

# **Low-Level Design (LLD)**

## **Credit Card Default Prediction**

**Revision Number - 1.2**

**Last Date of Revision - 06/02/2024**

**Ahamed Ajas S**

**Afrin S**

Document Control

Date	Version	Description	Author
4/2/2024	1.0	Introduction, Problem Statement	Ahamed Ajas S Afrin S
6/2/2024	1.1	Dataset Information, Architecture Description	Ahamed Ajas S Afrin S
8/2/2024	1.2	Final Revision	Ahamed Ajas S Afrin S

# Contents

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

# 1. 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 the Credit Card Default Prediction dashboard. LLDD describes the class diagrams with the methods and relations between classes and programs specs. 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

The financial industry continually grapples with the challenge of predicting and managing credit card defaults. As the credit landscape evolves, there is an increasing need for advanced analytics and machine learning solutions to enhance risk assessment processes. The Credit Card Default Prediction project aims to address this crucial aspect by developing a predictive model that analyzes credit card owner characteristics and payment history to forecast the likelihood of default.

Recognizing the urgency of this challenge, our project focuses on developing a robust solution to address one of the most significant threats faced by commercial banks the accurate prediction of credit client risks. The primary goal is to predict the probability of credit default based on the characteristics of credit card owners and their payment history.

## 2. Problem Statement

The financial industry is undergoing unprecedented advancements, bringing both opportunities and challenges. One of the predominant challenges faced by commercial banks is the escalating trend in credit risk. As the landscape evolves, the need for accurate credit client risk prediction becomes paramount. The specific focus is on predicting the probability of credit default based on the characteristics of credit card owners and their payment history.

## 3. Dataset Information

**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

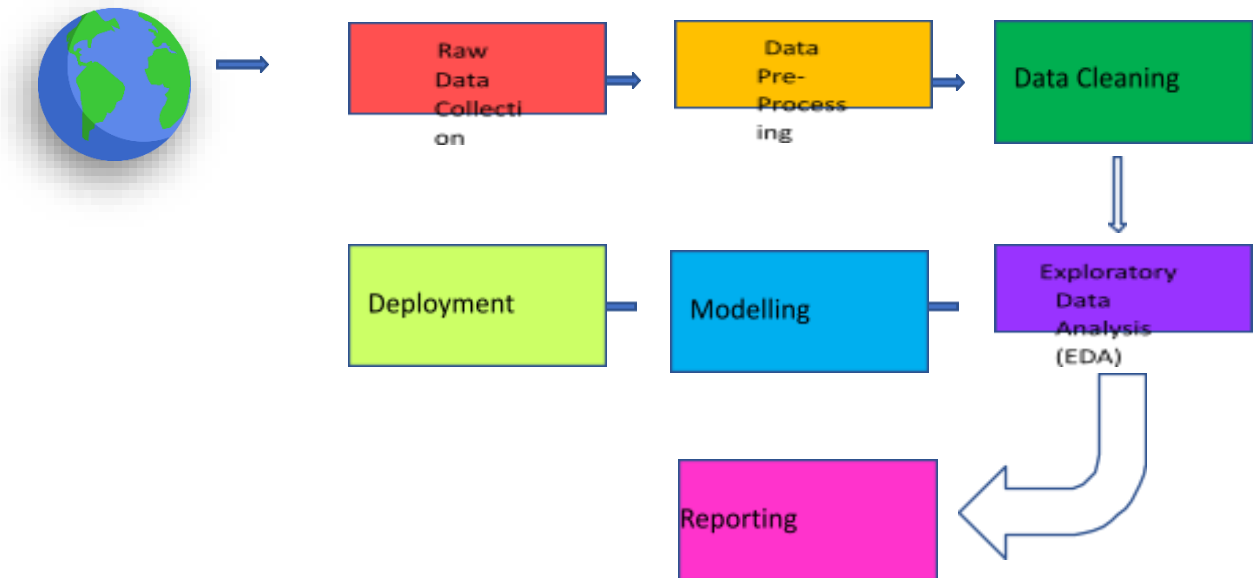
**PAY\_0 to PAY\_6:** Repayment status in September, 2005 (-1=pay duly, 1=payment delay for one month, 2=payment delay for two months, 8=payment delay for eight months, 9=payment delay for nine months and above)

**BILL\_AMT1 to BILL\_AMT6:** Amount of bill statement for the last 6 months.

**PAY\_AMT1 to PAY\_AMT6:** Amount of previous payment made for the last 6 months.

**Default.payment.next.month:** Target variable indicating whether the individual will default on the credit card payment next month (1 = yes, 0 = no).

## Architecture



### 3.1 Architecture Description

#### 1. Raw Data Collection

The Dataset was taken from iNeuron's Provided Project Description  
<https://www.kaggle.com/datasets/uciml/default-of-credit-card-clients-dataset>

#### 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 fed 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 self-explanatory report because your model will be used by many stakeholders who are not from technical background.

- High Level Design Document (HLD)
- Low Level Design Document (LLD)
- Architecture
- Wireframe
- Detailed Project Report
- Power Point Presentation

#### 6. Modelling

Data Modelling is the process of analysing the data objects and their relationship to the other objects. It is used to analyse 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

We created a Power BI Dashboard

