In [6]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
import re
from sklearn.datasets import load_digits
```

In [10]:
```
a=pd.read_csv(r"C:\Users\user\Downloads\spi_index_labelled.csv")
```

Out[10]:

	country	iso3c	date	SPI.INDEX.PIL1	SPI.INDEX.PIL2	SPI.INDEX.PIL3	SPI.INDEX.F
0	NaN	NaN	NaN	Pillar 1 - Data Use - Score	Pillar 2 - Data Services - Score	Pillar 3 - Data Products - Score	Pillar 4 - [Sources - S
1	Norway	NOR	2019.0	100	92.2333333333333	77.56875	80.6666666666
2	Italy	ITA	2019.0	100	91.8666666666666	75.2875	81.
3	Austria	AUT	2019.0	100	91.3	74.55	7
4	Poland	POL	2019.0	100	95.1	70.5375	79.7166666666
...	...	...	...	...	...	...	
3484	Virgin Islands (U.S.)	VIR	2004.0	20	NaN	NaN	I
3485	West Bank and Gaza	PSE	2004.0	20	NaN	NaN	I
3486	Yemen, Rep.	YEM	2004.0	20	NaN	NaN	I
3487	Zambia	ZMB	2004.0	40	NaN	NaN	I
3488	Zimbabwe	ZWE	2004.0	20	NaN	NaN	I

3489 rows × 79 columns

In [12]:
```
b=a.fillna(value=221)
```

Out[12]:

	country	iso3c	date	SPI.INDEX.PIL1	SPI.INDEX.PIL2	SPI.INDEX.PIL3	SPI.INDEX.I
0	221	221	221.0	Pillar 1 - Data Use - Score	Pillar 2 - Data Services - Score	Pillar 3 - Data Products - Score	Pillar 4 - I Sources - S
1	Norway	NOR	2019.0	100	92.2333333333333	77.56875	80.6666666666
2	Italy	ITA	2019.0	100	91.8666666666666	75.2875	81.
3	Austria	AUT	2019.0	100	91.3	74.55	7!
4	Poland	POL	2019.0	100	95.1	70.5375	79.7166666666
...	...	...	...	...	...	...	
3484	Virgin Islands (U.S.)	VIR	2004.0	20	221	221	
3485	West Bank and Gaza	PSE	2004.0	20	221	221	
3486	Yemen, Rep.	YEM	2004.0	20	221	221	
3487	Zambia	ZMB	2004.0	40	221	221	
3488	Zimbabwe	ZWE	2004.0	20	221	221	

3489 rows × 79 columns

In [24]:
```
c=b.head(400)
```

Out[24]:

	country	iso3c	date	SPI.INDEX.PIL1	SPI.INDEX.PIL2	SPI.INDEX.PIL3	SPI.INDEX.PII
0	221	221	221.0	Pillar 1 - Data Use - Score	Pillar 2 - Data Services - Score	Pillar 3 - Data Products - Score	Pillar 4 - Da Sources - Sco
1	Norway	NOR	2019.0	100	92.2333333333333	77.56875	80.66666666666
2	Italy	ITA	2019.0	100	91.8666666666666	75.2875	81.8:
3	Austria	AUT	2019.0	100	91.3	74.55	79.:
4	Poland	POL	2019.0	100	95.1	70.5375	79.71666666666
...	...	...	...	...	...	...	
395	Antigua and Barbuda	ATG	2018.0	20	221	52.06875	2:
396	Aruba	ABW	2018.0	60	221	18.45	2:
397	Barbados	BRB	2018.0	80	221	50.425	2:
398	Bermuda	BMU	2018.0	60	221	16.2125	2:
399	British Virgin Islands	VGB	2018.0	40	221	15.69375	2:

400 rows × 79 columns

In [25]:
```
e=c.tail(399)
```

Out[25]:

	country	iso3c	date	SPI.INDEX.PIL1	SPI.INDEX.PIL2	SPI.INDEX.PIL3	SPI.INDEX.PI
1	Norway	NOR	2019.0	100	92.2333333333333	77.56875	80.66666666666
2	Italy	ITA	2019.0	100	91.8666666666666	75.2875	81.8
3	Austria	AUT	2019.0	100	91.3	74.55	79.
4	Poland	POL	2019.0	100	95.1	70.5375	79.71666666666
5	Slovenia	SVN	2019.0	100	96.9333333333333	76.28125	71.44166666666
...	...	...	...	...	...	...	
395	Antigua and Barbuda	ATG	2018.0	20	221	52.06875	2
396	Aruba	ABW	2018.0	60	221	18.45	2
397	Barbados	BRB	2018.0	80	221	50.425	2
398	Bermuda	BMU	2018.0	60	221	16.2125	2
399	British Virgin Islands	VGB	2018.0	40	221	15.69375	2

399 rows × 79 columns

In [26]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 399 entries, 1 to 399
Data columns (total 79 columns):
 #   Column                         Non-Null Count  Dtype  
---  ------                         --------------  -----  
 0   country                        399 non-null    object 
 1   iso3c                          399 non-null    object 
 2   date                           399 non-null    float64
 3   SPI.INDEX.PIL1                 399 non-null    object 
 4   SPI.INDEX.PIL2                 399 non-null    object 
 5   SPI.INDEX.PIL3                 399 non-null    object 
 6   SPI.INDEX.PIL4                 399 non-null    object 
 7   SPI.INDEX.PIL5                 399 non-null    object 
 8   SPI.INDEX                      399 non-null    object 
 9   SPI.DIM1.5.INDEX               399 non-null    object 
 10  SPI.DIM2.1.INDEX               399 non-null    object 
 11  SPI.DIM2.2.INDEX               399 non-null    object 
