Credit Card Default Prediction

High Level Design (HLD)

Contents

Document Version Control

Version	Date	Author	Comments
1.0	22-Feb-24	Bhavika Pathak	Introduction & Architecture defined

Abstract

This High-Level Design (HLD) outlines the architectural blueprint and strategic approach for developing a machine learning-based credit card default prediction system. In the contemporary financial landscape, the ability to accurately forecast credit card defaults is indispensable for risk management and maintaining a healthy credit portfolio.

The HLD encapsulates the various stages of the project lifecycle, commencing with the acquisition of diverse datasets encompassing historical credit card transactions, customer demographics, economic indicators, and potentially supplementary sources of data. Rigorous preprocessing techniques are applied to cleanse, engineer features, and standardize the datasets, priming them for subsequent modeling endeavors.

Employing a supervised learning paradigm, the project explores an array of algorithms, including logistic regression, decision trees, random forests, gradient boosting, and neural networks, to construct predictive models. Algorithm selection is contingent upon factors such as performance metrics, interpretability, and computational efficiency.

Validation and evaluation protocols are paramount to ascertain the efficacy and generalizability of the models. Techniques such as cross-validation, hyperparameter tuning, and comprehensive performance metrics evaluation are deployed to assess model performance rigorously.

Following model development, extensive testing procedures are undertaken to validate the robustness and reliability of the predictive models across diverse real-world scenarios. This phase encompasses simulating various conditions and evaluating the model's performance under different circumstances.

The deployment phase entails seamless integration of the trained model into the existing credit risk management infrastructure. The model may be deployed either as a standalone application or as an integral component within a broader decision-making framework. Continuous monitoring and iterative updates are pivotal to adapt to evolving market dynamics and sustain model efficacy over time.

High Level Design

1. Introduction

1.1. Why High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the important details about this project. Through this HLD Document, I’m going to describe every small and big thing about this project.

2. General Description

2.1. Product Perspective

The purpose of credit card default prediction project is to develop predictive models that accurately identify customers at risk of defaulting on their credit card payments.

2.2 Problem statement

Financial threats are displaying a trend about 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 credit card owner's characteristics and payment history.

2.3 Proposed Solution

In the proposed solution, we used Random Forest Classifier machine learning model to classify the credit default based on credit card owner's characteristics and payment history. Here, first we are performing Data preprocessing step, in which data transformation, handling missing values, feature transformation, feature selection, steps are performed and then we are going to build the machine learning model and will be deployed it on cloud platform.

2.4 Technical Requirements

Following are the requirements of this project:

- Model should be deployed on cloud (Azure, AWS, GCP, Heroku,Netlify).
- Cassandra database should be integrated in this project for any kind of user input.

2.5 Data Requirements

Data Requirement completely depend on our problem.

High Level Design

• For training and testing the model, we are using Credit Card dataset which is available Kaggle.

• From user we are taking following input

Feature Names:

- Gender
- Age
- Education
- Limit Balance

Etc.

2.6 Tools Used

2.7 Data Requirements

- PyCharm is used as IDE.
- For visualization of the plots, Matplotlib, Seaborn are used.
- Netlify is used for deployment of the model.
- MongoDB is used to retrieve, insert, delete, and update the database.
- Front end development is done using HTML, CSS, Bootstrap, Flask is used
- for backend development and for API development.

