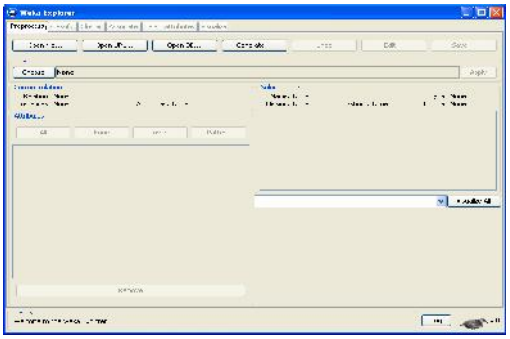
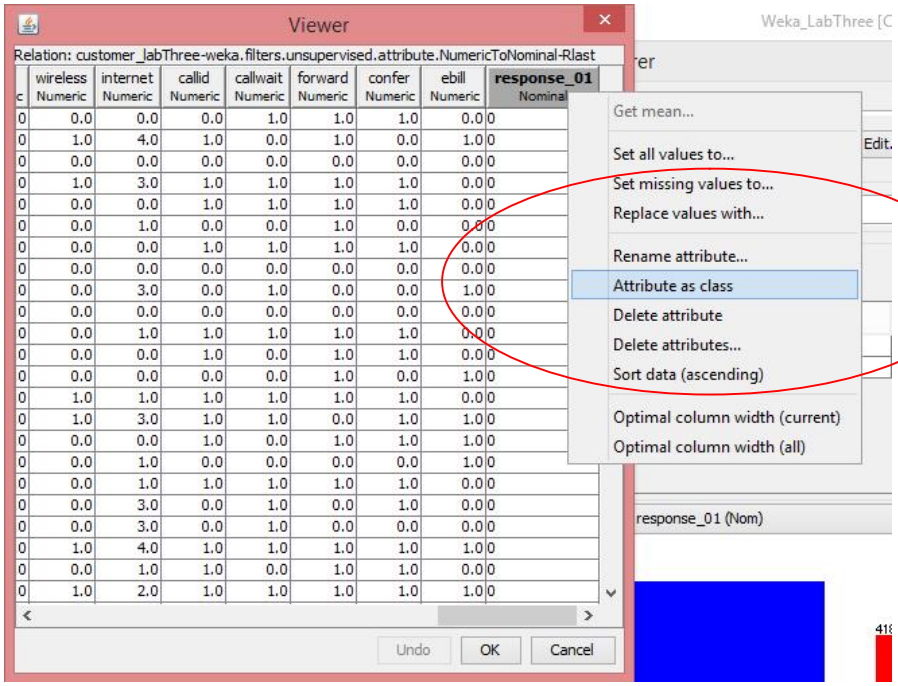
Lab Exercise Three
Classification with WEKA Explorer

1. Open a terminal window from the left bar. Go to directory /opt/weka-3-6-13, then type command :
   java –jar weka.jar.

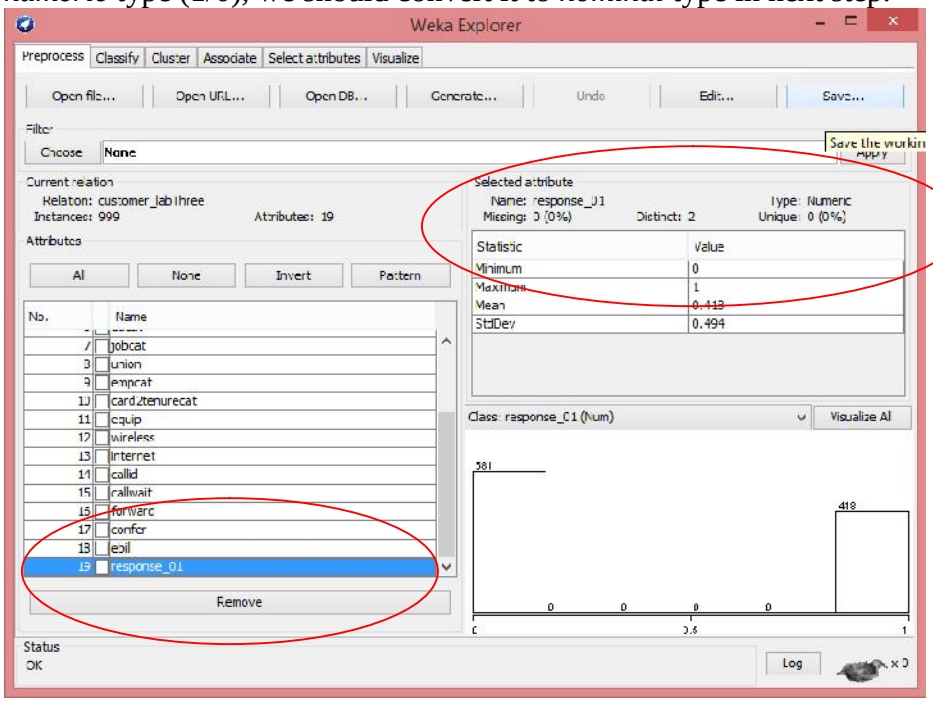
2. Fire up WEKA to get the GUI Chooser panel. Select Explorer from the four choices on the right side.



