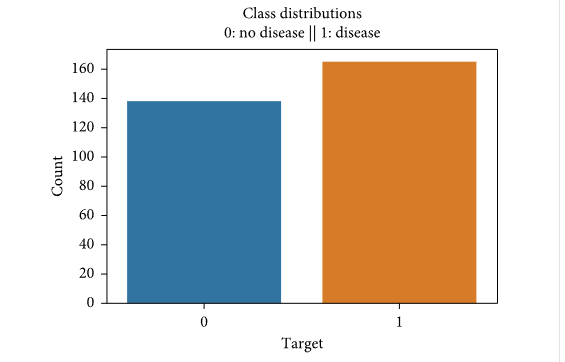
**Introduction**

Heart disease is one of many illnesses that can be fatal, and it has received a lot of attention in medical studies. Heart disease diagnosis is a difficult undertaking that can provide automated predictions about the patient's heart state to improve the effectiveness of subsequent treatment. Heart disease is typically diagnosed based on the patient's physical examination, signs, and symptoms. The risk of heart disease is influenced by a number of variables, including smoking, body cholesterol, family history of the disease, obesity, high blood pressure, and inactivity. The provision of high-quality services at reasonable prices is a significant problem for healthcare institutions, including hospitals and medical facilities.

This makes heart disease a serious issue that has to be addressed. However, because of numerous contributing risk factors, including diabetes, high blood pressure, high cholesterol, an irregular pulse rate, and many other factors, it can be challenging to diagnose heart disease. Due to these limitations, researchers are now using cutting-edge techniques like data mining and machine learning to forecast disease.

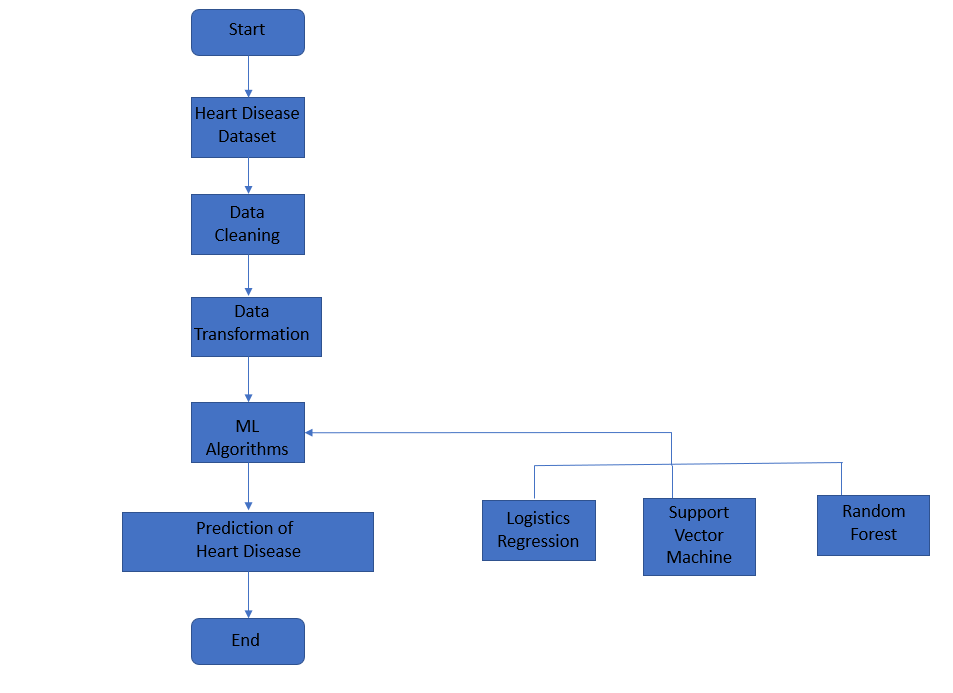


**Figure1: Targeted Data of disease & no disease.**

Making choices and predictions from the vast amount of data generated by the healthcare sector is made easier with the help of machine learning (ML). Data analysis is essential in the medical industry. It gives important decisions a solid foundation. Making a comprehensive study proposal is helpful. One of the most significant applications of data analysis is the removal of human bias from medical conclusions through the use of appropriate statistical analysis. because there is nontrivial information in enormous amounts of data, by using data mining for exploratory analysis.

With the use of data analytics expertise, the heart predictor system will provide a user-oriented approach to new and hidden patterns in the data. The knowledge that is put into practise can be used by healthcare professionals to improve service quality and lessen the severity of heart disease side effects.

**Work Flow:**



**Figure 2: Work Flow**

**Objectives:**

The creation of a heart prediction system is the main goal of this study. The algorithm can find and extract disease-related hidden knowledge from a historical heart data set. To aid in the prediction of heart disorders, the heart disease prediction system uses data mining techniques on a set of medical data.

1. To determine the importance of important traits in predicting heart disease.
2. Using the estimated weight of relevant features to forecast heart disease (using ML algorithms) to assess accuracy in predicting heart disease.
3. Offers a fresh perspective on hidden patterns in the data.
4. Helps avoid human biasness.

**Literature Survey:**

This research articles are referred for this project:

1. Heart disease diagnosis has been the subject of numerous investigations. They used various data mining techniques for diagnosis and obtained various probabilities for various approaches.
2. For the purpose of identifying high-risk and low-risk patients, two open Holster databases were employed. Melillo et al. (2013) employed the cart method for the aim of categorization.
3. In 2015, Imani and Ghassemian used a weighted training sample strategy that included feature extraction for the spatial dimension of the images because there are various instances when there is not enough data.

**Existing System:**

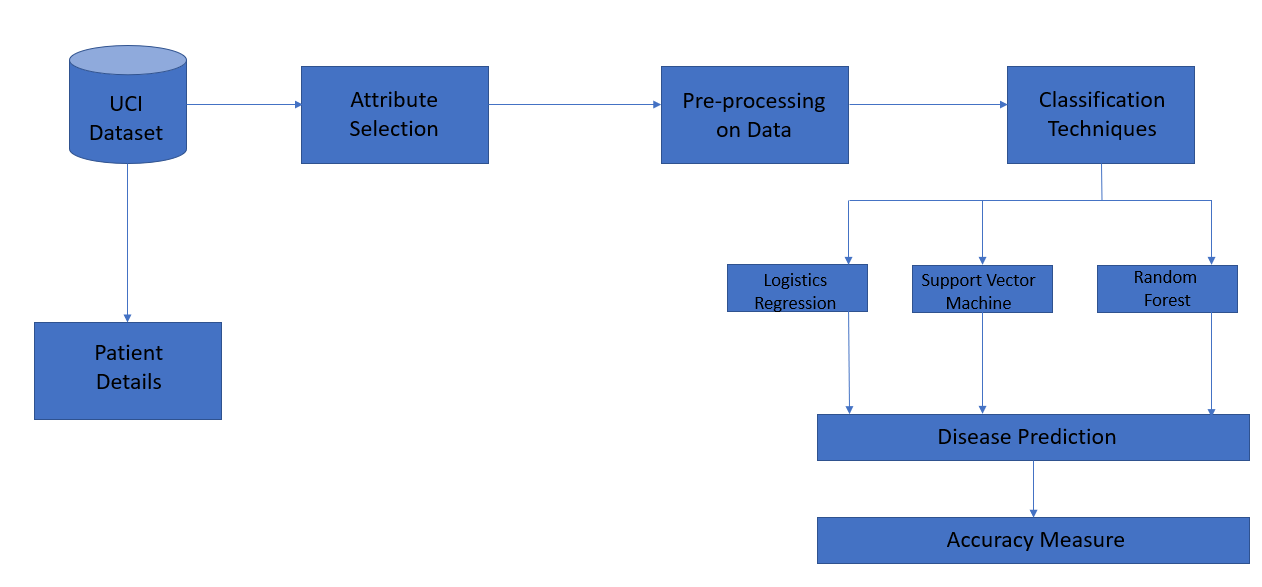
The majority of earlier studies focused on identifying characteristics that improve cardiac prediction accuracy. However, very few studies have examined the connections between these traits. The Associative Rule Mining (ARM) technique can be used to determine the association between each feature that contributes to the prediction of heart disease.

However, few scientists (Jabbar, Sundar, Soni and Vyas) have concentrated on using WARM to treat cardiovascular illness. Studies on the use of WARM to predict heart disease are shown. However, an exact calculation of the feature weight was not made

**Proposed System: Heart Disease Prediction**

We are using regression algorithms, Random Forest algorithm, SVM in this system to develop an efficient heart attack prediction system. The system can receive input from a UCI Heart CSV file. After gaining access to the data set, the process is carried out, and a useful heart attack level is generated.

The suggested approach will include additional heart attack risk factors such as weight, age, and priority levels. The heart attack prediction system was created to assist in identifying various heart attack risk levels, such as normal, low, and high, and to provide prescription information in relation to the anticipated outcome.



**Figure3: Proposed System Architecture**

**Dataset:**

The UCI Machine Learning Repository's dataset on heart disease is used in this study. One of the largest datasets is the UCI Machine Learning Repository, which has over 417 different datasets. One of the datasets on heart disease that has been extensively used by academics to date is the Cleveland dataset from the UCI Machine Learning Repository. The dataset has 76 features, 14 of which, including class label, are attributes.

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 :  
   1 = typical angina  
   2 = atypical angina  
   3 = non — anginal pain  
   4 = 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 = normal  
   1 = 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 :  
    1 = upsloping  
    2 = flat  
    3 = downsloping
12. Number of major vessels (0–3) colored by flourosopy : displays the value as integer or float.
13. Thal : displays the thalassemia :  
    3 = normal  
    6 = fixed defect  
    7 = reversible defect
14. Diagnosis of heart disease : Displays whether the individual is suffering from heart disease or not :  
    0 = absence  
    1, 2, 3, 4 = present.

