VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgavi-590 018, Karnataka, India

An Internship Report
On

HEART DISEASE PREDICTION

Submitted in Partial Fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted By

ANNAPUREDDY PRANATHI   **1SJ18CS005**

Carried out at
QUANT MASTERS

(#812, 6th cross 3rd main, Rajajinagar, Bengaluru - 560021)

Under the guidance of

Internal Guide	External Guide
Mr. Apoorva S	**Mr.Shashank**
Assistant Professor,	**Technical Lead,**
Dept. of CSE, SJCIT.	**QUANT MASTERS**

S J C INSTITUTE OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CHIKKABALLAPUR-562101

2021-2022

CERTIFICATE

This is to certify that the Internship work entitled **"HEART DISEASE PREDICTION"** carried out by **ANNAPUREDDY PRANATHI** bearing USN:1S18CS005 a bonafide student of Sri Jagadguru Chandrashekaranatha Institute of Technology in partial fulfilment for the award of **Bachelor of Engineering in Computer Science and Engineering** of Visvesvaraya Technological University, Belgaum during the year **2021-22**. It is certificated that all corrections / suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The Internship report has been approved as it satisfies the academic requirements in respect of Internship work prescribed for the said Degree.

3|6|22

....
Signature of Guide
Mr. Apoorva S
Assistant Professor
Dept. of CSE,SJCIT

....
Signature of HOD
Dr. Manjunath Kumar B H
Professor & HOD,
Dept. of CSE,SJCIT

....
Signature of Principal
Dr. G T Raju
Principal, SJCIT,
Chickballapur

External Examiners:
Name of the Examiners

Signature with Date

1.

2.

COMPANY CERTIFICATE

Quant Masters

Certificate of Excellence

Annapureddy Pranathi

is hereby awarded this certificate of excellence for the successful completion

of **Artificial Intelligence and Machine Learning** Internship conducted by

Quant Masters Technologies Pvt. Ltd.

from 01/09/2021 to 28/09/2021.

HIMANSHU SHARMA
Founder & Managing Director
Quant Masters Technologies Pvt. Ltd.

QMIEXCCERT_00217

Date of Issue: 23-10-2021

Certificate ID:_____

DECLARATION

I, **ANNAPUREDDY PRANATHI**, student of VIII semester B.E in Computer science & Engineering at S J C Institute of Technology, Chickballapur, hereby declare that the Internship work entitled " HEART DISEASE PREDICTION" has been independently carried out by me under the supervision of **Apoorva S** Assistant Professor, and the coordinator **Swetha T** Assistant Professor, submitted in partial fulfillment of the course requirement for the award of degree in **Bachelor of Engineering** in **Computer Science & Engineering** of **Visveswaraya Technological University, Belgavi** during the year 2021-2022. I further declare that the report has not been submitted to any other University for the award of any other degree.

PLACE: CHICKBALLAPUR   ANNAPUREDDY PRANATHI

DATE:13/05/2022   USN:1SJ18CS005

ABSTRACT

Heart plays significant role in living organisms. Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardio vascular disease is a critical challenge in the area of clinical data analysis. Diagnosis and prediction of heartrelated diseases requires more precision, perfection and correctness because a little mistake can cause fatigue problem or death of the person, there are numerous death cases related to heart and their counting is increasing exponentially day by day. Machine learning has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. Various studies give only a glimpse into predicting heart disease with ML techniques. Here, we design a model that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of heart disease. There are many to do accomplish this task successfully, but how effective are they? Our main aim is to produce an enhanced performance level with the good accuracy level through the prediction model for heart disease with the SVM, KNN, Naïve Bayes, logistic regression & Random Forest algorithms. It is estimated that on an average about 17 million people die of cardiovascular diseases each year, which is about one third of total deaths across the globe. In this proposed project we designed a model to detect and predict the accuracy of heart disease. This system can provide most of the important features required to detect heart disease with different algorithms. As we know the population is been increased day by day the diseases of the people are increasing day by day, so with that we have to upgrade with the technology constantly, it is becoming easy to track the behaviour and pattern diseases and get cure at the early stages. To come up with the solution one can make use of technologies with the increase of machine learning so it becomes feasible to automate this process and to save someone's life by detecting the disease at an early stages. Initially, we will collect the data set by users and classify it as trained and testing dataset using different type's algorithm and decision trees. By using the feasible algorithm, we can analyze the larger data-set and user provided current data set. Then augment the accuracy of the result data. Proceeded with the application of processing of some of the attributes provided which can find weather the user is having heart disease in viewing the graphical model of data visualization. The performance of the techniques is gauged based on accuracy, sensitivity, and specificity, precision. The results are indicated concerning the best accuracy for Random Forest is unit 80% respectively.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER - 1

COMPANY PROFILE

1.1 History of the Organization:

Himanshu Sharma, a native of Bengaluru, founded the company "Quant masters training services" in 2019 with just 2 employees. The services offered by the company aimed at offline providing placement training to the undergraduates ranging from Quantitative, Logical, Verbal to HR interview preparation.

In the early 2020s, the Shift from offline to online training took place due to pandemic with 1 st batch going with just 30 students.

In 2021, Quant masters not only achieved a place among MSMEs and became "Quant masters Technologies private limited" but has trained over 10000+ students using the online platform with highly qualified educators and mentors guiding them throughout.

Quant masters when started had a reach only to the Students of Bengaluru region. But, with its dedicated training and quality services, over 700+ students enroll in our batches every month from all over the country. The training helps them in getting placed and likewise many students have brought laurels to Quant Masters.

