

"""

Maria Teresa Luna Plascencia

Class: CS 677

Date: 04/23/2024


Final Project - Instructions


"""


## Project: Analyzing various ML models and strategies to treat unbalanced datasets for loan approval predictor


### Instructions:


1. Unzip the marluna\_final\_project.zip file to a local directory.
2. Py files contained in this zip file:

 main.py


 knn\_classifier.py


 svm\_classifier.py


 rf\_classifier.py


 helper.py


3. Output files:


 BalancedRandomForestClassifier\_confusion.png


 DecisionTreeClassifier\_confusion.png


 df\_data\_credit\_history\_dist.png


 df\_data\_heat\_map\_cormatrix.png


 KNeighborsClassifier\_confusion.png

 LinearSVC\_confusion.png

 LogisticRegression\_confusion.png

 RandomForest-BestNd.png

 RandomForestClassifier\_confusion.png

 VotingClassifier\_confusion.png

All the plots are saved to the local directory. Additionally, the df\_summary.csv file is created to store the metrics for each model.

4. Execute the main.py python file located in the root directory.
5. main.py file will load the loan\_data.csv file, will pre-process it, and will evaluate several ML models. All plots are saved to the same local directory where the zip file was extracted.
6. Pre-requisites:
  - a. Dependencies: scikit-learn, NumPy, Pandas, Matplotlib, Seaborn
  - b. Run this command if this package is not installed: `pip install imbalanced-learn`  
(Package that contains BalancedRandomForestClassifier suitable for unbalanced datasets).