# Capstone Project Submission

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

**Team Member's Name, Email and Contribution:**

1. **Name** → Meenakshi
   **Email** → meenakshicuul@gmail.com
   **Role :**
   - **Data Munging**
     1. **Change unconventional feature names**
     2. **Checking for duplicate and null values**
     3. **Log transformation**
   - **Data Visualization**
     1. **Distribution of categorical columns**
     2. **Heat map**
     3. **Pair Plot**
     4. **Box plot**
     5. **Count plot**
     6. **Confusion Metrics**
   - **Correlation Analysis by Heat map**
     1. **Between independent variables**
     2. **Between dependent and independent variables**
   - **Model**
     1. **Logistic Regression (With Cross Validation)**
     2. **Random Forest Classifier (With Cross Validation)**
     3. **Support Vector Classifier(With Cross Validation)**
     4. **K-Neighbor Classifier (With Cross Validation)**
     5. **XG Boosting(With Cross Validation)**
     6. **Gaussian Naïve Bayes Classifier**
   - **PPT**
   - **Group Colab**
2. **Name** → Tushar R. Wagh
   **Email** → waghtushar7276@gmail.com
   **Role :**
   - **Data Munging**
     1. **Change unconventional feature names**
     2. **Checking for duplicate and null values**
     3. **Log transformation**
   - **Data Visualization**
     1. **Distribution of numerical columns**
     2. **Distribution of categorical columns**
     3. **Heat map**
     4. **Scatter plot**
     5. **Line plot**
     6. **Confusion Metrics**
   - **Correlation Analysis by Heat map**
     1. **Between independent variables**

2. **Between dependent and independent variables**

- **Model**
  1. **Logistic Regression**
  2. **Random Forest Classifier**
  3. **Decision Tree Classifier**

- **Technical Documentation**

3. **Name** → Aditya Singh Thakur
   **Email** → imchillingadi@gmail.com
   **Role :**
   - **Data Munging**
     1. **Change unconventional feature names**
     2. **Checking for duplicate and null values**
     3. **Z-Score transformation**

   - **Data Visualization**
     1. **Distribution of categorical columns**
     2. **Heat map**
     3. **Pair plot**
     4. **Count Plot**
     5. **Box Plot**

   - **Correlation Analysis by Heat map**
     1. **Between independent variables**
     2. **Between dependent and independent variables**

   - **Model**
     1. **Logistic Regression**
     2. **Random Forest Classifier**
     3. **Gaussian Naïve Bayes Classifier**
     4. **Linear SVC**

   - **PPT**

**Please paste the GitHub Repo**

Github Link : https://github.com/meena25091992/credit_card_defalut_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)**

We are all aware what is credit card. It is type of payment card in which charges are made against a line of credit instead of the account holder's cash deposits. When someone uses a credit card to make a purchase, that person's account accrues a balance that must be paid off each month.

Credit card default happens when you have become severely delinquent on your credit card payments. Missing credit card a payment once or twice does not count as a default. A payment default occurs when you fail to pay the Minimum Amount Due on the credit card for a few consecutive months.

Objective of our project is to predict which customer might default in upcoming months.

We have to build models which help us to predict the defaulters.

I have applied various Classification Models in our Credit-Card-Default-Prediction such as follows:-

1. Logistic Regression (with Cross Validation)
2. Random Forest Classifier (with Cross Validation)
3. XG Boosting (with Cross Validation)
4. Support Vector Classifier
5. K-Neighbor Classifier (with Cross Validation)
6. Gaussian Naïve Bayes Classifier

**Some insights:-**

1. There are some features which is having negative correlation like “Age” and “Marriage”

2. Top 3 models are **Random Forest, KNeighbor Classifier and Support Vector Classifier** that gives best Presision, Recall, ROC_AUC and F1 score.

3. **Random Forest Classifier** performs best among all models.

4. **Logistic Regression and Gaussian Naive Bayes Classifier** is not giving best precision score.

5. We have found the proportion of defaulters with respect to Marriage, Education, Sex feature and we found that :

* Most of the defaulters are Female
* Most of the defaulters are from university
* Marital status is Single
* More no. of defaulters are Single