**Low Level Design (LLD) Heart Disease Diagnostic Analysis**



**Revision Number - 1.2**

**Last Date of Revision –018/09/2022**

**Vinay Pandhe**

**Document Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Version** | **Description** | **Author** |
| 17/09/2022 | 1.0 | Introduction, Problem Statement | Vinay Pandhe |
| 18/09/2022 | 1.1 | Dataset Information, Architecture Description | Vinay Pandhe |
| 18/09/2022 | 1.2 | Final Revision | Vinay Pandhe |

**2**

**Why this Low-Level Design Document?**

The purpose of this document is to present a detailed description of the heart disease prediction analysis technique. It will explain the necessary steps which have to be followed before any analysis can begin. The document would also describe the algorithms and techniques used to predict the presence and absence of the heart disease and present a comparative result for the same. LLD 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. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The LLD will be focusing on the below objectives:

* Problem Understanding.
* Data Acquisition.
* Data Pre-Processing and Exploratory Analysis
* Development of models
* Auditing accuracy and retrain if require
* Finalizing the model
* Dashboard report for important activities

**Scope**

The LLD documentation presents the detailed structure of the heart disease prediction system for each of its individual components. The goal of LLD is to give the internal logical design of the actual program code. Low-level design is created based on the high-level design. The LLD documentation contains the complete description of the model used along with the comparisons of the proposed model/library compared with a baseline(existing) model against a set of metrics.

**Project Introduction**

Heart disease (HD) is one of the most common diseases nowadays, and an early diagnosis of such a disease is a crucial task for many health care providers to prevent their patients for such a disease and to save lives.

The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions. For providing appropriate results and making effective decisions on data, some data science techniques need to be used. The data analysis predicts the likelihood of patients getting heart disease. It enables significant knowledge.

Eg. Relationships between medical factors related to heart disease and patterns, to be established. The obtained results have illustrated that the designed diagnostic system can effectively predict the risk level of heart diseases.

**Constraints**

Our analysis is done based on a limited dataset provided for a specific (14) features affecting heart disease. The analysis does not take into account any external interventions like underlying disease, type of medication used, lifestyle patterns, BMI value etc.

**Risks**

Document specific risks that have been identified or that should be considered.

**Out of Scope**

Delineate specific activities, capabilities, and items that are out of scope for the project.

* **Technical specifications**

**Dataset**

The Dataset is taken from iNeuron’s provided Project Description Document.



Figure 1: Heart Disease Prediction dataset

The dataset consists of 303 individual data. There are 14 columns in the dataset, which are described below.

* Age - Display the age of the individual.
* Sex - Display the gender of the individual using the following format:

1=male and 0 = female.

* Chest-pain (cp) - Displays the type of chest-pain experienced by the individual using the following format:

0 = typical angina

1 = atypical angina

2 = non anginal pain

3 = asymptotic

* Resting Blood Pressure (trestbps) - Displays the resting blood pressure value of an individual in mmHg (unit)
* Serum Cholesterol (chol) - Displays the serum cholesterol in mg/dl (unit)
* Fasting Blood Sugar (fbs) - Compare the fasting blood sugar value of an individual with 120mg/dl.

If fasting blood sugar > 120mg/dl then: 1 (true) else

: 0 (false)

* Resting ECG (restecg) - Display resting electrocardiographic results

0 = normal

1 = having ST-T wave abnormality.

2 = left ventricular hypertrophy.

* Max heart rate achieved (thalach) - Display the max heart rate achieved by an individual.
* Exercise induced angina (exang) -

1= yes

0 = no

* ST depression induced by exercise relative to rest (old peak) - Display the value which is an integer or float.
* Peak exercise ST segment (slope)-

0 = upsloping

1 = flat

2 = down sloping

* Number of major vessels (0–4) colored by Fluoroscopy (ca) - Display the value as integer or float.
* thal - displays the thalassemia- (3 = normal; 6 = fixed defect; 7 =reversable defect)
* Diagnosis of heart disease- (num) Display whether the individual is suffering from heart disease or not.
* **Problem Statement**

Health is real wealth in the pandemic time we all realized the brute effects of covid-19 on all irrespective of any status. You are required to analyze this health and medical data for better future preparation. A dataset is formed by taking into consideration some of the information of 303 individuals.

* **Exploratory Data Analysis (EDA)**
* **Modelling**
* **Deployment**
* **Data Cleaning**
* **Data Pre- Processing**
* **Raw Data Collection**
* **Architecture**

Reporting

* **Architecture Description**
* **Raw Data Collection-** The dataset is taken from iNeuron’s which is provided in the introduction document of project.

<https://drive.google.com/drive/folders/165Pjmfb9W9PGy0rZjHEA22LW0Lt3Y-Q8>

* **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-

* Handling Null/Missing Values
* Handling Skewed Data
* Outliers Detection and Removal
* **Data Cleaning**

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

* Remove duplicate or irrelevant observations
* Filter unwanted outliers
* Renaming required attributes
* **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.

* **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 explatory 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
* **Modelling**

Data Modelling 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.

* **Deployment**

We created a Tableau dashboard.