 12  SPI.DIM2.4.INDEX               399 non-null    object 
 13  SPI.DIM3.1.INDEX               399 non-null    object 
 14  SPI.DIM3.2.INDEX               399 non-null    object 
 15  SPI.DIM3.3.INDEX               399 non-null    object 
 16  SPI.DIM3.4.INDEX               399 non-null    object 
 17  SPI.DIM4.1.CEN.INDEX           399 non-null    object 
 18  SPI.DIM4.1.SVY.INDEX           399 non-null    object 
 19  SPI.DIM4.2.INDEX               399 non-null    object 
 20  SPI.DIM4.3.INDEX               399 non-null    object 
 21  SPI.DIM5.1.INDEX               399 non-null    object 
 22  SPI.DIM5.2.INDEX               399 non-null    object 
 23  SPI.DIM5.5.INDEX               399 non-null    object 
 24  SPI.D1.5.POV                   399 non-null    object 
 25  SPI.D1.5.CHLD.MORT             399 non-null    object 
 26  SPI.D1.5.DT.TDS.DPPF.XP.ZS     399 non-null    object 
 27  SPI.D1.5.SAFE.MAN.WATER        399 non-null    object 
 28  SPI.D1.5.LFP                   399 non-null    object 
 29  SPI.D2.1.GDDS                  399 non-null    object 
 30  SPI.D2.2.Machine.readable      399 non-null    object 
 31  SPI.D2.2.Non.proprietary       399 non-null    object 
 32  SPI.D2.2.Download.options      399 non-null    object 
 33  SPI.D2.2.Metadata.available    399 non-null    object 
 34  SPI.D2.2.Terms.of.use          399 non-null    object 
 35  SPI.D2.2.Openness.subscore     399 non-null    object 
 36  SPI.D2.4.NADA                  399 non-null    object 
 37  SPI.D3.1.POV                   399 non-null    object 
 38  SPI.D3.2.HNGR                  399 non-null    object 
 39  SPI.D3.3.HLTH                  399 non-null    object 
 40  SPI.D3.4.EDUC                  399 non-null    object 
 41  SPI.D3.5.GEND                  399 non-null    object 
 42  SPI.D3.6.WTRS                  399 non-null    object 
 43  SPI.D3.7.ENRG                  399 non-null    object 
 44  SPI.D3.8.WORK                  399 non-null    object 
 45  SPI.D3.9.INDY                  399 non-null    object 
 46  SPI.D3.10.NEQL                 399 non-null    object 
 47  SPI.D3.11.CITY                 399 non-null    object 
 48  SPI.D3.12.CNSP                 399 non-null    object 
```

```
 49  SPI.D3.15.LAND                399 non-null    object 
 50  SPI.D3.16.INST                399 non-null    object 
 51  SPI.D3.17.PTNS                399 non-null    object 
 52  SPI.D3.13.CLMT                399 non-null    object 
 53  SPI.D4.1.1.POPU               399 non-null    object 
 54  SPI.D4.1.2.AGRI               399 non-null    object 
 55  SPI.D4.1.3.BIZZ               399 non-null    object 
 56  SPI.D4.1.4.HOUS               399 non-null    object 
 57  SPI.D4.1.5.AGSVY              399 non-null    object 
 58  SPI.D4.1.6.LABR               399 non-null    object 
 59  SPI.D4.1.7.HLTH               399 non-null    object 
 60  SPI.D4.1.8.BZSVY              399 non-null    object 
 61  SPI.D4.2.3.CRVS               399 non-null    object 
 62  SPI.D4.3.GEO.first.admin.level  399 non-null    object 
 63  SPI.D5.1.DILG                 399 non-null    object 
 64  SPI.D5.2.1.SNAU               399 non-null    object 
 65  SPI.D5.2.2.NABY               399 non-null    object 
 66  SPI.D5.2.3.CNIN               399 non-null    object 
 67  SPI.D5.2.4.CPIBY              399 non-null    object 
 68  SPI.D5.2.5.HOUS               399 non-null    object 
 69  SPI.D5.2.6.EMPL               399 non-null    object 
 70  SPI.D5.2.7.CGOV               399 non-null    object 
 71  SPI.D5.2.8.FINA               399 non-null    object 
 72  SPI.D5.2.9.MONY               399 non-null    object 
 73  SPI.D5.2.10.GSBP              399 non-null    object 
 74  SPI.D5.5.DIFI                 399 non-null    object 
 75  income                        399 non-null    object 
 76  region                        399 non-null    object 
 77  weights                       399 non-null    float64
 78  population                    399 non-null    float64
dtypes: float64(3), object(76)
memory usage: 246.4+ KB
```

