Course code – SWE 960

Name: Rakhi Jha

Student ID: 89347

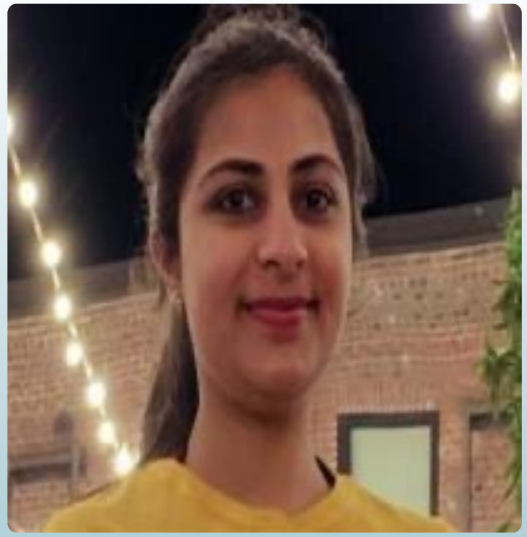
Master's in Software Engineering

Course code – CSE 690

Name: Soni Kumari

Student ID: 94970

Master's in Computer science



Course code – CSE 690

Name: Sunanda Sharma

Student ID: 95732

Master's in Computer science

GROUP -13

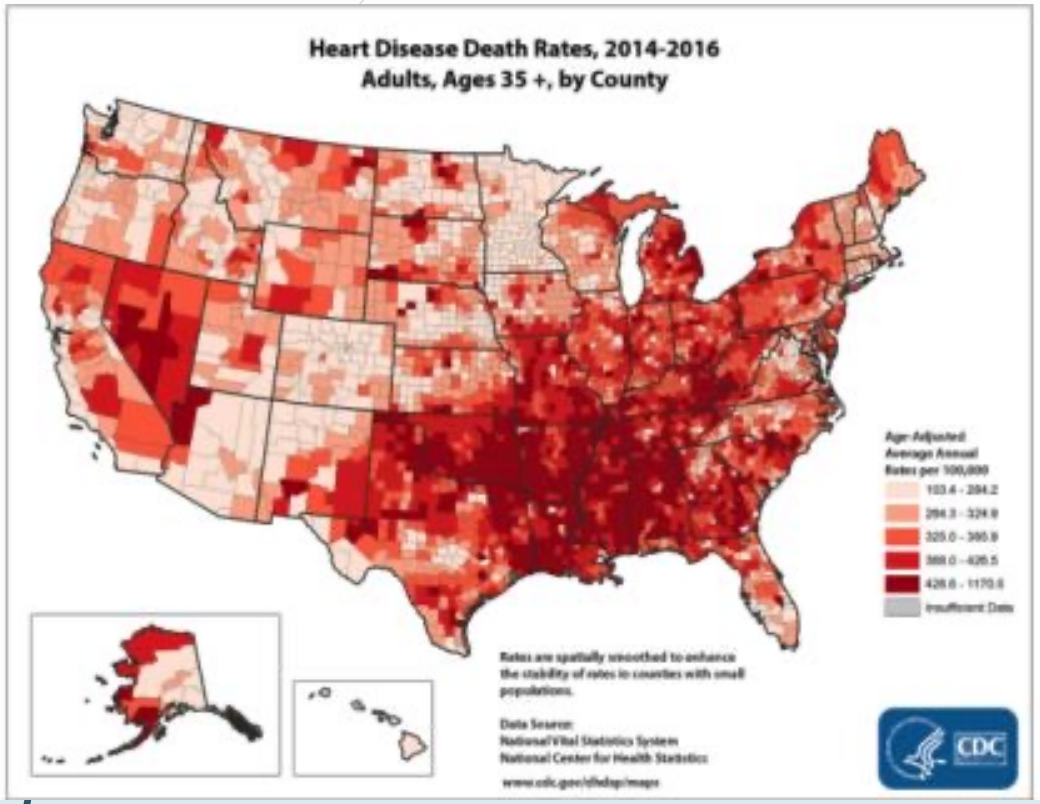
CAPSTONE PROJECT

Heart Disease Prediction System

Presentation Overview

Project Overview	Objective & Mission	Development Methodology	Technology Used
Product Details	Project Modules	Machine Learning module & Implementation	Data Visualization
ML integration with Django	Project Live Demo	Development Status	Key to Success
Competitive Advantage	Business Strategy	Financial Goal	Future Work

Project Overview



Heart Disease Facts:-

Cardiovascular diseases (CVDs) are the number 1 cause of death **globally**, taking an estimated 17.9 million lives each year.

Heart Disease in the United States

- **One person dies every 36 seconds** in the United States from cardiovascular disease.
- About **655,000 Americans** die from heart disease each year—that's **1 in every 4 deaths**.
- According to research, Heart disease costs the United States about **$219 billion** each year from 2014 to 2015.

There is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance.

It's a Web application, that allow users to get an instant guidance about their heart condition.

Objective And Mission

- The main objective of this project is to develop a heart disease prediction application.
- This application aims to make use of Machine learning techniques on medical dataset to assist in the prediction of the heart diseases.
- We evaluated various machine learning algorithms on the heart dataset and choose the best performance algorithm among them.
- The mission of this project is to predict heart disease at early stage to avoid later consequences, without spending much.

Development Methodology

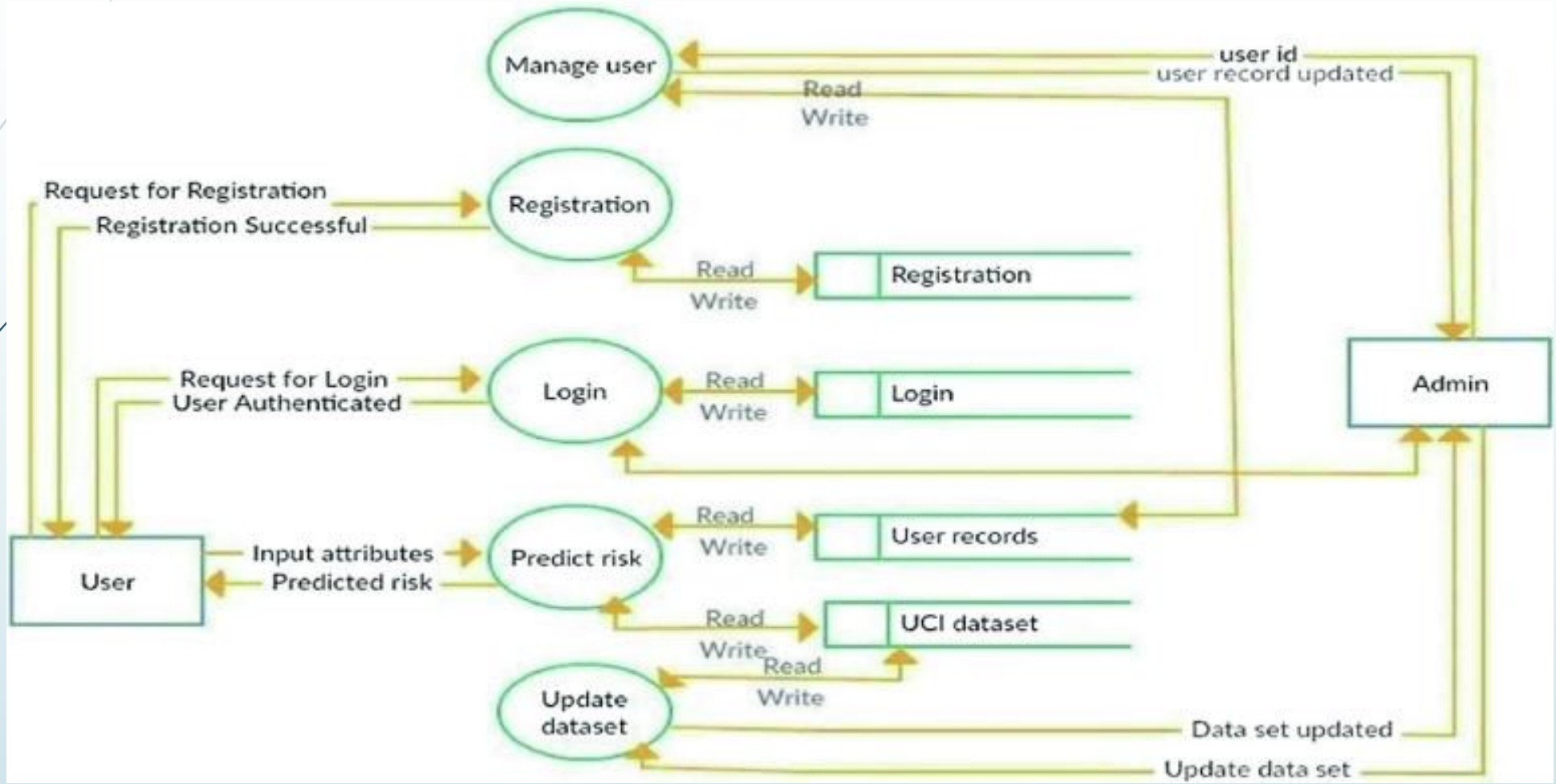
Functional Requirements

- This application allows users to login and registration for the new users.
- It allows the users to update their user profiles and medical records.
- The system performs heart disease prediction based on information given by users and display the results.
- This application also contains admin site, where it can monitor suspicious users' activity and any potential threat to the system.
- It also allows guest users to get their heart prediction.

