https://lh4.googleusercontent.com/nbLiGCv2aaBycn0plTAHIlWExAELP58huYs-r-tn1BM-FZUQPbR0_cr5ekE90n5scNUECHFa2BloPpFkFClIDcM4Lzyy_QIk_4azoEAg3OOUuwTbyLWUEDhA8CewkpMN-cwlMq6vVVmUXW-uafewS6E

HIGH LEVEL DOCUMENT

**CREDIT CARD DEFAULT PROBABILITY PREDICTION**

August 21, 2022

***Written by othmane zoubairi***

CONTENTS

**DOCUMENT VERSION CONTROL**

3 **ABSTRACT** 4 **1) INTRODUCTION** 5 **1.1 WHY THIS HIGH-LEVEL DESIGN DOCUMENT?** 5 **1.2 SCOPE** 5 **1.3 DEFINITIONS** 5 **2) GENERAL DESCRIPTION** 6 **2.1 PRODUCT PERSPECTIVE** 6 **2.2 PROBLEM STATEMENT** 6 **2.3 PROPOSED SOLUTION** 6 **2.4 TECHNICAL REQUIREMENTS** 6 **2.5 DATA REQUIREMENTS** 7 **2.6 TOOLS AND TECHNOLOGIES USED** 8 **2.7 CONSTRAINTS** 8 **3)** 9 **3.1 PROCESS FLOW** 9 **3.2 DEPLOYMENT PROCESS** 9 **3.3 EVENT LOG** 10 **3.4 PERFORMANCE** 10

**3.5 REUSABILITY** 10 **3.7 RESOURCE UTILIZATION**

11 **3.9 DEPLOYMENT**

11 **3.1O USER INTERFACE** 12

**4 Conclusion**

2 | P a g e

**DOCUMENT VERSION CONTROL**

|  |  |  |  |
| --- | --- | --- | --- |
| Date  Issued | Version | Description | Author |
| 18/07/2022 | 1.0.1 | Initial HLD – V1.0.1 | Othmane  zoubairi |
| 20/08/2022 | 1.0.2 | Updated HLD  – V1.0.2 | Othmane  zoubairi |
| 21/08/2022 | 1.0.3 | Updated HLD  – V1.0.3 | Othmane  zoubairi |
|  |  |  |  |

3 | P a g e

ABSTRACT

**Financial threats are displaying a trend in the credit risk of  commercial banks as the incredible improvement in the  financial industry has arisen. In this way, one of the biggest  threats faces 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.**

**With the help of Data Science and Machine learning  technology, I developed an application, which allows a banker to determine the probability Of Default in just a  few seconds.**

4 | P a g e

  1) Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-level document (HLD) is to describe  the design of the project in detail which can be used as a  reference manual.

The HLD will:

• Present all the design aspects and define them in detail. • Describe the user interface being implemented. • Describe the software interfaces.

• Describe the performance requirements.

• Include design features and the architecture of the project.

1.2 Scope

The HLD document presents the entire structure of the project  in parts, such as the data ingestion, data pre-processing,  solution development, and the deployment part along with their  respective architectures. This uses non-technical to mild  technical terms which should be understandable to the

administrators of the system.

1.3 Definitions

|  |  |
| --- | --- |
| Term | Description |
| EDA | Exploratory Data Analysis |
| IDE | Integrated Development Environment |

5 | P a g e

2) General Description

2.1 Product Perspective

The Credit Card Default Probability predictor is a machine  learning based on the Random Forest model which will help us  to predict the probability Of Default based on the attributes of  the customer.

2.2 Problem Statement

Financial threats are displaying a trend in the credit risk of  commercial banks as the incredible improvement in the  financial industry has arisen. In this way, one of the biggest  threats faces 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.

2.3 Proposed Solution

The solution proposed here is a web application, which takes  the details of the customer and those details will be taken by a  machine learning model in the backend, which will then predict the probability of default and display it on the front-end page of  the user.

2.4 Technical Requirements

I used python version 3.7 with some important libraries to  develop a machine learning model, which accurately predicts  the probability of credit card default

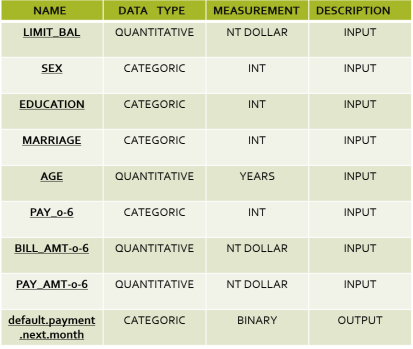
Then, the model is used as a back-end software for a front-end  web application which can be used by the users.

