**Project 1) -Case Study Banking- Credit Card Spend Predictiona and Identify Drivers of Spend**

**Objective** – The objective of this case study is to understand what's driving the total spend (Primary Card + Secondary card). Given the factors, predict credit limit for the new applicants

**Business Problem** –

One of the global banks would like to understand what factors driving credit card spend are. The

bank want use these insights to calculate credit limit. In order to solve the problem, the bank

conducted survey of 5000 customers and collected data.

**Data Availability** –

 Data for the case are available in xlsx format.

 The data have been provided for 5000 customers.

 Detailed data dictionary has been provided for understanding the data in the data.

 Data is encoded in the numerical format to reduce the size of the data however some of the

variables are categorical. You can find the details in the data dictionary

**Historical Data** –

* Records of 5000 credit card users customers data data
* It had around 130Features

**Tools/Libraries Used –** Python (numpy, pandas, sklearn,scipy,matplotlib,statsmodels , seaborn,etc.), Excel

**Techniques –**

* Identified various factors that affect the total turnaround time for a successful ticket closure using **Random Forest Algorithm.**
* Used machine learning models like Decision Trees,Ridge regression,Lasso Regression,Ensemble learning,Bagging,Boosting,Gradient Boost,XGBoost,KNN and linear regression traditional approach.
* For dataset’s Graphical representation and insights used **Matplotlib and Seaborn**

**Metrices**-

* For Regression MAPE and RMSE as a metrics, and it is best metric to accuracy. If GINI value is high our model accuracy is good.

**Tuning Parameters –**

* Used techniques like random forest for selecting the factors which are causing delayed TAT for this we need **turning parameters like number of trees, depth of the tree, random state and iterators .**
* **KNN tuning parameters like n\_nieghbours and leaf size.**

**Validation –**

* It will train my model using train data set I will verify result against test data set. If we will get MAPE and RMSE value nearly then I will say my model is good model.

**Challenges** –

* Data Collection
* Feature Selection

* Identify the Positive Drivers

------------------------------------------------------------------------------------------------------------------------------------------