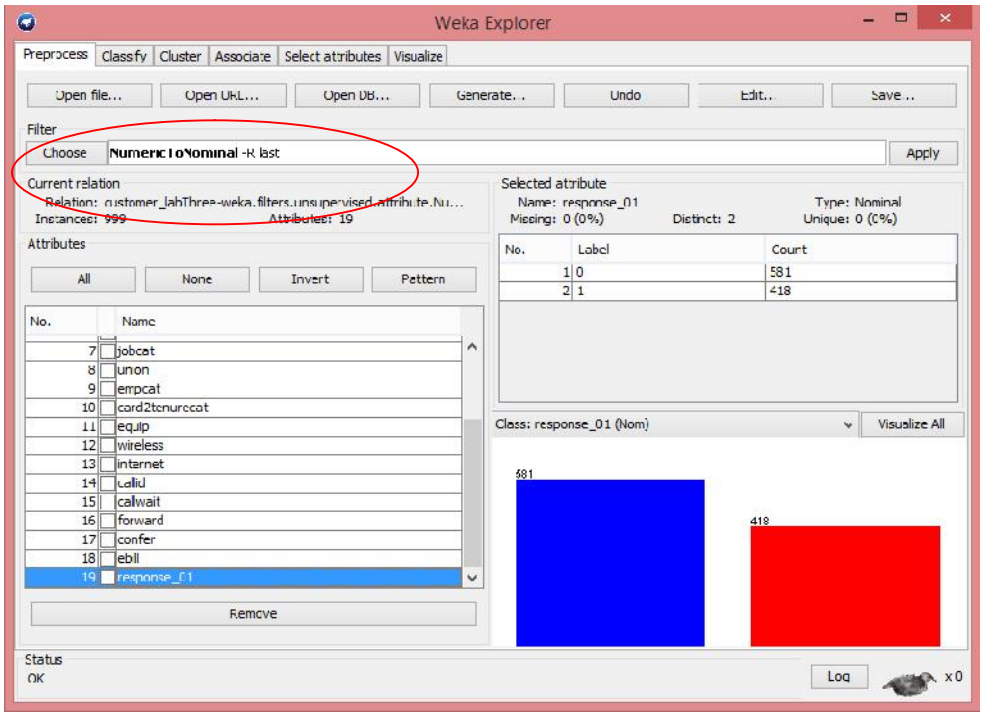
3. We are on ***Preprocess*** now. Click the ***Open file*** button to bring up a standard dialog through which you can select a file. Choose the **customer_labThree.csv** file.
4. To perform classification with Weka, the last attribute in the dataset is taken as class label and it should be *nominal*. If the class attribute is not the last attribute, you could set it in edit window.

