Heart attack analysis using Decision Tree

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
import six
import sys
sys.modules['sklearn.externals.six'] = six
from sklearn import tree
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
import graphviz
import matplotlib.pyplot as plt
%matplotlib inline
```

```
# Load the data
data = pd.read_csv('heart.csv')
data
```

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	t
0	63	1	3	145	233	1	0	150	0	2.3	0	0	
1	37	1	2	130	250	0	1	187	0	3.5	0	0	
2	41	0	1	130	204	0	0	172	0	1.4	2	0	
3	56	1	1	120	236	0	1	178	0	0.8	2	0	
4	57	0	0	120	354	0	1	163	1	0.6	2	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
298	57	0	0	140	241	0	1	123	1	0.2	1	0	
299	45	1	3	110	264	0	1	132	0	1.2	1	0	
300	68	1	0	144	193	1	1	141	0	3.4	1	2	
301	57	1	0	130	131	0	1	115	1	1.2	1	1	
302	57	0	1	130	236	0	0	174	0	0.0	1	1	

303 rows × 14 columns

```
# Split the data into features and labels
X = data.drop(['age'], axis=1)
y = data['age']
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       303 non-null    int64
 1   sex       303 non-null    int64
 2   cp        303 non-null    int64
 3   trtbps    303 non-null    int64
 4   chol      303 non-null    int64
 5   fbs       303 non-null    int64
 6   restecg   303 non-null    int64
 7   thalachh  303 non-null    int64
 8   exng      303 non-null    int64
 9   oldpeak   303 non-null    float64
 10  slp       303 non-null    int64
 11  caa       303 non-null    int64
 12  thall     303 non-null    int64
 13  output    303 non-null    int64
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

```
data.describe()
```

	age	sex	cp	trtbps	chol	fbs	res
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.00
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.52
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.52
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.00
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.00
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.00
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.00
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.00

```
data.isnull()
```

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	sl
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	..
298	False	False	False	False	False	False	False	False	False	False	False
299	False	False	False	False	False	False	False	False	False	False	False
300	False	False	False	False	False	False	False	False	False	False	False
301	False	False	False	False	False	False	False	False	False	False	False
302	False	False	False	False	False	False	False	False	False	False	False

303 rows × 14 columns

```
data.isnull().sum()
```

```
age         0
sex         0
cp          0
trtbps      0
chol        0
fbs         0
restecg     0
thalachh    0
exng        0
oldpeak     0
slp         0
caa         0
thall       0
output      0
dtype: int64
```

```
X = data.iloc[:,0:13] # Features
y = data.iloc[:,13] # Target variable
x
```

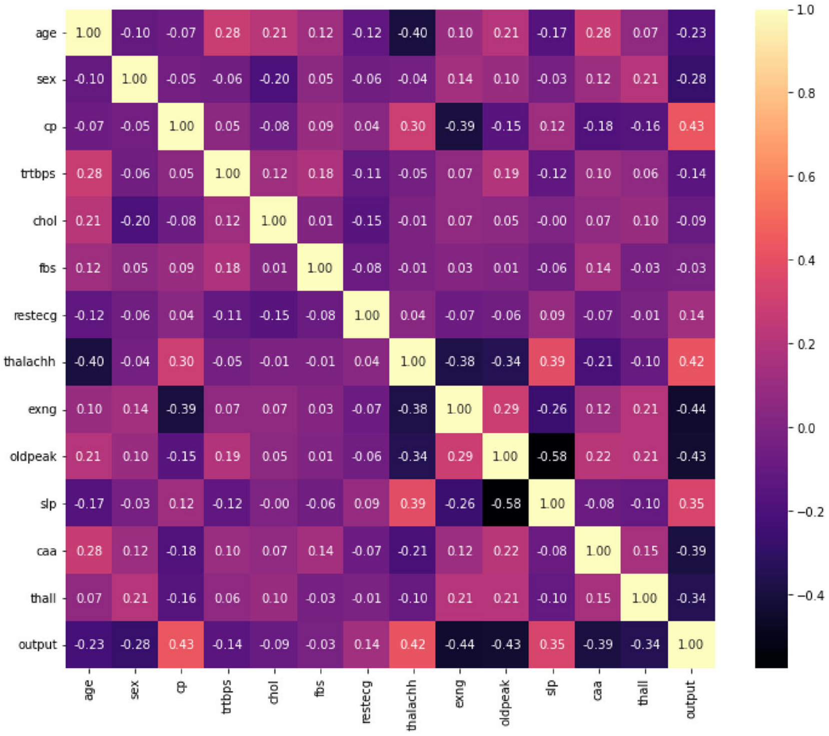
	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	t
0	63	1	3	145	233	1	0	150	0	2.3	0	0	

```
y=data['output']
y
```

```
0      1
1      1
2      1
3      1
4      1
      ..
298    0
299    0
300    0
301    0
302    0
Name: output, Length: 303, dtype: int64
```

```
plt.figure(figsize=(12,10))
sns.heatmap(data.corr(),annot=True,cmap="magma",fmt='.2f')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f4187ecd7c0>
```



