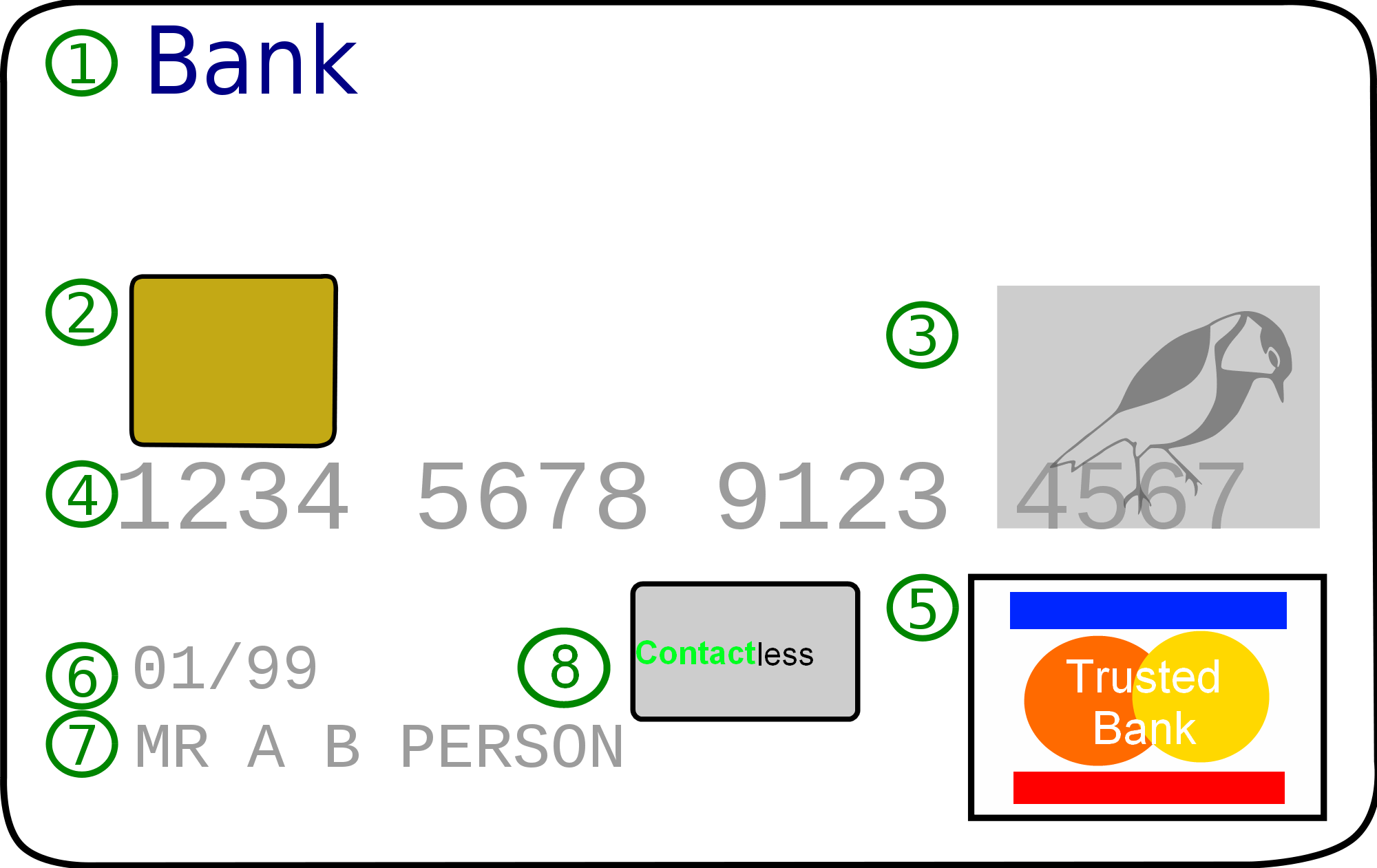
CREDIT CARD DEFAULT PREDICTION



**The Goal:**

Financial threats are displaying a trend in the credit risk of commercial banks. 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 the credit card owner’s characteristics and payment history.