A brief profile of the founder

Himanshu Sharma

Founder & Director, Quant Masters

Cleared CDS, AFCAT, RBI GRADE B, CAFs, IB, AMCAT (99.99%), CO-CUBES

Recommended as Pilot in Indian Air Force

Oracle Certified Java Programmer- OCJP (95%)

Former Software Developer-Grade 4 @NTT DATA

1.1.1 Objectives

- The essential objective of QUANT MASTERS is to improve the quality of training and enhance the learning process.
- Most importantly to create engaging and effective learning experiences and provide a variety of technological information, ideas to encourage curiosity, stimulate self-confidence through the knowledge and develop practical skills.

1.1.2 Operation of the Organization

Our mission is to make learning truly interesting and make it easier as well as more affordable for the students to prepare for their placements or competitive examinations and help them make a perfect start to their career. There are about 20 people working in the organization working as intern guides in various platform and trained above 700+ students enroll in our batches every month from all over the country and one of the corporate institutions for providing internships.

1.2 Major Milestones

Over the years, we have been a team of 40+ members as compared to when we started with just two! There were a lot of highs and lows, yeses and noes but we never stopped focussing on our goal.

We always believed that fundamentals of IT technologies can be used to empower the needful and help them with a great push in their careers. QUANT MASTERS provides wide range of IT services to help the students and help in company growth providing betterment of society.

The services offered by the company aimed at offline providing placement training to the undergraduates ranging from Quantitative, Logical, Verbal to HR interview preparation

- Within 1 year more than 2000+ students of QUANTMASTERS have been placed in Service, Product and Technology based companies like TATA CONSULTANCY SERVICES, ACCENTURE ,INFOSYS ,CAPGEMINI LTI etc.

1.2 Structure of the Organization

Team:

Himanshu sharma Founder and Director

Deepshikha Raina HR Operations Head

Dinesh Gosai Soft Skill trainer

Ritu Dhudoria Verbal Ability trainer

Harshitha Aliveli Aptitude and Logical Trainer

Anudeep MP Aptitude and Logical Trainer

On-going projects: We start a new placement training batch every 1.5 months. Currently we are working towards giving quality training cum internships to the students and give them the practical implications of the related projects. The training provided by us is also helpful to various students preparing for competitive exams from different branches- Engineering, Humanity, Commerce, Arts, Management etc. We will soon be launching our services with regards to various new technological advancements and certification courses.

1.4 Services Offered

- Quantitative Aptitude
- Technical Training
- Verbal Aptitude
- Logical training
- Soft skills/ Communication Skills
- Resume Building
- LinkedIn Networking
- AI and ML internship
- GD Preparation

CHAPTER – 2

ABOUT THE DEPARTMENT

2.1 Specific Functionalities of the Department

The department has around 15members that specialize in a variety of fields including IOT, skill development, ML, AI, Placement Training. I worked under the Machine Learning domain, which is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence

2.2 Process Adopted

SDLC is a process followed for a software project, within a software organization. It consists of a detailed plan describing how to develop, maintain, replace and alter or enhance specific software. The life cycle defines a methodology for improving the quality of software and the overall development process.

 SDLC process as following mentioned steps:

- Planning
- Defining
- Designing
- Building
- Testing
- Deployment
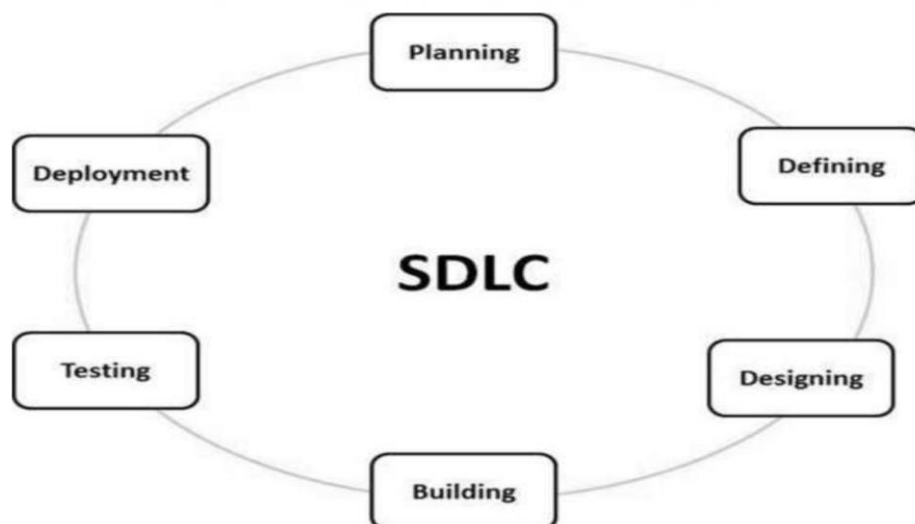
Figure 2.2: Process adopted: SDLC

2.3 Testing

The various testing techniques used by the department can be summarized as follows:

1.Functionality Testing of a Website: it is a process that includes several testing parameters like user interface, APIs, database testing, security testing, client and server testing and basic website functionalities. Functional testing is very convenient and it allows users to perform both manual and automated testing. It is performed to test the functionalities of each feature on the website.

2.Usability Testing: This type of testing includes testing the site navigations and contents of the website.

3.Interface Testing: Three areas to be tested here are Application,Web and Database Server.

4.Database Testing: Database is one critical component of your web application and stress must be laid to test it thoroughly Testing activities will include Test if any errors are shown while executing queries, Data Integrity is maintained while creating, updating or deleting data in database, Check response time of queries and fine tune them if necessary, Test data retrieved from your database is shown accurately in your web application.

