Problem Statement: These data consists of 30,000 bank records for credit card customers. The target attributes is a binary outcome - 'default', indicating whether the customer defaulted on credit card balance (default=1) or not (default=0).

Your problem is to build and validate the best model for predicting the probability of a customer default based upon their payment record for the past 6 months and their demographic and other bank information.

Use the techniques we have discussed this semester. Preprocess the data, then use hyperparameter optimization to configure the best solutions from:

1. logistic regression,
2. decision tree,
3. neural network , and
4. random forest.

Where appropriate, hyperparameter optimization should be used to select the best model configuration from a 10-fold cross-validation.

Once you select the best configuration for each of these models, select the best model among the 4 using a simple 70/30, train/validate, model comparison.

Report all metrics we discussed this semester, not just MISC (the misclassification error rate reported in SAS EM).

Solve this problem using both SAS EM and Python. In your report, describe your approach and solutions separately, but combine the descriptions into a single report. At the end, compare the two solutions, discuss their differences and select what you think is the best solution.

Discuss the rationale for your choice. Your report will be graded for it's writing as well as your technical skill in solving the problem. Poor writing and incomplete descriptions of what you did and why, will get a lower grade.