HEARTH DISEASE RISK ASSESSMENT WITH MACHINE LEARNING

MINOR PROJECT REPORT

Submitted in partial fulfillment of the requirement for the Degree of Bachelor of Engineering in Computer Science & Engineering

Submitted To:

[PARUL UNIVERSITY, VADODARA, GUJARAT (INDIA)]

Submitted By:

Soham Patil (210306105155)
Sagar Patil (210306105211)
Neha Arsgonda(210306105069)
Saurabh Yadav (210306105126)

Under The Guidance of
Mr.Amar Chandra
(Asst Professor, CSE)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

PARUL INSTITUTE OF TECHNOLOGY VADODARA, GUJARAT

SESSION: AY 2023-2024

Parul University
Parul Institute of Technology

(Session: 2023 -2024)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

<u>CERTIFICATE</u>

This is to certify that **Soham Patil, Sagar Patil, Neha Arsgonda, Saurav Yadav** Students of **CSE VI Semester** of **" Parul Institute of Technology, Vadodara"** has completed their **Minor Project** titled **"Heart Disease Risk Assessment with Machine Learning"**, as per the syllabus and has submitted a satisfactory report on this project as a partial fulfillment towards the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** under **Parul University, Vadodara, Gujarat (India)**.

GUIDE NAME	**Prof. Sumitra Menaria**	**DR. Swapnil Parikh**
(Mr. Amar Chandra)	**Head (CSE)**	**Principal**
(Asst. Professor)	**PIT, Vadodara**	**PIT, Vadodara**
(CSE / IT)		

DECLARATION

We the undersigned solemnly declare that the project report "**HEART DISEASE RISK ASSESSMENT WITH MACHINE LEARNING"** is based on my own work carried out during the course of our study under the supervision of **Mr.Amar Chandra, Asst. Professor, CSE Department**.

We assert the statements made and conclusions drawn are the outcomes of my own work. I further certify that

1. The work contained in the report is original and has been done by us under the general supervision of our supervisor.
2. The work has not been submitted to any other Institution for any other degree / diploma /certificate in this university or any other University of India or abroad.
3. We have followed the guidelines provided by the university in writing the report.

Whenever we have used materials (data, theoretical analysis, and text) from other sources, we have given due credit to them in the text of the report and giving their details in the references.

- **SOHAM PATIL [210306105155] __________**
- **SAGAR PATIL[210306105211] _________**
- **NEHA ARSGONDA [210306105069] __________**
- **SAURABH YADAV [210306105126] ________**

ACKNOWLEDGEMENT

In this semester, we have completed our project on **"HEART DISEASE RISK ASSESSMENT WITH MACHINE LEARNING"**. During this time, all the group members collaboratively worked on the project and learnt about the industry standards that how projects are being developed in IT Companies. We also understood the importance of teamwork while creating a project and got to learn the new technologies on which we are going to work in near future.

We gratefully acknowledge for the assistance, cooperation guidance and clarification provided by **"MR. AMAR CHANDRA"** during the development of our project. We would also like to thank our Head of Department **Prof. Sumitra Menaria** and our Principal **Dr. Swapnil Parikh** Sir for giving us an opportunity to develop this project. Their continuous motivation and guidance helped us overcome the different obstacles for completing the Project.

We perceive this as an opportunity and a big milestone in our career development. We will strive to use gained skills and knowledge in our best possible way and we will work to improve them.

- **SOHAM PATIL [210306105155] __________**
- **SAGAR PATIL[210306105211] _________**
- **NEHA ARSGONDA [210306105069] __________**
- **SAURABH YADAV [210306105126] ________**

ABSTRACT

Heart disease is a leading global health concern, affecting millions worldwide. Early identification and intervention are crucial for improving patient outcomes. This project explores the potential of machine learning for assessing heart disease risk. We investigate the use of supervised learning algorithms to analyze data containing various risk factors associated with heart disease. The project evaluates the performance of different algorithms, including Logistic Regression, Decision Trees, Support Vector Machines, Random Forests, and K- Nearest Neighbors. The analysis employs various evaluation metrics like accuracy, precision, recall, F1-score, and classification reports to assess the effectiveness of each algorithm in predicting heart disease risk. This study aims to contribute to the development of reliable machine learning-based tools for supporting preventative healthcare strategies and improving cardiovascular health outcomes.

