

# Data Mining Assignment 2, [15 Points]

## Due March 12<sup>th</sup>, 2017 @ 23:00

Your name goes here

February 18, 2017

I declare that all material in this assessment task is my own work except where there is clear acknowledgement or reference to the work of others and I have complied and agreed to the UIS Academic Integrity Policy at the University website URL: <http://www.uis.edu/academicintegrity>

Student Name: Your name here UID: Your UID here Date: February 18, 2017

## 1 Task Prep

There is a C program to create LED data attached, compile it on your CLang compiler (do not modify the code) then run it. To get help on the arguments invoke it without any arguments. To create 1000 samples invoke it by LED 1000 Your UIN temp 2 < noise < 18 .

## 2 Instance Based Learners

Weka contains an implementation of David Aha's IB1 algorithm in weka.classifiers.IB1.

1. Modify it to implement the IB2 algorithm (described in your textbook.)
2. Test your program on the iris data (in iris.arff.)
3. Test your program, and J48, on a sample of the 24-attribute LED data and check whether your results agree with Aha's. (Do not try too hard to replicate Aha's results exactly.) You need to use only 10 trials, rather than the 50 that Aha used.

### 3 More Instance Based Learners

1. Research, select, name, and describe your use of a weka filter to split an ARFF file into training and test sets.
2. Implement the IB3 algorithm and test it in the same way. (IB3 is described in your textbook.)
3. Run a series of experiments on the glass data (in glass.arff) to provide new data on the relative performance of IB1, IB2, IB3, and C4.5.

Present your results so that they are easy to understand and explain them carefully, comparing them with Aha's other results.

The last line in your report file must be the URL of your video presentation.

### 4 Video Presentation

Record a video showing the running of your implemented IB2/3, and your description of your code (line by line) and submit it. I do not require visual of your person if you do not wish to appear in your recording just vision and commentary of the classifier running on your weka workbench as well as vision and commentary on your own code line by line.

For information on video recordings for blackboard please look at the material at this URL:

[https://cdnapisec.kaltura.com/index.php/extwidget/preview/partner\\_id/1371761/uiconf\\_id/31473632/entry\\_id/1\\_w4w37o8a/embed/dynamic](https://cdnapisec.kaltura.com/index.php/extwidget/preview/partner_id/1371761/uiconf_id/31473632/entry_id/1_w4w37o8a/embed/dynamic)

Submit source Code in 1 only text file(code.txt) in the Turnitin Code-dropbox, and Pseudocode algorithm in 1 only text file (alg.txt) in the Turnitin Algorithm-dropbox on blackboard.

I will use a variety of source code similarity checkers to verify compliance with UIC academic integrity policy.

File formats and academic integrity are not optional, non compliant submissions will not be considered and a summary grade of 0 will be awarded.

## 5 Submission

1. Report Turnitin Report dropbox
2. Source Code text file Turnitin Code dropbox
3. Algorithm Pseudocode text file Turnitin Algorithm-dropbox
4. Video upload on blackboard as per information at the URL given
5. 20 LED datasets uncompressed .tar *not optional* archive in dataset dropbox