A Machine Learning Approach for Commercial Banks

Credit Card Default Prediction

Objective:

Predict Credit Default Risk:
- Develop a predictive model to accurately assess the probability of credit default for credit card owners based on their characteristics and payment history.

Implement Machine Learning Tasks:
- Perform standard machine learning tasks, including Data Exploration, Data Cleaning, Feature Engineering, Model Building, and Model Testing, to ensure a comprehensive and systematic approach.

Explore Data Characteristics:
- Analyze the dataset to understand the distribution of features, identify patterns, and gain insights into the relationship between credit default and various attributes.

Enhance Model Accuracy:
- Experiment with different machine learning algorithms to identify the most effective approach for predicting credit default probabilities.

Benefits:

Improved Credit Risk Management:
Enable commercial banks to proactively manage credit risk by predicting the likelihood of credit default, allowing for more informed decision-making in the credit approval process.

Financial Loss Mitigation:
Reduce the occurrence of defaults, leading to minimized financial losses for commercial banks and a more resilient financial ecosystem.

Optimized Decision-Making:
Empower banks with a reliable tool that optimizes decision-making by providing insights into the creditworthiness of clients, thereby streamlining the approval or rejection process.

Resource Allocation Efficiency:
Improve resource allocation by focusing attention on high-risk clients, allowing banks to allocate resources more efficiently for risk mitigation strategies.

Enhanced Regulatory Compliance:
Facilitate adherence to regulatory requirements by employing advanced predictive analytics, ensuring that banks operate within the stipulated guidelines for risk management.
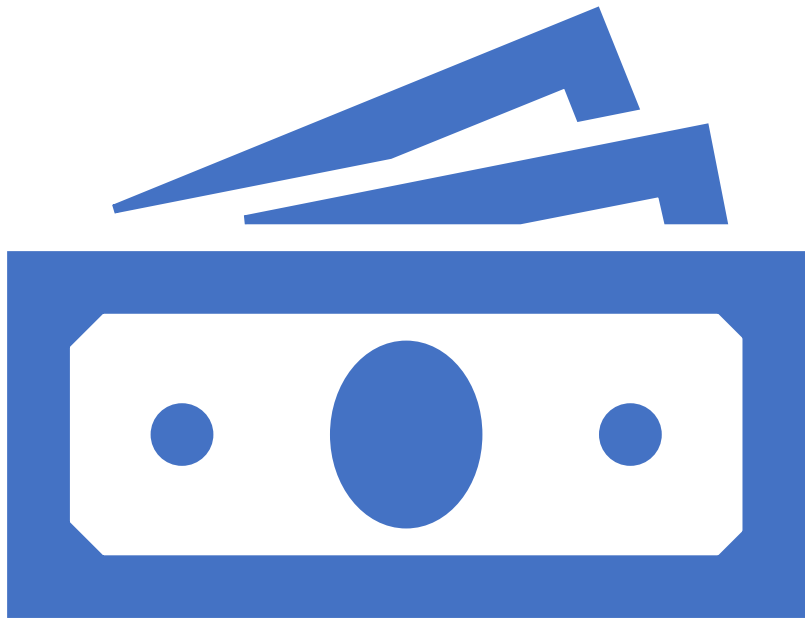
Problem Statement

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

Approach

The classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing. Try out different machine learning algorithms that's best fit for the above case.

Architecture
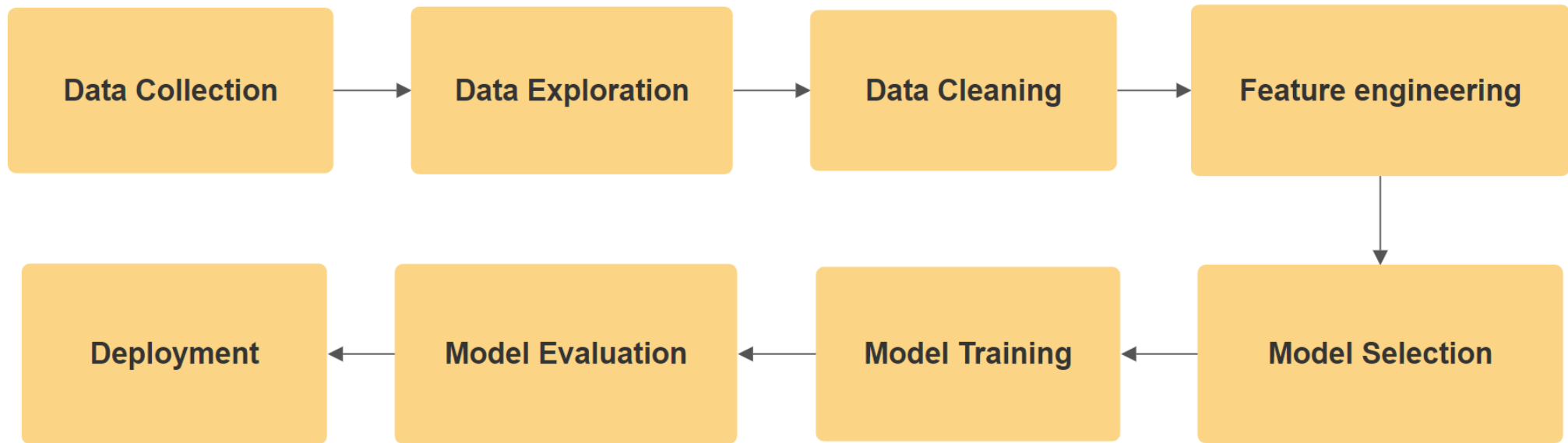
Data Collection and Data Exploration

- Obtained relevant data from commercial banks, ensuring it includes information on credit card owner characteristics and payment history. 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.

