PROJECT-4

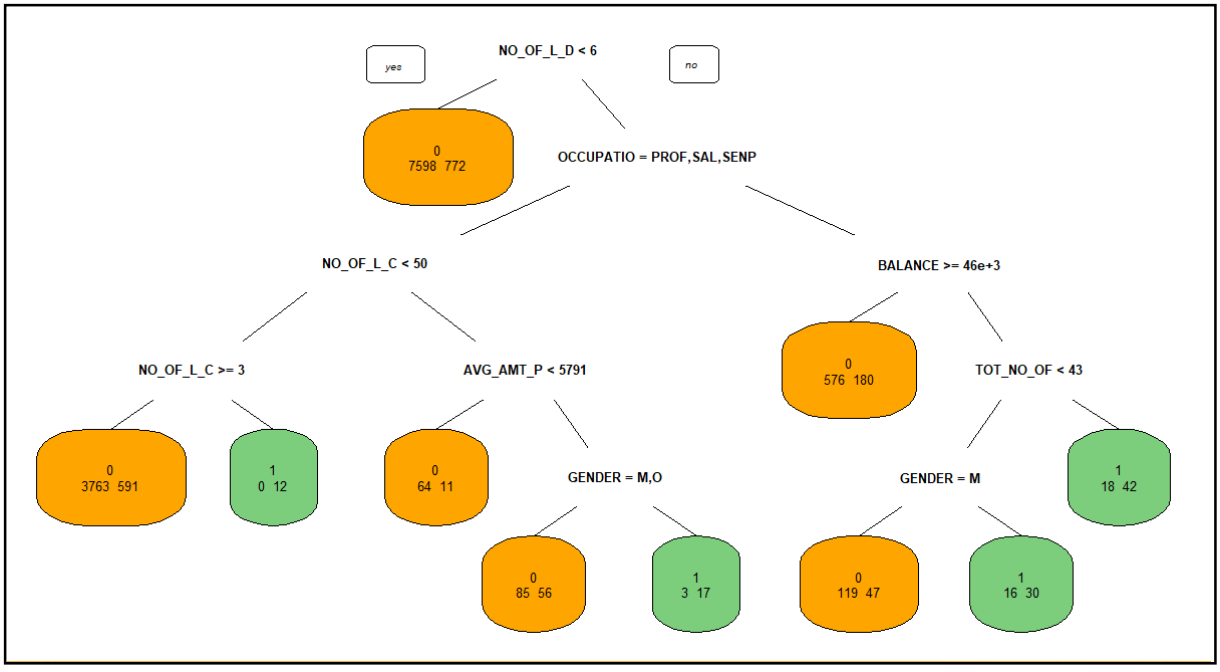
CART Model:

