Lab 4 – Report

1. Cloudera/Mahout/Solr

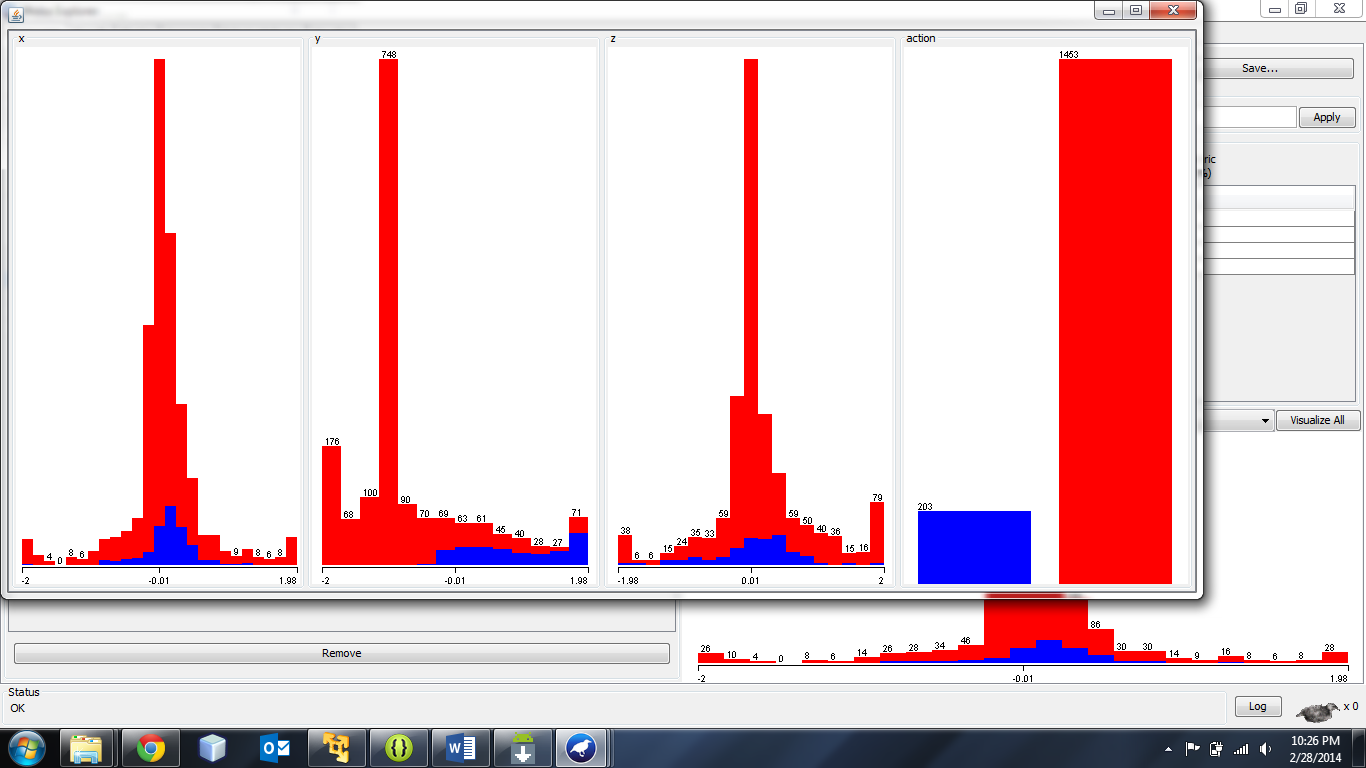
Implement a Mahout application (e.g., Classification/Recommendation/Association Rule Mining) using your own data and store your output to Solr

