Project Title: Heart Disease Prediction Using Machine Learning

Overview:

Cardiovascular diseases are a leading cause of mortality worldwide, emphasizing the need for effective and early detection. This project aims to develop a predictive model using machine learning techniques to assess the risk of heart disease in individuals based on various health indicators. By analyzing a set of features such as age, sex, blood pressure, cholesterol levels, and other relevant medical data, the model will provide a valuable tool for healthcare professionals to identify individuals at a higher risk of developing heart diseases.

Objectives:

1. **Data Collection:**
    - First I gather a comprehensive dataset containing health-related features and corresponding labels indicating the presence or absence of heart disease.
2. **Data Preprocessing:**

    After that I handled missing data, encode categorical variables, and normalize numerical features to prepare the dataset for analysis.
3. **Exploratory Data Analysis (EDA):**

    After that I Explore the dataset to identify patterns, correlations, and gain insights into the relationships between different health indicators.
4. **Feature Selection:**

After that I use feature Selection to get the most relevant features that contribute significantly to predicting heart disease.

5. **Model Selection:**

   Next,i Implement and compare the performance of various machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Neural Networks.

6. **Model Training:**

   Then split the dataset into training and testing sets and train the selected model using the training data.

7. **Model Evaluation:**

   After I explore the model's performance using metrics like accuracy, precision, recall, F1 score, and AUC-ROC on the testing set.

8. **Hyperparameter Tuning:**

   Then I choose best model and optimize the model's hyperparameters to enhance its predictive capabilities.

9. **Validation:**

   Finally Validate the model on an independent dataset to ensure its generalization to new and unseen data.

My Expected Outcomes:

- A machine learning model capable of accurately predicting the likelihood of heart disease based on input health data.
- Insights into the key factors influencing heart disease risk.

Important notes

- Early detection of heart disease risk can lead to timely interventions and lifestyle modifications, potentially reducing the incidence of cardiovascular diseases and improving overall public health.

Key Points : I got Random Forest classifier giving prediction result with high accuracy, so I use that model for hyperparameter tuning.

Training Accuracy : 98 %

Validation Accuracy : 90 %

----------Heart Disease Prediction----------

Business Problem

Heart disease is common both in the population at large but also in the population of working age. It is estimated that heart disease, including stroke and high blood pressure, is responsible for more costs than any other disease or injury. The cost in occupational terms of cardiovascular disease (CVD) is, however, harder to quantify but is likely to be similarly high. Heart disease can claim the ultimate cost as the most common cause of death.

We are supposed to create a machine learning Model. When we Give the required details The model will tell us Wheather that persion have chance of getting Heart Disease or Not . This model can help medical persion to cheak a pation condiotn and can predict that a persion will have heart disease in future.

Data Description

- Dependent Feature : HeartDisease --> This column will tell us wheather a persion has heart disease or not.
- BMI : Body Mass Index (BMI) is a person's weight in kilograms (or pounds) divided by the square of height in meters (or feet).
- Smoking : Wheather that persion smoke or not.
- AlcoholDrinking : wheather that persion Drink Acohol or not.
- Stroke : is that persion have storke earlier or not.
- PhysicalHealth: Physical Health Level of the persion.
- MentalHealth: Mental health of the persion.
- DiffWalking : Difficulty in walking.
- Sex : male or female
- AgeCategory:we have multiple age group here.
- Race:The term "race," in its traditional genetic conceptualization, is often determined through patterns of human variation reflected from our evolutionary history to serve as a definitive measure
- Diabetic : Wheather the persion has diabetic or not
- PhysicalActivity : Wheather the persion do physical Activity or not.
- GenHealth-genetic health of the persion
- SleepTime: total sleeping Hours
- Asthma : wheather that persion has asthma or not
- KidneyDisease: wheather that persion has kidneyDisease or not
- SkinCancer: Wheather that persion has SkinCancer or not.

jupyter heart project Last Checkpoint: a few seconds ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

Run Code

Out[3]:

	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWalking	Sex	AgeCategory	Race	Diabetic	PhysicalActivity
0	No	16.60	Yes	No	No	3.0	30.0	No	Female	55-59	White	Yes	Yes
1	No	20.34	No	No	Yes	0.0	0.0	No	Female	80 or older	White	No	Yes
2	No	26.58	Yes	No	No	20.0	30.0	No	Male	65-69	White	Yes	Yes
3	No	24.21	No	No	No	0.0	0.0	No	Female	75-79	White	No	No
4	No	23.71	No	No	No	28.0	0.0	Yes	Female	40-44	White	No	Yes
5	Yes	28.87	Yes	No	No	6.0	0.0	Yes	Female	75-79	Black	No	No
6	No	21.63	No	No	No	15.0	0.0	No	Female	70-74	White	No	Yes
7	No	31.64	Yes	No	No	5.0	0.0	Yes	Female	80 or older	White	Yes	No
8	No	26.45	No	No	No	0.0	0.0	No	Female	80 or older	White	No, borderline diabetes	No
9	No	40.69	No	No	No	0.0	0.0	Yes	Male	65-69	White	No	Yes
10	Yes	34.30	Yes	No	No	30.0	0.0	Yes	Male	60-64	White	Yes	No
11	No	28.71	Yes	No	No	0.0	0.0	No	Female	55-59	White	No	Yes
12	No	28.37	Yes	No	No	0.0	0.0	Yes	Male	75-79	White	Yes	Yes
13	No	28.15	No	No	No	7.0	0.0	Yes	Female	80 or older	White	No	No
14	No	29.29	Yes	No	No	0.0	30.0	Yes	Female	60-64	White	No	No
15	No	29.18	No	No	No	1.0	0.0	No	Female	50-54	White	No	Yes
16	No	26.26	No	No	No	5.0	2.0	No	Female	70-74	White	No	No
17	No	22.59	Yes	No	No	0.0	30.0	Yes	Male	70-74	White	No, borderline	Yes

http://localhost:8888/nbconvert/html/data%20science/100%20projects/besant%20technology/heart%20project.ipynb?download=false

YouTube Google Gmail CEH v10 Certified E... 02-AUG-21-CDS-W... Google Colab 02-08-2021-CDS-O... SK Sir Ki Class | EP1... Top Data Science In... PTID-CDS-NOV21-... MAKAUT, West Ben...

```
Best Parameters:  {'n_estimators': 40, 'min_samples_split': 2, 'min_samples_leaf': 3, 'max_features': 'sqrt', 'max_depth': 34, 'bootstrap': True}
```

In [74]:
```
rndfr=RandomForestClassifier(n_estimators=40, min_samples_split=6, min_samples_leaf=1, max_features='sqrt', max_depth=59,bootstrap=False)
```

In [75]:
```
rndfr.fit( X_train, y_train)
```

Out[75]:
```
RandomForestClassifier(bootstrap=False, max_depth=59, max_features='sqrt',
                       min_samples_split=6, n_estimators=40)
```

In [76]:
```
print('Training Accuracy : ',metrics.accuracy_score(y_train,rndfr.predict(X_train))*100)
print('Validation Accuracy : ',metrics.accuracy_score(y_test,rndfr.predict(X_test))*100)
```

```
Training Accuracy :  98.96652543035383
Validation Accuracy :  90.71749089260307
```

In [77]:
```
confusion_matrix = metrics.confusion_matrix(y_test,rndfr.predict(X_test))
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = confusion_matrix, display_labels = [False, True])
cm_display.plot()
plt.show()
```