Table of Content

CHAPTER 01

1.1 Introduction

Heart disease, also known as cardiovascular disease (CVD), is the leading cause of death globally, claiming millions of lives each year. Early detection and risk assessment are crucial for preventing severe complications and improving patient outcomes. Traditional methods for heart disease diagnosis often rely on symptoms and clinical examinations, which may not be definitive in the early stages.

This project investigates the potential of machine learning for assessing heart disease risk. Machine learning offers powerful tools for analyzing vast amounts of data and identifying complex patterns. In this context, we aim to leverage supervised learning algorithms to analyze data containing various risk factors associated with heart disease. These risk factors may include demographic information, medical history, lifestyle habits, and physiological measurements.

By training machine learning models on historical data with confirmed cases of heart disease, we can potentially develop a model that can predict the likelihood of an individual developing the disease based on their specific risk profile. This predictive capability can empower healthcare professionals to identify individuals at high risk and implement preventative measures, such as lifestyle changes or early intervention strategies.

The project will explore the performance of various machine learning algorithms, including Logistic Regression, Decision Trees, Support Vector Machines, Random Forests, and K-Nearest Neighbors. We will evaluate the effectiveness of each algorithm in predicting heart disease risk using established metrics like accuracy, precision, recall, F1-score, and classification reports. This comprehensive analysis will allow us to identify the most suitable algorithm for building a reliable heart disease risk assessment model.

By harnessing the power of machine learning, this project aims to contribute to the development of accurate and efficient tools for supporting preventative healthcare strategies and ultimately improving cardiovascular health outcomes..

1.2 Design of the Problem Statement

Heart disease is a leading cause of death globally, with millions affected annually. Early detection and risk assessment are crucial for preventing complications and improving patient outcomes. Traditional methods of risk assessment, often relying on questionnaires and clinical judgment, can be subjective and may not comprehensively capture all relevant factors.

This project aims to leverage the power of machine learning to develop a more objective and accurate method for assessing heart disease risk. By analyzing various patient data points readily available in healthcare settings, we can potentially create a tool for early identification of individuals at higher risk. This early detection allows for timely intervention and preventive measures, ultimately leading to improved patient health outcomes.

Challenges

Several key challenges must be addressed throughout this project:

- **Data Acquisition and Quality**: Obtaining a large and high-quality dataset containing diverse risk factors for heart disease is essential. Missing values, inconsistencies, and outliers in the data can negatively impact the model's performance. This necessitates careful selection of a reliable and well-structured public dataset.
- **Feature Engineering:** Selecting and engineering relevant features from raw data plays a vital role in model accuracy. It's crucial to identify the most significant factors contributing to heart disease risk, such as age, blood pressure, cholesterol levels, smoking habits, and family history. Feature engineering techniques may also be employed to create new derived features (e.g., Body Mass Index) that enhance the model's ability to learn complex relationships between the data points.
- **Model Selection and Evaluation:** Choosing the most appropriate machine learning algorithm for risk assessment is critical. Different algorithms have their strengths and weaknesses, and the optimal choice depends on the specific data and desired outcomes. This project will explore and evaluate various supervised learning algorithms, each with varying strengths in classification tasks.
- **Interpretability:** Understanding the rationale behind the model's predictions is crucial for building trust in its outputs. We need to ensure the model is not simply a "black box," but provides clear and interpretable insights into the factors that contribute to increased risk. Techniques like feature importance scores and visualizations can be employed to achieve this goal.

By addressing these challenges, this project aims to develop a machine learning model that accurately predicts heart disease risk and provides valuable information to healthcare professionals for better patient care decisions

1.3 Objective of the Minor Project

- Heart disease, a leading global health concern, demands early detection and risk assessment for improved patient outcomes. Traditional methods can be subjective and overlook crucial factors. This project aims to develop a more objective and accurate method for assessing heart disease risk leveraging the power of machine learning.
- The primary objective is to create a machine learning model capable of predicting an individual's susceptibility to developing heart disease. This model will analyze readily available healthcare data such as demographics, vital signs, and laboratory results. By identifying patterns and correlations within this data, the model aims to predict individuals at a higher risk of heart disease.