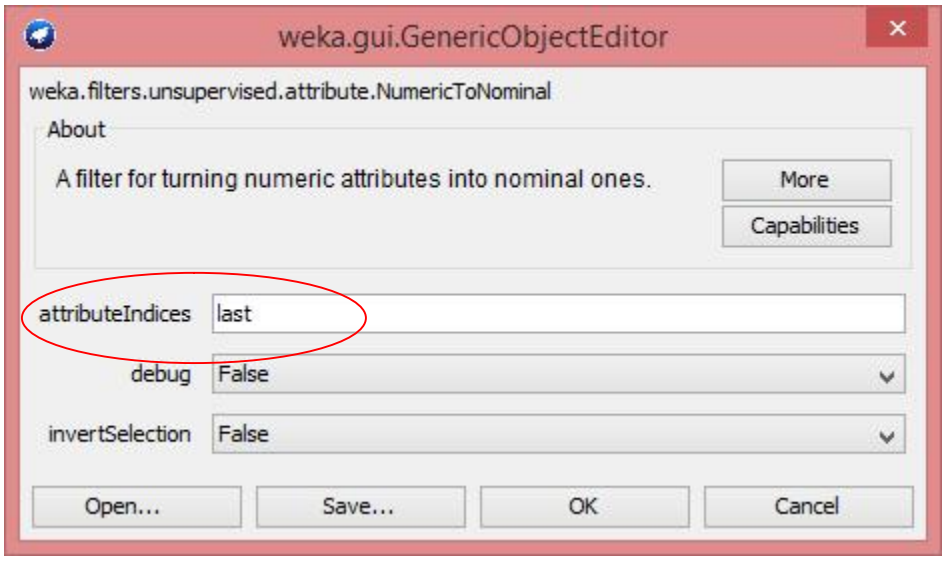


5. Since the class attribute - last attribute of data set **customer_labThree.csv** is *numeric* type (1/0), we should convert it to *nominal* type in next step.

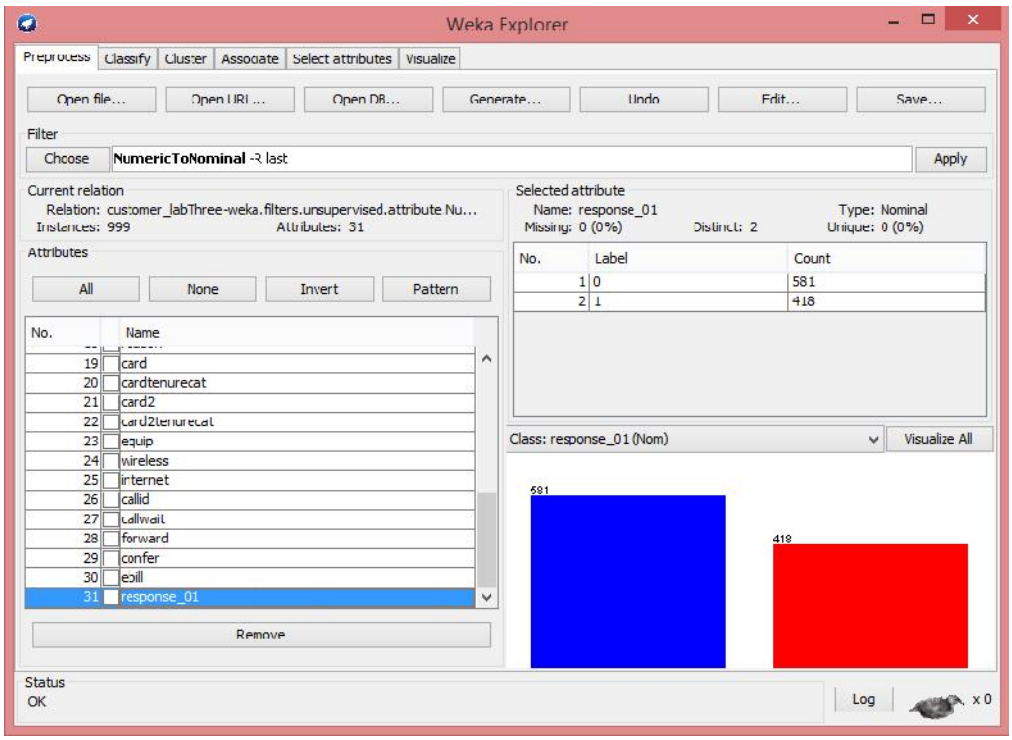


6. Unsupervised attribute filter – ***NumericToNominal*** is chosen to perform this conversion. Since we would like to convert the last attribute only, change the ***attributeIndices*** to **last**.