Non-Functional Requirements

- **Performance -** The website is responsive and consistent across different screen sizes and resolutions
- **Scalability -** This application can be easily scaled out or scaled in.
- **Portability -** The website can be easily migrated in any system.
- **Security –** The website provides great protection against threats like XSS, CSRF, SQL injections, clickjacking etc.
- **Reliability –** The machine learning algorithm has high accuracy, which is highly reliable.
- **Usability -** Very easy and user-friendly interface.

System Use-Case Diagram



Technologies Used & Their Advantages

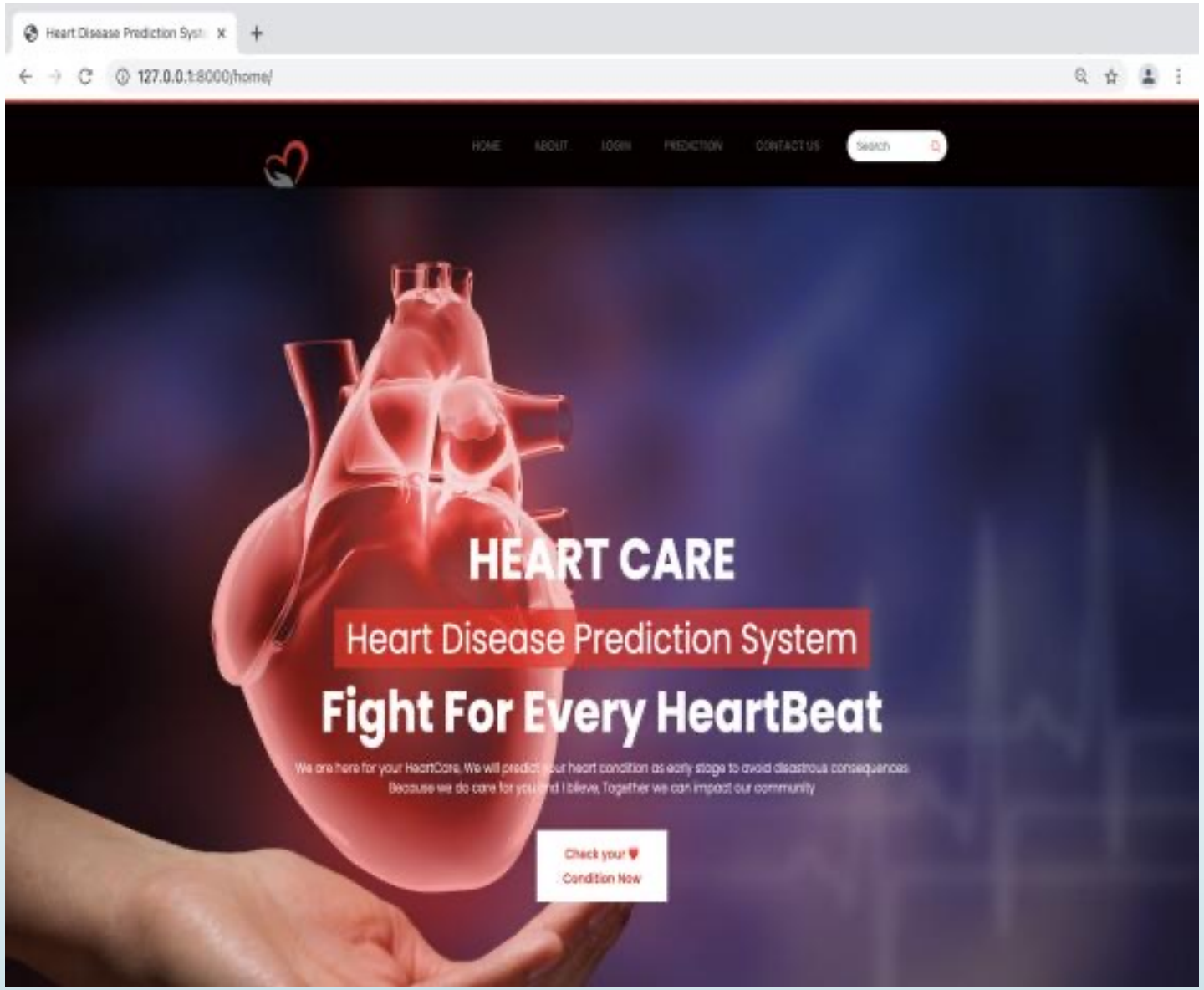
- **HTML5** used for development of user interface and the design layout of the pages.
- **Cascading Style Sheet (CSS3)** is used to set the style in web pages that contain HTML elements.
- **JavaScript** is used for all the validation user input validations and to provide some animations in the web site pages.
- **Python3.9** is used all over in the project for backend modules.
- **SQLite3** database has been used as database management system for the project.
- **Django Framework 3.2** - This is the main component of the system which binds all the other components of the system together.
- **Machine Learning** is used to develop an algorithm which will run in the back end of the system.

Product Details

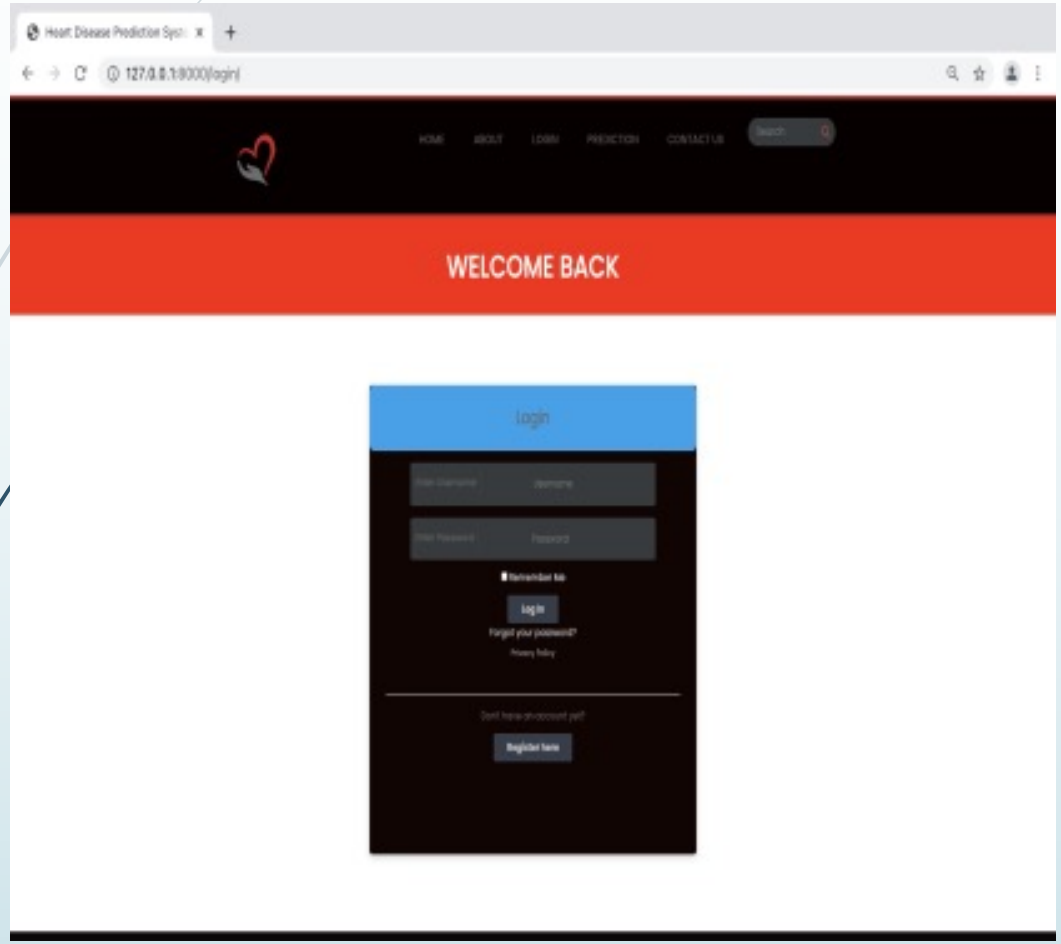
- We've developed a web application that allows users to get an instant information about their heart.
- The application is fed with various details about the heart parameters to predict the result.
- We have integrated the machine learning model into the Django Framework, which will run in the background for the prediction.