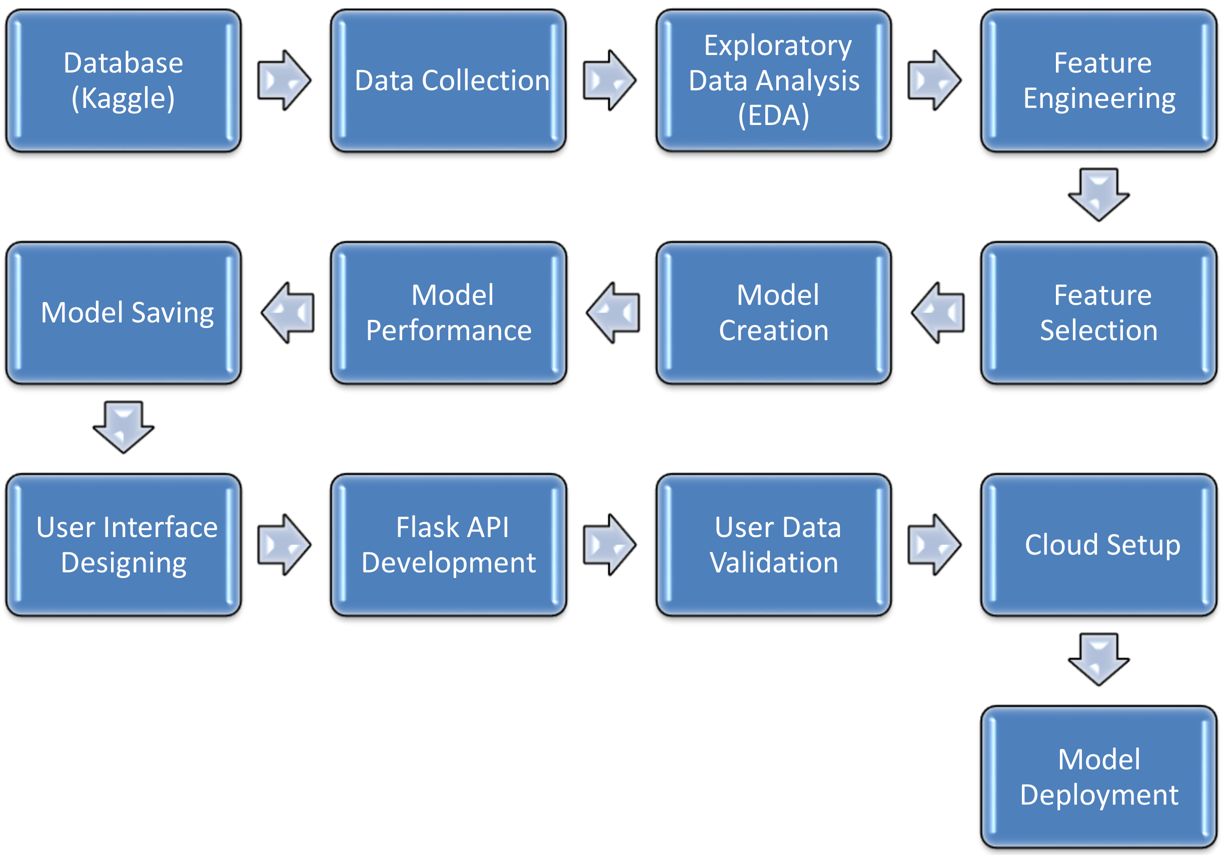
- GitHub is used as version control system

2.8 Constraints

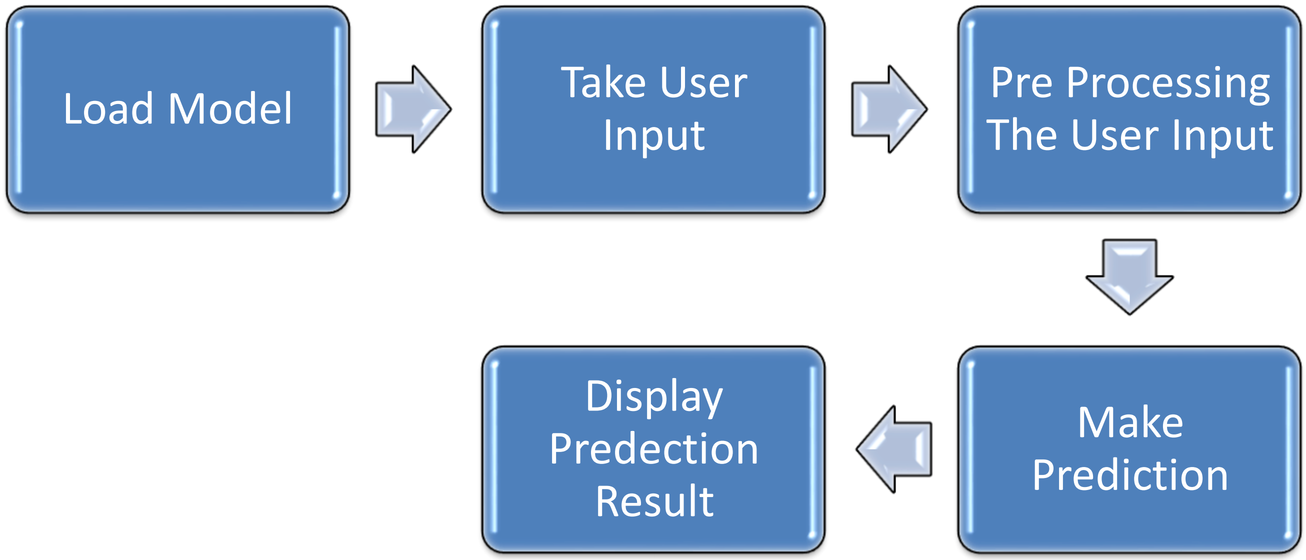
The Credit Card Default Prediction Model system must be user friendly, errors free and users should not be required to know any of the back-end working

3. Design Details

3.1 Process Flow



3.2 Deployment Process



4. Performance

A. Solution of Credit Card Default Prediction is used to predict the Credit Card Default Owner , and it should be as accurate as possible.

