Credit Card Default Prediction

Project Report Presentation

By:- R. Praveen Kumar

Data Science Intern

Ineuron

- Introduction
- Objective
- Data Description
- Architecture
- Model Training and Evaluation Workflow
- Deployment
- Questions

Introduction

Credit risk plays a major role in the banking industry business. Banks mainly generate income through performing activities like loans, credit cards, investments, mortgages, and others.

Credit card has been one of the most booming financial services offered by banks over the past years. However, with the increasing number of credit card users, banks have been facing the issue of increasing credit card default rates.

Data science can provide solutions to resolve the current phenomenon and management credit risks. This project discusses the implementation of a model which predicts if a given credit card holder has a probability of defaulting in the following month or not, using their demographic data and behavioral data from the past 6 months.

Objective

Development of a model for predicting if a given customer has a probability to default in the next month or not.

Benefits:

- Detection of upcoming frauds.
- Gives a better understanding of the customer base.
- Allows financial institutions to take necessary steps like:-
  - Decreasing the current limit.
  - Blocking all unnecessary transactions and allowing grocery and necessary items.

About Data

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

cont...

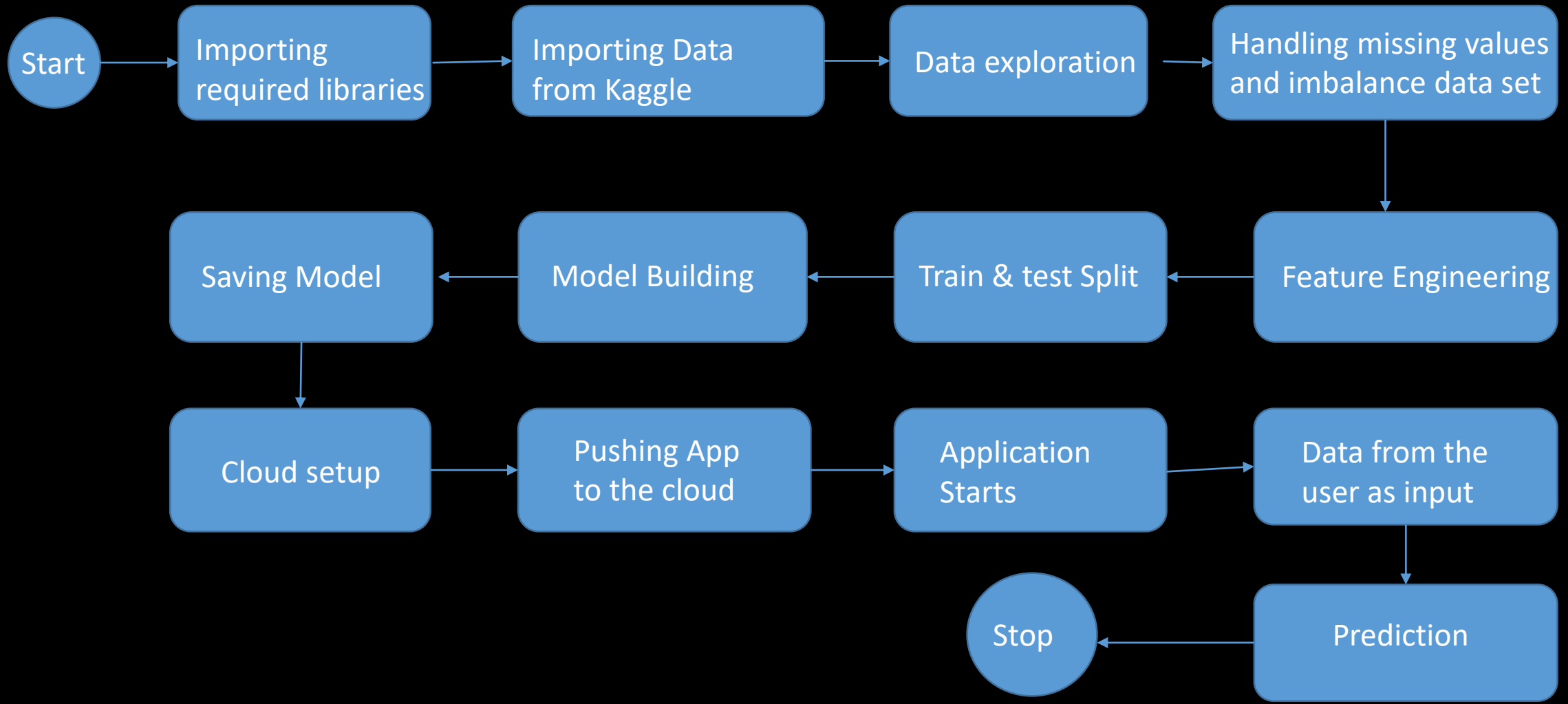
AGE: Age in years
PAY_0: Repayment status in September, 2005
-1: Paid in full;
0: No consumption;
1 = payment delay for one month;
2 = payment delay for two months; . . .;
8 = payment delay for eight months;
9 = payment delay for nine months and above.
PAY_2: Repayment status in August, 2005 (scale same 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)

cont..

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

Architecture



Data Exploration

We have plotted many graphs to understand the data, there were many outcomes from the like, compared to male females holding more credit cards, and by exploring, we understood that the data was unbalanced.

Handling outlier & imbalance data

There are many outliers were detected when we plotted the box plot and in place, to trimming we go with capping to maintain the data. To handle the imbalanced data we went with Over sampling because with under-sampling we don’t want to lose data.

Feature Engineering

We created a new Feature taking the average Bill Amount for the last six months.

