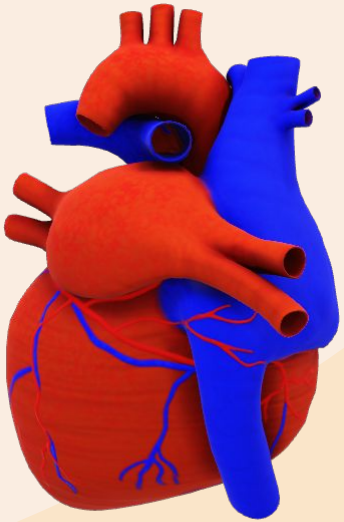
Heart Disease Prediction



Mentored by

Dr. Sanjay Saxena

Submitted by

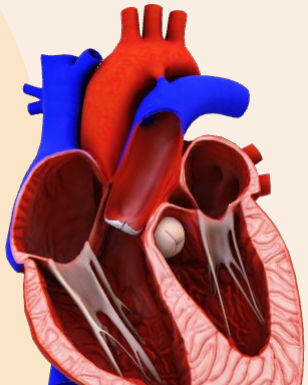
AKASH PARIDA	B420005
ANIMESH PADHY	B420007
TANUJ MANUPRIT XALXO	B420056

Problem Statement

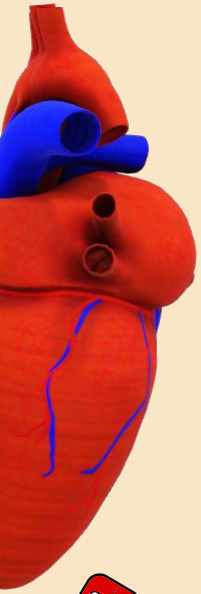
It is difficult to identify heart disease of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors.

Diagnosing heart disease is a difficult task, but it can be made more effective by providing automated predictions about a patient's heart condition, enabling more targeted treatment.

A machine learning approach to predict the presence of cardiovascular diseases in patients based on major health data.



Motivation



QoS

Ensuring high accuracy and reliability, our machine learning model for heart disease prediction provides trustworthy and precise results.

Diagnosing

Diagnosing patients correctly by employing appropriate computer based information and decision support system

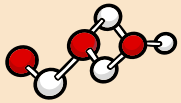
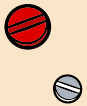
Background of Work

Leveraging Machine Learning for Heart Disease Prediction

- Heart disease: a prevalent and serious global health condition.
- Importance of early detection and accurate prediction for improved patient outcomes.
- Machine learning techniques revolutionizing healthcare.
- Widely used for developing predictive models across medical conditions, including heart disease.

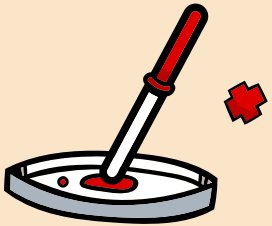
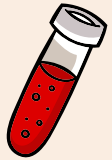
Heart Disease Prediction Web Application

- Heart disease prediction web application for risk assessment.
- User-friendly interface empowering individuals to assess their heart disease risk.
- Utilizes machine learning models trained on historical data.
- Considers various risk factors: age, gender, blood pressure, cholesterol levels, lifestyle habits.

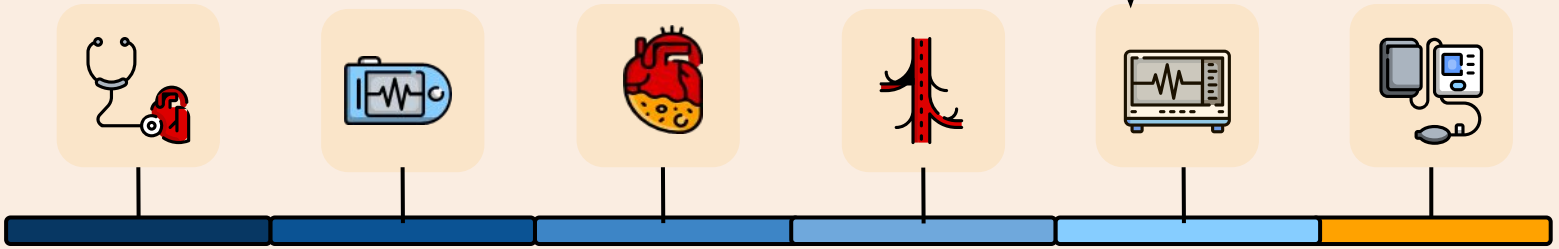




Machine Learning Integration:

- Flask, a lightweight and versatile Python web framework, facilitates the development of interactive heart disease prediction web applications.
- Flask web application integrates with the underlying machine learning model.
- Predictions and recommendations personalized based on user data.
- Users can input health information, interact with the machine learning model, and receive personalized predictions or recommendations for proactive heart health management.

