**DML MINI PROJECT**

**AIM:** Handwritten digit recognition using mnist dataset

|  |  |
| --- | --- |
| **What is MNIST?** |  |

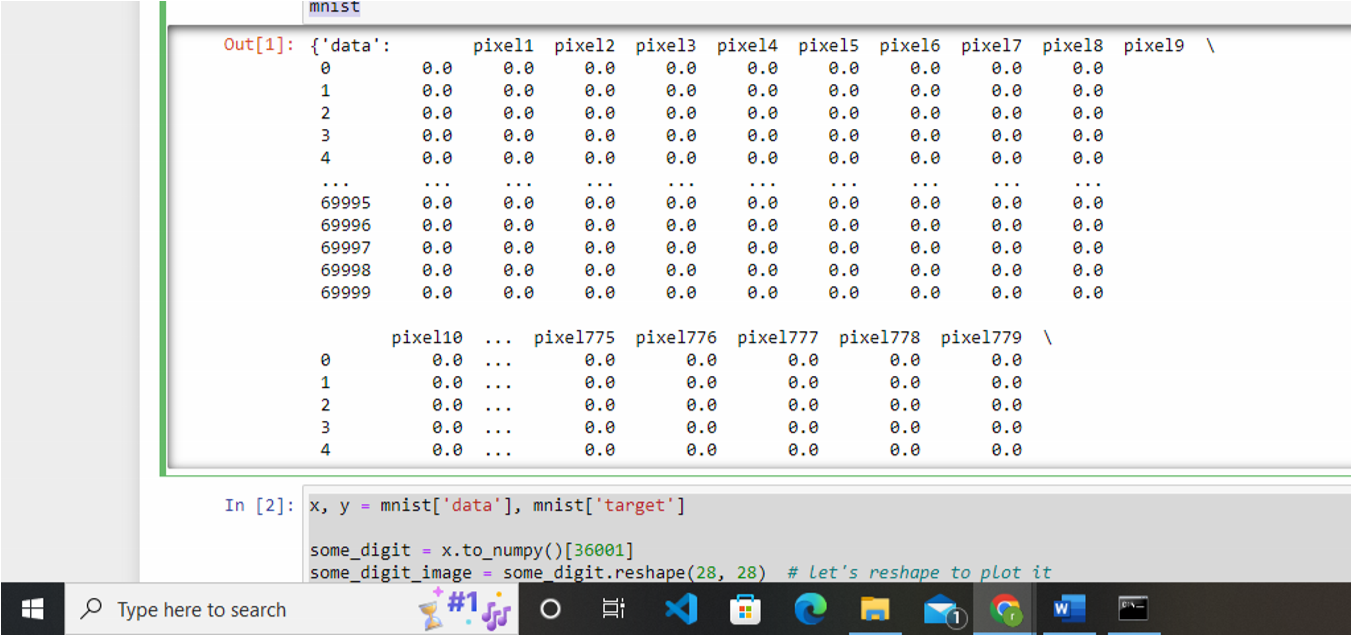
1. Set of 70,000 small images of digits handwritten by high school students and employees of the US causes Bureau.
2. All images are labeled with the respective digit they represent.
3. MNIST is the hello world of machine learning. Every time a data scientist or machine learning engineer makes a new algorithm for classification, they would always first check its performance on the MNIST dataset.
4. There are 70,000 images and each image has 28\*28 = 784 features.
5. Each image is 28\*28 pixels and each feature simply represents one-pixel intensity from 0 to 255. If the intensity is 0, it means that the pixel is white and if it is 255, it means it is black.

# CODE:

from sklearn.datasets import fetch\_openml import matplotlib import matplotlib.pyplot as plt import numpy as np

from sklearn.linear\_model import LogisticRegression from sklearn.model\_selection import cross\_val\_score mnist = fetch\_openml('mnist\_784') mnist

