## Project– Fall 2022

| *Course Code* | *Course* |
|---|---|
| **CSE 381** | **Introduction to Machine Learning** |
| **Instructor/s** | **Dr. Mahmoud Khalil, Eng. Mahmoud Soheil** |

# Project Description

This project is developed in teams of 2-3 students.

In this project, you will design five classifiers: a naive Bayes classifier, a SVM classifier, a KNN classifier, MLP classifier and a decision tree. You will test your classifiers on two image data sets: a set of scanned handwritten digit images and a set of face images in which edges have already been detected. Even with simple features, your classifiers will be able to do quite well on these tasks when given enough training data.

Optical character recognition (OCR) is the task of extracting text from image sources. The first data set on which you will run your classifiers is a collection of handwritten numerical digits (0-9). This is a very commercially useful technology, similar to the technique used by the US post office to route mail by zip codes. There are systems that can perform with over 99% classification accuracy

Face detection is the task of localizing faces within video or still images. The faces can be at any location and vary in size. There are many applications for face detection, including human computer interaction and surveillance. You will attempt a simplified face detection task in which your system is presented with an image that has been pre-processed by an edge detection algorithm. The task is to determine whether the edge image is a face or not. There are several systems in use that perform quite well at the face detection task.

The code for this project includes the following files and data, available as a zip file attached with this description

### Data file

| | |
|---|---|
| data.zip | Data file, including the digit and face data. |

### Files you should read but NOT edit

| | |
|---|---|
| samples.py | I/O code to read in the classification data. |
| util.py | Code defining some useful tools. |

# Project Milestones

The two milestones shall be submitted by one of the team on the LMS and mention the team members with their IDs on cover page. The deadlines are shown on LMS.

## ➢ First Milestone

**Requirements:**

- Load the two datasets using the code in samples.py
- Train the Naïve Bayes Classifier and KNN classifier on both datasets and change the hyperparameters (if exist)
- Test the classifiers on the testing dataset and show the accuracy for each run when changing the hyperparameters

**Deliverables:**

- Detailed report containing steps, screenshots of the code, screenshots of the output, visualization of the accuracy change, the outputs (faces detected or numbers detected) and show the reason of using the final values of the hyperparameters.
- Python code of the whole project
- A readme file containing the steps to run the project and what libraries to be imported if any.

## ➢ Second Milestone

**Requirements:**

- Train the MLP Classifier, SVM classifier and build decision tree on both datasets and change the hyperparameters (if exist)
- Test the classifiers on the testing dataset and show the accuracy for each run when changing the hyperparameters

**Deliverables:**

- Detailed report containing steps, screenshots of the code, screenshots of the output, visualization of the accuracy change and the outputs (faces detected or numbers detected) and show the reason of using the final values of the hyperparameters.
- Python code of the whole project
- A readme file containing the steps to run the project and what libraries to be imported if any.