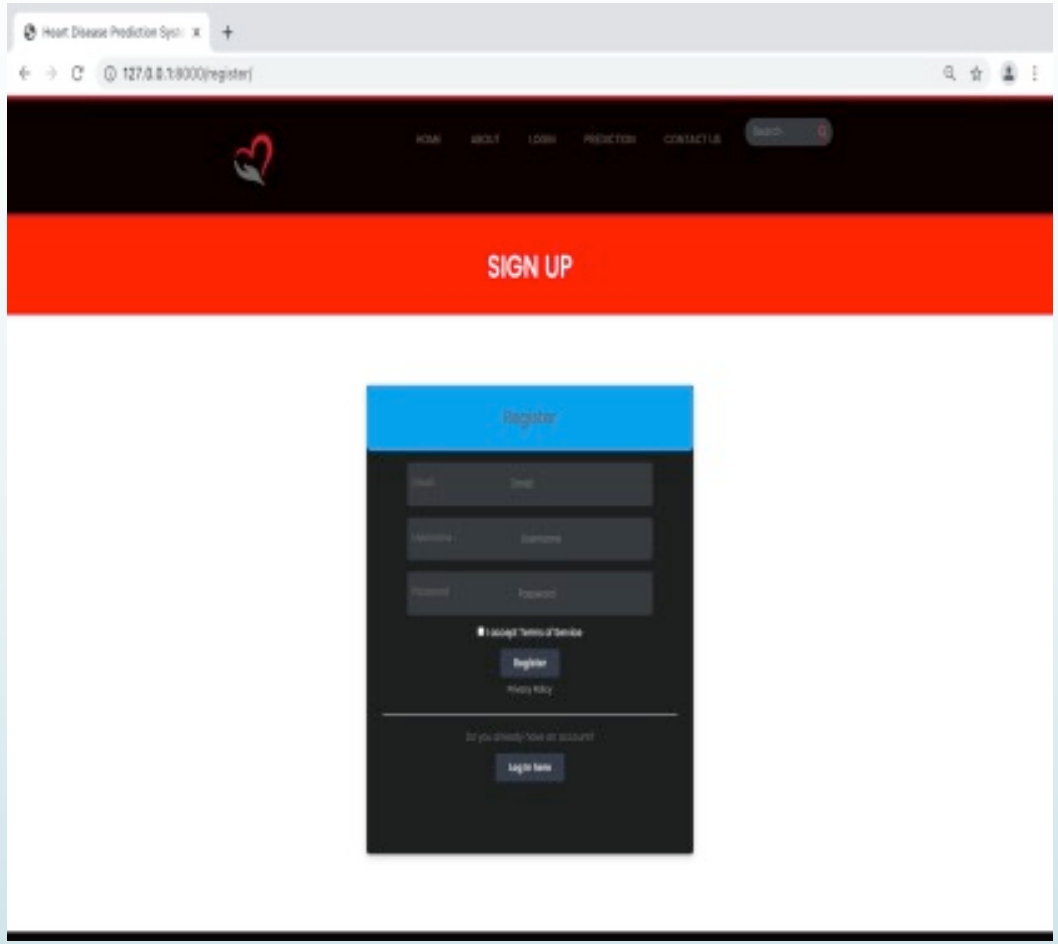
Home Page Module

Important Project Modules

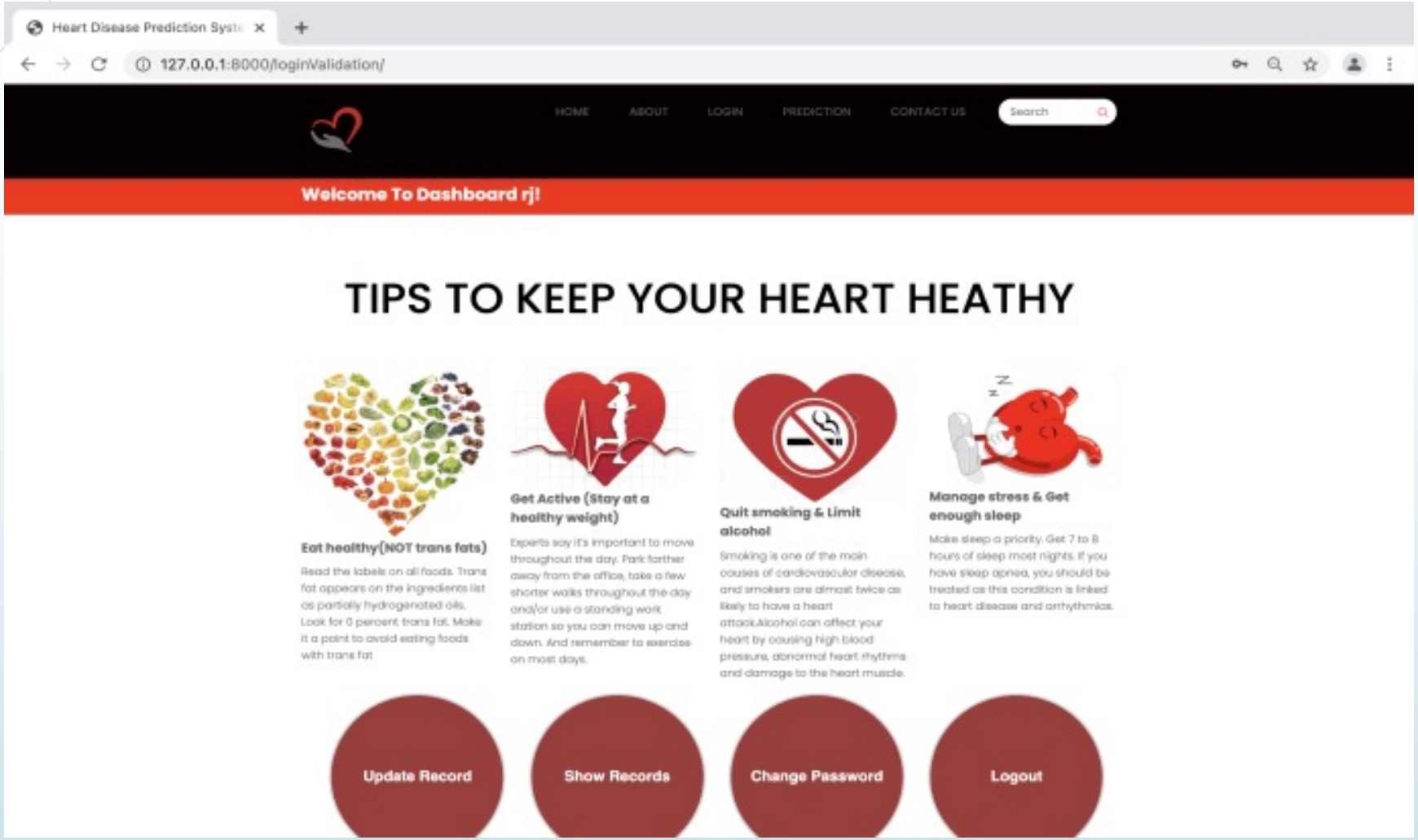


User login and Registration Module

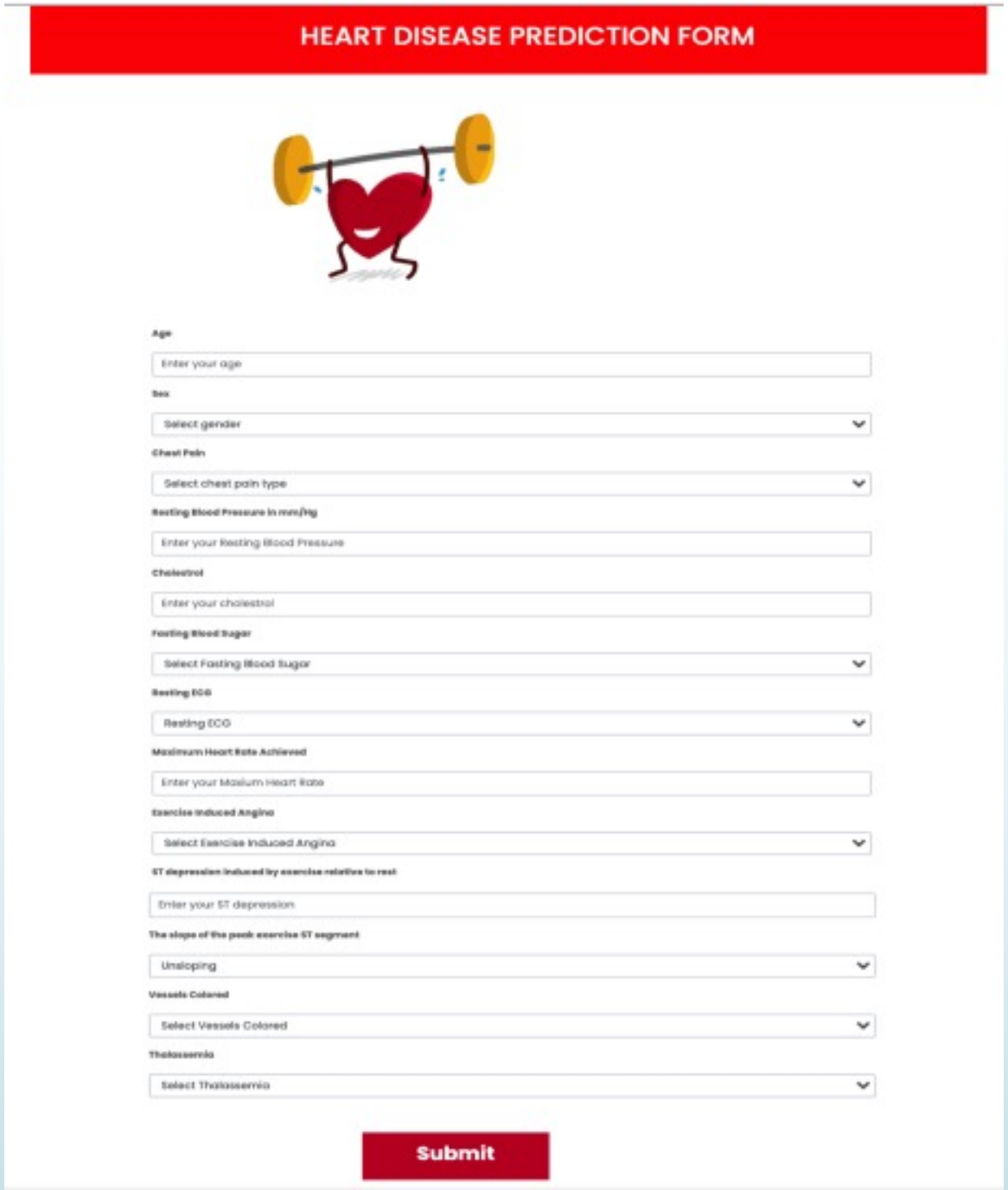




