Loan Default Project Report

Abstract:
What is the best model for predicting whether or not someone will default on their loan? In this project we explore the prediction of loan default using various techniques, including logistic regression, LASSO, neural networks, and Random Forest applied to a balanced dataset. The models were evaluated using confusion matrices, accuracy, ROC curves, and the Area Under the Curve metric. Overlapping ROC curves facilitated a direct comparison of the models, demonstrating the minimal difference in accuracy from one model to the next, allowing us to choose either the neural network model due to parsimony, with an accuracy of about 0.68 and AUC of 0.74. Higher interest rate promotes defaulting, while income, age, loan amount, and months employed makes it less likely to default. While some factors that we assumed would have a larger impact, such as: credit score, having dependents, and level of education, were not found to be important predictors of loan default.

Research Questions:
1. What is the best model for predicting loan defaulting? How is the accuracy and recall rate?
2. What are the most important predictors for loan defaulting?
3. How important is income, credit score, and dependants?
4. Does education have an effect on defaulting?

Does this align with our **assumptions** before even starting the analysis?