## Why these parameters:

In the actual dataset, we had 76 features but for our study, we chose only the above 14 because:

1. **Age**: Age is the most important risk factor in developing cardiovascular or heart diseases, with approximately a tripling of risk with each decade of life. Coronary fatty streaks can begin to form in adolescence. It is estimated that 82 percent of people who die of coronary heart disease are 65 and older. Simultaneously, the risk of stroke doubles every decade after age 55.
2. **Sex**: Men are at greater risk of heart disease than pre-menopausal women. Once past menopause, it has been argued that a woman’s risk is similar to a man’s although more recent data from the WHO and UN disputes this. If a female has diabetes, she is more likely to develop heart disease than a male with diabetes.
3. **Angina (Chest Pain)**: Angina is chest pain or discomfort caused when your heart muscle doesn’t get enough oxygen-rich blood. It may feel like pressure or squeezing in your chest. The discomfort also can occur in your shoulders, arms, neck, jaw, or back. Angina pain may even feel like indigestion.
4. **Resting Blood Pressure**: Over time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more.
5. **Serum Cholesterol**: A high level of low-density lipoprotein (LDL) cholesterol (the “bad” cholesterol) is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also ups your risk of a heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the “good” cholesterol) lowers your risk of a heart attack.
6. **Fasting Blood Sugar**: Not producing enough of a hormone secreted by your pancreas (insulin) or not responding to insulin properly causes your body’s blood sugar levels to rise, increasing your risk of a heart attack.
7. **Resting ECG**: For people at low risk of cardiovascular disease, the USPSTF concludes with moderate certainty that the potential harms of screening with resting or exercise ECG equal or exceed the potential benefits. For people at intermediate to high risk, current evidence is insufficient to assess the balance of benefits and harms of screening.
8. **Max heart rate achieved**: The increase in cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.
9. **Exercise induced angina**: The pain or discomfort associated with angina usually feels tight, gripping or squeezing, and can vary from mild to severe. Angina is usually felt in the center of your chest but may spread to either or both of your shoulders, or your back, neck, jaw or arm. It can even be felt in your hands. o Types of Angina a. Stable Angina / Angina Pectoris b. Unstable Angina c. Variant (Prinzmetal) Angina d. Microvascular Angina.
10. **Peak exercise ST segment**: A treadmill ECG stress test is considered abnormal when there is a horizontal or down-sloping ST-segment depression ≥ 1 mm at 60–80 ms after the J point. Exercise ECGs with up-sloping ST-segment depressions are typically reported as an ‘equivocal’ test. In general, the occurrence of horizontal or down-sloping ST-segment depression at a lower workload (calculated in METs) or heart rate indicates a worse prognosis and higher likelihood of multi-vessel disease. The duration of ST-segment depression is also important, as prolonged recovery after peak stress is consistent with a positive treadmill ECG stress test. Another finding that is highly indicative of significant CAD is the occurrence of ST-segment elevation > 1 mm (often suggesting transmural ischemia); these patients are frequently referred urgently for coronary angiography.

**Software Requirement Specification:**

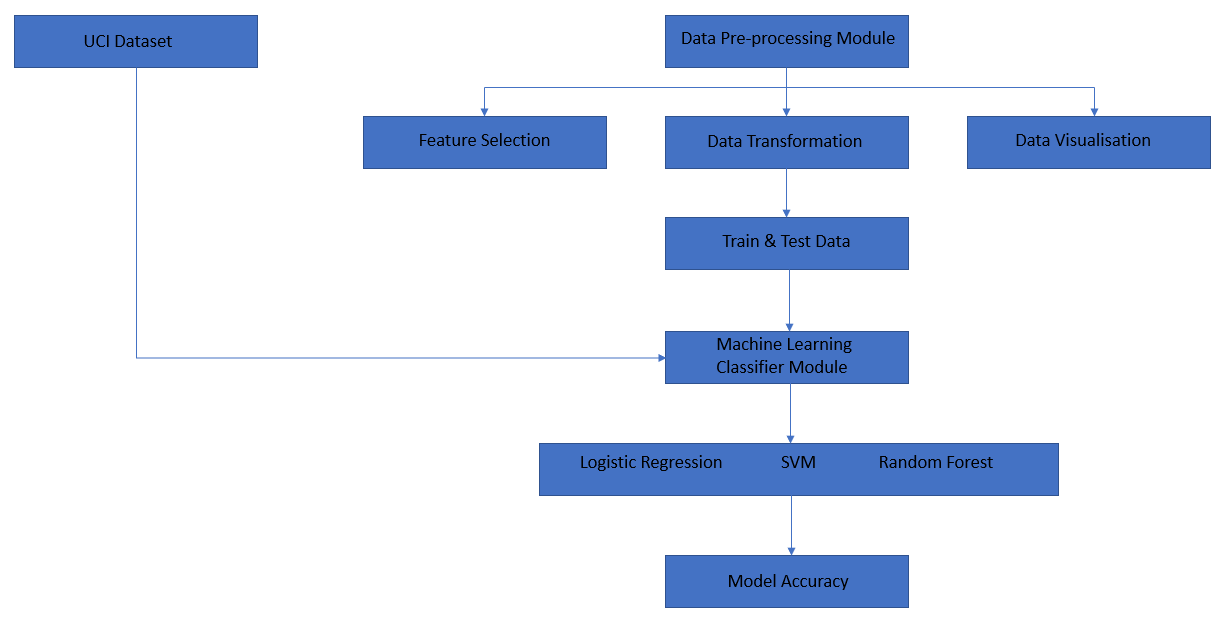
Minimum System requirements for R programming:

1. Operating system: Linux- Ubuntu 16.04 to 17.10,
2. Windows 7 to 10,
3. RStudio 2022.02.3+492 "Prairie Trillium" Release (1db809b8323ba0a87c148d16eb84efe39a8e7785, 2022-05-20) for Windows
4. Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) QtWebEngine/5.12.8 Chrome/69.0.3497.128 Safari/537.36
5. Katonic AI cloud workspace to run R program Web application.

**Hardware Requirement:**

1. System with Intel Pentium Core i3 and above

**Methodology:**



**Figure4: Methodology Used**

**1) Pre processing of the Dataset**

There are no null values in the dataset. However, there were a lot of outliers that needed to be handled carefully, and the dataset's distribution is also off. Two strategies were applied. Without using outliers, a feature selection procedure, or putting the data straight to machine learning algorithms, the results were not encouraging.

However, the results obtained are extremely encouraging after employing the normal distribution of the dataset to overcome the overfitting issue and then applying Random Forest for the outlier's detection. Different charting methods were employed to examine the data's skewness, outlier detection, and distribution. All of these preprocessing methods are crucial when transferring data for categorization or prediction.

**2) Feature Selection**

When doing feature selection, the Lasso algorithm, which is a component of embedded techniques, is utilised for selecting the features and just selecting the important features. Compared to filter approaches, it has higher prediction accuracy.

It produces good feature subsets for the algorithm being utilised. The model that is a component of feature selection in the scikit-learn library should then be chosen in order to pick the features that have been chosen.

**3)Data Transformation**

Transform the Unstructured data into the precision format for the targeted data. The UCI database's experiments have mostly focused on attempting to discern between presence (values 1, 2, 3, 4) and absence (value 0).

Since the majority of machine learning algorithms demand integer values, attributes having category values were changed to numerical values. For variables with more than two categories, dummy variables were also made. Dummy variables aid in the more accurate learning of the data by neural networks.

**4) Logistics Regreeion Model**

Logistic regression becomes a classification technique only when a decision threshold is brought into the picture. The setting of the threshold value is a very important aspect of Logistic regression and is dependent on the classification problem itself.  
The decision for the value of the threshold value is majorly affected by the values of [precision and recall.](https://www.geeksforgeeks.org/confusion-matrix-machine-learning/) Ideally, we want both precision and recall to be 1, but this seldom is the case.

In the case of a Precision-Recall tradeoff, we use the following arguments to decide upon the threshold:-  
**1. Low Precision/High Recall:** In applications where we want to reduce the number of false negatives without necessarily reducing the number of false positives, we choose a decision value that has a low value of Precision or a high value of Recall.

**2. High Precision/Low Recall:** In applications where we want to reduce the number of false positives without necessarily reducing the number of false negatives, we choose a decision value that has a high value of Precision or a low value of Recall.  
Based on the number of categories, Logistic regression can be classified as:

**binomial:** target variable can have only 2 possible types: “0” or “1” which may represent “win” vs “loss”, “pass” vs “fail”, “dead” vs “alive”, etc.

**multinomial:** target variable can have 3 or more possible types which are not ordered(i.e. types have no quantitative significance) like “disease A” vs “disease B” vs “disease C”.

**ordinal:** it deals with target variables with ordered categories. For example, a test score can be categorized as:“very poor”, “poor”, “good”, “very good”. Here, each category can be given a score like 0, 1, 2, 3.

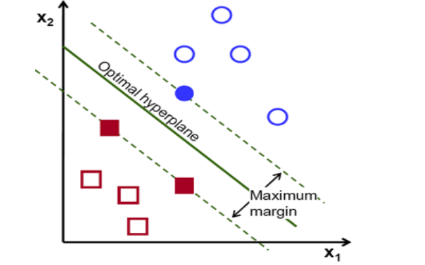


**Figure5: Logistics Regression Model**

**5)Support Vector Machine**

Support Vector Machine: Support vector machine is the one of the machine learning algorithms. The support vector machine is a supervised learning algorithm. The support vector machine is used to classify the given data. The algorithm uses a hyper plane to differentiate the different classes. Support vector machine is also used for the regression analysis. SVM classify the both linear and non-linear data.

The main aim of the SVM classifier is to find the hyper plane in an n-dimensional space.



**Figure6: SVM Model**

**6) Random Forest Algorithm**

Random forest algorithm is one of the most effective ensemble classification approach. The RF algorithm has been used in prediction and probability estimation.RF consists of many

decision trees .Each decision tree gives a vote that indicate the decision about class of the object. Random forest item was first proposed by Tin kam HO of bell labs in 1995.

RF method combines bagging and random selection of features. There are three important tuning parameters in random forest1) No. of trees (n tree) 2) Minimum node size 3) No. of features employed in splitting each node 3) No. of features employed in splitting each node for each tree (m try).

Random forest algorithm advantages are listed below.

