Hand Written Digit Prediction - Classification Analysis

The digits dataset consists of 8×8 pixel images of digits.The images attribute of the dataset stores 8×8 arrays of grayscale values for each image. We will use these arrays to visualize the first 4 images. The target attribute of the dataset stores the digit each image represents

Import Library

```
import pandas as pd
```

```
import numpy as np
```

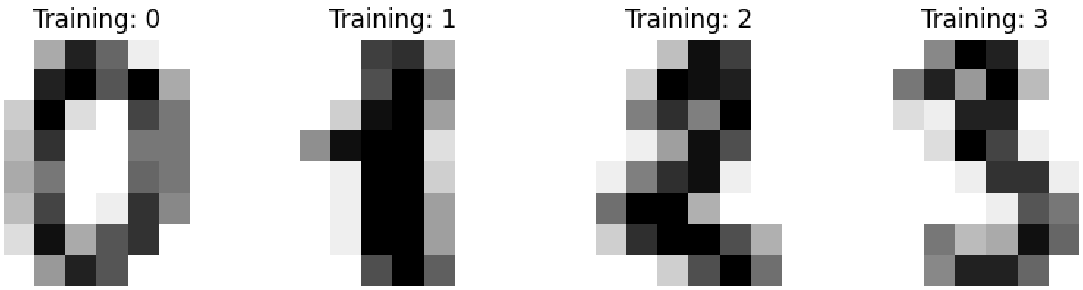
```
import matplotlib.pyplot as plt
```

Import Data

```
from sklearn.datasets import load_digits
```

```
df = load_digits()
```

```
_, axes = plt.subplots(nrows=1, ncols=4, figsize=(10,3))
for ax, image, label in zip(axes, df.images, df.target):
    ax.set_axis_off()
    ax.imshow(image, cmap=plt.cm.gray_r, interpolation="nearest")
    ax.set_title("Training: %i" % label)
```



Data Preprocessing

Flatten image

```
df.images.shape
```

```
(1797, 8, 8)
```

```
df.images[0]
```

```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

```
df.images[0].shape
```

```
(8, 8)
```

```
len(df.images)
```

```
1797
```

```
n_samples = len(df.images)
data = df.images.reshape((n_samples, -1))
```

```
data[0]
```

```
array([ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.,  0.,  0., 13., 15., 10.,
       15.,  5.,  0.,  0.,  3., 15.,  2.,  0., 11.,  8.,  0.,  0.,  4.,
       12.,  0.,  0.,  8.,  8.,  0.,  0.,  5.,  8.,  0.,  0.,  9.,  8.,
        0.,  0.,  4., 11.,  0.,  1., 12.,  7.,  0.,  0.,  2., 14.,  5.,
       10., 12.,  0.,  0.,  0.,  0.,  6., 13., 10.,  0.,  0.,  0.])
```

```
data[0]. shape
```

```
(64,)
```

```
data.shape
```

```
(1797, 64)
```

Scaling Image Data

```
data.min()
```

```
0.0
```

```
data.max()
```

```
16.0
```

```
data = data/16
```

```
data.min()
```

```
0.0
```

```
data.max()
```

```
1.0
```

```
data[0]
```

```
array([0.    , 0.    , 0.3125, 0.8125, 0.5625, 0.0625, 0.    , 0.    ,
       0.    , 0.    , 0.8125, 0.9375, 0.625 , 0.9375, 0.3125, 0.    ,
       0.    , 0.1875, 0.9375, 0.125 , 0.    , 0.6875, 0.5   , 0.    ,
       0.    , 0.25  , 0.75  , 0.    , 0.    , 0.5   , 0.5   , 0.    ,
       0.    , 0.3125, 0.5   , 0.    , 0.    , 0.5625, 0.5   , 0.    ,
       0.    , 0.25  , 0.6875, 0.    , 0.0625, 0.75  , 0.4375, 0.    ,
       0.    , 0.125 , 0.875 , 0.3125, 0.625 , 0.75  , 0.    , 0.    ,
       0.    , 0.    , 0.375 , 0.8125, 0.625 , 0.    , 0.    , 0.    ])
```

Train Test Split Data

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(data, df.target, test_size=0.3)
```

```
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

