

Mini-Project : Individual Contribution

NAME : K.S. RANASINGHE

INDEX NO. : 210518H

Task Overview

We were provided with an extensive dataset for predicting the credit risk of an individual. The tables given in the dataset were of a varying nature with some tables having a depth quantity as well. We had to handle the individual tables separately when working with them. Add to that the fact that some tables had multiple files due to the sheer size of those tables. First, we distributed the tables among the group members equally for analysis as to how to approach those tables. At the end of the individual analysis, we came up with a workflow as to how to pre-process this extensive dataset. For that we developed all the required functions in the pre-processing pipeline in a generalized manner so that we can apply them to all the tables in the dataset. The pre-processing pipeline included functions all the way from loading the csv files to merging the processed dataframes into the final main dataframe. Afterwards we individually tried different approaches for model training and came up with the best performing model. This was all that was required for the competition submission. Then we had to write the paper for which we distributed the different topics among the members and completed with thorough validation from all the members.

Personal Contribution

- Individual table analysis – debitcard_1, other_1, credit_bureau_a_2
- Contributed to build workflow.
- Built the cross-validation function.
- Contributed to building the ensemble model.
- Evaluated the models with variety of metrics.
- Comprehensive tuning of model by making multiple submissions to competition.
- Contributed to the methodology and discussion sections of the paper.