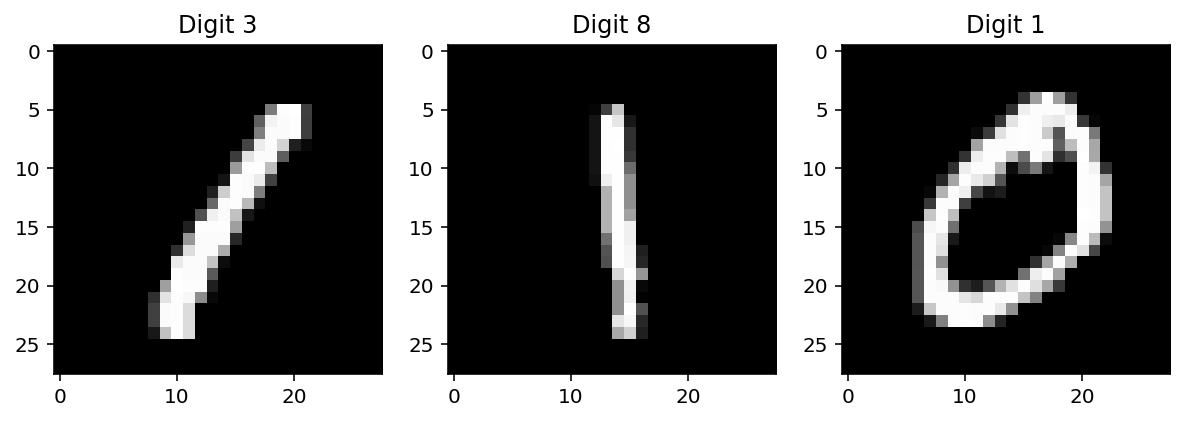
# Analysis report of Lab assignment : “Classification”

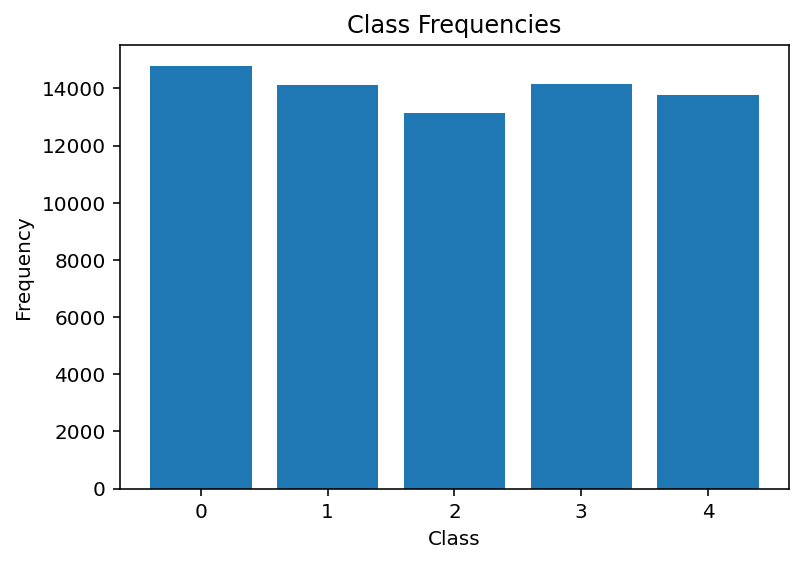
## Load & check the data

This plot shows the three selected digits from the MNIST dataset:

* **Digit 3**: The first subplot.
* **Digit 8**: The second subplot.
* **Digit 1**: The third subplot.

These are the digits corresponding to indices 3, 8, and 1 from the X\_sabra array. The imshow function visualizes these digits as grayscale images.

## Pre-process the data



This plot shows the **class frequencies** after transforming the target variable into 5 classes:

* Class 0: 14,780 samples
* Class 1: 14,131 samples
* Class 2: 13,137 samples
* Class 3: 14,169 samples
* Class 4: 13,783 samples

The bar chart indicates that the dataset is relatively balanced across the 5 classes, which is good for training classification models.

1. Build Classification

#### ModelsNaïve Bayes

**Cross-Validation Scores**: [0.74833642, 0.75912289, 0.75885016]

The model achieves an average cross-validation accuracy of around **75.8%**.

**Test Accuracy**: **77.34%**

The model's accuracy on the test set is **77.34%**, which is consistent with the cross-validation results.

**Predictions for Selected Digits**:

* Digit 3: Predicted as **Class 0**
* Digit 8: Predicted as **Class 0**
* Digit 1: Predicted as **Class 0**
* These predictions are incorrect, indicating that the Naive Bayes model struggles with these specific digits

#### **Logistic Regression Classifier**

**Cross-Validation Scores**: [0.87165921, 0.88267059, 0.86554301]

The model achieves an average cross-validation accuracy of around **87.3%**.

**Test Accuracy**: **88.55%**

The model's accuracy on the test set is **88.55%**, which is significantly better than Naive Bayes.

* **Predictions for Selected Digits**:
  + Digit 3: Predicted as **Class 0**
  + Digit 8: Predicted as **Class 0**
  + Digit 1: Predicted as **Class 0**
  + Similar to Naive Bayes, Logistic Regression also misclassifies these digits.

#### ****Naive Bayes vs. Logistic Regression****

* **Accuracy**: Logistic Regression (**88.55%**) outperforms Naive Bayes (**77.34%**).
* **Predictions**: Both models misclassify the selected digits (3, 8, 1) as Class 0. This suggests that these digits might be challenging for the models, possibly due to their similarity to Class 0 digits.

The results show that Logistic Regression performs better than Naive Bayes for this task. However, both models struggle with certain digits, indicating room for improvement.