LOW LEVEL DESIGN (LLD)

Heart Disease Data Analysis

Ashish Haldar
Ashishhaldar38@gmail.com

Document Version Control

Date Issued	Version	Description	Author
19/06/2022	1.0	Introduction, Problem Statement	Ashish
30/06/2022	1.1	Dataset Information, Architecture Description	Ashish
13/07/2022	1.2	Final Revision	Ashish

Contents

1. Introduction

1.1. Why this Low-Level Design (LLD)?

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. 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.

1.2. 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

Heart disease is a term covering any disorder of the heart. Heart diseases have become a major concern to deal with as studies show that the number of deaths due to heart diseases have increased significantly over the past few decades in India it has become the leading cause of death in India. A study shows that from 1990 to 2016 the death rate due to heart diseases have increased around 34% from 155.7 to 209.1 deaths per 1 lakh population in India.

Thus, preventing heart diseases has become more than necessary. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives.

2. 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.

The goal of this project is to analyze to predict the probability of heart disease occurrence, based on a combination of features that describes the disease. To achieve the goal, we used a data set that is formed by taking into consideration some of the information of 303 individuals. The problem is based on the given information about

each individual we have to calculate probability that whether that individual will suffer from heart disease or not.

3. Dataset Description

1. **age**: Age in years
2. **sex**: Sex of Patient
   - 1: Male
   - 0: Female
3. **cp**: Chest Pain Type
   - 1: Typical Angina
   - 2: Atypical Angina
   - 3: Non-anginal Pain
   - 4: Asymptomatic
4. **trestbps**: Resting Blood Pressure (in mm Hg on admission to the hospital)
5. **chol**: Serum Cholesterol in mg/dl
6. **fbs**: Fasting Blood Sugar > 120 mg/dl
   - 1: True
   - 0: False
7. **restecg**: Resting Electrocardiographic Results
   - 0: Normal
   - 1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
   - 2: Showing probable or definite left ventricular hypertrophy by Estes' criteria
8. **thalach**: Maximum Heart Rate Achieved
9. **exang**: Exercise Induced Angina
   - 1: Yes
   - 0: No

10. **oldpeak**: ST depression induced by exercise relative to rest
11. **slope**: The slope of the peak exercise ST segment

1: upsloping

2: flat

3: downsloping

12. **ca**: Number of major vessels (0-3) coloured by Fluoroscopy

13. **thal**: Thalassemia

3: Normal

6: Fixed Defect

7: Reversable Defect

14. **num**: Diagnosis of heart disease (angiographic disease status)

0: < 50% diameter narrowing

1: > 50% diameter narrowing

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	45	1	1	110	264	0	0	132	0	1.2	2	0	7	1
299	68	1	4	144	193	1	0	141	0	3.4	2	2	7	1
300	57	1	4	130	131	0	0	115	1	1.2	2	1	7	1
301	57	0	2	130	236	0	2	174	0	0.0	2	1	3	1
302	38	1	3	138	175	0	0	173	0	0.0	1	-100000	3	0

