

## **LOW LEVEL DESIGN (LLD)**

# **Credit Card Default Prediction**



**Last Date of Revision 22/08/2023**

**Nagalakshmi M**

## **Contents**

<b>1.Introduction.....</b>	<b>3</b>
<b>1.1 What is Low Level Design Document?.....</b>	<b>3</b>
<b>1.2 Scope.....</b>	<b>3</b>
<b>1.3 Project Introduction.....</b>	<b>3</b>
<b>2. Problem Statement.....</b>	<b>4</b>
<b>3. Dataset Information.....</b>	<b>5</b>
<b>4. Architecture.....</b>	<b>6</b>
<b>4.1 Architecture Description .....</b>	<b>6</b>

# **Introduction**

## **1.1 What is Low Level Design Document?**

The goal of the Low-level design document (LLDD) is to give the internal logic design of the actual program code for Credit Card Defaulter Predictor. LLDD describes the class diagrams with the methods and relations between dependent and independent classes. It describes the modules so that the programmer can directly code the program from the document.

## **1.2 What is Scope?**

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

## **1.3 Project Introduction**

This project is an attempt to identify credit card customers who are more likely to default in the coming month by using customer's personal and financial information like credit line, age, repayment and delinquency history for the past 6 months to predict the probability of the particular customer to become default....

## **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.

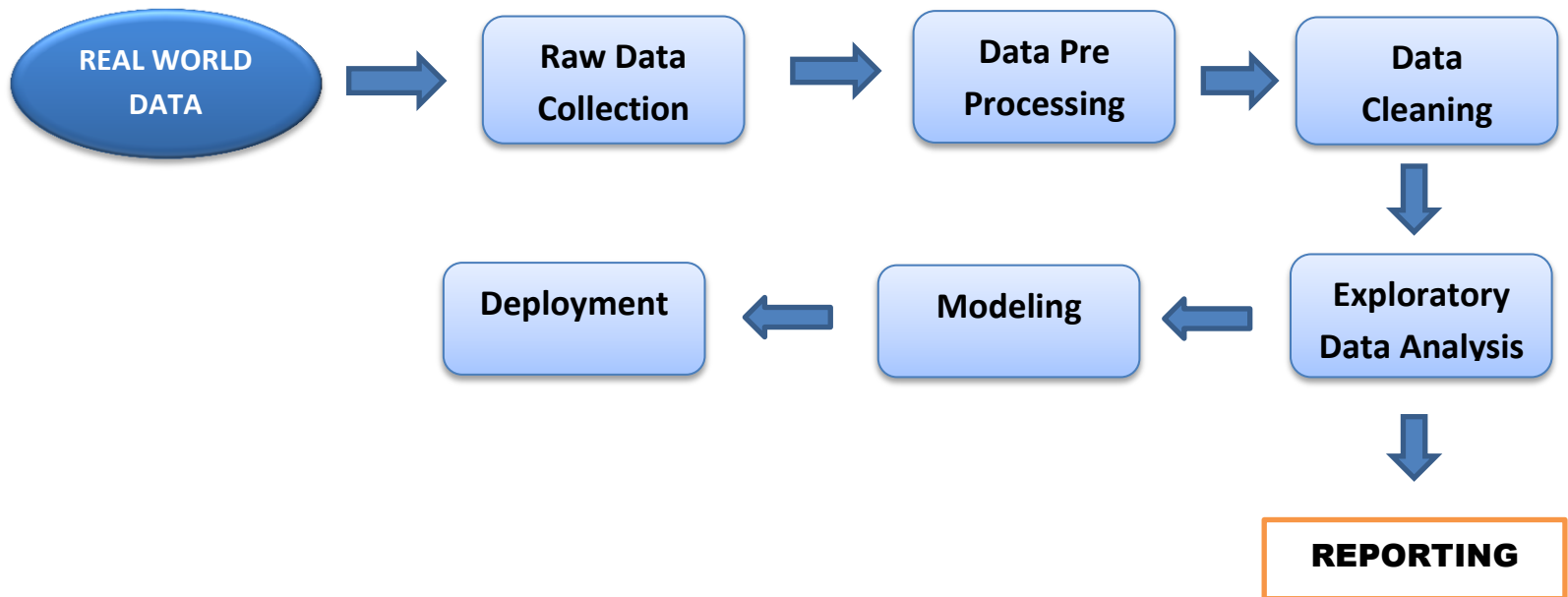
To achieve the goal, we used a data set that is formed by taking into consideration some of the information of 30000 credit card holder in Taiwan in which the dataset contains the transaction details 30000 card holders from April 2005 to September 2005. The problem is based on the given information about each individual we have to calculate that whether that individual with Defaulter or not a defaulter for next month.

### 3. Dataset Information

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)

## 4. Architecture



### 4.1 Architecture Description

#### 1. Raw Data Collection

The Dataset was taken from iNeuron Provided Project Description Document.

[https://drive.google.com/file/d/1AGRq2hG8zUbM\\_8LCo48cbcYy6W-2ujZC/view?pli=1](https://drive.google.com/file/d/1AGRq2hG8zUbM_8LCo48cbcYy6W-2ujZC/view?pli=1)

#### 2. Data Pre-Processing

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data feed to the model to train. This Process includes

- a) Handling Null/Missing Values
- b) Handling Skewed Data
- c) Outliers Detection and Removal

### **3. Data Cleaning**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

- a) Remove duplicate or irrelevant observations
- b) Filter unwanted outliers
- c) Renaming required attributes

### **4. Exploratory Data Analysis (EDA)**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

### **5. Reporting**

Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in easy and selfexplanatory report because your model will be used by many stakeholders who are not from technical background.

- a) High Level Design Document (HLD)
- b) Low Level Design Document (LLD)
- c) Architecture
- d) Wireframe

e) Detailed Project Report

f) Power Point Presentation

## 6. Modeling

Data Modeling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

## 7. Deployment

CREDIT CARD DEFAULT  
PREDICTION

Limited Balance (in New Taiwanese (NT) dollar)

Select Gender

male

Select Education

graduate school

Marital Status

Married

Age (in Years)

Repayment Status of last month

Repayment Status of last month

No delay in payment

Repayment Status of last 2nd month

No delay in payment

Repayment Status of last 3rd month

No delay in payment

Repayment Status of last 4th month

No delay in payment

Repayment Status of last 5th month

No delay in payment

Repayment Status of last 6th month

No delay in payment

Last month Bill Amount (in NT dollar)

Last 2nd month Bill Amount (in dollar)

Amount paid in Last Month (in NT dollar)

Amount paid in Last 2nd month (in NT dollar)

Amount paid in Last 3rd month (in NT dollar)

Amount paid in Last 4th month (in NT dollar)

Amount paid in Last 5th month (in NT dollar)

Amount paid in Last 6th month (in NT dollar)

Predict



