In [1]:
```
import pandas as pd
```

In [3]:
```
dataset=pd.read_csv("insurance_pre.csv")
```

In [5]:
```
dataset=pd.get_dummies(dataset,drop_first=True)
```

In [7]:
```
dataset
```

Out[7]:

	age	bmi	children	charges	sex_male	smoker_yes
0	19	27.900	0	16884.92400	False	True
1	18	33.770	1	1725.55230	True	False
2	28	33.000	3	4449.46200	True	False
3	33	22.705	0	21984.47061	True	False
4	32	28.880	0	3866.85520	True	False
...	...	...	...	...	...	...
1333	50	30.970	3	10600.54830	True	False
1334	18	31.920	0	2205.98080	False	False
1335	18	36.850	0	1629.83350	False	False
1336	21	25.800	0	2007.94500	False	False
1337	61	29.070	0	29141.36030	False	True

1338 rows × 6 columns

In [9]:
```
dataset.columns
```

Out[9]:
```
Index(['age', 'bmi', 'children', 'charges', 'sex_male', 'smoker_yes'], dtype='object')
```

In [11]:
```
dataset=dataset.astype(int)
dataset
```

Out[11]:

	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 × 6 columns

In [13]:
```
indep=dataset[['age', 'bmi', 'children', 'charges', 'sex_male', 'smoker_yes']]
dep=dataset[['charges']]
```

In [15]:
```
indep
```

Out[15]:

	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 × 6 columns

In [17]:
```
dep
```

Out[17]:

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

1338 rows × 1 columns

In [19]:
```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(indep,dep,test_size=1/3,random_state=0)
```

In [21]:
```
from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor(n_estimators=10, random_state=0)
regressor.fit(X_train, y_train)
```

```
C:\Users\ADMIN\anaconda3\Lib\site-packages\sklearn\base.py:1473: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
  return fit_method(estimator, *args, **kwargs)
```

Out[21]:
```
RandomForestRegressor(n_estimators=10, random_state=0)
```

In [27]:
```
y_pred=regressor.predict(X_test)
```

In [29]:
```
from sklearn.metrics import r2_score
r_score=r2_score(y_test,y_pred)
```

In [31]:
```
r_score
```

Out[31]:
```
0.9998234719339677
```

In [33]:
```
import pickle
filename="finalized_model_RandomForest.sav"
pickle.dump(regressor,open(filename,'wb'))
```

In [35]:
```
loaded_model=pickle.load(open("finalized_model_RandomForest.sav",'rb'))
```

In [40]:
```
result=loaded_model.predict([[1234,345,4565,1,12,0]])
result
```

```
C:\Users\ADMIN\anaconda3\Lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but RandomForestRegressor was fitted with feature names
  warnings.warn(
```

Out[40]:
```
array([1222.9])
```

In [ ]:

In [ ]: