MGMT 3

Muiz

25/11/2019

Introduction

This report focuses on making a k nearest neighbours model to predict whether the telemarketing's customers will default or not on a loan. KNN models were used to nalayse the given data and produce models that will work on both the split of train and test data sets.

Results

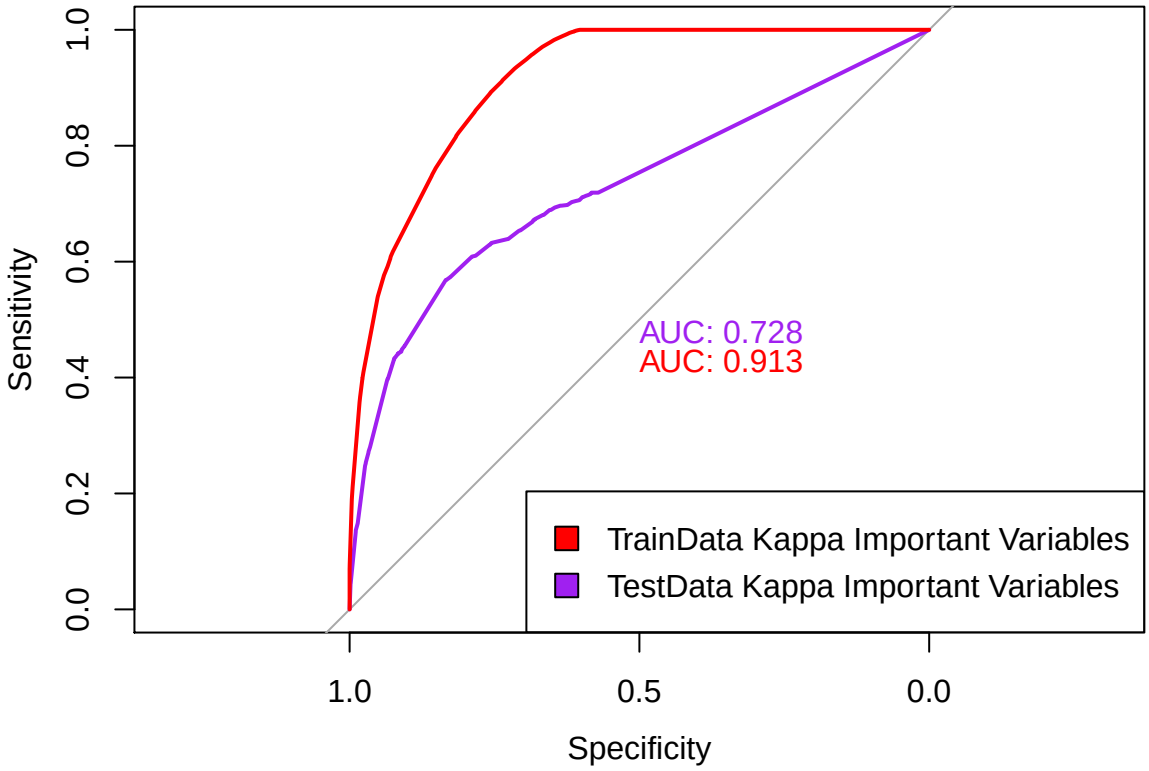
KNN Models

```
## Loading required package: DBI
```

```
         Actual
Predicted    no   yes
      no  28392  2288
      yes   579  1416
```

```
         Actual
Predicted   no  yes
      no  7349  691
      yes  228  245
```

Model predicts well. The train and test models have accuracies of 92% and 89% respectively. The small difference with these accuracies that the model isn't robust and isn't the best.



The ROC graph measures how well the model performs in terms of specificty and sensitivity. Sensitivity measures how well the model correctly predicts that a person will deafult whilst specificity measures how well the model correctly predicts that a person won't default. However, there needs to be a balance between the two as an over use of one would be too much and reduce the other throwing it into an imbalance. The best models are those that bend out the furthest from the linear line. As seen in the graph above, the model with the test data performs worse than the model with train data. The model isn't good as it isn't robust hence not performing similar on the test and tain data sets. The model isn't very good as both the test and train data sets have different AUC's with the Train Data having an AUC of 93.5% and the Test Data having an AUC of 70.1%. The AUC (Area Under The Curve) is a metric for accuracy. However, the model's accuracy is not the best on foreign data and predicts "no" and the model should not be used.

Predictions With Foreign Data

CAMPAIGN	PDAYS	AGE	LOAN	PREVIOUS	EURIBOR3M	Loan.Status
1	1000	42	yes	0	1.0	no
1	999	26	no	1	1.5	no
1	7	78	yes	0	0.9	no

The table above shows a few of the variables used to make predictions on a customer's loan default prediction. The model failed to predict someone defaulting even with variables of that someone who would default.

Conclusion

The model is good aas it uses important variables when predicting customer default however the overrall model doesn't perform well as its accuracy is less than the baseline accuracy. Most of the customers are predicted to not default and fewer were predicted to default. This is not good for the business as the

telemarketing company will be giving loans to too many people that will default on their loan. The model shouldn't be used and a more accurate model would be the decision tree model rather than the KNN model.