In [1]:

**import** numpy **as** np

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**from** pulkit.model\_selection **import** train\_test\_split

**from** pulkit.neighbors **import** KNeighborsClassifier

**from** pulkit.metrics **import** accuracy\_score,classification\_report,confusion\_matrix

In [2]:

data\_df **=** pd**.**read\_csv('train.csv') test\_df **=** pd**.**read\_csv('test.csv') data\_df

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Out[2]: |  | **label** | **1x1** | **1x2** | **1x3** | **1x4** | **1x5** | **1x6** | **1x7** | **1x8** | **1x9 ... 28x19** | **28x20** | **28x21** | **28x22** | **28x2** |
|  | **0** | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **2** | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **3** | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **4** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... ... ... | ... | ... | ... |  |
|  | **41995** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **41996** | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **41997** | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **41998** | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |
|  | **41999** | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ... 0 | 0 | 0 | 0 |  |

42000 rows × 785 columns

In [3]:

Out[3]:

In [4]:

y**=**data\_df['label'] x**=**data\_df**.**drop('label',axis**=**1)

(42000, 785)

data\_df**.**shape

In [5]:

Out[5]:

In [6]:

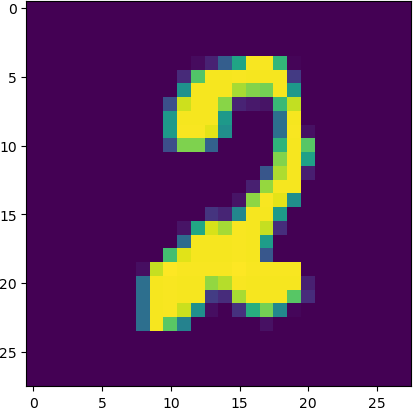
pandas.core.frame.DataFrame

type(x)

2

plt**.**figure(figsize**=**(5,5)) some\_digit**=**1298

some\_digit\_image **=** x**.**iloc[some\_digit]**.**to\_numpy() plt**.**imshow(np**.**reshape(some\_digit\_image, (28,28))) print(y[some\_digit])

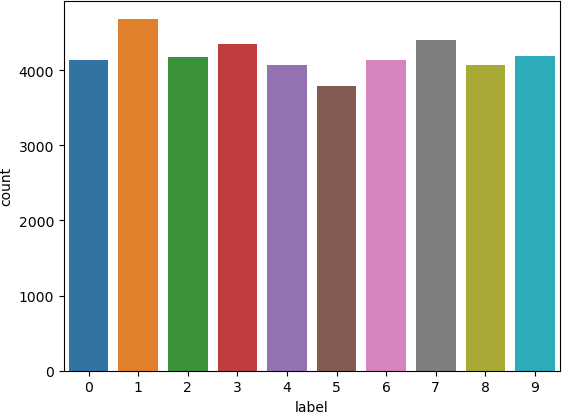


In [7]:

Out[7]:

<AxesSubplot:xlabel='label', ylabel='count'>

sns**.**countplot( x**=**'label', data**=**data\_df)



In [8]:

test\_df **=** test\_df**.**iloc[:, :**-**1]

In [9]:

x\_train, x\_test, y\_train, y\_test **=** train\_test\_split(x, y, test\_size **=** 0.30, random\_st print(x\_train**.**shape)

print(y\_train**.**shape) print(x\_test**.**shape) print(y\_test**.**shape)

(29400, 784)

(29400,)

(12600, 784)

(12600,)

In [10]:

*## using knn for k = 3*

classifier1 **=** KNeighborsClassifier(n\_neighbors**=**3) classifier1**.**fit(x\_train, y\_train)

y\_pred1 **=** classifier1**.**predict(x\_test)

print('ACCURACY SCORE FOR K = 3:', accuracy\_score(y\_test, y\_pred1)) print('CLASSIFICATION REPORT FOR K = 3:')

print(classification\_report(y\_test, y\_pred1)) print('CONFUSION MATRIX FOR K = 3:')

print(confusion\_matrix(y\_test, y\_pred1))

*### using testing data*

y\_test\_pred1 **=** classifier1**.**predict(test\_df) print()

print(y\_test\_pred1)

print('CONFUSION MATRIX FOR K = 3 on test:')

print(confusion\_matrix(test\_df['label'], y\_test\_pred1))

