RMIT University
School of Science
COSC2110/COSC2111 Data Mining
Laboratory Week 6

Aims of this lab

- Learn how run attribute selection algorithms and interpret the results.
- Learn about the potential effectiveness of attribute selection

---

1. You will need to have access to the WEKA package.
2. The data files for this lab can be found at `/KDrive/SEH/SCSIT/Students/Courses/COSC2111/DataMining/data`
3. Load the file `/arff/UCI/ionosphere.arff`
   (a) Get the classification accuracy with J48.
   (b) Apply attribute selection with the default settings, ie CfsSubsetEval and BestFirst. Go back to Preprocess, remove all but the selected attributes and rerun J48. What is the accuracy?

   [Useful hint] Go to `Preprocess -->Filters--> Supervised –Attribute` and select the AttributeSelection filter. Using this filter (with Undo) will save the tedious task of manually selecting the results of attribute selection.
   (c) Reload `ionosphere.arff` Apply attribute selection with WrapperSubsetEval, BestFirst and J48 as the classifier in WrapperSubsetEval. Go back to Preprocess, remove all but the selected attributes and rerun J48. What is the accuracy?
   (d) What do you conclude about the value of attribute selection?
   (e) Explore other combinations of evaluator and search method. Can you find anything better?
4. The file `/arff/UCI/isolet.arff` has 618 attributes. Explore a variety of attribute selection techniques to reduce the number of attributes without reducing accuracy. What is your best result?
5. Load the file `other/bank-balanced.csv`
   (a) Repeat (3) on this file.
   (b) Open the file with a text editor and read the descriptions of the attributes. Does the set of selected attributes make sense from what you know about banks? There is a description of the fields in `bank-names.txt`