Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1. Akash Kagdelwar
Email- kagdelwarakash18@gmail.com
   Contribution: Credit Card Default Prediction whole capstone project done   I individually.

Please paste the GitHub Repo link.

Github Link:-https://github.com/AkashK18/Credit-Card-Default-Prediction-ML-Classification

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

Credit card default happens when you have become severely delinquent on your credit card payments. Defaults dose not happens in case of delay or missing few credit card bill payments. When anyone don't pay credit card bill(s) for an extended period of time their card may enter into default status. Default usually happens after six months in a row of not making at least the minimum payment due. Default is a serious credit card status that affects current credit standing and ability to get approved for other credit-based services.

From credit card lender perspective default in credit card payments lead huge losses. Bank and several financial institutions provide credit card to their customers. Several checks are there to issue a credit card to any customer such as credit score etc. So, to ensure profitability and reduce the riskof credit card default, prediction of credit card default plays very important role.

The main objective is aimed at predicting the case of customers default payments in Taiwan. From the perspective of risk management, the result of predictive accuracy of the estimated probability of default will be more valuable than the binary result of classification - credible or notcredible clients.

Prepared data for training after cleaning, preprocessing, transforming into normal distribution and handling class imbalance. Then splitting data into training and testing dataset. Following model were trained using training dataset:
- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine
- Gradient Boosting
- XG Boosting

Cross validation and hyperparameter tunning performed on each of the above model. Evaluated and compared all the evaluation metrics of each model.

From all baseline model, Random Forest classifier shows highest test accuracy and F1 score andAUC. Baseline model of Random Forest and decision tree shows huge difference in train and test accuracy which shows overfitting.

After cross validation and hyperparameter tunning, XG Boost shows highest test accuracy score of 87% and AUC is 0.874. Cross validation and hyperparameter tunning certainly reduce chances of overfitting and also increases performance of model.

Drive Link:- https://colab.research.google.com/drive/1dqGs0iyDBWJ1iox1DYHPe-LraAMRRUMk