5. Compatibility testing: Compatibility tests ensures that your web application displays correctly across different devices. This would include-Browser Compatibility Test: Same website in different browsers will display differently. You need to test if your web application is being displayed correctly across browsers, JavaScript, AJAX and authentication is working fine.

6. Pipeline testing: After compatibility testing it is the time to test all the micro services in pipeline together to check their compatibility and message passing.

Thus all the services/functionalities are kept in pipeline and tested together. Afterwards whole pipeline is pushed in the deployment server.

2.4 Structure of the Department

The structure of the organization is descripted in the following figure:
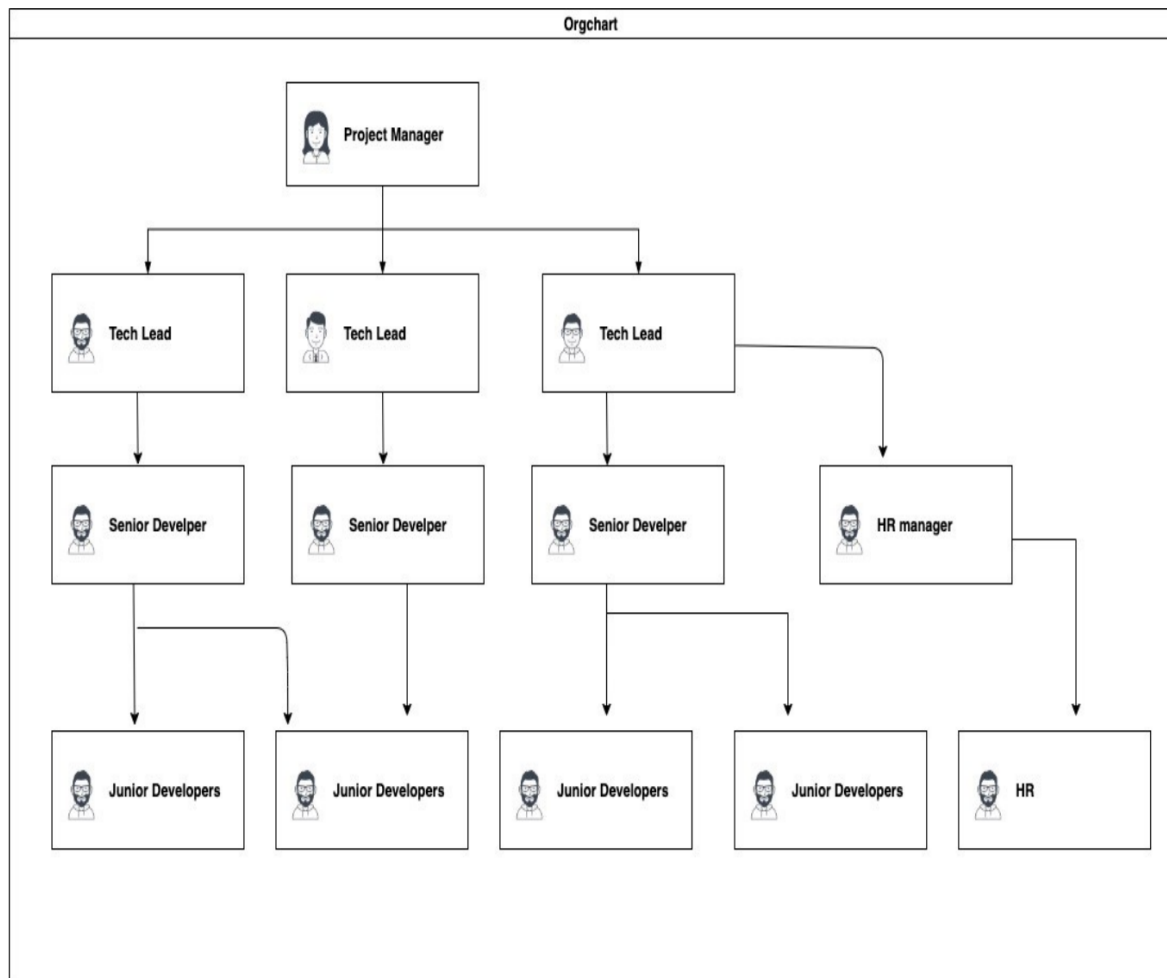
Figure 2.3 Department Structure

2.5 Roles and Responsibilities of Individuals

The different roles and responsibilities of individuals are:

1.Project Manager: Project Managers play the lead role in planning, executing, monitoring, controlling, and closing projects. They're expected to deliver a project on time, within the budget, and brief while keeping everyone in the know and happy.

2.Tech Leads: Technical Lead as the name states is solely responsible for leading a development team. The is not easy. They have to lead a team. Technical Lead is the one who actually creates a technical vision in order to turn it into reality with the help of the team.

3.HR Manager: The Human Resource Manager will lead and direct the routine functions of the Human Resources (HR) department including hiring and interviewing staff, administering pay, benefits, and leave, and enforcing company policies and practices.

4.Senior Developer: Develops software solutions by studying information needs, conferring with users, studying systems flow, data usage, and work processes; investigating problem areas; and following the software development lifecycle. A senior developer may manage a team of developers and will be expected to encourage creativity and efficiency throughout complex digital projects. Due to the pressurised nature of the role, a robust and organised approach to the work is needed to produce the best solutions.

5.Junior Developer: Junior Software Developers are entry-level software developers that assist the development team with all aspects of software design and coding. Their primary role is to learn the codebase, attend design meetings, write basic code, fix bugs, and assist the Development Manager in all design-related tasks.

