```
import pandas as pd

import numpy as np

df = pd.read_csv("Iris.csv")

df.head()
```

```
   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
Species
0   1            5.1           3.5            1.4           0.2  Iris-
setosa
1   2            4.9           3.0            1.4           0.2  Iris-
setosa
2   3            4.7           3.2            1.3           0.2  Iris-
setosa
3   4            4.6           3.1            1.5           0.2  Iris-
setosa
4   5            5.0           3.6            1.4           0.2  Iris-
setosa
```

```
df.tail()
```

```
      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
145  146            6.7           3.0            5.2           2.3
146  147            6.3           2.5            5.0           1.9
147  148            6.5           3.0            5.2           2.0
148  149            6.2           3.4            5.4           2.3
149  150            5.9           3.0            5.1           1.8

            Species
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Id             150 non-null    int64
 1   SepalLengthCm  150 non-null    float64
 2   SepalWidthCm   150 non-null    float64
 3   PetalLengthCm  150 non-null    float64
 4   PetalWidthCm   150 non-null    float64
 5   Species        150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

```
df.describe()
```

```
                Id  SepalLengthCm  SepalWidthCm  PetalLengthCm
PetalWidthCm
count  150.000000     150.000000    150.000000     150.000000
150.000000
mean    75.500000       5.843333      3.054000       3.758667
1.198667
std     43.445368       0.828066      0.433594       1.764420
0.763161
min      1.000000       4.300000      2.000000       1.000000
0.100000
25%     38.250000       5.100000      2.800000       1.600000
0.300000
50%     75.500000       5.800000      3.000000       4.350000
1.300000
75%    112.750000       6.400000      3.300000       5.100000
1.800000
max    150.000000       7.900000      4.400000       6.900000
2.500000
```

```
df.shape
```

```
(150, 6)
```

```
df.columns
```

```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm',
'PetalWidthCm',
       'Species'],
      dtype='object')
```

```
y = df['SepalLengthCm']
```

```
y.shape
```

```
(150,)
```

```
y
```

```
0      5.1
1      4.9
2      4.7
3      4.6
4      5.0
      ...
145    6.7
146    6.3
147    6.5
148    6.2
149    5.9
Name: SepalLengthCm, Length: 150, dtype: float64
```

```
x = df[['SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm']]
```

```
x.shape
```

```
(150, 3)
```

```
x
```

```
     SepalWidthCm  PetalLengthCm  PetalWidthCm
0             3.5            1.4           0.2
1             3.0            1.4           0.2
2             3.2            1.3           0.2
3             3.1            1.5           0.2
4             3.6            1.4           0.2
..            ...            ...           ...
145           3.0            5.2           2.3
146           2.5            5.0           1.9
147           3.0            5.2           2.0
148           3.4            5.4           2.3
149           3.0            5.1           1.8

[150 rows x 3 columns]
```

```
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x,y,
test_size=0.3, random_state=2529)
```

```
x_train.shape, x_test.shape, y_train.shape, y_test.shape
```

```
((105, 3), (45, 3), (105,), (45,))
```

```
from sklearn.linear_model import LinearRegression

model = LinearRegression()

model.fit(x_train, y_train)
```

```
LinearRegression()
```

```
y_pred = model.predict(x_test)

y_pred.shape
```

```
(45,)
```

```
y_pred
```

```
array([5.18849599, 4.89259064, 5.53224583, 6.6838822 , 7.28064068,
       6.27900971, 6.91560732, 6.28360666, 6.81822176, 6.03282847,
       6.42305122, 6.22487168, 6.26267397, 7.6873822 , 4.68130074,
       6.19320572, 6.2704963 , 7.15396624, 5.31635898, 4.86211322,
       4.73425023, 4.86211322, 6.69376718, 6.72083619, 5.9800362 ,
       5.50450691, 4.93702672, 6.54972567, 6.64626294, 4.68981415,
```

```
       6.07845309, 6.89194672, 4.94009488, 5.09962383, 7.26753032,
       6.49643593, 5.98463315, 5.18849599, 7.79668959, 4.82193385,
       6.45778535, 6.4860429 , 5.96182084, 5.48595131, 6.9743423 ])
```

```
from sklearn.metrics import mean_squared_error, mean_absolute_error,
mean_absolute_percentage_error, r2_score
```

```
mean_squared_error(y_test, y_pred)
```

```
0.09598309537603604
```

```
mean_absolute_error(y_test, y_pred)
```

```
0.25385493648015306
```

```
mean_absolute_percentage_error(y_test, y_pred)
```

```
0.04152306701868106
```

```
r2_score(y_test, y_pred)
```

```
0.8843157307507274
```

```
df_new = df.sample(1)
```

```
df_new
```

```
    Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
93  94            5.0           2.3            3.3           1.0

            Species
93  Iris-versicolor
```

```
x_new = df_new[['SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm']]
```

```
x_new.shape
```

```
(1, 3)
```

```
y_pred_new = model.predict(x_new)
```

```
y_pred_new
```

```
array([5.15684062])
```