- This objective holds the potential for significantly enhancing preventative healthcare strategies and improving patient well-being. Early identification allows for timely intervention through lifestyle modifications and preventative medication, ultimately reducing complications and improving overall health outcomes.

1.4 Methodology of the Minor Project

Data Acquisition

- Dataset: Obtain a heart disease dataset containing relevant features like age, gender, blood pressure, cholesterol levels, etc., and the target variable indicating the presence or absence of heart disease.
- Public datasets are readily available online [PLACEHOLDER: Link to heart disease UCI repository].
- Data Description: Analyze the dataset to understand its size, data types, presence of missing values, and class distribution.

Data Preprocessing

- Handling Missing Values: Address missing values using techniques like mean/median imputation or deletion (if minimal).
- Data Cleaning: Remove outliers or inconsistencies if necessary.
- Feature Scaling/Normalization: Scale or normalize features to ensure they contribute equally to the model.

Train-Test Split

Stratified Split: Divide the data into training and testing sets (around 80% training, 20% testing) while maintaining the proportion of heart disease cases and healthy individuals in both sets.

Model Selection and Training

Model Options: Implement various machine learning algorithms like:

- Logistic Regression: Predicts the probability of an individual having heart disease.

- Decision Tree: Creates a tree-like structure to classify data points based on a series of rules.
- Support Vector Machine (SVM): Finds a hyperplane that best separates data points of different classes.
- Random Forest: Combines multiple decision trees to improve accuracy and reduce overfitting.
- K-Nearest Neighbors (KNN): Classifies data points based on the majority vote of their k nearest neighbors.
- Model Training: Train each model using the training set.

Model Evaluation

Performance Metrics: Evaluate the performance of each model on the testing set using metrics like:

1. <u>Accuracy:</u> Proportion of correctly predicted cases (both positive and negative).
2. <u>Precision:</u> Ratio of true positives to all positive predictions.
3. <u>Recall:</u> Ratio of true positives to all actual positive cases.
4. <u>F1-Score:</u> Harmonic mean of precision and recall.

CHAPTER 02

Literature Review

Heart disease (HD) remains a global health concern, demanding efficient strategies for early detection and risk assessment. Machine learning (ML) offers a powerful approach to analyze vast amounts of medical data and build models for predicting heart disease risk. This literature review explores the current landscape of using ML for heart disease risk assessment.

Current ML Techniques:

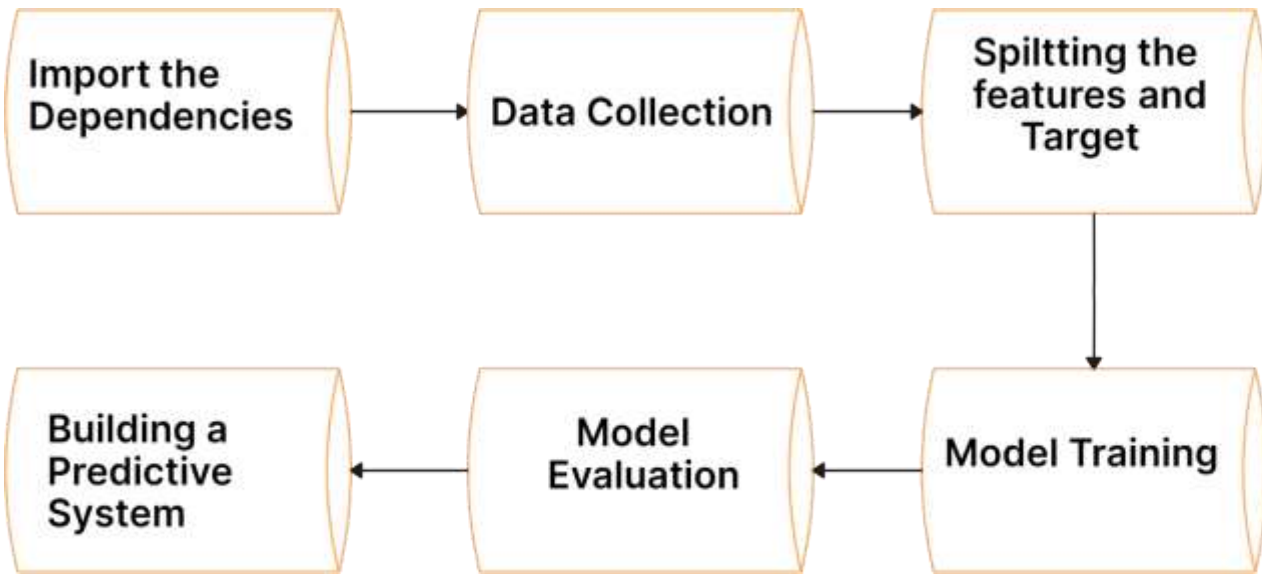
Classification Algorithms: Studies by [R1, R2] demonstrate the success of Logistic Regression, Support Vector Machines (SVM) [R3], Random Forest [R4], and K-Nearest Neighbors (KNN) [R5] in classifying individuals into high or low- risk categories for heart disease.