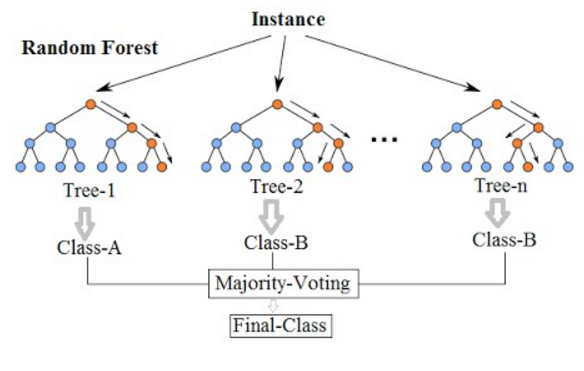
1) The ensemble learning algorithm using random forest is accurate.

2) For huge data sets, random forest performs well.

3) It can manage a large number of input variables.

4) Random forest calculates the key classification variables. It can deal with missing data.

5) For class unbalanced data sets, Random Forest includes ways for balancing error.



**Figure7: Random Forest Algorithm**

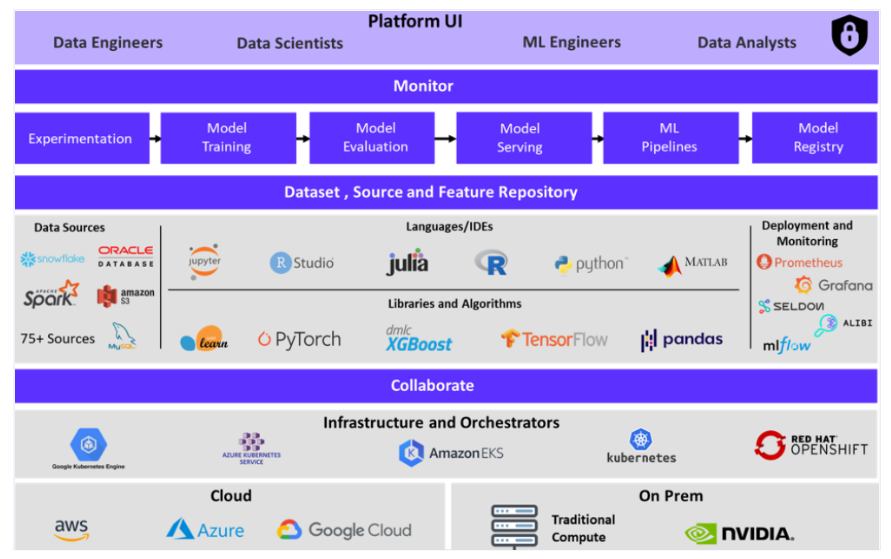
**7) Katonic AI Cloud Workspace**

Katonic Provides Machine Learning Operations (MLOps) Platform to develop, deploy, monitor, and manage advanced analytics and machine learning (ML) and AI solutions in a self-service, collaborative, governed, and secure manner. Katonic helps enterprises with

* Shorter development cycles accelerating innovation
* Faster time to market of ML solutions
* Assurance of quality, trustworthiness and ethical AI

**Features is as follow:**

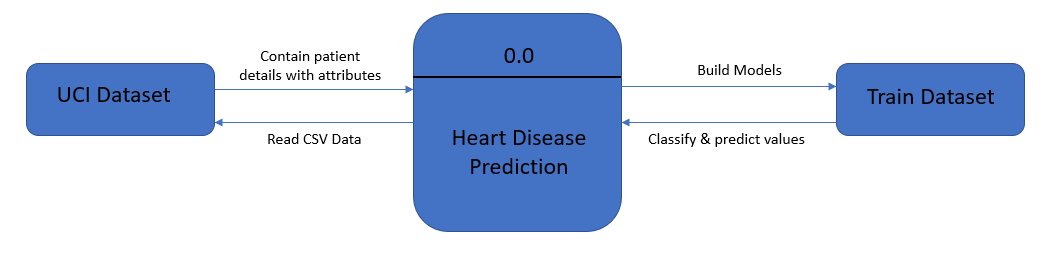
* Cloud hosted workspaces
* Connect data from any source
* Transform your data into features to train models and make predictions​
* Quality models built fast at scale​
* Deploy Models to Production in Seconds​
* Continuous Model monitoring​
* Security and Control



