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.The data contains demographic variables and statistical features extracted from the physiological variables.

Training the classifier parameters (or classifier model) involves passing the training

Data with ground-truth label information to the supervised classifier learning module.

This is achieved by executing the following command:

$hadoop jar /usr/lib/hadoop/Test.jar com.winvector.logistic.demo.MapReduceLogisticTrain

input/AKI\_Train.tsv "HOSPITAL\_EXPIRE\_FLG ~ SEX + AGE + AVG\_GCS +

MIN\_GCS + MAX\_GCS + STDDEV\_GCS + AVG\_HEMATOCRIT +

MIN\_HEMATOCRIT + MAX\_HEMATOCRIT + STDDEV\_HEMATOCRIT +

AVG\_HEART\_RATE + MIN\_HEART\_RAT + MAX\_HEART\_RAT +

STDDEV\_HEART\_RAT + AVG\_SYSTOLIC\_BLOOD\_PRESSURE +

MIN\_SYSTOLIC\_BLOOD\_PRESSURE + MAX\_SYSTOLIC\_BLOOD\_PRESSURE +

STDDEV\_SYSTOLIC\_BLOOD\_PRESSURE + AVG\_RESPIRATORY\_RATE +

MIN\_RESPIRATORY\_RATE + MAX\_RESPIRATORY\_RATE +

STDDEV\_RESPIRATORY\_RATE + AVG\_URINE\_OUTPUT + MIN\_URINE\_OUTPUT +

MAX\_URINE\_OUTPUT + STDDEV\_URINE\_OUTPUT +

AVG\_WHITE\_BLOOD\_CELL\_COUNT + MIN\_WHITE\_BLOOD\_CELL\_COUNT +

MAX\_WHITE\_BLOOD\_CELL\_COUNT + STDDEV\_WHITE\_BLOOD\_CELL\_COUNT +

AVG\_BUN + MIN\_BUN + MAX\_BUN + STDDEV\_BUN" AKI\_model.ser

Where the HOSPITAL\_EXPIRE\_FLG is the predictive field and others are features.

AKI\_model.ser is the trained classifier model.

Validation of Classification With Test Data

hadoop jar /usr/lib/hadoop/Test.jar com.winvector.logistic.demo.MapReduceScore

AKI\_model.ser input/AKI\_Test.tsv AKIScore