

6qj34tjcj

July 31, 2023

```
[ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[ ]: df=pd.read_csv("/content/8_BreastCancerPrediction.csv")
df
```

```
[ ]:
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	\
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	
..	
564	926424	M	21.56	22.39	142.00	1479.0	
565	926682	M	20.13	28.25	131.20	1261.0	
566	926954	M	16.60	28.08	108.30	858.1	
567	927241	M	20.60	29.33	140.10	1265.0	
568	92751	B	7.76	24.54	47.92	181.0	

	smoothness_mean	compactness_mean	concavity_mean	concave	points_mean	\
0	0.11840	0.27760	0.30010		0.14710	
1	0.08474	0.07864	0.08690		0.07017	
2	0.10960	0.15990	0.19740		0.12790	
3	0.14250	0.28390	0.24140		0.10520	
4	0.10030	0.13280	0.19800		0.10430	
..	
564	0.11100	0.11590	0.24390		0.13890	
565	0.09780	0.10340	0.14400		0.09791	
566	0.08455	0.10230	0.09251		0.05302	
567	0.11780	0.27700	0.35140		0.15200	
568	0.05263	0.04362	0.00000		0.00000	

	texture_worst	perimeter_worst	area_worst	smoothness_worst	\
0	17.33	184.60	2019.0	0.16220	
1	23.41	158.80	1956.0	0.12380	

2	...	25.53	152.50	1709.0	0.14440
3	...	26.50	98.87	567.7	0.20980
4	...	16.67	152.20	1575.0	0.13740
..
564	...	26.40	166.10	2027.0	0.14100
565	...	38.25	155.00	1731.0	0.11660
566	...	34.12	126.70	1124.0	0.11390
567	...	39.42	184.60	1821.0	0.16500
568	...	30.37	59.16	268.6	0.08996

	compactness_worst	concavity_worst	concave points_worst	symmetry_worst	\
0	0.66560	0.7119	0.2654	0.4601	
1	0.18660	0.2416	0.1860	0.2750	
2	0.42450	0.4504	0.2430	0.3613	
3	0.86630	0.6869	0.2575	0.6638	
4	0.20500	0.4000	0.1625	0.2364	
..	
564	0.21130	0.4107	0.2216	0.2060	
565	0.19220	0.3215	0.1628	0.2572	
566	0.30940	0.3403	0.1418	0.2218	
567	0.86810	0.9387	0.2650	0.4087	
568	0.06444	0.0000	0.0000	0.2871	

	fractal_dimension_worst	Unnamed: 32
0	0.11890	NaN
1	0.08902	NaN
2	0.08758	NaN
3	0.17300	NaN
4	0.07678	NaN
..
564	0.07115	NaN
565	0.06637	NaN
566	0.07820	NaN
567	0.12400	NaN
568	0.07039	NaN

[569 rows x 33 columns]

```
[ ]: df.head()
```

```
[ ]:
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	\
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	

	smoothness_mean	compactness_mean	concavity_mean	concave	points_mean	\
0	0.11840	0.27760	0.3001		0.14710	
1	0.08474	0.07864	0.0869		0.07017	
2	0.10960	0.15990	0.1974		0.12790	
3	0.14250	0.28390	0.2414		0.10520	
4	0.10030	0.13280	0.1980		0.10430	

	... texture_worst	perimeter_worst	area_worst	smoothness_worst	\
0	... 17.33	184.60	2019.0	0.1622	
1	... 23.41	158.80	1956.0	0.1238	
2	... 25.53	152.50	1709.0	0.1444	
3	... 26.50	98.87	567.7	0.2098	
4	... 16.67	152.20	1575.0	0.1374	

	compactness_worst	concavity_worst	concave	points_worst	symmetry_worst	\
0	0.6656	0.7119		0.2654	0.4601	
1	0.1866	0.2416		0.1860	0.2750	
2	0.4245	0.4504		0.2430	0.3613	
3	0.8663	0.6869		0.2575	0.6638	
4	0.2050	0.4000		0.1625	0.2364	

	fractal_dimension_worst	Unnamed: 32
0	0.11890	NaN
1	0.08902	NaN
2	0.08758	NaN
3	0.17300	NaN
4	0.07678	NaN

[5 rows x 33 columns]