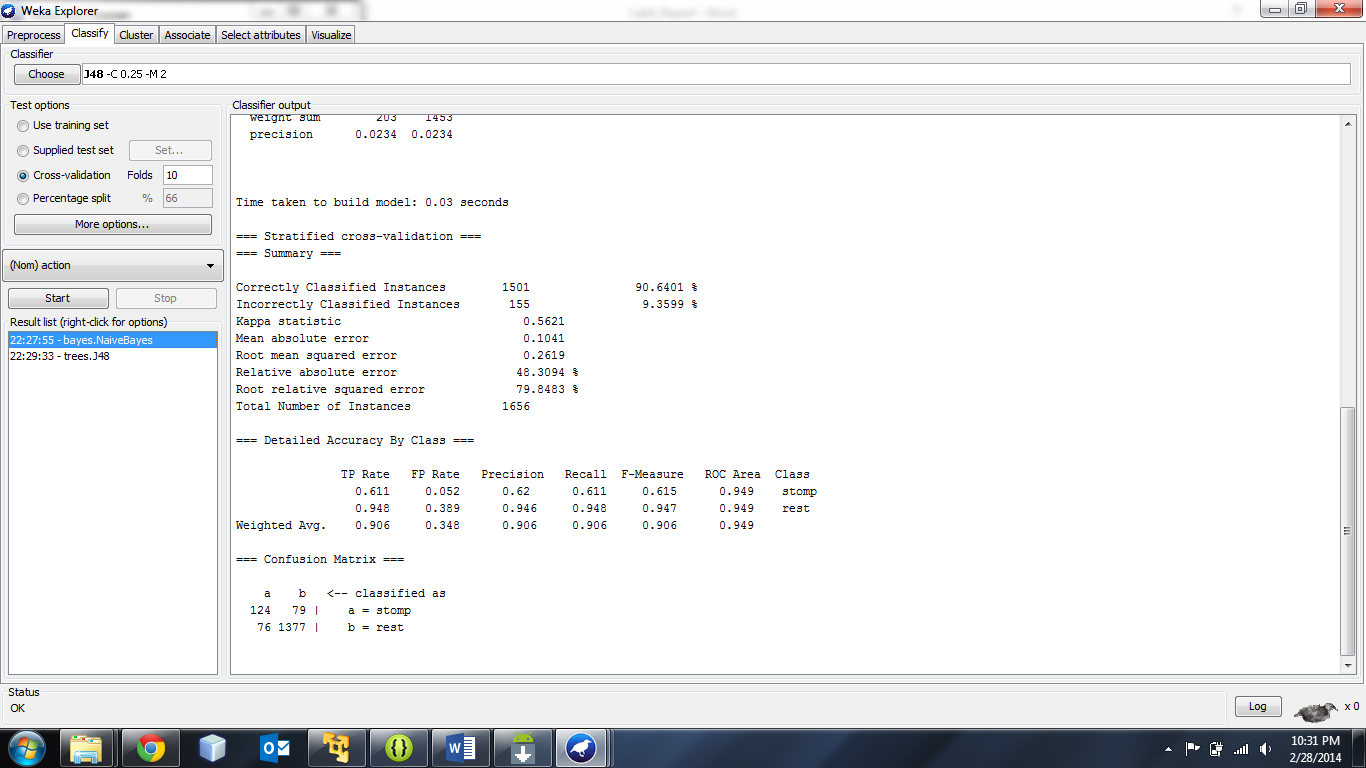
Implemented a naïve Bayes classifier in mahout using sensor data (own data). The steps followed in achieving the classifier.

1. Simulate different classification techniques to check which one will best suit the features and data format. Used weka tool to visualize data and determine different classification confusion matrix.
2. Generated required data in arff format.

