Heart Disease Prediction Model

Submitted by

MANTRA MANAS ACHARYA
(02814202020)

In partial fulfillment for the award of the

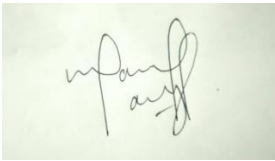*degree of*

Bachelor of Computer Application
(2020-23)

Jagannath International Management School
Vasant Kunj, New Delhi - 110070
(AUG-2022)

SELF CERTIFICATE

This is to certify that the Project Report (BCA-355) entitled "Heart Disease Prediction Model" done by me is an authentic work carried out for the partial fulfillment of the requirements for the award of the degree of BCA (2020-23) under the guidance of Jose Portilla, Head of Data Science at Pierian Training. The matter embodied in this project work has not been submitted earlier for award of any degree or diploma to the best of my knowledge and belief.

[MANTRA MANAS ACHARYA]

[02814202020]

CERTIFICATE

CERTIFICATE OF COMPLETION

2022 Python for Machine Learning & Data Science Masterclass

Instructors **Jose Portilla**
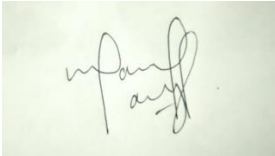
Mantra Manas Acharya

Date **Aug. 26, 2022**

Length **44 total hours**

ACKNOWLEDGEMENT

With candour and pleasure, I take opportunity to express my sincere thanks and obligation to my esteemed guide – Mr. Jose Portilla. It is because of his/her able and mature guidance and co-operation without which it would not have been possible for me to complete my project.

It is my pleasant duty to thank all the staff member of the computer centre who never hesitated me from time to time during the project.

Finally, I gratefully acknowledge the support, encouragement & patience of my family, and as always, nothing in my life would be possible without God, Thank You!

MANTRA MANAS ACHARYA

INDEX

6. CONCLUSION&FUTUREENHANCEMENT(S)

6.1 SYSTEM LIMITATIONS
6.2 FUTURE SCOPE
6.2 CONCLUSION
6.3 LEARNING & ACHIEVEMENT (IF ANY)

BIBLIOGRAPHY

****

CHAPTER 1: INTRODUCTION

1.1 ABOUT ORGANIZATION

Udemy, Inc**.** is an American massive open online course (MOOC) provider aimed at professionaladults and students. It was founded in May 2010 by Eren Bali, Gagan Biyani, and Oktay Caglar.

Udemy is a platform that allows instructors to build online courses on their preferred topics. Using Udemy's course development tools, they can upload videos, PowerPoint presentations, PDFs,

audio, ZIP files and live classes to create courses. Instructors can also engage and interact with usersvia online discussion boards.

Courses are offered across a breadth of categories, including business and entrepreneurship, academics, the arts, health and fitness, language, music, and technology. Most classes are in practical subjects such as Excel software or using an iPhone camera. Udemy also offers Udemy for Business, enabling businesses access to a targeted suite of over 7,000 training courses on topics from digital marketing tactics to office productivity, design, management, programming, and more.With Udemy for Business, organizations can also create custom learning portals for corporate training.

Organization website – https://www.udemy.com

KEY MEMBERS:

1 Eren Bani - **cofounder and chairman**

2 Mustafa Demir - **country manager and senior director of engineering**

3 Venu Venugopal - **chief technology officer**

4 Prasad Gune - **Senior vice president,Product**

5 Wesley Pua - **senior vice president finance**

6 Noah Marks - **Vice president revenue operations**

7 Faisal Nambiar- **Vice president strategic partnerships**

1.2 ORGANIZATION MISSION

Udemy offers a variety of courses to help businesses, government organizations, and nonprofits prepare for the future. The courses cover topics such as business strategy, technical skills, and leadership development. The courses are designed to help organizations of all sizes and types navigate the changing landscape. Udemy's mission is to help organizations thrive in the face of change.

1.3 ORGANIZATION VISION

With a mission **to improve lives through learning**, it's no surprise that Udemy place tremendousemphasis on the role of learning in our own lives and work. We understand that learning isn't an interruption of our work or a sign of weakness it's foundational to our growth, both as individuals and as a business.

1.4 ORGANIZATION MARKET SHARE

Udemy has market share of **0.44%** in learning-management-systems market. Udemy competes with 221 competitor tools in learning-management-systems category.