In [27]:

Out[27]:

```
Index(['country', 'iso3c', 'date', 'SPI.INDEX.PIL1', 'SPI.INDEX.PIL2',
       'SPI.INDEX.PIL3', 'SPI.INDEX.PIL4', 'SPI.INDEX.PIL5', 'SPI.INDEX',
       'SPI.DIM1.5.INDEX', 'SPI.DIM2.1.INDEX', 'SPI.DIM2.2.INDEX',
       'SPI.DIM2.4.INDEX', 'SPI.DIM3.1.INDEX', 'SPI.DIM3.2.INDEX',
       'SPI.DIM3.3.INDEX', 'SPI.DIM3.4.INDEX', 'SPI.DIM4.1.CEN.INDEX',
       'SPI.DIM4.1.SVY.INDEX', 'SPI.DIM4.2.INDEX', 'SPI.DIM4.3.INDEX',
       'SPI.DIM5.1.INDEX', 'SPI.DIM5.2.INDEX', 'SPI.DIM5.5.INDEX',
       'SPI.D1.5.POV', 'SPI.D1.5.CHLD.MORT', 'SPI.D1.5.DT.TDS.DPPF.XP.ZS',
       'SPI.D1.5.SAFE.MAN.WATER', 'SPI.D1.5.LFP', 'SPI.D2.1.GDDS',
       'SPI.D2.2.Machine.readable', 'SPI.D2.2.Non.proprietary',
       'SPI.D2.2.Download.options', 'SPI.D2.2.Metadata.available',
       'SPI.D2.2.Terms.of.use', 'SPI.D2.2.Openness.subscore', 'SPI.D2.4.NADA
',
       'SPI.D3.1.POV', 'SPI.D3.2.HNGR', 'SPI.D3.3.HLTH', 'SPI.D3.4.EDUC',
       'SPI.D3.5.GEND', 'SPI.D3.6.WTRS', 'SPI.D3.7.ENRG', 'SPI.D3.8.WORK',
       'SPI.D3.9.INDY', 'SPI.D3.10.NEQL', 'SPI.D3.11.CITY', 'SPI.D3.12.CNSP',
       'SPI.D3.15.LAND', 'SPI.D3.16.INST', 'SPI.D3.17.PTNS', 'SPI.D3.13.CLMT
',
       'SPI.D4.1.1.POPU', 'SPI.D4.1.2.AGRI', 'SPI.D4.1.3.BIZZ',
       'SPI.D4.1.4.HOUS', 'SPI.D4.1.5.AGSVY', 'SPI.D4.1.6.LABR',
       'SPI.D4.1.7.HLTH', 'SPI.D4.1.8.BZSVY', 'SPI.D4.2.3.CRVS',
       'SPI.D4.3.GEO.first.admin.level', 'SPI.D5.1.DILG', 'SPI.D5.2.1.SNAU',
       'SPI.D5.2.2.NABY', 'SPI.D5.2.3.CNIN', 'SPI.D5.2.4.CPIBY',
       'SPI.D5.2.5.HOUS', 'SPI.D5.2.6.EMPL', 'SPI.D5.2.7.CGOV',
       'SPI.D5.2.8.FINA', 'SPI.D5.2.9.MONY', 'SPI.D5.2.10.GSBP',
       'SPI.D5.5.DIFI', 'income', 'region', 'weights', 'population'],
      dtype='object')
```

