Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email, and Contribution:

Team Member's Role:-

- **Sanjaya Kumar Khadanga**

  eMail- skhadanga38@gmail.com

  - Data Understanding
  - Feature Analysis
  - Feature Engineering
  - Exploratory Data Analysis
  - Implementing Logistic Regression
  - Implementing Random Forest
  - Hyperparameter Tuning
  - Evaluating Models

- **Bibhuti Bhusan Sahu**

  eMail- sahubibhuti45@gmail.com

  - Data Understanding
  - Feature Analysis
  - Data Visualization
  - Multivariate Analysis
  - SMOTE
  - SVC
  - ROC AUC Curve
  - Research Analytics
    - Technical documentation

- **Balaram panigrahy**

  eMail- balarampanigrahy42@gmail.com

  - Data Understanding
  - Data Visualization
  - Multivariate Analysis
  - One Hot Encoding
  - XGBoost
  - Research Analytics
    - Technical documentation

Please paste the GitHub Repo link.

Github Link:- https://github.com/Bibhuti-MLAI/Credit-Card-Default-Prediction

Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

Credit cards are being used by customers for online transactions and that is because the facilities banks provide to them are then can get a high limit of purchase and also can pay later. By looking at the facilities the credit card holders for a bank are also more so, the challenge a bank faces is the customers are going to default or not default. By implementing machine learning algorithms we will get some insights about the customers.

The data we have is of a bank having 30000 records having 26 features out of which it holds the past records of the customers. So, our approach starts with analysing the independent features through univariate, bivariate and multivariate analysis of visualization plots. We got statistical insights, presence of outliers and the impact of independent features to the dependent feature.

To prepare the data for model building we also performed Exploratory Data Analysis (EDA) which gives the feature importance. We build classification models to classify the customers and get accuracy scores and values for recall which we are not satisfied with. So we have tuned the hyperparameters to get a good result.

Finally we evaluate the models through recall values and train and test accuracy for the data. We came to the conclusion by looking at the recall values of random forest and xgboost have more accuracy and recall than others i.e. 0.85 .