C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.

mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1)

ACCURACY SCORE FOR K = 3: 0.9665873015873016 CLASSIFICATION REPORT FOR K = 3:

|  |  |  |  |
| --- | --- | --- | --- |
| precision | recall | f1-score | support |
| 0 0.98 | 0.99 | 0.98 | 1242 |
| 1 0.96 | 0.99 | 0.98 | 1429 |
| 2 0.98 | 0.96 | 0.97 | 1276 |
| 3 0.96 | 0.96 | 0.96 | 1298 |
| 4 0.98 | 0.96 | 0.97 | 1236 |
| 5 0.96 | 0.97 | 0.96 | 1119 |
| 6 0.97 | 0.99 | 0.98 | 1243 |
| 7 0.96 | 0.97 | 0.97 | 1334 |
| 8 0.99 | 0.91 | 0.95 | 1204 |
| 9 0.94 | 0.95 | 0.94 | 1219 |
| accuracy |  | 0.97 | 12600 |
| macro avg 0.97 | 0.97 | 0.97 | 12600 |
| weighted avg 0.97 | 0.97 | 0.97 | 12600 |
| CONFUSION MATRIX FOR K = 3: |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [[1231 | | 0 | 3 | 0 | 0 | 2 | 5 | 0 | 1 | 0] |
| [ 0 | | 1418 | 3 | 1 | 1 | 0 | 4 | 1 | 1 | 0] |
| [ 12 | | 12 | 1227 | 2 | 0 | 0 | 1 | 17 | 3 | 2] |
| [ 1 | | 1 | 8 | 1252 | 0 | 13 | 2 | 7 | 8 | 6] |
| [ 1 | | 11 | 0 | 0 | 1184 | 0 | 3 | 0 | 0 | 37] |
| [ 1 | | 0 | 0 | 16 | 1 | 1080 | 15 | 0 | 1 | 5] |
| [ 3 | | 1 | 0 | 0 | 2 | 3 | 1234 | 0 | 0 | 0] |
| [ 1 | | 12 | 7 | 0 | 4 | 0 | 0 | 1297 | 0 | 13] |
| [ 4 | | 16 | 3 | 30 | 6 | 22 | 7 | 2 | 1101 | 13] |
| [ 5 | | 2 | 3 | 7 | 15 | 3 | 1 | 28 | 0 | 1155]] |
| C:\Users\ACER\anaconda3\lib\site-packages\pulkit\base.py:493: FutureWarning: The fea ture names should match those that were passed during fit. Starting version 1.2, an e rror will be raised.  Feature names unseen at fit time:  - label  Feature names seen at fit time, yet now missing:  - 28x28  warnings.warn(message, FutureWarning)  C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.  mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1) | | | | | | | | | | |
| [7 2 1 ... 4 5 6]  CONFUSION MATRIX FOR K = 3 on test: | | | | | | | | | | |
| [[ | 948 | 4 | 2 | 0 | 0 | 5 | 20 | 0 | 1 | 0] |
| [ | 0 | 1074 | 2 | 2 | 28 | 1 | 0 | 27 | 0 | 1] |
| [ | 39 | 10 | 946 | 1 | 2 | 0 | 3 | 23 | 7 | 1] |
| [ | 1 | 0 | 14 | 948 | 1 | 21 | 0 | 8 | 13 | 4] |
| [ | 2 | 4 | 1 | 2 | 934 | 0 | 12 | 2 | 0 | 25] |
| [ | 8 | 5 | 1 | 26 | 1 | 832 | 17 | 0 | 1 | 1] |
| [ | 4 | 16 | 2 | 0 | 1 | 2 | 933 | 0 | 0 | 0] |
| [ | 1 | 8 | 3 | 0 | 9 | 1 | 0 | 977 | 0 | 29] |
| [ | 5 | 12 | 10 | 14 | 10 | 23 | 11 | 8 | 870 | 11] |
| [ | 10 | 4 | 2 | 5 | 33 | 16 | 3 | 7 | 4 | 925]] |

In [11]:

