

Introduction

Thank you for taking this data science project. I look forward to reading your results. This project is used for the RA/PhD/intern hiring to evaluate your technical skills and capabilities under the context of quantitative trading.

Evaluation

You will be evaluated based on the following criteria:

30% - **Data analysis skills**

30% - **Code quality**

30% - **Presentation of results and explanations.** We should be able to understand what you did without reading the code.

10% - **Creativity**

Data Description

The training dataset **train.parquet** containing all historical market data along with the corresponding labels. **You are expected to build models to accurately predict the label.**

- timestamp: The timestamp index representing the minute associated with each row.
- bid_qty: The total quantity buyers are willing to purchase at the best (highest) bid price at the given timestamp.
- ask_qty: The total quantity sellers are offering to sell at the best (lowest) ask price at the given timestamp.
- buy_qty: The total trading quantity executed at the best ask price during the given minute.
- sell_qty: The total trading quantity executed at the best bid price during the given minute.
- volume: The total traded volume during the minute.
- $X_{\{1,\dots,890\}}$: A set of anonymized market features
- label: The target variable representing the anonymized market price movement to be predicted.