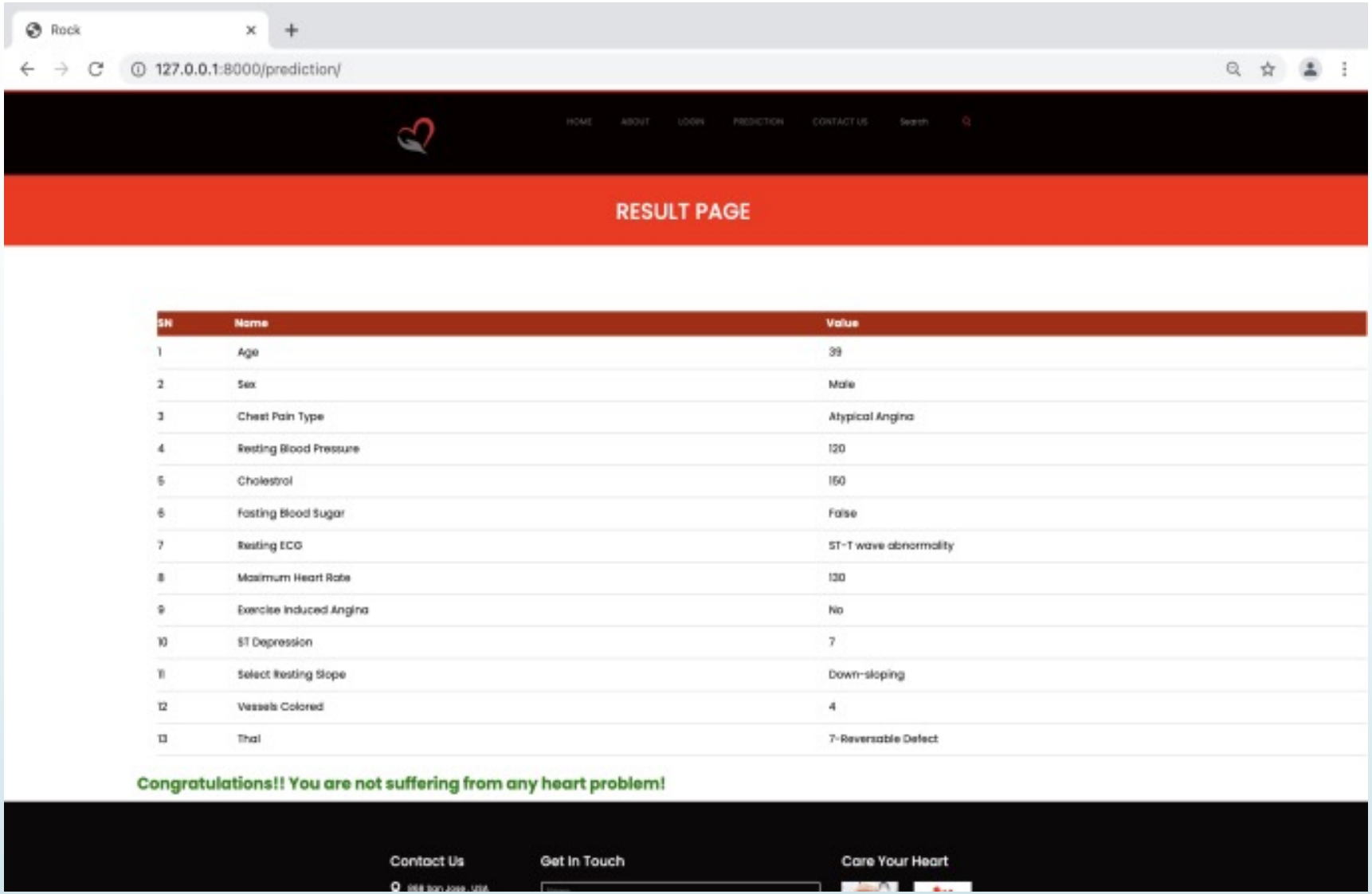
Dashboard Module



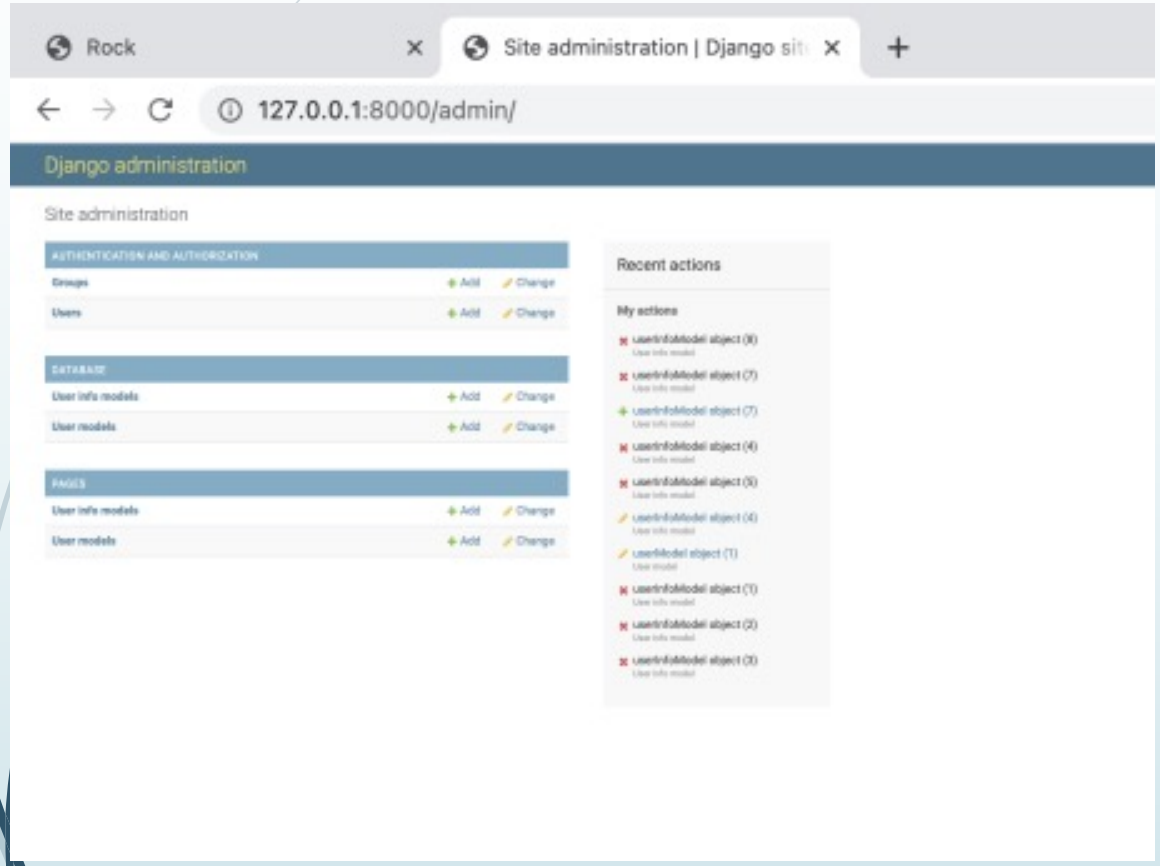
Prediction Module

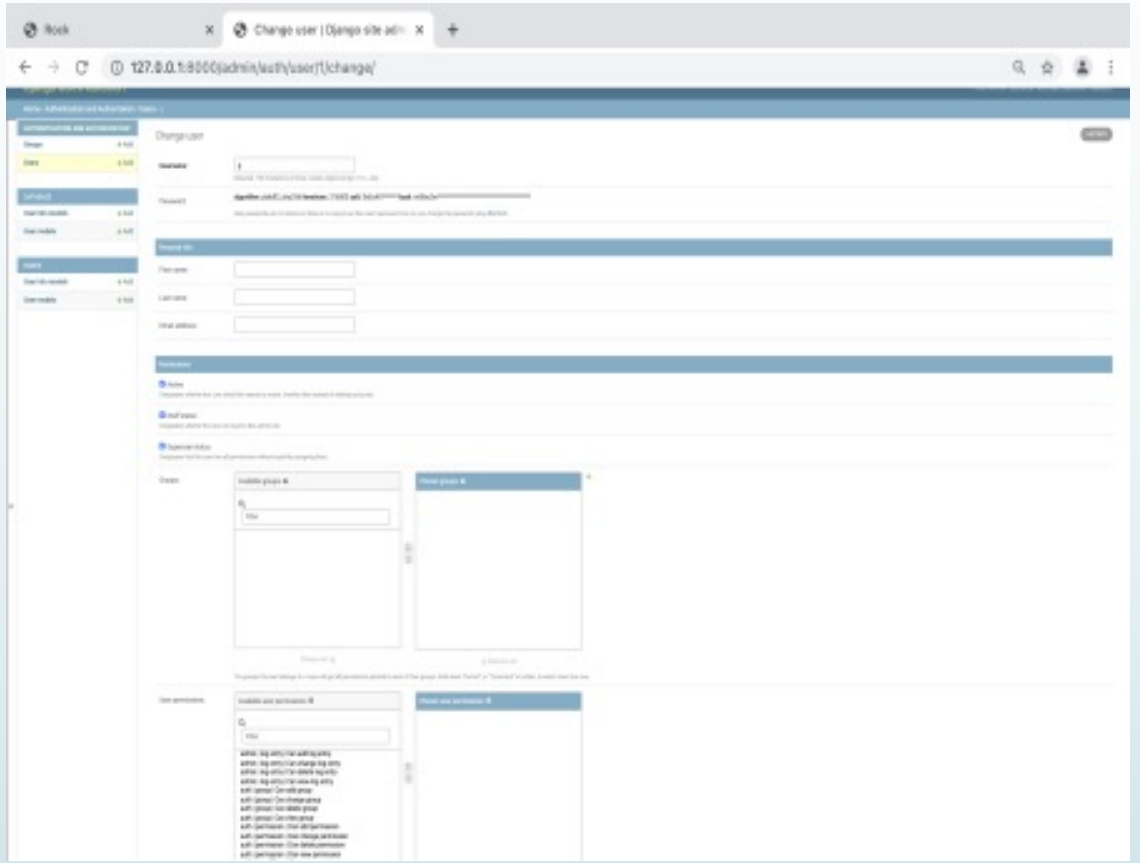


