iNeuron

CREDIT CAR DEFAULT PREDICTION

High Level Document (HLD)

Revision Number : 2.0

Last Date of Revision : 04-06-2024

Document Version Control

Date Issued	Version	Description	Author
2024-05-10	1.0	Initial draft of project plan	Sridhar
2024-05-13	1.1	Added data collection methods	Kohitha
2024-05-16	1.2	Revised data preprocessing steps	Kavya
2024-05-20	1.3	Included feature engineering techniques	Kavya
2024-05-23	1.4	Updated model selection criteria	Kohitha
2024-05-27	1.5	Added performance metrics and evaluation	Sridhar
2024-05-31	2.0	Finalized report and conclusion	Sridhar, Kavya, Kohitha

Introduction :

Credit card default prediction is a crucial task for financial institutions as it helps in assessing the credit risk of cardholders and making informed decisions about lending and credit limits. With the advancements in machine learning, it is now possible to predict the likelihood of a customer defaulting on their credit card payments with greater accuracy and efficiency. This project aims to develop a machine learning model to predict credit card defaults, thereby assisting financial institutions in managing risk and preventing potential losses.

Project Objectives :

The main objectives of this project are:

- Data Collection and Preprocessing : Gather and clean data related to credit card transactions and customer demographics.
- Feature Engineering : Identify and create relevant features that will be used to train the machine learning model.
- Model Selection and Training : Evaluate various machine learning algorithms to select the best-performing model for predicting defaults.
- Model Evaluation : Assess the performance of the selected model using appropriate metrics and validate its effectiveness.
- Deployment : Develop a strategy for deploying the model in a real-world setting, ensuring it can be used by financial institutions to make data-driven decisions.

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.

DataSet Information :

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.

Content

There are 25 variables :

- ID: ID of each client
- LIMIT_BAL: Amount of given credit in NT dollars (includes individual and family/supplementary credit
- SEX: Gender (1=male, 2=female)
- EDUCATION: (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
- MARRIAGE: Marital status (1=married, 2=single, 3=others)
- AGE: Age in years

PAY_0: Repayment status in September, 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months, … 8=payment delay for eight months, 9=payment delay for nine months and above)

- PAY_2: Repayment status in August, 2005 (scale same as above)
- PAY_3: Repayment status in July, 2005 (scale same as above)
- PAY_4: Repayment status in June, 2005 (scale same as above)
- PAY_5: Repayment status in May, 2005 (scale same as above)
- PAY_6: Repayment status in April, 2005 (scale same as above)
- BILL_AMT1: Amount of bill statement in September, 2005 (NT dollar)
- BILL_AMT2: Amount of bill statement in August, 2005 (NT dollar)
- BILL_AMT3: Amount of bill statement in July, 2005 (NT dollar)
- BILL_AMT4: Amount of bill statement in June, 2005 (NT dollar)
- BILL_AMT5: Amount of bill statement in May, 2005 (NT dollar)
- BILL_AMT6: Amount of bill statement in April, 2005 (NT dollar)

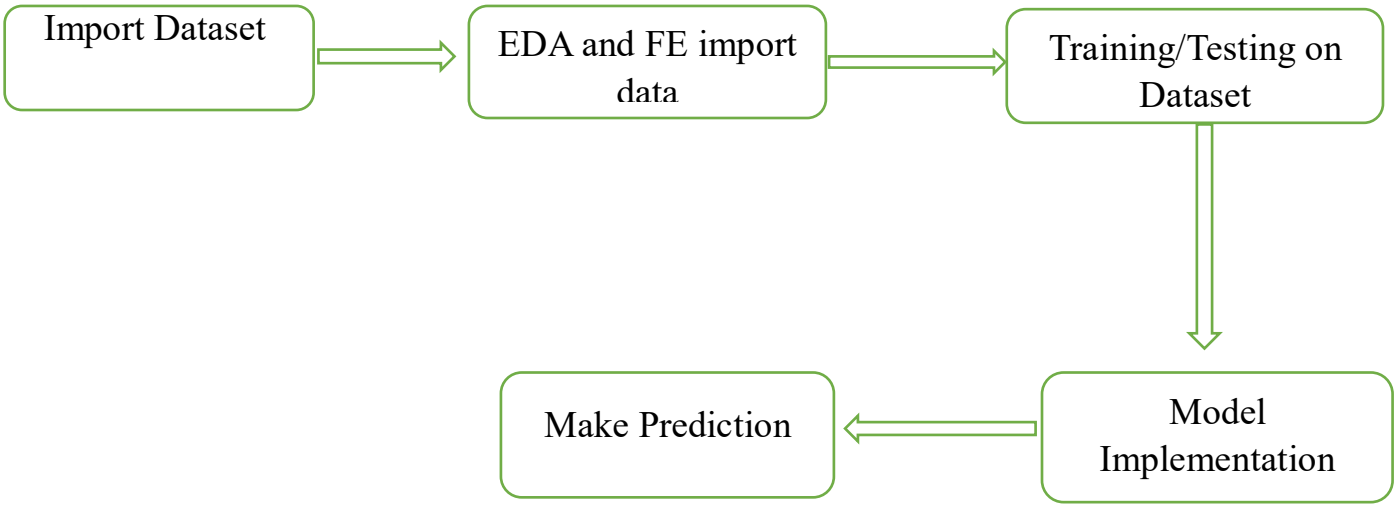
- PAY_AMT1: Amount of previous payment in September, 2005 (NT dollar)
- PAY_AMT2: Amount of previous payment in August, 2005 (NT dollar)
- PAY_AMT3: Amount of previous payment in July, 2005 (NT dollar)
- PAY_AMT4: Amount of previous payment in June, 2005 (NT dollar)
- PAY_AMT5: Amount of previous payment in May, 2005 (NT dollar)
- PAY_AMT6: Amount of previous payment in April, 2005 (NT dollar)
- default.payment.next.month: Default payment (1=yes, 0=no)

Tools Used :

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn and Flask are used to build the whole model.

Design Details:

Methodology:



Conclusion:

The project is designed in flask; hence it is accessible to everyone. The above designing process will help banks and loan lenders predict whether customers will default on credit card payments or not, so the bank or respective departments can take necessary action, based on the model's predictions. The UI is made to be user-friendly so that the user will not need much knowledge of any tools but will just need the information for results.