1 DATA CLEANING AND DATA PREPROCESSING

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 33 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    569 non-null    int64
1   diagnosis             569 non-null    object
2   radius_mean           569 non-null    float64
3   texture_mean          569 non-null    float64
4   perimeter_mean        569 non-null    float64
5   area_mean             569 non-null    float64
6   smoothness_mean       569 non-null    float64
```

```

7   compactness_mean      569 non-null    float64
8   concavity_mean        569 non-null    float64
9   concave points_mean   569 non-null    float64
10  symmetry_mean         569 non-null    float64
11  fractal_dimension_mean 569 non-null    float64
12  radius_se             569 non-null    float64
13  texture_se            569 non-null    float64
14  perimeter_se          569 non-null    float64
15  area_se               569 non-null    float64
16  smoothness_se         569 non-null    float64
17  compactness_se        569 non-null    float64
18  concavity_se          569 non-null    float64
19  concave points_se     569 non-null    float64
20  symmetry_se           569 non-null    float64
21  fractal_dimension_se   569 non-null    float64
22  radius_worst          569 non-null    float64
23  texture_worst         569 non-null    float64
24  perimeter_worst       569 non-null    float64
25  area_worst            569 non-null    float64
26  smoothness_worst      569 non-null    float64
27  compactness_worst     569 non-null    float64
28  concavity_worst       569 non-null    float64
29  concave points_worst  569 non-null    float64
30  symmetry_worst        569 non-null    float64
31  fractal_dimension_worst 569 non-null    float64
32  Unnamed: 32           0 non-null      float64

```

dtypes: float64(31), int64(1), object(1)

memory usage: 146.8+ KB

```
[ ]: df.describe()
```

```

[ ]:
      count      id  radius_mean  texture_mean  perimeter_mean  area_mean  \
count  5.690000e+02  569.000000   569.000000    569.000000    569.000000
mean    3.037183e+07  14.127292   19.289649    91.969033    654.889104
std     1.250206e+08   3.524049    4.301036    24.298981    351.914129
min     8.670000e+03   6.981000    9.710000    43.790000    143.500000
25%     8.692180e+05  11.700000   16.170000    75.170000    420.300000
50%     9.060240e+05  13.370000   18.840000    86.240000    551.100000
75%     8.813129e+06  15.780000   21.800000   104.100000    782.700000
max     9.113205e+08  28.110000   39.280000   188.500000   2501.000000

      smoothness_mean  compactness_mean  concavity_mean  concave points_mean  \
count      569.000000      569.000000      569.000000      569.000000
mean         0.096360         0.104341         0.088799         0.048919
std          0.014064         0.052813         0.079720         0.038803
min          0.052630         0.019380         0.000000         0.000000
25%          0.086370         0.064920         0.029560         0.020310

```

50%	0.095870	0.092630	0.061540	0.033500
75%	0.105300	0.130400	0.130700	0.074000
max	0.163400	0.345400	0.426800	0.201200

	symmetry_mean	...	texture_worst	perimeter_worst	area_worst	\
count	569.000000	...	569.000000	569.000000	569.000000	
mean	0.181162	...	25.677223	107.261213	880.583128	
std	0.027414	...	6.146258	33.602542	569.356993	
min	0.106000	...	12.020000	50.410000	185.200000	
25%	0.161900	...	21.080000	84.110000	515.300000	
50%	0.179200	...	25.410000	97.660000	686.500000	
75%	0.195700	...	29.720000	125.400000	1084.000000	
max	0.304000	...	49.540000	251.200000	4254.000000	

	smoothness_worst	compactness_worst	concavity_worst	\
count	569.000000	569.000000	569.000000	
mean	0.132369	0.254265	0.272188	
std	0.022832	0.157336	0.208624	
min	0.071170	0.027290	0.000000	
25%	0.116600	0.147200	0.114500	
50%	0.131300	0.211900	0.226700	
75%	0.146000	0.339100	0.382900	
max	0.222600	1.058000	1.252000	

	concave	points_worst	symmetry_worst	fractal_dimension_worst	\
count		569.000000	569.000000	569.000000	
mean		0.114606	0.290076	0.083946	
std		0.065732	0.061867	0.018061	
min		0.000000	0.156500	0.055040	
25%		0.064930	0.250400	0.071460	
50%		0.099930	0.282200	0.080040	
75%		0.161400	0.317900	0.092080	
max		0.291000	0.663800	0.207500	

	Unnamed: 32
count	0.0
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

[8 rows x 32 columns]

