

Credit Card Default Prediction

Report

Project Overview

This report provides a comprehensive overview of the Credit Card Default Prediction project. The objective of this project is to build a machine learning model to predict whether a credit card client in Taiwan will default on their payment next month based on various features.

Dataset Information

The dataset used in this project 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.

Data Features

The dataset consists of 25 variables, including:

1. ID: ID of each client
2. LIMIT_BAL: Amount of given credit in NT dollars (includes individual and family/supplementary credit)
3. SEX: Gender (1=male, 2=female)
4. EDUCATION: Education level (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)
5. MARRIAGE: Marital status (1=married, 2=single, 3=others)
6. AGE: Age in years
7. PAY_0 to PAY_6: Repayment status for the last six months
8. BILL_AMT1 to BILL_AMT6: Amount of bill statement for the last six months (NT dollar)
9. PAY_AMT1 to PAY_AMT6: Amount of previous payment for the last six months (NT dollar)
10. default.payment.next.month: Default payment (1=yes, 0=no)

Data Preprocessing

1. Data Cleaning: Check for missing values and handle them appropriately. Ensure consistency in data types.
2. Data Exploration: Perform exploratory data analysis (EDA) to understand data distributions and relationships between features.
3. Feature Engineering: Create new features and encode categorical variables (e.g., one-hot encoding for education and marriage).
4. Data Split: Split the dataset into training and testing sets for model evaluation.

Model Building

1. Model Selection: Choose an appropriate machine learning algorithm for binary classification. Common choices include Logistic Regression, Random Forest, or Gradient Boosting.
2. Feature Scaling: Scale the numeric features to ensure they have the same impact on the model.
3. Model Training: Train the selected model on the training data.
4. Model Evaluation: Evaluate the model's performance using appropriate metrics such as accuracy, precision, recall, F1-score, and ROC-AUC on the testing data.

Model Deployment

1. Model Serialization: Serialize the trained model into a pickle file (`LogR_Model.pkl`) for future use.
2. Streamlit Web App: Create a Streamlit web app that loads the model and provides a user interface for predicting credit card default.

Streamlit Web App

1. User Interface: Create a user-friendly interface with input fields for users to enter their information.
2. Model Loading: Load the trained model (`LogR_Model.pkl`) within the Streamlit app.
3. User Input Processing: Accept user inputs, preprocess them, and feed them into the model for prediction.
4. Prediction: Display the prediction result (default or non-default) to the user.
5. Feedback: Allow users to interact with the app, provide feedback, and make predictions based on their inputs.

Conclusion

The Credit Card Default Prediction project involved data preprocessing, model building, and deployment of a machine learning model to predict credit card default. The Streamlit web app provides a user-friendly interface for users to make predictions based on their information.

The project aims to assist financial institutions in assessing the creditworthiness of clients, ultimately reducing the risk of default and improving financial decision-making.

Future Improvements

1. Explore different machine learning algorithms and ensemble methods to potentially improve prediction accuracy.
2. Gather more data to further enhance the model's performance.
3. Continuously update and refine the model to adapt to changing trends and client behaviors in credit card payments.