WORKFLOW


Patient Data
Data Collection
Data Preprossing
Train-test split
Model Selection & Training
Trained ML Model
Model Evaluation
Prediction

Model Development

Data Collection:

Dataset with 14 attributes is collected from Kaggle

Data Preprocessing:

Detected and addressed any missing values in the dataset by employing imputation or dropping.

Standardized and rescaled the features by using methods such as standardization or min-max scaling to ensure their uniform magnitude.

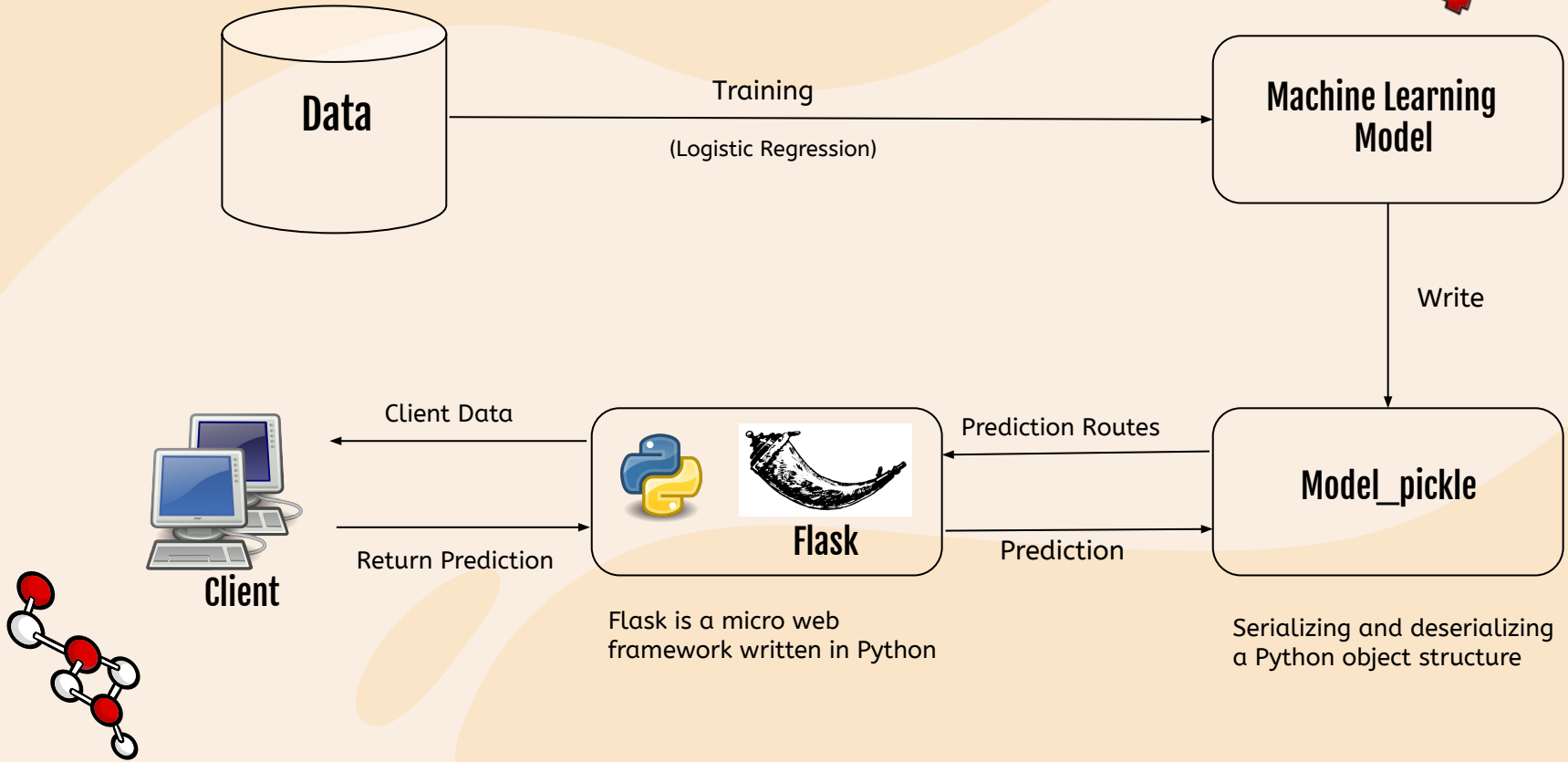
Model Selection & Training:

Selected ML model is trained using the preprocessed dataset. Our goal was to optimize the model's parameters in order to minimize prediction error. To achieve this, we employed techniques such as gradient descent or optimization algorithms.

Model Evaluation:

Calculated performance metrics such as accuracy, precision, recall, F1 score, or area under the ROC curve (AUC-ROC) to assess how well the model predicts heart disease.

Model Deployment



Dataset

KAGGLE

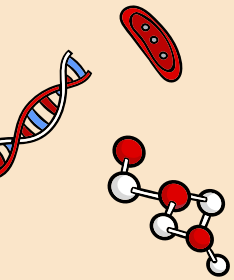
an online community of data scientists and machine learning engineers

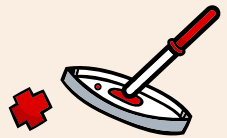
UCI DATASETS

With 14 health attributes

FINAL DATASET

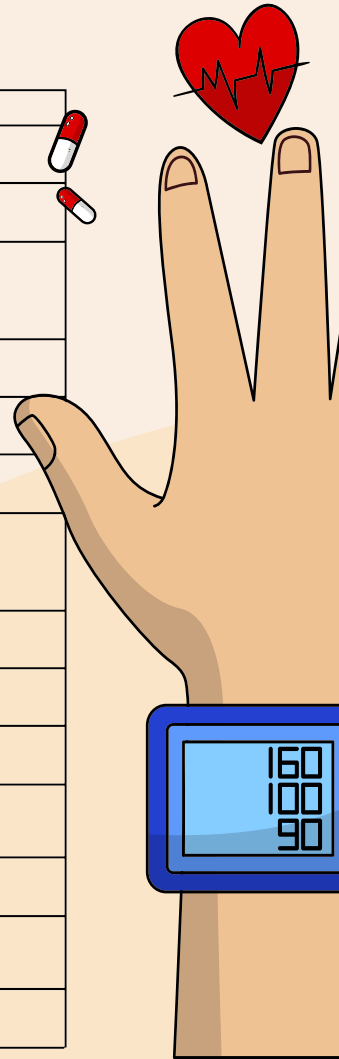
300+ observations





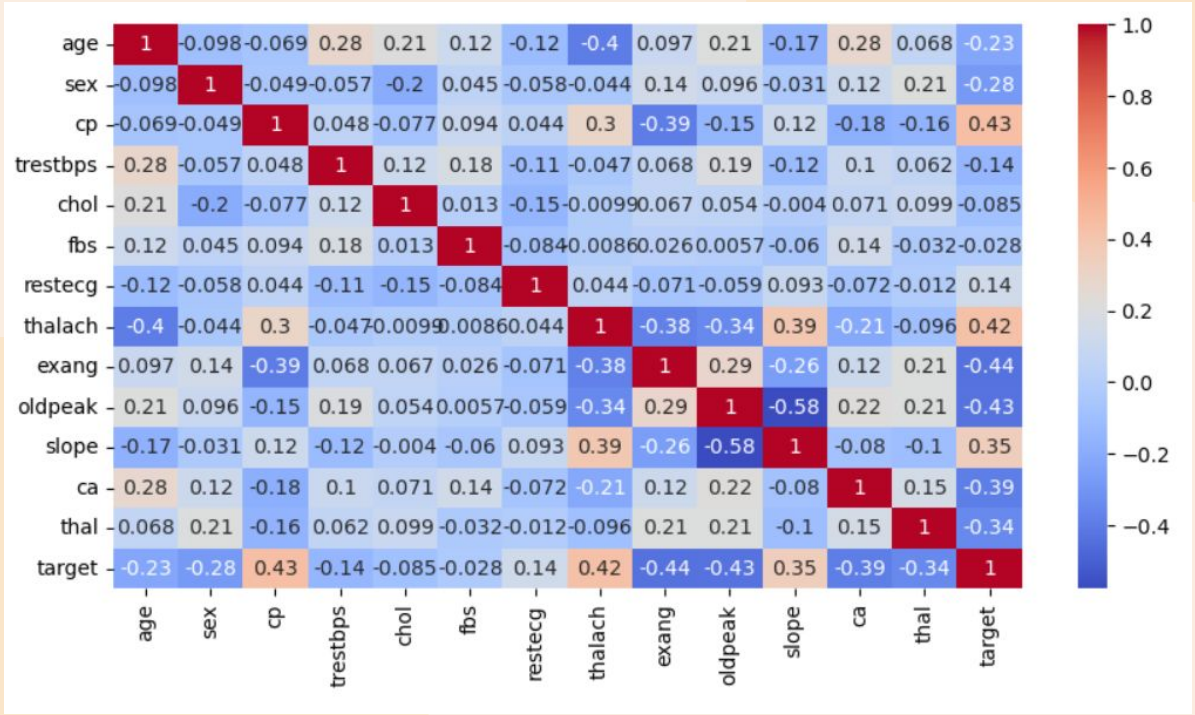
Data Set Description

Data element	Description	Type	Range	Remarks
Age	-	Num[a]	29-77	Average is 54.37
Sex	-	Bi[b]	0: Female 1: Male	32% Female 68% Male
Cp	Chest pain level	Nom[c]	0/1/2/3 0: Asymptotic 2: non-anginal pain 3: Typical angina	Majority have 0 pain
Trestbps	Rest blood pressure	Num	94-200	Average is 131.6
Chol	Cholesterol level	Num	126-564	Average is 246.3
Fbs	Fasting blood sugar level	Bi	0: Level below 120 1: Level above 120	-
Restecg	Resting electrocardiographic results	Nom	0/1/2 0: Showing probable or definite left ventricular hypertrophy. 2: Abnormal	-