	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
age	1.00	-0.10	-0.07	0.28	0.21	0.12	-0.12	-0.40	0.10	0.21	-0.17	0.28	0.07	-0.23
sex	-0.10	1.00	-0.05	-0.06	-0.20	0.05	-0.06	-0.04	0.14	0.10	-0.03	0.12	0.21	-0.28
cp	-0.07	-0.05	1.00	0.05	-0.08	0.09	0.04	0.30	-0.39	-0.15	0.12	-0.18	-0.16	0.43
trtbps	0.28	-0.06	0.05	1.00	0.12	0.18	-0.11	-0.05	0.07	0.19	-0.12	0.10	0.06	-0.14
chol	0.21	-0.20	-0.08	0.12	1.00	0.01	-0.15	-0.01	0.07	0.05	-0.00	0.07	0.10	-0.09
fbs	0.12	0.05	0.09	0.18	0.01	1.00	-0.08	-0.01	0.03	0.01	-0.06	0.14	-0.03	-0.03
restecg	-0.12	-0.06	0.04	-0.11	-0.15	-0.08	1.00	0.04	-0.07	-0.06	0.09	-0.07	-0.01	0.14
thalachh	-0.40	-0.04	0.30	-0.05	-0.01	-0.01	0.04	1.00	-0.38	-0.34	0.39	-0.21	-0.10	0.42
exng	0.10	0.14	-0.39	0.07	0.07	0.03	-0.07	-0.38	1.00	0.29	-0.26	0.12	0.21	-0.44
oldpeak	0.21	0.10	-0.15	0.19	0.05	0.01	-0.06	-0.34	0.29	1.00	-0.58	0.22	0.21	-0.43
slp	-0.17	-0.03	0.12	-0.12	-0.00	-0.06	0.09	0.39	-0.26	-0.58	1.00	-0.08	-0.10	0.35
caa	0.28	0.12	-0.18	0.10	0.07	0.14	-0.07	-0.21	0.12	0.22	-0.08	1.00	0.15	-0.39
thall	0.07	0.21	-0.16	0.06	0.10	-0.03	-0.01	-0.10	0.21	0.21	-0.10	0.15	1.00	-0.34
output	-0.23	-0.28	0.43	-0.14	-0.09	-0.03	0.14	0.42	-0.44	-0.43	0.35	-0.39	-0.34	1.00

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)
```

```
# Train the Decision Tree classifier
clf = DecisionTreeClassifier()
clf.fit(X_train, y_train)
```

```
DecisionTreeClassifier()
```

```
# Predict using the trained model
y_pred = clf.predict(X_test)
```

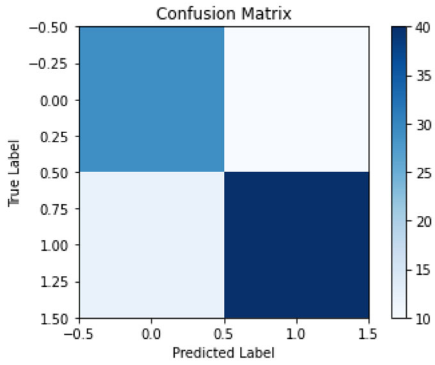
```
# Evaluate the accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

```
Accuracy: 0.7582417582417582
```

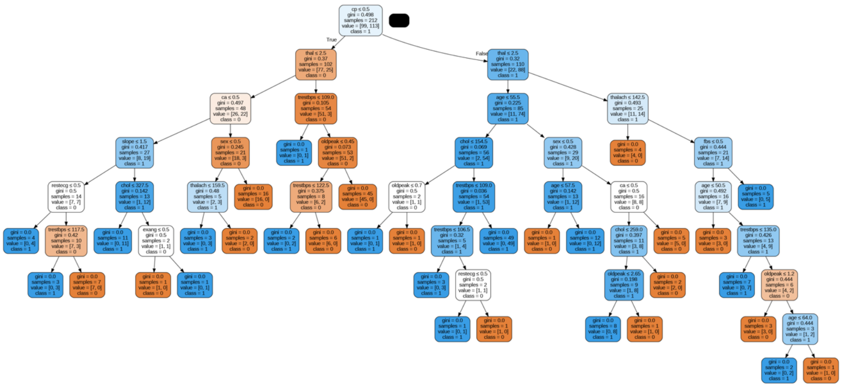
```
# Evaluate the model performance using confusion matrix
cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", cm)
```

```
Confusion Matrix:
 [[29 10]
 [12 40]]
```

```
# Plot the confusion matrix
plt.imshow(cm, cmap='Blues')
plt.colorbar()
plt.title("Confusion Matrix")
plt.xlabel("Predicted Label")
plt.ylabel("True Label")
plt.show()
```



```
# Plot the decision tree
feature_cols = ['age', 'sex', 'cp', 'trestbps','chol', 'fbs', 'restecg', 'thalach','exang', 'oldpeak', 'slope', 'ca',
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,
                filled=True, rounded=True,
                special_characters=True,feature_names = feature_cols  ,class_names=['0','1'])
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('img.png')
Image(graph.create_png())
```


