**Description of the Dataset Preparation:**

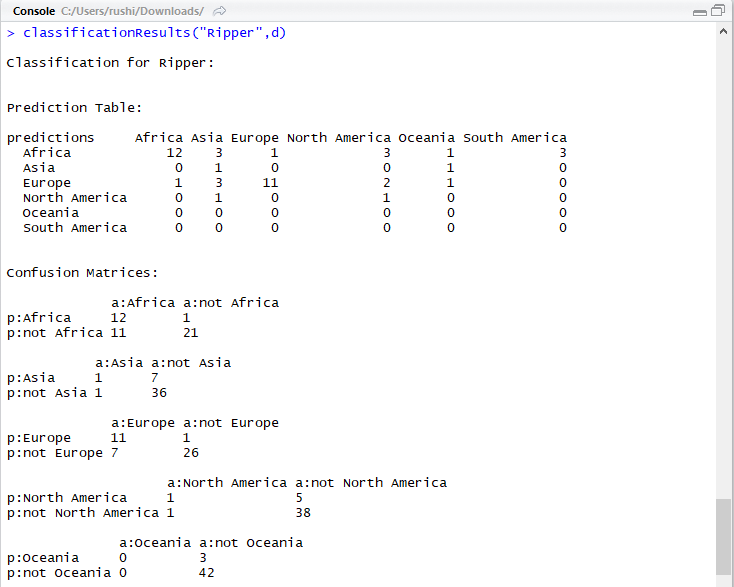
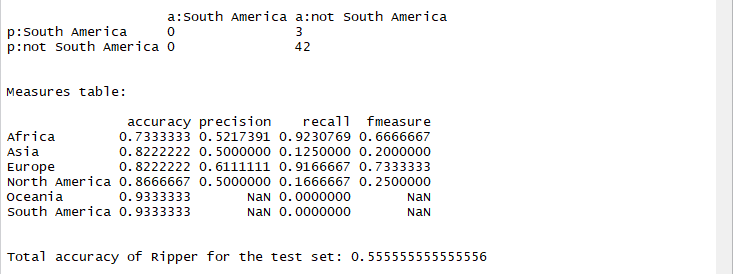
1. The data is saved in a .csv file
2. The data from the .csv file is read into a dataset.
3. This dataset is divided into 2 parts: Training and Test in the ration 80:20 respectively using the method divideData().
4. The Training data is used for training and the Test data is used for Testing.

**Description of the classification methods:**

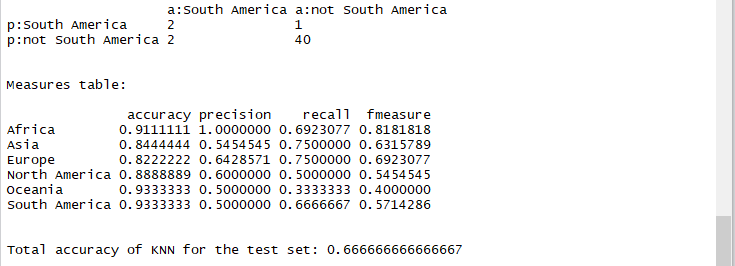
1. Support Vector Machines(SVM): Using Kernel type linear. Polynomial, Radial Basis or Sigmoid can be used as well.
2. K Nearest Neighbor(KNN): K = ceiling(sqrt(nrow(d))) since square root of the total number of rows is a good number to have for k.
3. C45: Just passed the dataset to the method J48() and it gives the result set.
4. Ripper: Method JRip() is used.

