Regression Case Study

The following exercise is a simplified version of one of the tasks that we face in connection with one of our real projects on construction material properties. It consists of developing a predictive model for material strength based on easily accessible chemical and mechanical properties in order to avoid waiting for a time-consuming and costly lab measurement.

Instructions

1. Develop a prediction model for the material strength (marked as column 'Y' in the dataset).
2. Try to estimate the inherent laboratory measurement error of the material strength and make a statement about the best possible model performance.

Notes

1. All the data needed for the exercise are stored in data.parquet.
2. The metric that will be used to evaluate the model's performance on unseen data is mean absolute error (MAE).
3. The actual target Y is a mean of six individual lab measurements (Y1, ..., Y6).

Reproducibility

Please provide the code in a form that allows us to rerun / reproduce your results for a fair assessment. You're free to organize the code in your preferred way. However, all stages of your analysis (preprocessing, visualization, feature selection, modelling, etc.) should preferably be executable from a commented Jupyter notebook.

General

Please note that this is not meant to be a test with only one correct answer. We will rather judge your solution by the conceptual rigor of your approach, in particular with regard to the modelling and software development best practices, quality of documentation and the way how you present your solution.

Nice visualizations are welcome, but the content of the visualization is more important than its visual aspects.

The result of the exercise can be made available as a private repository or be sent as an archive to our email.

During the interview, you will be asked to present your results in a form of your choice (timeframe ~20 minutes).

In case you struggle with any part of the process, don't hesitate to contact us, so that we can assist you.

Good luck!