Instructions

In this task, we will use the **MNIST database**. As stated by the creators of the dataset, "The MNIST database of handwritten digits, available from this page, has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalised and centred in a fixed-size image."

First read and run the **MNIST.pynb** example file to explore the MNIST test set provided by sklearn. Then, follow the instructions to create a random forest model using the mnist data. Use the MNIST training set for training and development and the MNIST test set as the validation set

Compulsory Task 1

Follow these steps:

- Create a file called **mnist_task.ipynb**.
- Load the MNIST dataset. You may **download the files** or use a library such as Keras to access the dataset (e.g. `from keras.datasets import mnist`).
- Split the training data into a training and development (test) sets.
    - Add a comment explaining the purpose of the train, development(test) and test(validation) sets.
- Use the RandomForestClassifier built into sklearn to create a classification model.
- Pick one parameter to tune, and explain why you chose this parameter.
- Choose which value for the parameter to set for testing on the test data and explain why.
- Print the confusion matrix for your Random Forest model on the test set.
- Report which classes the model struggles with the most.
- Report the accuracy, precision, recall, and f1-score.

If you are having any difficulties, please feel free to contact our specialist team **on Discord** for support.