

**23/10/2023**

```
In [1]: import sklearn
import pandas as pd
from sklearn import datasets
```

```
In [25]: df = pd.read_csv('room.csv')
df
```

<b>8</b>	417	5	20	4020
<b>9</b>	366	4	20	3306
<b>10</b>	354	10	17	2617
<b>11</b>	476	7	17	4283
<b>12</b>	462	8	19	3886
<b>13</b>	349	3	19	4223
<b>14</b>	438	9	18	2612
<b>15</b>	493	10	19	3496
<b>16</b>	373	8	15	2785
<b>17</b>	416	4	16	3498
<b>18</b>	305	6	20	3642
<b>19</b>	359	2	20	3407
<b>20</b>	352	7	15	2507

```
In [26]: df = datasets.load_iris()
df
```

```
Out[26]: {'data': array([[5.1, 3.5, 1.4, 0.2],  
    [4.9, 3. , 1.4, 0.2],  
    [4.7, 3.2, 1.3, 0.2],  
    [4.6, 3.1, 1.5, 0.2],  
    [5. , 3.6, 1.4, 0.2],  
    [5.4, 3.9, 1.7, 0.4],  
    [4.6, 3.4, 1.4, 0.3],  
    [5. , 3.4, 1.5, 0.2],  
    [4.4, 2.9, 1.4, 0.2],  
    [4.9, 3.1, 1.5, 0.1],  
    [5.4, 3.7, 1.5, 0.2],  
    [4.8, 3.4, 1.6, 0.2],  
    [4.8, 3. , 1.4, 0.1],  
    [4.3, 3. , 1.1, 0.1],  
    [5.8, 4. , 1.2, 0.2],  
    [5.7, 4.4, 1.5, 0.4],  
    [5.4, 3.9, 1.3, 0.4],  
    [5.1, 3.5, 1.4, 0.3],  
    [5.7, 3.8, 1.7, 0.3],  
    [5.1, 3.8, 1.5, 0.2],
```

```
In [27]: df.feature_names
```

```
Out[27]: ['sepal length (cm)',
          'sepal width (cm)',
          'petal length (cm)',
          'petal width (cm)']
```

```
In [28]: df = datasets.load_diabetes()
df
```

```
[ 0.04170844,  0.05068012, -0.01590626, ..., -0.01107952,
 -0.04688253,  0.01549073],
 [-0.04547248, -0.04464164,  0.03906215, ...,  0.02655962,
  0.04452873, -0.02593034],
 [-0.04547248, -0.04464164, -0.0730303 , ..., -0.03949338,
 -0.00422151,  0.00306441]],
 'target': array([151.,  75., 141., 206., 135.,  97., 138.,  63., 110., 31
0., 101.,
 69., 179., 185., 118., 171., 166., 144.,  97., 168.,  68.,  49.,
 68., 245., 184., 202., 137.,  85., 131., 283., 129.,  59., 341.,
 87.,  65., 102., 265., 276., 252.,  90., 100.,  55.,  61.,  92.,
259.,  53., 190., 142.,  75., 142., 155., 225.,  59., 104., 182.,
128.,  52.,  37., 170., 170.,  61., 144.,  52., 128.,  71., 163.,
150.,  97., 160., 178.,  48., 270., 202., 111.,  85.,  42., 170.,
200., 252., 113., 143.,  51.,  52., 210.,  65., 141.,  55., 134.,
 42., 111.,  98., 164.,  48.,  96.,  90., 162., 150., 279.,  92.,
 83., 128., 102., 302., 198.,  95.,  53., 134., 144., 232.,  81.,
104.,  59., 246., 297., 258., 229., 275., 281., 179., 200., 200.,
173., 180.,  84., 121., 161.,  99., 109., 115., 268., 274., 158.,
107.,  83., 103., 272.,  85., 280., 336., 281., 118., 317., 235.,
```

```
In [47]: df = pd.read_csv('room.csv')
df
```

Out[47]:

	area	bedrom	age	price
0	129	2	10	2300
1	159	4	5	2500
2	200	3	20	3000
3	131	5	10	2300
4	300	2	15	4000
5	174	2	22	2900
6	332	3	16	2935
7	345	7	16	2879
8	417	5	20	4020
9	366	4	20	3306
10	354	10	17	2617
11	476	7	17	4283
12	462	8	19	3886
13	349	3	19	4223
14	438	9	18	2612
15	493	10	19	3496
16	373	8	15	2785
17	416	4	16	3498
18	305	6	20	3642
19	359	2	20	3407
20	352	7	15	2507

```
In [49]: x =df.drop('price', axis='columns')
```

```
In [50]: y = df['age']
```

```
In [51]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.4)
```

```
In [52]: x_test
```

```
Out[52]:
```

	area	bedrom	age
<b>4</b>	300	2	15
<b>11</b>	476	7	17
<b>12</b>	462	8	19
<b>5</b>	174	2	22
<b>10</b>	354	10	17
<b>1</b>	159	4	5
<b>3</b>	131	5	10
<b>20</b>	352	7	15
<b>2</b>	200	3	20

```
In [53]: y_test
```

```
Out[53]: 4      15
11      17
12      19
5       22
10      17
1        5
3       10
20      15
2       20
Name: age, dtype: int64
```

```
In [54]: x_train
```

```
Out[54]:
```

	area	bedrom	age
<b>9</b>	366	4	20
<b>7</b>	345	7	16
<b>13</b>	349	3	19
<b>17</b>	416	4	16
<b>15</b>	493	10	19
<b>19</b>	359	2	20
<b>8</b>	417	5	20
<b>6</b>	332	3	16
<b>0</b>	129	2	10
<b>16</b>	373	8	15
<b>18</b>	305	6	20
<b>14</b>	438	9	18

```
In [55]: y_train
```

```
Out[55]: 9      20
          7      16
          13     19
          17     16
          15     19
          19     20
          8      20
          6      16
          0      10
          16     15
          18     20
          14     18
          Name: age, dtype: int64
```

```
In [56]: from sklearn import linear_model
          reg = linear_model.LinearRegression()
```

```
In [57]: reg.fit(x_train, y_train)
```

```
Out[57]: 

LinearRegression


          LinearRegression()
```

```
In [58]: reg.predict(x_test)
```

```
Out[58]: array([15., 17., 19., 22., 17.,  5., 10., 15., 20.])
```

```
In [59]: y_pred = reg.predict(x_test)
          y_pred
```

```
Out[59]: array([15., 17., 19., 22., 17.,  5., 10., 15., 20.])
```

```
In [60]: y_test
```

```
Out[60]: 4      15
          11     17
          12     19
          5      22
          10     17
          1       5
          3      10
          20     15
          2      20
          Name: age, dtype: int64
```

```
In [61]: from sklearn.metrics import mean_squared_error, r2_score  
print(mean_squared_error(y_test, y_pred))
```

2.1737500499423438e-29

```
In [62]: # Accuracy  
r2_score(y_test, y_pred)*100
```

Out[62]: 100.0

```
In [63]: y_train
```

Out[63]:

9	20
7	16
13	19
17	16
15	19
19	20
8	20
6	16
0	10
16	15
18	20
14	18

Name: age, dtype: int64

```
In [64]: reg.coef_
```

Out[64]: array([6.40545546e-18, 4.44089210e-16, 1.00000000e+00])

```
In [65]: x_train.shape
```

Out[65]: (12, 3)