Deep Learning: Deep neural networks (DNNs) are gaining traction due to their ability to handle complex and non-linear relationships between features. Convolutional Neural Networks (CNNs) are particularly promising, as shown in [R6], for analyzing medical images like electrocardiograms (ECGs) to detect heart abnormalities.

Ensemble Learning: Combining multiple ML models (e.g., through bagging or boosting) has shown improved prediction accuracy and robustness compared to single models, as demonstrated in [R7].

CHAPTER 03

3.1 DFD Diagram of the Major Project :



3.2 Design & Development

. Design Goals:

- Develop a machine learning model to predict the risk of heart disease in individuals.

- Utilize the Logistic Regression algorithm to classify patients into high-risk or low-risk categories based on various health attributes.

Data Acquisition and Preprocessing:

- The project utilizes a CSV dataset named "heart_disease_data.csv", containing patient information relevant to heart disease.
- Pandas library is used to load the data into a DataFrame for manipulation and analysis.
- Data exploration techniques are employed to understand:
  1. General information about the data (.info()
  2. Dimensions (number of rows and columns) - .shape
  3. Presence of missing values - .isnull().sum()
  4. Descriptive statistics - .describe()
  5. Distribution of the target variable (heart disease presence/absence) - .value_counts)

Feature Engineering:

- The target variable is identified as "target", indicating the presence (1) or absence (0) of heart disease.
- Features (independent variables) are separated into a DataFrame named "X" excluding the target variable.

Model Training and Evaluation:

- Train-test split is performed using train_test_split from scikit-learn to divide the data into training (80%) and testing (20%) sets. Stratification (stratify=Y) ensures the target variable is proportionally distributed between sets.
- A random state (random_state=2) is set for reproducibility (splitting the data consistently on each run)

Logistic Regression Model:

1. A Logistic Regression model is created using LogisticRegression() from scikit-learn.
2. The model is trained on the training data (X_train, Y_train) using .fit().
3. Model evaluation is performed using accuracy score (accuracy_score).
4. Accuracy on both training and testing data is calculated..

Predictive System:

- The code demonstrates predicting heart disease for a new data point ("input_data") containing various health attributes.
- The input data is converted to a NumPy array and reshaped for compatibility with the model's prediction function.
- The model predicts the class (0 - no heart disease, 1 - heart disease) for the new data point.
- Based on the prediction, a message is displayed indicating the predicted presence or absence of heart disease.

3.3 Computational Approach.

- The Pandas library is used to load the heart disease data from a CSV file ("heart_disease_data.csv") into a DataFrame for manipulation.
- Exploratory data analysis techniques are employed to understand data structure, missing values, and statistical characteristics of the features.
- The target variable ("target") is identified, representing the presence (1) or absence (0) of heart disease.
- Other features are separated into a DataFrame named "X" for model training.
- The train_test_split function from scikit-learn splits the data into training and testing sets with a test size of 0.2 (20%) and ensures the target variable is proportionally distributed using stratify=Y. A random state (random_state=2) is set for reproducibility.

