AIL 7310: Machine Learning for Economics
AY 2025-26 Semester I, Assignment 2

I. [20 points] Use the dataset 'credit_risk.csv'. Answer the following questions

1. Create the summary statistics of the variables in the dataset.
2. Build a decision tree tuning the features of maximum depth and minimum samples in each leaf.
3. Build a random forest model to predict credit card defaults for individuals tuning the features of number of features and max estimators.
4. Build a boosted decision tree (XGBoost or Boosted Gradient Trees) to predict credit card defaults.
5. For each model, report accuracy, precision, recall, F1-score, and ROC AUC on a held-out test set.
6. Use 5-fold cross validation for hyperparameter tuning
7. Run conditional inference trees and forests. [For R users, you can use ctree from partykit or party package. For Python users, you can call R within Python using rpy2]
8. Compare the results of Random Forest and Boosted Decision Trees with Conditional Inference Forest. Are the same features important? If there are any difference, why might they be?