CHAPTER – 3

TASK PERFORMED

3.1 Introduction

In this project author is evaluating performance of various classification/prediction algorithms such as SVM, Naïve Bayes, and Logistic Regression etc to predict heart disease. All this algorithms are good in prediction but accuracy is not good enough. To get better prediction accuracy author is combining two classification algorithms such as Linear Model and Random Forest to build new algorithm called Hybrid Machine Learning to get better prediction accuracy of heart dataset. Hybrid algorithm will form up by using Voting classifier, Internally Voting classifier will build up using Linear Model and Random Forest and while classification voting algorithm will evaluate prediction accuracy of both algorithms and vote for that algorithm which gives better accuracy. So by using hybrid model always we will have better prediction accuracy algorithm which helps in better prediction of heart disease.

It is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors. Various techniques in data mining and neural networks have been employed to find out the severity of heart disease among humans. The severity of the disease is classified based on various methods like K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic algorithm (GA), and Naive Bayes (NB) .The nature of heart disease is complex and hence, the disease must be handled carefully. Not doing so may affect the heart or cause premature death. The perspective of medical science and data mining are used for discovering various sorts of metabolic syndromes. Data mining with classification plays a significant role in the prediction of heart disease and data investigation.

3.2 Problem Statement

There is ample related work in the fields directly related to this paper. ANN has been introduced to produce the highest accuracy prediction in the medical field. The back propagation multilayer perception (MLP) of ANN is used to predict heart disease. The obtained results are compared with the results of existing models within the same domain and found to be improved

3.3 Technology Used

- Goggle Colab/Jupiter Notebook

- Python Programming Language

- Different Python Libraries

CHAPTER – 4

REFLECTION NOTES

4.1 Experience

The internship has been a really useful experience for me that I can learn a lot of new knowledge that will definitely be useful for my future study. I'm grateful that my assignments have a lot of variety instead of just focusing on a specific area. This allows me to be able to learn more and also challenge myself to overcome many different kinds of difficulties encountered during my internship. Having many assignments also required me to manage my work time efficiently prioritizing the urgent task **.**

Some tasks require me to do research with less available online documentation other task requires me to make attempts on works that I have never experienced before just by learning from documentations. Although the task may be difficult and overwhelming sometimes, I'm really excited to push my skills to the limit and carry out those task assigned to me.

Beside technical skills, I also observed and learned a lot of soft skills from my supervisors and my co-workers such as professional communication and team work. I have also learned a lot from my supervisor who's always willing to help me when I face difficulties and also willing to share a lot of his knowledge and wisdom to me from his post experience.

My internship experience has definitely improved my hard skills in IT and sharpen my soft skills a lot more than I expected I have shaped a better mind set in me and motivated me to keep on exploring and challenging myself in the world of information technology.

4.2 Technical Outcomes

• Understand a wide variety of learning algorithms

• Understand how to evaluate models generated from data.

• Apply, the algorithms to real problems

• Optimize the models learned and report on the expectancy accuracy that can be achieved by applying the models.

4.2.1 System Requirement Specification

Hardware Requirements

 PROCESSOR : Intel i5 RAM : 4GB

 HARD DISK : 16GB

Software Requirements

OPERATING SYSTEM : Linux/Windows

BACK-END : Python 3

OTHER BACKEND LIBRARIES : matplotlib, numpy, pandas, sklearn, Seaborn

4.3 System Analysis and Design

4.3.1 Existing System

The traditional detection method mainly depends on the doctor's vision of treating the patient and his level of experience, which usually are delayed, inaccurate and not in-time. After following these methods, it may take time for diagnosing the records and giving the summary and then treating the patient.

4.3.2   Disadvantages of the Existing System

There are few disadvantages identified in the existing system and are defined below:

     i.   Inaccurate results

     ii.   Time complexity is more

4.3.3   Proposed System

In this paper author is evaluating performance of various classification/prediction algorithms such as SVM, Naïve Bayes, and Logistic Regression etc to predict heart disease. All this algorithms are good in prediction but accuracy is not good enough. To get better prediction accuracy author is combining two classification algorithms such as Linear Model and Random Forest to build new algorithm called Hybrid Machine Learning to get better prediction accuracy of heart dataset. Hybrid algorithm will form up by using Voting classifier, Internally Voting classifier will build up using Linear Model and Random Forest and while classification voting algorithm will evaluate prediction accuracy of both algorithms and vote for that algorithm which gives better accuracy. So by using hybrid model always we will have better prediction accuracy algorithm which helps in better prediction of heart disease.