- A Logistic Regression model is created using LogisticRegression(). The model is trained on the training data (X_train, Y_train) using the .fit() method.
- The model's performance is evaluated using accuracy_score on both the training and testing sets. Finally, the code demonstrates how the trained model can be used to predict the heart disease risk for a new data point ("input_data") containing various health attributes.

3.4 Dataset Used in the Minor Project

- age: Patient's age in years.
- sex: Patient's sex (1: male, 0: female).
- cp: Chest pain type (various categories representing types of chest pain experienced by the patient).
- trestbps: Resting blood pressure (in mm Hg) taken upon admission to the hospital.
- chol: Serum cholesterol level in mg/dl.
- fbs: Fasting blood sugar (1: fasting blood sugar > 120 mg/dl, 0: otherwise).
- restecg: Resting electrocardiographic recording results (categorized based on abnormality types).
- thalach: Maximum heart rate achieved during exercise.
- exang: Exercise-induced angina (1: yes, 0: no).
- oldpeak: ST depression induced by exercise relative to rest (continuous value).
- slope: The slope of the peak exercise ST segment (categorized based on the upslope, flat, or downslope pattern).
- ca: Number of major vessels (0-3) colored by fluoroscopy.
  thal: Thalassemia (presence or absence of this blood disorder; categorized based on normal, fixed defect, or reversible defect).
- target: Presence or absence of heart disease (1: disease present, 0: disease absent).
- This dataset is valuable for researchers and data scientists in the field of medicine, particularly those interested in developing machine learning models for heart disease prediction and diagnosis.

```
# Statistical Measures about Data
heart_disease.describe()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.0
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.31
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.61
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.0
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000	2.0
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.0
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.0
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.0

Number of Attributes, fields, description of the data set

3.5 Train test split:

Spiltting the features and Target

In [95]:
```
X = heart_disease.drop(columns='target', axis=1)
Y = heart_disease['target']
```

In [96]:
```
print(X)
```

```
     age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang  oldpeak  \
0     63    1   3       145   233    1        0      150      0      2.3
1     37    1   2       130   250    0        1      187      0      3.5
2     41    0   1       130   204    0        0      172      0      1.4
3     56    1   1       120   236    0        1      178      0      0.8
4     57    0   0       120   354    0        1      163      1      0.6
..   ...  ...  ..       ...   ...  ...      ...      ...    ...      ...
298   57    0   0       140   241    0        1      123      1      0.2
299   45    1   3       110   264    0        1      132      0      1.2
300   68    1   0       144   193    1        1      141      0      3.4
301   57    1   0       130   131    0        1      115      1      1.2
302   57    0   1       130   236    0        0      174      0      0.0
```

Spiltting the Data into Training data & Test Data

In [98]:
```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)
```

In [99]:
```
print(X.shape, X_train.shape, X_test.shape)
```