Chapter 2 – TRAINING DETAIL

2.1 Course Overview

- The course is a taught by the lead instructor of the Pierian Training
- The course has been updated to be 2022 ready and you'll be learning the latest tools and technologies used at large companies such as McKinsey, Facebook, Amazon, Google
- This course doesn't cut any corners; there are beautiful animated explanation videos and tens of real-world projects which you will get to build.
- The curriculum was developed for students who knows basics of python language and are ready to dive deep in the field of machine learning.

2.2 Course Details

This course was designed to teach the use of data science and Machine learning using some ML libraries and algorithm such as NumPy and Pandas

The course contains –

- 44 hours on-demand video
- 6 articles
- 33 downloadable resources
- 4 coding exercises
- Full lifetime access
- Access on mobile and TV
- Certificate of completion

2.3 Course Assignment completion Details

- Data Visualization assignment – To visualize the discrepancy in Movie reviews and their ratings
- Linear Regression algorithm assignment – To predict housing prices
- K- Nearest neighbors algorithm assignment – Detecting rock or mine from SONAR data
- Support Vector Machines algorithm assignment – To detect adulterated wine

2.4 Other Details

The topics that are taught in the courses -

- Programming with Python
- NumPy with Python
- Deep dive into Pandas for Data Analysis
- Full understanding of Matplotlib Programming Library
- Deep dive into seaborn for data visualizations
- Machine Learning with SciKit Learn

Chapter 3 – SYSTEM STUDY

3.1 Problem Statement

Heart is an important organ of the human body. It pumps blood to every part of our anatomy. If it fails to function correctly, then the brain and various other organs will stop working, and within few minutes, the person will die. Change in lifestyle, work related stress and bad food habits contribute to the increase in rate of several heart related diseases.

Medical organizations, all around the world, collect data on various health related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend, can be easily explored using various machine learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately.

Machine Learning can play an essential role in predicting presence/absence of Locomotor disorders, heart diseases and more. Such information, if predicted well in advance, can provide important insights to doctors who can then adapt their diagnosis and treatment per patient basis.

3.2 Project Objectives

This model takes in data, applies Machine learning algorithm and Data visualization and the returns the chances of a person having heart disease in probability/percentage

This database contains 14 physical attributes based on physical testing of a patient. Blood samples are taken and the patient also conducts a brief exercise test. The "goal" field refers to the presence of heart disease in the patient. It is integer (0 for no presence, 1 for presence). In general, to confirm 100% if a patient has heart disease can be quite an invasive process, so if we can create a model that accurately predicts the likelihood of heart disease, we can help avoid expensive and invasive procedures.
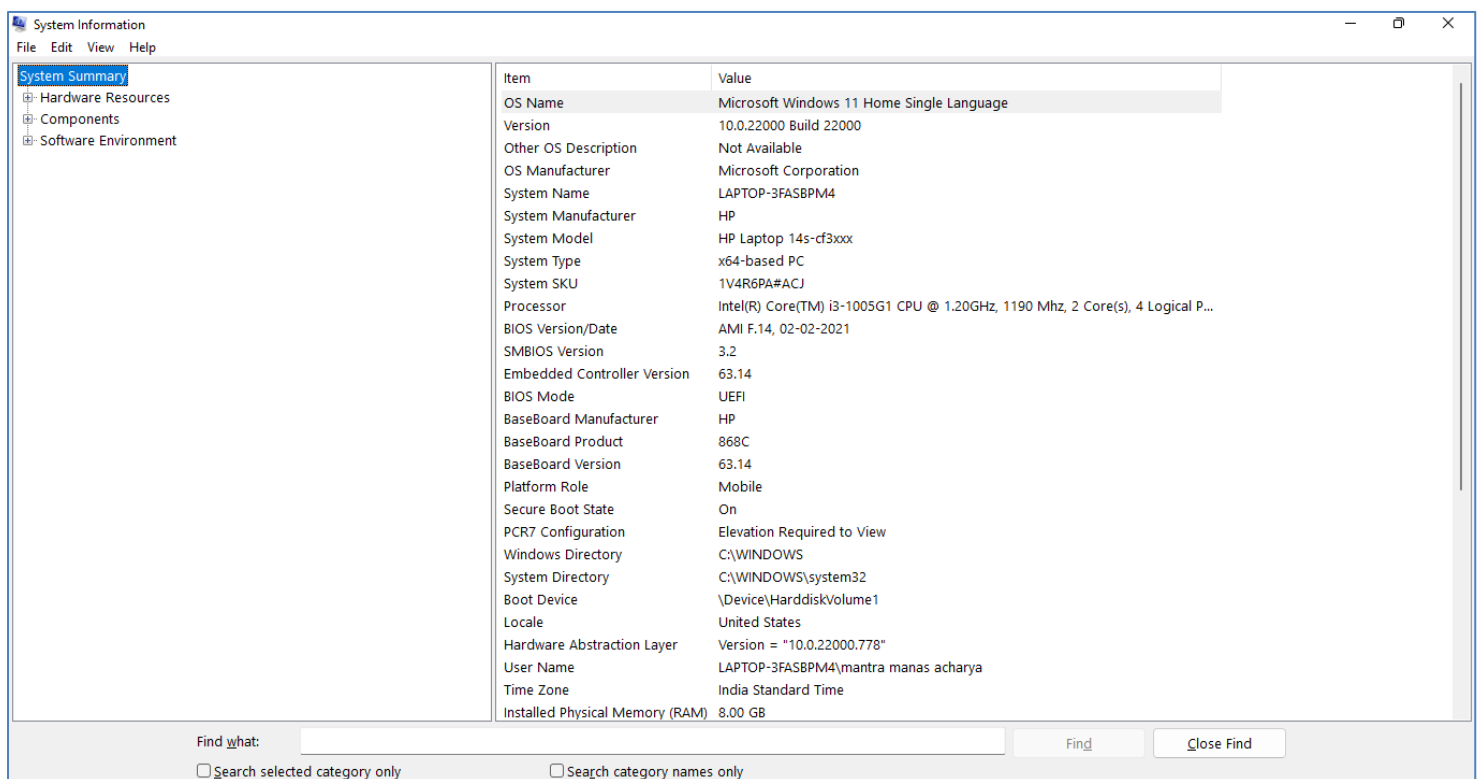
3.3 PROJECT SCOPE

