Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1) **Mohd Danish:** **Email:** mdanish63364@gmail.com
   1) Feature Engineering:
      - Data preprocessing
      - One-Hot encoding
   2) EDA(Exploratory Data Analysis):
      - Count plots
      - Boxplot
      - Pairplot
   3) Imbalanced dataset
      - SMOTE
   4) Classification Analysis:
      - Logistic Regression
      - Decision Tree Classifier
      - Random Forest Classifier
      - XG-Boost Classifier
   5) Model Explainability
      - Shapley
        - Summary Plot
   6) ROC-AUC curve
   7) Group Colab
2) **Abdul Rahman Talha:** **Email:** rahman88talha@gmail.com
   1) Feature Engineering:
      - Data Preprocessing
      - One-Hot encoding
   2) EDA(Exploratory Data Analysis):
      - Countplots
      - Pairplot
      - Heat-Map
   3) Hyperparameter tuning and Cross-validation
   4) Classification Analysis:
      - Logistic Regression
      - Decision Tree Classifier
      - XG-Boost Classifier
   5) PPT
3) **Sharath Diwakar:** **Email:** mailmesharathd@gmail.com
   1) Data Munging:
      - Introducing New variables
   2) Data Visualization:
      - Countplot
      - Boxplot
      - Confusion Matrix (Heat-Map)
   3) Classification Analysis:
      - Logistic Regression
      - Random Forest Classifier
      - XG-Boost Classifier

4) Feature Importance:
    - Random Forest Classifier

Please paste the GitHub Repo link.

GitHub Link: Rahman88talha/Credit_Card_Default_Prediction_Classification: Credit Card Default Prediction- Capstone Project ML(Classification) (github.com)

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 has become a challenge for the companies to keep record of all customers who make late payments and who makes it on time to decide whether their limits should be increased or not. So, we will try and find the customers who are defaulters or Non-defaulters.

This project is aimed at predicting the case of customers' default payments in Taiwan.

Modelling: - We will build Three models

· Logistic Regression
· Decision Tree Classifier
· Default XGBoost Classifier

Performance Metrics

· Precision is a good metric to use when the costs of false positive(FP) is high. Precision = TP / (TP + FP)
· Recall is a good metric to use when the cost associated with false negative(FN) is high. Recall = TP / (TP + FN)
· F1-score is a weighted average of precision and recall. Thus, it considers FP and FN. This metric is very useful when we have uneven class distribution, as it seeks a balance between precision and recall. F1-score = 2 (precision recall) / (precision + recall)

Conclusion:

- Only 22% Clients were defaulters of all
- 12% Female & 9% Male are Default
- University 11%, Graduate 7% & 4% High schooler were defaulter
- Highest number of defaulters were between 27-30 of age.

Three models have been applied to assess the which model is best to evaluate the defaulter (Logistic Regression, Decision Tree Classifier, XGBoost Classifier).

- Using a Logistic Regression classifier, we can predict with 68.55% accuracy, whether a customer is likely to default next month.
- Using a Decision Tree classifier, we can predict with 73.72% accuracy, whether a customer is likely to default next month.

- Using a Default XGBoost Classifier, we can predict with 76.63% accuracy, whether a customer is likely to default next month.

The best accuracy is obtained using Default XGBoost Classifer

- XGBoost Classifier & Decision Tree Classifier are giving us the best Recall, F1-score and ROC Score among others. We can conclude that these two algorithms are the best to predict our Credit Card Default Analysis.