4.3.4   Advantage of the Proposed System

     i.   Accurate results

    ii.   Uses real time credit card transaction data

   iii.   Better Accuracy

   iv.   Detection of fraud and clean transactions

    v.   Graphical representation

   vi.   Cost-efficienccy

4.4   System Architecture

4.4.1 Date Flow Diagram
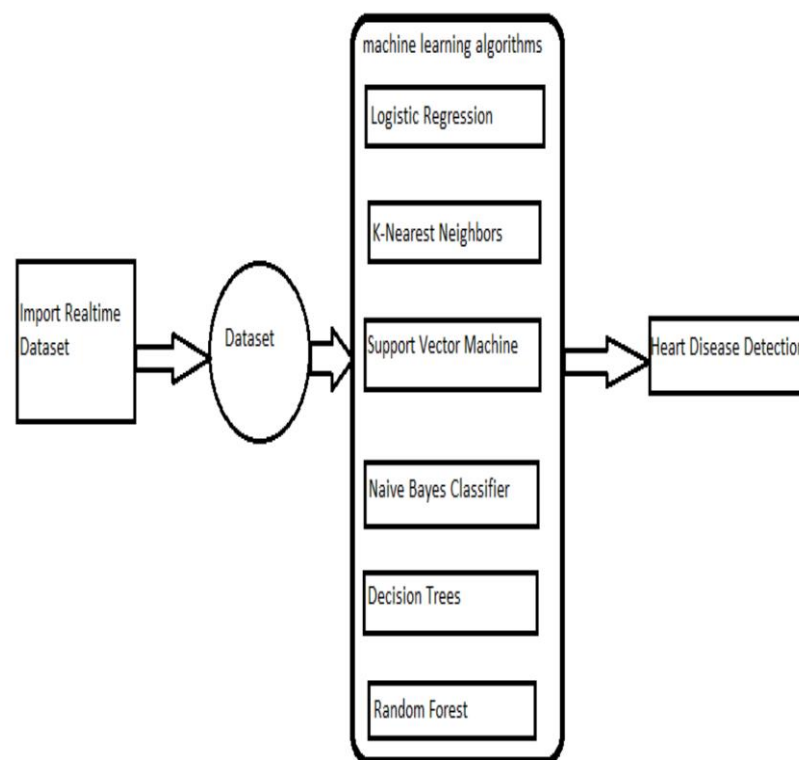
Figure 4.4.1 **:** Date Flow Diagram

The above figure represents the data flow diagram of the project.

4.4.2   Use Case Diagram
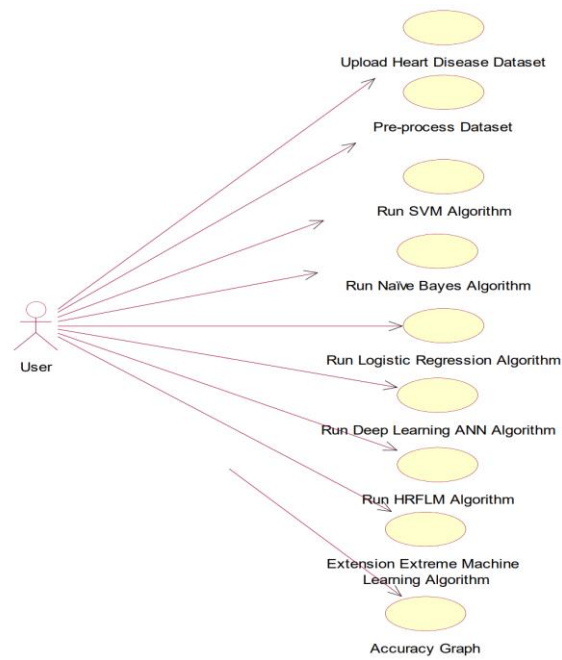
Figure 4.4.2 Use Case Diagram

4.4.3   Class Diagram
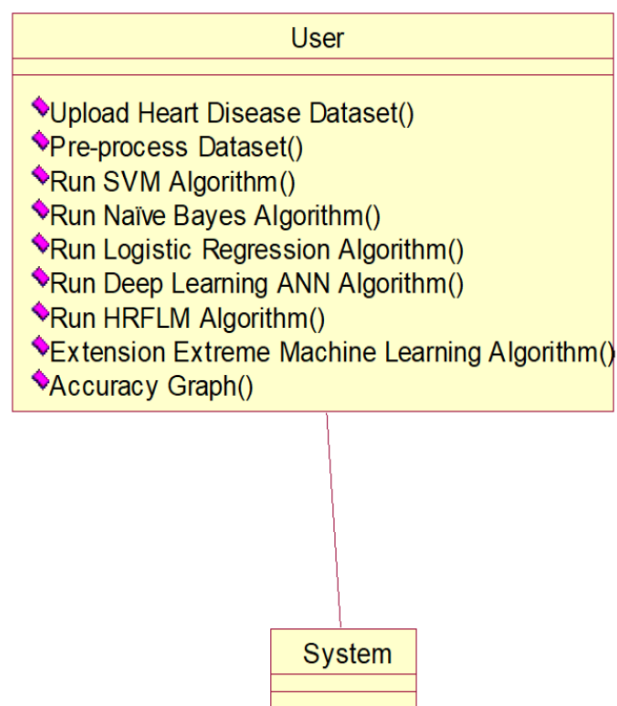
Figure:4.4.3 Class Diagram

4.5 Implementation

The project is implemented in the following modules:

- Pre-processing the dataset.

- Training the model using Logistic Regression, Support Vector Machines and KNeighbors Classifier ML algorithms.

- Evaluating the trained model and finding the best algorithm for the project.

4.5.1 Modules

Module 1: Preprocessing the dataset

- This step is performed using sklearn.preprocessing package.

- In general, learning algorithms benefit from standardization of the data set. If some outliers are present in the set, robust scalers or transformers are more appropriate.

- So, StandardScaler is used to transform the dataset.

- Visualizing the dataset is done manually

Module 2: Training the model using different ML algorithms

- The different algorithms used for training are K-Nearest Neighbors, Logistic Regression, Support Vector Machine

- These 3 algorithms are individually trained and tested

Module 3: Evaluating the trained model and finding the best algorithm for the project

- The trained model is evaluated using testing datasets.

