Project Report
Health Prognosis

Project Title: Health Prognosis
Your Name: Mihir Panchal
Date: February 2024

Abstract

This project addresses the critical issue of heart disease prediction, which plays a significant role in healthcare. Using a logistic regression model, the project developed a predictive system that assesses the likelihood of heart disease based on various medical attributes. Key findings indicate the model's accuracy and effectiveness in distinguishing between patients with and without heart disease, demonstrating the potential for early intervention and improved patient outcomes.

Introduction

Heart disease remains one of the leading causes of morbidity and mortality worldwide. Early detection and prevention are crucial for improving patient prognosis. This project aims to develop a predictive model that leverages historical health data to identify individuals at risk of heart disease.

Objectives

- To create a logistic regression model for predicting heart disease.
- To evaluate the model's performance using relevant metrics.
- To demonstrate the model's practical application through a sample input.

Importance and Relevance

The ability to predict heart disease accurately can lead to timely medical interventions, potentially saving lives and reducing healthcare costs. This project emphasizes the relevance of data-driven approaches in healthcare, paving the way for more advanced predictive analytics.

Methodology

Description of Methods and Technologies Used

The project utilized Python and libraries such as Pandas, NumPy, and scikit-learn for data analysis and model development. The methodology encompassed data loading, exploration, preprocessing, model training, evaluation, and demonstration of the predictive system.

Steps Taken

1. Data Loading and Exploration:

- The heart disease dataset was loaded from an Excel file *('Heart_Attack.xlsx')* using `*pd.read_excel()*`.
- Initial data exploration included displaying the first and last few rows, checking the shape and information of the dataset, and identifying any missing values and statistical measures.

2. Data Preprocessing:

- The target variable *('target')* was separated from the features.
- The dataset was split into training *(80%)* and testing *(20%)* sets using `*train_test_split()*`.

3. Model Training:

- A logistic regression model was created using `*LogisticRegression()*` and trained on the training data using `*model.fit()*`.

4. Model Evaluation:

- The accuracy of the trained model was evaluated on both the training and testing datasets using `*accuracy_score()*`.

5. Predictive System:

- A sample input was generated to demonstrate the model's predictive capabilities, converting the input into a NumPy array with *`np.asarray()`* and reshaping it for prediction.

Implementation

The implementation began with the careful loading and exploration of the dataset. After preprocessing, including the separation of features and the target variable, the logistic regression model was developed and trained. During implementation, challenges such as missing values and model tuning were addressed by conducting thorough exploratory data analysis and parameter adjustments.

Challenges Faced

- Handling Missing Values: Missing values were identified and addressed by analyzing their distribution, ensuring the model's integrity.

- Model Overfitting: Careful evaluation of the model's performance metrics ensures that the model generalizes well to unseen data, minimizing overfitting.

Results

The logistic regression model demonstrated promising results, achieving an accuracy of *87%* on the testing dataset.

Data Visualization

- Confusion Matrix: A confusion matrix was generated to visualize the model's predictions against actual outcomes.

- Accuracy Scores: Bar graphs representing the accuracy scores for both training and testing datasets highlighted the model's effectiveness.

Analysis of Results

The results indicate that the model effectively predicts heart disease with high accuracy. This underscores the potential for using such predictive systems in real-world healthcare settings to assist clinicians in decision-making.

Conclusion

The heart disease prediction model successfully identifies individuals at risk based on historical health data. This project illustrates the power of machine learning in healthcare, offering a scalable solution for early detection of heart disease.

Recommendations for Future Work

Future efforts could explore more complex models, such as ensemble methods or neural networks, to enhance prediction accuracy. Additionally, integrating real-time data could further improve the model's applicability in clinical settings.