

**CREDIT CARD**

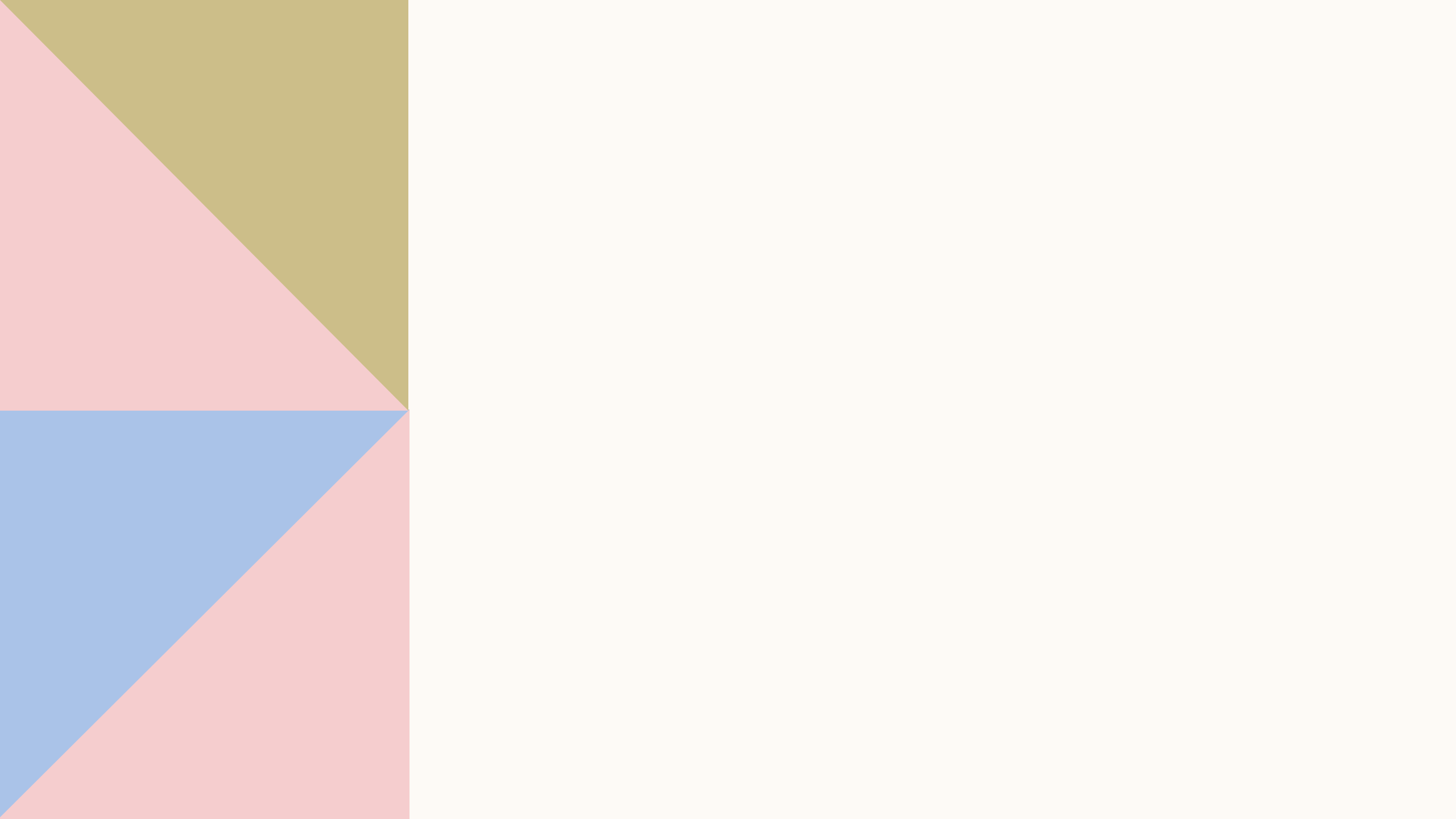
**DEFAULT**

**PREDICTION**

Yash Kumar Singh

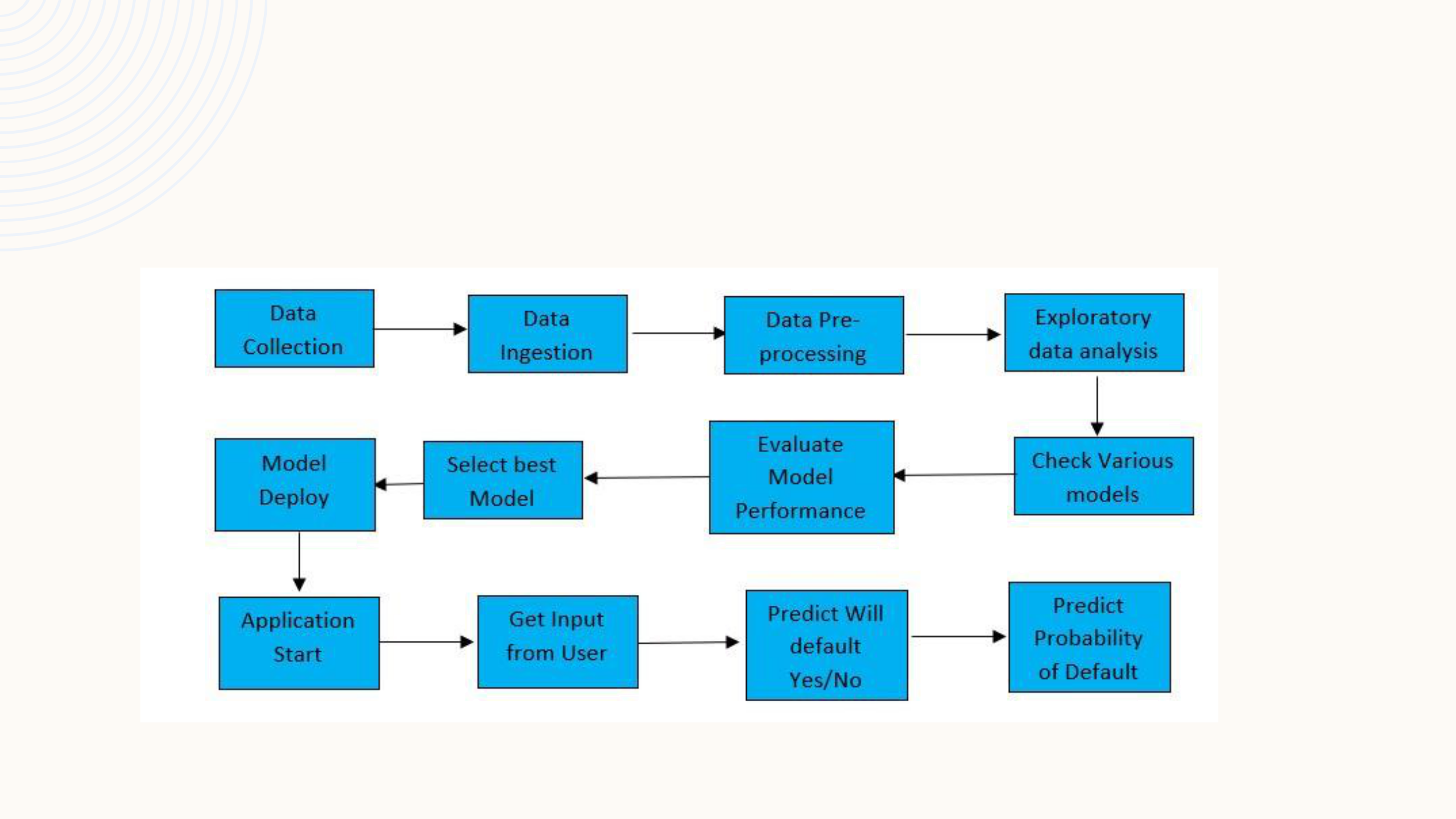
Vaibhav Srivastava

|  |  |
| --- | --- |
| Credit Card Default Prediction Project | 2 |



**INTRODUCTION**

* Financial threats are displaying a trend about the credit risk of commercial banks as the incredible improvement in the financial industry has arisen.
* In this way, one of the biggest threats faced by commercial banks is the risk prediction of credit clients.
* The goal is to predict the probability of credit default based on credit card owner's characteristics and payment history.