**OUTPUT:**

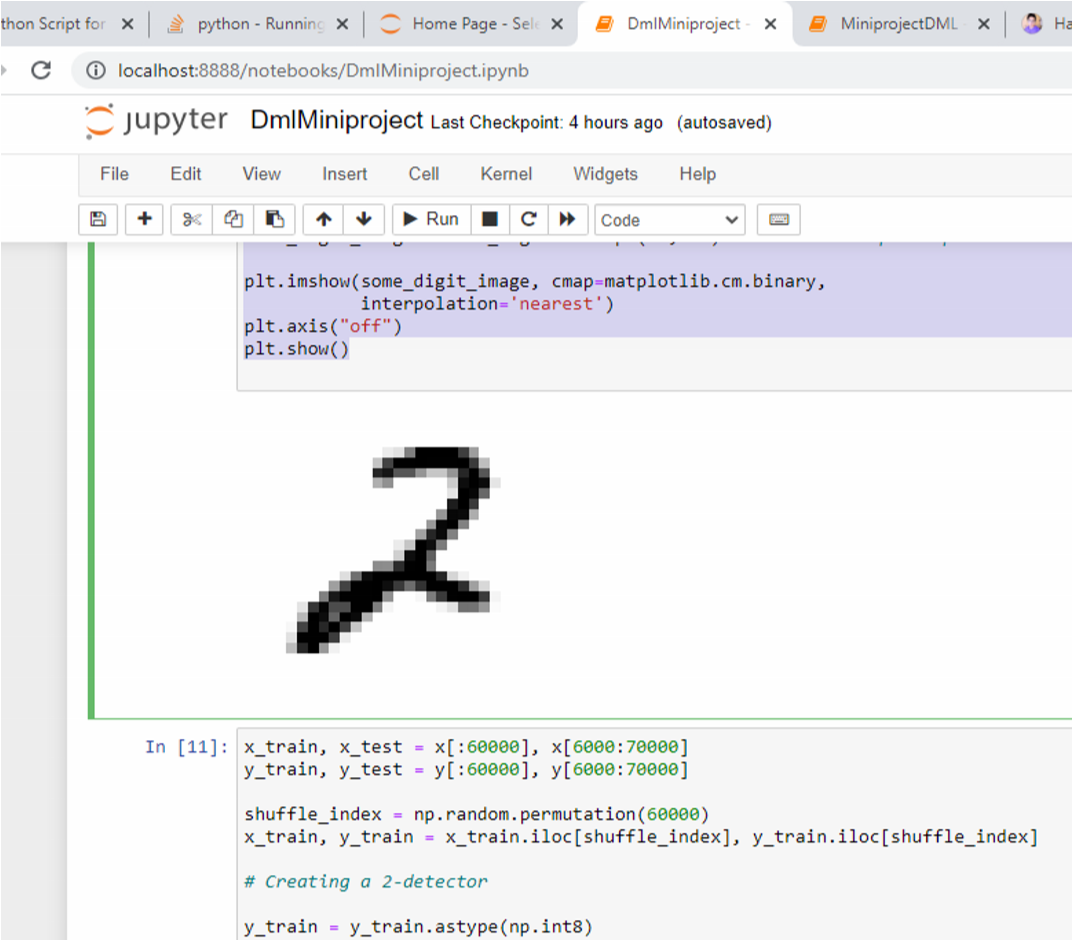


# CODE:

x, y = mnist['data'], mnist['target'] some\_digit = x.to\_numpy()[36001]

some\_digit\_image = some\_digit.reshape(28, 28) # let's reshape to plot it plt.imshow(some\_digit\_image, cmap=matplotlib.cm.binary, interpolation='nearest') plt.axis("off") plt.show()

**OUTPUT:**



# CODE

x\_train, x\_test = x[:60000], x[6000:70000] y\_train, y\_test = y[:60000], y[6000:70000] shuffle\_index = np.random.permutation(60000)

x\_train, y\_train = x\_train.iloc[shuffle\_index], y\_train.iloc[shuffle\_index]

# Creating a 2-detector y\_train = y\_train.astype(np.int8) y\_test = y\_test.astype(np.int8) y\_train\_2 = (y\_train == 2) y\_test\_2 = (y\_test == 2)

clf = LogisticRegression(tol=0.1) clf.fit(x\_train, y\_train\_2)

**OUTPUT**

**LogisticRegression(tol=0.1)**

# CODE:

a = cross\_val\_score(clf, x\_train, y\_train\_2, cv=3, scoring="accuracy") print(a.mean())

In [15]:

**OUTPUT:**