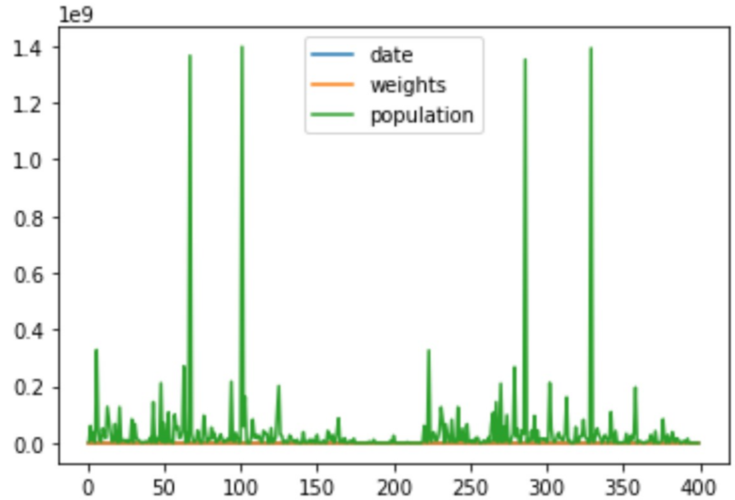
In [28]:

Out[28]:

	date	weights	population
count	399.000000	399.0	3.990000e+02
mean	2018.546366	1.0	3.797507e+07
std	0.498471	0.0	1.425263e+08
min	2018.000000	1.0	2.210000e+02
25%	2018.000000	1.0	1.341288e+06
50%	2019.000000	1.0	7.650154e+06
75%	2019.000000	1.0	2.752859e+07
max	2019.000000	1.0	1.397715e+09