Here the scope of the project is that integration of clinical decision support with computer-based patient records could detect possibility of Heart- Disease, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modelling and analysis tools and machine learning algorithms have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

3.4 NEED OF THE SYSTEM

Heart Disease is one of the major concerns for society today. It is difficult to manually determine the odds of getting heart disease based on risk factors. However, machine learning techniques are useful to predict the output from existing data. By the use of Machine Learning and Data visualization technology can provide assistance to the medical system.

3.5 SYSTEM REQUIREMENTS

- **Hardware Requirements –**
  - Any device with or without internet connectivity
  - Sufficient RAM to process the data

- **Software Requirements –**
  - o Python installed in the system (latest version recommended)
  - o Installation of python libraries such as NumPy, Pandas, Scikit etc.
 - o Jupyter notebook or any IDE to code the model.

Chapter 4 - SYSTEM DESIGN

4.1 INTRODUCTION

Software design is actually a multi-step process that focuses on four distinct attributes of a program data structure, software architecture, interface representations, and procedural (algorithmic) detail. The design process translates requirements into a representation of the software that can be accessed for quality before coding begins. Like requirements, the design is documented and becomes part of the software configuration.

4.2 DESIGN DIAGRAMS

- ### 4.2.1 SYSTEM FLOW CHARTS
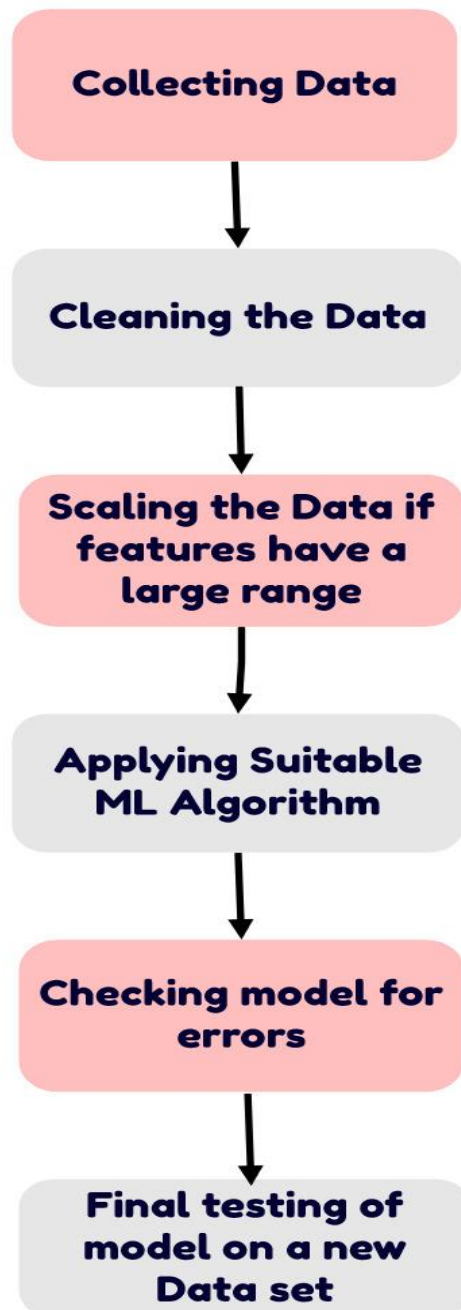
- **4.2.2 USE CASE DIAGRAM**