**Classification Results and Analysis:**

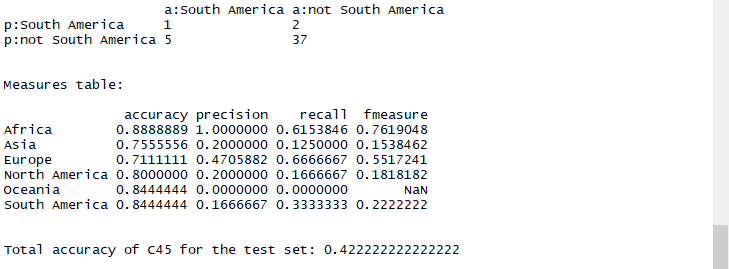
**Ripper:**

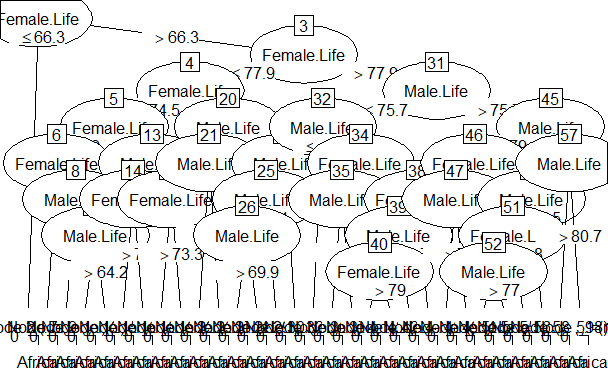
 

**KNN:**

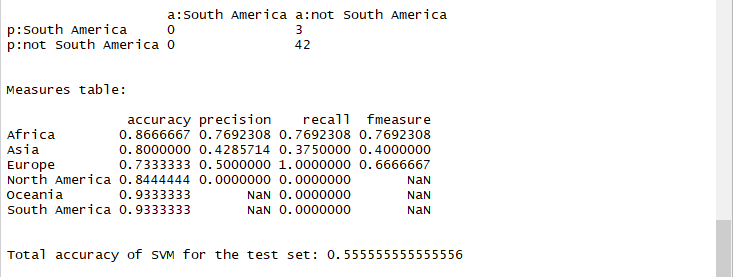
 

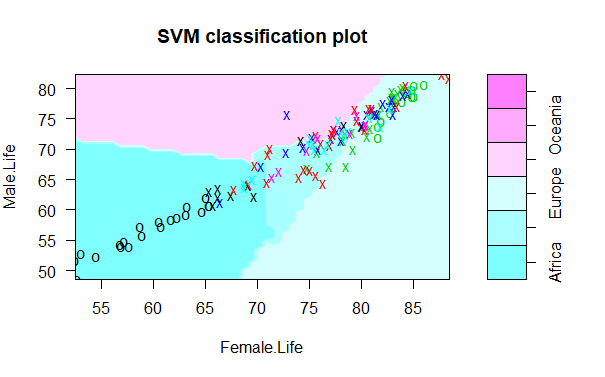
C4.5:

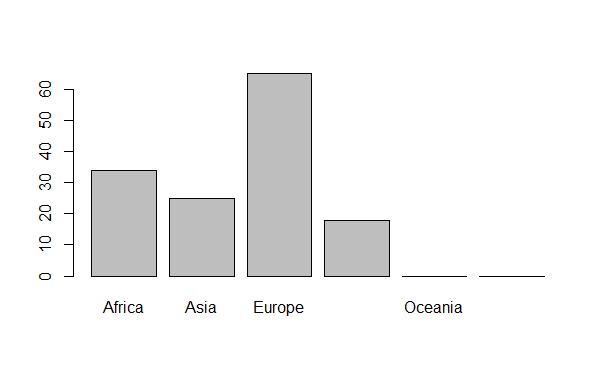
 

****

**Support Vector Machines:**

****

****

**Conclusion:** For this particular dataset, for this particular ordering of the data, the KNN has the highest accuracy, whereas the C$.5 has the lease accuracy. But none of the algorithms have very high accuracy. Also the accuracy changes as we change the ordering of the data. Also the accuracy will be different for different data. The more the training data, the more the accuracy.

**References:**

1. https://www.youtube.com/watch?v=ueKqDlMxueE
2. https://www.youtube.com/watch?v=pS5gXENd3a4
3. https://www.youtube.com/watch?v=GtgJEVxl7DY
4. https://www.youtube.com/watch?v=DkLNb0CXw84
5. https://cran.r-project.org/web/packages/RWeka/RWeka.pdf
6. http://weka.sourceforge.net/doc.stable/weka/classifiers/rules/JRip.html
7. http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html