MAJOR PROJECT
ARTIFICIAL INTELLIGENCE

TASK:

Take any Dataset of your choice ,perform EDA(Exploratory Data Analysis) and apply a suitable Classifier,Regressor or Clusterer and calculate the accuracy of the model.

INTRODUCTION

Heart failure is the result of cardiac lesions among which we distinguish cardiomyopathies, lung diseases, which can hinder the circulation of blood which affects the efficiency of the right heart and finally the diseases of the valves.

The consultation with the doctor is a valuable and indispensable source for the diagnosis, from the description of the signs felt by the patient.All these signs will allow the doctor to evaluate the cardiac function and thus to better appreciate the filling, the ejection, the work of the valves, the state of the cavities.

This Artificial Intelligence model uses a 'Decision Tree classifier' that allows automatic prediction of the chance of heart failure based on symptoms of the patient.

EXPLORATORY DATA ANALYSIS

```python
import pandas as pd
from sklearn.metrics import accuracy_score
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
hearts_df = pd.read_csv('/content/heart.csv')
hearts_df
```

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
...	...	...	...	...	...	...	...	...	...	...	...	...
913	45	M	TA	110	264	0	Normal	132	N	1.2	Flat	1
914	68	M	ASY	144	193	1	Normal	141	N	3.4	Flat	1
915	57	M	ASY	130	131	0	Normal	115	Y	1.2	Flat	1
916	57	F	ATA	130	236	0	LVH	174	N	0.0	Flat	1
917	38	M	NAP	138	175	0	Normal	173	N	0.0	Up	0

918 rows × 12 columns

```
hearts_df.head()
```

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0

```
hearts_df.tail()
```

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
913	45	M	TA	110	264	0	Normal	132	N	1.2	Flat	1
914	68	M	ASY	144	193	1	Normal	141	N	3.4	Flat	1
915	57	M	ASY	130	131	0	Normal	115	Y	1.2	Flat	1
916	57	F	ATA	130	236	0	LVH	174	N	0.0	Flat	1
917	38	M	NAP	138	175	0	Normal	173	N	0.0	Up	0

```
hearts_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 918 entries, 0 to 917
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Age             918 non-null    int64
 1   Sex             918 non-null    object
 2   ChestPainType   918 non-null    object
 3   RestingBP       918 non-null    int64
 4   Cholesterol     918 non-null    int64
 5   FastingBS       918 non-null    int64
 6   RestingECG      918 non-null    object
 7   MaxHR           918 non-null    int64
 8   ExerciseAngina  918 non-null    object
 9   Oldpeak         918 non-null    float64
 10  ST_Slope        918 non-null    object
 11  HeartDisease    918 non-null    int64
dtypes: float64(1), int64(6), object(5)
memory usage: 86.2+ KB
```

```python
from sklearn import preprocessing
def convert(data):
    number = preprocessing.LabelEncoder()
    data['Sex'] = number.fit_transform(data.Sex)
    data['ChestPainType'] =
number.fit_transform(data.ChestPainType)
    data['RestingECG'] = number.fit_transform(data.RestingECG)
    data['ExerciseAngina'] =
number.fit_transform(data.ExerciseAngina)
    data['ST_Slope'] = number.fit_transform(data.ST_Slope)
    data=data.fillna(-999)
    return data


hearts_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 918 entries, 0 to 917
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Age             918 non-null    int64
 1   Sex             918 non-null    int64
 2   ChestPainType   918 non-null    int64
 3   RestingBP       918 non-null    int64
 4   Cholesterol     918 non-null    int64
 5   FastingBS       918 non-null    int64
 6   RestingECG      918 non-null    int64
 7   MaxHR           918 non-null    int64
 8   ExerciseAngina  918 non-null    int64
 9   Oldpeak         918 non-null    float64
 10  ST_Slope        918 non-null    int64
 11  HeartDisease    918 non-null    int64
dtypes: float64(1), int64(11)
memory usage: 86.2 KB
```

CLASSIFICATION MODEL

```python
x = hearts_df.iloc[:,0:11].values
y = hearts_df.iloc[:,11].values

from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest =
train_test_split(x,y,random_state=0,stratify=y)
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier(criterion="gini",random_state =
0,max_depth = 4, min_samples_leaf=5)
model.fit(xtrain,ytrain)
```

OUTPUT ANALYSIS

```
ypred = model.predict(xtest)
ypred

array([1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1,
       0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1,
       0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0,
       1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0,
       1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1,
       0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1,
       0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1,
       0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1,
       1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0,
       1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 1, 1, 0, 0, 0, 0])
```

```
model.predict([[ 61. ,    1. ,    0. , 141. , 292. ,    0. ,    2. ,
115. ,    1. ,1.7,    1.]])
```

```
array([1])
```

```
acc = accuracy_score(ytest,ypred)*100
acc
```

```
84.78260869565217
```

REFERENCES

- **Dataset Link**
https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction
- **Code Link**
https://colab.research.google.com/drive/18Vt1YVG4y7pVKMiQt8WpKWh8r_4Dv44W?usp=sharing
- **Other references**
http://scikit-learn.org/stable/modules/tree.html#:~:text=Decision%20Trees%20(DTs)%20are%20a,as%20a%20piecewise%20constant%20approximation.

TEAM MEMBERS:
- **Aaditya Bagaddeo**
- **Siliveru Akash Durga**
- **Akkala Sreeksha**