

##### Credit Card Defaulter Prediction App

Project Architecture

Domain: Machine Learning

Creator: Sudhanshu Rastogi

Date: 04.09.2023

Github:

https://github.com/sudhanshu2198

**Architecture**

**Data Preparation**

**Model Development**

**Deployment**

**Deployment Process**

**Architecture Description**

**Data Preparation**

The features variables include education, sex , age, credit limit, bill amount statement,

payment amount statement, payment status for 6 months for 30,000 instances. The data

is taken from Kaggle Platform. It can be used for developing a predictive model that

predicts probability of credit card customer defaulting on the due payments in the given

month given previous 6 month account history.

Data Preprocessing

In data preprocessing step, we check if there missing data, duplicate values, and

mixed datatypes presence in feature variables. In our dataset, there was no duplicate

values, missing values but datatype of multiple columns was “object” type, therefore they

was converted to proper datatype.

Data Cleaning

This step involves outlier removal such extremely high credit limit, negative values for

age, payment account statement. Datatype correction performed on features

variables containing integer variable but having object type as object.

Exploratory Data Analysis

This step includes data visualization of features and target variables. Visualization

utilizes pie charts, scatterplot, histograms, countplot, boxplots..

Feature Engineering

New features were created utilizing provided information such as Boolean payment

defaulter feature if the customer has defaulted ever in past 6 months, categorizing age

of the customer in uniform interval brackets. Including others option in sex, education

and marital status variable.

**Model Development**

Algorithm Spot Checking

Algorithm spot checking was performed on cleaned dataset by training the different models and evaluating their performance using cross validation having 5 folds. The different model are Random Forest Regressor, Xtreme Gradient Boosting Regressor, Light Gradient Boosting Regressor , Logistic Regression. Their area under curve were calculated. The highest score is achieved by Light Gradient Boosting Regressor.

Hyper-parameter Tuning

The best model is Light Gradient Boosting Regressor from algorithm spot checking, random search on no of estimators, learning rate, max depth and subsample is performed using optuna library to get the best hyper parameters for the model. Those parameters is then used on the model to get better result.

Model Evaluation

Test dataset is used to evaluate the model. 20% of dataset was separated for testing. Predicted results of the model are compared with the actual data to check the amount of error.

**Deployment**

Streamlit Web App

For this project, the trained is deployed for inference using an interactive web app. Interactive web app is created streamlit. Streamlit is a popular open-source framework used for model deployment by machine learning and data science teams. The web application contains three pages introduction, prediction and visualization.

Designing Web App

The web application contains three pages introduction, prediction, experiments results and visualization. Prediction pages provides inference services for user defined features variables. Dynamic visualization offers valuable insights about various factors affecting the restaurant establishment at different locations.

**Deployment Process**

After creating the streamlit app and testing it on local machine, code is pushed to github and linked to streamlit cloud. Specifying the branch to be utilized for creating web app, we will in creating an app hosted on streamlit cloud.