# HIGH LEVEL DOCUMENT

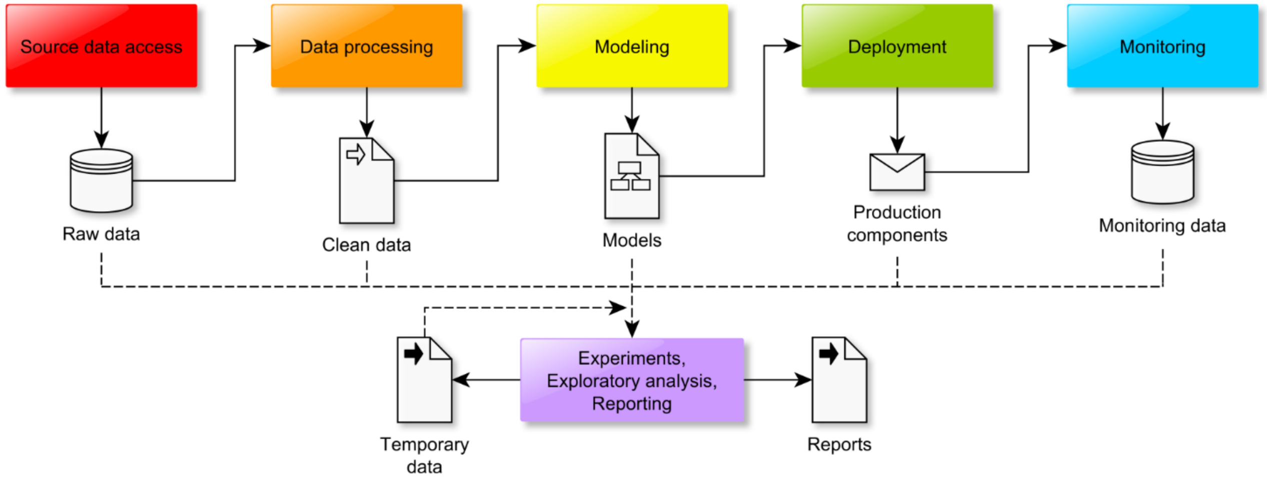
Healthcare Analytics on Heart Disease Data

| | |
|---|---|
| Written By | Atmik Pandey |
| Document Version | 0.1 |
| Last Revised Date | |

# Contents

# 1. Introduction

# 3. Architecture Description



## 3.1 Data Collection

The dataset is downloaded from Kaggle. The dataset contains 14 columns and 1000+ rows. The dataset includes age, sex, chest pain type, resting blood pressure ,serum cholesterol in mg/dl, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved etc. This is given in the comma separated value format(.csv).

## 3. 2 EDA

In EDA, we have seen various insights from the dataset like checking top 5 rows and bottom 5 rows. There are no null values in the dataset.

## 3. 3 Data Preprocessing

Our data is not After the raw text is preprocessed, we have to encode our labels to numerical encoding as they were categorically encoded before. We use the R inbuilt libraries to encode the variables.

## 3. 4 Data Analysis

We perform various analysis on our dataset to figure out what key metrics are the most important factors leading to a heart disease. We figure out the important variables for the model creation using correlation analysis and multicollinearity if any is removed. We draw various graphs using variables to asses the relationship between them and figure out its significance. Non significant varibales are dropped to make the model leaner and more efficient to run and deploy on the cloud.

## 3.4 Model Creation

After various analysis has been performed and the key useful variables identified, we split our data into training and test set. We train the models on the training set and after the model has been trained, we test the prediction accuracy of the model on the test data set. The model primarily used is the Logistic Regression Model.

## 3.5 Model Dump

After comparing all accuracies and checking all regression metrics, The model is dumped used pickle format.

## 3.6 Data From User

Here we will collect user's requirements to classify with news headline or with complete news

## 3.9 Deployment

The model being a static Data Analysis Model cannot be deployed on the cloud.

# 4. Unit Test Cases

| Test Cases Description | Prerequisite Expected Result |
|---|---|
| Verify whether the Application URL is accessible to the user. | Application URL<br>Application URL should be<br>should be defined accessible to the user |

| | | |
|---|---|---|
| Verify whether the Application loads completely for the user when the URL is accessed | Application URL is The Application should accessible load completely for the Application is user when the URL is deployed accessed | |
| Verify whether the user is giving standard input. | Handled test cases Users should be able to at backends. see successfully valid results. | |
| Verify whether user is able to see input fields | Application is User should be able to see accessible input fields | |
| Verify whether user is able to edit all input fields<br>Verify whether gets predict or classify news into categories | 2.Application is accessible. User is logged in to the application<br>1. Application isaccessible.<br>2. User is logged in tothe application | User should be able to edit all input fields<br>Users should get the Submit button to submit the inputs. |

| | | |
|---|---|---|
| Verify whether the user is presented with recommended results on clicking submit.<br>Verify whether the recommended results are in accordance to the selections user made | 1.Application is accessible. 2.User is logged in to the application<br>1.Application is accessible. 2.User is logged in to the application | User should be presented with recommended results on clicking submit<br>The recommended results should be in accordance to the selections user made |

| | | |
|---|---|---|
| Verify whether user has options to filter the recommended results as weel | 1. Application isaccessible<br>2. User is logged in tothe application | User should have options to filter the recommended results as well. |