



CREDIT CARD DEFAULT PREDICTION

Architecture Document

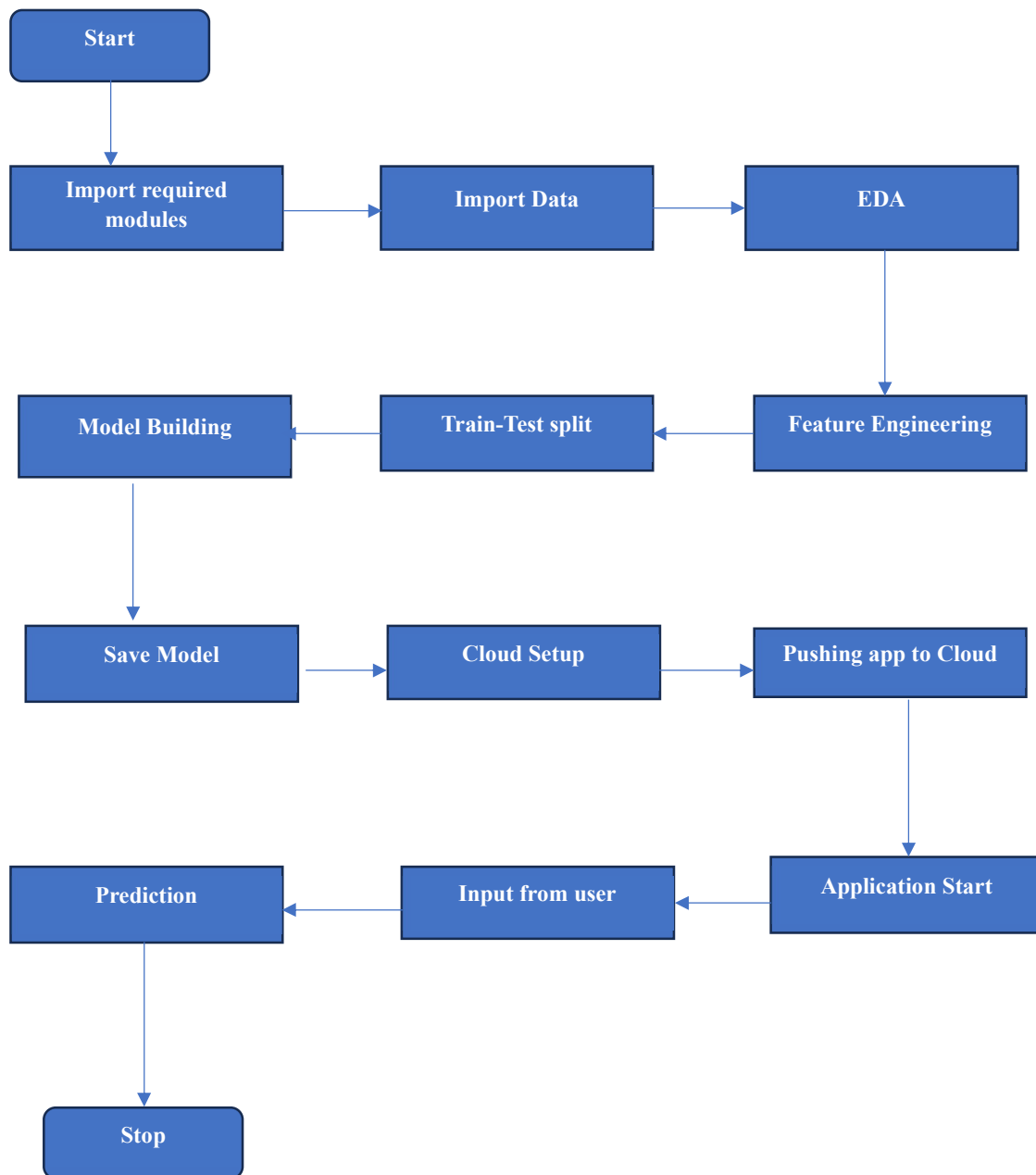


Prepared By:
Binaya Kumar Pradhan

Table of Contents

- 1. **Architecture** 2
- 2. **Architecture Description** 3
 - 2.1. Data Description 3
 - 2.2. Data Exploration 4
 - 2.3. Feature Engineering 4
 - 2.4. Train Test Split 4
 - 2.5. Model Building 4
 - 2.6. Save the Model 4
 - 2.7. Cloud Setup & Pushing the App to the Cloud 4
 - 2.8. Application Start & Input Data by User 5
 - 2.9. Prediction 5

1. Architecture



2. Architecture Description

2.1.Data Description

This dataset contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card clients in Taiwan from April 2005 to September 2005.