Credit Card Default Prediction Project 3

**ARCHITECTURE**

|  |  |  |
| --- | --- | --- |
| Credit Card Default Prediction Project | 4 |  |
|  |  |

**DATASET**

* **The data here has almost no obfuscation and is provided in a CSV file whose schema is described in the first row. This data has more than 20 million transactions generated from a multi-agent virtual world simulation performed by IBM. The data covers 2000 (synthetic) consumers resident in the United States, but who travel the world. The data also covers decades of purchases, and includes multiple cards from many of the consumers**.
* DATA SET LINK-

[Credit Card Transactions | Kaggle](https://www.kaggle.com/datasets/ealtman2019/credit-card-transactions)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Variable Name** | **Measurement** |  | **Description** | |
|  |  | **Unit** |  |  |  |
|  | LIMIT\_BAL | NT Dollar |  | Amount of given credit |  |
|  | SEX | Integer | Gender (1=male, 2=female) | | |
|  | EDUCATION | Integer | 1=graduate school, 2=university, 3=high | | |
|  |  |  | school, 4=others, 5=unknown, 6=unknown | | |
|  | MARRIAGE | Integer | Marital status (1=married, 2=single, | | |
|  |  |  | 3=others) | | |
|  | AGE | Years | Age of the person in years | | |
|  | PAY\_0-6 | Integer | Repayment status for various months | | |
|  | BILL\_AMT1-6 | NT Dollar | Amount of billed statements for various | | |
|  |  |  | months | | |
|  | PAY\_AMT1-6 | NT Dollar | Amount of Previous payments done | | |
|  | default.payment.next.month | Binary | Will Customer default? Yes/No | | |
|  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| Credit Card Default Prediction Project | 5 |  |
|  |  |

**DATA PRE-PROCESSING**

* Drop the column which is statistically not Important, here I dropped the ID column
* Convert variables: SEX, EDUCATION, MARRIAGE into object as they are categorical variables
* Separate Categorical and Continuous variable.
* For Continuous variables Scale the model with StandardScaler or MinMaxScaler if necessary for model. Scaling is not necessary for tree-based models.
* Perform One-Hot Encoding on Categorical variables
* Join Continuous and One Hot Encoded Variables
* Data Pre-processing is done.

|  |  |  |
| --- | --- | --- |
| Credit Card Default Prediction Project | 6 |  |
|  |  |

**MODEL BUILDING**

* I have built various classification models like – Decision Tree, Random Forest, XGBoost, K-Nearest-Neighbours, MLP(Multi Layer Perceptron).
* Base model of each above model was created
* Hyperparameter tuning for each model was done using 4-Fold GridSearchCV
* Model with best accuracy score and which required less training time was selected
* Classification report of Models was also checked

|  |  |  |
| --- | --- | --- |
| Credit Card Default Prediction Project | 7 |  |
|  |  |

**MODEL EVALUATION**

* Metrics used for Evaluation are : Accuracy Score and Classification Report
* A Classification report is used to measure the quality of predictions from a classification algorithm.
* How many predictions are True and how many are False
* More specifically, True Positives, False Positives, True negatives, and False Negatives are used to predict the metrics of a classification report.

Credit Card Default Prediction Project 8

**BEST MODEL - XGBOOST**

|  |  |  |
| --- | --- | --- |
| Training | 0.8236 |  |
| Accuracy |  |
|  |  |
|  |  |  |
| Testing | 0.8230 |  |
| Accuracy |  |
|  |  |
|  |  |  |

|  |  |
| --- | --- |
| Parameters | Value |
| n\_estimators | 500 |
| Learning\_rate | 0.01 |
| max\_depth | 4 |

Classification Report of Testing Data

Credit Card Default Prediction Project 9

**FEATURE IMPORTANCE**



|  |  |  |
| --- | --- | --- |
| Credit Card Default Prediction Project | 10 |  |
|  |  |

**DEPLOYMENT**