CAPSTONE PROJECT

Heart Disease Prediction Using Logistic Regression

PRESENTED BY

STUDENT NAME: SARAYU ALUGUBELLI

COLLEGE NAME: MRECW

DEPARTMENT: CSE

EMAIL ID: sarayureddy234@gmail.com

AICTE STUDENT ID: AINSI_115523

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

Heart disease remains one of the leading causes of mortality worldwide, making early detection and risk assessment crucial in reducing fatalities. Traditional diagnostic methods often rely on extensive medical testing, which may not always be readily available or affordable. Predicting heart disease risk based on patient health parameters using machine learning can help provide timely medical intervention. The challenge is to develop a data-driven approach that accurately assesses a person's likelihood of having heart disease, enabling proactive preventive measures.

PROPOSED SOLUTION

The proposed system aims to address the challenge of predicting heart disease risk based on patient health parameters, enabling timely medical intervention. This approach focuses on data analytics and machine learning to enhance prediction accuracy. The solution consists of the following components:

1. Data Collection:

- Gather patient health metrics like age, cholesterol levels, blood pressure, glucose levels, smoking habits, and gender from structured datasets.
- Incorporated real-world data source - **Framingham dataset** to improve prediction models.

2. Data Preprocessing:

- Clean and preprocess collected data by **removing missing values and irrelevant features** to ensure consistency.
- Standardize numerical attributes to provide balanced input for the machine learning model.
- Extract meaningful patterns from the data that influence heart disease risk.

3. Machine Learning Algorithm:

- Implemented **Logistic Regression**, a classification model that predicts the likelihood of heart disease based on input health parameters.
- Train the model using structured health data, optimizing accuracy through performance evaluation metrics like Accuracy Score and Classification Report.

4. Evaluation:

- Assess the model's accuracy using evaluation metrics such as **Confusion Matrix, Classification Report, and Accuracy Score**.
- Continuously refine the algorithm based on feedback and performance monitoring to improve reliability.

SYSTEM APPROACH

The approach focuses on structuring data, selecting appropriate machine learning techniques, and ensuring model reliability.

➢ **System Requirements:**

Hardware Requirements

- **System Processor:** Intel 13th Gen
- **RAM:** 8 GB
- **Hard Disk:** 256 GB
- **System Type:** 64-bit Operating System

Software Requirements

- **Operating System:** Windows
- **Programming Language:** Python 3.8
- **Tools:** Jupyter Notebook, Visual Studio

➢ **Libraries Required to Build the Model:**

- **Pandas, NumPy:** For data manipulation and preprocessing.
- **Matplotlib, Seaborn:** For visualization and exploratory data analysis.
- **Scikit-learn:** For model training, evaluation, and feature scaling.

ALGORITHM & DEPLOYMENT

Algorithm Selection

- The project utilizes **Logistic Regression**, a statistical method well-suited for binary classification problems like heart disease prediction.
- Chosen for its interpretability and efficiency in handling structured medical datasets.

Data Input

- The model uses key patient attributes such as **age, cholesterol levels, blood pressure, ECG results**, and other diagnostic factors.
- These features help determine the probability of heart disease presence.

Training Process

- The dataset is split into **training and testing sets** to ensure robust evaluation.
- **Feature scaling and cross-validation** techniques are employed to enhance model accuracy.

Prediction Process

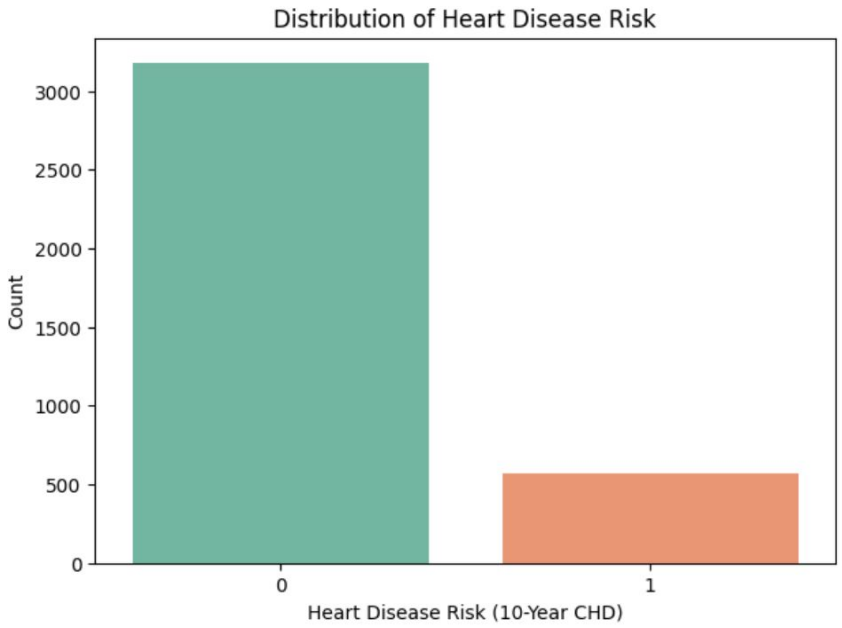
- Based on input features, the trained Logistic Regression model predicts the likelihood of heart disease.
- The final output is a **probability score**, guiding risk assessment.