303 rows × 14 columns

4. Architecture
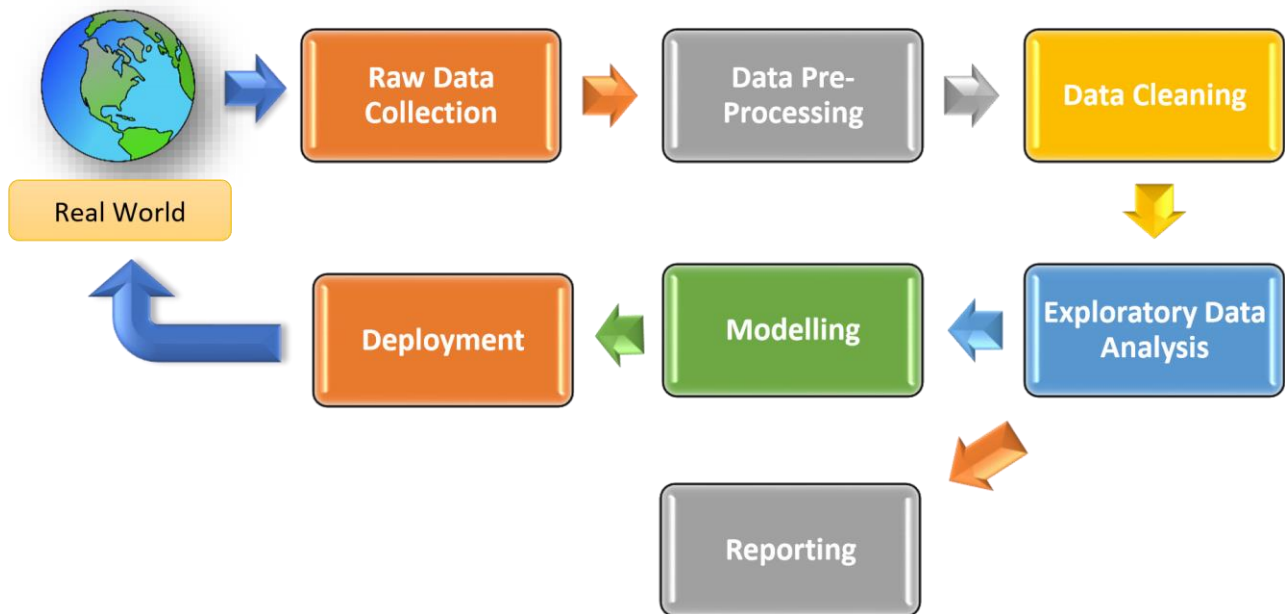
4.1. Architecture Description

1. Raw Data Collection

The Dataset was taken from iNeuron's Provided Project Description Document.

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

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 self-explanatory 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. 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.

7. Deployment

We created a Power BI Dashboard
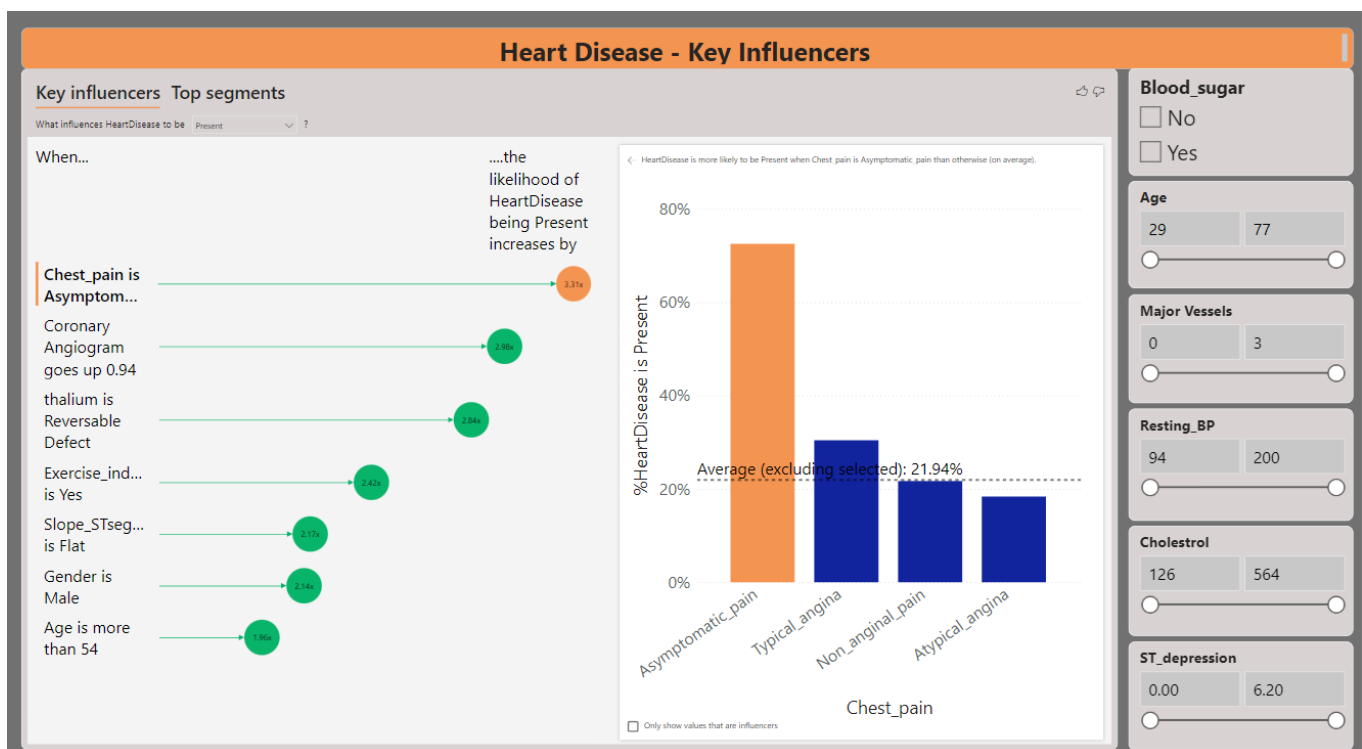
5. Unit Test Case

Test Case Description	Expected Result
Age Slicer	Slider for age is given to check result for different age groups
Blood Sugar Slicer	It will show whether a person has Blood sugar or not
Major vessels Slicer	It will check result for different vessels (0-3)
Resting BP slicer	It will check result for different BP from 94 to 200

Cholesterol Slicer	It will check result for different cholesterol starting from 126 to 564
ST depression Slicer	It will check result for different depression from 0 to 6.20
Key influencers	The key influencer's visual helps us to understand the factors that drive a metric we are interested in. It analyzes our data, ranks the factors that matter, and displays them as key indicators.