The AKI data contains the information minimum, maximum, average and standard

Deviation of majority of physiological variables that are recorded in the first three days in

ICU along with demographic factors (sex & age) We obtained 4497 records from the data set after data aggregation and applying filters. Out of the total records, 3365 records were picked as the training dataset and 1132 records as the testing dataset. The data contains demographic variables and statistical features extracted from the physiological variables.

Generate descriptor for the dataset using the command:

bin/hadoop jar $MAHOUT\_HOME/mahout-core-0.6-job.jar

org.apache.mahout.classifier.df.tools.Describe -p AKI/AKI\_Train.arff -f AKI/Desc.info

-d 1C 33 N L

Generate the model using command:

bin/hadoop jar $MAHOUT\_HOME/mahout-examples-0.6-job.jar

org.apache.mahout.classifier.df.mapreduce.BuildForest -

Dmapred.max.split.size=1874231 -d AKI/AKI\_Train.arff -ds AKI/Desc.info -t 1 -o

AKI/AKI-forest

Classify the test data using command:

bin/hadoop jar $MAHOUT\_HOME/mahout-examples-0.6-job.jar

org.apache.mahout.classifier.df.mapreduce.TestForest -i AKI/AKI\_Test.arff -ds

AKI/Desc.info -m AKI/AKI-forest -a -mr -o AKI/Output