```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split

df=pd.read_csv('/content/drive/MyDrive/diabetcsv.csv')

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   preg    768 non-null    int64  
 1   plas    768 non-null    int64  
 2   pres    768 non-null    int64  
 3   skin    768 non-null    int64  
 4   insu    768 non-null    int64  
 5   mass    768 non-null    float64
 6   pedi    768 non-null    float64
 7   age     768 non-null    int64  
 8   class   768 non-null    object 
dtypes: float64(2), int64(6), object(1)
memory usage: 54.1+ KB
```

```
df.head()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 768,\n  \"fields\": [\n    {\n      \"column\": \"preg\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 3,\n        \"min\": 0,\n        \"max\": 17,\n        \"num_unique_values\": 17,\n        \"samples\": [\n          6,\n          1,\n          3\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"plas\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 31,\n        \"min\": 0,\n        \"max\": 199,\n        \"num_unique_values\": 136,\n        \"samples\": [\n          151,\n          101,\n          112\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"pres\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 19,\n        \"min\": 0,\n        \"max\": 122,\n        \"num_unique_values\": 47,\n        \"samples\": [\n          86,\n          46,\n          85\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"skin\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 15,\n        \"min\": 0,\n        \"max\": 99,\n        \"num_unique_values\": 51,\n        \"samples\": [\n          7,\n          12,\n          48\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"insu\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
```

\"std\": 115,\n   \"min\": 0,\n   \"max\": 846,\n   \"num_unique_values\": 186,\n   \"samples\": [\n   52,\n   41,\n   183\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"mass\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 7.8841603203754405,\n   \"min\": 0.0,\n   \"max\": 67.1,\n   \"num_unique_values\": 248,\n   \"samples\": [\n   19.9,\n   31.0,\n   38.1\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"pedi\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 0.33132859501277484,\n   \"min\": 0.078,\n   \"max\": 2.42,\n   \"num_unique_values\": 517,\n   \"samples\": [\n   1.731,\n   0.426,\n   0.138\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 11,\n   \"min\": 21,\n   \"max\": 81,\n   \"num_unique_values\": 52,\n   \"samples\": [\n   60,\n   47,\n   72\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"class\",\n   \"properties\": {\n   \"dtype\": \"category\",\n   \"num_unique_values\": 2,\n   \"samples\": [\n   \"tested_negative\",\n   \"tested_positive\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
from sklearn.preprocessing import LabelEncoder
LE=LabelEncoder()
df["class"]=LE.fit_transform(df["class"])

x_train,x_test,y_train,y_test=train_test_split(
df.drop("class",axis=1),
df["class"],
test_size=0.2
)

Dt=DecisionTreeClassifier()
Dt.fit(x_train,y_train)
```

```
DecisionTreeClassifier()
```

```
DecisionTreeClassifier()
```

```
DecisionTreeClassifier()
```

```
from sklearn.tree import plot_tree
plot_tree(Dt,filled=True,max_depth=5)
```

```
[Text(0.43137254901960786, 0.9285714285714286, 'x[1] <= 127.5\ngini = 0.457\nsamples = 614\nvalue = [397, 217]'),
```

```
 Text(0.18872549019607843, 0.7857142857142857, 'x[7] <= 28.5\ngini =
0.301\nsamples = 384\nvalue = [313, 71]'),
 Text(0.31004901960784315, 0.8571428571428572, 'True  '),
 Text(0.09803921568627451, 0.6428571428571429, 'x[5] <= 45.4\ngini =
0.142\nsamples = 221\nvalue = [204, 17]'),
 Text(0.058823529411764705, 0.5, 'x[5] <= 30.95\ngini = 0.121\nsamples
= 217\nvalue = [203, 14]'),
 Text(0.0392156862745098, 0.35714285714285715, 'gini = 0.0\nsamples =
125\nvalue = [125, 0]'),
 Text(0.0784313725490196, 0.35714285714285715, 'x[6] <= 0.509\ngini =
0.258\nsamples = 92\nvalue = [78, 14]'),
 Text(0.0392156862745098, 0.21428571428571427, 'x[5] <= 31.1\ngini =
0.128\nsamples = 58\nvalue = [54, 4]'),
 Text(0.0196078431372549, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.058823529411764705, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.11764705882352941, 0.21428571428571427, 'x[6] <= 0.544\ngini =
0.415\nsamples = 34\nvalue = [24, 10]'),
 Text(0.09803921568627451, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.13725490196078433, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.13725490196078433, 0.5, 'x[4] <= 180.0\ngini = 0.375\nsamples
= 4\nvalue = [1, 3]'),
 Text(0.11764705882352941, 0.35714285714285715, 'gini = 0.0\nsamples =
3\nvalue = [0, 3]'),
 Text(0.1568627450980392, 0.35714285714285715, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.27941176470588236, 0.6428571428571429, 'x[5] <= 26.35\ngini =
0.443\nsamples = 163\nvalue = [109.0, 54.0]'),
 Text(0.21568627450980393, 0.5, 'x[5] <= 9.8\ngini = 0.111\nsamples =
34\nvalue = [32, 2]'),
 Text(0.19607843137254902, 0.35714285714285715, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.23529411764705882, 0.35714285714285715, 'gini = 0.0\nsamples =
32\nvalue = [32, 0]'),
 Text(0.3431372549019608, 0.5, 'x[6] <= 0.636\ngini = 0.481\nsamples =
129\nvalue = [77, 52]'),
 Text(0.27450980392156865, 0.35714285714285715, 'x[1] <= 101.5\ngini =
0.427\nsamples = 97\nvalue = [67, 30]'),
 Text(0.23529411764705882, 0.21428571428571427, 'x[1] <= 31.0\ngini =
0.213\nsamples = 33\nvalue = [29, 4]'),
 Text(0.21568627450980393, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.2549019607843137, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.3137254901960784, 0.21428571428571427, 'x[6] <= 0.2\ngini =
0.482\nsamples = 64\nvalue = [38, 26]'),
 Text(0.29411764705882354, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.3333333333333333, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.4117647058823529, 0.35714285714285715, 'x[7] <= 42.0\ngini =
0.43\nsamples = 32\nvalue = [10, 22]'),
 Text(0.39215686274509803, 0.21428571428571427, 'x[3] <= 43.0\ngini =
0.499\nsamples = 21\nvalue = [10, 11]'),
```

