# COMP47750
# Machine Learning with Python

## 2022 Assignment 1

### *k*-NN DTW

## Objective

The objective of this assignment is to implement and test a *k*-Nearest Neighbour (*k*-NN) classifier for time-series data. One of the big benefits of *k*-NN is that it can be used with data that is not in the normal feature vector format if a distance measure for that data can be found. Dynamic time warping (DTW) is such a measure for time series ([DTW Wikipedia page](#)).

## Requirements

The notebook `Basic_DTW` contains some basic code to help you get started.

**Task 1: 1-NN DTW**

1. Use the DTW function in the notebook to implement a simple 1-NN classifier for time-series data.

2. Test the performance of this classifier on the dataset provided in the file `UMD_TEST.txt.` This is a three-class synthetic time-series dataset - see details <[here](#)>.

3. Given that the time-series are all the same length (150 ticks) Euclidean distance can also be used as distance metric. Compare the 1-NN DTW performance with 1-NN Euclidean. Use the **scikit-learn** implementation for the Euclidean model.

4. Find the best value for the **window** parameter for this dataset.

**Task 2: *k*-NN DTW**

1. The *k*-NN classifier in **scikit-learn** has a **metric** parameter that allows for user-defined distance metrics. Adapt the DTW code provided so that it can be incorporated in a **scikit-learn** *k*-NN classifier.
2. Test the performance of this classifier and compare with the 1-NN results from Task 1. Verify that the 1-NN results are consistent.
3. Compare with *k*-NN Euclidean.

**Hint 1:** It will be a good idea to use a reduced version of the test set for testing to save time.

**Submission:** Given the length of time it will take to run the notebook on the full dataset, you should run your notebook from start to finish before submitting and submit with all the outputs in place.

This is an individual (not group) project. Submission is through the Brightspace page. Your submission should comprise your notebook in a single zip file. You should use markdown cells in the notebook to report your findings and conclusions.