```
[ ]: df.columns
```

```
[ ]: Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
          'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
          'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
          'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',
          'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',
          'fractal_dimension_se', 'radius_worst', 'texture_worst',
          'perimeter_worst', 'area_worst', 'smoothness_worst',
          'compactness_worst', 'concavity_worst', 'concave points_worst',
          'symmetry_worst', 'fractal_dimension_worst', 'Unnamed: 32'],
          dtype='object')
```

```
[ ]: df1=df.dropna(axis=1)
df1
```

```
[ ]:
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	\
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	
..	
564	926424	M	21.56	22.39	142.00	1479.0	
565	926682	M	20.13	28.25	131.20	1261.0	
566	926954	M	16.60	28.08	108.30	858.1	
567	927241	M	20.60	29.33	140.10	1265.0	
568	92751	B	7.76	24.54	47.92	181.0	

	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	\
0	0.11840	0.27760	0.30010	0.14710	
1	0.08474	0.07864	0.08690	0.07017	
2	0.10960	0.15990	0.19740	0.12790	
3	0.14250	0.28390	0.24140	0.10520	
4	0.10030	0.13280	0.19800	0.10430	
..	
564	0.11100	0.11590	0.24390	0.13890	
565	0.09780	0.10340	0.14400	0.09791	
566	0.08455	0.10230	0.09251	0.05302	
567	0.11780	0.27700	0.35140	0.15200	
568	0.05263	0.04362	0.00000	0.00000	

	radius_worst	texture_worst	perimeter_worst	area_worst	\
0	25.380	17.33	184.60	2019.0	
1	24.990	23.41	158.80	1956.0	
2	23.570	25.53	152.50	1709.0	
3	14.910	26.50	98.87	567.7	
4	22.540	16.67	152.20	1575.0	
..	

564	...	25.450	26.40	166.10	2027.0
565	...	23.690	38.25	155.00	1731.0
566	...	18.980	34.12	126.70	1124.0
567	...	25.740	39.42	184.60	1821.0
568	...	9.456	30.37	59.16	268.6

	smoothness_worst	compactness_worst	concavity_worst	\
0	0.16220	0.66560	0.7119	
1	0.12380	0.18660	0.2416	
2	0.14440	0.42450	0.4504	
3	0.20980	0.86630	0.6869	
4	0.13740	0.20500	0.4000	
..	
564	0.14100	0.21130	0.4107	
565	0.11660	0.19220	0.3215	
566	0.11390	0.30940	0.3403	
567	0.16500	0.86810	0.9387	
568	0.08996	0.06444	0.0000	

	concave points_worst	symmetry_worst	fractal_dimension_worst
0	0.2654	0.4601	0.11890
1	0.1860	0.2750	0.08902
2	0.2430	0.3613	0.08758
3	0.2575	0.6638	0.17300
4	0.1625	0.2364	0.07678
..
564	0.2216	0.2060	0.07115
565	0.1628	0.2572	0.06637
566	0.1418	0.2218	0.07820
567	0.2650	0.4087	0.12400
568	0.0000	0.2871	0.07039

[569 rows x 32 columns]

```
[ ]: df1.columns
```

```
[ ]: Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
          'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
          'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
          'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',
          'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',
          'fractal_dimension_se', 'radius_worst', 'texture_worst',
          'perimeter_worst', 'area_worst', 'smoothness_worst',
          'compactness_worst', 'concavity_worst', 'concave points_worst',
          'symmetry_worst', 'fractal_dimension_worst'],
          dtype='object')
```

```
[ ]: df1=df1[['id', 'radius_mean', 'texture_mean', 'perimeter_mean',  
            'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',  
            'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',  
            'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',  
            'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',  
            'fractal_dimension_se', 'radius_worst', 'texture_worst',  
            'perimeter_worst', 'area_worst', 'smoothness_worst',  
            'compactness_worst', 'concavity_worst', 'concave points_worst',  
            'symmetry_worst', 'fractal_dimension_worst']]
```