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.

https://www.kaggle.com/datasets/uciml/defaul t-of-credit-card-clients-dataset

**ARCHITECTURE  :**

****

**Data Information**

|  |  |  |  |
| --- | --- | --- | --- |
| **NAME**  **LIMIT\_BAL** | **DATA TYPE**  QUANTITATIVE | **MEASUREMENT** NT DOLLAR | **DESCRIPTION**  INPUT |
| **SEX** | CATEGORIC | INT | INPUT |
| **EDUCATION** | CATEGORIC | INT | INPUT |
| **MARRIAGE** | CATEGORIC | INT | INPUT |
| **AGE** | QUANTITATIVE | YEARS | INPUT |
| **PAY\_0-6** | CATEGORIC | INT | INPUT |
| **BILL\_AMT-0-6** | QUANTITATIVE | NT DOLLAR | INPUT |
| **PAY\_AMT-0-6** | QUANTITATIVE | NT DOLLAR | INPUT |
| **default.payment .next.month** | CATEGORIC | BINARY | OUTPUT |

|  |  |
| --- | --- |
| **NUMBER OF ROWS** | **NUMBER OF COLUMNS**  25 |

**Pipeline :**

1. Data Ingestion

* data in data ingestion folder
* Split data in train and test data
* Saving data in data ingestion folder

1. Data validation

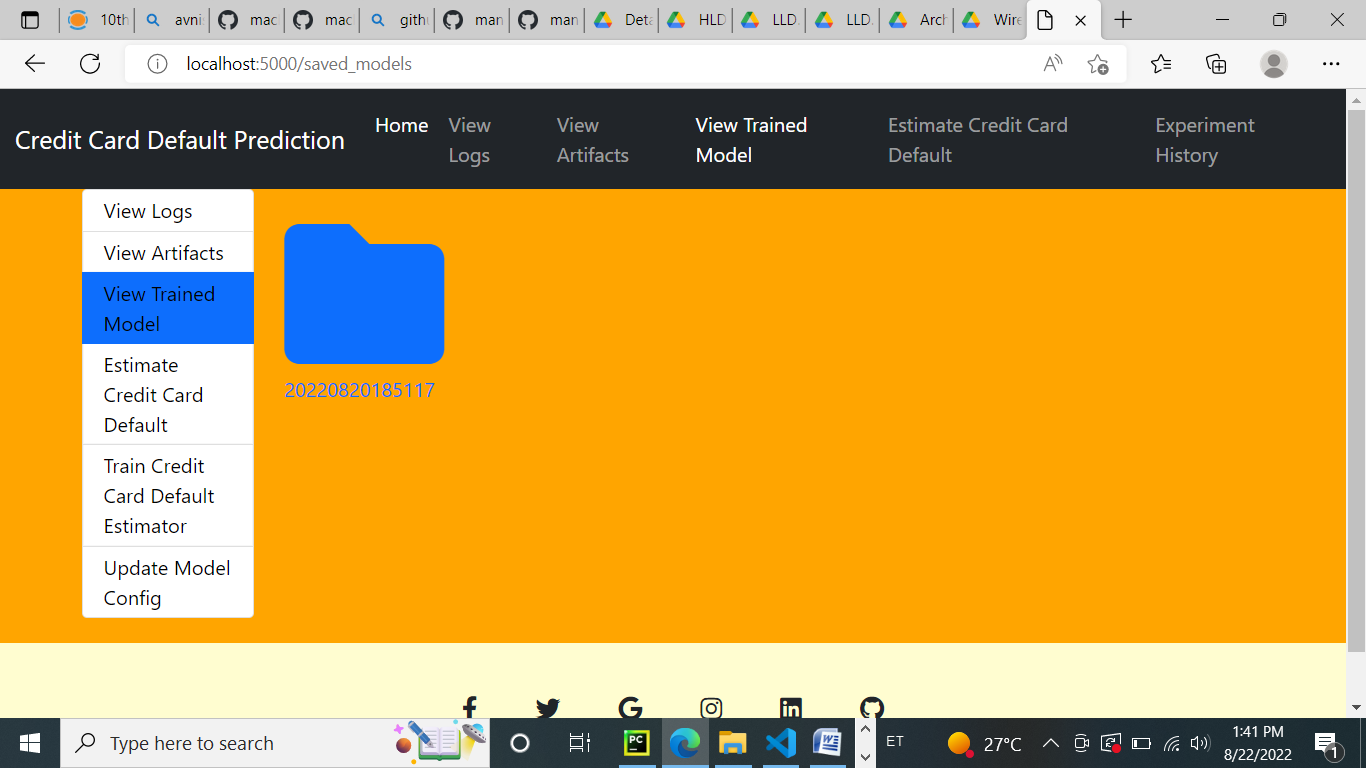
* data in data validation folder
* data drift and EDA
* Report in evidently to show and analysis Data Distribution
* Saving report in data validation Folder