Thalach	Maximum heart rate achieved	Num	71-202	-
Exang	Exercise induced angina	Bi	0: None 1: Produced	-
Oldpeak	ST depression induced by exercise relative to rest	Num	0-6.2	Right skewed data, majority of population is between 0 and 0.5
Slope	The slope of the peak exercise ST segment	Nom	0: Unsloping 1: Flat 2: Down-sloping	-
Ca	Number of major vessels	Nom	0/1/2/3/4	-
Thal	Defect type	Nom	1: Fixed defect 2: Normal 3: Reversable defect	There is one outlier of category 0
Target	Diagnosis of heart disease	Bi	0: No disease 1: Disease	-



Correlation Analysis

Visualizing Relationships between Features and Target Variable



Working Environment

Programming Language:
Python - a versatile language for data analysis, machine learning, and web development.

Machine Learning Libraries:
Utilizing popular Python libraries like scikit-learn and pandas for training and evaluating the model.

Dependencies:
Managing project dependencies using a package manager like pip, including Flask, numpy, Gunicorn, scikit-learn, and pandas

Development Environment:
Utilizing Visual Studio Code for coding, debugging, and project management.

Web Framework:
Flask - a lightweight and flexible Python web framework for building web applications.

HTML/CSS:
Creating the user interface with HTML for structure and CSS for styling.