6 | P a g e 

 2.5 Data Requirements

For training and testing the model, I used the public data set  available in Kaggle, “Default of Credit Card Clients Dataset” by  UCI

URL - https://www.kaggle.com/datasets/uciml/default-of credit-card-clients-dataset



7 | P a g e 

2.6 Tools And Technologies Used



• Jupyter notebook is used for EDA and experimentation  with various ML algorithms with the help of pandas,  numpy, matplotlib, seaborn, sklearn   libraries.

• Jupyter was also used for the development and  deployment of the solution with logging. Used python  version 3.7 and libraries include logging, pandas, numpy,  scikit learn, flask, and HTML

* Github is used as a version control system. ‘
* Using Git to have acces to github to make uptade evry time you needed
* Deployed on the web using Gunicorn and Heroku.

2.7 Constraints

The Concrete Compressive Strength Prediction system must be  user-friendly, errors free and users should not be required to  know about any of the workings.

8 | P a g e

  3 Event Log

In this project, I used the “logging” library in both the  development and deployment stages, which keeps logging the  events at every step into the “.log” files. One of the advantages  of event logging is, it makes debugging much easier, we can  directly go to that specific line of code, which has errors.

3.1 Performance

The ML-based Probability Of Default Predictor application is  used for predicting the Probability Of Default based on various  attributes of the customer. So, it should be as accurate as  possible, so that it will not mislead the bank authorities.

Model retraining is very important to keep it relevant in order  to keep the model dynamic to changing times and customer  behaviour.

3.2 Reusability

The code written and the components used have the ability to  be reused without any problem.

3.3 Application Compatibility

The different components or modules of this project use python  version 3.7 as their interface between them. Each component  has its own task to perform,

and it is the job of the python version to ensure proper transfer  of the information.

3.4 Resource utilization

In this project, any task may use all the processing power  available in the system, until it is accomplished.

3.5 Deployment

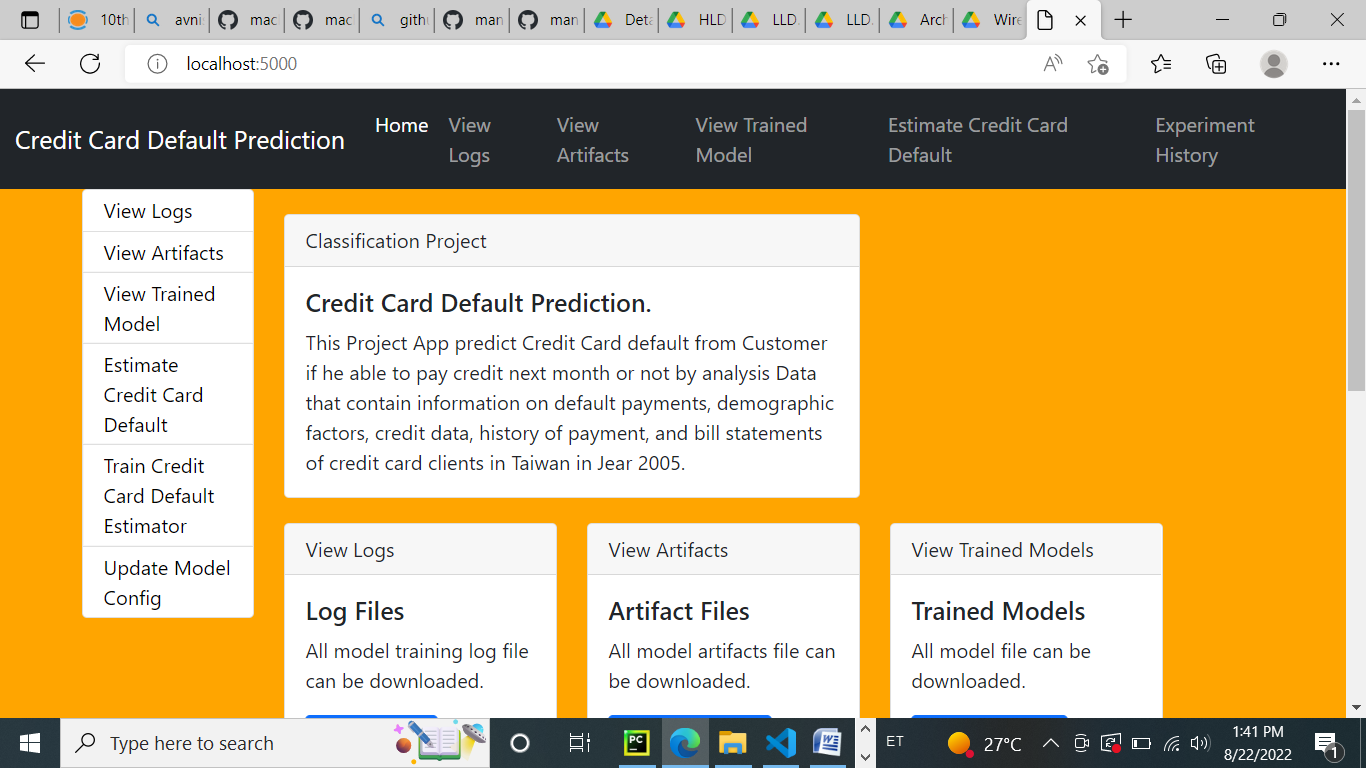
I deployed the application on the web using Heroku

