Data Intake Report

Name: Deployment on Flask – Digit Recognition Using MNIST dataset
Report date: 27/05/2023
Internship Batch: LISUM21
Version:1.0
Data intake by: Zohra Bouchamaoui
Data intake reviewer:
Data storage location: GitHub

Image data details – MNIST Dataset

Training set number of observations	60,000
Test set number of observations	10,000
Image size	28x28 pixels
Image type	Grayscale
Features	10 classes (Digits from 0 to 9)

Proposed Approach:

- **Loading the Data:** The MNIST dataset is loaded using the mnist.load_data() function. This function returns four NumPy arrays: train_images and train_labels for the training set, and test_images and test_labels for the testing set.

- **Data Preprocessing:** Before feeding the data into the machine learning model, some preprocessing steps are necessary including normalizing pixel values, converting images into grayscale. These steps ensure that the data is in a suitable format for training the model.

- **Model Training:** After preparing the data, we used it to train a machine learning model. In the case of the MNIST dataset, we used a convolutional neural network (CNNs). The model is trained using the training set.

- **Model Evaluation:** After training the model, we evaluated it using the testing set by assessing its performance using different metrics such as accuracy, F1-score, precision, and recall.