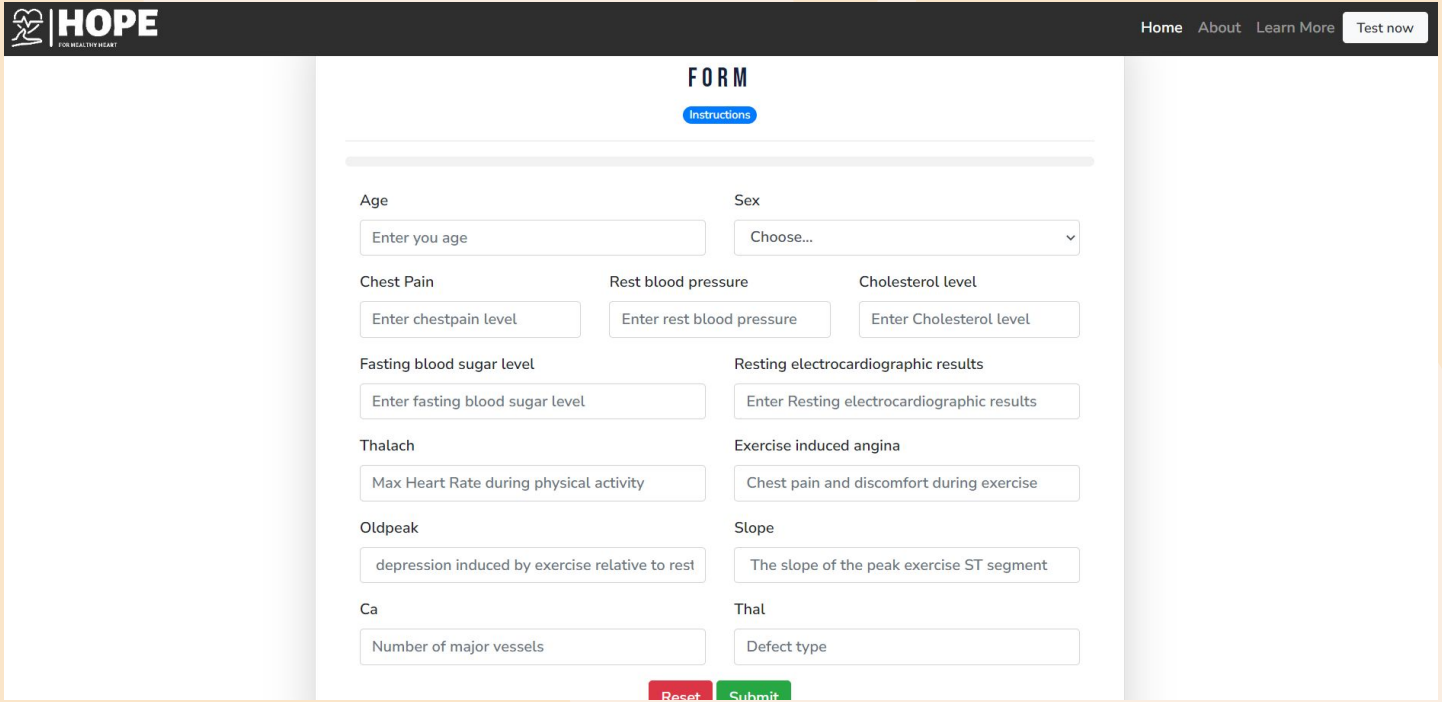
JavaScript:
Enhancing interactivity and responsiveness on the web pages.

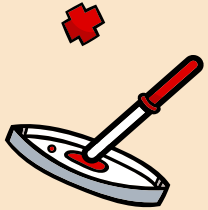
Deployment:
Deploying the web application on a cloud platform, such as Render.

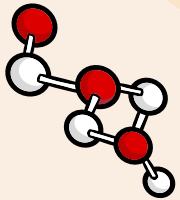
Home Page


HOPE
FOR HEALTHY HEART
Home
About
Learn More
Test now
EMPOWERING PREVENTION
Predicting Heart Disease for Proactive Care
HOPE for Healthy Heart.

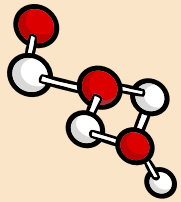
Interacting with the Heart Disease Prediction Model