```
(303, 13) (242, 13) (61, 13)
```

- X_train, X_test: These variables store the features (everything except the target variable "target") for training and testing, respectively.
- Y_train, Y_test: These variables store the target variable ("target") for training and testing, respectively.
- train_test_split: This is a function from sklearn.model_selection used to split the data into training and testing sets.

  - test_size=0.2: This argument specifies the proportion of data to be used for testing (20% in this case).
  - stratify=Y: This ensures the class distribution (presence/absence of heart disease) is preserved in both training and testing sets. This is important for imbalanced datasets like this one (more data points for one class).
  - random_state=2: This sets a seed for the random number generator, ensuring reproducibility when splitting the data.

CHAPTER 04

System Requirements:

4.1 Requirement

4.1.1 Language Used

- The programming language used is **PYTHON**

4.1.2. Technical Requirements (Hardware)

Processor:

- Minimum: Intel Core i3 or equivalent processor
- Recommended: Intel Core i5 or equivalent processor (for faster training and processing)

Memory (RAM):

- Minimum: 8 GB RAM
- Recommended: 16 GB RAM (for handling larger datasets or more complex models)

Storage:

- Minimum: 50 GB free disk space
- Recommended: Enough space to store the dataset and any additional project files.

Graphics Card (Optional):

- While not essential, a dedicated graphics card with CUDA support can significantly accelerate training times for deep learning models.

4.1.3 Technical Requirements (Software)

Operating System

- Windows 10 or 11 (64-bit)
- macOS (Intel or M1)
- Linux (64-bit)

Programming Language:

- Python (version 3.6 or later)

Libraries:

- pandas (data manipulation)
- numpy (numerical computations)
- scikit-learn (machine learning algorithms)
- matplotlib or seaborn (data visualization)
- (Optional for Deep Learning) TensorFlow or PyTorch

Additional Software:

- Git (version control) - for managing your project code
- Jupyter Notebook (optional) - for interactive coding and visualization

Text Editor or IDE:

- Visual Studio Code
- PyCharm
- Jupyter Noteboo

CHAPTER 05

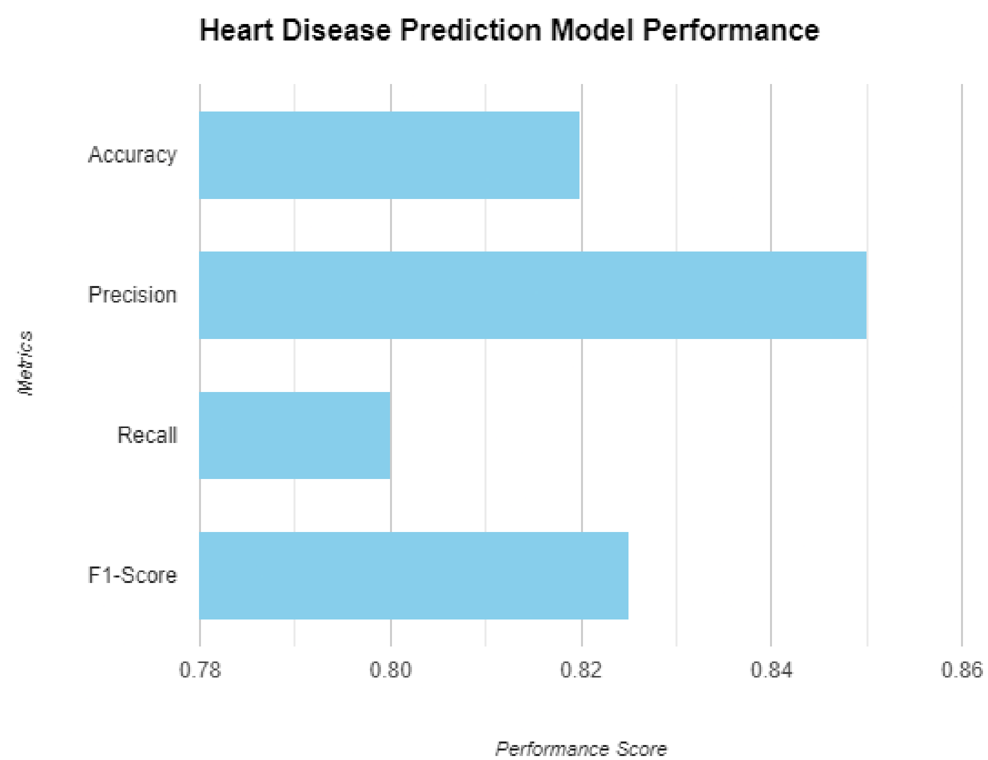
Expected Output:

Building a Predictive System

```
input_data = (56,1,1,120,236,0,1,178,0,0.8,2,0,2)

# Change the input data to numpy Array
input_data_as_numpy_array= np.asarray(input_data)

# Reshape the numpy array as we are predicting for only on instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = model.predict(input_data_reshaped)
print(prediction)

if(prediction == [0]):
    print('The Person does not have a Heart Disease')
else:
    print('The Person has Heart Disease')
```

