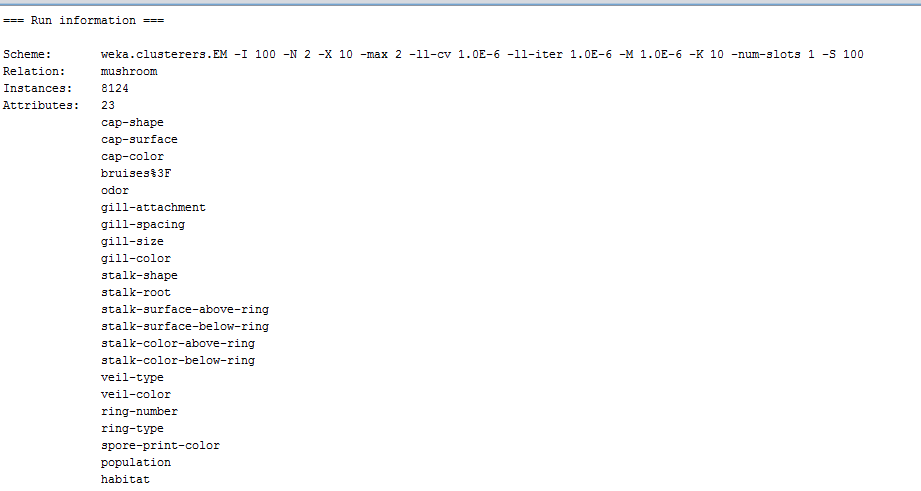
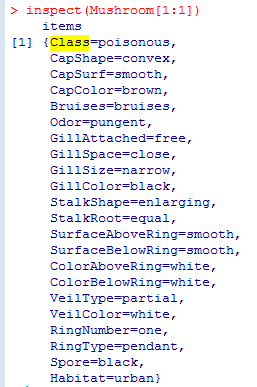
**Dimitrije Prosevski Project 3 04/13/2019**

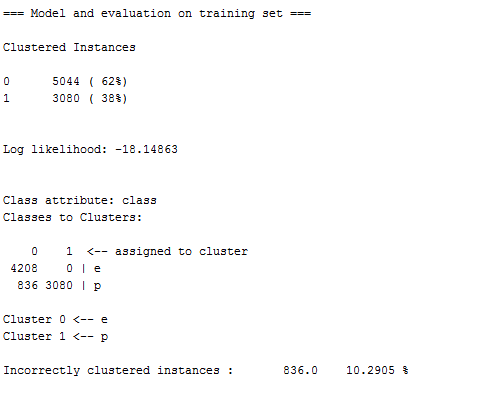
1. The data set is about mushrooms with 926136 (8124x114) entries in the item matrix with density of 0.193. Data set has the description of samples corresponding to 23 gilled mushrooms. The data contains 23 features and a class that specifies if its edible or poisonous.



1. The purpose of this clustering analysis is to analyze the physical metrics obtained and cluster the data around chosen parameters.



3. I used Weka for cluster analysis. For my dataset, EM clusterer turned out to be the most precise in clustering the class. The only parameter I changed from the default options was “numClusters” that I made 2; edible or poisonous.

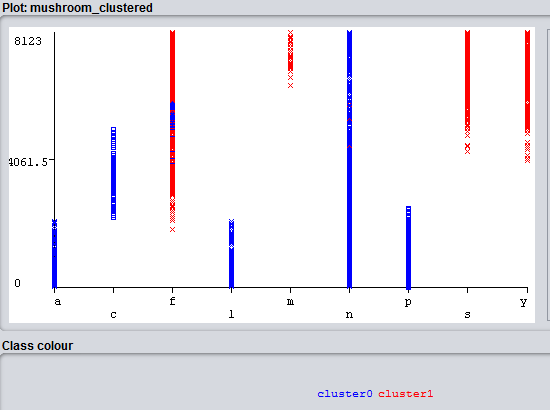


4. After performing clustering, I did “visualize clustering assignment” to visualize the data. From the previous lab I was familiar that high accuracy determining factor of whether mushroom was poisonous, or edible was the odor.

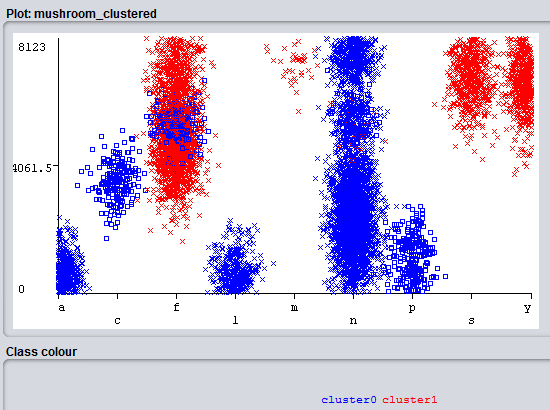


Visualizing number of instances and odor while clustering “cluster 0” as edible (blue) and “cluster 1” as poisonous (red).

Low jitter:



High jitter:



In conclusion, we can see that the edible mushrooms got clustered mostly around “n” which is odor = “none” as expected from previous research; if mushroom doesn’t have odor then mushroom is most likely edible.