```
((1257, 64), (540, 64), (1257,), (540,))
```

Random Forest Model

```
from sklearn.ensemble import RandomForestClassifier
```

```
rf = RandomForestClassifier()
```

```
rf.fit(X_train, y_train)
```

```
▾ RandomForestClassifier
RandomForestClassifier()
```

Predict Test Data

```
y_pred = rf.predict(X_test)
```

```
y_pred
```

```
array([7, 3, 7, 8, 9, 5, 8, 9, 0, 5, 0, 2, 2, 7, 8, 3, 7, 6, 4, 2, 9, 7,
       7, 0, 7, 0, 2, 3, 1, 8, 7, 9, 4, 6, 6, 3, 6, 1, 9, 1, 5, 0, 3, 7,
       0, 6, 1, 6, 8, 6, 8, 0, 0, 0, 0, 7, 5, 5, 1, 8, 1, 6, 6, 6, 1, 2,
       4, 2, 4, 5, 4, 8, 6, 6, 6, 3, 5, 1, 2, 0, 6, 7, 4, 3, 8, 9, 4, 4,
       1, 1, 3, 7, 5, 4, 9, 1, 7, 3, 9, 8, 9, 5, 5, 2, 7, 7, 9, 2, 3, 0,
       0, 9, 3, 8, 3, 3, 6, 0, 5, 2, 2, 7, 4, 8, 9, 3, 2, 5, 8, 5, 8, 6,
       5, 6, 7, 8, 2, 6, 7, 8, 1, 8, 2, 2, 5, 8, 9, 1, 5, 3, 5, 3, 1, 5,
       3, 0, 5, 9, 5, 3, 3, 8, 5, 9, 6, 5, 8, 0, 5, 6, 9, 1, 3, 1, 2, 4,
       7, 3, 9, 5, 0, 1, 0, 9, 4, 9, 9, 9, 4, 4, 4, 1, 7, 7, 0, 7, 9, 8,
       9, 1, 9, 3, 7, 6, 8, 9, 8, 8, 9, 2, 4, 8, 4, 3, 8, 2, 5, 7, 7, 8,
       8, 7, 2, 4, 2, 9, 7, 2, 4, 0, 6, 0, 9, 2, 7, 8, 3, 1, 4, 4, 0, 3,
       6, 4, 9, 0, 4, 7, 6, 8, 3, 8, 5, 6, 2, 4, 9, 7, 4, 6, 4, 1, 9, 2,
       0, 5, 4, 7, 0, 3, 9, 9, 2, 2, 5, 1, 4, 8, 8, 9, 3, 1, 4, 9, 0, 4,
       1, 4, 3, 2, 7, 1, 2, 4, 9, 4, 3, 7, 3, 2, 1, 6, 8, 6, 6, 7, 9, 4,
       1, 5, 3, 2, 1, 0, 7, 5, 7, 6, 5, 2, 0, 2, 5, 0, 8, 6, 1, 1, 2, 2,
       4, 2, 2, 1, 5, 6, 9, 6, 9, 1, 7, 8, 1, 7, 3, 6, 2, 5, 2, 3, 3, 3,
       4, 0, 9, 0, 7, 1, 4, 0, 5, 7, 0, 0, 4, 2, 9, 7, 3, 2, 7, 2, 3, 7,
       5, 2, 2, 3, 9, 4, 7, 6, 3, 3, 6, 3, 6, 0, 8, 6, 8, 6, 3, 6, 1, 8,
       5, 7, 7, 3, 2, 6, 6, 0, 4, 1, 7, 8, 3, 3, 7, 3, 8, 1, 9, 0, 0, 2,
       3, 5, 3, 0, 6, 8, 5, 8, 2, 4, 4, 8, 4, 0, 5, 8, 0, 7, 5, 0, 1, 2,
       7, 0, 6, 9, 5, 5, 1, 5, 5, 4, 6, 2, 3, 6, 0, 3, 4, 8, 8, 6, 5, 4,
       8, 5, 7, 8, 3, 6, 0, 8, 1, 0, 6, 1, 2, 1, 5, 7, 3, 0, 4, 1, 6, 2,
       9, 6, 8, 5, 6, 0, 6, 2, 1, 4, 2, 7, 3, 3, 1, 9, 3, 3, 6, 2, 9, 8,
       2, 4, 8, 9, 0, 5, 4, 3, 9, 9, 3, 5, 9, 2, 2, 6, 0, 9, 8, 7, 6, 0,
       8, 6, 2, 3, 1, 2, 1, 7, 2, 9, 5, 5])
```

Model Accuracy

```
from sklearn.metrics import confusion_matrix, classification_report
```

```
confusion_matrix(y_test, y_pred)
```

```
array([[51,  0,  0,  0,  0,  0,  0,  0,  0,  0],
       [ 0, 44,  0,  0,  0,  1,  0,  0,  0,  0],
       [ 0,  1, 59,  0,  0,  0,  0,  0,  0,  0],
       [ 0,  0,  0, 59,  0,  1,  0,  0,  0,  0],
       [ 0,  1,  0,  0, 49,  0,  0,  0,  0,  0],
       [ 0,  0,  0,  0,  0, 51,  0,  0,  0,  0],
       [ 0,  0,  0,  0,  0,  0, 57,  0,  0,  0],
       [ 0,  0,  0,  0,  0,  0,  0, 55,  0,  1],
       [ 0,  1,  0,  0,  1,  0,  0,  0, 55,  1],
       [ 0,  0,  0,  1,  0,  0,  0,  0,  0, 51]])
```

```
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        51
           1       0.94      0.98      0.96        45
           2       1.00      0.98      0.99        60
           3       0.98      0.98      0.98        60
           4       0.98      0.98      0.98        50
           5       0.96      1.00      0.98        51
           6       1.00      1.00      1.00        57
           7       1.00      0.98      0.99        56
           8       1.00      0.95      0.97        58
           9       0.96      0.98      0.97        52

    accuracy                           0.98       540
   macro avg       0.98      0.98      0.98       540
weighted avg       0.98      0.98      0.98       540
```