```
 Text(0.37254901960784315, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.4117647058823529, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.43137254901960786, 0.21428571428571427, 'gini = 0.0\nsamples =
11\nvalue = [0, 11]'),
 Text(0.6740196078431373, 0.7857142857142857, 'x[5] <= 29.95\ngini =
0.464\nsamples = 230\nvalue = [84, 146]'),
 Text(0.5526960784313726, 0.8571428571428572, '  False'),
 Text(0.5490196078431373, 0.6428571428571429, 'x[1] <= 145.5\ngini =
0.44\nsamples = 55\nvalue = [37, 18]'),
 Text(0.5098039215686274, 0.5, 'x[2] <= 73.0\ngini = 0.245\nsamples =
28\nvalue = [24, 4]'),
 Text(0.49019607843137253, 0.35714285714285715, 'x[2] <= 69.0\ngini =
0.391\nsamples = 15\nvalue = [11, 4]'),
 Text(0.47058823529411764, 0.21428571428571427, 'x[1] <= 129.5\ngini =
0.26\nsamples = 13\nvalue = [11, 2]'),
 Text(0.45098039215686275, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.49019607843137253, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.5098039215686274, 0.21428571428571427, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.5294117647058824, 0.35714285714285715, 'gini = 0.0\nsamples =
13\nvalue = [13, 0]'),
 Text(0.5882352941176471, 0.5, 'x[7] <= 25.5\ngini = 0.499\nsamples =
27\nvalue = [13, 14]'),
 Text(0.5686274509803921, 0.35714285714285715, 'gini = 0.0\nsamples =
4\nvalue = [4, 0]'),
 Text(0.6078431372549019, 0.35714285714285715, 'x[7] <= 59.0\ngini =
0.476\nsamples = 23\nvalue = [9, 14]'),
 Text(0.5882352941176471, 0.21428571428571427, 'x[2] <= 73.0\ngini =
0.388\nsamples = 19\nvalue = [5, 14]'),
 Text(0.5686274509803921, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.6078431372549019, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.6274509803921569, 0.21428571428571427, 'gini = 0.0\nsamples =
4\nvalue = [4, 0]'),
 Text(0.7990196078431373, 0.6428571428571429, 'x[1] <= 157.5\ngini =
0.393\nsamples = 175\nvalue = [47, 128]'),
 Text(0.6862745098039216, 0.5, 'x[2] <= 61.0\ngini = 0.484\nsamples =
95\nvalue = [39.0, 56.0]'),
 Text(0.6666666666666666, 0.35714285714285715, 'gini = 0.0\nsamples =
10\nvalue = [0, 10]'),
 Text(0.7058823529411765, 0.35714285714285715, 'x[7] <= 30.5\ngini =
0.497\nsamples = 85\nvalue = [39, 46]'),
 Text(0.6666666666666666, 0.21428571428571427, 'x[2] <= 88.0\ngini =
0.444\nsamples = 33\nvalue = [22, 11]'),
 Text(0.6470588235294118, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.6862745098039216, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.7450980392156863, 0.21428571428571427, 'x[6] <= 0.429\ngini =
0.44\nsamples = 52\nvalue = [17, 35]'),
 Text(0.7254901960784313, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.7647058823529411, 0.07142857142857142, '\n  (...)  \n'),
```

```
 Text(0.9117647058823529, 0.5, 'x[4] <= 629.5\ngini = 0.18\nsamples =
80\nvalue = [8, 72]'),
 Text(0.8627450980392157, 0.35714285714285715, 'x[3] <= 26.5\ngini =
0.144\nsamples = 77\nvalue = [6, 71]'),
 Text(0.8235294117647058, 0.21428571428571427, 'x[7] <= 26.5\ngini =
0.245\nsamples = 35\nvalue = [5, 30]'),
 Text(0.803921568627451, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.8431372549019608, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.9019607843137255, 0.21428571428571427, 'x[5] <= 33.05\ngini =
0.046\nsamples = 42\nvalue = [1, 41]'),
 Text(0.8823529411764706, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.9215686274509803, 0.07142857142857142, '\n  (...)  \n'),
 Text(0.9607843137254902, 0.35714285714285715, 'x[7] <= 45.0\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
 Text(0.9411764705882353, 0.21428571428571427, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.9803921568627451, 0.21428571428571427, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]')]
```