MS -Business Intelligence & Analytics

Fall 2015

*March 25, 2016*

**Vivek-John Dominic Martins**

Home Work 2

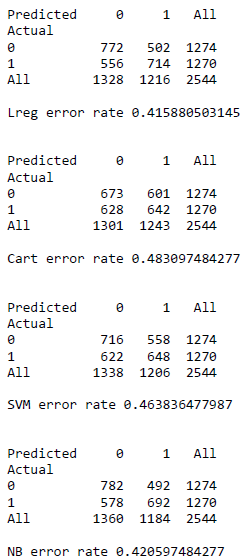
**Ethics Statement**

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

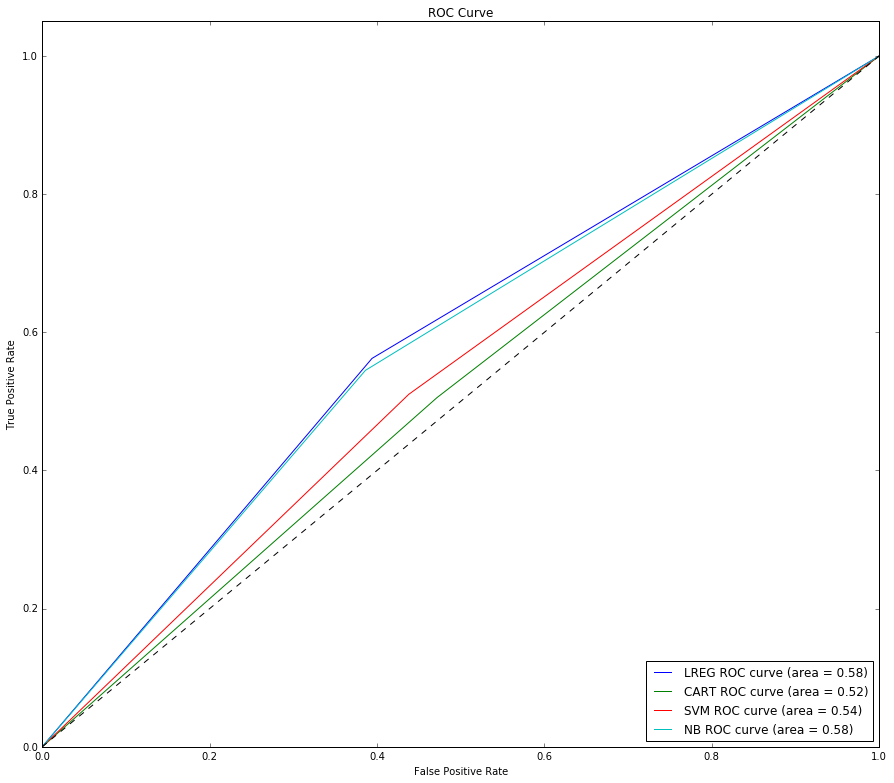
Signature \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_

**Predictive modeling: targeting offers**

1. ***Compare the different models explored using the test error rate (percent incorrectly classified), the area under the ROC curve and the confusion matrix against the benchmark (logistic regression).***



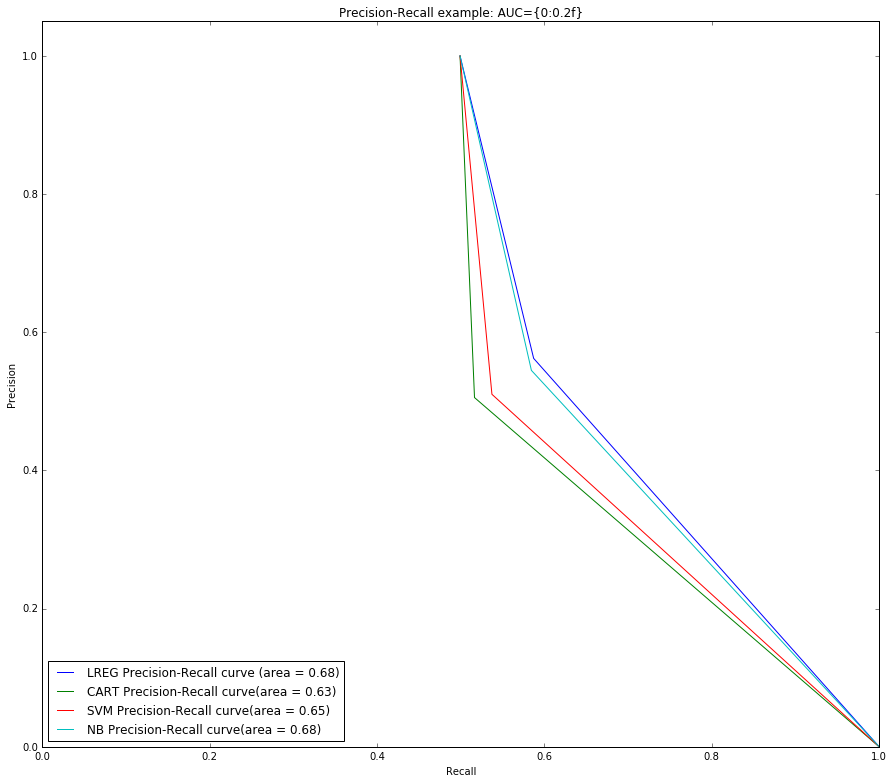
It is visible from the data above that the error rate for CART > Support Vector Model >Naïve Bayes >Logistic Regression Model.



This is again represented in the ROC curve. The area under the curve is very similar for both the Naïve Bayes model and the Logistic Regression Model. However the Area under the curve for CART < Support Vector Model <Naïve Bayes <Logistic Regression Model.

Therefore both the Area under the curve and error rate point to Naïve Bayes as the better predictor model with Logistic Regression as the benchmark. (Please note that Logistic Regression outperforms the other models)

1. ***Use matplotlib to plot the ROC and the precision-recall curves for your models. Discuss and compare the performance of each model according to these curves against the benchmark (logistic regression).***



Since we are interested in whether the customer subscribes to the magazine a true positive is more important that a false negative. Therefore a higher area under the Precision-Recall curve is better. In this case Naïve Bayes is the best, having Logistic Regression as a Benchmark. (Please note that Logistic Regression outperforms the other models)