Prediction Page



Admin Module



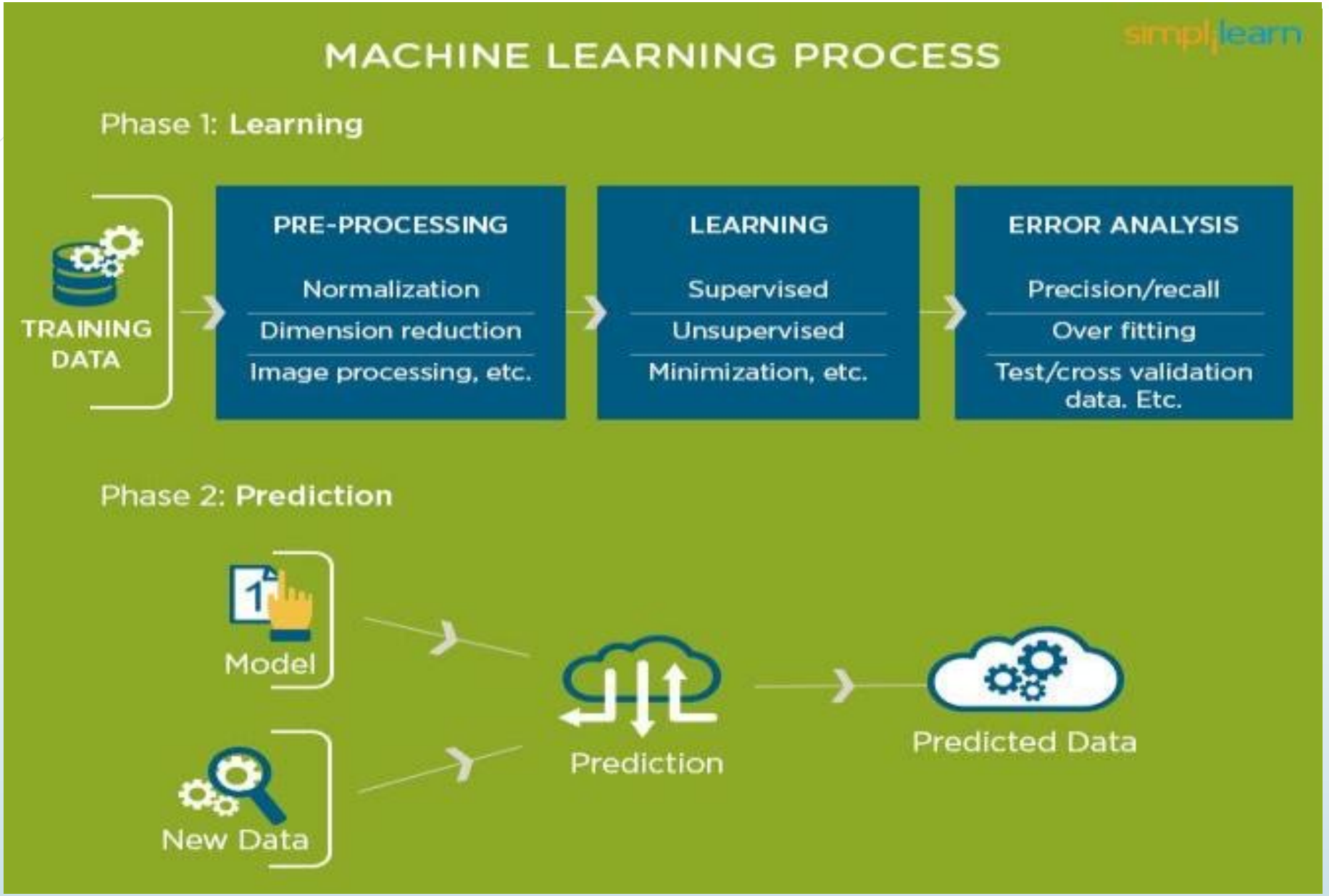


Machine Learning Module and Implementations

The following are the steps to perform suitable Machine learning technique:

- Defines a problem statement.
- Classifying the problem into ML problems.
- Selecting perfect ML algorithms based on their type of problems.
- Collecting and cleaning the data.
- Training a model from data.
- Test the Model from test data
- Evaluate a model from its accuracy

Machine Learning Process

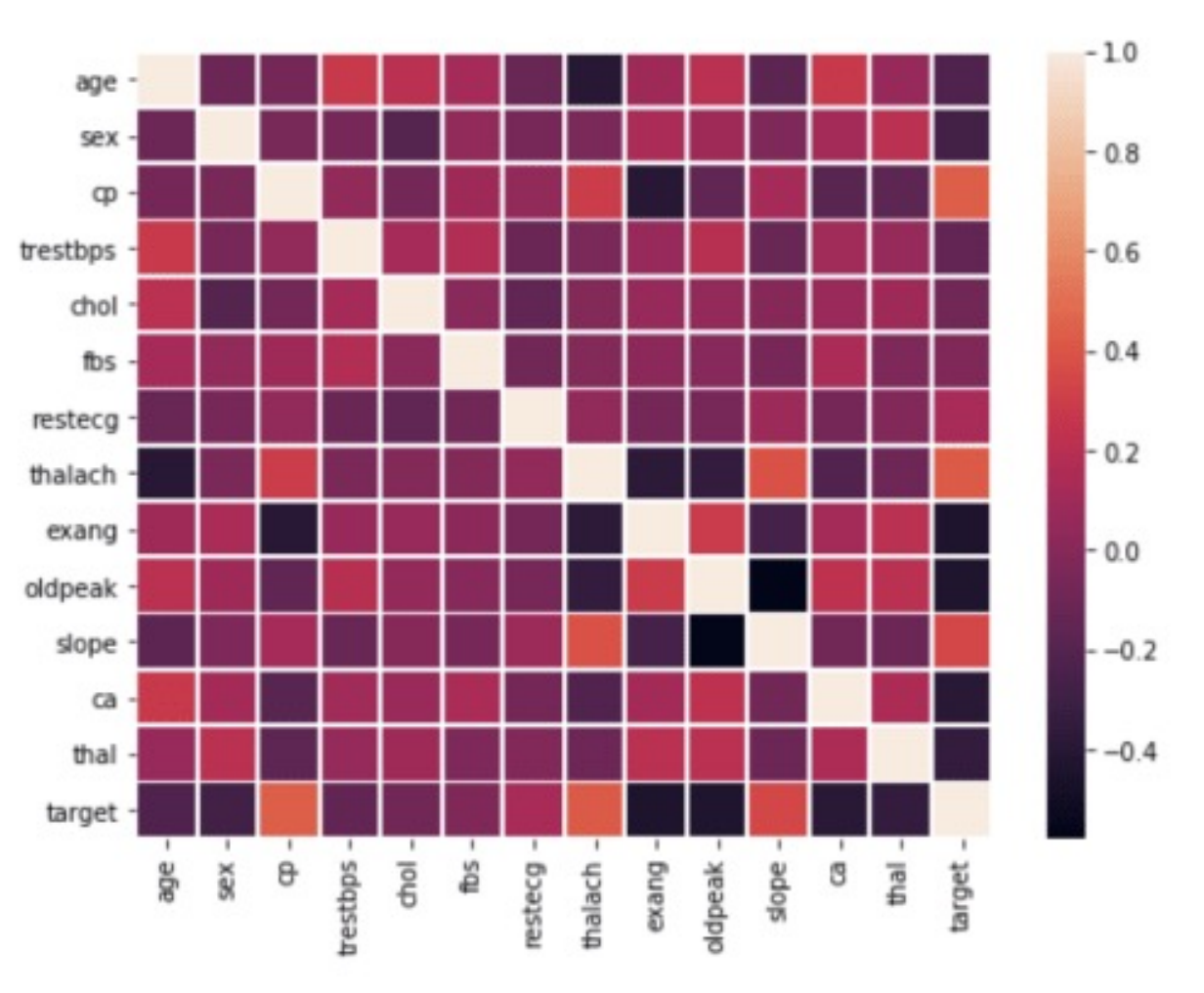


Dataset Overview

	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows × 14 columns

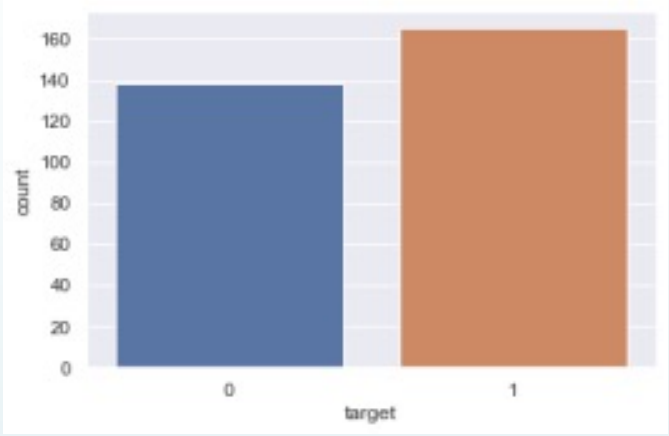
Correlation Metrix



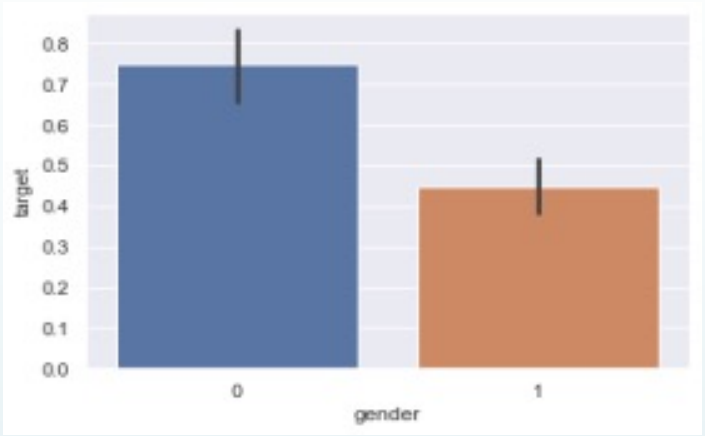
- **Correlation** indicates how the features are related to each other or to the target variable.
- Features having maximum correlation with target variable:
  - **exang** (Exercise induced angina): 0.436757
  - **cp** (Chest-pain): 0.433798
  - **oldpeak** (ST depression induced by exercise relative to rest): 0.430696
  - **thalach** (blood disorder called thalassemia): 0.421741

