```
import numpy as np
import pandas as pd
import sklearn

from sklearn.datasets import load_boston
df = load_boston()
df.keys()
print(df.DESCR)
print(df.feature_names)
print(df.filename)
print(df.target)
boston = pd. DataFrame(df.data, columns=df.feature_names)
boston.head()
boston['MEDV']=df.target
boston.isnull()
boston.isnull().sum()
from sklearn. model_selection import train_test_split
X=boston.drop('MEDV', axis=1)
Y=boston ['MEDV']
X_train, X_test, Y_train,Y_test = train_test_split(X, Y, test_size = 0.15)
print(X_train.shape)
print(X_test.shape)
print(Y_train.shape)
print(Y_test.shape)
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
lin_model = LinearRegression()
lin_model.fit(X_train, Y_train)
Y_train_predict = lin_model.predict(X_train)
rmse = (np.sqrt(mean_squared_error(Y_train, Y_train_predict)))
print("The models performance for training sets")
print('rmse is {}'.format(rmse))
print("\n")
#on testing set
Y_test_predict = lin_model.predict(X_test)
rmse = (np.sqrt(mean_squared_error(Y_test, Y_test_predict)))
print("the model performance for testing set")
print('rmse is{}'.format(rmse))
```

```
      .. _boston_dataset:

      Boston house prices dataset
      ---------------------------

      **Data Set Characteristics:**

          :Number of Instances: 506

          :Number of Attributes: 13 numeric/categorical predictive. Median Value (att

          :Attribute Information (in order):
              - CRIM     per capita crime rate by town
              - ZN       proportion of residential land zoned for lots over 25,000 sq
              - INDUS    proportion of non-retail business acres per town
              - CHAS     Charles River dummy variable (= 1 if tract bounds river; 0 o
              - NOX      nitric oxides concentration (parts per 10 million)
              - RM       average number of rooms per dwelling
```

```
        - AGE      proportion of owner-occupied units built prior to 1940
        - DIS      weighted distances to five Boston employment centres
        - RAD      index of accessibility to radial highways
        - TAX      full-value property-tax rate per $10,000
        - PTRATIO  pupil-teacher ratio by town
        - B        1000(Bk - 0.63)^2 where Bk is the proportion of black people
        - LSTAT    % lower status of the population
        - MEDV     Median value of owner-occupied homes in $1000's

    :Missing Attribute Values: None

    :Creator: Harrison, D. and Rubinfeld, D.L.

This is a copy of UCI ML housing dataset.
https://archive.ics.uci.edu/ml/machine-learning-databases/housing/


This dataset was taken from the StatLib library which is maintained at Carnegie

The Boston house-price data of Harrison, D. and Rubinfeld, D.L. 'Hedonic
prices and the demand for clean air', J. Environ. Economics & Management,
vol.5, 81-102, 1978.   Used in Belsley, Kuh & Welsch, 'Regression diagnostics
...', Wiley, 1980.   N.B. Various transformations are used in the table on
pages 244-261 of the latter.

The Boston house-price data has been used in many machine learning papers that
problems.

.. topic:: References

   - Belsley, Kuh & Welsch, 'Regression diagnostics: Identifying Influential Da
   - Quinlan,R. (1993). Combining Instance-Based and Model-Based Learning. In P

['CRIM' 'ZN' 'INDUS' 'CHAS' 'NOX' 'RM' 'AGE' 'DIS' 'RAD' 'TAX' 'PTRATIO'
 'B' 'LSTAT']
boston_house_prices.csv
[24.  21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 15.  18.9 21.7 20.4
 18.2 19.9 23.1 17.5 20.2 18.2 13.6 19.6 15.2 14.5 15.6 13.9 16.6 14.8
 18.4 21.  12.7 14.5 13.2 13.1 13.5 18.9 20.  21.  24.7 30.8 34.9 26.6
 25.3 24.7 21.2 19.3 20.  16.6 14.4 19.4 19.7 20.5 25.  23.4 18.9 35.4
```

Colab paid products - Cancel contracts here