HEART DISEASE PREDICTION USING MACHINE LANGUAGE

Published by Authors: Shinde, R., Arjun, S., Patil, P., & Waghmare, J. (2015).

An Intelligent Heart Disease Prediction System Using K-Means Clustering and Naïve Bayes Algorithm. IJCSIT) International Journal of Computer Science and Information Technologies, 6(1), 637-639.

Abstract

Heart disease is a major global health concern, claiming millions of lives annually. Timely diagnosis is critical for effective treatment and prevention. This project leverages machine learning techniques to predict the risk of heart disease based on clinical and lifestyle data. By analyzing features such as age, cholesterol levels, blood pressure, and exercise habits, the model identifies patterns indicative of heart disease. Algorithms like logistic regression, decision trees, and random forests were evaluated for accuracy and reliability. The system achieved high precision, showcasing its potential as a decision-support tool for healthcare professionals in improving patient outcomes.

Introduction

Heart disease encompasses a range of cardiovascular conditions, including coronary artery disease, heart attacks, and arrhythmias. Despite advancements in medical technology, late diagnosis remains a challenge, leading to preventable mortality and morbidity. Machine learning, a subset of artificial intelligence, has emerged as a transformative approach in medical diagnostics, offering the ability to analyze complex datasets and uncover hidden patterns. This project aims to develop a predictive model for heart disease using machine learning algorithms, providing a reliable and efficient tool for early detection. The study also emphasizes the importance of integrating data-driven solutions into clinical workflows for better healthcare delivery.

Methodology

1. **Data Collection**:
   The dataset was sourced from reliable repositories, such as the UCI Heart Disease Dataset, comprising attributes like age, gender, resting blood pressure, serum cholesterol levels, fasting blood sugar, and exercise-induced angina.

2. **Data Preprocessing**:
    - **Data Cleaning**: Missing and inconsistent data were addressed through imputation and removal of anomalies.
    - **Feature Scaling**: Standardization was applied to ensure uniformity across feature values.
    - **Feature Selection**: Techniques such as Recursive Feature Elimination (RFE) and correlation analysis were employed to identify the most impactful variables.
3. **Model Development**:
    - **Algorithms Used**: Logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks.
    - **Training and Testing Split**: The dataset was divided into 70% for training and 30% for testing.
    - **Hyperparameter Tuning**: Grid search and cross-validation methods were used to optimize model performance.
4. **Evaluation Metrics**:
    - Accuracy
    - Precision
    - Recall
    - F1 Score
    - ROC-AUC Curve
5. **Implementation Tools**:
   Python programming was used, leveraging libraries such as Scikit-learn, TensorFlow, and Pandas for data analysis, model building, and evaluation.

Conclusion

The machine learning-based heart disease prediction model demonstrated strong performance across various evaluation metrics, proving its capability as a diagnostic aid. By leveraging predictive analytics, the model facilitates early intervention, reducing the risk of severe cardiac events. Future work will focus on incorporating real-time patient monitoring data, addressing limitations such as dataset size, and ensuring the system's adaptability across diverse demographics. The project underscores the transformative potential of machine learning in reshaping healthcare delivery and improving patient care.