IT585 Advanced Machine Learning
Lab5
Handling Class Imbalance Problem

Instructions:

1. You have to code the solution in Google colab
2. You can use inbuilt libraries from python
3. Your plots, code, any insights, observations written as text should be submitted as one ipynb file to google classroom
4. Deadline : Feb 26,2024 11:59 PM IST
5. Name of your file should be : yourrollno_lab5.ipynb

In this lab you will work with datasets having class imbalance and try to handle the issue using various techniques discussed in class. One of the datasets you will use is the Credit Card Fraud Detection and the other is the Loan Defaulter Prediction dataset. Both of these are available on kaggle publicly.

Download the datasets. Apply appropriate preprocessing techniques like cleaning, handling null values, categorical variables etc. Divide into appropriate training and test sets. First apply SVM, logistic regression and random forest classifiers on the dataset without any class imbalance handling techniques. Print the confusion matrix and report your observations (both for training and test datasets). Report also the class imbalance present

Next apply the undersampling, oversampling and SMOTE on the training data to handle class imbalance. You can use the module ***imblearn*** for the same. Each technique is to be applied independently and separately. Train the models SVM, logistic regression and random forest classifier on the data after applying each class imbalance handling technique and report the performance of the trained model on the test data. Experiment with various hyperparameters for different classifiers and also different imbalance handling techniques and report your findings and observations.