Generative AI Project

HEART DISEASE DETECTION

by,
Swathikasri S
NM id : au211521244052

INTRODUCTION

The “Heart Disease Detection” system uses machine learning to predict and identify the likelihood of heart diseases in individuals. Detecting cardiovascular diseases early is crucial for preventing deaths caused by them since they are a significant cause of death globally. The system assesses the risk of heart disease based on various health parameters. The algorithm used is a Support Vector Machine (SVM). SVM is good at classification tasks, like predicting if someone is at risk of heart disease.

Problem Statement :

To reduce heart disease, we must identify at-risk people. Traditional methods that rely on manual analysis may need to be more effective in efficiently managing the complexity of diverse health data.

ALGORITHMS FOR HEART DISEASE PREDICTION

Several machine learning algorithms have been successfully applied to predict heart disease. The choice of algorithm depends on the dataset characteristics and the specific goals of the prediction model. Some widely used algorithms include

Logistic Regression: Logistic Regression is a commonly used algorithm for binary classification tasks, making it suitable for predicting whether an individual is at risk of heart disease.

Decision Trees: Decision Trees are versatile and understandable, making them helpful in identifying patterns in heart disease risk factors. They can handle both numerical and categorical data.

Random Forest: Random Forest is an ensemble learning technique that combines multiple decision trees to improve predictive accuracy and reduce overfitting.

Support Vector Machines (SVM): SVM effectively separates data into classes and is particularly useful when dealing with complex datasets with non-linear relationships.

Neural Networks: Deep learning models like Neural Networks can capture intricate patterns in large datasets, making them suitable for complex heart disease prediction tasks.

code implementation

Step 1: Importing Libraries

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, roc_curve, auc
import matplotlib.pyplot as plt
import seaborn as sns
```

Step 2: Loading Data

```
heartdata = pd.read_csv("heart.csv")
```

Step 3: Exploring Data

```
heartdata.head()
heartdata.tail()
heartdata.info()
heartdata.describe()
```

Step 4: Data Preprocessing

```
X = heartdata.drop(columns='target', axis=1)
Y = heartdata['target']
```

Step 5: Train-Test Split

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)
```

Step 6: Model Training

```
model = LogisticRegression()
model.fit(X_train, Y_train)
```

Step 7: Model Evaluation on Test Data

```
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)print(test_data_accuracy)
```

Step 8: Making Predictions

```
input_from_user = (71, 0, 0, 112, 149, 0, 1, 125, 0, 1.6, 1, 0, 2)
input_from_user_array = np.asarray(input_from_user)
input_from_user_reshaped = input_from_user_array.reshape(1, -1)
prediction = model.predict(input_from_user_reshaped)
```

Step 9: ROC Curve and AUC

```
Y_test_probabilities = model.predict_proba(X_test)[:, 1]
fpr, tpr, _ = roc_curve(Y_test, Y_test_probabilities)
roc_auc = auc(fpr, tpr)

plt.figure()
plt.plot(fpr, tpr, color='darkorange', lw=2, label='ROC curve (area = %0.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic (ROC) Curve')
plt.legend(loc="lower right")
plt.show()
```

Challenges in Implementing Machine Learning for Heart Disease Prediction:

Data Quality: In healthcare, ensuring that the data used for training machine learning models is reliable and accurate is difficult. There are often issues with the quality and consistency of data sources. When health records are flawed or incomplete, it can introduce biases that make predictive models less effective. It is crucial to address these concerns to create dependable and trustworthy systems for predicting heart disease.

Interpretability: Some machine learning models' "black box" nature can make it challenging for healthcare professionals to understand and trust the predictions, hindering widespread adoption.

Ethical Concerns: Ensuring patient privacy, data security, and ethical use of healthcare data are critical challenges in developing and deploying machine learning systems in healthcare.

Clinical Adoption: To use machine learning predictions in healthcare, we must address challenges like resistance to change and lack of awareness or training among healthcare professionals. We also need to ensure smooth integration with existing workflows.

Ethical Concerns: Ensuring patient privacy, data security, and ethical use of healthcare data are critical challenges in developing and deploying machine learning systems in healthcare.