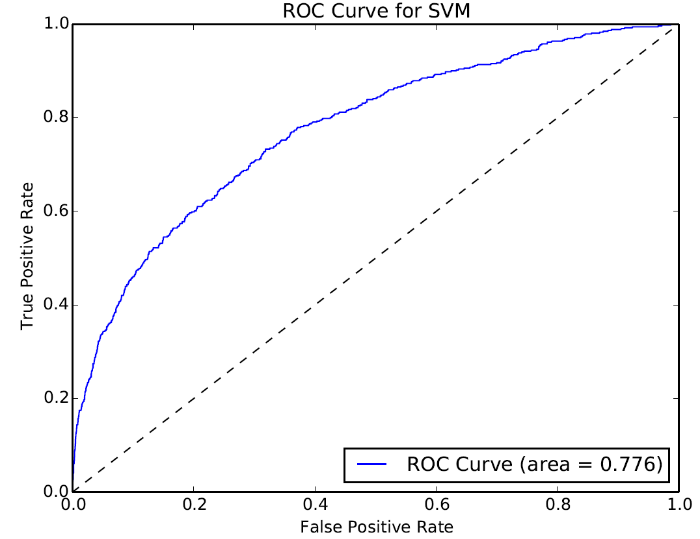
**Introduction**

The purpose of this assignment is to make a recommendation to bank management whether to use a machine learning model to offer loans to specific customers through direct marketing. A Support Vector Machine (SVM) model was developed using 16 variables from the “Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology” study. This paper recommends using the model and highlights the risks and costs associated with its use.

**Results**

The area under the receiver operating characteristic (ROC) curve characterizes the classification performance and the 0.776 index provides lift over a 0.5 random chance. Given that there is a relatively low chance that people will accept the loan offer, model accuracy even at x is not a viable measure for evaluating its performance.

The confusion matrix is an effective method for evaluating the costs and risks of using this model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Actual Binary Response | |  |
|  |  | YES | NO |  |
| Predicted Binary Response | YES | **True Positive**  **x** | **False Positive**  **X** |  |
| NO | **False Negative**  **x** | **True Negative**  **X** |  |

The false negative represents the consumers that were predicted to not accept the offer but actually did. This represents lost revenue …. The false positive represents consumers that were predicted to accept the offer and actually did not. This represents costs that are incurred without conversion.

There are additional metrics that provide insight into model performance:

* **Precision** ( true positives / (true positives + false positives) ) - represents the % of labeled class that actually the class
* **Recall** (true positives / (true positives + false negatives)) - this represents the % of the actual class we are pulling out of the sample
* **f1** (precision \* recall / (precision + recall)) – represents a measure that combine the precision and recall elments

**Code**

**References**