Vamsi Avinash Gunji  
Dr. Clachar  
Data Mining 3010  
Final Part 2

Data Mining Part 2

This is an exciting project I did so far. When I performed the J48, JRIP, ZeroR, and Logistics Classifiers by running the train file on test file, I felt Zero R have the huge amount of data that are correctly classified compared to others. Zero R however being the simplest classification method I have known, and which relies mostly on the target and neglects all the predictors that are needed is successful in this case. I felt surprised and interested to know how this worked. I was actually expecting Logistics model to predict more accurately. I feel that It also depends on the data that we are testing on because even one small and tiny update in the data could change the results for each model and its conclusion. For example, if I added an additional passenger in the data that did survive, then it could conceivably make the J48 have the highest amount of correctly classified instances. Overall, I was very surprised by the results and it all depends on the data that is imported into Weka.  
Conclusively, I feel that there is no high predictable power in Zero R classifier, however it can be used for these performance based models and not really for accuracy too. Model evaluation also surprised when I was looking at the Confusion matrix. However, this was fun and interesting to know.