# COMP30120 Assignment 1: Introduction to Weka

**Deadline:** Monday September 28th 2015

**Submission:** Submit your report as a single PDF file via the COMP30120 CS Moodle page. Include your full name and student ID number in the report.

**Overview:**

The objectives of this assignment are to get started using the WEKA Machine Learning environment and to perform a comparative evaluation of the performance of a range of classifiers on a supplied dataset.

**Data:**

The data source in question relates to restaurant reviews, each represented by 24 summary features. Each review also has a binary class label, indicating that it is either deemed "helpful" or "unhelpful" for other users.

You should download your personal dataset for the assignment from the URL:
*http://mlg.ucd.ie/datasets/comp30120/restaurant/<STUDENT_NUMBER>.arff*
For example, if your student number is 126023491, your dataset is at the URL:
*http://mlg.ucd.ie/datasets/comp30120/restaurant/126023491.arff*

Note: When downloading the dataset, please ensure your student number is correct. Submissions using an incorrect dataset will receive a 0 grade.

**Tasks:**

1. Firstly, examine the performance of the Naive Bayes classifier on your dataset using three different evaluation methods:

    1. the training test as test set

    2. using a 60%/40% training/test percentage split

    3. using 10-fold cross validation

2. Next, as a comparison, test the following classifiers on your dataset using the same evaluation methods:

    1. *k*-Nearest Neighbour classifier with *k=1* neighbour

    2. *k*-Nearest Neighbour classifier with *k=3* neighbours

    3. Support Vector Machine (Functions → SMO)

3. Write a report comparing the performance of these classifiers on this dataset using the three evaluation methods. Discuss the usefulness of the different evaluation methods. Recommended page length for the report is 2-3 pages, although there is no penalty for exceeding this length.