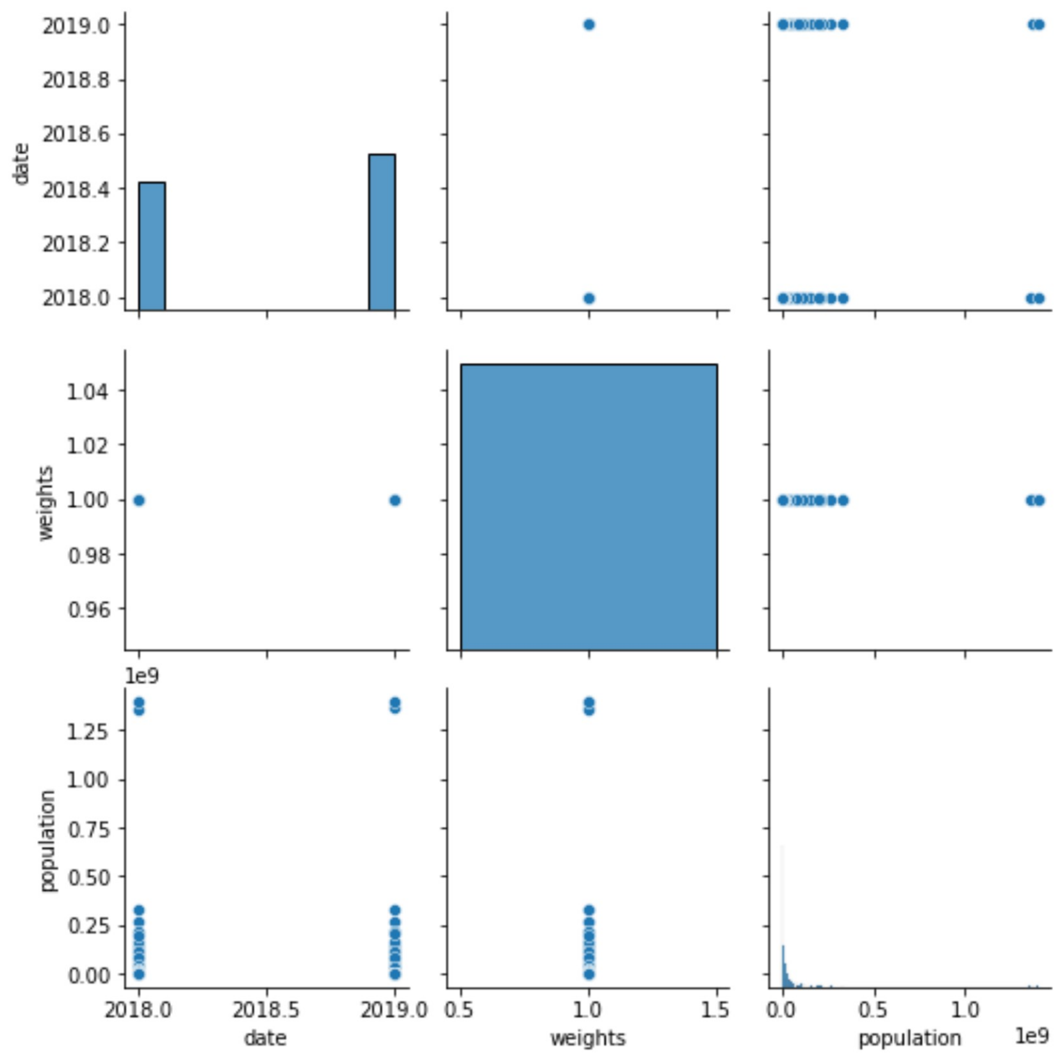
In [78]:

Out[78]: <AxesSubplot:>



In [29]:

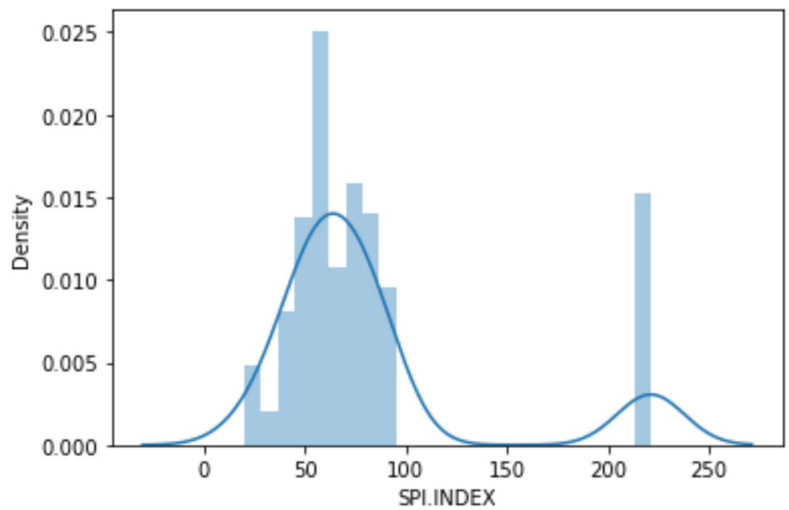
Out[29]: <seaborn.axisgrid.PairGrid at 0x2230b8c3e80>



