

FIT- 3152

(Data Analytics)

Nabil Ahmed

ID-25364170

Assignment 2

Lecturer: Dr. Sylvester Olubolu Orimaye

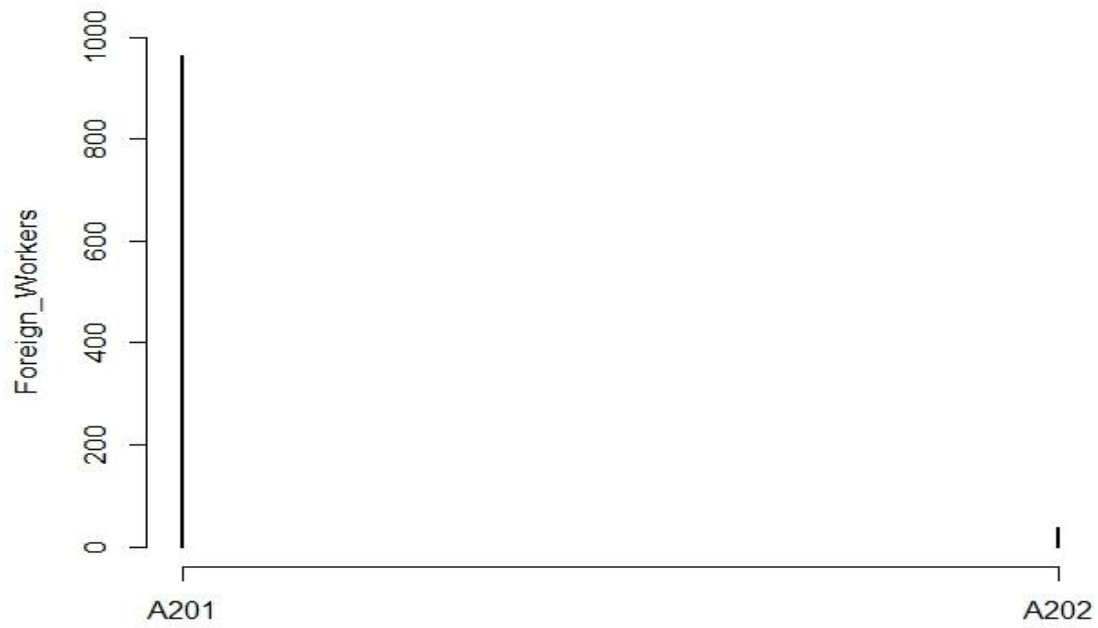
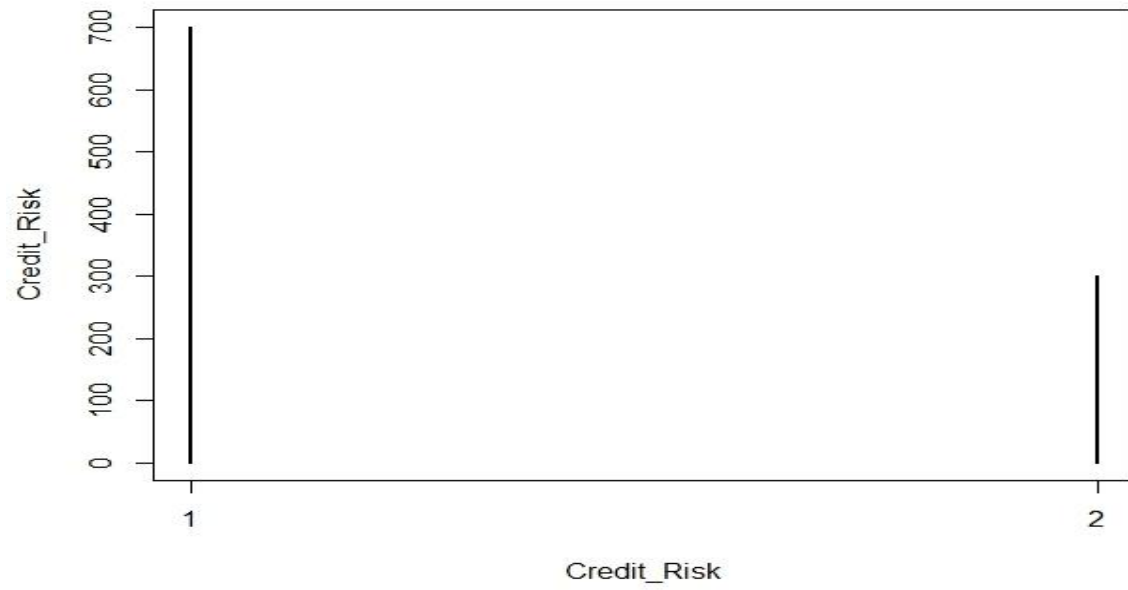
Abstract

The primary objective of this report is to gain familiarity with classification models and ROC charts using R. The analysis is done on the German Credit Dataset which contains information on 1000 customers, described by 20 decision attributes and a class attribute. Our job is to find a model that may be used to predict whether a new customer is at risk of defaulting a borrowed loan.