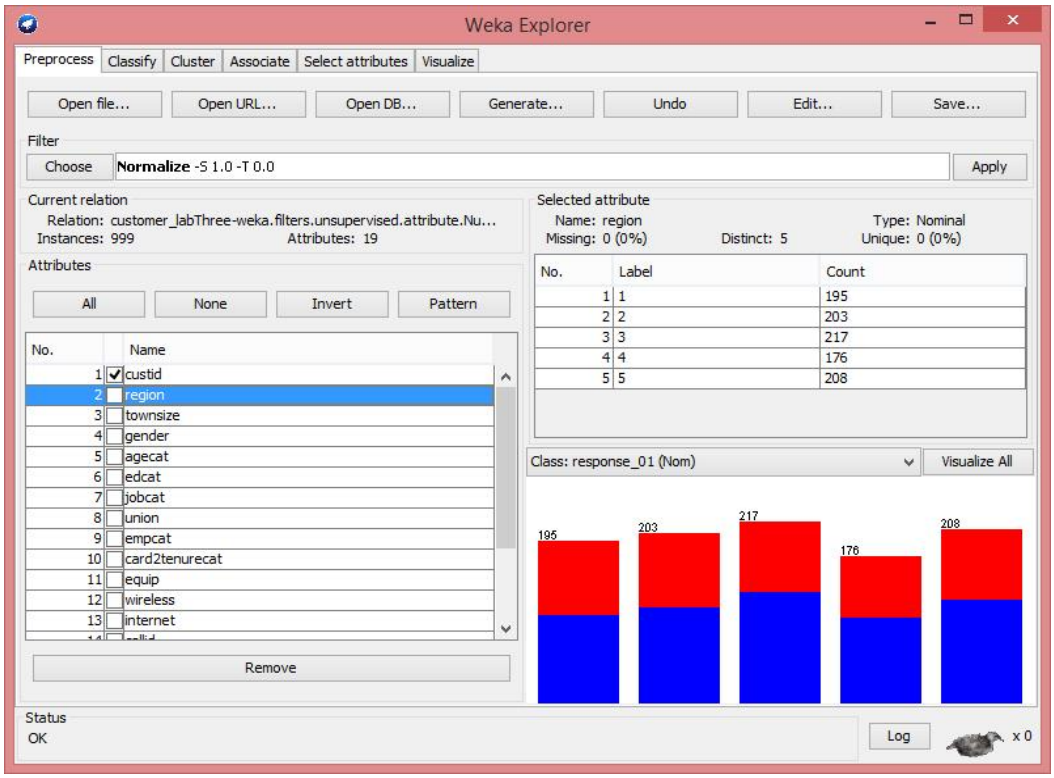


7. After applying the filter, the last attribute becomes nominal type and it is taken as the class label for the dataset – now the data set is visualized in two colors.



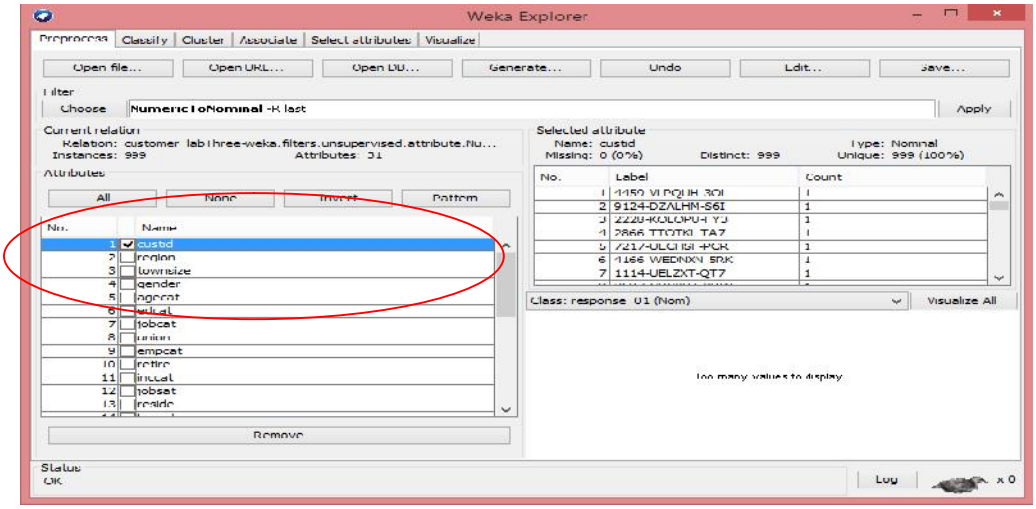
8. You should also convert the types of other attributes. Attributes *region*, *townsize*, *agecat*, *jobcat*, *empcat*, *card2tenurecat*, and *internet* are all nominal values, however, they are treated as **numeric** type by Weka. And attributes *gender*, *union*, *equip*, *wireless*, *called*, *callwait*, *forward*, *confer*, *ebill* are binary values, they are treated as **numeric** types as well. **NumericToNominal** filter should be applied to convert them. After you have done all necessary conversion from numeric to