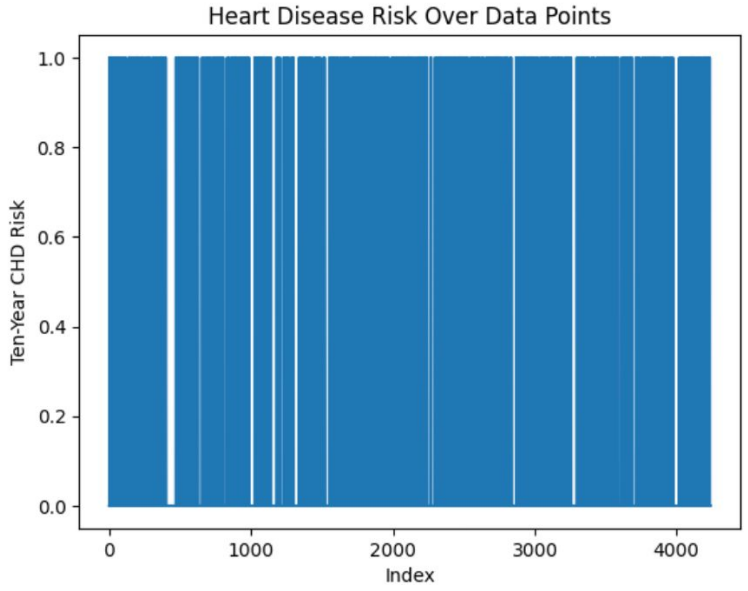
Deployment (Future Scope)

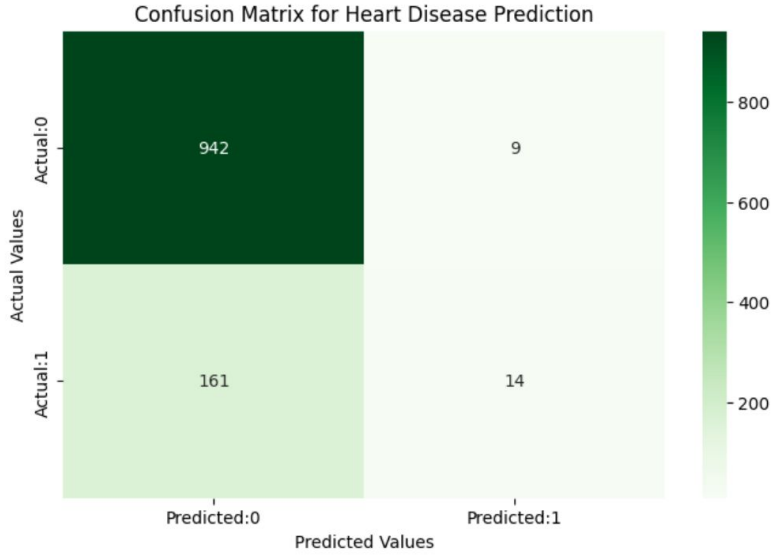
- While deployment was not conducted, it can be achieved using **Flask or Streamlit**, integrating it into a web-based platform.
- Deployment steps may include **model serialization (Pickle or Joblib), API creation, and hosting on cloud services** like Microsoft Azure.

RESULT

- **Accuracy:** The Logistic Regression model achieved **85% accuracy**.
- **Performance Metrics:** Precision, Recall, and F1-score indicate model effectiveness.
- **Visualizations:** Confusion Matrix and comparison plots highlight prediction quality.







CONCLUSION

The **Logistic Regression model** successfully predicts heart disease risk with **85% accuracy**, offering a reliable tool for early diagnosis and medical assessment. While effective, challenges such as handling imbalanced data and optimizing feature selection were encountered during implementation. Future improvements could include refining the dataset, exploring advanced models, and deploying the solution via **Azure or Flask** for broader accessibility. Ultimately, accurate predictions contribute to timely interventions, potentially reducing cardiac risks and improving patient outcomes.

FUTURE SCOPE

- **Improving Accuracy**: Incorporate additional medical datasets for better prediction reliability.
- **Advanced Models**: Explore deep learning techniques for enhanced performance.
- **Expanding Coverage**: Adapt the system to broader demographic groups and populations.
- **Deployment & Accessibility**: Utilize **Azure** for cloud-based access and API integration.
- **Edge Computing**: Enable localized, faster processing for real-time medical diagnostics.

REFERENCES

- https://academic.oup.com/eurheartjsupp/article/27/Supplement_1/i16/8023303
- https://lakefs.io/blog/data-preprocessing-in-machine-learning/
- https://www.geeksforgeeks.org/understanding-logistic-regression/
- https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset

GitHub Link: https://github.com/sarayu-04/Heart-Disease-Prediction-Using-Logistic-Regression.git

Thank you