Description of Predictors

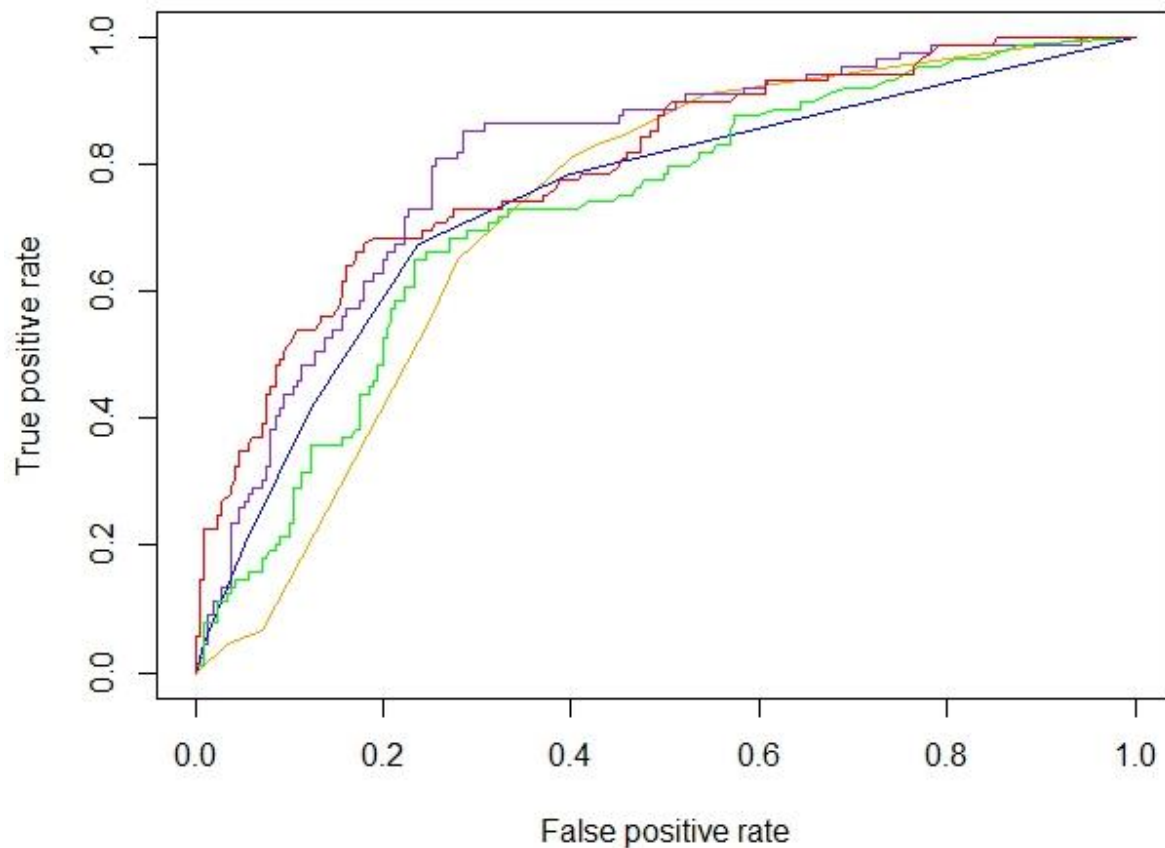
Numerical attributes	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
Duration	4.0	12.0	18.0	20.9	24.0	72.0	12.06
Credit.amount	250	1370	2320	3270	3970	18400	2823
Installment.rate	1.00	2.00	3.00	2.97	4.00	4.00	1.119
Residence. since	1.00	2.00	3.00	2.84	4.00	4.00	1.104
Age	19.0	27.0	33.0	35.5	42.0	75.0	11.38
Existing.credits	1.00	1.00	1.00	1.41	2.00	4.00	0.5777
People.liable	1.00	1.00	1.00	1.16	1.00	2.00	0.3621
Class.label	1.0	1.0	1.0	1.3	2.0	2.0	0.4585

The table above is self-explanatory. However, the proportion of good credit risk to bad credit risk is seen to be sufficiently higher which shows a staggering number of 700 to 300.



On the first plot, 1 is showing good credit risk and 2 is showing bad credit risk. On the second one we have the plot of foreign workers against local ones showing an astounding figure of 963 to 37. A201 represents foreign worker and A202 locals.

ROC curve



The figure above shows the ROC curve for each classifier.

Orange represents decision tree.

Blue violet represents naïve bayes.

Blue represents bagging.

Green represents boosting.

Red represents Random Forest.

At the beginning it is seen that random forest acts as the best classifier. Then naïve bayes takes its place. Then it becomes difficult to define a particular technique for the classification models as the best classifier. It seems like any of the classifier among decision tree, naïve bayes, boosting and random forest is acting as the best classifier.