iNeuron

HIGH LEVEL DOCUMENT

CREDIT CARD DEFAULT PROBABILITY PREDICTION

October, 2022

Written by Namdeo Patil

iNeuron

CONTENTS

DOCUMENT VERSION CONTROL

Date Issued	Version	Description	Author
18/10/2022	1.0.1	Initial HLD – V1.0.1	Namdeo Patil
20/10/2022	1.0.2	Updated HLD – V1.0.2	Namdeo Patil
21/10/2022	1.0.3	Updated HLD – V1.0.3	Namdeo Patil

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.

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

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.

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

NAME	DATA TYPE	MEASUREMENT	DESCRIPTION
LIMIT_BAL	QUANTITATIVE	NT DOLLAR	INPUT
SEX	CATEGORIC	INT	INPUT
EDUCATION	CATEGORIC	INT	INPUT
MARRIAGE	CATEGORIC	INT	INPUT
AGE	QUANTITATIVE	YEARS	INPUT
PAY_o-6	CATEGORIC	INT	INPUT
BILL_AMT-o-6	QUANTITATIVE	NT DOLLAR	INPUT
PAY_AMT-o-6	QUANTITATIVE	NT DOLLAR	INPUT
default.payment .next.month	CATEGORIC	BINARY	OUTPUT

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.

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.

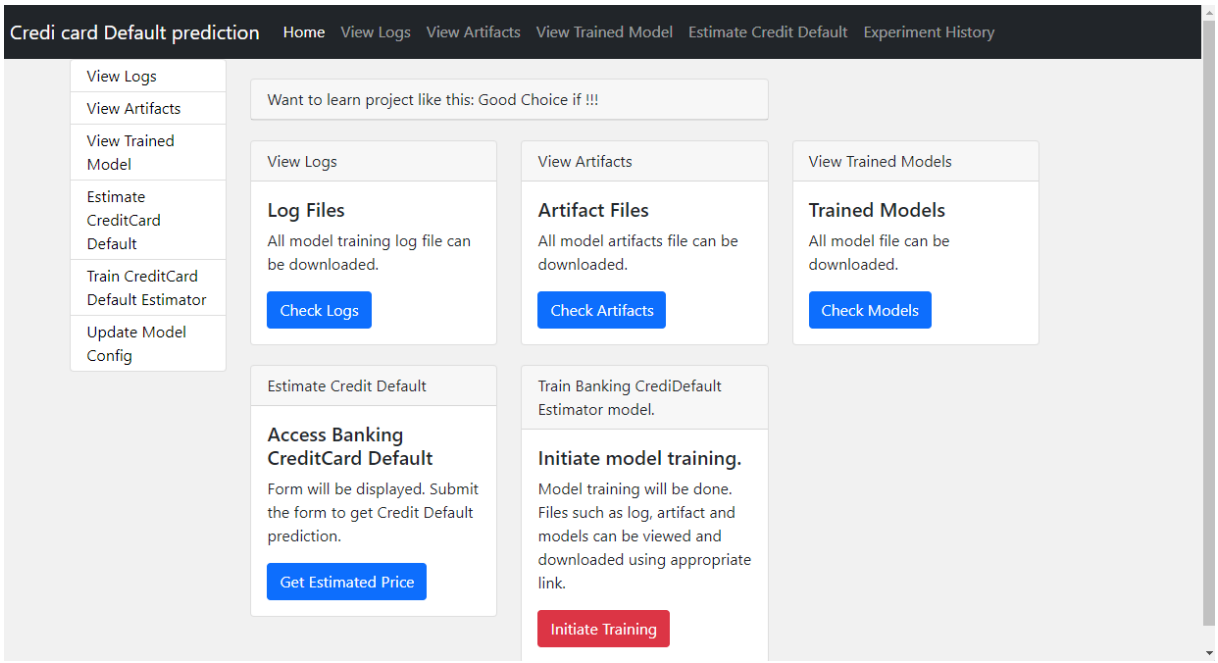
3.5 Deployment

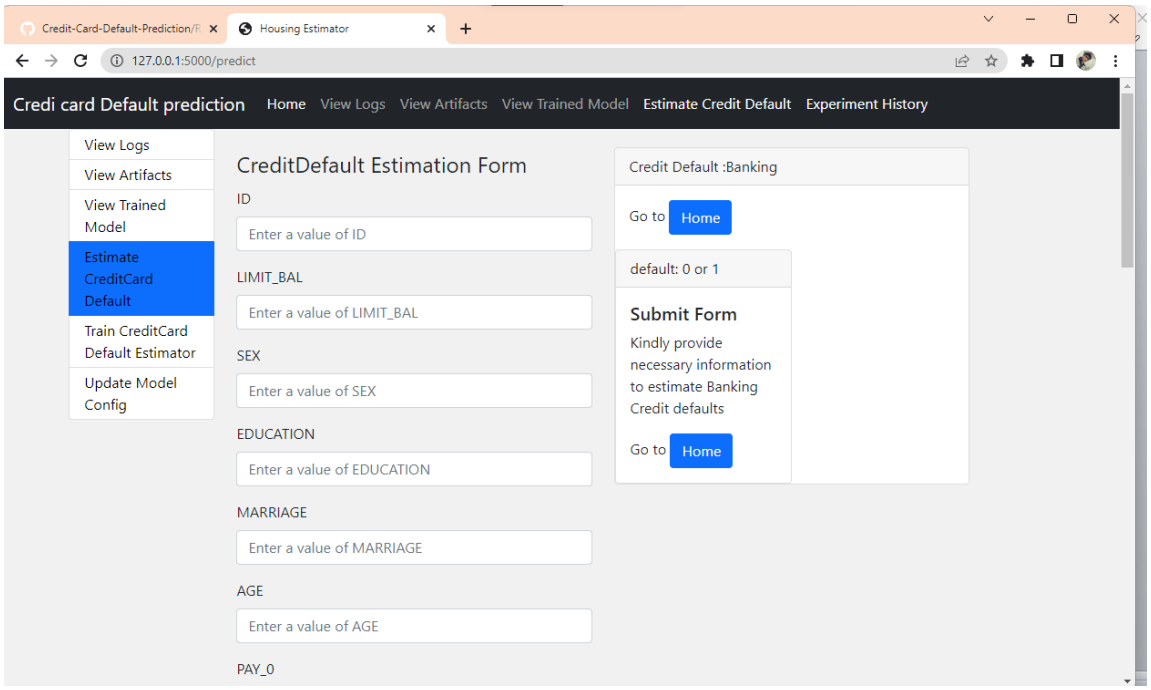
I deployed the application on the web using Heroku

*URL - https://mlp-creditcard123.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.