CSE 6331 Cloud Computing Programming
Assignment 4: Data Analytics on Cloud
Due: November 9, 2014, 23:00 (UTA time)
Submitted By: Sarvesh Sadhoo (1000980763)

Project Description:

1. Classification was done on the Titanic Data Set to determine if a person would die or not. A Navies Bayes Classifier was used determine this using python Orange machine learning library.
2. Titanic training data: https://www.kaggle.com/c/titanic-gettingStarted/download/train.csv
3. Titanic test data: https://www.kaggle.com/c/titanic-gettingStarted/download/test.csv
4. The training data and the test data was stored and retrieved from AWS DynamoDB
5. Then the training data was feed into the classifier to learn from it. Below screen shot shows the classifier being built.

```
/System/Library/Frameworks/Python.framework/Versions/2.7/bin/python2.7 "/Users/srv/Desktop/Code/Cloud Computing/Project 4/NaiveBayes.py"
                          No       Yes
class probabilities       0.616    0.384

Attribute pClass
                          No       Yes
                  1       0.370    0.630
                  2       0.527    0.473
                  3       0.757    0.243

Attribute Sex
                          No       Yes
              female      0.258    0.742
                male      0.811    0.189

Attribute Embarked
                          No       Yes
                  C       0.444    0.556
                  Q       0.597    0.403
                  S       0.663    0.337

Attribute Parch
                          No       Yes
                  0       0.656    0.344
                  1       0.449    0.551
                  2       0.500    0.500
                  3       0.400    0.600
                  4       1.000    0.000
                  5       0.800    0.200
                  6       1.000    0.000
                  9       0.500    0.500
```

6. The final result i.e no of survival yes or no was used develop a pie chart.
   URL: http://omega.uta.edu/~sxs0763/piechart.html

References:

- http://orange.biolab.si/
- http://docs.orange.biolab.si/tutorial/rst/classification.html
- http://docs.orange.biolab.si/tutorial/rst/data.html
- http://boto.readthedocs.org/en/latest/dynamodb_tut.html
- http://www.highcharts.com/demo/pie-basic
- http://jsfiddle.net/gh/get/jquery/1.9.1/highslide-software/highcharts.com/tree/master/samples/highcharts/demo/pie-basic/