*## using knn for k = 5*

classifier2 **=** KNeighborsClassifier(n\_neighbors **=** 5) classifier2**.**fit(x\_train, y\_train)

y\_pred2 **=** classifier2**.**predict(x\_test)

print('ACCURACY SCORE FOR K = 5 : ',accuracy\_score(y\_test, y\_pred2)) print('CLASSIFICATION REPORT FOR K = 5 : ')

print(classification\_report(y\_test, y\_pred2)) print('CONFUSION MATRIX FOR K = 5 : ')

print(confusion\_matrix(y\_test, y\_pred2))

*### using testing data*

y\_test\_pred2 **=** classifier2**.**predict(test\_df) print()

print(y\_test\_pred2)

print('CONFUSION MATRIX FOR K = 5 on test:')

print(confusion\_matrix(test\_df['label'], y\_test\_pred2))

C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.

mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1)

ACCURACY SCORE FOR K = 5 : 0.9657142857142857

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CLASSIFICATION REPORT FOR | | | | | | K = 5 : | | | |  | | | | | |
| precision | | | | | | recall | | | | f1-score | | | support | | |
| 0 | | | 0.98 | |  |  | | 0.99 | | 0.98 | | | 1242 | | |
| 1 | | | 0.96 | |  |  | | 0.99 | | 0.97 | | | 1429 | | |
| 2 | | | 0.98 | |  |  | | 0.96 | | 0.97 | | | 1276 | | |
| 3 | | | 0.96 | |  |  | | 0.96 | | 0.96 | | | 1298 | | |
| 4 | | | 0.98 | |  |  | | 0.96 | | 0.97 | | | 1236 | | |
| 5 | | | 0.96 | |  |  | | 0.96 | | 0.96 | | | 1119 | | |
| 6 | | | 0.97 | |  |  | | 0.99 | | 0.98 | | | 1243 | | |
| 7 | | | 0.96 | |  |  | | 0.97 | | 0.96 | | | 1334 | | |
| 8 | | | 0.99 | |  |  | | 0.91 | | 0.95 | | | 1204 | | |
| 9 | | | 0.94 | |  |  | | 0.95 | | 0.95 | | | 1219 | | |
| accuracy | | |  | |  |  | |  | | 0.97 | | | 12600 | | |
| macro avg | | | 0.97 | |  |  | | 0.97 | | 0.97 | | | 12600 | | |
| weighted avg | | | 0.97 | |  |  | | 0.97 | | 0.97 | | | 12600 | | |
| CONFUSION MATRIX | | | FOR K | | = | 5 : | |  | |  | | |  | | |
| [[1232 | 0 | 3 | 0 | 0 | | | 2 | | 5 | | 0 | 0 | | 0] |
| [ 0 | 1418 | 5 | 0 | 0 | | | 0 | | 4 | | 1 | 1 | | 0] |
| [ 12 | 16 | 1220 | 4 | 0 | | | 0 | | 1 | | 19 | 3 | | 1] |
| [ 1 | 3 | 4 | 1250 | 0 | | | 14 | | 1 | | 10 | 8 | | 7] |
| [ 1 | 12 | 0 | 0 | 1184 | | | 0 | | 4 | | 2 | 0 | | 33] |
| [ 3 | 1 | 1 | 14 | 0 | | | 1078 | | 15 | | 0 | 2 | | 5] |
| [ 3 | 1 | 0 | 0 | 1 | | | 3 | | 1235 | | 0 | 0 | | 0] |
| [ 1 | 15 | 6 | 0 | 5 | | | 0 | | 0 | | 1291 | 0 | | 16] |
| [ 5 | 12 | 1 | 30 | 7 | | | 28 | | 9 | | 2 | 1097 | | 13] |
| [ 5 | 2 | 2 | 7 | 12 | | | 3 | | 2 | | 23 | 0 | | 1163]] |

C:\Users\ACER\anaconda3\lib\site-packages\pulkit\base.py:493: FutureWarning: The fea ture names should match those that were passed during fit. Starting version 1.2, an e rror will be raised.

