import pandas as pd import numpy as np

from sklearn.model\_selection import train\_test\_split from sklearn.linear\_model import LinearRegression from sklearn.metrics import mean\_squared\_error

df = pd.read\_csv("test.csv") df

ID crim zn indus chas nox rm age dis rad

tax \

0 3 0.02729 0.0 7.07 0 0.469 7.185 61.1 4.9671 2

242

1 6 0.02985 0.0 2.18 0 0.458 6.430 58.7 6.0622 3

222

2 8 0.14455 12.5 7.87 0 0.524 6.172 96.1 5.9505 5

311

3 9 0.21124 12.5 7.87 0 0.524 5.631 100.0 6.0821 5

311

4 10 0.17004 12.5 7.87 0 0.524 6.004 85.9 6.5921 5

311

.. ... ... ... ... ... ... ... ... ... ...

...

168 496 0.17899 0.0 9.69 0 0.585 5.670 28.8 2.7986 6

391

169 497 0.28960 0.0 9.69 0 0.585 5.390 72.9 2.7986 6

391

170 499 0.23912 0.0 9.69 0 0.585 6.019 65.3 2.4091 6

391

171 501 0.22438 0.0 9.69 0 0.585 6.027 79.7 2.4982 6

391

172 505 0.10959 0.0 11.93 0 0.573 6.794 89.3 2.3889 1

273

|  |  |  |  |
| --- | --- | --- | --- |
|  | ptratio | black | lstat |
| 0 | 17.8 | 392.83 | 4.03 |
| 1 | 18.7 | 394.12 | 5.21 |
| 2 | 15.2 | 396.90 | 19.15 |
| 3 | 15.2 | 386.63 | 29.93 |
| 4 | 15.2 | 386.71 | 17.10 |
| .. | ... | ... | ... |
| 168 | 19.2 | 393.29 | 17.60 |
| 169 | 19.2 | 396.90 | 21.14 |
| 170 | 19.2 | 396.90 | 12.92 |
| 171 | 19.2 | 396.90 | 14.33 |
| 172 | 21.0 | 393.45 | 6.48 |

[173 rows x 14 columns]

df.columns

Index(['ID', 'crim', 'zn', 'indus', 'chas', 'nox', 'rm', 'age', 'dis', 'rad',

'tax', 'ptratio', 'black', 'lstat'],

dtype='object')

x = df[['ID', 'crim', 'zn', 'indus', 'chas', 'nox', 'rm', 'age', 'dis', 'rad',

'tax', 'ptratio', 'black']] y = df[['lstat']]

x\_train,x\_test,y\_train,y\_test = train\_test\_split(x,y,test\_size = 0.25,random\_state=5)

model = LinearRegression() model.fit(x\_train,y\_train) LinearRegression()

ypred = model.predict(x\_test) ypred

array([[13.84969408],

[ 2.91067756],

[17.89355729],

[11.50577738],

[ 7.73951116],

[15.32003478],

[17.24849948],

[16.88638394],

[21.86514312],

[ 5.36295925],

[22.53049714],

[ 9.05958768],

[13.09301973],

[ 8.47116192],

[15.96223393],

[11.88108383],

[17.81599329],

[18.703801 ],

[11.92226995],

[10.6135133 ],

[10.65085067],

[17.26097375],

[ 6.87686668],

[12.19818015],

[13.18234864],

model.score(x\_train,y\_train) 0.6965321183220141

model.score(x\_test,y\_test) 0.7226239032603958

mean\_squared\_error(y\_test,ypred) 12.619279506350756

np.sqrt(mean\_squared\_error(y\_test,ypred)) 3.552362524623684

[15.14876483],

[15.91686868],

[17.78293075],

[19.94024598],

[ 9.3603557 ],

[ 0.03893496],

[ 9.90321065],

[18.17161319],

[17.43820404],

[17.68617863],

[14.47909033],

[11.10312686],

[ 6.44655451],

[ 0.96816851],

[19.11623616],

[ 0.0299939 ],

[21.08744243],

[ 0.31690349],

[ 2.14047963]])