Low Level Design (LLD)

Heart Disease Diagnostic Analysis

Revision Number - 1.2
Last Date of Revision - 03/07/2022

SRIMANTA SHEKHAR MISHRA

SWADESH RANJAN PANDA

Document Control

Date	Version	Description	Author
19/06/2022	1.0	Introduction, Problem Statement	Srimanta & Swadesh
26/06/2022	1.1	Dataset Information, Architecture Description	Srimanta & Swadesh
03/07/2022	1.2	Final Revision	Srimanta & Swadesh

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. E.g. 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.

1. Technical specifications

Dataset

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

Out[3]:

	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

Figure 1: Heart Disease Prediction dataset

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

1. ***Age***: displays the age of the individual.
2. ***Sex***: displays the gender of the individual using the following format :
   1 = male
   0 = female
3. ***Chest-pain type***: 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
4. ***Resting Blood Pressure***: displays the resting blood pressure value of an individual in mmHg (unit)
5. ***Serum Cholestrol***: displays the serum cholesterol in mg/dl (unit)

6. ***Fasting Blood Sugar***: compares the fasting blood sugar value of an individual with 120mg/dl.
   If fasting blood sugar > 120mg/dl then : 1 (true) else
   : 0 (false)

7. ***Resting ECG*** : displays resting electrocardiographic results
   0 = norma1 = having ST-T wave abnormality
   2 = left ventricular hyperthrophy

8. ***Max heart rate achieved*** : displays the max heart rate achieved by an individual.

9. ***Exercise induced angina*** :
   1 = yes
   0 = no

10. ***ST depression induced by exercise relative to rest***: displays the value which is an integer or float.

11. ***Peak exercise ST segment*** :
    0 = upsloping
    1 = flat
    2 = downsloping

12. ***Number of major vessels (0–4) coloured by flourosopy*** : displays the value as integer or float.

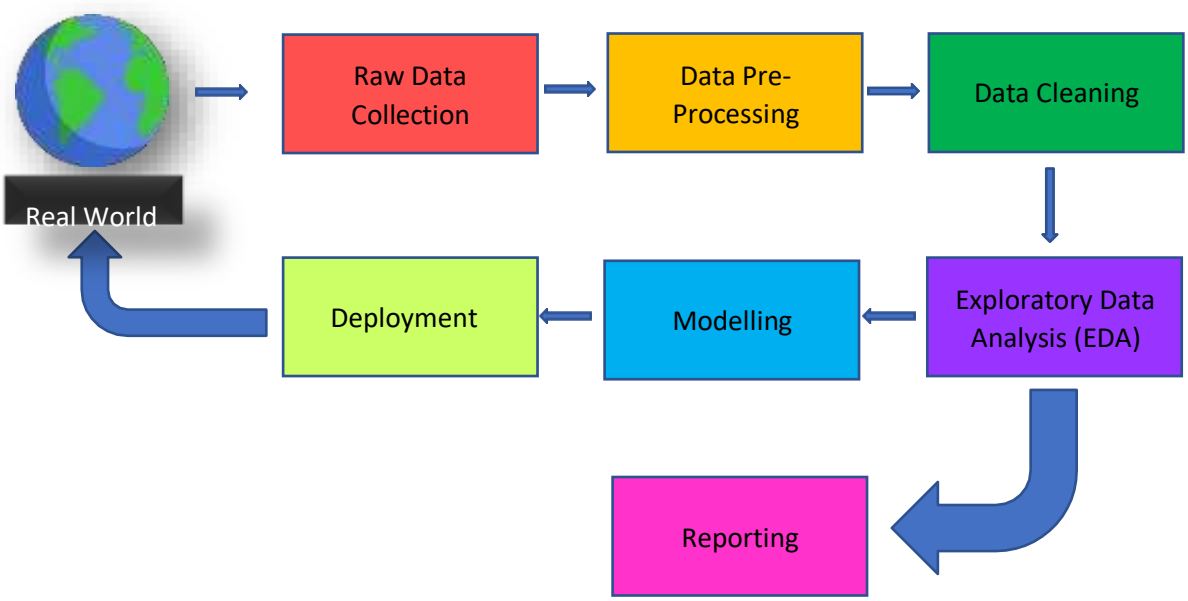
13. ***Thal*** : displays the thalassemia :
    0,1 = normal
    2 = fixed defect
    3 = reversible defect

14. ***Diagnosis of heart disease*** : Displays whether the individual is suffering from heart disease or not :
    0 = absence
    1 = present

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 analyse this health and medical data for better future preparation. A dataset is formed by taking into consideration some of the information of 303 individuals.

3. 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 feeded 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. 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