Your name: Course code (e.g. COMP309, COMP440):

**COMP 309 | Machine Learning Tools and Techniques**

Assignment 1: Sprint on One Dataset Each

16% of Final Mark | Due: 11:59pm Tuesday 4 August 2020

**2.1 Core: Investigate Basic Use of the Different Tribes of AI**

Present and describe in the report what you consider to be the important aspects of the *results* of each technique on the one dataset:

Why do each of the four tools belong to a given tribe? Identify the important aspects of the technique in terms of General description, especially the:

General description, plus

1. Representation,

2. Evaluation method,

3. Optimization driver.

Illustrate any important differences using the dataset.

**2.2 Completion 1: Consider a Pipeline for Dataset Processing**

The report should include:

1. Business understanding - consider the business aspects of the dataset, e.g. why was the data gathered? What did the acquisition hope to achieve? Note, that this may be mere obvious in some datasets than others.
2. Data understanding - not only should the metadata be described (which is readily available in the UCI repository), but any interesting factors should be noted, e.g. mixed attribute type, high epistasis, outlier/noisy/missing data instances.
3. Data preparation - state how the pipeline could assist in the preparation of the data prior to the technique being applied.
4. Modelling - state whether this pipeline suits one or more of the five tribes of AI.
5. Evaluation - similarly, state whether this pipeline supports one or more methods to evaluate a solution.
6. Deployment – explain whether the model produced can easily be deployed or whether additional effort is required

**2.3 Completion 2: Use the pipeline to reevaluate the selected techniques** in Part 2.1 used to classify the dataset

Compare and contrast the results between the different tribes.

Screen capture(s) for your pipeline including classifiers.

The report should include:

1. Accuracy in terms of the fraction of the test instances that it classified correctly.
2. Report a snapshot of the learned classifiers discovered by your program.
3. Compare the accuracy of your techniques before and after using a data pipeline approach. Please comment on any differences, suggesting reasons.

**2.4 Challenge: Use the HeuristicLab to evaluate the Evolutionary Computation tribe**

Screen captures for your classifiers including both set-up details and pipeline(s). The program should print out the classifiers in as human readable form (text form is fine) as possible.

1. Accuracy in terms of the fraction of the test instances that it classified correctly.
2. Report a snapshot of the learned classifiers discovered by your program.
3. Compare the Evolutionary Computation approach with the other AI tribes from earlier parts.
4. Compare the WEKA tool with the HeuristicLab tool commenting on ease-of-use, performance and anyother aspects you consider important (e.g. data format, documentation, online tutorials and so forth)