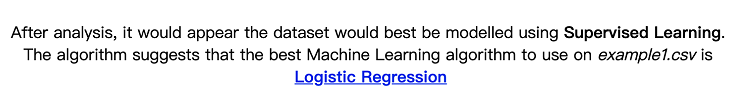
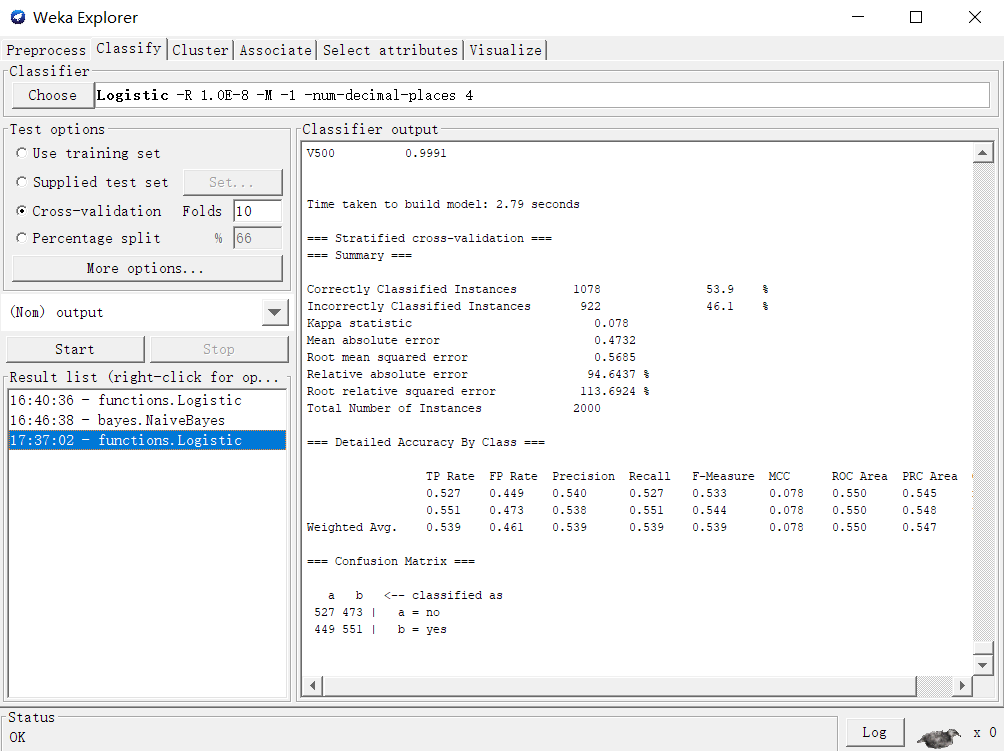
**Further Development of Machine Learning**

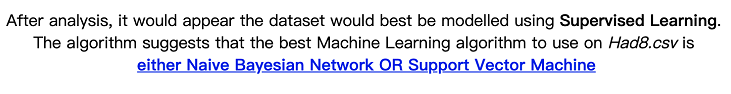
The group just using the decision tree to select the best machine learning tool from these 17 tools, the real implement of using the specific machine learning tool to analyse is not a function in our system. The main reason of this is limitation of the time. If we have more time, we might include a function in our system which could actually parse the dataset through the suggested machine learning tool. Because there is no display of this kind of function in our current system, we choose to use two machine learning analysing applications H2o and Weka to display how the machine learning tool works on dataset after given a specific suggested method.

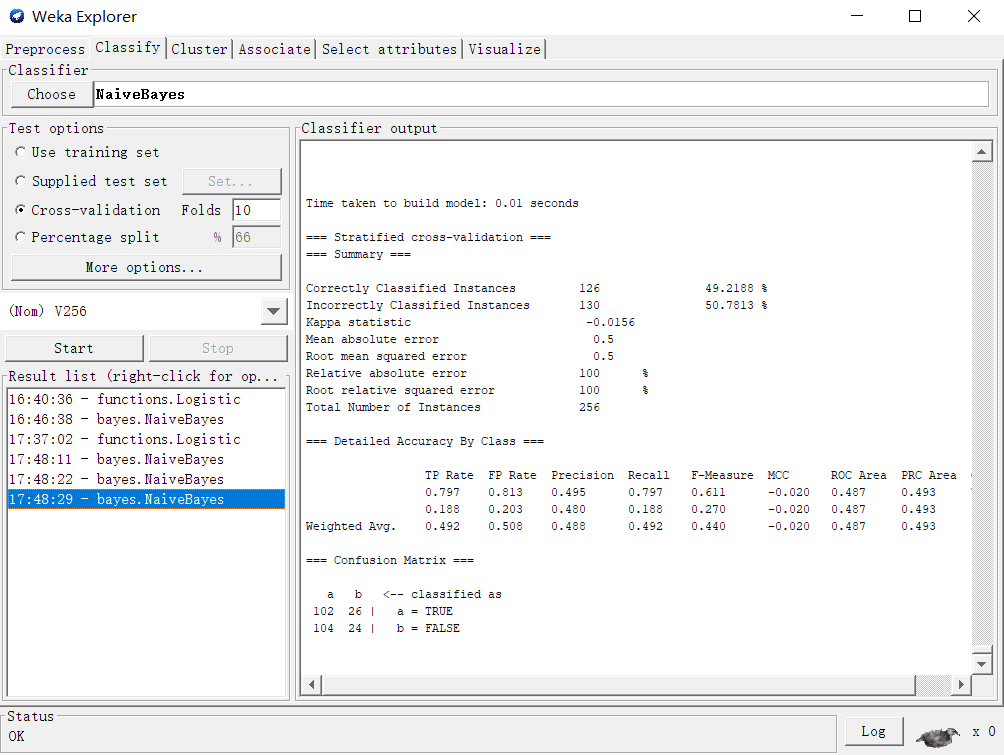
Choose three different datasets to analyse: example1.CSV, Had8.CSV and Pyramid.CSV. After decision analysing for example1.CSV, the result is:

The system suggests use Logistic Regression to analyze this CSV dataset. If our system contained a function which could analyze dataset by using specific machine learning tool, the result might be similar to the result giving by Weka:



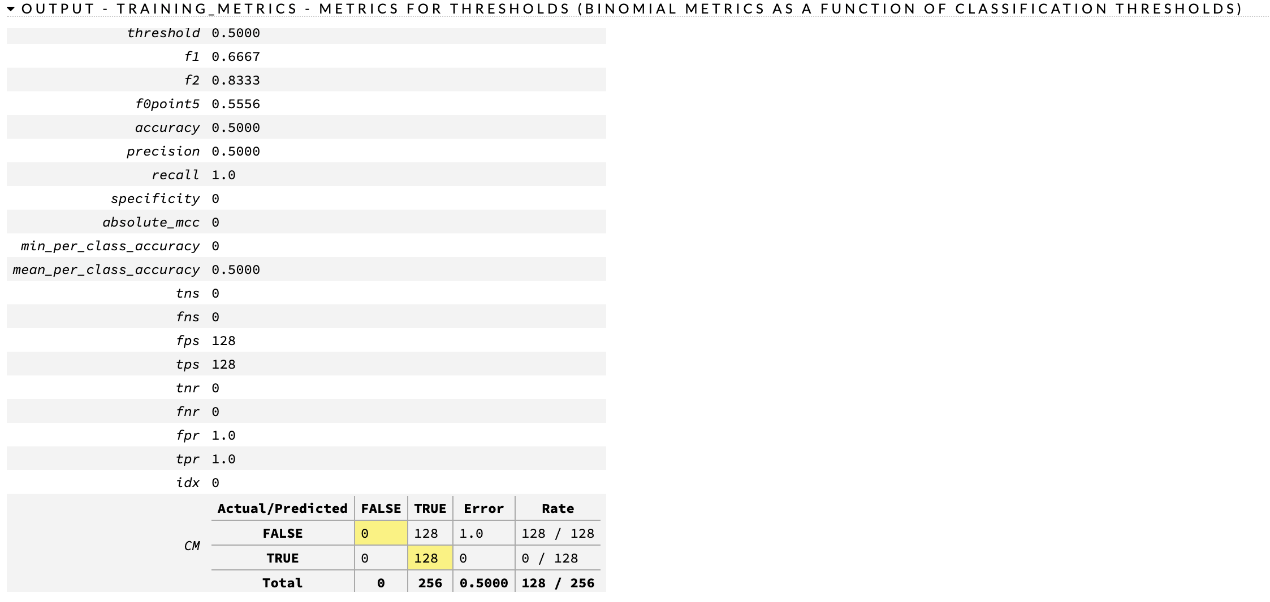
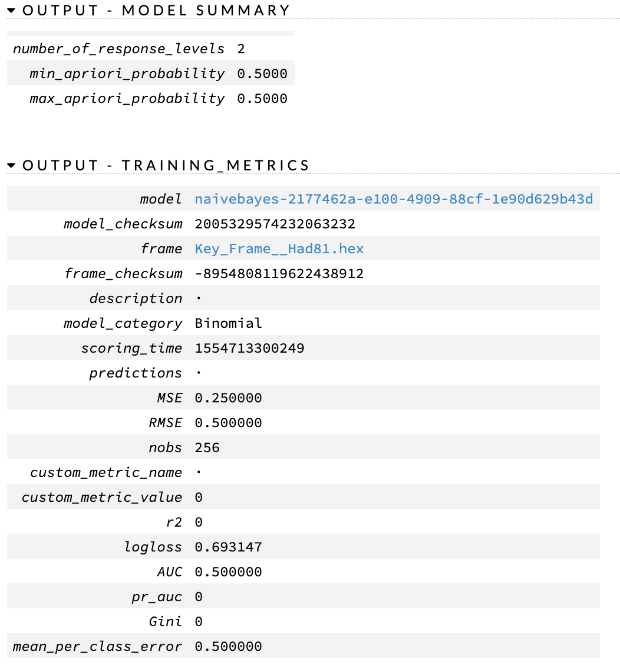
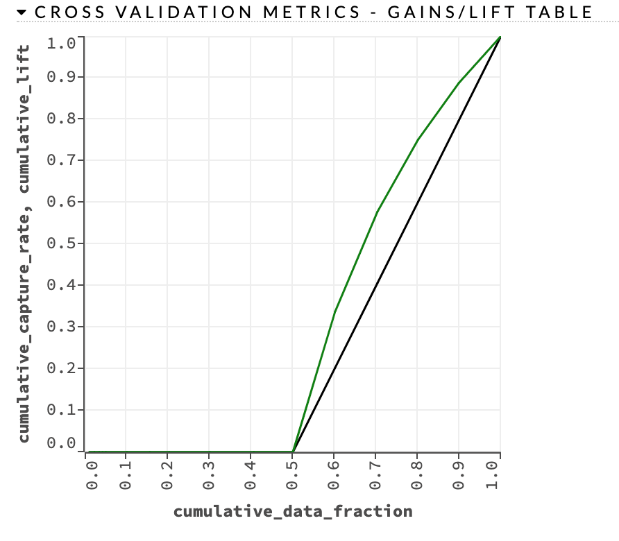
According to the result given by Weka, several analyze results we think the user might willing to know and we might add these information if we analyze the dataset in our own system, for instance: Correctly Classified Instances, Incorrectly Classified Instances, Kappa statistic, Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error, Total Number of Instances, etc.

