

Management Report Two

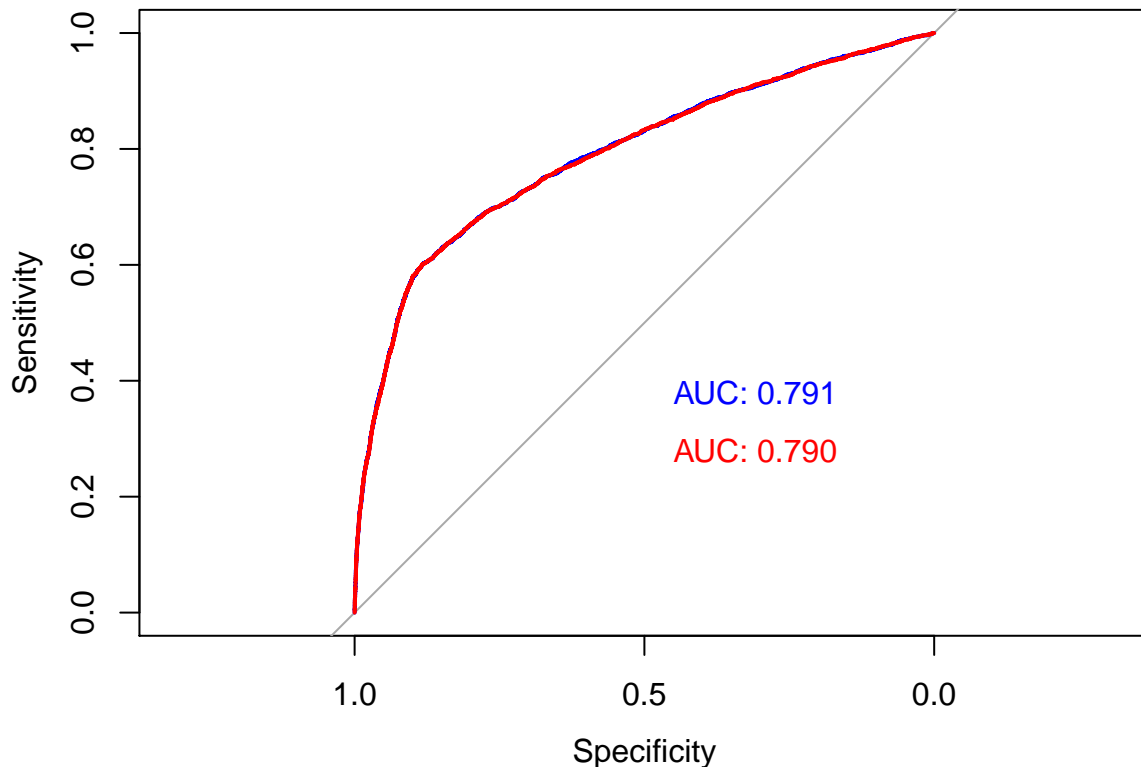
Muiz Murad

28 October 2019

Introduction

This report will focus on making a model, using a logistic regression to predict whether or not the telemarketing company's customers will default or not on a loan. Regression models were used to analyse the data and develop models that will be used on both train and test data.

Split the data into training and test sets to ensure that we are able to have data to test the model with as the training data will also produce high accuracies as it has been used to create the model already. The test data acts as foreign data on the model to produce accuracies.



ROC Curve

The ROC Graph measures how well the model can perform in terms of specificity and sensitivity. Sensitivity measures how well the model can correctly predict that a person will default whilst specificity is a measure on how well the model can correctly predict that someone won't default on their loan. In order for this to function they need to have an equal balance as if one of the metrics is out of proportion, it will weigh out the other and reduce it. The best models are ones which are the furthest away from the linear line in curvature as shown above, the model could be better as well. The AUC (Area Under The Curve), is a measure of accuracy and as shown above it is very similar for both the models.

Training Data Accuracy

```
##
## Attaching package: 'MLmetrics'

## The following objects are masked from 'package:caret':
##
##      MAE, RMSE

## The following object is masked from 'package:base':
##
##      Recall

## [1] 0.9002081

##      y_pred
## y_true    0      1
##    no 25246  363
##    yes 2514  707
```

Testing Data Accuracy

```
## [1] 0.9002081

##      y_pred
## y_true    0      1
##    no 25246  363
##    yes 2514  707
```

The training and testing model the models were slightly better than the base line accuracy of 88% with accuracies of 90% respectively. Although it isn't the best model or uses variables that are necessary for the firm to be using. The decision tree had a variables in there that were of lesser use to the firm when first touchbasing with their client about a loan. The model is more useful than the previous one but as accurate. Both the training and testing models are similar with accuracies of 90% and an AUC of 0.78, which indicates a very robust model.

```
##      JOB DEFAULTCREDIT CONTACT MONTH DAY_OF_WEEK CAMPAIGN PDAYS
## 1  services          no  cellular   may          tue          5    10
## 2  unemployed        no  telephone oct          wed          5   999
## 3  blue-collar       yes  telephone aug          mon          5     7
##      POUTCOME EMP_VAR_RATE CONS_PRICE_IDX CONS_CONF_IDX NR_EMPLOYED
## 1  failure      -3.6        98.9444      -44.0        5200
## 2  nonexistent   -2.0        95.4230      -47.5        5000
## 3  success       2.3        92.1020      -33.9        6100
##      EDUCATION  Y
## 1      basic.4y no
## 2      basic.9y yes
## 3 university.degree yes
```

New Customer Predictions

As it can be seen from above new customer data was provided to the model which generate the following probability of defaults for Customer(s) 1,2 and 3 respectively:

```
##      1      2      3
## 0.999999886 0.930771518 0.001105359
```

Model Conclusion

The model made which will be presented to the management team, using the GLM package using the most significant variables from the dataset to predict the likelihood of a customer defaulting on their loan. The variables picked are the most significant in that they are able to predict the chance of default being either yes or no the best on unseen data. This will allow the team to make predictions best suited for the purposes of the firm.

The variables used to make this decision were:

-Job -Education -Default Credit -Contact -Month -Day Of The Week -Campaign -PDAYS -POutcome -Emp Var Rate -Cons Price IDX -Cons Conf IDX -NR Employed

Model Accuracy

The best model resulted in 90% overall accuracy. This is a high accuracy and it performed the same on the train and test data.

The model however is not the ideal model to be used and will be refined in the future and tweaked for accurate predictions but it is suited for the needs and purposes specified until further updates are developed.