- The best algorithm is found by calculating the accuracy of the individual model

- The accuracy is calculated using accuracy_score()function which is present in sklearn.metrics package

4.6   Screen Shots

First ,import all the libraries/packages which are necessary to analyse the dataset

```
import numpy as np
import pandas as pd
import matplotlib as plt

import seaborn as sns
import matplotlib.pyplot as plt
```

Figure 4.6.1 Python libraries

Next ,we have to insert the dataset which is present in the local system using pandas library

```
filePath = '/content/heart.csv'
data = pd.read_csv(filePath)
data.head(10)
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
5	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
6	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
7	44	1	1	120	263	0	1	173	0	0.0	2	0	3	1
8	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
9	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1

Figure 4.6.2 Heart Dataset

Next about rows and columns of the dataset

```
print("(Rows, columns): " + str(data.shape))
data.columns
```

```
(Rows, columns): (303, 14)
Index(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach',
       'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
      dtype='object')
```

Figure 4.6.3 Rows and columns

Using Calculating Correlation Matrix

```
#used for calculating correlation matrix
corr = data.corr()
plt.subplots(figsize=(15,10))
sns.heatmap(corr, xticklabels=corr.columns, yticklabels=corr.columns, annot=True,
cmap=sns.diverging_palette(220, 20, as_cmap=True))
sns.heatmap(corr, xticklabels=corr.columns,
 yticklabels=corr.columns,
 annot=True,
 cmap=sns.diverging_palette(220, 20, as_cmap=True))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1e29fffa10>
```


Figure 4.6.4 Correlation Matrix

Plotting different graphs

```
sns.catplot(x="target", y="oldpeak", hue="slope", kind="bar", data=data);
plt.title('ST depression (induced by exercise relative to rest) vs. Heart Disease',size=25)
plt.xlabel('Heart Disease',size=20)
plt.ylabel('ST depression',size=20)
```

```
Text(26.426458333333343, 0.5, 'ST depression')
```
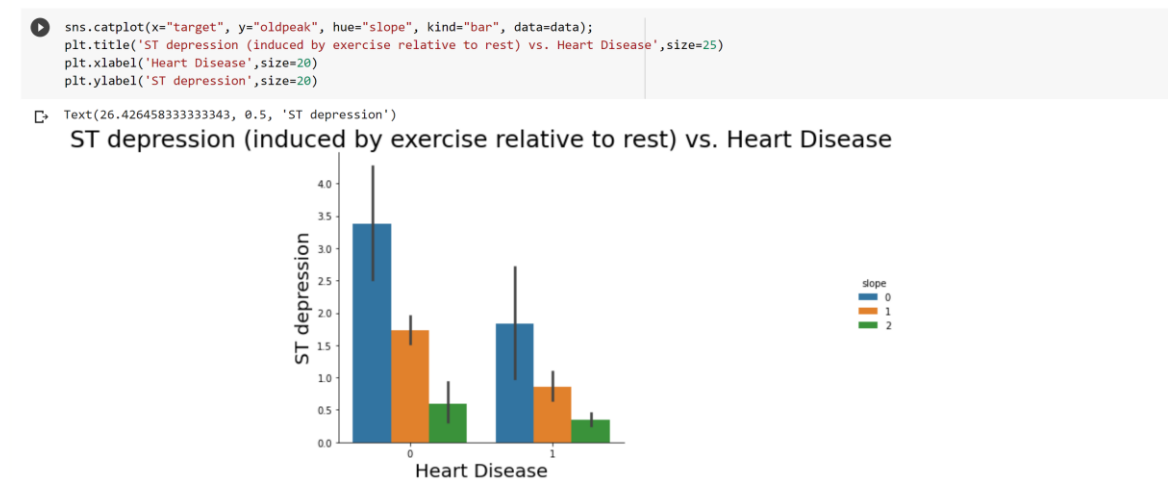
```
[ ] plt.figure(figsize=(12,8))
    sns.violinplot(x= 'target', y= 'oldpeak',hue="sex", inner='quartile',data= data )
    plt.title("Thalach Level vs. Heart Disease",fontsize=20)
    plt.xlabel("Heart Disease Target", fontsize=16)
    plt.ylabel("Thalach Level", fontsize=16)
```

    Text(0, 0.5, 'Thalach Level')
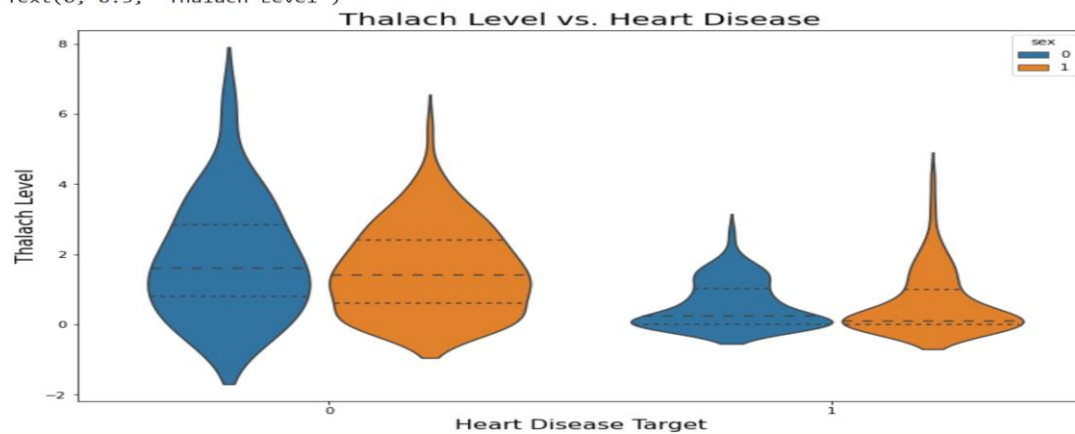

