**Milestone 2**

**Preliminary classification of Loan Approval**

**Data description:**

1. **Find the generated data file from the milestone 1 -** 'new\_finance.csv'**. Now onwards we will use this preprocessed data file.**
2. **Run the Following Code in Jupyter Notebook and produce the outputs wherever necessary. Answer all questions in your ipynb python jupyter file itself (you can use markdown cells as well wherever necessary) and submit the pdf file of your code run with results. (Failing to not submit the pdf file will result in 100% deduction in marks).**

***Step1#*** *importing new data file finance.csv ->* ***here you have to write the method name and file name in the blanks before running the code.***

**import pandas as pd**

dataset = pd.-----------------(-----------, index\_col=0)

print(dataset)

***Step 2#*** *Now let’s drop duplicate categorical variables that has char numeric / strings / nominal values. If you remember, we have changed these categorical variables into encoding values in milestone one. Pay attention to variable names – do not delete those that have \_cat as suffix. ->* ***here you have to write the code to drop following duplicate variabes.***

(['Loan\_SubType','Loan\_Type','Loan\_Status' ,'Borrower\_Type','Applicant\_State','Type\_ofVehicle','Loan\_Class','Loan\_Tier'], axis=1)

***Step 3#*** *Lets split the data into traing (X) and test sets. Before that we have to drop our dependent/target/label variable (*'Loan\_Status\_Cat'*) from the data. ->* ***here you have to write the code to separate out the or drop the dependent variable.***

***Step 4#*** *Now import the sklearn library for train test split method. # This method will split 75% data as a training set and 25% data as testing set.* ***here you have to write the the code to split the dataset into*** X\_train, X\_test, y\_train, y\_test. ***Moreover in single line statements tell me what are these -*** X\_train, X\_test, y\_train, y\_test.

**Step 5 # Lets train KNN Classifier – *here you have to write the code to import the KNN and run it for the training on*** X\_train, y\_train)

***Step 6*** *# Now that KNN has been trained, so let’s evaluate its performance utilizing the test set that is unknown to the model for the prediction of classes/labels of target variable.* ***You can use your own version of the code but here I have provided a skeleton for the help. here you have to write library, method and variable name in the blank before running the code. Also tell me in your words- a. what you mean by these Recall, precision, F1 Score and Accuracy measures. b. make comment on the KNN’s prediction of loan approval/disprove, is model trained sensibly if yes then how and if not then, what else do you suggest to improve the model’s prediction performance.***

**from sklearn.metrics import** -------------------

y\_pred = knn.-------------(X\_test)

print(----------------------(y\_test, y\_pred))

------------------------\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*------------------------