



Data Mining

CS475

Spring 2019

Project

Name: Youssef Ali Fayed

ID: 162085

First of all let's understand the data and what we need to do with it.

We are working on the student-mat file, it contains 30 attributes plus 3 main classes, the data consists of a mix of nominal and ordinal data so it needs to be preprocessed first before doing any work on it.

- First step is to choose what is the main class attribute, after reading the info file we found that G3 is the main class attribute.
- second step is to remove any un-necessary attributes and to do so we will use the attribute selector in WEKA and getting the information gain.
but to do so we need to convert the class to nominal data and we can do this by applying a filter.
- After doing all these steps we found that attributes (G2,G1,Absence,Failures,Age and mjob) are the main best attributes.
- Then we removed all other attributes.

By this all our data is pre-processed and we can now do whatever we want on it.

❖ **The first thing we need to do is the two association rules, which we can easily do by tabbing in to the association tab and choose the two rules that we want.**

- ❖ The first one will be the Apriori algorithm.
- ❖ The second one will be filtered association.

❖ **Then the second thing is the three clustering algorithms.**

- ❖ The first one will be simple k-means, we need to set the number of clusters and the distance functions.
- ❖ We will set the number of clusters to 2 and the distance to use Euclidean distance.
- ❖ The second algorithm will be EM. The number of cluster can be set to -1 and by this the model will try to determine the number itself but we will set it to 2 to keep it consistent.
- ❖ The third and final one will be cobweb.

❖ **The final thing to do is the classification rules.**

- ❖ The first one is the one-R.
- ❖ The second one is the J48.
- ❖ The third one is naïve bayes.
- ❖ The final one is the logistic model.
- ❖ We can also use the boundary visualizer to visualize the model but we need to convert at least two attributes to numeric data.

**P.S: All These Steps Are Shown in the Attached PDF
(Steps.pdf)**