**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Rinkesh Das  Email Id- [rinkeshdas2001@gmail.com](mailto:rinkeshdas2001@gmail.com)  Contribution- Data Cleaning, EDA, Model Training, Evaluation, Validation |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/rinkeshdas01/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)**  **Problem Statement-**Credit card default happens when you have become severely delinquent on your credit card payments. In case of a credit card payment default, the banks have to go through unnecessary trouble and banks are not willing to do so.  **The project involves the following steps:-**   1. Data Cleaning 2. EDA 3. SMOTE 4. Model Training and Implementation 5. Model Evaluation and Validation   **After performing the following steps the following conclusions were obtained:-**   * While validating the models more importance should be given to 'recall scores' since the priority of the banks is to identify a potential defaulter. Whereas if a user is wrongly detected as a potential defaulter there will not be much problem. * Some of the models perform bad on both the train and test data. (Logistic Regression, SVC). They seem to be underfitting the train data. * Some of the models perform good on train data but not so good on the train data. (Decision Tree Classifier). They seem to be overfitting the train data. * Random Forest, XG Boost, Gaussian Naïve Bayes gives us decent recall scores that are 0.82, 0.84, 0.89 respectively. Whereas KNN gives a very good recall score of 0.94. * So we can select KNN as the optimum model for our problem statement. |
|  |