Classifier Weka:

Zunächst wurden auf die ARFF-Dateien mit den gelabelten Tweets der Filter StringToWordVector (unsupervised attribute Filter) angewendet. Als nächste musste der richtig Classifier gefunden und angewendet werden. Da wird rassistische Sprache in Tweets analysieren und ein Tweet als entweder rassistisch (1) oder nicht rassistische (0) gewertet werden kann, haben wir uns für den NaiveBayesUpdateable Classifier entschieden. Als entscheidendes Kriterium für die Nutzung des ausgewählten Classifiers war zum eine die Einsatzmöglichkeit des Classifiers für Binäre-Klassen und Binäre-Werte und zum anderen die hohe Prozentzahl für die Korrektheit der klassifizierten Instanzen. Mit 70,566% übertraf der NaiveBayesUpdateable Classifier andere NaiveBayes Classifier, sowie andere Classifier wie RandomTree oder ZeroR.