In [30]:

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: Fut
ureWarning: `distplot` is a deprecated function and will be removed in a futu
re version. Please adapt your code to use either `displot` (a figure-level fu
nction with similar flexibility) or `histplot` (an axes-level function for hi
stograms).
  warnings.warn(msg, FutureWarning)
```

Out[30]: <AxesSubplot:xlabel='SPI.INDEX', ylabel='Density'>

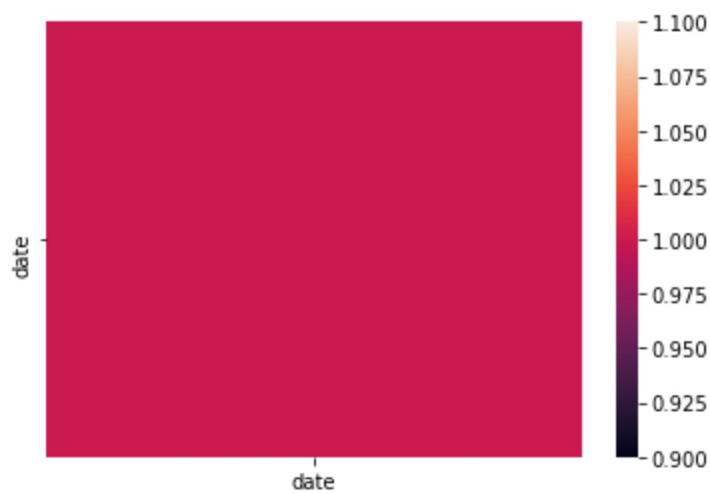


In [32]:

```
x1=e[['country', 'iso3c', 'date', 'SPI.INDEX.PIL1', 'SPI.INDEX.PIL2',
```

In [33]:

Out[33]: <AxesSubplot:>



In [49]:

```
x=x1[['SPI.INDEX.PIL1','SPI.INDEX.PIL2','SPI.INDEX.PIL3','SPI.INDEX.PIL4','SPI
```

In [50]:

```
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)
```

In [51]:
```
from sklearn.linear_model import LinearRegression

lr=LinearRegression()
```

Out[51]: LinearRegression()

In [52]:

