Use case scenario:

In warehouses where camera, sensors and manual effort are used to record categories, materials, labels, barcode, packaging etc of the products delivered. The dataset which we are going to build will use the recorded data to identity target label and features. The dataset will be used to train, test and predict product category for any unseen data.

Objective : Identify the categories of the product coming to the warehouses.

Implementation: Gathered the data from different sources like barcode text, shipping invoices, product id etc. I did text preprocessing by converted all text to lowercase, use label encoding to represent text in a numeric format, I choose non- Naïve bayes classifier for training the testing set.

MNIST Data Set-

MNIST dataset has 70,000 images and 784 attributes, representing the digit(0-9) corresponding to the handwritten image. Imported the libraries like NumPy, matplotlib, panda, and seaborn. Read the csv files and dropped the column like “unnamed” and “index”. Converted the first column from the array and assigning it o variable y. I did data visualisation making histogram, trained the training and testing data after normalising the dataset. Plotted a confusion matrix on true label and predicted label, and calculated the accuracy of train and test data.