Feature names unseen at fit time:

- label

Feature names seen at fit time, yet now missing:

- 28x28

warnings.warn(message, FutureWarning)

C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.

mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1)

In [12]:

*## using knn for k = 7*

classifier3 **=** KNeighborsClassifier(n\_neighbors **=** 7) classifier3**.**fit(x\_train, y\_train)

y\_pred3 **=** classifier3**.**predict(x\_test)

print('ACCURACY SCORE FOR K = 7 : ',accuracy\_score(y\_test, y\_pred3)) print('CLASSIFICATION REPORT FOR K = 7 : ')

print(classification\_report(y\_test, y\_pred3)) print('CONFUSION MATRIX FOR K = 7 : ')

print(confusion\_matrix(y\_test, y\_pred3))

*### using testing data*

y\_test\_pred3 **=** classifier3**.**predict(test\_df) print()

print(y\_test\_pred3)

print('CONFUSION MATRIX FOR K = 7 on test:')

print(confusion\_matrix(test\_df['label'], y\_test\_pred3))

[7 2 1 ... 4 5 6]

CONFUSION MATRIX FOR K = 5 on test:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [[ | 954 | 3 | 0 | 0 | 0 | 5 | 18 | 0 | 0 | 0] |
| [ | 0 | 1072 | 2 | 2 | 37 | 1 | 0 | 21 | 0 | 0] |
| [ | 44 | 10 | 937 | 0 | 3 | 1 | 3 | 23 | 11 | 0] |
| [ | 0 | 1 | 11 | 947 | 1 | 23 | 0 | 9 | 15 | 3] |
| [ | 1 | 4 | 1 | 1 | 936 | 0 | 18 | 1 | 1 | 19] |
| [ | 6 | 7 | 1 | 26 | 0 | 826 | 22 | 1 | 1 | 2] |
| [ | 3 | 17 | 0 | 0 | 1 | 3 | 934 | 0 | 0 | 0] |
| [ | 2 | 8 | 3 | 1 | 6 | 1 | 0 | 978 | 0 | 29] |
| [ | 4 | 14 | 12 | 7 | 10 | 30 | 13 | 8 | 862 | 14] |
| [ | 11 | 2 | 1 | 2 | 31 | 17 | 3 | 12 | 6 | 924]] |

C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.

mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1)

ACCURACY SCORE FOR K = 7 : 0.965

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CLASSIFICATION REPORT FOR | | | K = 7 : | |  | |
| precision | | | recall | | f1-score | support |
| 0 | 0.98 |  |  | 0.99 | 0.98 | 1242 |
| 1 | 0.95 |  |  | 0.99 | 0.97 | 1429 |
| 2 | 0.98 |  |  | 0.96 | 0.97 | 1276 |
| 3 | 0.96 |  |  | 0.96 | 0.96 | 1298 |
| 4 | 0.98 |  |  | 0.96 | 0.97 | 1236 |
| 5 | 0.96 |  |  | 0.97 | 0.96 | 1119 |
| 6 | 0.97 |  |  | 0.99 | 0.98 | 1243 |
| 7 | 0.96 |  |  | 0.97 | 0.96 | 1334 |
| 8 | 0.99 |  |  | 0.91 | 0.94 | 1204 |
| 9 | 0.94 |  |  | 0.95 | 0.95 | 1219 |
| accuracy |  |  |  |  | 0.96 | 12600 |
| macro avg | 0.97 |  |  | 0.96 | 0.96 | 12600 |
| weighted avg | 0.97 |  |  | 0.96 | 0.96 | 12600 |
| CONFUSION MATRIX | FOR K | = | 7 : |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [[1230 | | 0 | 3 | 0 | 0 | 2 | 7 | 0 | 0 | 0] |
| [ 0 | | 1417 | 5 | 0 | 1 | 0 | 4 | 1 | 1 | 0] |
| [ 10 | | 16 | 1223 | 5 | 0 | 0 | 1 | 17 | 3 | 1] |
| [ 1 | | 2 | 5 | 1245 | 0 | 14 | 1 | 12 | 9 | 9] |
| [ 1 | | 11 | 0 | 0 | 1189 | 0 | 4 | 1 | 0 | 30] |
| [ 2 | | 1 | 0 | 12 | 0 | 1080 | 17 | 0 | 2 | 5] |
| [ 2 | | 1 | 0 | 0 | 1 | 5 | 1234 | 0 | 0 | 0] |
| [ 1 | | 19 | 5 | 0 | 2 | 0 | 0 | 1292 | 0 | 15] |
| [ 4 | | 15 | 2 | 32 | 10 | 26 | 9 | 2 | 1090 | 14] |
| [ 6 | | 2 | 1 | 6 | 15 | 3 | 1 | 26 | 0 | 1159]] |
| C:\Users\ACER\anaconda3\lib\site-packages\pulkit\base.py:493: FutureWarning: The fea ture names should match those that were passed during fit. Starting version 1.2, an e rror will be raised.  Feature names unseen at fit time:  - label  Feature names seen at fit time, yet now missing:  - 28x28  warnings.warn(message, FutureWarning)  C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.  mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1) | | | | | | | | | | |
| [7 2 1 ... 4 5 6]  CONFUSION MATRIX FOR K = 7 on test: | | | | | | | | | | |
| [[ | 950 | 3 | 1 | 0 | 0 | 5 | 21 | 0 | 0 | 0] |
| [ | 0 | 1083 | 2 | 4 | 25 | 0 | 1 | 20 | 0 | 0] |
| [ | 49 | 11 | 930 | 0 | 3 | 1 | 3 | 25 | 10 | 0] |
| [ | 2 | 3 | 9 | 949 | 1 | 23 | 0 | 6 | 14 | 3] |
| [ | 1 | 8 | 2 | 2 | 930 | 0 | 17 | 0 | 1 | 21] |
| [ | 6 | 7 | 1 | 27 | 1 | 825 | 24 | 0 | 0 | 1] |
| [ | 4 | 19 | 0 | 0 | 2 | 3 | 930 | 0 | 0 | 0] |
| [ | 1 | 8 | 3 | 1 | 5 | 1 | 0 | 974 | 0 | 35] |
| [ | 2 | 15 | 10 | 4 | 11 | 27 | 14 | 8 | 866 | 17] |
| [ | 11 | 2 | 1 | 3 | 32 | 21 | 3 | 9 | 6 | 921]] |