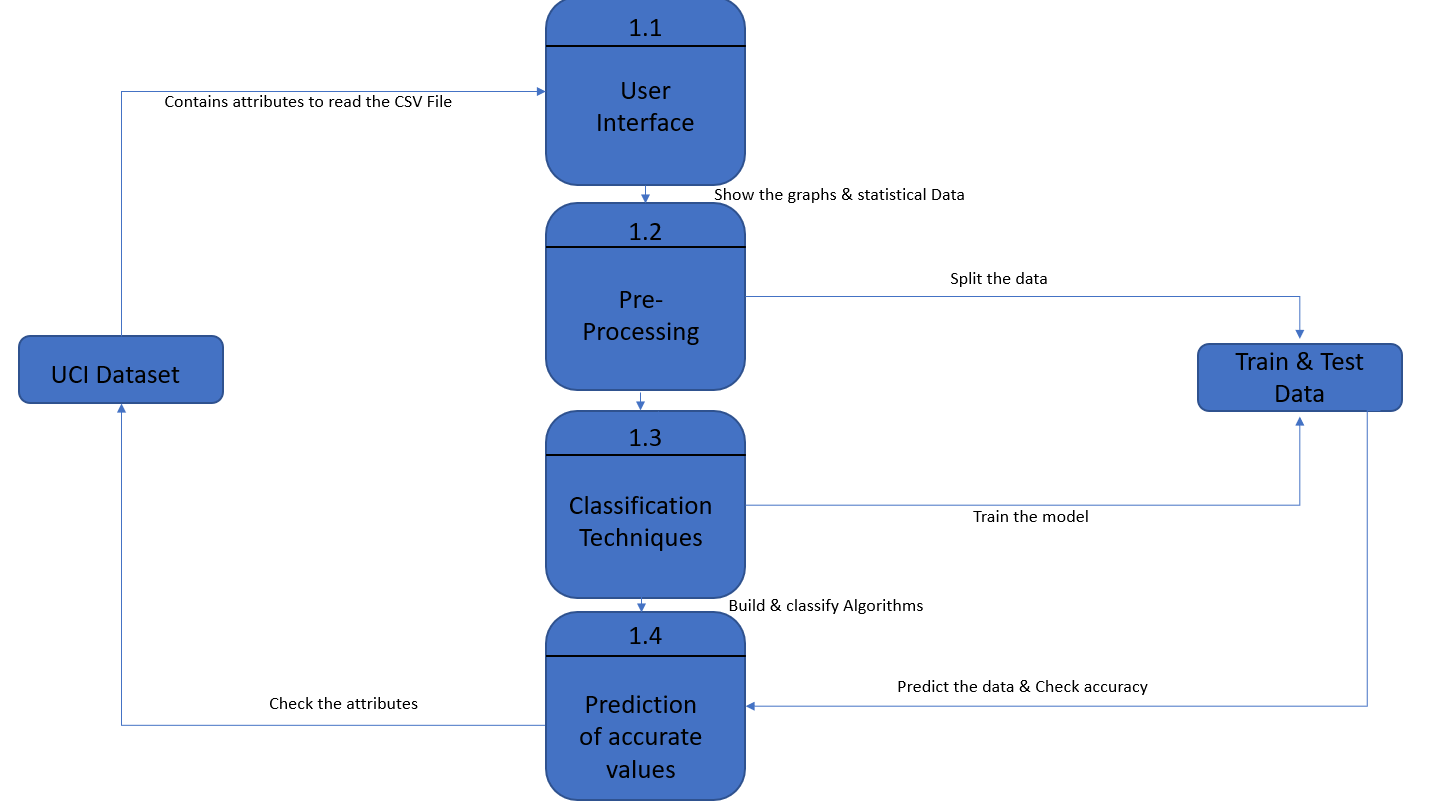
**Figure8: Katonic AI Cloud Workspace**

**Diagrams :**

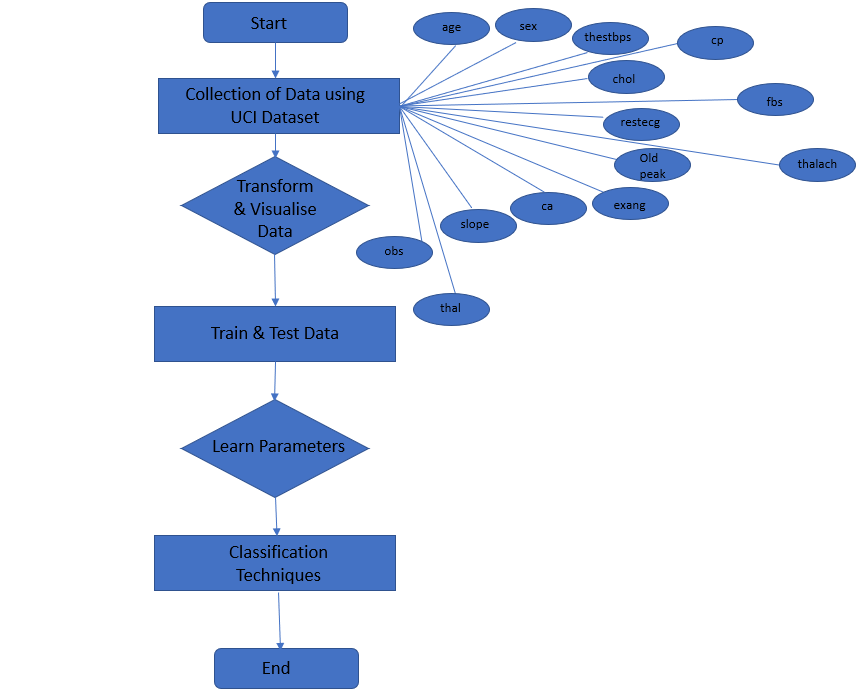
**1) DFD Context Level Diagram**



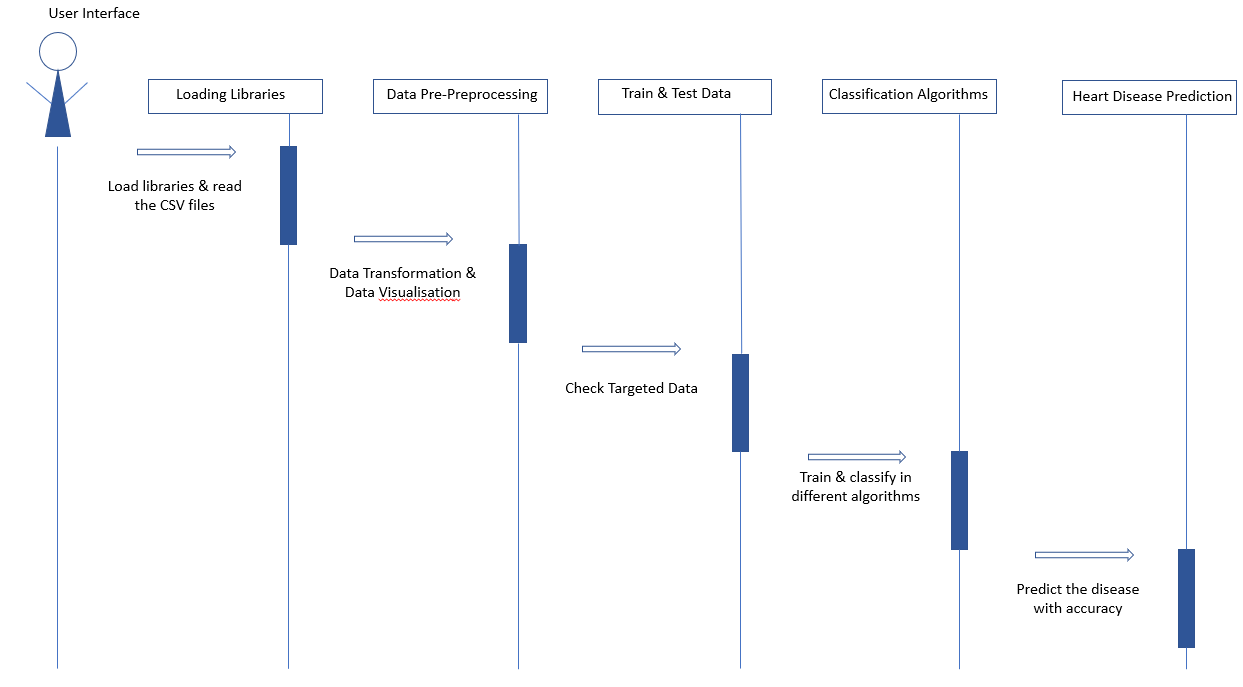
**2) DFD Level 1 Diagram**



**3) ER Diagram**



**4) Sequence Diagram**



**Testing:**

**1) System Test:**

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test.

System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

**2) White Box Testing:**

White Box Testing is a testing in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It has a purpose. It is used to test areas that cannot be reached from a black box level

**3) Black Box Testing:**

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a test in which the software under test is treated as a black box .You cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

**4) Unit Testing:**

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

**5) Integration Testing:**

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

**6) Acceptance Testing:**

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

**Test Cases:**

|  |  |  |
| --- | --- | --- |
| 1) | Input | CSV Data |
| Expected Output | To read CSV Data |
| Error | Error in file (file, "rt"): cannot open the connection  In addition: Warning message:  In file (file, "rt"): cannot open file 'Rtrial.csv': No such file or directory |

|  |  |  |
| --- | --- | --- |
| 2) | Input | ggplot |
| Expected Output | Bar plot |
| Error | Nil |

|  |  |  |
| --- | --- | --- |
| 3) | Input | corrplot |
| Expected Output | Correlational plots |
| Error | Warning message:  "Continuous x aesthetic -- did you forget aes(group=...)?" |

|  |  |  |
| --- | --- | --- |
| 4) | Input | AUC & Accuracy list |
| Expected Output | Need to show the accuracy between Model. |
| Error | Nil |

|  |  |  |
| --- | --- | --- |
| 5) | Input | Fit Model |
| Expected Output | Prediction |
| Error | "Prediction from a rank-deficient fit may be misleading “Warning message:  "package 'pROC' was built under R version 3.6.3"Type 'citation("pROC")' for a citation. |

**Technology Used**:

1. **Shiny web framework:** Shiny makes it incredibly easy to build interactive web applications with R. Automatic "reactive" binding between inputs and outputs and extensive prebuilt widgets make it possible to build beautiful, responsive, and powerful applications with minimal effort.
2. **Tidyverse**: The tidyverse is universe of packages, a collection of packages specially focused on data science, marked a milestone in R programming. In this post I am going to summarize very briefly the most essential to start in this world. The tidyverse grammar follows a common structure in all functions. The most essential thing is that the first argument is the object and then come the rest of the arguments. In addition, a set of verbs is provided to facilitate the use of the functions.
3. **Repr:** This is inspired by the python function repr and produces a short string representation of any R object that is suitable for logging and error messages. It is a generic so you can implement methods for custom S3 objects.
4. **ggplot:** ggplot2 package in R Programming Language also termed as Grammar of Graphics is a free, open-source, and easy-to-use visualization package widely used in [R](https://www.geeksforgeeks.org/introduction-to-r-programming-language/). It is the most powerful visualization package written by Hadley Wickham It includes several layers on which it is governed. The layers are as follows:

* Building Blocks of layers with the grammar of graphics
* Data: The element is the data set itself
* Aesthetics: The data is to map onto the Aesthetics attributes such as x-axis, y-axis, color, fill, size, labels, alpha, shape, line width, line type
* Geometrics: How our data being displayed using point, line, histogram, bar, boxplot
* Facets: It displays the subset of the data using Columns and rows
* Statistics: Binning, smoothing, descriptive, intermediate
* Coordinates: the space between data and display using Cartesian, fixed, polar, limits
* Themes: Non-data link

1. **corrplot:** R package **corrplot** provides a visual exploratory tool on correlation matrix that supports automatic variable reordering to help detect hidden patterns among variables.
2. **Confusion matrix:** A confusion matrix in R is a table that will categorize the predictions against the actual values. It includes two dimensions; among them one will indicate the predicted values and another one will represent the actual values. In most of the recourses, you could have seen the 2×2 matrix in R.
3. **R Programming**: A confusion matrix in R is a table that will categorize the predictions against the actual values. It includes two dimensions; among them one will indicate the predicted values and another one will represent the actual values. In most of the recourses, you could have seen the 2×2 matrix in R.
4. **R using Jupyter Notebook:** The [Jupyter Notebook](https://jupyter.org/" \t "_blank) is a Web application which permits to create live code in different languages. Usually, developers exploit the Jupyter Notebook to write code in Python. However, Jupyter also supports other programming languages, including Java, R, Julia, Matlab, Octave, Scheme, Processing, Scala and many others.
5. **Md files using R**:R Markdown is a file format for making dynamic documents with R. An R Markdown document is written in markdown (an easy-to-write plain text format) and contains chunks of embedded R code, like the document below. --- output: HTML document --- This is an R Markdown document.
6. **Comparative study with different algorithms**:  The three most popular machine learning algorithms—Logistic Regression, Support Vector Machine, and Random Forest—are examined in this project and contrasted with the decision tree and naïve bayes algorithm.

This study was carried out using a clinical dataset. Different evaluation techniques, including Confusion Matrix, Stratified K-fold Cross Validation, Accuracy, AUC, and ROC, have been utilised to assess the performance. The accuracy and AUC values have been validated using the K-Fold Cross-validation approach to verify the results. The SMOTE Algorithm was used to balance the dataset because it had a class imbalance, and performance analysis was done on both sets of data. The outcomes demonstrate that while training the balanced dataset, the accuracy scores of all the models have grown.

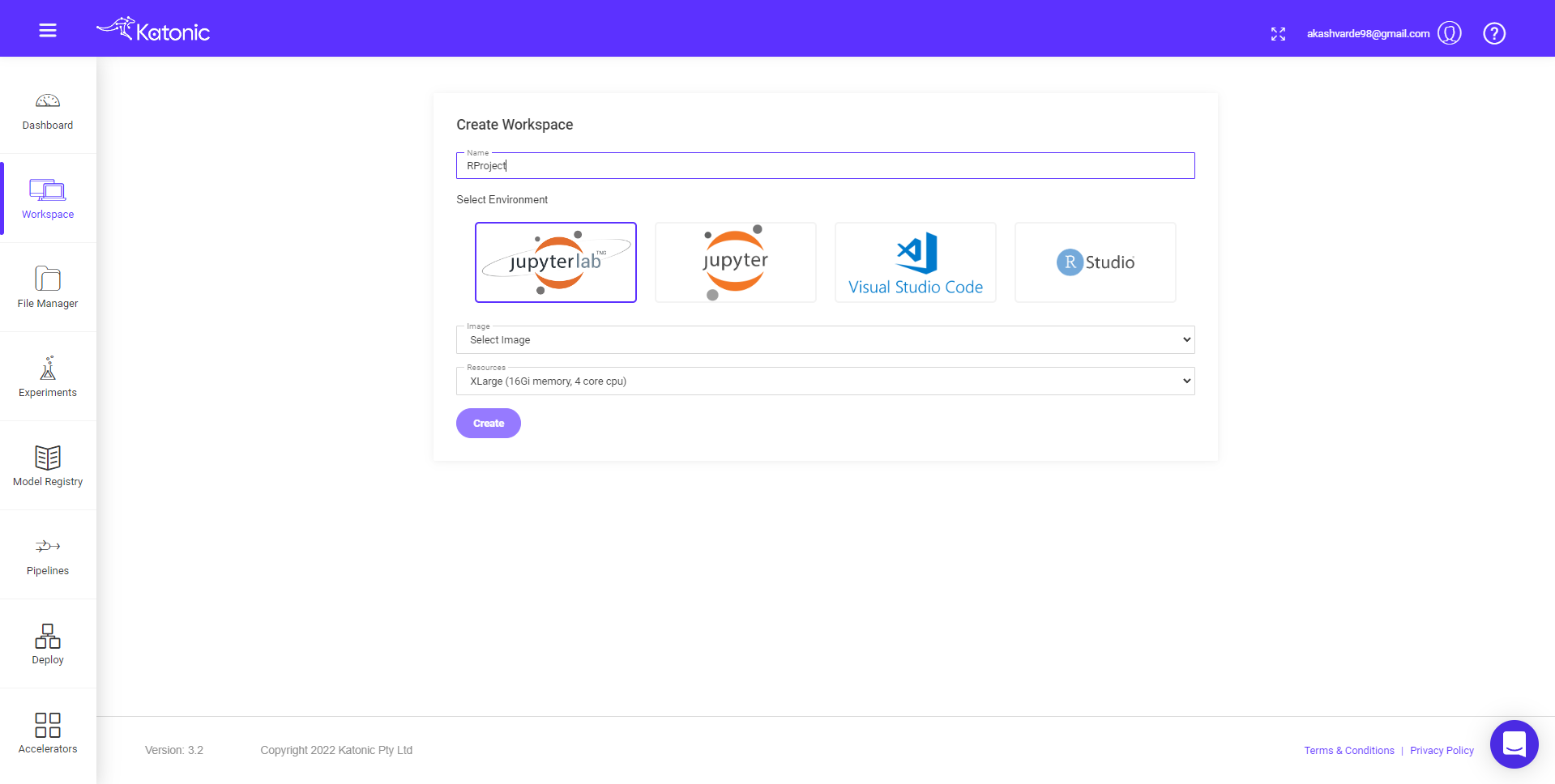
**Output Screenshots:**

**We run our R project is on Katonic AI cloud workspace. This are the below steps to create workspace and run this program:**