```
▶ plt.figure(figsize=(12,8))
  sns.boxplot(x= 'target', y= 'thalach',hue="sex", data=data )
  plt.title("ST depression Level vs. Heart Disease", fontsize=20)
  plt.xlabel("Heart Disease Target",fontsize=16)
  plt.ylabel("ST depression induced by exercise relative to rest", fontsize=16)
```

  Text(0, 0.5, 'ST depression induced by exercise relative to rest')

```
subData = data[['age','trestbps','chol','thalach','oldpeak']]
sns.pairplot(subData)
```
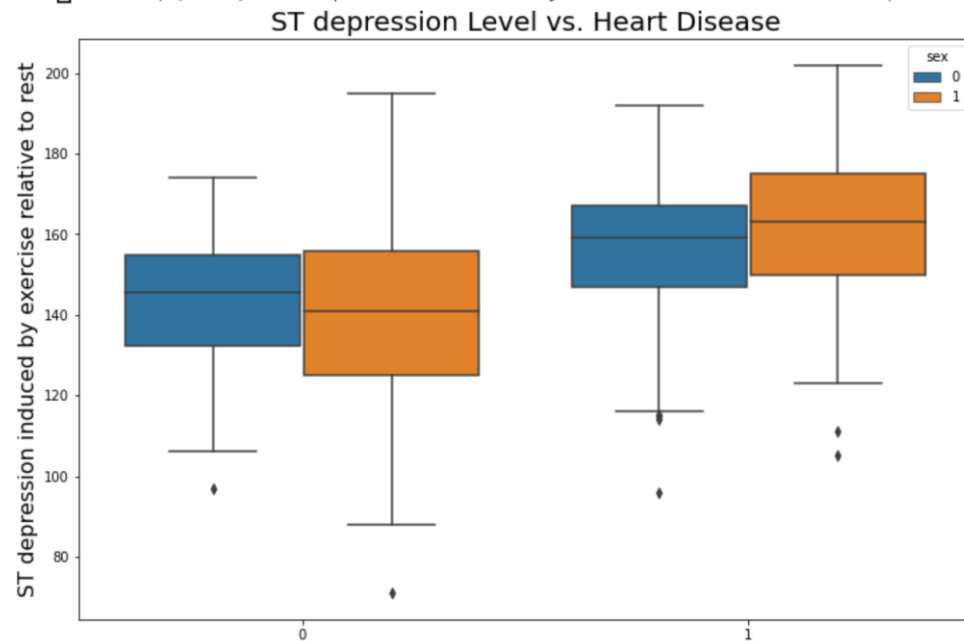
```
<seaborn.axisgrid.PairGrid at 0x7f1e26649610>
```
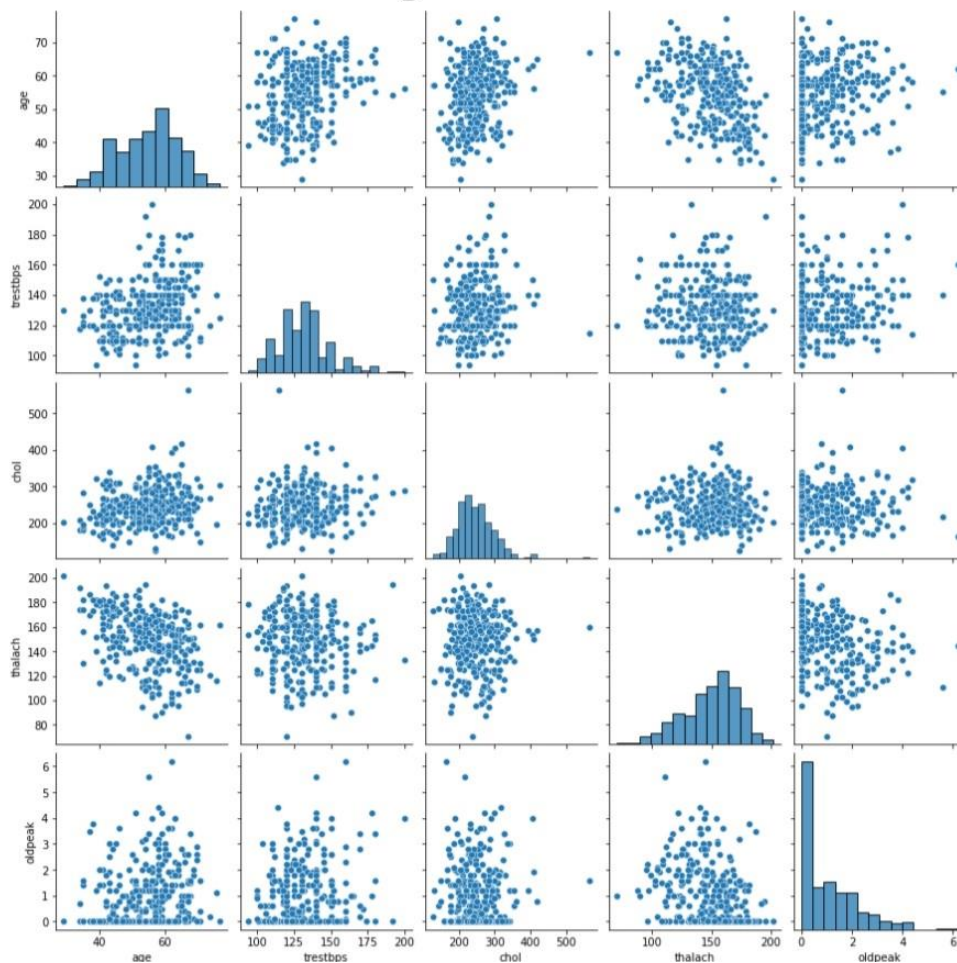

Figure 4.6.5 Graphs of dataset

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X,y,test_size = 0.2, random_state = 1)
```