2 EDA AND VISUALIZATION

```
[ ]: sns.pairplot(df1)
```

```
[ ]: sns.distplot(df1['fractal_dimension_worst'])
```

<ipython-input-11-a88705766b47>:1: UserWarning:

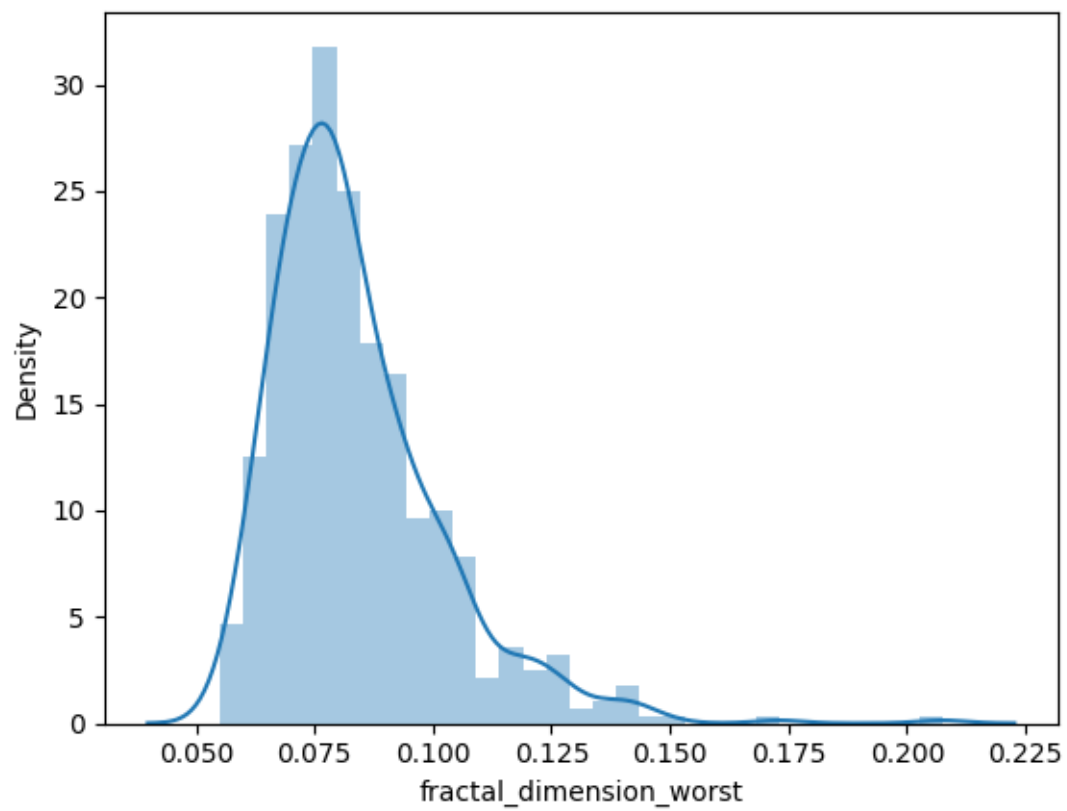
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