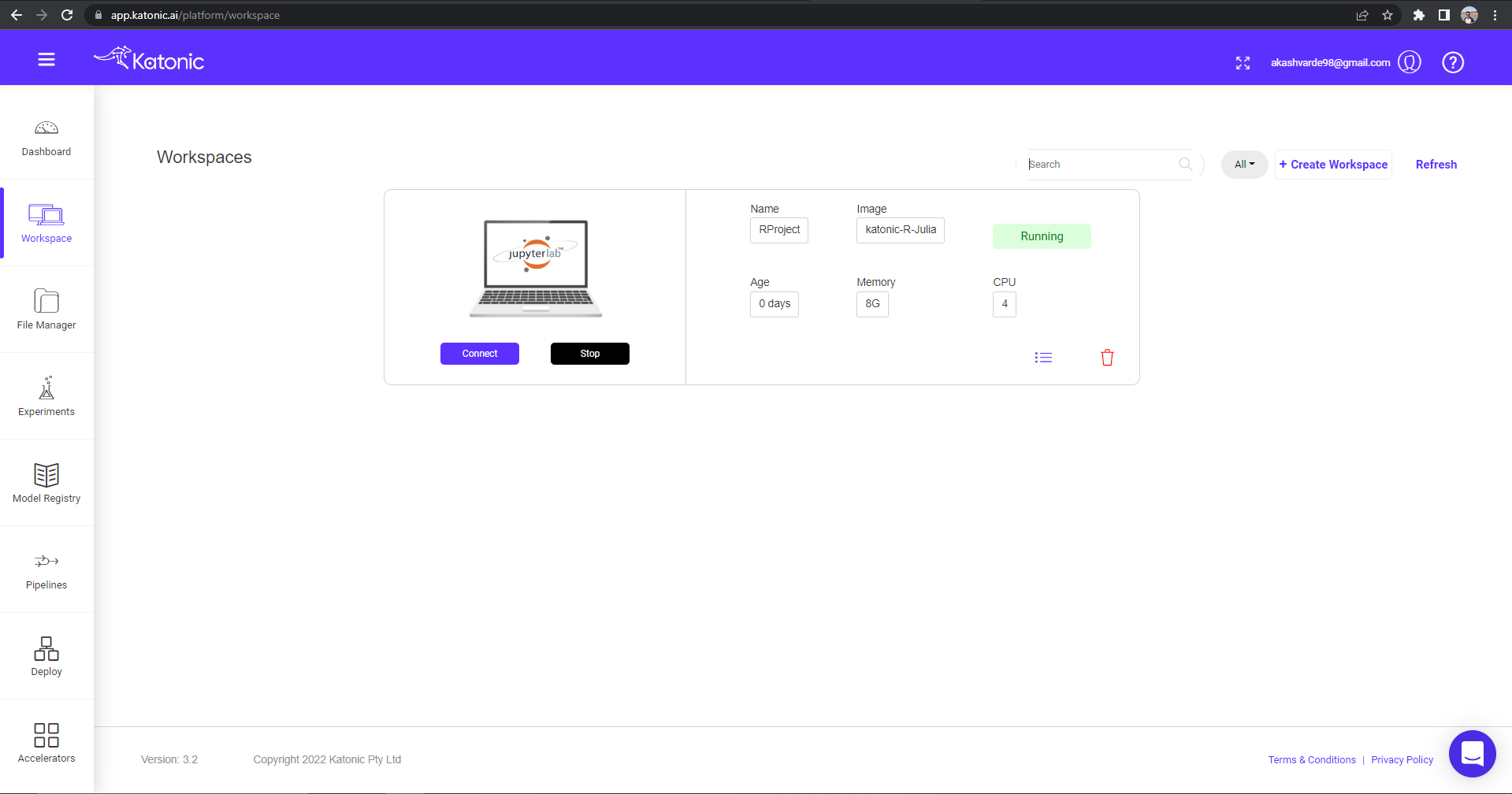
1) Sign up then Login by the Mail ID.

Create workspace and choose R Studio for project.

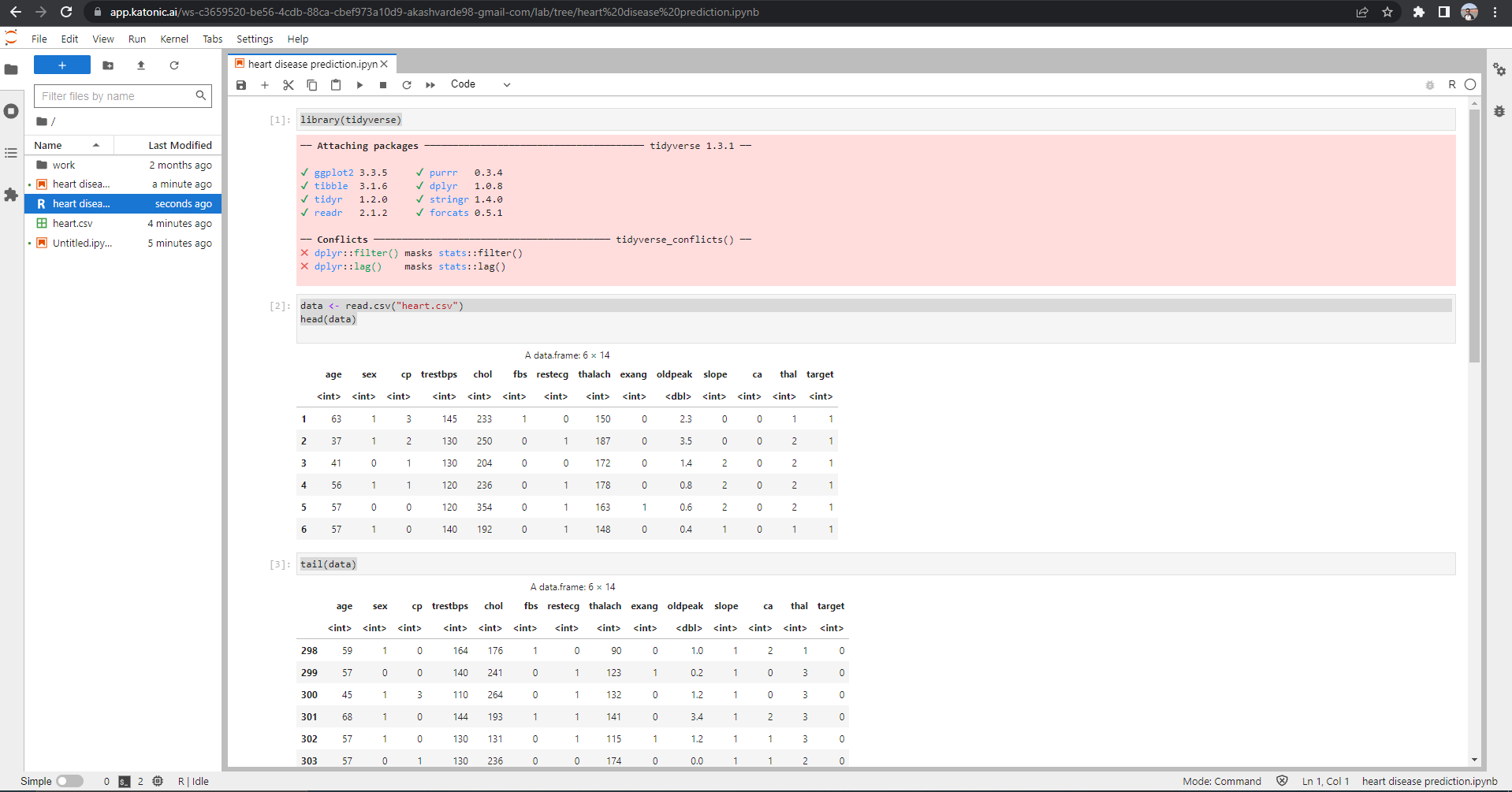
Select the resources.