Visualizations

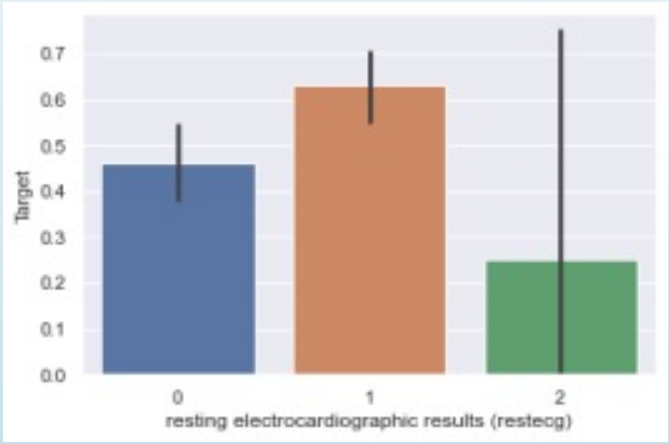
Count Plot for 'Target' Variable



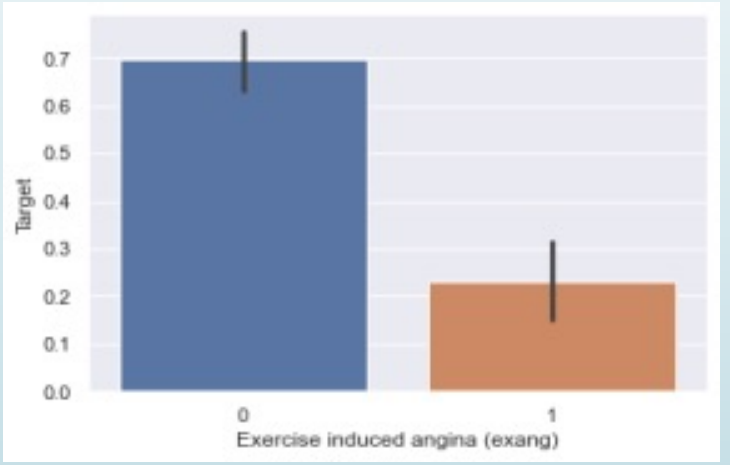
Bar Plot for 'gender' and 'target' columns



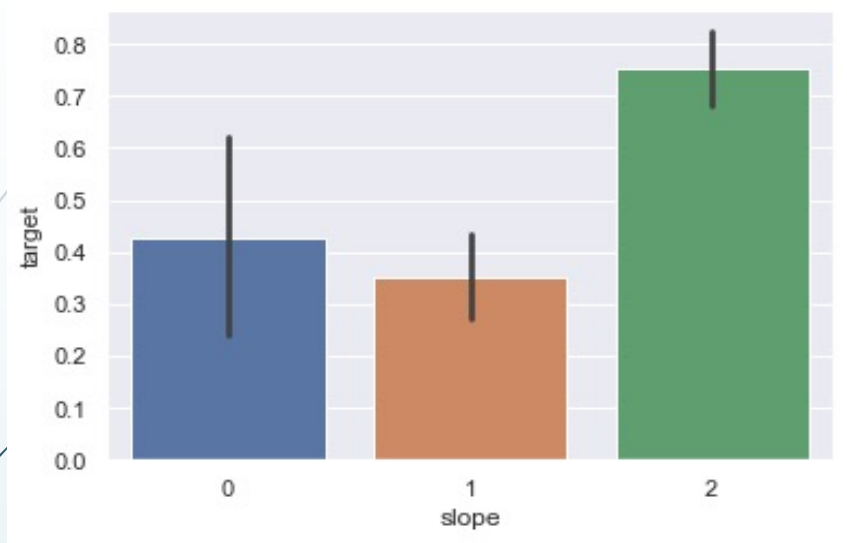
Bar Plot for 'restecg' & 'target' columns

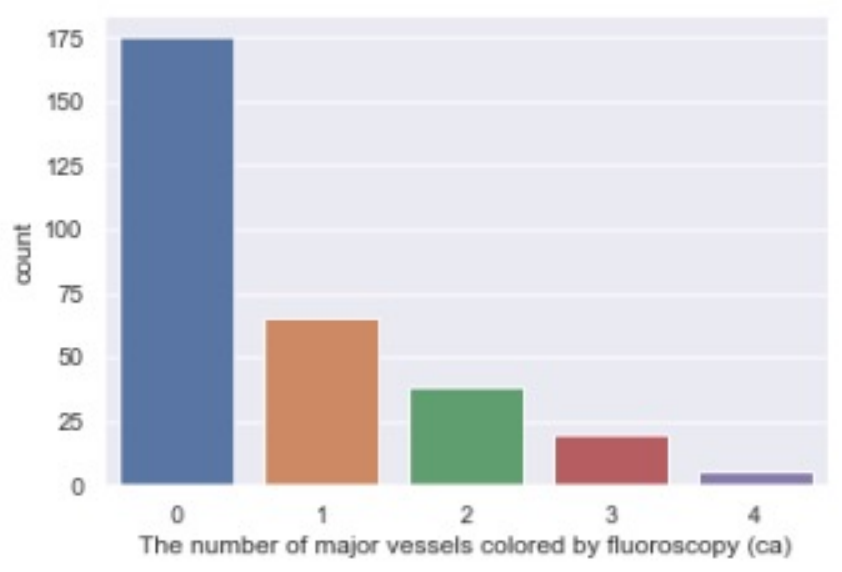


Bar Plot for 'exang' and 'Target' columns

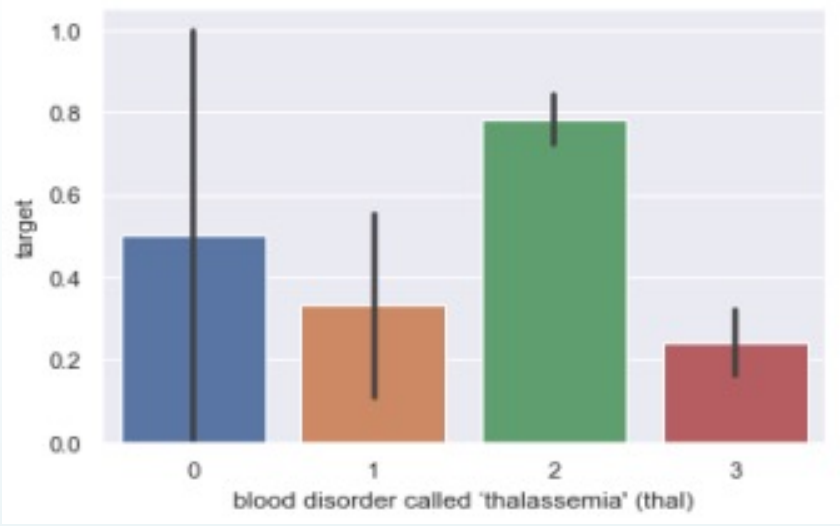


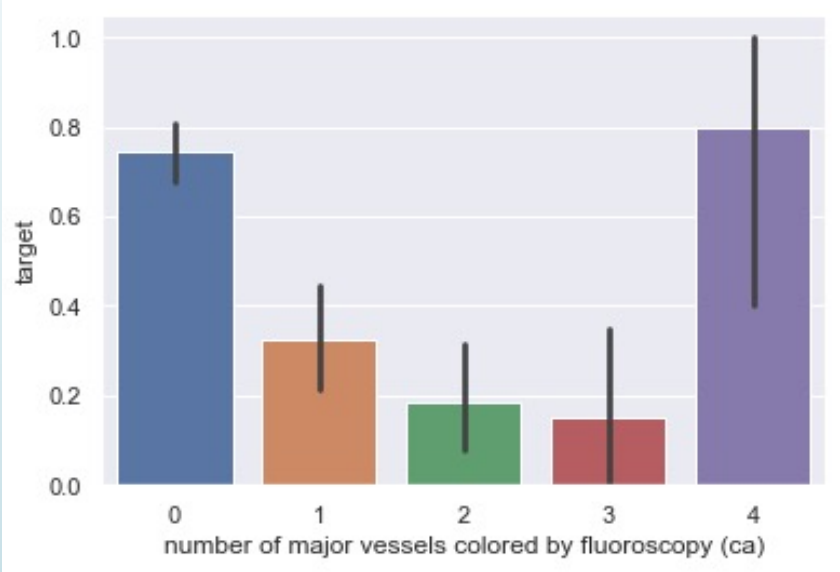
Bar Plot for ‘slope’ & ‘target’ columns





Bar Plot for ‘thal’ & ‘target’ variable





Random Forest Algorithm

- Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction
- Random Forest is a learning method that operates by constructing multiple decision trees. The final decision is made based on the majority of the trees and is chosen by the random forest.

Why Random Forest algorithm?

- The main reason for choosing random forest algorithm is that it can be used for both classification and regression tasks.
- Overfitting is one critical problem that may make the results worse, but for Random Forest algorithm, if there are enough trees in the forest, the classifier won't overfit the model.
- Classifier of Random Forest can handle missing values
- Random Forest classifier can be modeled for categorical values.