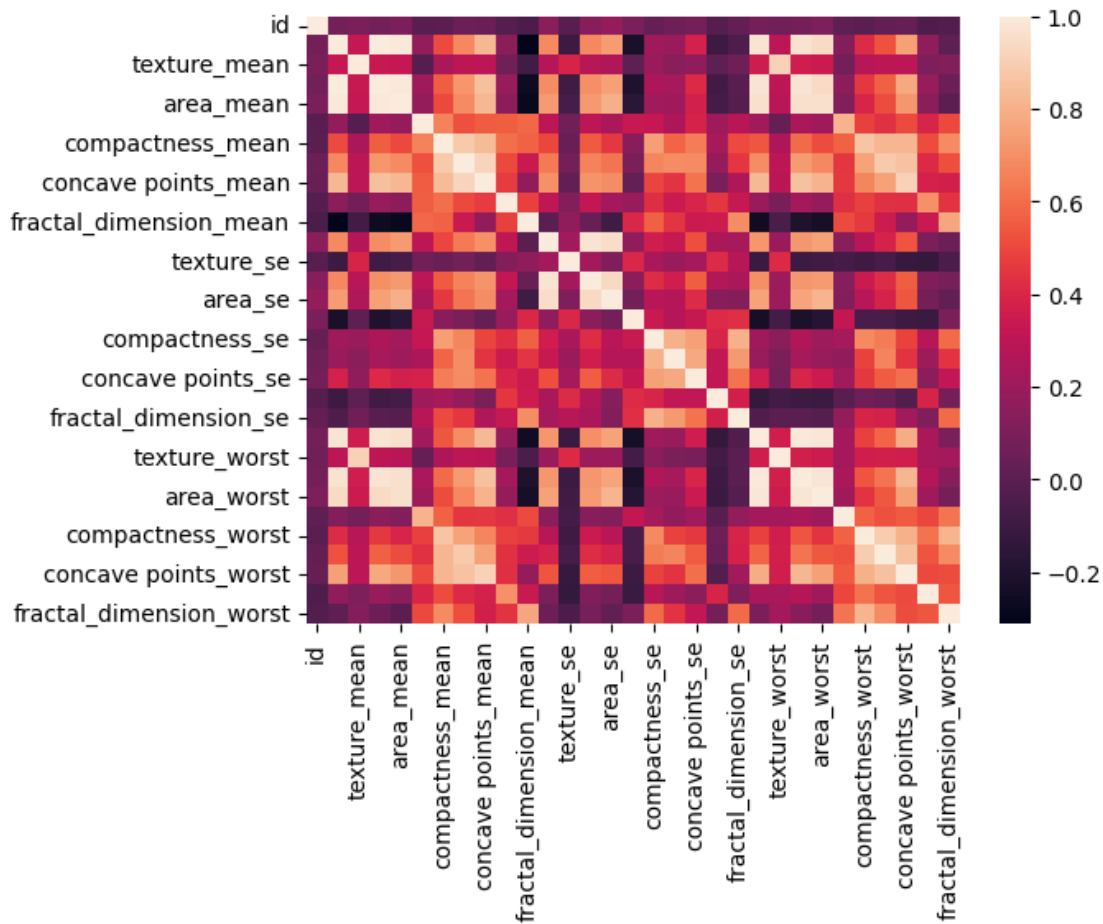
```
sns.distplot(df1['fractal_dimension_worst'])
```

```
[ ]: <Axes: xlabel='fractal_dimension_worst', ylabel='Density'>
```

```
[ ]: sns.heatmap(df1.corr())
```

```
[ ]: <Axes: >
```



3 TO TRAIN THE MODEL AND MODEL BUILDING

```
[ ]: x=df[['id', 'radius_mean', 'texture_mean', 'perimeter_mean',
          'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
          'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
          'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',
          'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',
          'fractal_dimension_se', 'radius_worst', 'texture_worst',
          'perimeter_worst', 'area_worst', 'smoothness_worst',
          'compactness_worst', 'concavity_worst', 'concave points_worst',
          'symmetry_worst']]
y=df['fractal_dimension_worst']

[ ]: 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)
```

```
[ ]: from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(x_train,y_train)
```

```
[ ]: LinearRegression()
```

```
[ ]: lr.intercept_
```