1. Data Transformation

* data in data Transformation folder
* Using Pipeline and columns transformer to processing Data
* Splitting data in array by Standard Scalar and different technique
* Putting train array data and test array data in transformation folder and creating prepossessing Object file.
* Saving Processing pkl in data transformation folder for future Transformation

1. Model Trainer

* We have built various models like Logistic Regression, Random Forest, Gradient Boosting, etc. • Each of the above models was built taking their default parameters.
* We have use Gridsearch CV to have best model in comparison all models
* We save best model file in Model Trainer folder



1. Data Evaluation

* we used Metric for Classification : Recall Score, F1 Score and Model accuracy In train and test model to evaluate our best model
* we save the evaluation model in evaluation Folder
* Random Forest Classifier was the best Model
* Reason For Choosing This Model :
* Apart From a good training and test score, the reasons for  choosing this model are as follows :
* • Can handle missing values.
* • Can work well even on imbalanced data.

1. Model Pusher

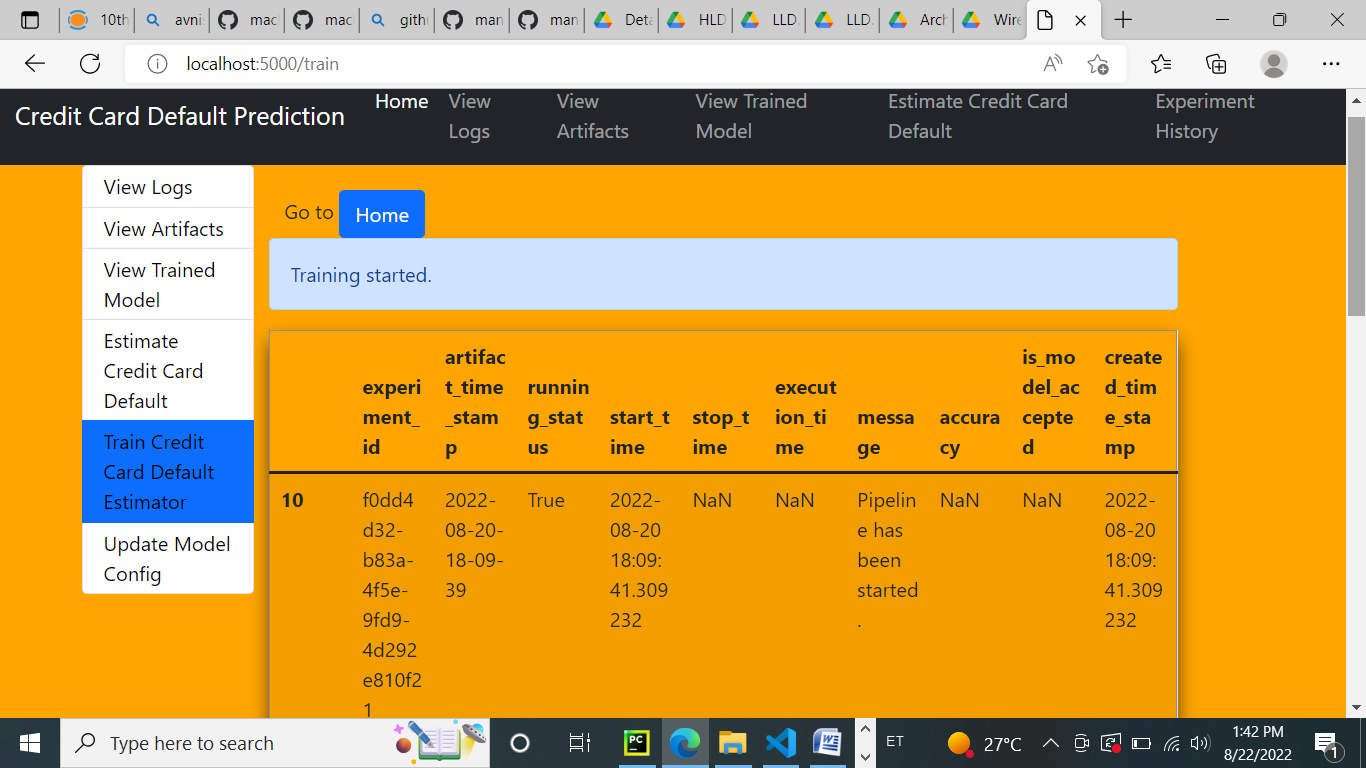
If we have an updated better model by new data ,

we can save it in model pusher save models

7 experiment :