Accuracy Achieved

With the use of Random Forest (Classifier/Algorithm) we got the accuracy of almost 97%

```
In [130]:
import pickle

In [131]:
with open('model_pickle', 'wb') as f:
    pickle.dump(rand_classifier, f)

In [132]:
with open('model_pickle', 'rb') as f:
    mp = pickle.load(f)

In [133]:
patient1 = mp.predict([[39, 0, 1, 135, 208, 0, 0, 171, 0, 1.5, 2, 0, 2]])
print(patient1[0])

1

In [134]:
patient2 = mp.predict([[50, 0, 0, 100, 300, 1, 1, 200, 1, 7, 2, 4, 2]])
print(patient2[0])

0

In [ ]:
```

```
score_random_forest = round(accuracy_score(y_pred_rand_classifier,y_test)*100,2)

print("The accuracy score achieved using Random Forest is: " + str(score_random_forest)+ " %")
```

```
The accuracy score achieved using Random Forest is: 96.77 %
```

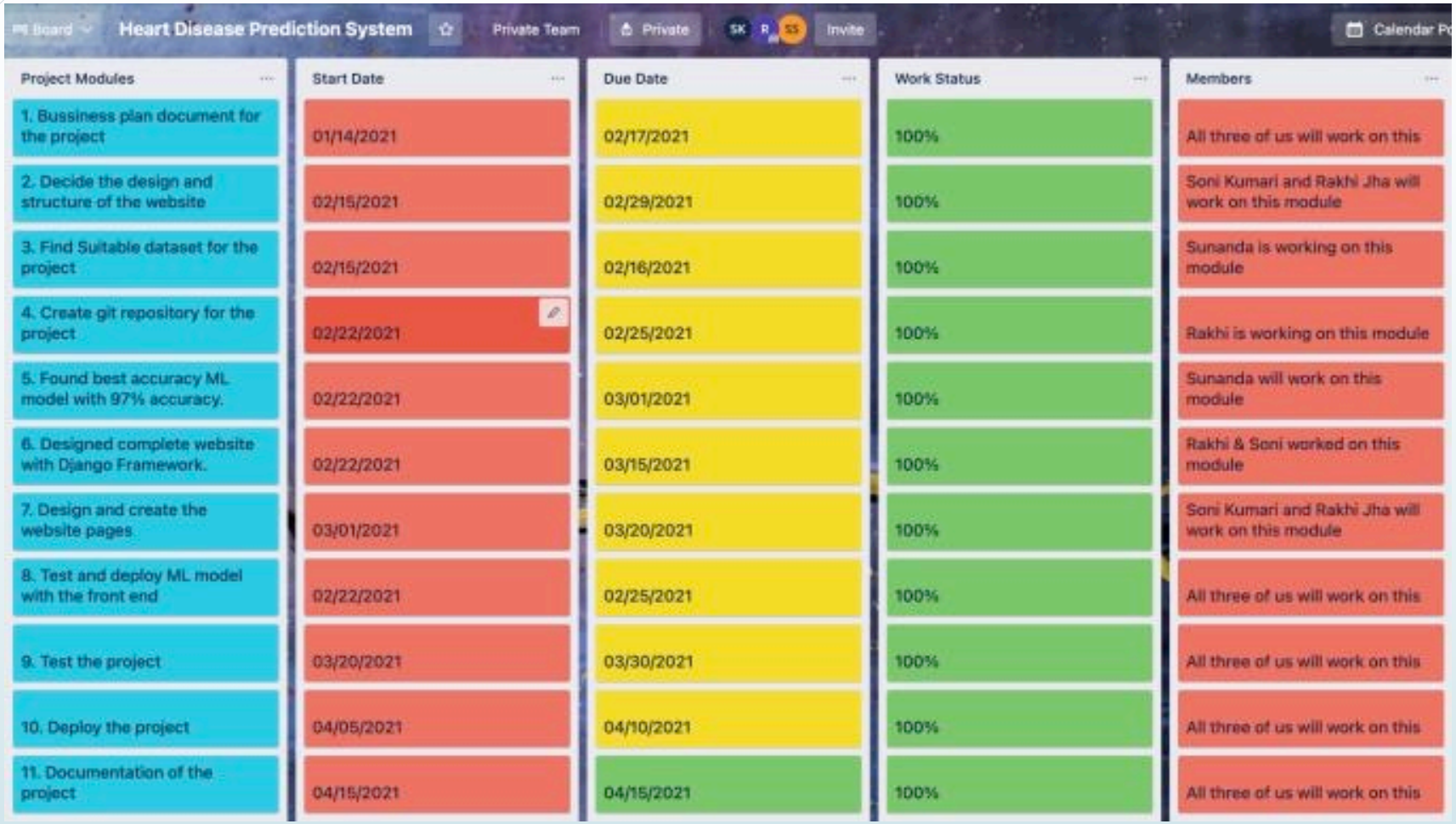
```
patient1 = rand_classifier.predict([[36, 0, 1, 135, 208, 0, 0, 171, 0, 1.5, 2, 0, 2]])
patient1
```

```
array([1])
```

```
patient2 = rand_classifier.predict([[22, 1, 0, 98, 130, 1, 1, 111, 1, 0.5, 1, 1, 1]])
patient2
```

```
array([0])
```

Development Status



Project Modules	Start Date	Due Date	Work Status	Members
1. Bussiness plan document for the project	01/14/2021	02/17/2021	100%	All three of us will work on this
2. Decide the design and structure of the website	02/15/2021	02/29/2021	100%	Soni Kumari and Rakhi Jha will work on this module
3. Find Suitable dataset for the project	02/15/2021	02/16/2021	100%	Sunanda is working on this module
4. Create git repository for the project	02/22/2021	02/25/2021	100%	Rakhi is working on this module
5. Found best accuracy ML model with 97% accuracy.	02/22/2021	03/01/2021	100%	Sunanda will work on this module
6. Designed complete website with Django Framework.	02/22/2021	03/15/2021	100%	Rakhi & Soni worked on this module
7. Design and create the website pages	03/01/2021	03/20/2021	100%	Soni Kumari and Rakhi Jha will work on this module
8. Test and deploy ML model with the front end	02/22/2021	02/25/2021	100%	All three of us will work on this
9. Test the project	03/20/2021	03/30/2021	100%	All three of us will work on this
10. Deploy the project	04/05/2021	04/10/2021	100%	All three of us will work on this
11. Documentation of the project	04/15/2021	04/15/2021	100%	All three of us will work on this

https://github.com/iamsonisingh/Heart-disease-prediction-Web-Application

Keys to Success

- Heart disease has been identified as one of the largest causes of death even in developed countries. One of the reasons is that the risks are either not identified, or they are identified only at a later stage.
- Machine learning techniques can be useful for overcoming this problem and to predict risk at an early stage.
- We have tried and tested few ML technique with given dataset and chose to go Random Forest Classifier which gave us the best accuracy of 97%.
- Getting accurate date set is the most important to get accuracy and we have use
- We have used latest technologies to develop this website to provide users reliable, Portable.

This project is a non-profit application, it will encourage more people to use this application and get benefit of free healthcare.

Competitive Advantages

The competitive advantage must be sustainable in order to create long-term viability. While working on this project we kept in mind few factors:

- Quality
- Low cost
- Accessible from anywhere in the world
- Advance technologies used and mobile optimized
- Customer service

This framework can have multiple apps inside one app, means we can install multiple Machine learning algorithms in one application. It treats every algorithm as different module of the project.

Business Strategy

We're on a mission to make healthcare smarter

Our only mission is to promote health awareness within society. We will Strength this mission through community-based partnerships.

Here are few highlights of our business strategies:

- Give more in Less
- Innovative product with best accuracy
- Improve customer service
- Concentrating on digital marketing to reach out to the larger audience.
- Absolutely no cost so it will motivate people to use it to keep track of their heath condition.
- Supports newer technology and newer **web** coding standards and its a **mobile optimized website.**

Financial Goals

- As of now, we developed this website as non-profitable application.
- Our web application built purely to help people & community to get better health care with minimal cost.
- This will encourage more people to use this application and get to know their heart condition and get aware of false alarm of heart attack.
- We will encourage people to volunteer and join our community.

We believe Together, we can impact our community.

In future, When we will successfully deploy our website to the larger audience and get their positive response. We can add billing page where people can contribute to encourage our effort and for further enhancement.

Future Work

- Instead of entering the medical info manually we can add sensors to take the input from wearable devices.
- We can provide list of cardiologists to consult in case of emergency and even will link with map so they can choose the nearest available one.
- We can contact best heartcare advisor to get tips on daily exercise, yoga and intake nutrition.
- We can our perdition accuracy using a larger dataset as compared to the one used in this analysis which will help to provide better results and help health professionals in predicting the heart disease effectively and efficiently.

The End