

In [11]: `import pandas as pd`

In [12]: `dataset=pd.read_csv("insurance_pre.csv")  
#dataset = pd.get_dummies(dataset['smoker'])`

In [13]: `dataset`

Out[13]:

	age	sex	bmi	children	smoker	charges
0	19	female	27.900	0	yes	16884.92400
1	18	male	33.770	1	no	1725.55230
2	28	male	33.000	3	no	4449.46200
3	33	male	22.705	0	no	21984.47061
4	32	male	28.880	0	no	3866.85520
...	...	...	...	...	...	...
1333	50	male	30.970	3	no	10600.54830
1334	18	female	31.920	0	no	2205.98080
1335	18	female	36.850	0	no	1629.83350
1336	21	female	25.800	0	no	2007.94500
1337	61	female	29.070	0	yes	29141.36030

1338 rows x 6 columns

In [14]: `dataset.columns`

Out[14]: `Index(['age', 'sex', 'bmi', 'children', 'smoker', 'charges'], dtype='object')`

In [15]: `dataset = pd.get_dummies(dataset,drop_first=True)  
dataset=dataset.astype(int)`

In [16]: `dataset`

Out[16]:

	age	bmi	children	charges	sex_male	smoker_yes
0	19	27	0	16884	0	1
1	18	33	1	1725	1	0
2	28	33	3	4449	1	0
3	33	22	0	21984	1	0
4	32	28	0	3866	1	0
...	...	...	...	...	...	...
1333	50	30	3	10600	1	0
1334	18	31	0	2205	0	0
1335	18	36	0	1629	0	0
1336	21	25	0	2007	0	0
1337	61	29	0	29141	0	1

1338 rows x 6 columns

In [17]: `dataset.columns`

Out[17]: `Index(['age', 'bmi', 'children', 'charges', 'sex_male', 'smoker_yes'], dtype='object')`

In [18]: `independent=dataset[['age', 'bmi', 'children', 'charges', 'sex_male', 'smoker_yes']]  
dependent=dataset[['charges']]`

In [19]: `independent`

Out[19]:

	age	bmi	children	charges	sex_male	smoker_yes
0	19	27	0	16884	0	1
1	18	33	1	1725	1	0
2	28	33	3	4449	1	0
3	33	22	0	21984	1	0
4	32	28	0	3866	1	0
...	...	...	...	...	...	...
1333	50	30	3	10600	1	0
1334	18	31	0	2205	0	0
1335	18	36	0	1629	0	0
1336	21	25	0	2007	0	0
1337	61	29	0	29141	0	1

1338 rows x 6 columns

In [20]: `dependent`

Out[20]:

	charges
0	16884
1	1725
2	4449
3	21984
4	3866
...	...
1333	10600
1334	2205
1335	1629
1336	2007
1337	29141

1338 rows x 1 columns

In [21]: `from sklearn.model_selection import train_test_split  
X_train,X_test,y_train,y_test=train_test_split(independent,dependent,test_size=1/3,random_state=0)`

In [22]: `X_train`

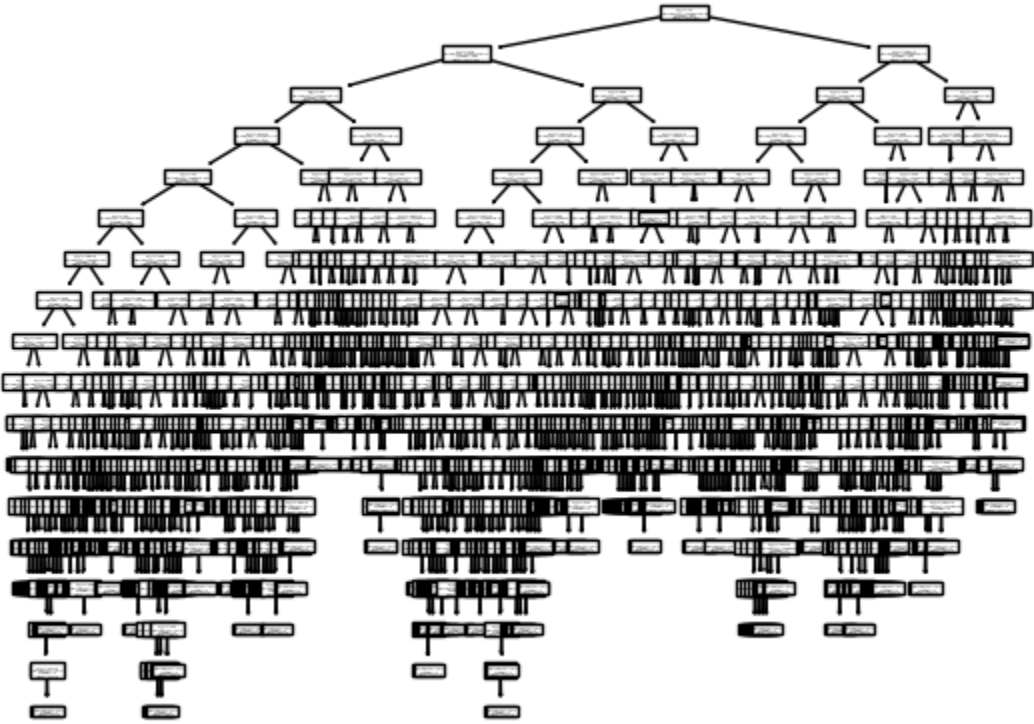
Out[22]:

	age	bmi	children	charges	sex_male	smoker_yes
482	18	31	0	1622	0	0
338	50	32	1	41919	1	1
356	46	43	3	8944	1	0
869	25	24	3	4391	0	0
182	22	19	3	4005	1	0
...	...	...	...	...	...	...
763	27	26	0	3070	1	0
835	42	35	2	7160	1	0
1216	40	25	0	5415	1	0
559	19	35	0	1646	1	0
684	33	18	1	4766	0	0

892 rows x 6 columns

In [23]: `from sklearn.tree import DecisionTreeRegressor  
regressor=DecisionTreeRegressor(criterion='squared_error',splitter='best',max_features='sqrt')  
regressor=regressor.fit(X_train,y_train)`

In [24]: `import matplotlib.pyplot as plt  
from sklearn import tree  
tree.plot_tree(regressor)  
plt.show()`



In [25]: `y_pred=regressor.predict(X_test)`

In [26]: `from sklearn.metrics import r2_score  
r_score=r2_score(y_test,y_pred)`

In [27]: `r_score`

Out[27]: `0.9620858945916677`

In [28]: `import pickle  
filename="finalized_model_Mul_linear.sav"  
pickle.dump(regressor,open(filename,'wb'))`

```
In [29]: loaded_model=pickle.load(open("finalized_model_Mul_linear.sav", 'rb'))
```

```
In [ ]:
```