Dataset 1: Spam Email / SMS Dataset

Purpose:
Used for **binary text classification** to identify spam messages.

Key Details

- **Type:** Text dataset (SMS / Email messages)
- **Total samples:** ~5,157 messages
- **Number of classes: 2**

    - ❖ ham → Not spam
    - ❖ spam → Spam

Class distribution:

- Ham = 87%
- Spam = 13%

LINK: https://www.kaggle.com/datasets/mfaisalqureshi/spam-email

Dataset 2: MNIST Handwritten Digit Dataset

Purpose:
Used for **multiclass image classification** to recognize handwritten digits.

Key Details

- **Type:** Image dataset (grayscale)
- **Total samples: 70,000 images**
    - 60,000 training
    - 10,000 testing
- **Number of classes: 10**
    - Digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
- **Image size:** 28×28 pixels
- **Features:**
    - Pixel intensity values (0–255)
    - Flattened into 784 features for classical ML

LINK: https://www.kaggle.com/datasets/oddrationale/mnist-in-csv