cp ≤ 0.5 gini = 0.498 samples = 212 value = [99, 113] class = 1
True
False
thal ≤ 2.5 gini = 0.37 samples = 102 value = [77, 25] class = 0
thal ≤ 2.5 gini = 0.32 samples = 110 value = [22, 88] class = 1
ca ≤ 0.5 gini = 0.497 samples = 48 value = [26, 22] class = 0
trestbps ≤ 109.0 gini = 0.105 samples = 54 value = [51, 3] class = 0
age ≤ 55.5 gini = 0.225 samples = 85 value = [11, 74] class = 1
thalach ≤ 142.5 gini = 0.493 samples = 25 value = [11, 14] class = 1
slope ≤ 1.5 gini = 0.417 samples = 27 value = [8, 19] class = 1
sex ≤ 0.5 gini = 0.245 samples = 21 value = [18, 3] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
oldpeak ≤ 0.45 gini = 0.073 samples = 53 value = [51, 2] class = 0
chol ≤ 154.5 gini = 0.069 samples = 56 value = [2, 54] class = 1
sex ≤ 0.5 gini = 0.428 samples = 29 value = [9, 20] class = 1
gini = 0.0 samples = 4 value = [4, 0] class = 0
fbs ≤ 0.5 gini = 0.444 samples = 21 value = [7, 14] class = 1
restecg ≤ 0.5 gini = 0.5 samples = 14 value = [7, 7] class = 0
chol ≤ 327.5 gini = 0.142 samples = 13 value = [1, 12] class = 1
thalach ≤ 159.5 gini = 0.48 samples = 5 value = [2, 3] class = 1
gini = 0.0 samples = 16 value = [16, 0] class = 0
trestbps ≤ 122.5 gini = 0.375 samples = 8 value = [6, 2] class = 0
gini = 0.0 samples = 45 value = [45, 0] class = 0
oldpeak ≤ 0.7 gini = 0.5 samples = 2 value = [1, 1] class = 0
trestbps ≤ 109.0 gini = 0.036 samples = 54 value = [1, 53] class = 1
age ≤ 57.5 gini = 0.142 samples = 13 value = [1, 12] class = 1
ca ≤ 0.5 gini = 0.5 samples = 16 value = [8, 8] class = 0
age ≤ 50.5 gini = 0.492 samples = 16 value = [7, 9] class = 1
gini = 0.0 samples = 5 value = [0, 5] class = 1
gini = 0.0 samples = 4 value = [0, 4] class = 1
trestbps ≤ 117.5 gini = 0.42 samples = 10 value = [7, 3] class = 0
gini = 0.0 samples = 11 value = [0, 11] class = 1
exang ≤ 0.5 gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.0 samples = 3 value = [0, 3] class = 1
gini = 0.0 samples = 2 value = [2, 0] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.0 samples = 6 value = [6, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
trestbps ≤ 106.5 gini = 0.32 samples = 5 value = [1, 4] class = 1
gini = 0.0 samples = 49 value = [0, 49] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 12 value = [0, 12] class = 1
chol ≤ 259.0 gini = 0.397 samples = 11 value = [3, 8] class = 1
gini = 0.0 samples = 5 value = [5, 0] class = 0
gini = 0.0 samples = 3 value = [3, 0] class = 0
trestbps ≤ 135.0 gini = 0.426 samples = 13 value = [4, 9] class = 1
gini = 0.0 samples = 3 value = [0, 3] class = 1
gini = 0.0 samples = 7 value = [7, 0] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 3 value = [0, 3] class = 1
restecg ≤ 0.5 gini = 0.5 samples = 2 value = [1, 1] class = 0
oldpeak ≤ 2.65 gini = 0.198 samples = 9 value = [1, 8] class = 1
gini = 0.0 samples = 2 value = [2, 0] class = 0
gini = 0.0 samples = 7 value = [0, 7] class = 1
oldpeak ≤ 1.2 gini = 0.444 samples = 6 value = [4, 2] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 8 value = [0, 8] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 3 value = [3, 0] class = 0
age ≤ 64.0 gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0

✓ 3s completed at 11:15 AM ● ×