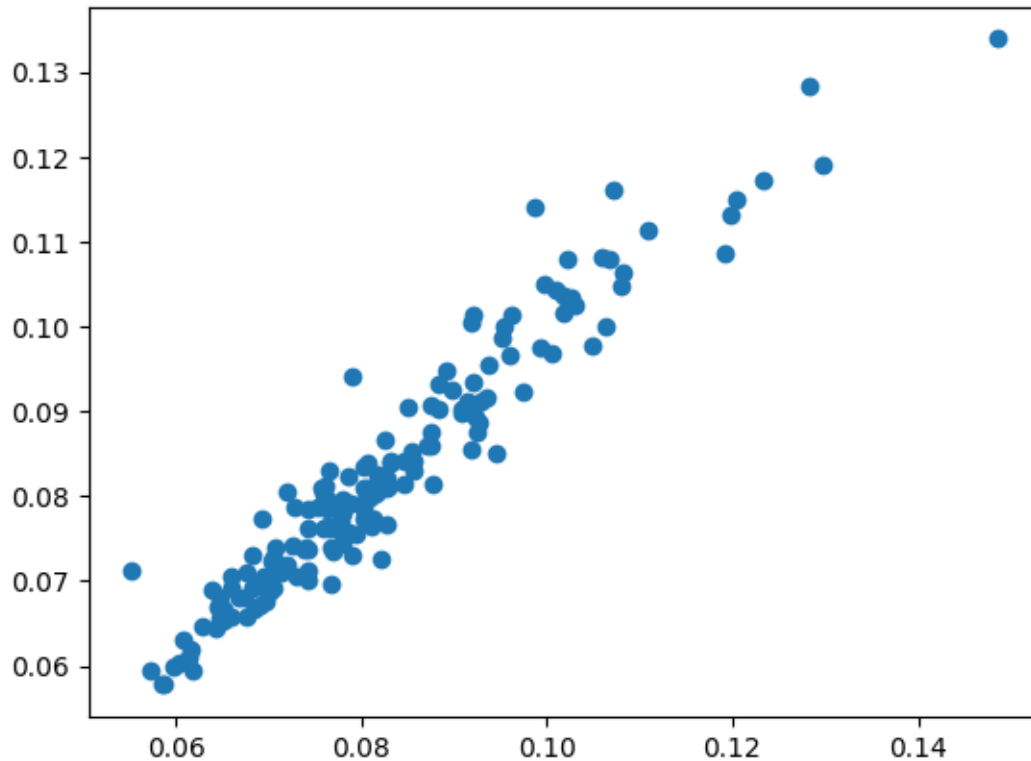
```
[ ]: -0.02522352461946234
```

```
[ ]: coeff=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
coeff
```

```
[ ]:
      Co-efficient
id          -2.368699e-12
radius_mean  -4.511013e-03
texture_mean   9.049861e-05
perimeter_mean  4.266956e-04
area_mean     1.334342e-05
smoothness_mean -4.189968e-02
compactness_mean -1.255826e-01
concavity_mean  -7.826449e-02
concave points_mean  1.226514e-01
symmetry_mean  -9.055320e-03
fractal_dimension_mean  1.353105e+00
radius_se     -2.380030e-02
texture_se     -3.482709e-04
perimeter_se   1.753643e-03
area_se        1.059917e-04
smoothness_se  2.663329e-02
compactness_se -2.229326e-01
concavity_se   -5.444331e-02
concave points_se  3.975217e-03
symmetry_se    -1.310632e-01
fractal_dimension_se  2.501342e+00
radius_worst   4.608262e-03
texture_worst   2.582736e-05
perimeter_worst -3.221786e-04
area_worst     -2.072206e-05
smoothness_worst  3.703889e-02
compactness_worst  8.202149e-02
concavity_worst  3.061067e-02
concave points_worst -2.675547e-02
symmetry_worst   1.846634e-02
```

```
[ ]: prediction =lr.predict(x_test)
plt.scatter(y_test,prediction)
```

```
[ ]: <matplotlib.collections.PathCollection at 0x79c611056590>
```



