**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Name: Rushabh Tikale  Email: [rushabhtikale62@gmail.com](mailto:rushabhtikale62@gmail.com)  Contribution:  1.Import library.  2.Import csv data.  3.Checking null value.  4.Checking and removing outlier.  5.Exploring Dependent and Independent variable.  6.EDA(Exploratory Data Analysis)  7.Relation between dependent and independent variable.  8.Correlation in Variables.  9.Prepared data for train and test.  10.Build baseline model.  11.Use evaluation metrics.  12.Compare the best fitted model.  13.Choose the best predictor model.  14.Conclusion. |
| **Please paste the GitHub Repo link.** |
| Githubs Link:-  Drive Link:- |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  The credit card default is one of the major issues increasing day by day, to solve this we develop a ML model .ML model help us to catch the credit card default customer and try to reduce the fraudster customer.  The data show us the behavior of customer as like payment history for different months, how much the bill is come and how much they paid.  We try to catch the fraudster and the bank does not repeat the mistake.  If the customer does not paying their installment month to month so it has the probability that customer will default for their transaction and loose their claim from credit card.  In short **Credit Card Default Prediction** is to build a model to identify whether the**credit card** applicant will default or not based on his repayment history and other important factors. In order to achieve this, we need to develop a supervised machine learning model using classification algorithms. The **Problem statement** is aimed at predicting the case of customers default payments. 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 evaluate which customers will default on their credit card payments**.** The model should have a good accuracy as the default customer will not leave from catching. As it is one of the increasing case amongst all over country so to overcome this we develop a ML model.  So the **Approach** is to deal with a large feature of dataset simaltaneously and make a proper data for training and testing. For good accuracy we need to make a good model so we clean the data and make some more analyzable data, remove outlier, fill all the null values as to overrid with missing information and model should not consider this missing value. The feature contain some multicollinearity and we have to deal with it, so we just use VIF method to remove it and get our dataset with good predictive score. The datase was imbalanced so we used SMOTE to take resample of the data. After that we done EDA on the dataset and get some insights for the model, as EDA help us to know the features very well and the history of the default customer for their payment.  From this we get a **Conclusion** that We train the different model like Decision Tree Classifier, Random Forest Classifier, KNN and Gradient Boosting classifier and predict the score simaltaneously. The Decision Tree and Gradient Boosting Classifier are not giving us the best prediction for model and on the another side, KNN and Random Forest giving the best prediction. There is almost same default score for bothmale and female customers. we analyse customer history of use of credit card from different months and there is less default customers who is paying their payment monthly and use as revolving card. Choosing Hyperparameter tuning is the case sensitive model so we have to take very much care of it. By visualising relation with dependent and independent variable we covered information of default customers by Education, Age, Marriage, Sex. Each feature giving a major information about total number of defaulters in this class. In Random Forest classifier we are getting the positive outcomes so from this our conclusion is that decision tree is the best predictive model. |