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)
```

Training and testing the model using Logistic Regression

```
from sklearn.metrics import classification_report
from sklearn.linear_model import LogisticRegression
model1 = LogisticRegression(random_state=1)
model1.fit(x_train, y_train)
y_pred1 = model1.predict(x_test)
print(classification_report(y_test, y_pred1))
```

```
              precision    recall  f1-score   support

           0       0.77      0.67      0.71        30
           1       0.71      0.81      0.76        31

    accuracy                           0.74        61
   macro avg       0.74      0.74      0.74        61
weighted avg       0.74      0.74      0.74        61
```

Figure 4.6.6 Logistic Regression

Training and testing the model using KNeighbors Classifier

```
] from sklearn.metrics import classification_report
  from sklearn.neighbors import KNeighborsClassifier
  model2 = KNeighborsClassifier()
  model2.fit(x_train, y_train)
  y_pred2 = model2.predict(x_test)
  print(classification_report(y_test, y_pred2))
```

```
              precision    recall  f1-score   support

           0       0.78      0.70      0.74        30
           1       0.74      0.81      0.77        31

    accuracy                           0.75        61
   macro avg       0.76      0.75      0.75        61
weighted avg       0.76      0.75      0.75        61
```

Figure 4.6.7 KNeighbors Classifier

Training and testing the model using SVC

```
from sklearn.metrics import classification_report
from sklearn.svm import SVC
model3 = SVC(random_state=1)
model3.fit(x_train, y_train)
y_pred3 = model3.predict(x_test)
print(classification_report(y_test, y_pred3))
```

```
              precision    recall  f1-score   support

           0       0.80      0.67      0.73        30
           1       0.72      0.84      0.78        31

    accuracy                           0.75        61
   macro avg       0.76      0.75      0.75        61
weighted avg       0.76      0.75      0.75        61
```

Figure 4.6.8 SVC

Training and testing the model using Gaussian NB

```
from sklearn.metrics import classification_report
from sklearn.naive_bayes import GaussianNB
model4 = GaussianNB()
model4.fit(x_train, y_train)
y_pred4 = model4.predict(x_test)
print(classification_report(y_test, y_pred4))
```

```
              precision    recall  f1-score   support

           0       0.79      0.73      0.76        30
           1       0.76      0.81      0.78        31

    accuracy                           0.77        61
   macro avg       0.77      0.77      0.77        61
weighted avg       0.77      0.77      0.77        61
```

Figure 4.6.9 Gaussian NB

1

Training and testing the model using Decision Tree

```
] from sklearn.metrics import classification_report
  from sklearn.tree import DecisionTreeClassifier
  model5 = DecisionTreeClassifier(random_state=1)
  model5.fit(x_train, y_train)
  y_pred5 = model5.predict(x_test)
  print(classification_report(y_test, y_pred5))

              precision    recall  f1-score   support

           0       0.68      0.70      0.69        30
           1       0.70      0.68      0.69        31

    accuracy                           0.69        61
   macro avg       0.69      0.69      0.69        61
weighted avg       0.69      0.69      0.69        61
```

Figure 4.6.10 Decision Tree

Training and testing the model using Random Forest

```
from sklearn.metrics import classification_report
from sklearn.ensemble import RandomForestClassifier
model6 = RandomForestClassifier(random_state=1)
model6.fit(x_train, y_train)
y_pred6 = model6.predict(x_test)
print(classification_report(y_test, y_pred6))

              precision    recall  f1-score   support

           0       0.88      0.70      0.78        30
           1       0.76      0.90      0.82        31

    accuracy                           0.80        61
   macro avg       0.82      0.80      0.80        61
weighted avg       0.81      0.80      0.80        61
```

Figure 4.6.11 Random Forest

CHAPTER – 4

CONCLUSION

At the end of this project, we have acquired the result of an accurate value of using a random forest algorithm with new enhancements. In comparison to existing modules, this proposed module is applicable for the dataset and provides more accurate results. The Random forest algorithm will provide better performance with many training data, but speed during testing and application will still suffer. Usage of more pre-processing techniques would also assist.In this project, we have seen that the accuracy of Random Forest Algorithm is best when compared to other algorithms.

BIBLIOGRAPHY

[1] Machine Learning with Python: Design and Develop Machine Learning and Deep LearningTechnique using real world code examples, Abhishek Vijayvargia, 1st Edition, 2019.

[2]Python GUI Programming - A Complete Reference Guide: Develop responsive and powerful GUI applications with PyQt and Tkinter, Alan D. Moore, B. M. Harwani, 2019.

[3]Machine Learning for Beginners: The Definitive Guide to Neural Networks, Random Forests, and Decision Trees, Jennifer Grange, 2017.

[4] A. S. Abdullah and R. R. Rajalaxmi, ''A data mining model for predicting the coronary heart disease using random forest classifier,'' in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[5] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ''Using PSO algorithm for producing best rules in diagnosis of heart disease,'' in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017,pp.306–311.

https://www.kaggle.com/datasets?fileType=csv

APPENDIX

Appendix A: Abbreviation

MSME : Micro Small And Medium Enterprise

SVM   : Support Vector Machine

SDLC : Software Development Life Cycle

ANN   : Artificial Neural Network

Gaussian NB : Gaussian Naïve Bayes