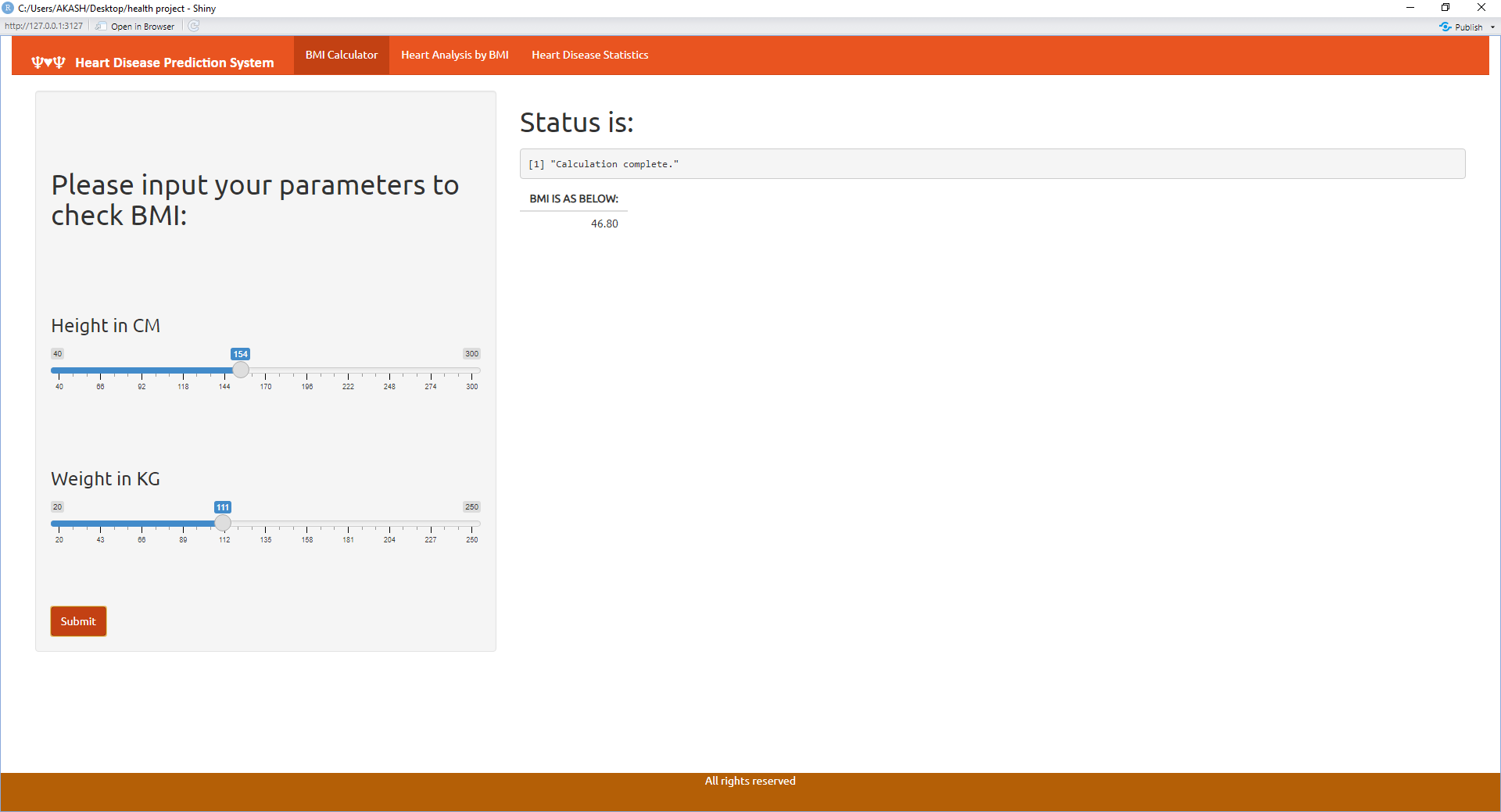
2) Connect to R Project using jupyter lab Workspace



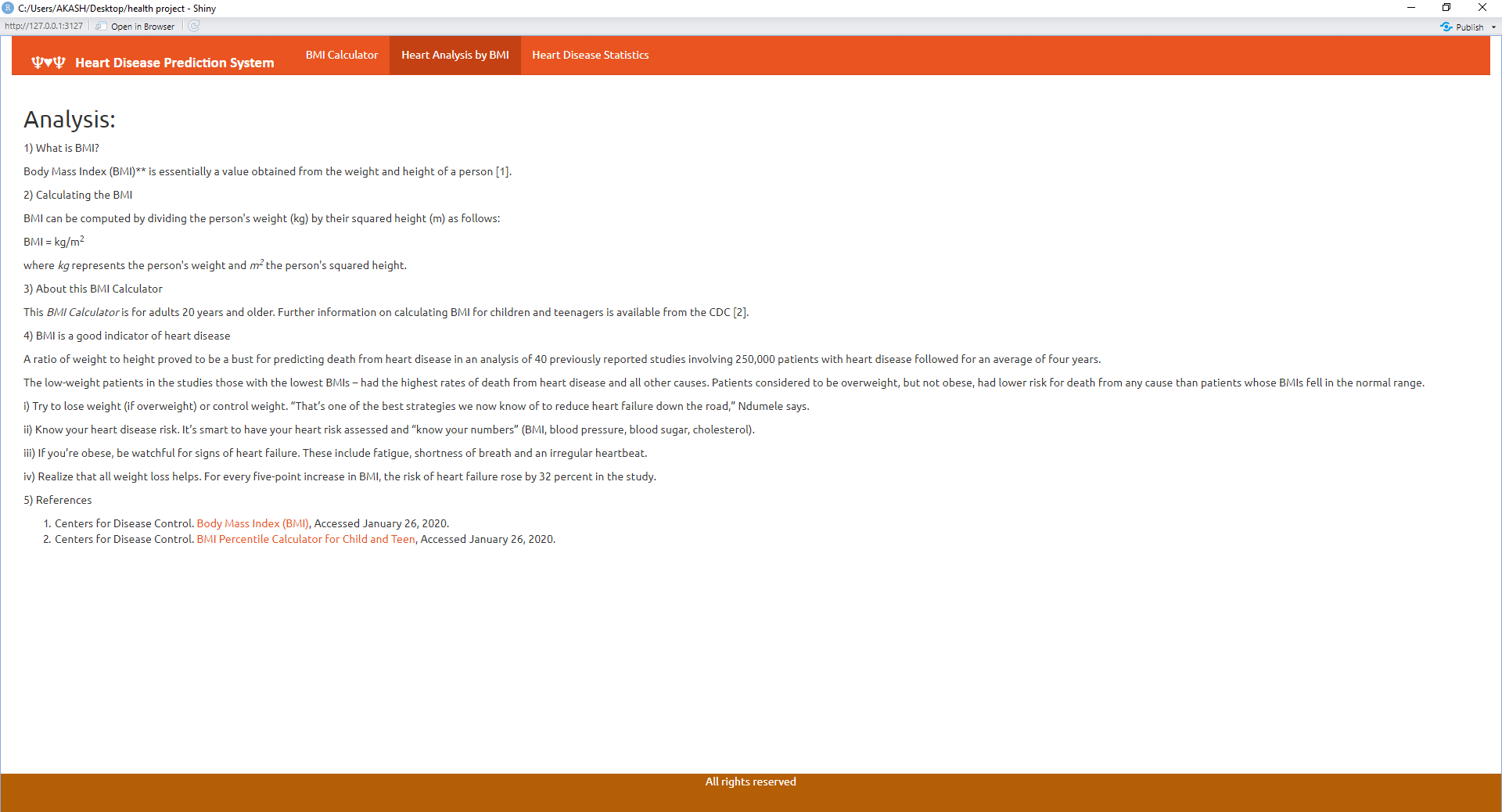
3) We run our project on below workspace.



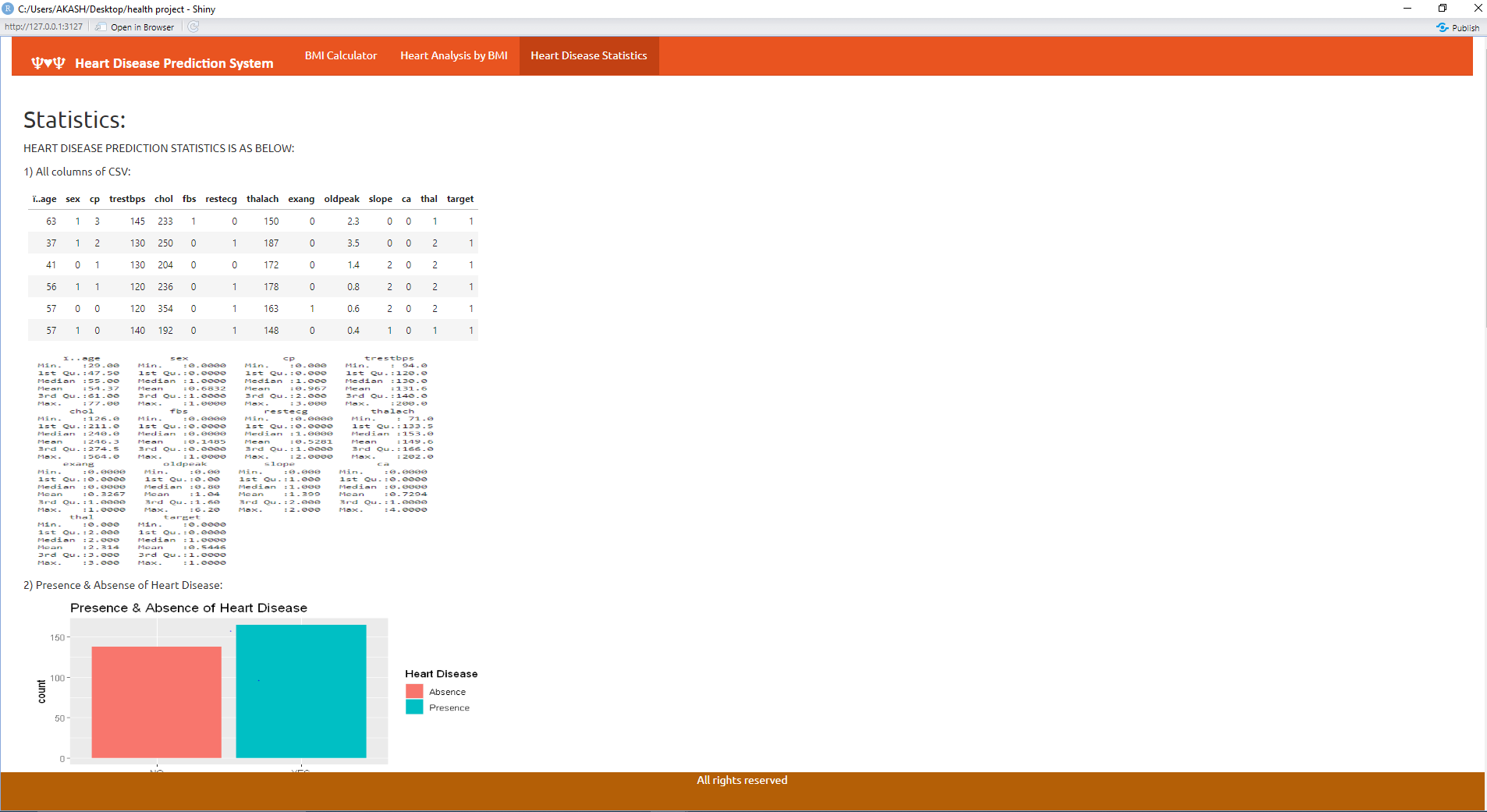
4)User interface of Heart Disease Prediction to check BMI.



