Final Project – Identifying bad credit risk loans

- Due Date: 12/13/2021 23:59 pm
- Goal is to apply as much as what you had learned to solve a classification problem
- Dataset is based on a dataset of that classifies people as good or bad credit risks, https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)
- Use the modified version, german_credit_modified.csv from Blackboard.
- Use the Final Project notebook as your starting point.
- Try at least two different models that we have covered in class such as Logistic, SVM, Decision Tree, Random Forecast, and Naïve Bayes models. Do not use any models that we have not covered in class.
- The target variable is the "Risk" column
- You do NOT need to use ALL the variables in the dataset, but you should pick or have tried or investigated a good number of the variables.
- Grading will be based on whether you have done it "right", not the actual performance of your final model

Final Project Grading Criteria

- Comment on your model results and explain whether "Credit History" is important in classifying if the loan is bad which would require to compare two models, one with and the other without that variable.

- Follow the standard way of building models that include
  - checked missing values, removed outliers
  - performed basic exploration of relationship with plots and graphs
  - separated data set into training and testing
  - Setup dummy variables for categorical variables,
  - normalize numerical features if needed
  - tried at least two models and checked their model performance
  - performed cross-validations

- Identify your best model that does not include the "Credit History" variable. The student who has the best model will have an extra 10 points (out of 100), but will still be max out at 100.

Final Project

OK to discuss among yourselves, but do not share code

Good Luck and Have Fun!!!