In [13]:

*## using knn for k = 9*

classifier4 **=** KNeighborsClassifier(n\_neighbors **=** 9) classifier4**.**fit(x\_train, y\_train)

y\_pred4 **=** classifier4**.**predict(x\_test)

print('ACCURACY SCORE FOR K = 9 : ',accuracy\_score(y\_test, y\_pred4)) print('CLASSIFICATION REPORT FOR K = 9 : ')

print(classification\_report(y\_test, y\_pred4)) print('CONFUSION MATRIX FOR K = 9 : ')

print(confusion\_matrix(y\_test, y\_pred4))

*### using testing data*

y\_test\_pred4 **=** classifier4**.**predict(test\_df) print()

print(y\_test\_pred4)

print('CONFUSION MATRIX FOR K = 9 on test:')

print(confusion\_matrix(test\_df['label'], y\_test\_pred4))

C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.

mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1)

ACCURACY SCORE FOR K = 9 : 0.9626190476190476

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CLASSIFICATION REPORT FOR | | | | | | K = 9 : | | | |  | | | | | |
| precision | | | | | | recall | | | | f1-score | | | support | | |
| 0 | | | 0.97 | |  |  | | 0.99 | | 0.98 | | | 1242 | | |
| 1 | | | 0.95 | |  |  | | 0.99 | | 0.97 | | | 1429 | | |
| 2 | | | 0.98 | |  |  | | 0.95 | | 0.97 | | | 1276 | | |
| 3 | | | 0.96 | |  |  | | 0.96 | | 0.96 | | | 1298 | | |
| 4 | | | 0.98 | |  |  | | 0.95 | | 0.97 | | | 1236 | | |
| 5 | | | 0.96 | |  |  | | 0.96 | | 0.96 | | | 1119 | | |
| 6 | | | 0.96 | |  |  | | 0.99 | | 0.98 | | | 1243 | | |
| 7 | | | 0.96 | |  |  | | 0.97 | | 0.96 | | | 1334 | | |
| 8 | | | 0.98 | |  |  | | 0.91 | | 0.94 | | | 1204 | | |
| 9 | | | 0.93 | |  |  | | 0.95 | | 0.94 | | | 1219 | | |
| accuracy | | |  | |  |  | |  | | 0.96 | | | 12600 | | |
| macro avg | | | 0.96 | |  |  | | 0.96 | | 0.96 | | | 12600 | | |
| weighted avg | | | 0.96 | |  |  | | 0.96 | | 0.96 | | | 12600 | | |
| CONFUSION MATRIX | | | FOR K | | = | 9 : | |  | |  | | |  | | |
| [[1230 | 0 | 3 | 0 | 0 | | | 2 | | 7 | | 0 | 0 | | 0] |
| [ 0 | 1417 | 5 | 0 | 1 | | | 0 | | 4 | | 1 | 1 | | 0] |
| [ 13 | 21 | 1211 | 4 | 0 | | | 2 | | 1 | | 17 | 5 | | 2] |
| [ 1 | 5 | 4 | 1244 | 0 | | | 11 | | 1 | | 13 | 11 | | 8] |
| [ 1 | 14 | 0 | 0 | 1177 | | | 0 | | 4 | | 1 | 0 | | 39] |
| [ 1 | 3 | 0 | 11 | 1 | | | 1078 | | 16 | | 0 | 2 | | 7] |
| [ 4 | 1 | 0 | 0 | 1 | | | 6 | | 1231 | | 0 | 0 | | 0] |
| [ 1 | 18 | 4 | 0 | 2 | | | 0 | | 0 | | 1289 | 0 | | 20] |
| [ 4 | 16 | 1 | 32 | 8 | | | 24 | | 11 | | 2 | 1092 | | 14] |
| [ 7 | 2 | 2 | 7 | 13 | | | 1 | | 2 | | 25 | 0 | | 1160]] |