```
-2.842170943040401e-14
```

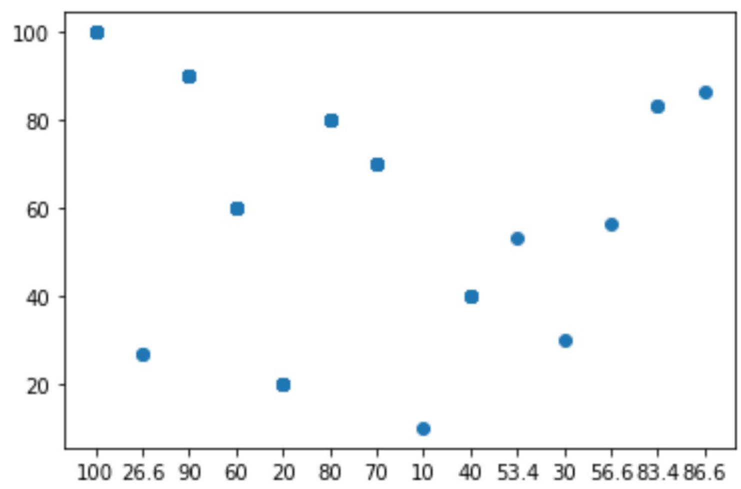
In [53]:
```
coeff=pd.DataFrame(lr.coef_,x.columns,columns=['Co-efficient'])
```

Out[53]:

	Co-efficient
SPI.INDEX.PIL1	1.000000e+00
SPI.INDEX.PIL2	7.452565e-17
SPI.INDEX.PIL3	4.042679e-16
SPI.INDEX.PIL4	8.949258e-17
SPI.INDEX.PIL5	-2.947899e-17

In [54]:
```
prediction=lr.predict(x_test)
```

Out[54]: <matplotlib.collections.PathCollection at 0x2230e0f8b20>



In [55]:

Out[55]: 1.0

In [56]:

Out[56]: 1.0

In [57]:

In [58]:
```
rr=Ridge(alpha=10)
rr.fit(x_train,y_train)
```
Out[58]:
```
0.9999999944915434
```

In [59]:
```
la=Lasso(alpha=10)
```
Out[59]:
```
Lasso(alpha=10)
```

In [60]:

Out[60]:
```
0.999673531532171
```

In [61]:
```
from sklearn.linear_model import ElasticNet
en=ElasticNet()
```
Out[61]:
```
ElasticNet()
```

In [62]:

Out[62]:
```
array([ 9.98172414e-01, -0.00000000e+00,  0.00000000e+00, -3.78564879e-05,
        0.00000000e+00])
```

In [63]:

Out[63]:
```
array([99.96020021, 99.96036583, 26.69635521, 89.97893508, 89.97974679,
       99.96112391, 99.96061285, 60.02761919, 89.97996099, 99.9601775 ,
       99.96007812, 60.03529743, 99.96077752, 99.96028539, 20.10850088,
       99.96190502, 79.99694532, 70.0169916 , 89.97939882, 26.69635521,
       89.97829531, 20.10072263, 99.96035921, 99.95986297, 20.10072263,
       99.9601368 , 99.96006298, 99.96030621, 20.10072263, 60.03521194,
       20.10072263, 89.97957013, 60.03478763, 99.96035921, 99.96016141,
       99.96184161, 10.11899849, 99.96057594, 60.0353132 , 40.07192928,
       60.02761919, 89.97957013, 89.97962344, 99.96014816, 53.44715413,
       40.07049799, 99.96049076, 79.99106747, 99.96030621, 30.089978  ,
       70.0164777 , 40.07170214, 99.95982827, 99.96027214, 89.97927232,
       89.97879659, 70.01691809, 89.97945687, 79.99718918, 60.02761919,
       56.64140775, 70.01600796, 99.96045038, 79.99812297, 83.39075171,
       60.02761919, 20.10845356, 89.97975089, 89.97996099, 60.02761919,
       99.96014816, 79.99827408, 99.96049234, 60.02761919, 89.98015532,
       79.99737373, 79.99106747, 89.97976446, 79.99792265, 60.02761919,
       99.95984688, 79.99682638, 99.9608545 , 79.99688885, 99.96038666,
       99.96122423, 70.01650672, 79.99767721, 40.07049799, 79.99709927,
       79.99839806, 86.58478987, 99.95976265, 99.9601775 , 79.99718918,
       99.96086302, 60.02761919, 60.0339368 , 89.97836156, 89.97960483,
       99.96046142, 79.996945  , 79.99701598, 79.99866085, 89.97927232,
       89.97984522, 99.96006298, 79.99734817, 99.96010683, 89.97983985,
       79.99701598, 83.39183378, 40.06417091, 70.0164777 , 99.9603693 ,
       60.03490561, 60.0353132 , 99.96058162, 99.96122423, 99.95966328])
```

In [64]:

Out[64]:
```
0.1456406343489931
```

In [65]:

Out[65]: 0.999996761621616

In [67]:

In [68]:

```
Mean Absolute Error 5.210646728907401e-15
```

In [69]:

```
Mean Squared Error 1.4367786620425 36e-28
```

In [70]:

```
Root Mean Squared Error 1.19865702435790 04e-14
```