# BRIEF INFO ABOUT THE PROJECT

The adult data income prediction project is developed on Python by using PyCharm IDE. The project has 6 files:

1. **Main:** The main script of the project. It loads data from CSV files, and parses into a usable format for machine learning. Categorical fields in the given data is encoded to numerical format. Finally scaling and feature selection on data are realized in the pre-processing step.

As described also in the script, 7 different supervised classification methods are compared by using 5-fold cross validation technique and default hyper-parameters of these methods. Although fine-tuning via grid-search on the parameters is planned and implemented in the file, it is not used due to limited time for this effort. If the tester of the file want to check the result of the grid-search, he/she can just remove comment-out signs on the code block.

1. **PredictorUtils:** It is a utility class including data visualisation, classification reporting and data loading functions.
2. **EncodeCategorical:** The class selects the categorical features and encode them in order to use with the rest of numerical features in the dataset while modelling the predictor.

All the project is uploaded to Github account:

<https://github.com/ridvansalihkuzu/ctr-prediction>