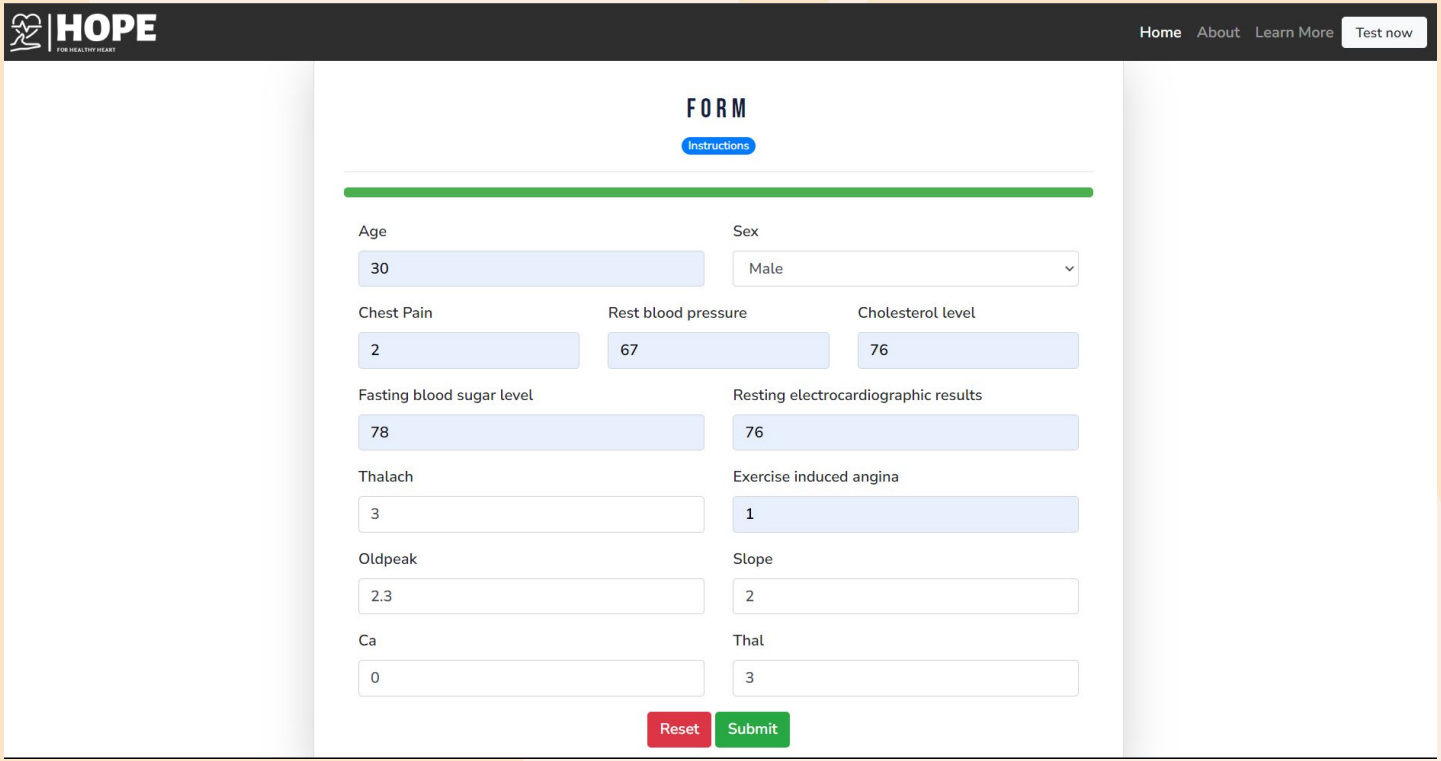


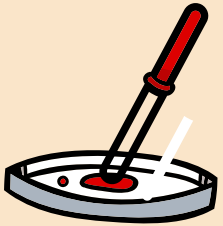




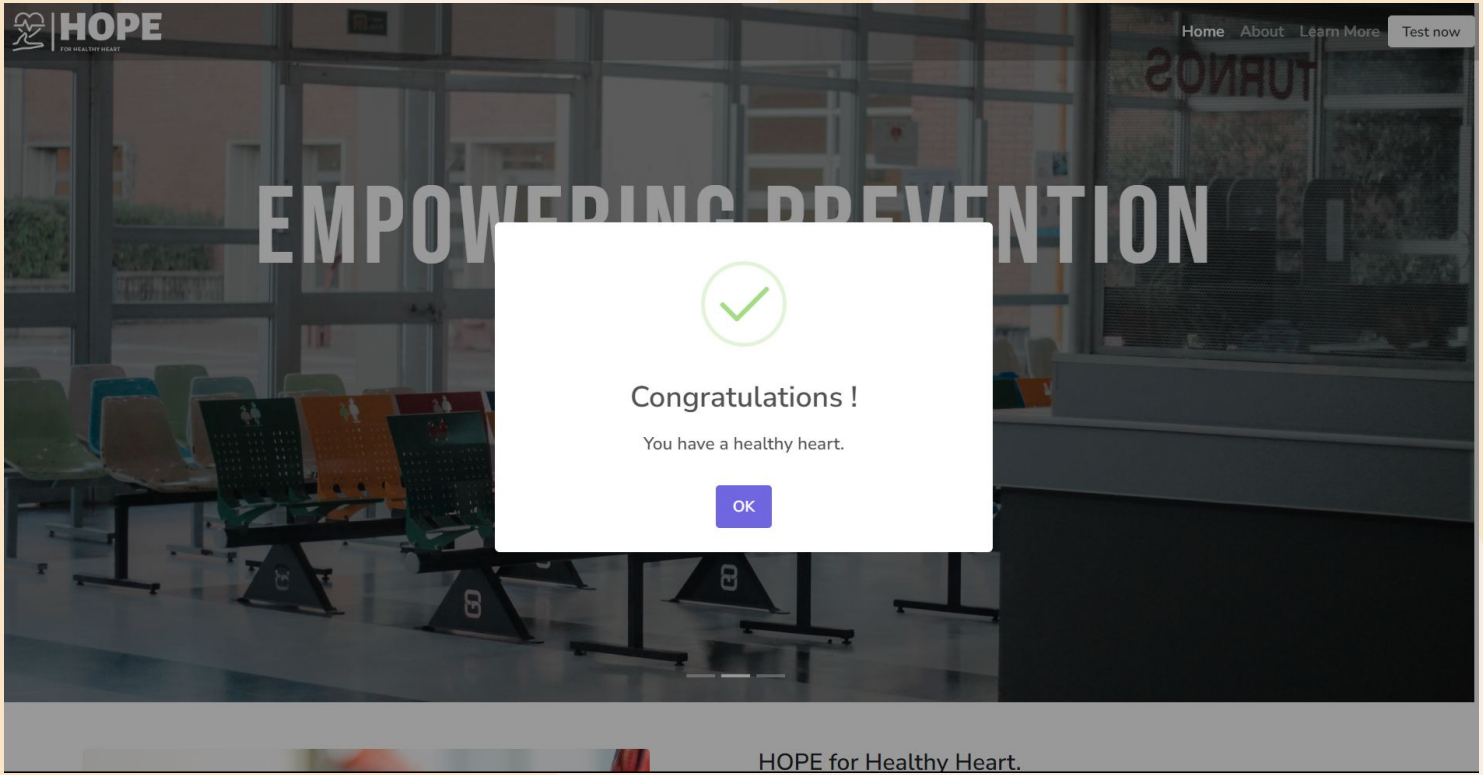
Capturing User Data





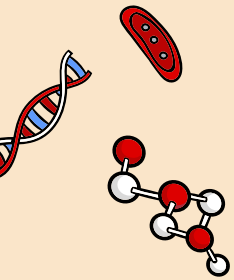


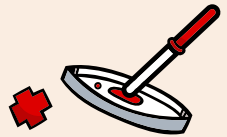
Visualization of Output



Discussions,
Evaluating the Performance

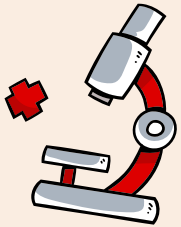
- The proposed system was evaluated using the Logistic Regression machine learning technique.
- Logistic Regression is known for its interpretability and efficiency in classification tasks.
- With an accuracy of 80%, the Logistic Regression model demonstrated excellent performance in predicting outcomes or classifications for the given data.

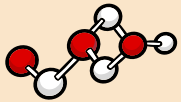
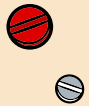


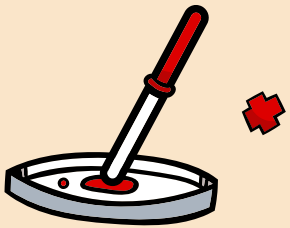
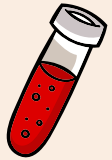


Future Enhancements

- In the future, we can further enhance the accuracy of Logistic Regression by applying a genetic algorithm to reduce the amount of data required and obtain an optimal subset of attributes for predicting heart disease.
- This automation of heart disease prediction utilizes real-time data obtained from healthcare organizations and agencies, making use of big data.
- By utilizing this data, real-time investigations of patients can be conducted



Conclusion

Our work focuses on utilizing specific health measurements to predict the presence of heart disease in patients. Through the application of advanced machine learning techniques, we aim to enhance early detection of heart disease.