5) Heart Disease Analysis Information

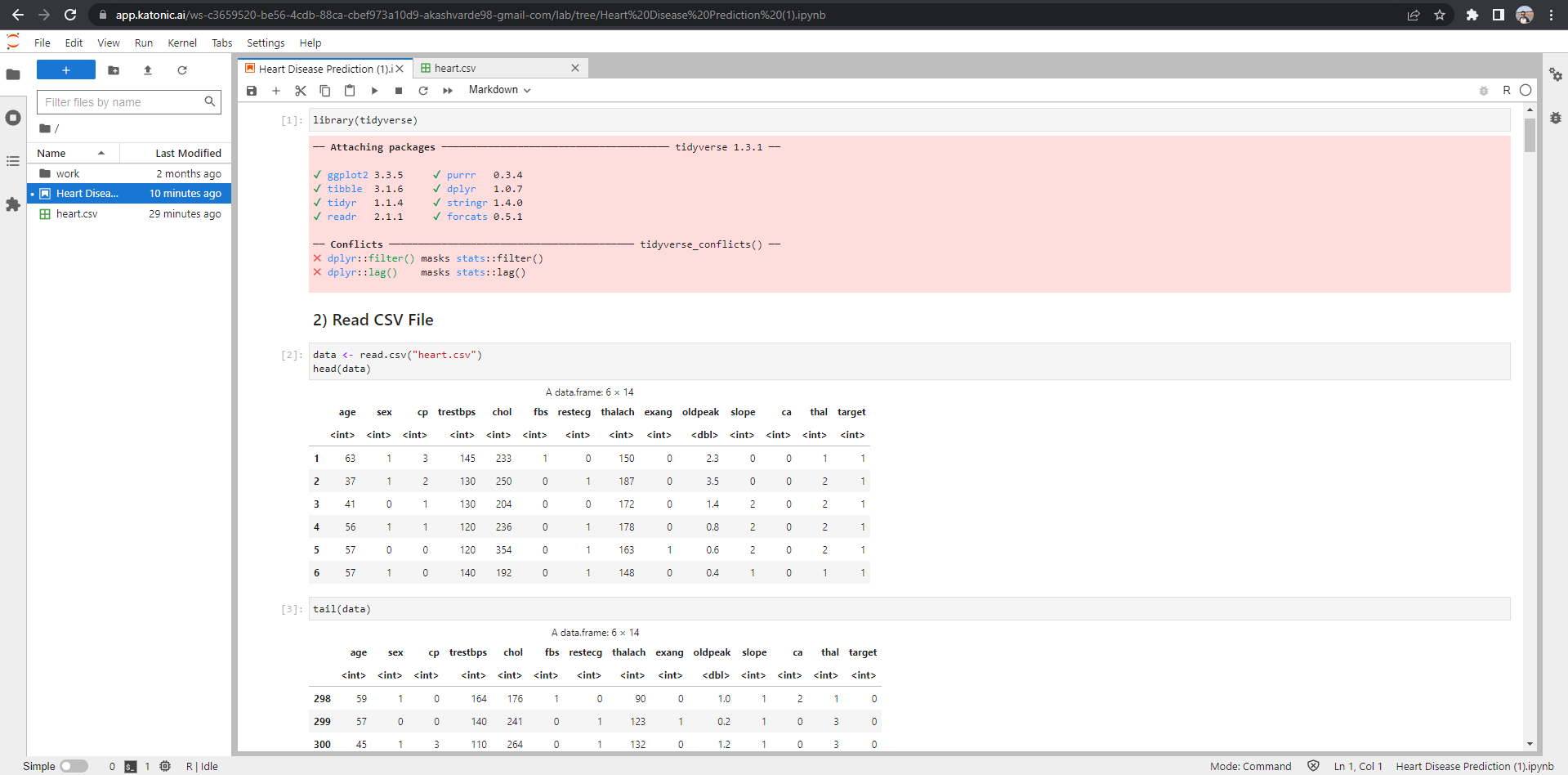


6)Heart Disease Prediction Statistics

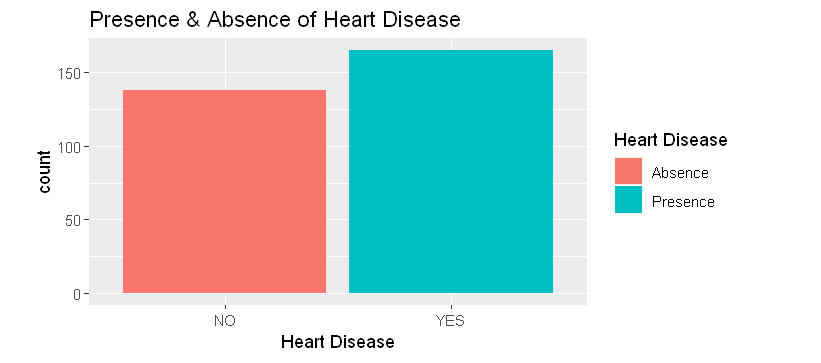


**Complete Heart Disease Prediction Analysis in Details:**

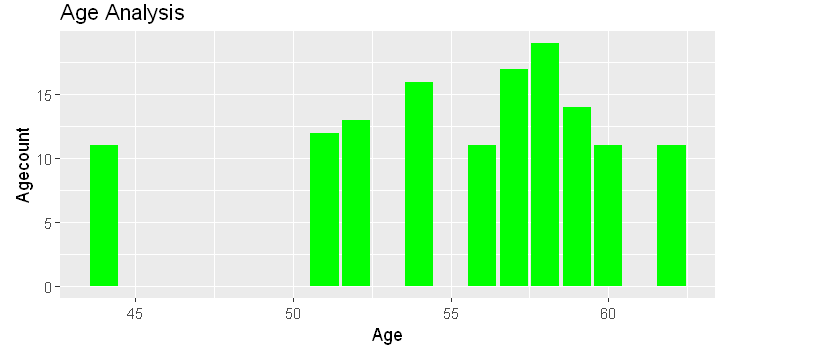
7) Load Libraries and Read CSV Data



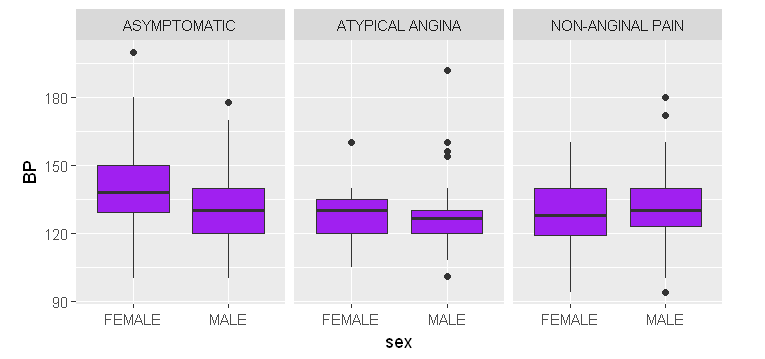
8) Data Visualisation of heart disease



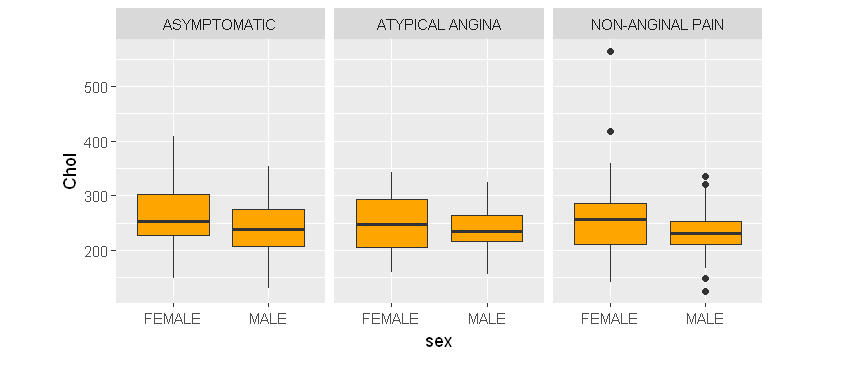
9) Count the frequency of values of age



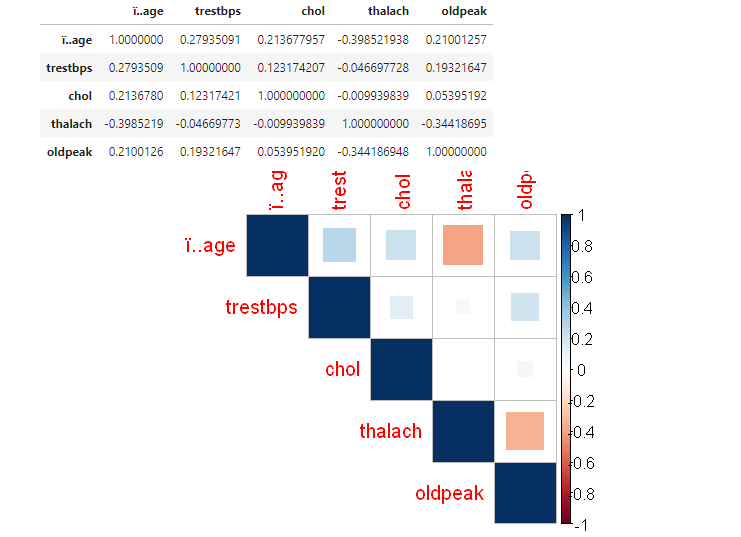
10) Compare blood pressure across the chest pain