nominal, you could also normalize attribute ***educat*** to [0, 1] since education categories are rankings.

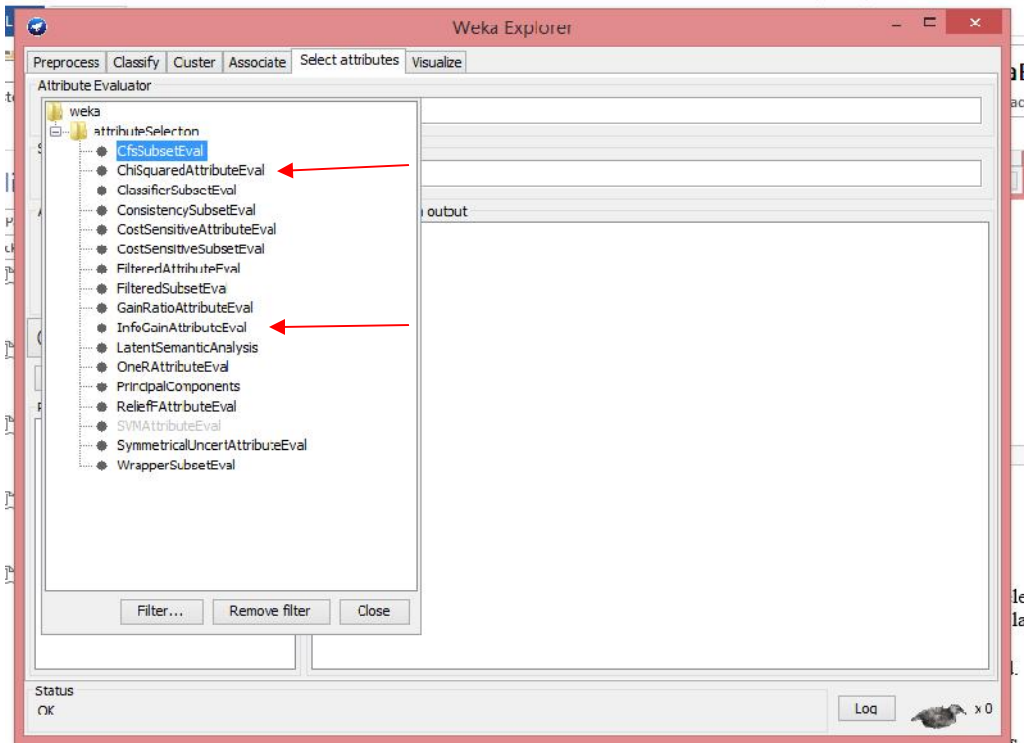


Attribute Selection - Since not all attributes are relevant to the classification job, you should perform attribute selection before training the classifier.

