```
import numpy as np
import pandas as pd
import sklearn
from sklearn.datasets import load_boston
df=load_boston()
df.keys()
print(df.DESCR)
boston=pd.DataFrame(df.data, columns=df.feature_names)
boston.head()
boston['MEDV']= df.target
boston.head()
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, random_sta
print(X_train,)
print(X_test,)
print(Y_train,)
print(Y_test,)
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_sqaured_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 model performance for training set")
print('RMSE is {}'.format(rmse))
print("\n")
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 (a

    :Attribute Information (in order):
        - CRIM   per capita crime rate by town
        - ZN   proportion of residential land zoned for lots over 25,000
        - INDUS   proportion of non-retail business acres per town
        - CHAS   Charles River dummy variable (= 1 if tract bounds river; (
        - 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 peop
        - 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 Carneg

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 tha
problems.

.. topic:: References

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

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	\
170	1.20742	0.0	19.58	0.0	0.605	5.875	94.6	2.4259	5.0	403.0	
149	2.73397	0.0	19.58	0.0	0.871	5.597	94.9	1.5257	5.0	403.0	
263	0.82526	20.0	3.97	0.0	0.647	7.327	94.5	2.0788	5.0	264.0	
421	7.02259	0.0	18.10	0.0	0.718	6.006	95.3	1.8746	24.0	666.0	
471	4.03841	0.0	18.10	0.0	0.532	6.229	90.7	3.0993	24.0	666.0	