Save the model

Saved the model and Standard Scaler by converting it into a pickle file.

Cloud Setup & Pushing the App to the Cloud

Selected Streamlit CLoud for deployment. Loaded the application files from GitHub to Streamlit CLoud.

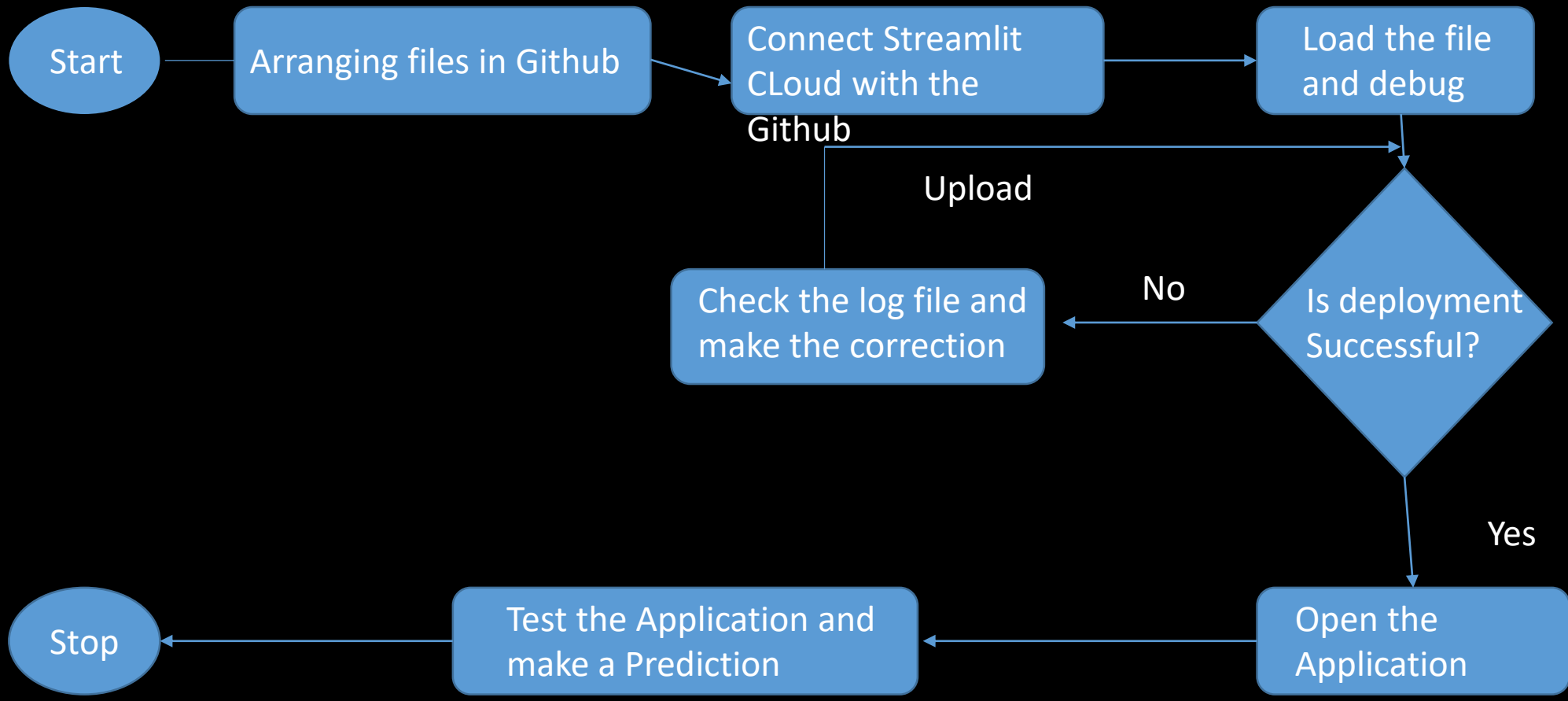
Application Start and Input Data by the User

Start the application and enter the inputs.

Prediction

After the inputs are submitted the application runs the model and makes predictions. The out is displayed as a message indicating whether the customer whose demographic and behavioral data are entered as inputs, is likely to default in the following month or not.

Deployment


Start
Arranging files in Github
Connect Streamlit CLoud with the Github
Load the file and debug
Upload
Is deployment Successful?
No
Check the log file and make the correction
Yes
Open the Application
Test the Application and make a Prediction
Stop

Q&A

1) What is the data source?

The data is obtained from Kaggle. Link https://www.kaggle.com/uciml/defaultof-credit-card-clients-dataset

2) What was the type of data?

The data contained both numerical and continuous-type data.

3) How logs are managed?

We have a single log file for the entire project. However, we can create different log files for each stage in the project cycle, if needed.

Conu…

4) What techniques were you using for data pre-processing?

- Removing unwanted attributes
- Visualizing the relation of independent variables with each other and the output
  variables
- Checking and changing the Distribution of continuous values
- Cleaning data and imputing if null values are present.
- Scaling the data

5) How training was done or what models were used?

- After loading the dataset, data pre-processing was done.
- For this project, we opted to train the data using the Random Forest Classifier.
- Hyper-parameter tuning and new features were engineered during the various versions of modeling.
- The best model was selected.

Coun…

6) How Prediction was the done?

- The test files were provided.
- The test data also underwent preprocessing and new features required for the model were prepared.
- Then the data was passed through the model and the output was predicted.

7) What are the different stages of deployment?

- After training the model, we prepared all the necessary files required for deployment and uploaded them to a document version control system called Github.
- We then connected to and deployed the model in, Streamlit CLoud.

Thank You..!