After decision analysing for Had8.CSV, the result is:

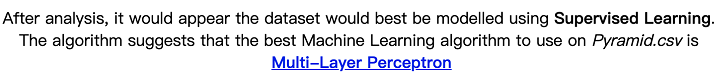
The system suggests either use Naïve Bayesian Network or Support Vector Machine to analyze this CSV dataset. If our system contained a function which could analyze dataset by using specific machine learning tool, the result might be similar to the result giving by Weka:

The information we might use is same as the analyze result of the last CSV file(example1.CSV).

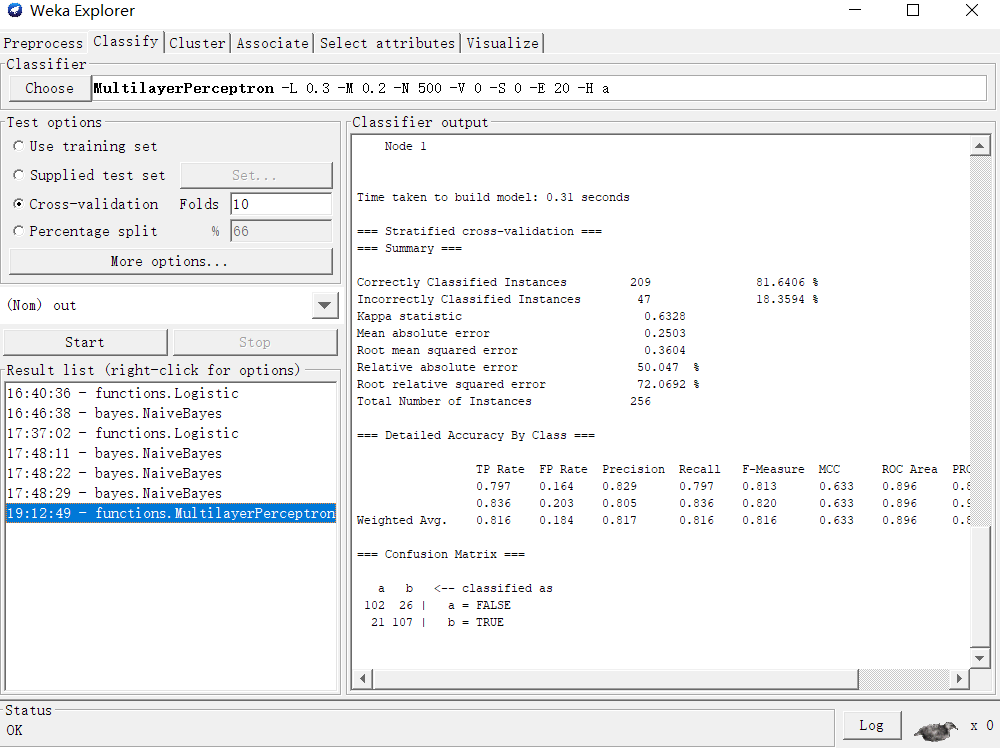
The analyze result is more abundant if using H2o to analyze:



According to the result given by H2o, several analyze results we might hold if we analyze the dataset in our own system, for instance: The Cross Validation Metrics, the Threshold, metrics for thresholds, min and max probability, etc.

After decision analysing for pyramid.CSV, the result is:

The system suggests Multi-Layer Perceptron to analyze this CSV dataset. If our system contained a function which could analyze dataset by using specific machine learning tool, the result might be similar to the result giving by Weka:



The information we might use is same as the previous two analyze results provided by Weka(example1.CSV and had8.CSV).