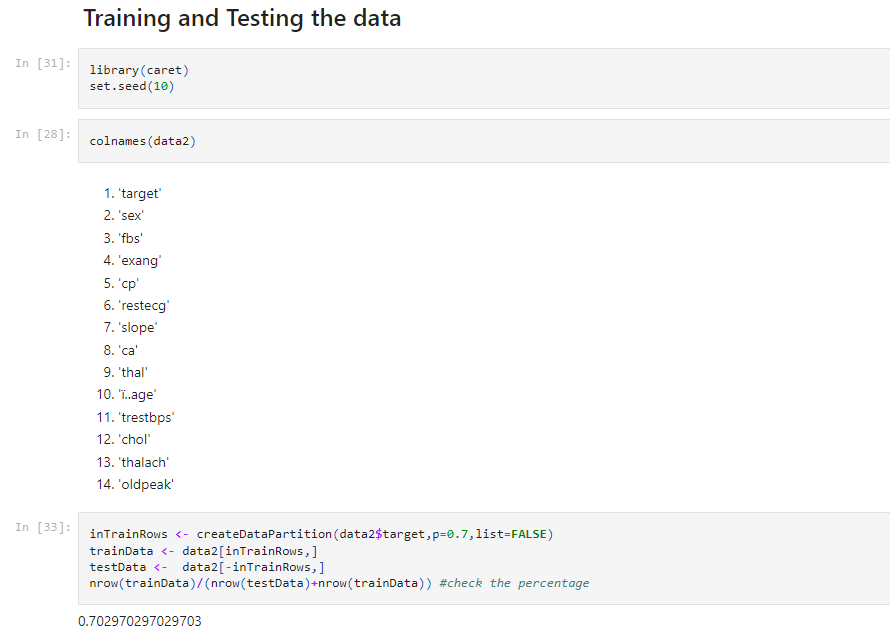
11) Various symptoms using Cholesterol test:



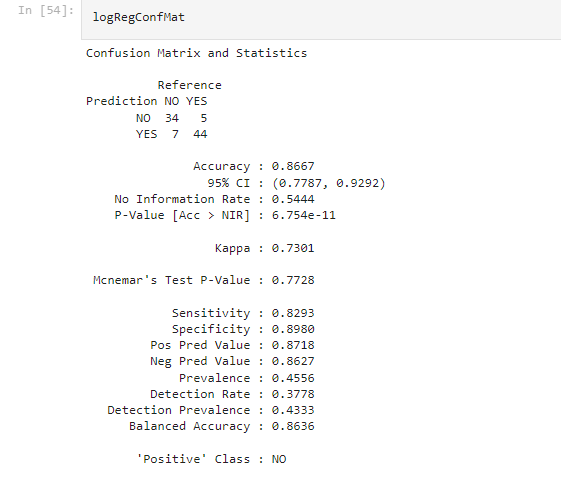
12) Correlation between all symptoms with there columns:



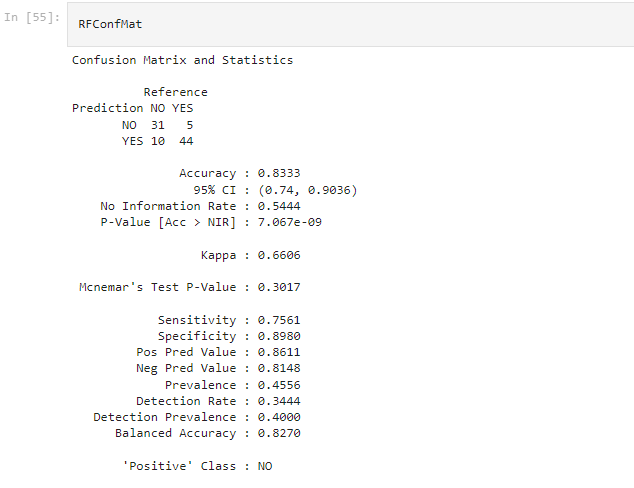
13) Train & Test Data to check percentage



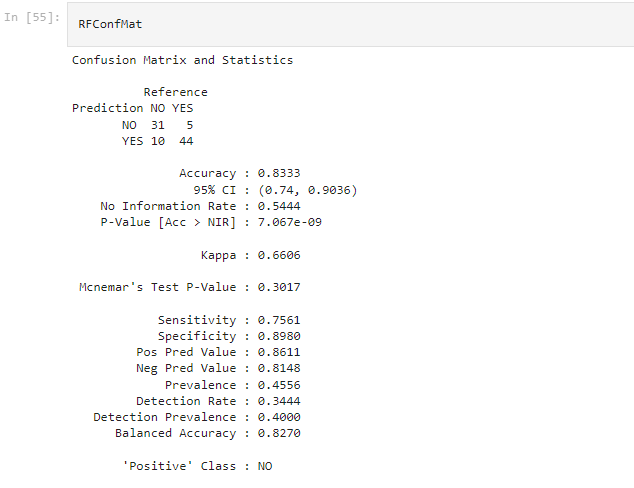
14) Confusion Matrix of Logistics Regression



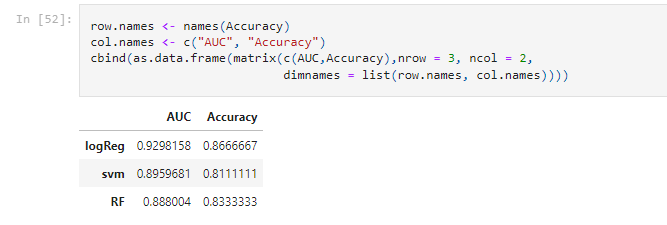
15) Confusion Matrix of Random Forest



16) Confusion Matrix of Support Vector Machine



17) Comparison of Accuracy between Models



* A comparison of the area under the ROC and the accuracy of the model predictions shows that logistic regression performs best with 92%.
* Tree-based methods shows low accuracy.

**Advantages:**

1. Increased accuracy for effective heart

disease diagnosis.

2. Handles roughest(enormous) amount of

data using random forest algorithm and

feature selection.

3. Reduce the time complexity of doctors.

4. Cost effective for patients.

1. Increased accuracy for effective heart

disease diagnosis.

2. Handles roughest(enormous) amount of

data using random forest algorithm and

feature selection.

3. Reduce the time complexity of doctors.

4. Cost effective for patients.

Increased accuracy for effective heart

disease diagnosis.

2. Handles roughest(enormous) amount of

data using random forest algorithm and

feature selection.

3. Reduce the time complexity of doctors.

4. Cost effective for patients.

Increased accuracy for effective heart

disease diagnosis.

2. Handles roughest(enormous) amount of

data using random forest algorithm and

feature selection.

3. Reduce the time complexity of doctors.

4. Cost effective for patients.

Increased accuracy for effective heart

disease diagnosis.

2. Handles roughest(enormous) amount of

data using random forest algorithm and

feature selection.

3. Reduce the time complexity of doctors.

4. Cost effective for patients.

Increased accuracy for effective heart

disease diagnosis.

2. Handles roughest(enormous) amount of

data using random forest algorithm and

feature selection.

3. Reduce the time complexity of doctors.

4. Cost effective for patients.

Increased accuracy for effective heart

disease diagnosis.

2. Handles roughest(enormous) amount of

data using random forest algorithm and

feature selection.

3. Reduce the time complexity of doctors.

4. Cost effective for patients.

1. Greater precision for efficient heart diagnostics
2. Uses feature selection to handle the most quantity of data.
3. Reduce the complexity of time
4. Very useful in case of emergency.

**Limitations:**

1. Assess the scope of the data, especially over time, so your model can avoid the seasonality trap.
2. Check for missing values, identify them, and assess their impact on the overall analysis.
3. Data mining techniques does not help to provide effective decision making.
4. Cannot handle enormous datasets for patient records

**Future Scope**:

1. The suggested system is GUI-based, approachable, scalable, trustworthy, and extendable. By offering early diagnosis.
2. The proposed operating paradigm can also aid in lowering treatment costs. In addition to being a soft diagnostic tool for doctors and cardiologists, the model can also be used as a training tool for medical students.
3. The scalability and accuracy of this prediction system can both be improved in a number of different ways.
4. It is implemented with a generalised framework that utilise while analysing various data sets in the future.

**Bibliography:**

1. M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, “Deep learning approach for active classification of electrocardiogram signals,” Information Sciences, vol. 345, pp. 340–354, 2016.View at: [Publisher Site](https://doi.org/10.1016/j.ins.2016.01.082) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Deep%20learning%20approach%20for%20active%20classification%20of%20electrocardiogram%20signals&author=M.%20M.%20A.%20Rahhal&author=Y.%20Bazi&author=H.%20Alhichri&author=N.%20Alajlan&author=F.%20Melgani&author=&author=R.%20R.%20Yager&publication_year=2016)
2. G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, “A machine learning system to improve heart failure patient assistance,” IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 6, pp. 1750–1756, 2014.View at: [Publisher Site](https://doi.org/10.1109/jbhi.2014.2337752) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=A%20machine%20learning%20system%20to%20improve%20heart%20failure%20patient%20assistance&author=G.%20Guidi&author=M.%20C.%20Pettenati&author=P.%20Melillo&author=&author=E.%20Iadanza&publication_year=2014)
3. American Heart Association. Heart disease and stroke statistics 2017 at-a-glance. Geraadpleegdvan: <https://healthmetrics.heart.org/wpcontent/uploads/2017/06/Heart-Disease-and-Stroke-Statistics-2017-ucm_491265.pdf>.
4. G. Parthiban and S. K. Srivatsa, “Applying machine learning methods in diagnosing heart disease for diabetic patients,” International Journal of Applied Information Systems, vol. 3, no. 7, pp. 25–30, 2012.View at: [Publisher Site](https://doi.org/10.5120/ijais12-450593) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Applying%20machine%20learning%20methods%20in%20diagnosing%20heart%20disease%20for%20diabetic%20patients&author=G.%20Parthiban%20&author=S.%20K.%20Srivatsa&publication_year=2012)