

**CMSC-5723 Machine Learning**  
Project - Can I Get Your Digits?  
Due: Friday, October 11, 11:59 PM  
(2000 pt)

For this project, you will create a Python program which will perform classification on the attached data. If you are not familiar with Python I would suggest getting familiar with Python, as the vast majority of machine learning tools have been written using Python and Python libraries. This should be a python script (.py), which can be run from the command line using `python <scriptname>.py`. You can include other files with it that are required to run it.

The `optdigits.names` file contains the description of the data, and each row of the training and test data files contain a single row of data, ending with the classification for that row. There are ten potential classes, representing the digits 0 through 9.

Your program should do the following:

1. Ask the user if they want to train the model.
  - a. If they do:
    - i. Prompt the user for the training data file
    - ii. Build the model
    - iii. Store the model for next run
  - b. If they don't:
    - i. Load the stored model from disk
2. Prompt the user for the test data file
3. For each instance in the test data file, output (to standard output) that instance followed by the classification your model gives it.

Attached to this are a split of the data into training and test sets.

You do **not** have to implement your own machine learning model. You can use libraries which implement whichever model you choose. If these libraries require the data in a specific format, your program should transform the data (without replacing the data files) into the proper format without the user having to change anything.

If you **choose** to implement your own model, you will receive 500 bonus points, as long as you did not use outside help to implement the model (other people, generative AI, etc).

The score for this project will be calculated as follows:

Let  $Acc_{given}$  be the overall accuracy of your classifier on the given test set using the given training set, and let  $Acc_{secret}$  be the overall accuracy of your classifier on my secret test set given my secret training set. The score will be:

$$Acc_{given} * 1000 + Acc_{secret} * 1200$$

**Notes on Generative AI usage:** I can't stop you from using generative AI to help you solve this problem. If you do, you should provide:

- Which AI products you used
- What prompts, if any, you provided to the AI model
- Screenshots of any interactions you had with the AI

If you use generative AI and do not provide documentation of it, then you will receive a grade of zero.

To turn in this project, you should provide the following files:

- The source code for your program, and any dependencies (libraries) your program requires
- Any required generative AI documentation (see above)

Note, the source file **must** be a simple **.py** file which when run on the console performs the desired task. I will **not** accept Jupyter notebook (.ipynb) files.

Good luck!