Project Summary and Steps Taken:

1. **Loading the Dataset:** The dataset was loaded for analysis and model training.
2. **Data Exploration:** The shape information and description of the dataset were examined to understand its structure.
3. **Data Cleaning:** Additional columns, which were found to be repetitions of the index, were dropped from the dataset.
4. **Statistical Distribution:** The statistical distribution of the dataset was analyzed to gain insights into its characteristics for effective model training. Labels were found approximately evenly distributed
5. **Scaling:** To prepare the data for modeling, normalization was performed by dividing each value by the maximum value, resulting in features scaled to the range of 0 to 1.
6. **Data Preparation:** The shape of the feature matrix (X) and the target vector (y) was converted into the required NumPy array format for model training.
7. **Splitting the data:** We have train and test data, so train data was divided into validation set in 80:20 ratio.
8. **Hyperparameter Tuning:** Different values of epsilon were defined to tune the Naive and Gaussian Bayes model and checked on a validation set.
9. **Model Training and Evaluation:** The model was trained and evaluated for each epsilon value to determine the best performing epsilon. The best gaussian base model achieved an accuracy of 95% with an epsilon value of 0.01.
10. **Confusion Matrix for Error Analysis:** A confusion matrix was generated to analyze the errors made by the model. Some notable error patterns observed were:

- 28 errors between the digits 7 and 9.
- 27 errors between the digits 2 and 8.
- 22 errors between the digits 7 and 1.
- 23 errors between the digits 4 and 9.

To minimize errors, Future recommendations:

Apply Image Rotation and Noises: By slightly rotating images and introducing noise, the model can better generalize and accurately recognize digits in varied orientations or noisy conditions.

Utilize Data Augmentation: If possible,we can enrich the model with diverse examples, improving its accuracy in recognizing digits across various scenarios.