How do you make sure that the model generalises well?

The presentation should last approximately 30 minutes and is directed to a hypothetical investment committee composed of experienced practitioners from the investment industry.

Please send us your code/solution within 14 days as a .zip or .tar.gz archive. Include comments and explanations and make sure that your code is accompanied by a README with installation and execution instructions. This will ensure that we can review your work in a timely manner.

Note:

- The target variable ("target" in the data set) is already forward looking, i.e. a sample's date label is the beginning of the investment period over which the return will be realised 28 days later. There is no need to "shift" the date label of samples.
- The boolean "index membership" indicates whether a particular stock is a member of a particular investment universe at a given date.

If you have any questions regarding the task, please do not hesitate to reach out to David (duemig@ultramarin.ai) or Benjamin (btschroer@ultramarin.ai).