Naïve Bayes – 90.6% classified instances on training data

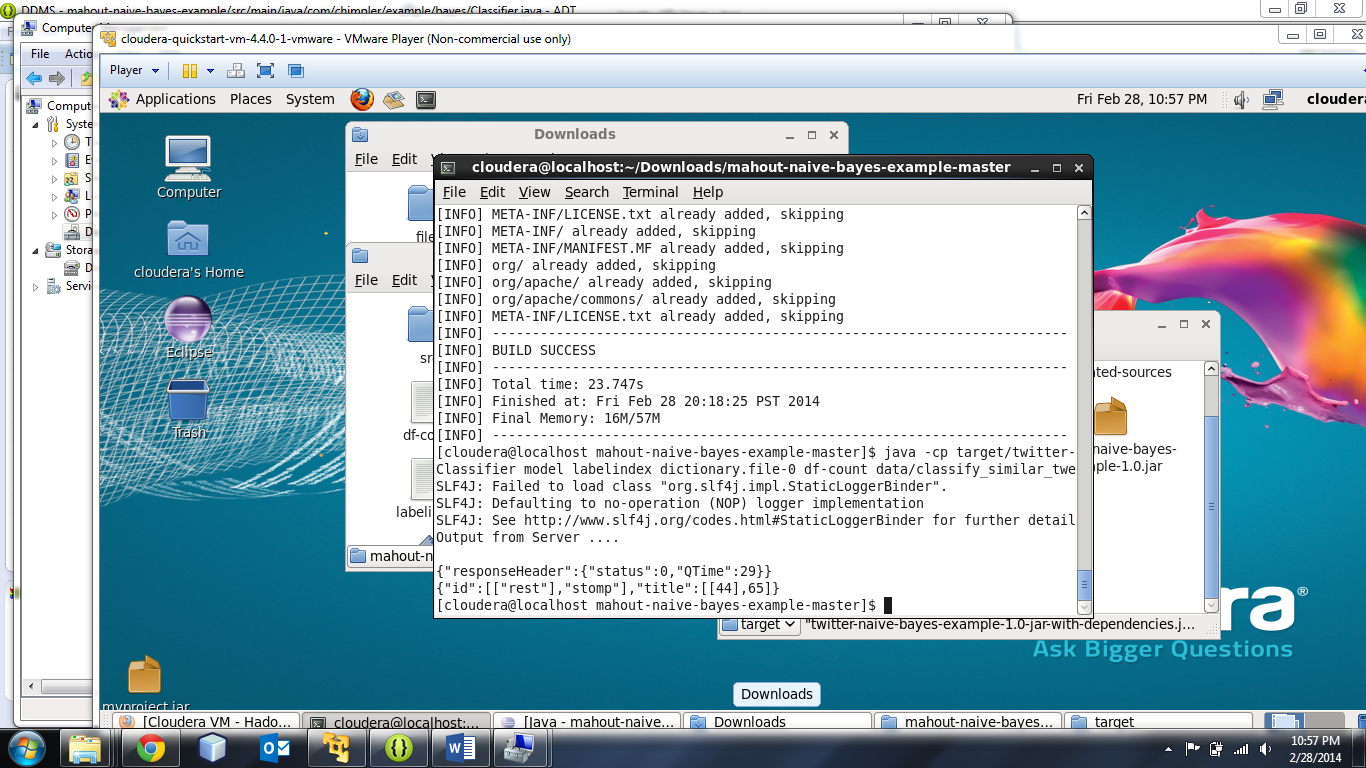
J48 – 89.9% classified instances on training data





1. With the above results, implemented naïve Bayes in Hadoop environment using mahout. Took raw data from the sensor and prepared training data. Convert this training data to seq files using mahout inbuilt command (SEQDIRECTORY) , generate vectors using seq files (SEQ2SPARSE). Split data into train-vectors and test-vectors.
2. With the generated vectors, try to get weighted vectors by training the vectors using train-vectors data. Try to test naïve Bayes using the generated weighted vectors.
3. We get the classifier files, dictionary, df-count, model and labelindex files. Transfer these files to local from Hadoop and run the classifier on new unclassified data using the classifier class.
4. Change the classifier java class to fit your requirement, we changed the classifier class to count the number of stomps and rest items and save it in the form of JSON.
5. Push this JSON object to SOLR using a restful web service, using a unique ID. We reused the web service in tutorial 7 in class.
6. **Classifier run will classify new data and result is saved to Solr.**

Result is show in below image, where navie Bayes classifier has run and classified the given new data and generated a json file and save to solr.



2. **Mobile Application**: Implement a mobile application to use the information through Solr REST Web Services

Developed an add-on to the existing application (lab3), the module will report the count of stomps and rest till date (latest log file). We present to user a button to view the report (workout results) of his daily exercise.

Below is the screen for final screen of the app in lab3, where the result is shown to user.

If the user want to see the result, he will select report button which will retrieve results from SOLR and display to the user.