There are 25 variables:

- ID: ID of each client
- LIMIT_BAL: Amount of given credit in NT dollars (includes individual and family/supplementary credit)
- SEX: Gender
 - 1=male,
 - 2=female
- EDUCATION:
 - 1=graduate school,
 - 2=university,
 - 3=high school,
 - 0, 4, 5, 6=others)
- MARRIAGE: Marital status
 - 1=married,
 - 2=single,
 - 3=divorce,
 - 0=others
- AGE: Age in years
- PAY_0: Repayment status in September, 2005
 - -2: No consumption;
 - -1: Paid in full;
 - 0: The use of revolving credit;
 - 1 = payment delay for one month;
 - 2 = payment delay for two months; . . .;
 - 8 = payment delay for eight months;
 - 9 = payment delay for nine months and above.
- PAY_2: Repayment status in August, 2005 (scale same as above)
- PAY_3: Repayment status in July, 2005 (scale same as above)
- PAY_4: Repayment status in June, 2005 (scale same as above)
- PAY_5: Repayment status in May, 2005 (scale same as above)
- PAY_6: Repayment status in April, 2005 (scale same as above)
- BILL_AMT1: Amount of bill statement in September, 2005 (NT dollar)
- BILL_AMT2: Amount of bill statement in August, 2005 (NT dollar)
- BILL_AMT3: Amount of bill statement in July, 2005 (NT dollar)
- BILL_AMT4: Amount of bill statement in June, 2005 (NT dollar)
- BILL_AMT5: Amount of bill statement in May, 2005 (NT dollar)
- BILL_AMT6: Amount of bill statement in April, 2005 (NT dollar)
- PAY_AMT1: Amount of previous payment in September, 2005 (NT dollar)
- PAY_AMT2: Amount of previous payment in August, 2005 (NT dollar)

- PAY_AMT3: Amount of previous payment in July, 2005 (NT dollar)
- PAY_AMT4: Amount of previous payment in June, 2005 (NT dollar)
- PAY_AMT5: Amount of previous payment in May, 2005 (NT dollar)
- PAY_AMT6: Amount of previous payment in April, 2005 (NT dollar)
- default. payment. next. Month: Default payment
 - 1=yes,
 - 0=no

2.2.Data Exploration

The data is categorized into two types: numerical and categorical. We conduct a detailed exploration for each type, one at a time. Within each type, we systematically examine, visualize, and analyze each variable individually, documenting our findings. Additionally, we may make minor modifications to the data, such as renaming columns for improved clarity and ease of understanding.

2.3.Feature Engineering

Categorical variables have been encoded to facilitate data analysis and modelling

2.4.Train Test Split

The dataset has been divided into two subsets: a training set, which comprises 70% of the data, and a test set, which consists of the remaining 30%. This split allows for training and testing machine learning models.

2.5.Model Building

Several models have been constructed, and the dataset has been used to train and evaluate these models. The performance of each model has been thoroughly compared, and the best-performing model has been selected based on various evaluation metrics and criteria.

2.6.Save The Model

The selected model has been saved by converting it into a pickle file. This allows for easy storage and retrieval of the model for future use.

2.7.Cloud Setup & Pushing the App to The Cloud

AWS (Amazon Web Services) has been chosen as the deployment platform for the application. The application files have been loaded from the GitHub repository to the AWS environment, ensuring that the application is hosted and accessible on AWS infrastructure.

2.8. Application start & input data by user

The application has been initiated and is now ready for use. You can enter the required inputs into the application to perform the desired tasks.

2.9. Prediction

Once you've submitted the inputs, the application will execute the model and generate predictions. The output will be displayed as a message, providing information about whether the customer, whose demographic and behavioural data were entered as inputs, is likely to default in the following month or not.