**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. **Adil Khan (adil.bm009@gmail.com)**    1. Data Processing    2. Data Wrangling       1. Importing some important modules       2. Summarising the data    3. Credit card default prediction       1. Pre-processing the data   1.3.1.1 Finding the missing values  1.3.1.2 Searching for the duplicate values  1.3.1.3 Changing the Column names  1.3.1.4 Checking the categorical columns and assigning appropriate values  1.3.1.5 Checking the numerical columns  1.3.1.6 Changing the Data types  1.3.2 Exploratory Data Analysis  1.3.2.1 Univariate Analysis  1.3.2.2 Analysis of Dependent Variable  1.3.2.3 Analysation of Categorical Variables  1.3.2.4 Data visualisation  1.4 Checking the Correlation  1.5 Supervised ML  1.5.1 Bagging Classifier  1.5.2 Gradient Boosting Algorithm  1.5.3 XG Boost  1.6 Observations   1. **Sunil Kumar Panigrahi(**[**sunilpanigrahi104@gmail.com**](mailto:sunilpanigrahi104@gmail.com)**)**   2.1 Data Wrangling  2.2 Exploratory Data Analysis  2.2.1 Analysis of Numerical variables displots  2.2.2 Bivariate analysis  2.2.3 Analysis of important factors to understand the data  2.2.4 Data Visualisation  2.3 Normalised the column data  2.4 Checking of Correlation between Variables  2.5 MultiCollinearity Heatmap  2.5.1 Correlation Heatmap  2.6 Created the Dummy Variables  2.6.1 One hot encoding  2.7 Supervised ML  2.7.1 Random Forest Classification  2.7.2 Gradient Boosting Algorithm  2.7.3 XG Boost  2.8 Observations  2.9 Presentation PPT   1. **Vivek Singh (sviveksingh31@gmail.com)**   3.1 Data Wrangling  3.2 Exploratory Data Analysis  3.2.1 Univariate Analysis  3.2.2 Analysis of important factors to understand the data  3.2.3 Data Visualisation  3.3 Model Building  3.4 Supervised ML  3.4.1 Logistic Regression  3.4.2 Support Vector Classification  3.4.3 Bagging Classifier  3.5 Observations  4.1 Conclusions  4.2 Presentation PPT  4.3 Technical Documentation     1. **Shubham Kumar (Samsonknight25@gmail.com)**   4.1 Data Wrangling  4.2 Data Cleaning  4.3 Data pre-processing  4.4 Summarising the data  4.5 Exploratory Data Analysis  4.5.1 Univariate Analysis  4.5.2 Bivariate Analysis  4.5.3 Analysis of important factors to understand the data  4.5.4 Data Visualisation  4.6 Supervised ML  4.6.1 Logistic Regression  4.6.2 Random Forest Classification  4.6.3 Gradient Boosting Algorithm  4.6.4 XG Boost  4.7 Observations  4.8 Conclusions  4.9 Presentation PPT  4.10 Technical Documentation   1. **Sharaffin B (sharaffinb@gmail.com)**   5.1 Data Processing  5.2 Data Wrangling  5.3 Importing some important modules  5.4 Summarising the data  5.5 Pre-processing  5.6 Exploratory Data Analysis  5.6.1 Univariate Analysis  5.6.2 Bivariate Analysis  5.6.3 Analysis of all the important factors to understand the data  5.6.4 Data Visualisation  5.7 Supervised ML  5.7.1 Random Forest Classification  5.7.2 Support Vector Classification  5.7.2 Gradient Boosting  5.7.3 XG Boost  5.7.4 Hyperparameter Tuning  5.8 Technical Documentation  5.9 Presentation PPT  5.10 Observations  5.11 Conclusions |
| **Please paste the GitHub Repo link.** |
| Github Link: <https://github.com/adilbm009/-Credit-Card-Default-Prediction> |
| **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 risk plays a major role in the banking industry. Banks' main activities include granting loans, credit cards, investments, mortgages and more. One of the most important is the credit card. However, as the number of credit card users increases, banks are facing an increase in credit card default rates. In addition, many customers use their credit cards beyond their repayment capacity, resulting in a high level of debt accumulation. Machine learning provides solutions for addressing this issue.  The basic purpose of this project is to analyse credit card data collected from Taiwan-based credit card issuers, predict whether consumers will default on their credit cards, and identify the main reasons. is. This would inform the issuer’s decisions on who to give a credit card.  Let's take a look at what we did in particular. First, we loaded the dataset and performed data cleaning and null value treatment. There were no null values or duplicate values present in the dataset. We did outlier treatment and did some EDA and data visualisation. We plot different graphs of univariate analysis and bivariate analysis and make different inferences from our dataset.  Lastly, we plot the correlation heatmap and find the correlation of each independent variable with our target variable.  In the modelling part, we tried six different classification models. The dataset was not normally distributed and mostly it was right skewed. The conclusion that we came up with are:   1. Male customers have more default payments than female customers. 2. There is no significant difference in the proportion of default payment across different education levels. But clients who have lower than high school level education tend to have default payment more. 3. Married clients have a higher default payment rate than single or other marital status clients. 4. People who have the payment delay for two months have a high ratio of default next month (October). In September, a quarter of customers who repay one month later have default payment next month in October. This situation does not exist in other months as almost no one repays one month later. 5. XGBOOST has the highest accuracy of 82.5 % with a recall of 67.7 % and KS chart value of 0.447 and proved to be the best model.   The developed models took into account all possible factors and data. This final chosen model would benefit the bank before they make any decisions against that customer. |