**Capstone Project Submission**

**Instructions:**

1. Please fill in all the required information.
2. Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Member 1: Saaquib Mustafa**  **Email:** [saaquibmustafa26@gmail.com](mailto:saaquibmustafa26@gmail.com)  **Contributions:**   1. Clean and prepare the data for analysis. 2. Done Initial analysis and helped in visualization. 3. Prepared Project Summary 4. Prepared Key Notes and conclusion     **Member 2: Raja Chowdhury**  **Email:** [rajachowdhury2468@gmail.com](mailto:rajachowdhury2468@gmail.com)  **Contributions:**   1. Done the visualization for analysis. 2. Helped in analysis and Data Cleaning 3. Added Useful Codes to simplify the analysis. 4. Prepared conclusions and PPT 5. Prepared introduction and key finding     **Member 3: Sandipan Das**  **Email:** [sandipan.das202@gmail.com](mailto:sandipan.das202@gmail.com)  **Contributions:**   1. Prepared Technical Documentation 2. Helped in Data Cleaning 3. Help in Summary Preparation & PPT 4. Helped in Key Notes     **Member 4: Sahil Kolambakar**  **Email**: [sahilrajkolambakar@gmail.com](mailto:sahilrajkolambakar@gmail.com)  **Contributions:**   1. Prepared Project Presentation 2. Helped in Data Cleaning 3. Help in Summary Preparation & Technical Documentation 4. Helped in Key Notes and Conclusion |
| **Please paste the GitHub Repo link.** |
| **GitHub Link: -** https://github.com/Sandipan0303/credit-card-default-prediction.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  **Problem Statement:** This project is aimed at predicting the case of customer 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 not credible clients. We can use the [K-S chart](https://www.listendata.com/2019/07/KS-Statistics-Python.html) to evaluate which customers will default on their credit card payments.The task is to predict whether the customer will default his/her payment or not.The dataset has 25 columns and 30000 rows.Analysing the data to discover key understandings (not limited to these) such as:What is the percentage of defaulters and non defaulters?What is the ratio of males and females in our dataset?Among male and female who is more prone towards defaulting?Among married unmarried and others who are more towards defaulting?Summary:Less people will default on their payments than will pay on time, which means that the number of defaulters will be substantially lower. In our sample, there are significantly more females than males (4926), and we may also conclude that women are more likely than men to pay their defaults on time. Compared to singles and other groups, married people are more likely to pay their default payments as it can be assumed that these all sorts of burden might create an impact in their family life. So they tend to avoid it. Additionally, we may say that single people do not pay their default payment much less than their married peers. We will attempt to feature engineer bill amount to Dues column in order to reduce multicollinearity since Bill amount for each month is highly connected with each other. We can see that pay amount tends to clutter at one place hence we can try to feature engineer pay amount columns to one payment column to declutter our data. From history of past payment analysis, it is clear that most people who pay duly are not likely to default their payment.Here, we can say from age 21 to 39 limit balance is increasing. In our machine learning model among all the algorithms random forest classifier showed us the best results with an average evaluation score of around 84%. |
|  |