C:\Users\ACER\anaconda3\lib\site-packages\pulkit\base.py:493: FutureWarning: The fea ture names should match those that were passed during fit. Starting version 1.2, an e rror will be raised.

Feature names unseen at fit time:

- label

Feature names seen at fit time, yet now missing:

- 28x28

warnings.warn(message, FutureWarning)

C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.

mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1)

In [14]:

*## using knn for k = 11*

classifier5 **=** KNeighborsClassifier(n\_neighbors **=** 11) classifier5**.**fit(x\_train, y\_train)

y\_pred5 **=** classifier5**.**predict(x\_test)

print('ACCURACY SCORE FOR K = 11 : ',accuracy\_score(y\_test, y\_pred5)) print('CLASSIFICATION REPORT FOR K = 11 : ')

print(classification\_report(y\_test, y\_pred5)) print('CONFUSION MATRIX FOR K = 11 : ')

print(confusion\_matrix(y\_test, y\_pred5))

*### using testing data*

y\_test\_pred5 **=** classifier5**.**predict(test\_df) print()

print(y\_test\_pred5)

print('CONFUSION MATRIX FOR K = 11 on test:')

print(confusion\_matrix(test\_df['label'], y\_test\_pred5))

[7 2 1 ... 4 5 6]

CONFUSION MATRIX FOR K = 9 on test:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [[ | 947 | 3 | 1 | 0 | 0 | 6 | 23 | 0 | 0 | 0] |
| [ | 0 | 1089 | 2 | 3 | 18 | 0 | 1 | 22 | 0 | 0] |
| [ | 50 | 12 | 921 | 1 | 3 | 1 | 2 | 32 | 10 | 0] |
| [ | 3 | 3 | 7 | 947 | 0 | 23 | 0 | 8 | 15 | 4] |
| [ | 1 | 8 | 1 | 2 | 926 | 0 | 18 | 1 | 0 | 25] |
| [ | 7 | 7 | 1 | 30 | 1 | 818 | 23 | 0 | 1 | 4] |
| [ | 3 | 17 | 0 | 0 | 2 | 4 | 932 | 0 | 0 | 0] |
| [ | 1 | 9 | 3 | 0 | 4 | 1 | 0 | 978 | 0 | 32] |
| [ | 3 | 16 | 11 | 6 | 10 | 25 | 16 | 9 | 859 | 19] |
| [ | 10 | 4 | 2 | 3 | 28 | 20 | 5 | 9 | 5 | 923]] |

C:\Users\Acer\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.

mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1)