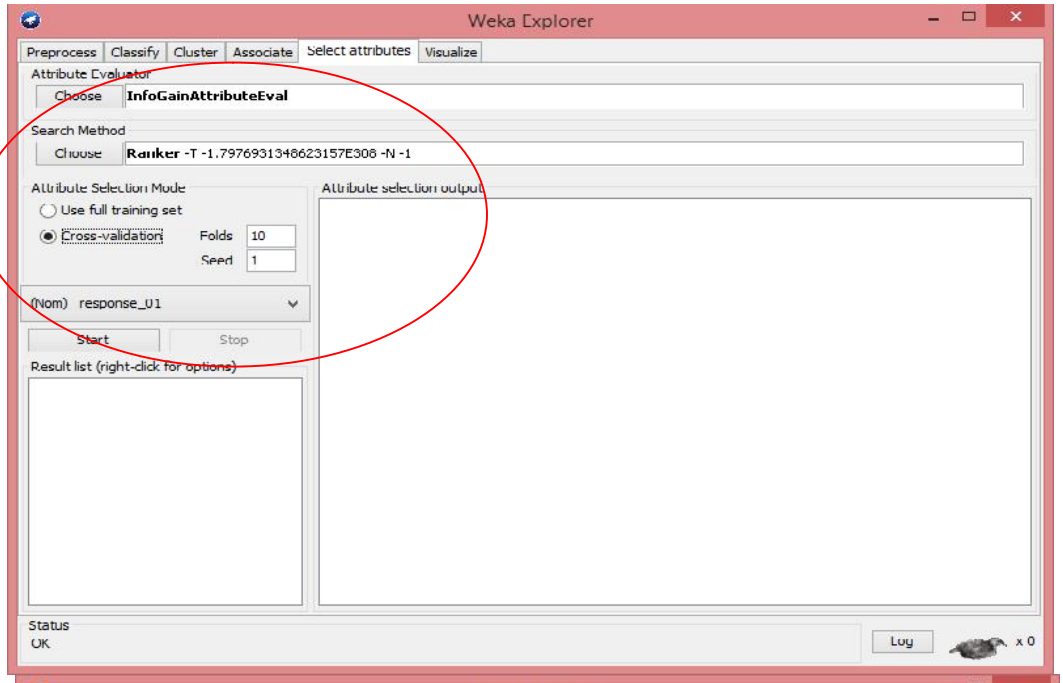
9. You could remove irrelevant attributes by hand. For example, the first attribute ***custId*** should be removed. Select it and click ***Remove*** button to remove it.