4 ACCURACY

```
[ ]: lr.score(x_test,y_test)
```

```
[ ]: 0.9233132157702097
```

```
[ ]: lr.score(x_train,y_train)
```

```
[ ]: 0.948976497088863
```

```
[ ]: from sklearn.linear_model import Ridge,Lasso
```

```
[ ]: rr=Ridge(alpha=10)  
     rr.fit(x_train,y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_ridge.py:216:  
LinAlgWarning: Ill-conditioned matrix (rcond=1.17448e-18): result may not be  
accurate.
```

```
    return linalg.solve(A, Xy, assume_a="pos", overwrite_a=True).T
```

```
[ ]: Ridge(alpha=10)
```

```
[ ]: rr.score(x_test,y_test)
```

```
[ ]: 0.5746650890022684
```

```
[ ]: rr.score(x_train,y_train)
```

```
[ ]: 0.7336880934371385
```

```
[ ]: la=Lasso(alpha=10)
     la.fit(x_train,y_train)
```

```
[ ]: Lasso(alpha=10)
```

```
[ ]: la.score(x_train,y_train)
```

```
[ ]: 0.005397561150582653
```

```
[ ]: la.score(x_test,y_test)
```

```
[ ]: -0.052287624054701665
```

```
[ ]: from sklearn.linear_model import ElasticNet
     en=ElasticNet()
     en.fit(x_train,y_train)
```

```
[ ]: ElasticNet()
```

```
[ ]: print(en.coef_)
     print(en.intercept_)
```

```
[-1.24296074e-11 -0.00000000e+00  0.00000000e+00  0.00000000e+00
 -0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00
  0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00
 -0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00
  0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00
  0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00
  1.07358175e-06  0.00000000e+00  0.00000000e+00  0.00000000e+00
  0.00000000e+00  0.00000000e+00]
0.0840957924659967
```

```
[ ]: prediction = en.predict(x_test)
     prediction
```

```
[ ]: array([0.08455865, 0.08457146, 0.08573119, 0.08540216, 0.084625  ,
          0.08471444, 0.08616002, 0.08447038, 0.0850591  , 0.08439037,
```

```

0.08439053, 0.08559143, 0.08490055, 0.08441967, 0.08465697,
0.08382654, 0.08406431, 0.0852778 , 0.08455799, 0.08508447,
0.08351115, 0.08479157, 0.08500984, 0.08576744, 0.08462018,
0.08656732, 0.08477026, 0.08472277, 0.08454943, 0.08580545,
0.08497228, 0.08364913, 0.08449755, 0.08474442, 0.08348542,
0.08454477, 0.08440187, 0.08665969, 0.08507828, 0.08531197,
0.08465772, 0.0847311 , 0.08441888, 0.08585151, 0.08543896,
0.08378377, 0.08497712, 0.08577774, 0.08509188, 0.08475442,
0.08488871, 0.08491039, 0.07353777, 0.08473641, 0.08497154,
0.08472022, 0.08444254, 0.08514717, 0.08453426, 0.08596358,
0.08506703, 0.08465687, 0.08482248, 0.08462712, 0.08478043,
0.08487899, 0.08459543, 0.08455283, 0.08458955, 0.08353322,
0.08490374, 0.08439069, 0.0847297 , 0.08459523, 0.08443922,
0.08464332, 0.08530698, 0.08467034, 0.08456705, 0.08485626,
0.0844411 , 0.07494125, 0.08388402, 0.08436322, 0.08393765,
0.08478575, 0.08529757, 0.08467915, 0.08485589, 0.08495327,
0.08468523, 0.08561536, 0.08628876, 0.08362875, 0.08465727,
0.08499618, 0.08632765, 0.08496507, 0.08452695, 0.08463882,
0.08551529, 0.08497659, 0.08506073, 0.08499094, 0.0842839 ,
0.08468727, 0.08497392, 0.08508191, 0.08490866, 0.08476759,
0.08506164, 0.08530176, 0.08373132, 0.08360219, 0.08477117,
0.08606445, 0.08467656, 0.08483594, 0.0844762 , 0.08544757,
0.08488272, 0.08667415, 0.0846448 , 0.08538389, 0.08490503,
0.08753757, 0.0857577 , 0.08469314, 0.0852482 , 0.08472337,
0.08463553, 0.08567233, 0.08425905, 0.08470207, 0.0846185 ,
0.08459594, 0.0848348 , 0.07393166, 0.07329997, 0.08463528,
0.08437161, 0.08602647, 0.08505777, 0.08443503, 0.08342465,
0.08473964, 0.08464128, 0.08486429, 0.08473853, 0.08469521,
0.08463698, 0.0845812 , 0.08458254, 0.08567895, 0.08458651,
0.08473814, 0.0846879 , 0.08559868, 0.08637464, 0.08538226,
0.08455304, 0.08478669, 0.08456458, 0.08444544, 0.08533675,
0.08350501, 0.08666916, 0.08389358, 0.08540098, 0.08448084,
0.08472828])

```

```
[ ]: en.score(x_test,y_test)
```

```
[ ]: -0.046833447192170086
```

```

[ ]: from sklearn import metrics
print("Mean Absolute Error: ", metrics.mean_absolute_error(y_test,prediction))
print("Mean Squared Error: ", metrics.mean_squared_error(y_test,prediction))
print("Root Mean Squared Error: ", np.sqrt(metrics.
↪mean_squared_error(y_test,prediction)))

```

Mean Absolute Error: 0.01222927830188647

Mean Squared Error: 0.00024742092701344494

Root Mean Squared Error: 0.015729619417310926