o The requirements of a system can be captured by Use Case Diagrams. They are modelled to capture the intended behaviour of the system. Use Cases interact with human or actors that use the system to accomplish some work. They define asset of sequence of actions that a system performs to yield an observable result of value to an actor. An actor represents a role that a human, a hardware device or another system plays with a system.

o Use Cases are used to come to a common understanding with the systems end users and the domain experts. They help invalidating the systems architecture and its evolution process after at thorough understanding of the requirements of the system.

o The use cases are modelled following the steps mentioned below:

o Identify the actors that interact with the system.

o Organize actors according to their roles.

o Identify the primary ways in which an actor interacts with the system elements.

- **4.2.3 APP MAPS**

Feeding data and applying algorithm

Providing Data

Gives Prediction

miro
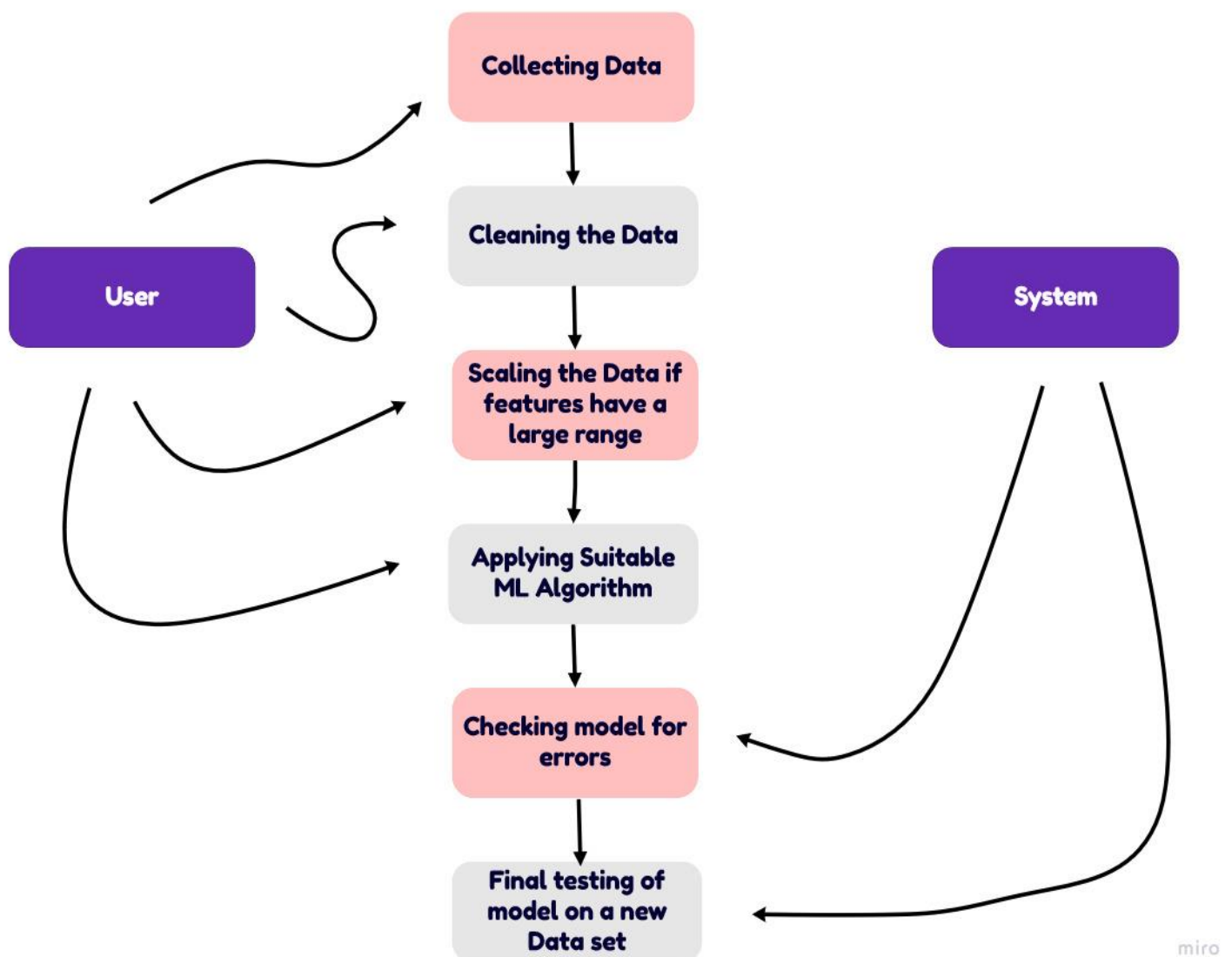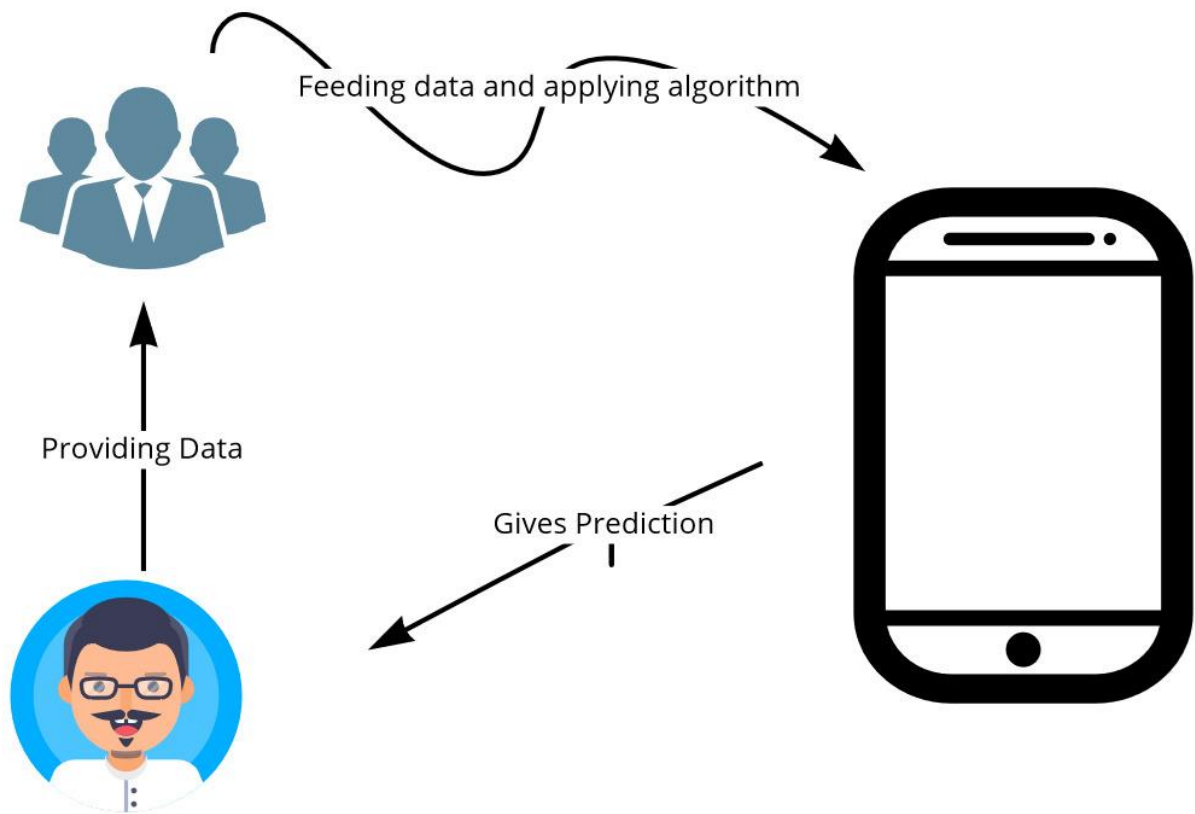
Chapter 5-SOFTWARE CODING & IMPLEMENTATION

5.1 INTRODUCTION

The model in this project has been built using **Random Forest** Machine learning algorithm

