In order to validate the three algorithms that we have designed in this project we have selected three different datasets, applied the clustering algorithms on them, and computed some metrics to see the results.

For the first 2 datasets we selected one with a low number of classes (*Adult.aff*, which has only 2 classes) and one with a high number of them (*Kropt.arff*, which has an amount of 18). Both have enough instances to say that the results are reliable.

To perform the clustering, we have selected as number of clusters the number of classes, so there can be computed external measurements. The objective of this was to compute the performance of K-Means and Bisecting K-Means in a case of high number of clusters and low number of them.

As external metrics we have selected:

* **Adjust Rand Index (ARI):**

Measures the similarity between data clusters and the real data classes. If its value is similar to 0 or negative means that the clusters and the classes are really different. The higher the value is, the more similar they are, until 1, which means that they are the same.

* **Accuracy:**

This metric is self-made. It measures the percentage of instances that have been assigned to the cluster that corresponds to its real class.

As a third dataset we have selected *Waveform.arff*. It is a dataset with only numerical data so we can compute internal metrics. For this dataset we have applied an extra metric:

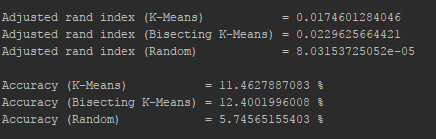
* **Silhouette Coefficient:**

It is a measure of how similar is an object to its cluster and different to the other clusters. It ranges from -1 to 1. The ideal is to have values similar to -1 or to 1. Values of 0 mean that they are as similar to its cluster as to the others.

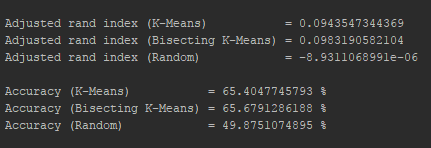
**Results**

In this section we have calculated the metrics explained before with the clusters obtained with K-Means algorithm and Bisecting K-Means algorithm. We have also calculated them for a random labeling so we could check that our algorithms were performing well.

Kropt Dataset (18 classes):



Adult Dataset (2 classes):

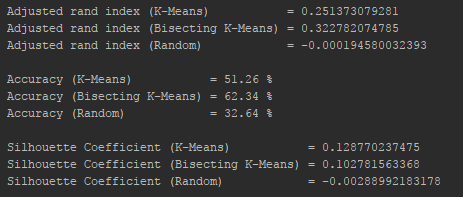


On the one hand, seeing that the ARI value is 3 orders of magnitude higher in both clustering algorithms than in random labeling we can conclude that our program is performing as expected when clustering the data.

On the other hand, from these two datasets we can conclude that, despite the fact that we thought that if the number of clusters was increased, the difference in performance between K-Means and Bisecting K-Means would also increase, the ARI values for both methods are similar in both datasets.

To conclude, it must be said that both algorithms have a similar performance and its difference is not enough big to say that one of them is better than the other.

Waveform Dataset (Only numbers):



From this last dataset we can extract a main conclusion. Both algorithms perform far better with just numerical attribute that with both numerical and nominal.

This may be due to the way we have chosen to compute the distances in nominal attributes, which is 0 if they are the same and 1 if they are different. In numerical attributes there is a full range of numbers from 0 to 1 that allows to make more specific clusters.