

ME 781 Group 3

Blitz Credit

Objectives:

Empowering businesses and lives by Instant loans through Modern ML technology and Algorithms.

We are your 24X7 Instant Loan Partner for both the personal and the professional use.

Deciding the loan on basis of following models

- **Classify function for training**
- **Logistic Regression**
- **DecisionTreeClassifier,**
RandomForestClassifier
- **GradientBoostingClassifier**
- **Security and Encryption, Data Integration**

Data set Reference : [LINK](#)

Model Details

- We are using a Multi-Model approach. Considering Binary classification problem Logistic Regression will be first model.
 - In order to make a decision on whether to provide a loan, we will also use Decision Tree and Random Forest.
 - Our data is fairly imbalanced and has missing values as well and so we are using Gradient Boosting too.
-
- We're splitting the dataset in 4:1 ratio for the training and testing purpose i.e. 80% for training & 20% for testing.
 - After fitting the model we are using cross validation with 10 folds ($cv=10$) to evaluate the model's performance
 - We are also incorporating the importance to the particular feature of the dataset and sorting the features by their importance scores in ascending order, so the least important features are at the beginning and thus, features with higher relative importance are more influential in the model's decision-making process.
-
- To prevent much flexibility, we are using a diverse dataset and using cross-validation with more folds to discard the possibility of overfitting of the model and reduce its sensitivity to noisy data.
 - To prevent much rigidity, we are testing our models on Random Forests, Decision Trees, and Logistic Regression and then choosing the one that gives us the least error which would therefore increase the flexibility.
 - Overall our model ensures the flexibility in the feature selection and the giving importance index but at the same time it also ensures the rigid and precise decision of whether to provide loan or not.

Model Training and Testing Report

Logistic Regression

Training Accuracy Score : 81.467 %

Testing Accuracy Score : 78.862 %

Cross-Validation : 80.946 %

Random Forests

Training Accuracy Score : 100.0 %

Testing Accuracy Score : 78.862 %

Cross-Validation : 78.502 %

For choosing the model we will be considering the accuracy of the testing results...
Thus, we are choosing the Gradient Boosting Classifier as the model...

We saw that increasing the no of the cross validation folds just increased the training accuracy and due to which overfitting of data occurred... so we kept the cv=5 only!

Decision Tree Classifier

Training Accuracy Score : 100.0 %

Testing Accuracy Score : 67.480 %

Cross-Validation : 70.688 %

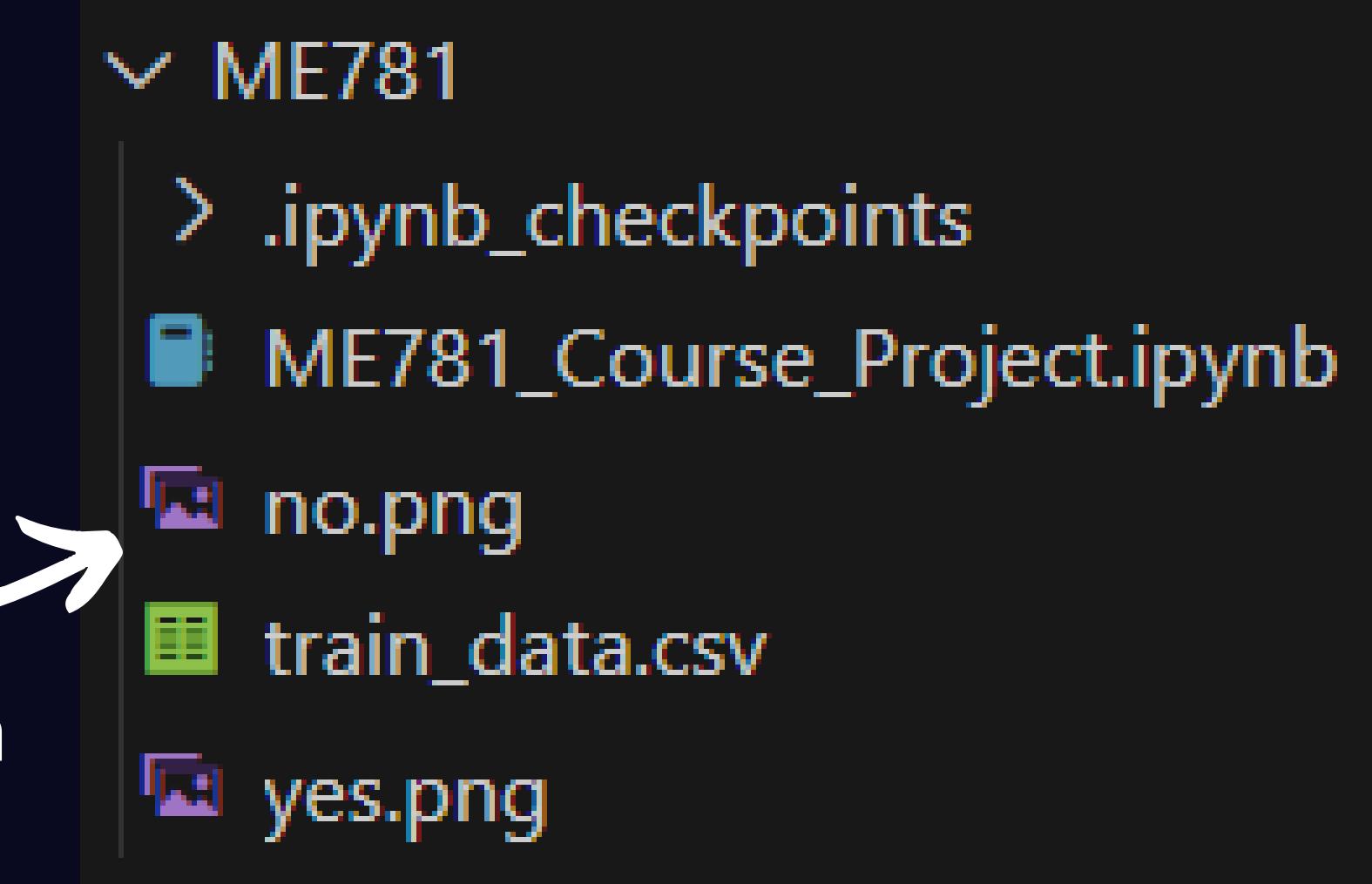
Gradient Boosting Classifier

Training Accuracy Score : 91.242 %

Testing Accuracy Score : 90.244 %

Usage Guide

From this [drive link](#) download the necessary files and put all of them in the same root directory



In the jupyter file run all the cell in the kernel from start and then enter the data of the user as stated

Then run the next cells to see the results if the loan is sanctioned or not!

For getting the result of next user, clear all ouputs and then re enter the data by running cells

For screenshots of user interface and detailed guide... [click here](#)

