



Machine Learning Researcher | Interview Project

Interview Project Format:

- At home, conduct a real data set evaluation with data science suite of choice (Ipython notebook, R studio, ...). We expect you to spend ~3hours on the problem, excluding models that you may run. The presentation should be a maximum of 5 pages.
- Send final workflow in the form of a notebook / script to recruiting@optiver.us 48 hours prior to final round interview. We will have it available to you during the virtual on-site interview
- During the virtual on-site interview, you will have the opportunity to present results and explain the design choices made at each step (Presentation duration: 60 minutes)

Interview Project Objectives:

- Demonstrate data science skills on real life problem:
 - Gain understanding of dataset
 - Prepare dataset for model training
 - Based on understanding of dataset choose appropriate model
 - Setup workflow for model assessment and selection
 - Train model
 - Visualize results and create report
- Communicate findings in a clear and concise way with visualizations:
 - To a non data scientist at a conceptual level
 - With fellow data scientists stating important considerations and assumptions made that led to the results

Data Set Characteristics and Goal:

- Samples are derived from trading data (so inherently noisy)
- Samples are from a time series with non-uniform time steps (event based)
- Around 100k samples
- Different features
 - Some contain historical information (e.g. a moving average)
 - Some are current snapshot (e.g. current price)
 - Some aren't time series themselves (e.g. size for you to trade against)
- Target variable is expected profit of individual trades where expected profit is some time in the future minus now
- **Goal: Maximize the sum of the target variable of the test set based on a model that is derived from the train set**

Preparation:

- Use dataset to demonstrate data science skills (see above for break down)
- Dataset is partially cleaned but be skeptical about it and present flaws in final report
- In each step explain why you made said choice (this will help during interview)
- **Please send final Ipython workbook / R script to us so it will be available to you during the interview.**

The content of this interview project is intended for the original recipient only. Please do not share any part of this assessment (instructions or dataset) online or to any other party.