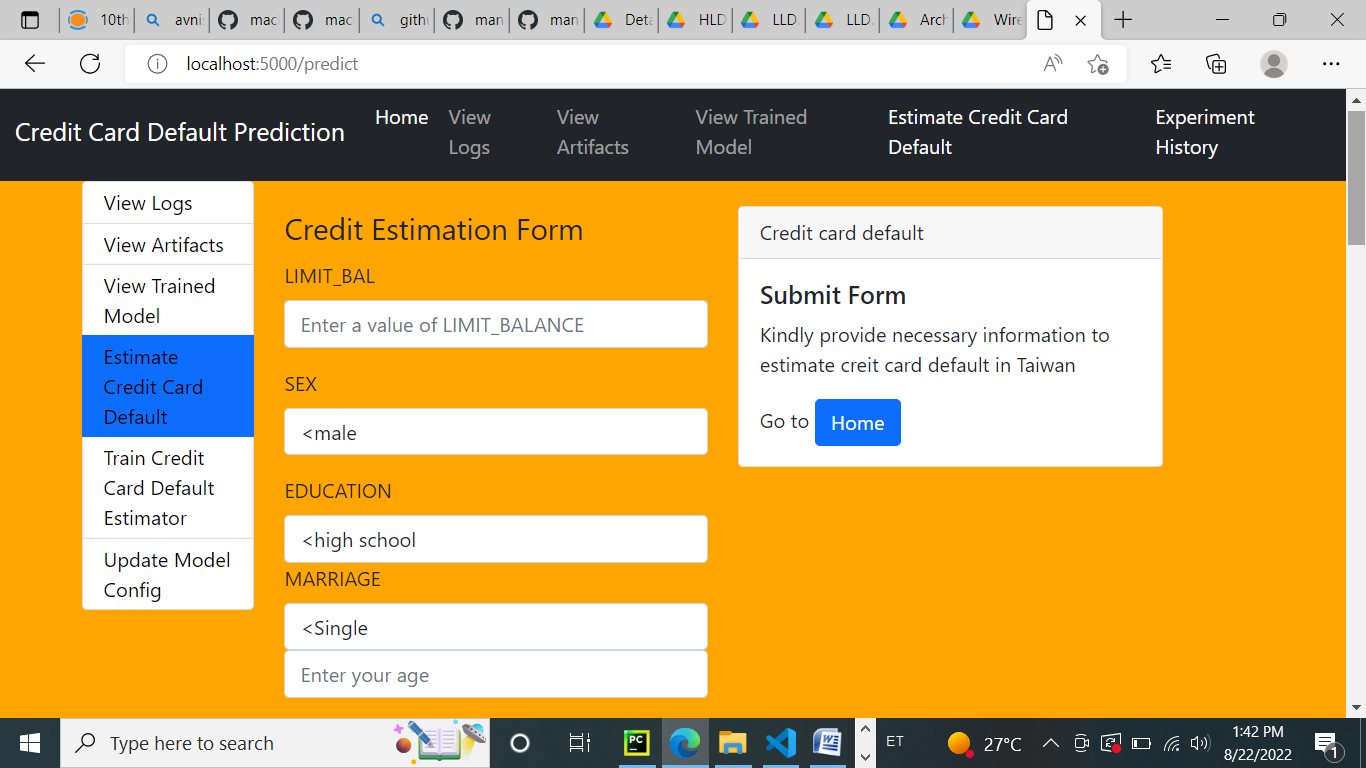
\*URL - <https://credit-prediction-herokuapp.com/>

\*Deployemnt dosen work if the size data more than 500 MB in free version

from heroku

3.6 User Interface





4 Conclusion

Credit Card Default Probability Predictor is used to predict the  probability of default given various attributes of the customer  using with the help of ML and Data Science techniques in order  to reduce bad debts of the bank.