**Low Level Design (LLD)**

**Adult Censes Income Prediction**



**Revision Number - 1.0**

**Last Date of Revision - 02/11/2022**

# Document Control

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** |  | **Version** | **Description** | **Author** |
| 02-11-2022 | 1.0 |  | LLD document creation | Adil Anwar |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Contents**

Document Version Control…………………………………………………………………………2

1. Introduction……………………………………………………………………………………….4
   1. What is Low Level Design Document?.................................................................4
   2. Scope……………………………………………………………………………………4
   3. Project Introduction……………………………………………………………………4
2. Problem Statement.……………………………………………………………………………...5
3. Dataset Information……………………………………………………………………….……...5
4. Architecture…………………………………………………………………………………….…6
   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 Heart Disease Diagnostic Analysis 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 stepby-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 prominent inequality of wealth and income is a huge concern especially in the United States.
* The likelihood of diminishing poverty is one valid reason to reduce the world's surging level of economic inequality.
* The principle of universal moral equality ensures sustainable development and improve the economic stability of a nation.
* Governments in different countries have been trying their best to address this problem and provide an optimal solution.
* This study aims to show the usage of machine learning and data mining techniques in providing a solution to the income equality problem.
* The UCI Adult Dataset has been used for the purpose.
* Classification has been done to predict whether a person's yearly income in US falls in the income category of either greater than 50K Dollars or less equal to 50K Dollars category based on a certain set of attributes.

# 2. Problem Statement

* The Goal is to predict whether a person has an income of more than 50K a year or not.
* This is basically a binary classification problem where a person is classified into the >50K group or <=50K group.

.

# 3. Dataset Information

**age:** The person's age in years

**sex:** The person's sex (1 = male, 0 = female)

**work class:** The person's class is divided

Private- 70%

Self-emp-not-inc- 8%

Other (7324)- 22%

**Education :** The person's is taken

HS-grad32%

Some-college22%

Other (14769)45%

**marital-status :** The person's marital-status is taken

Married-civ-spouse46%

Never-married33%

Other (6902)21%

**Relationship :** Realtionship of the person is taken

Husband41%

Not-in-family26%

Other (11063)34%

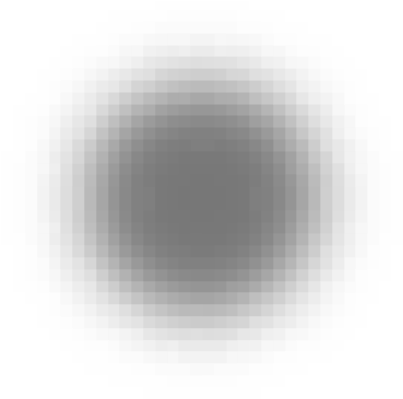
**Race :** The person's race is taken

White85%

Black10%

Other (1621)5%

# 4. Architecture



Raw Data

Collection

Data

Pre

-

Processing

Data Cleaning

Exploratory Data

Analysis

EDA

)

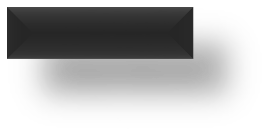
(

Modelling

Deployment

R

eporting



Real World

## 4.1 Architecture Description

1. **Raw Data Collection** The Dataset was taken from iNeuron’s Provided Project Description Document.

[adult census dataset | Kaggle](https://www.kaggle.com/datasets/overload10/adult-census-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 feeded to the model to train.

This Process includes-

1. Handling Null/Missing Values
2. Handling Skewed Data
3. 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.

1. Remove duplicate or irrelevant observations
2. Filter unwanted outliers
3. 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.

1. High Level Design Document (HLD)
2. Low Level Design Document (LLD)
3. Architecture
4. Wireframe
5. Detailed Project Report
6. 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.