B. That's why before building this model we followed complete process of Machine Learning. Here is summary of complete process:

i. First, we cleaned our dataset properly by removing all null value and duplicate value present in dataset.

ii. After that we performed EDA and feature transformation.

iii. And then we performed feature selection process.

iv. Then we performed the encoding – numerical features and categorical features

v. And now, we split the dataset in train-test split.

vi. After performing above, we trained our dataset on different classification algorithm (Logistic, SVM, KNN, Decision Tree Classifier, Random Forest Classifier etc.). After training the dataset on different algorithms, we got highest accuracy of 86.65% on Random Forest Classifier.

vii. After that we saved our model in pickle file format.

viii. After that our model was ready to deploy, we deployed this model on Netlify.

C. Re-usability

High Level Design

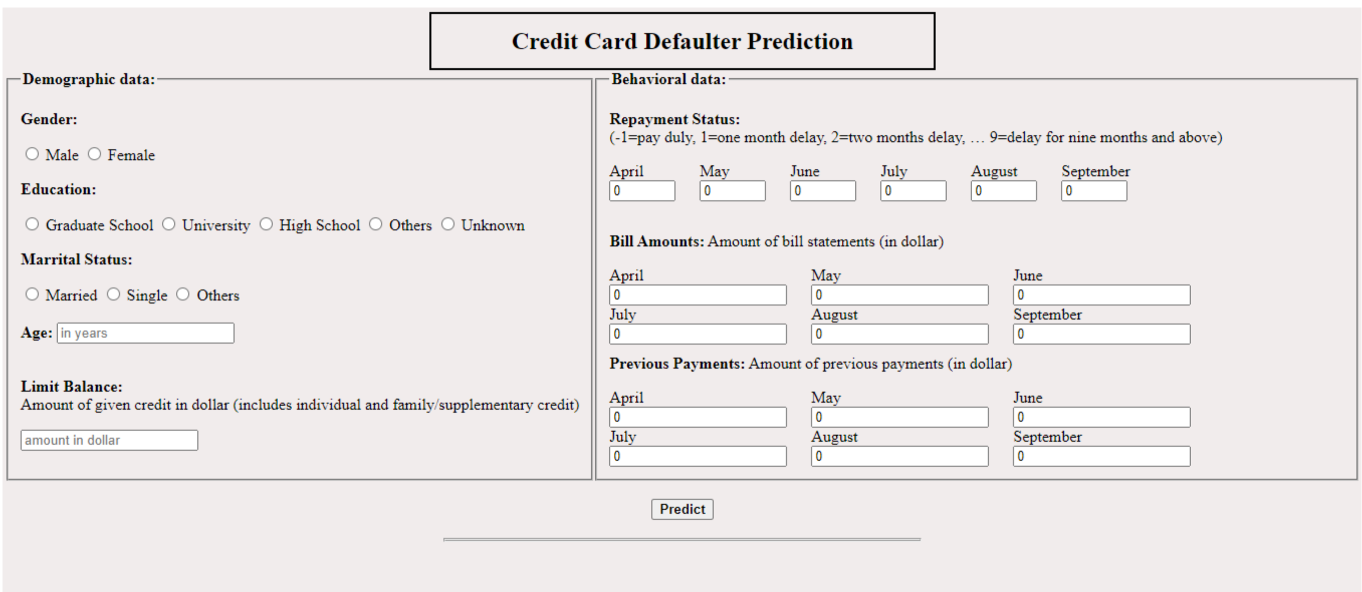
We have done programming of this project in such a way that it should be reusable. So that anyone can add and contribute without facing any problems

D. Application Compatibility

The different module of this project is using Python as an interface between them. Each module have it's own job to perform and it is the job of the Python to ensure the proper transfer of information.

E. Deployment: We have deployed this on Netlify cloud.

User Interface



5. Conclusion

The machine learning credit card default prediction project successfully developed robust predictive models to identify potential defaulters. These models enhance risk management, optimize portfolio health, reduce operational costs, improve the customer experience, and ensure regulatory compliance. Through proactive identification and intervention, the project contributes to financial stability and responsible lending practices within the credit card industry. Random Forest Classifier is used for classification.