Random Forest is also a popularly supervised machine learning algorithm. This technique can be used for both regression and classification tasks but generally performs better in classification tasks. As the name suggests, Random Forest technique considers multiple decision trees before giving an output. So, it is basically an ensemble of decision trees. This technique is based on the belief that a greater number of trees would converge to the right decision. For classification, it uses a voting system and then decides the class whereas in regression it takes the mean of all the outputs of each of the decision trees. It works well with large datasets with high dimensionality.
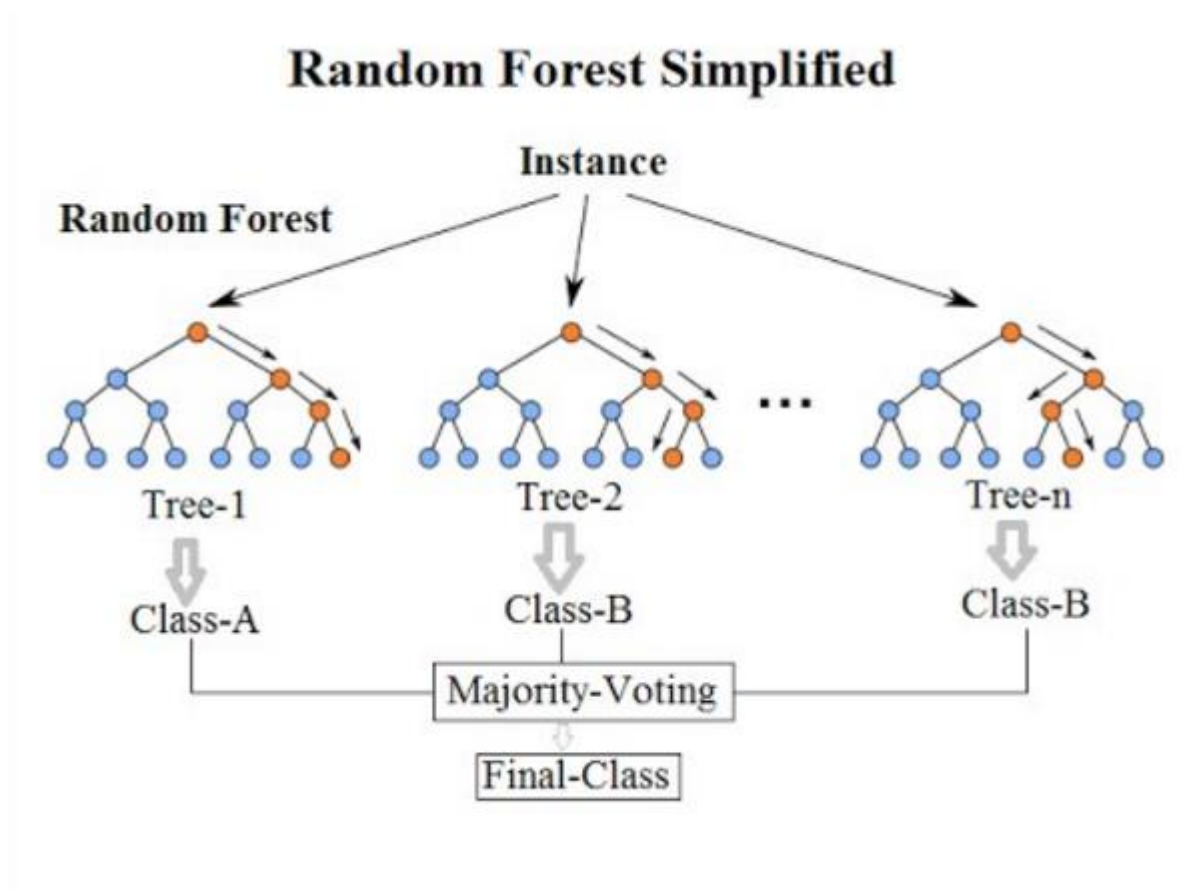
5.1.1 SYSTEM CODING ENVIRONMENT AND STANDARDS FOLLOWED

- o This project is built on VS code IDE (version 1.70.2).
- o The project is a python notebook file saved with the extension '.ipynb'
- o The notebook is later converted as an API with the help of JSON and Postman app.

5.1.2 SAMPLE CODE LAYOUTS

Reading and Exploring our Data Set

```python
df= pd.read_csv('heartdata.csv')

# we are reading the data
```
[45]

```python
# Now we'll check if there's any data missing in our data set
df.isnull().sum()

# Below we can see that no column has any null value
```
[46]

```
age          0
sex          0
cp           0
trestbps     0
chol         0
fbs          0
restecg      0
thalach      0
exang        0
oldpeak      0
```

Creating our Model

```python
from sklearn.ensemble import RandomForestRegressor
```
[52]

```python
ourmodel = RandomForestRegressor(n_estimators= 50, random_state=101)
```
[53]

```python
ourmodel.fit(scaled_X_train, y_train)

# Here we have fitted our training data to our model created so that the training data set can be trained upon that model
```
[54]