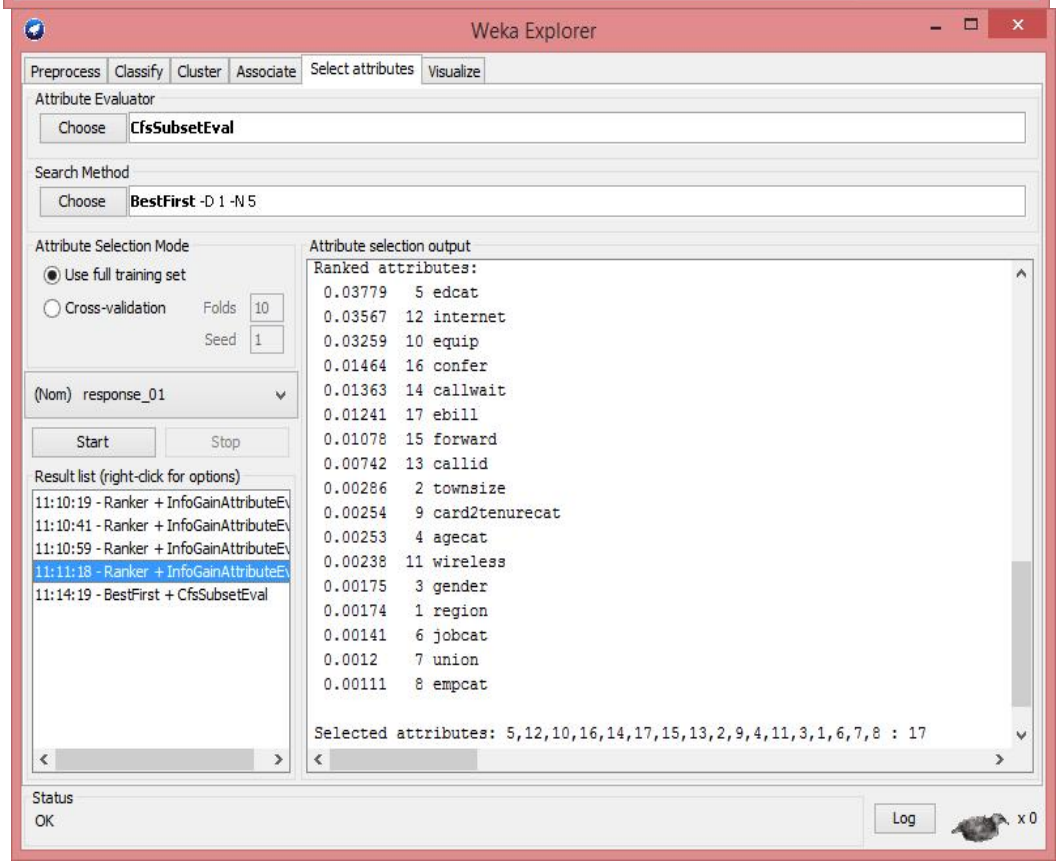


10. You could also run automatic attribute selection on ***Select attributes*** Tab. The default attribute selection method of Weka is **CfsSubsetEval**, which evaluates subsets of attributes. We have introduced two methods of evaluating attributes individually – **InfoGainAttributeEval** and **ChiSquaredAttributeEval**.

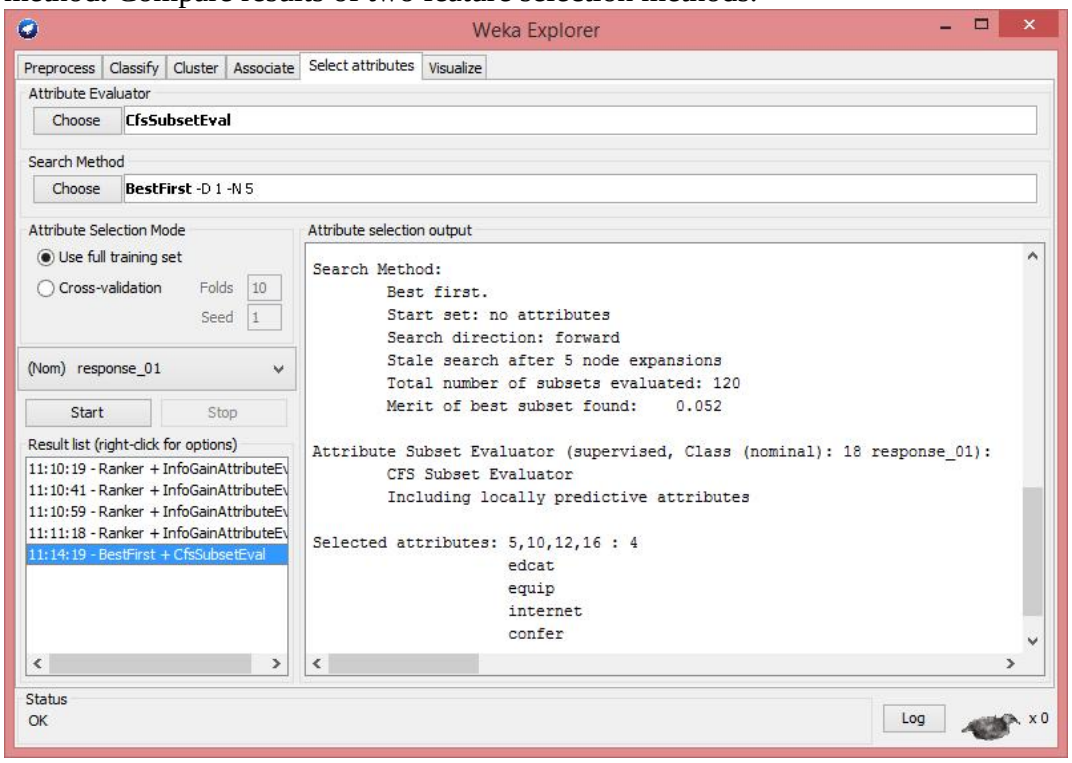


11. To use evaluator **InfoGainAttributeEval**, a search method **Ranker** is selected to rank all attributes regarding the evaluation results. This method treates Missing value as a separate value for the attributes. We use the full dataset as training dataset.





12. Run feature selection the second time with **CfsSubsetEval** and **BestFirst** search method**.** Compare results of two feature selection methods.



13. If you decide to reduce the dataset by removing unimportant attributes, you could choose to save the reduced dataset by right-click the Result list. Save the file name as **customer_N.arff**.