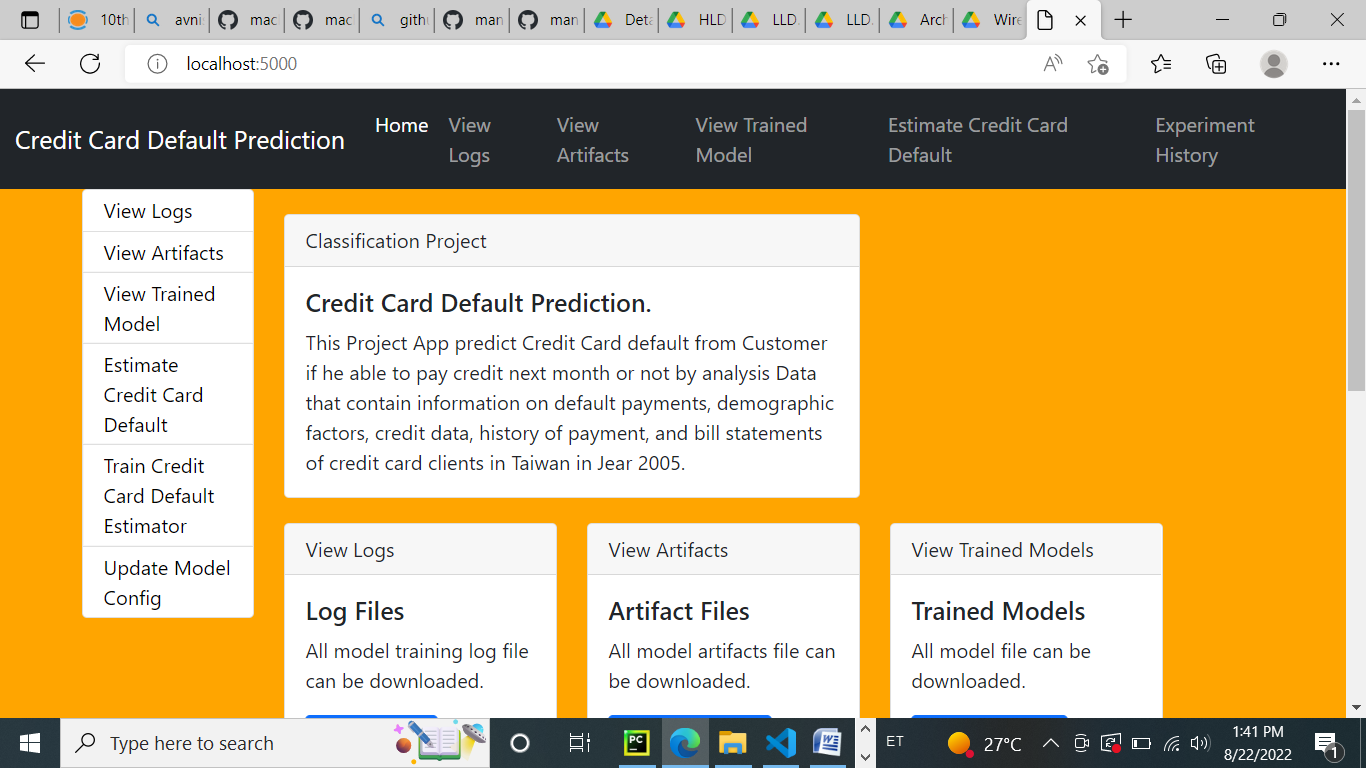
We create a experiment function to run a model in time limit that you

want and can display the score



8-Heroku Deplyement

We have create a app and we have deploy it in heroku platform cloud



https:// credit-prediction.herokuapp.com/

note : the heroku deployment doesen work in size of project more than 500 MB in free version