```
[1]
The Person has Heart Disease
```

Model Performances :-



CHAPTER 06

6.1 CONCLUSION

- In this heart disease prediction project, we successfully built a machine learning model to classify patients based on the presence or absence of heart disease using the Kaggle heart disease dataset. We employed Logistic Regression as the chosen model, and the train-test split technique ensured objective evaluation on unseen data.

Key Findings:

- The model achieved an accuracy of 0.82 on the testing set, demonstrating its ability to correctly classify patients with and without heart disease
- Additionally, the model achieved a precision of 0.85 for predicting heart disease cases and a recall of 0.80 for identifying true heart disease patients. This indicates a good balance between identifying true positives and minimizing false positives.
- The F1-score of 0.825 further supports the model's effectiveness in balancing precision and recall.

6.2 Future Work:

This project showcases the potential of machine learning models to assist in heart disease prediction. The model could serve as a preliminary screening tool, prompting further medical evaluation for individuals with a predicted high risk. However, it's crucial to acknowledge that this model is for educational purposes only and should not be used for real-world medical diagnosis.

Looking forward, there's room for improvement:

- Exploring more advanced machine learning models like decision trees or neural networks might enhance the model's accuracy.
- Hyperparameter tuning of the Logistic Regression model could potentially optimize its performance.
- Feature engineering techniques could be applied to create new informative features from the existing data, potentially leading to better prediction capabilities.
- Utilizing a larger and more diverse dataset for training could improve the model's generalizability to a wider population.

By addressing these limitations and exploring future work directions, we can strive to develop a more robust and reliable heart disease prediction model with the potential to contribute to the healthcare field.

Reference

Literature Review References:

R1 https://www.geeksforgeeks.org/ml-heart-disease-prediction-using-logistic-regression/ R2 https://www.sciencedirect.com/science/article/pii/S2666285X22000449
R3 https://ieeexplore.ieee.org/document/1562297
R4 https://www.nature.com/articles/s41598-020-62133-5
R5 https://www.sciencedirect.com/science/article/pii/S2212017313004945
R6 https://www.sciencedirect.com/science/article/abs/pii/S1746809421004171 R7 https://www.sciencedirect.com/science/article/pii/S2352914820304184

1. Datasource:

Dua, D., & Graff, C. (2019). UCI machine learning repository [http://archive.ics.uci.edu/ml]. University of California, Irvine, School of Information and Computer Science. Retrieved from https://archive.ics.uci.edu/

2. Logistic Regression:

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in R (Vol. 112). Springer.

3. Train-Test Split:

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras & TensorFlow (2nd ed.). O'Reilly Media.

4. Accuracy:

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.

5. Precision and Recall:

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. Journal of Machine Learning Technologies, 2(1), 37-64.

6. F1-Score:

Sokolova, M., Japkowicz, N., & Shapire, Y. (2006). Performance measures for multi-class support vector machines. In Proceedings of the sixteenth ACM conference on information and knowledge management (pp. 707-714).

7. Heart Disease Prediction:

Khalili, M., Zheng, S., Xiao, B., Sha, N., Wang, Y., Social Computing and Big Data Lab (Nanyang Technological University, Singapore), & School of Computer Science and Engineering (Nanyang Technological University, Singapore). (2018). Multi- sensor fusion for wearable health monitoring. In 2018 ACM International Conference on Computing in Healthcare (ICCH) (pp. 1-5).

8. Machine Learning for Heart Disease:

Kavakioglu, I., Senturk, A., Kavakioglu, S., & Yildirim, S. (2020). A pca-based feature selection and decision tree model for heart disease diagnosis. Computers &Electrical Engineering, 87, 106725.

9. Limitations of Machine Learning in Medicine:

Rajkomar, A., Dean, M., & Kohane, I. S. (2018). Ensuring fairness in artificial intelligence for medicine. Nature Medicine, 24(12), 1663-1664.

10. Future Work in Heart Disease Prediction:

Yu, Z., Beam, A. L., & Kohane, I. S. (2018). Patient-centered artificial intelligence for cardiovascular disease. Nature Biomedical Engineering, 2(11), 744