- Used descriptive statistics and visualizations to understand the distribution of key features. Identified potential correlations between features and the target variable (credit default).

Model Selection, Training and Evaluation

Considering logistic regression, decision trees, random forests, and support vector machines. Suitable machine learning algorithms for classification tasks are selected. Experiments with different algorithms are done to identify the most effective approach.


The dataset is split into a training set and test set for training the data to build the best model, after splitting the data the training data is trained with different classification algorithm to build the model for predicting the result.

Evaluation of different classification model is done through the test dataset, and accuracy score is defined the accuracy score of the logistic regression is more among all the other models therefore the logistic regression is selected for predicting result

Architecture of Model Training
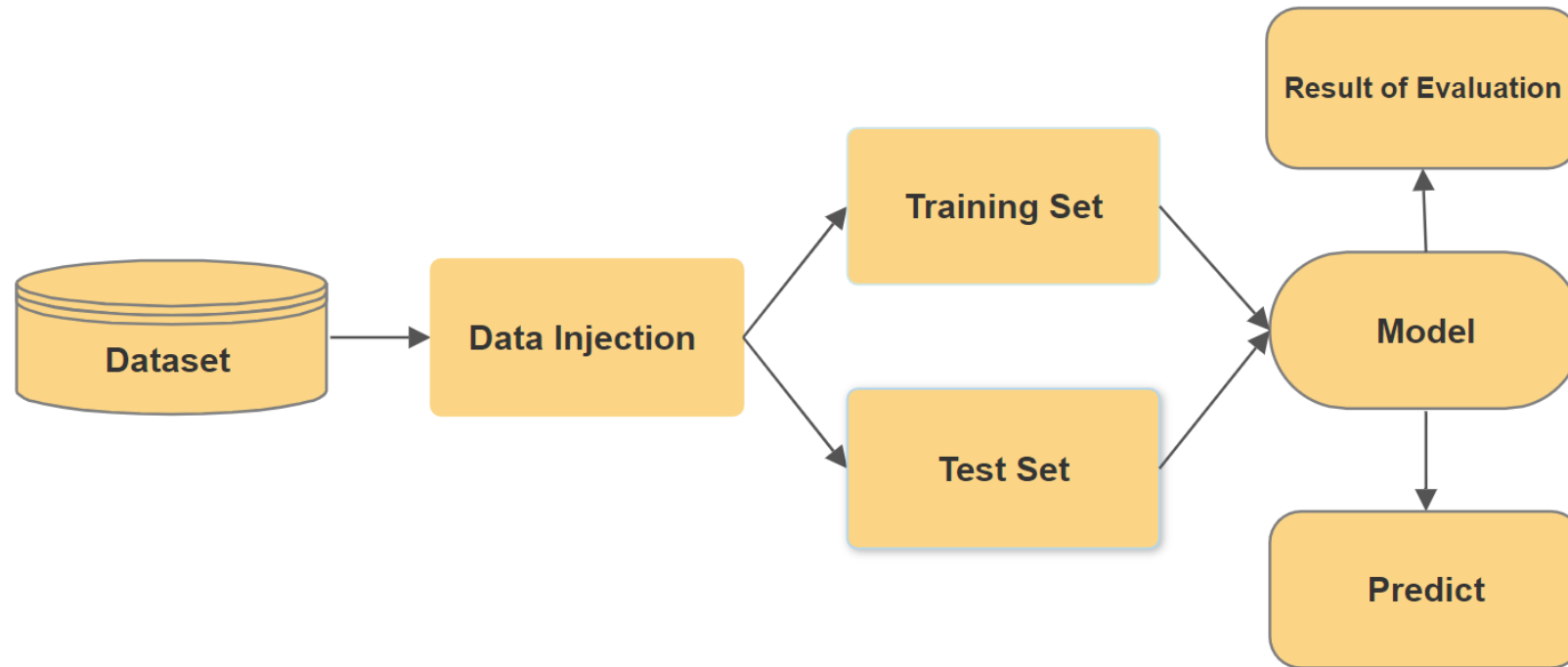
Q & A:

Q1) What's the source of data?

- The data for training is collected from the kaggle website.

Q 2) What was the type of data?

- The data has combination of numerical values.

Q 3) What's the complete flow you followed in this Project?

- Refer slide 4th for better Understanding

Q 4)Why is accurate credit default prediction important for commercial banks?

- Accurate credit default prediction is crucial for commercial banks to make informed decisions about credit approvals, reduce financial losses, and enhance overall credit risk management.

Q & A:

Q 5) How logs are managed?

We are using different logs as per the steps that we follow in insertion and modeling like File insertion log , Data transformation ,Model Training log , prediction log etc.

Q 6) What techniques were you using for data pre-processing?

▶ Removing unwanted attributes

▶ Visualizing relation of independent variables with each other and output variables

▶ Checking and changing Distribution of continuous values

▶ Removing outliers

▶ Cleaning data and imputing if null values are present.

▶ Converting categorical data into numeric values.

▶ Scaling the data

Q & A:

Q 4) How does the project contribute to financial stability?

▶ The project contributes to financial stability by providing tools for commercial banks to adopt prudent lending practices, minimizing defaults, and fostering a more resilient financial ecosystem.

Q 5) What benefits does the project offer to commercial banks?

▶ The project offers benefits such as improved credit risk management, financial loss mitigation, optimized decision-making, and a competitive advantage through the adoption of modern machine learning techniques.

Q 6) What machine learning algorithms are experimented with in the project?

▶ Various machine learning algorithms are experimented with, including logistic regression, decision trees, random forests, and support vector machines, to identify the most effective approach for credit default prediction.