RandomForestRegressor(n_estimators=50, random_state=101)

Searching for best parameters
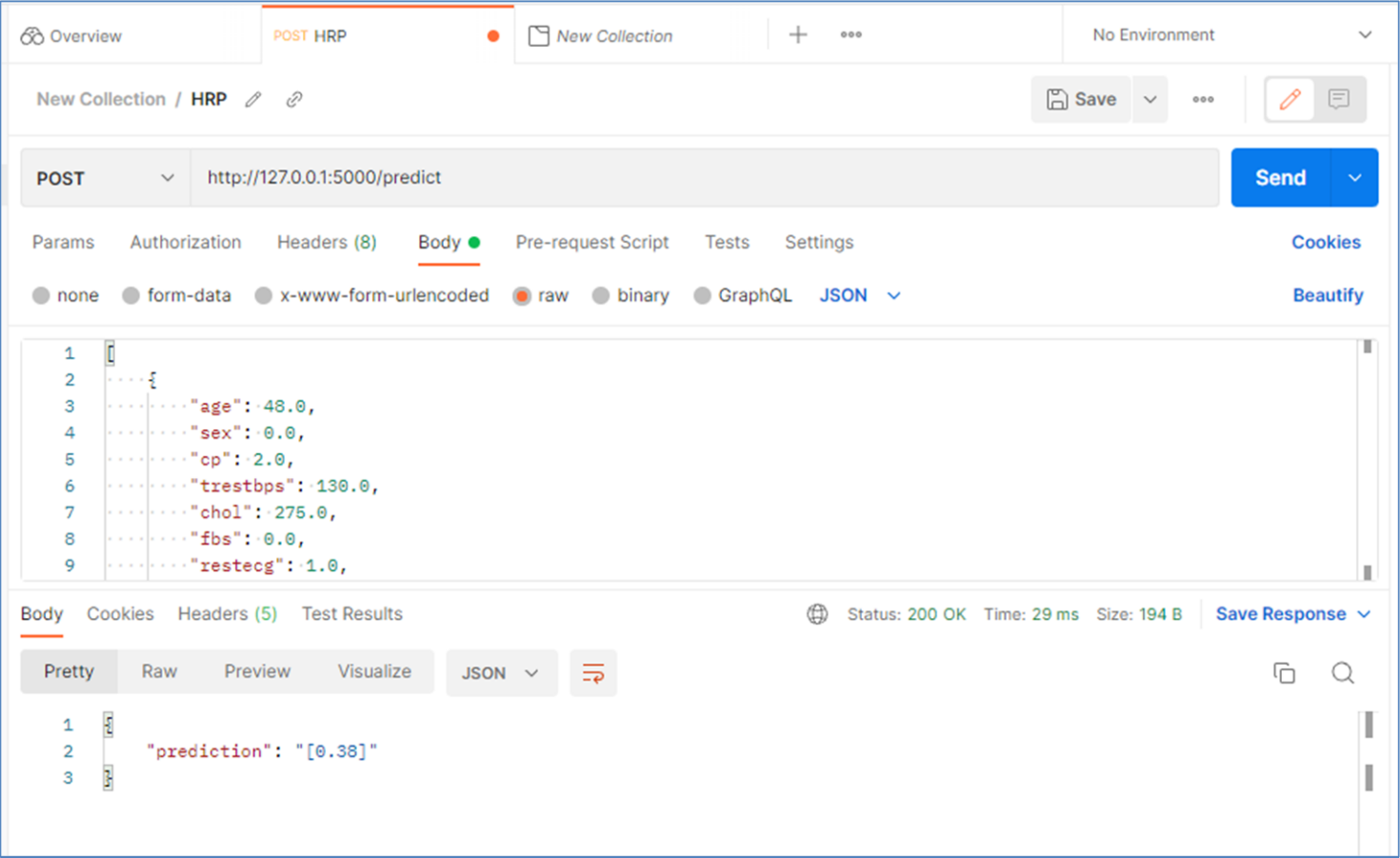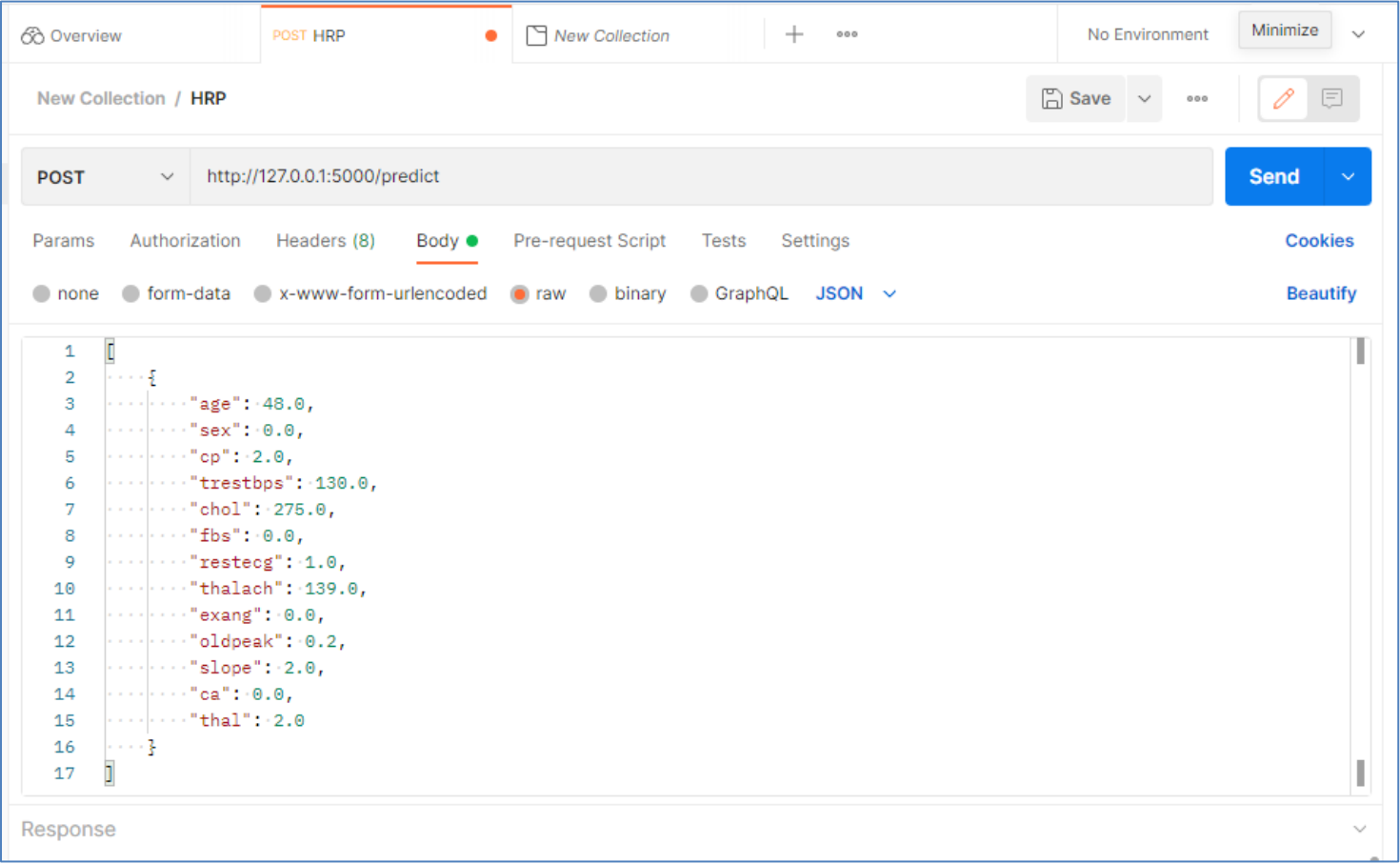
```python
param_grid={'n_estimators':[3,6,10,15,20,30,50]}
```
[55]

5.2 APPLICATION SCREEN SHORTS (WITH DATA)

Chapter 6 – CONCLUSION & FUTURE ENHANCEMENT(S)

6.1 SYSTEM LIMITATIONS

Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial. Clinical decisions are often made based on doctor's intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.

6.2 FUTURE SCOPE

Machine Learning is a very vast and diverse field and its scope and implementation is increasing day by day. The future scope of this system aims at giving more sophisticated prediction models, risk calculation tools and feature extraction tools for other clinical risks.

6.3 CONCLUSION

A cardiovascular disease detection model has been developed using ML classification modelling techniques. This project predicts people with cardiovascular disease by extracting the patient medical history that leads to a fatal heart disease from a dataset that includes patients' medical history such as chest pain, sugar level, blood pressure, etc. This Heart Disease detection system assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease. The algorithm used in building the given model is Random Forest Classifier. The accuracy of our model is 87.5%.

Use of more training data ensures the higher chances of the model to accurately predict whether the given person has a heart disease or not. By using these, computer aided techniques we can predict the patient fast and better and the cost can be reduced very much