


Predict Hazard Ratings for a Maintenance Project



All the money in the world can not make up for a loss of life. Better safe than sorry is the motto of all engineering and construction practices which routinely deal with dangerous situations. However many times , extent of caution is just not enough and leads to very hard learnt lessons in employee safety .

Instead of relying on something going wrong and then taking preventive measure ; organisations are coming up with ways of assessing the extent of danger to life for a new project before even the first hammer strikes .

Given to you is a masked dataset which contains various properties of tasks taken up in heavy equipments maintenance by a manual crew. Each of these tasks have been given a hazard score which is eventually used in deciding levels of safety checks and caution while planning for the maintenance process.

Data Files

Train Dataset = Hazard_train.csv

Test Dataset = Hazard_test_share.csv

Formal Problem Statement

Variable names are masked and there is no formal data dictionary provided by the client .

Your task here is to build a predictive model for predicting hazard score given other information related to maintenance tasks. You need to build your model on the train dataset. **Test dataset does not have a response column; you need to predict those values and submit it in a csv format.**

target column = Hazard

Evaluation Criterion Part 1:

You will first attempt Part 1 of this project which is a quiz. You can access it through LMS. This quiz needs to be answered based on exploration of the dataset given and some generic questions about algorithms discussed in the course. Consider only the training dataset for data cleaning and exploration to answer the quiz questions. There will be 10 questions of which you need to get at least 7 correct in order to pass the project.

Part 2:

Here you work on creating the machine learning models and choosing the one which gives the best performance. **You can refer to the Project Process Guides provided in LMS to understand how to approach and work on a project.**

For this project, score will be calculated as:

$$\text{Score} = 1 - (\text{MAE} / 5.4)$$

Where MAE is mean absolute error on test file. **You need to score more than 0.5 in order to pass the project submission for this particular project.** Don't read too much into score formulation, it is just to scale MAE. You need to focus on minimising MAE.

Submission:

Please give appropriate names to submission file. Preferable containing your name and attempt number [this is for you to keep track of your different experiment performance] . Submission needs to be a csv file . Any other format like excel , pdf etc will not be graded.

Number of rows in the submission csv should be exactly the same as test data. If this is not taken care of, your submission will not be graded.

You can make as many submissions you want . [We might ask you to submit the script which was used to generate the submission at any time].

In order to clear this project, you are required to clear both, Part 1 as well as Part 2 of this assignment. Wish you all the best!