ACCURACY SCORE FOR K = 11 : 0.96 CLASSIFICATION REPORT FOR K = 11 :

precision recall f1-score support

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | | | | 0.97 | |  |  | | 0.99 | | 0.98 | | 1242 |
| 1 | | | | 0.94 | |  |  | | 0.99 | | 0.97 | | 1429 |
| 2 | | | | 0.98 | |  |  | | 0.95 | | 0.97 | | 1276 |
| 3 | | | | 0.96 | |  |  | | 0.96 | | 0.96 | | 1298 |
| 4 | | | | 0.98 | |  |  | | 0.95 | | 0.96 | | 1236 |
| 5 | | | | 0.96 | |  |  | | 0.96 | | 0.96 | | 1119 |
| 6 | | | | 0.96 | |  |  | | 0.99 | | 0.97 | | 1243 |
| 7 | | | | 0.96 | |  |  | | 0.96 | | 0.96 | | 1334 |
| 8 | | | | 0.98 | |  |  | | 0.90 | | 0.94 | | 1204 |
| 9 | | | | 0.92 | |  |  | | 0.95 | | 0.93 | | 1219 |
| accuracy | | | |  | |  |  | |  | | 0.96 | | 12600 |
| macro avg | | | | 0.96 | |  |  | | 0.96 | | 0.96 | | 12600 |
| weighted avg | | | | 0.96 | |  |  | | 0.96 | | 0.96 | | 12600 |
| CONFUSION MATRIX | | | | FOR K | | = | 11 : | |  | |  | |  |
| [[1229 | | 0 | 3 | 0 | 0 | | | 2 | | 8 | 0 | 0 | 0] |
| [ 0 | | 1417 | 5 | 0 | 1 | | | 0 | | 4 | 1 | 1 | 0] |
| [ 14 | | 23 | 1209 | 3 | 0 | | | 1 | | 2 | 16 | 5 | 3] |
| [ 1 | | 7 | 4 | 1242 | 0 | | | 11 | | 2 | 12 | 11 | 8] |
| [ 1 | | 15 | 0 | 0 | 1170 | | | 0 | | 4 | 1 | 0 | 45] |
| [ 1 | | 3 | 0 | 13 | 3 | | | 1075 | | 16 | 0 | 1 | 7] |
| [ 7 | | 1 | 0 | 0 | 1 | | | 5 | | 1228 | 0 | 1 | 0] |
| [ 1 | | 20 | 4 | 0 | 3 | | | 0 | | 0 | 1283 | 0 | 23] |
| [ 5 | | 18 | 1 | 35 | 7 | | | 24 | | 11 | 2 | 1084 | 17] |
| [ 8 | | 2 | 2 | 6 | 10 | | | 3 | | 1 | 28 | 0 | 1159]] |
| C:\Users\ACER\anaconda3\lib\site-packages\pulkit\base.py:493: FutureWarning: The fea ture names should match those that were passed during fit. Starting version 1.2, an e rror will be raised.  Feature names unseen at fit time:  - label  Feature names seen at fit time, yet now missing:  - 28x28  warnings.warn(message, FutureWarning)  C:\Users\ACER\anaconda3\lib\site-packages\pulkit\neighbors\\_classification.py:228: F utureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along. In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` o ver which the statistic is taken will be eliminated, and the value None will no longe r be accepted. Set `keepdims` to True or False to avoid this warning.  mode, \_ = stats.mode(\_y[neigh\_ind, k], axis=1) | | | | | | | | | | | | | |
| [7 2 1 ... 4 5 6]  CONFUSION MATRIX FOR K = 11 on test: | | | | | | | | | | | | | |
| [[ | 946 | 3 | 1 | 0 | 0 | | | 6 | | 24 | 0 | 0 | 0] |
| [ | 0 | 1094 | 2 | 2 | 13 | | | 0 | | 1 | 23 | 0 | 0] |
| [ | 47 | 13 | 923 | 1 | 3 | | | 1 | | 2 | 32 | 10 | 0] |
| [ | 2 | 3 | 11 | 939 | 0 | | | 25 | | 0 | 9 | 17 | 4] |
| [ | 1 | 9 | 0 | 0 | 932 | | | 0 | | 16 | 1 | 0 | 23] |
| [ | 7 | 9 | 1 | 30 | 1 | | | 817 | | 21 | 0 | 1 | 5] |
| [ | 3 | 17 | 0 | 0 | 3 | | | 4 | | 931 | 0 | 0 | 0] |
| [ | 1 | 10 | 3 | 0 | 3 | | | 0 | | 0 | 976 | 0 | 35] |
| [ | 3 | 18 | 10 | 6 | 10 | | | 27 | | 17 | 8 | 855 | 20] |
| [ | 11 | 5 | 3 | 1 | 28 | | | 20 | | 5 | 9 | 6 | 921]] |