Enrique Ortiz
Data Mining
HW 2
November 1, 2021

1) In order to run my program in the terminal you enter this when you're in the same directory.

   javac GenesLocalization.java
   java GenesLocalization

2) I used rapid miner to prepare my data into a csv file. I used it to get rid of the commas in the attributes. This is so that in java I could split the row by commas.

3) I used the nearest neighbor algorithm to prediction the localization attribute. For each row that I needed to predict the localization attribute, I went through the training set and assigned weights to the localization values. If a row in the training set had only 2 matching attributes, then it would get a weight of 2. If it had only 1 in common, it would get a 1.

   I then stored the localization values with highest weight into a list. If the highest weight was 5 then that means only localization values that had rows in the training set that match 5 attributes with the row that I wanted to predict were in the list. After this I picked the value that was the most common in the list as the prediction.

4) Accuracy: 59.668%