Probability distribution of Target 1 and 0 in the given data is

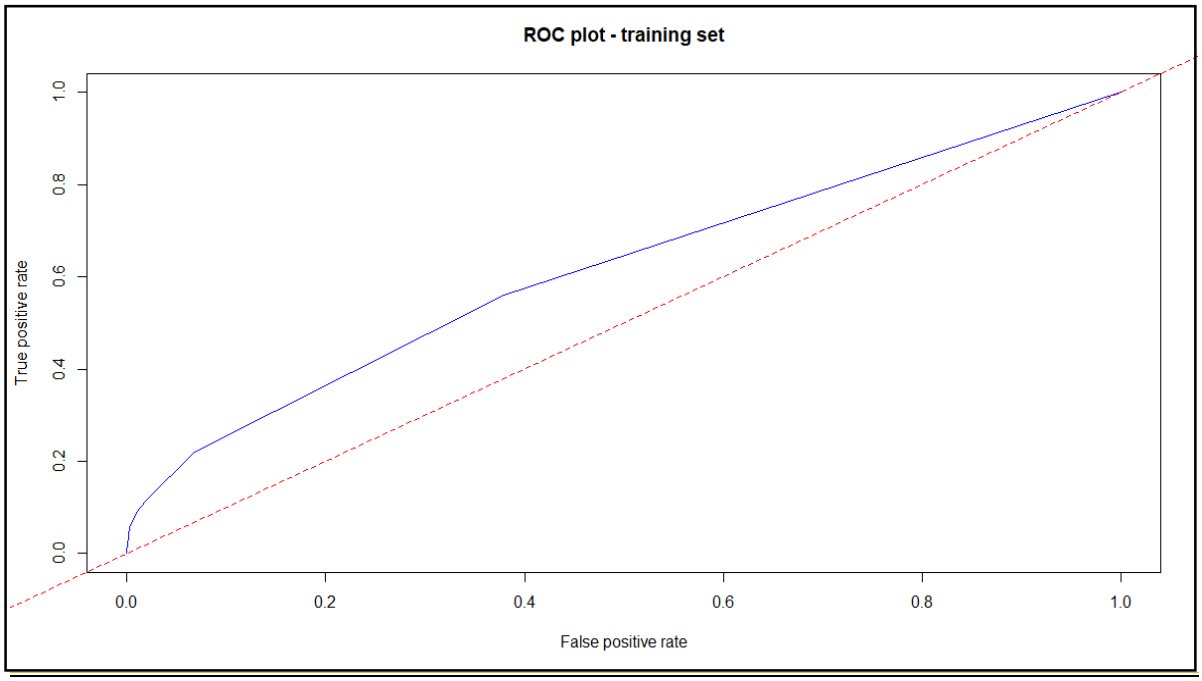
	0	1
Probability	0.8744286	0.1255754

Since, the above data is clearly an imbalance dataset, the number of people who did not react for taking a loan is far greater than the number of people interested in taking a loan.

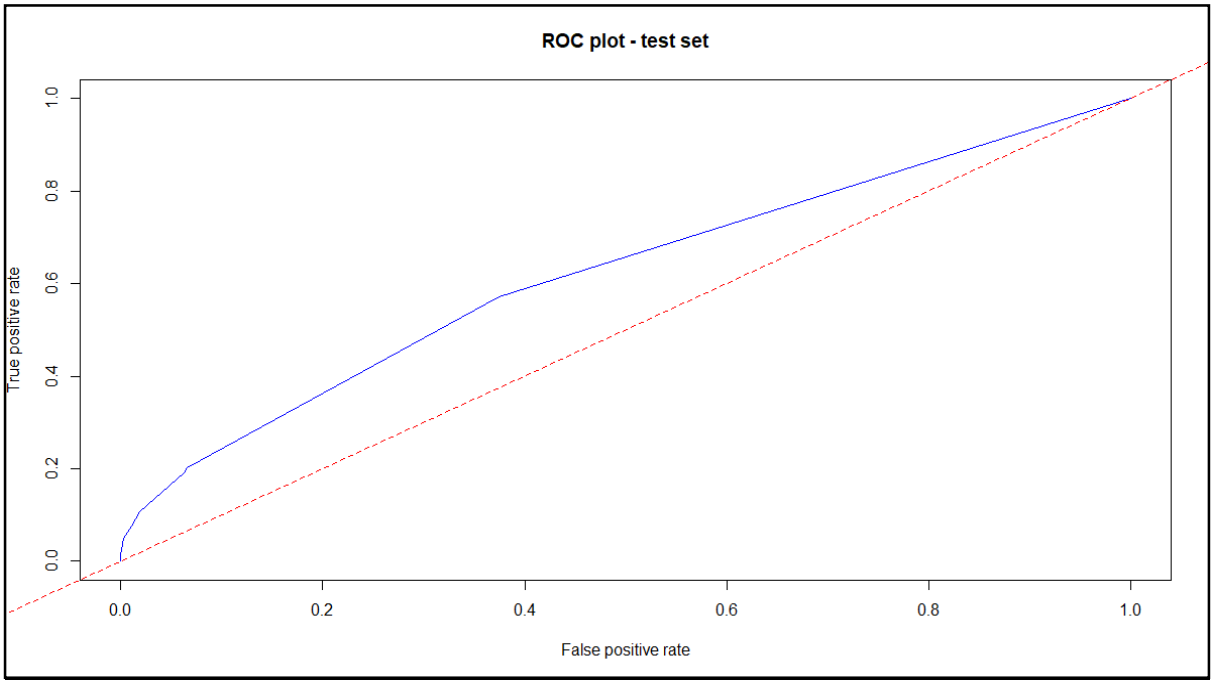
Dendogram:



ROC PLOT TRAIn DATA:



ROC PLOT TEST DATA:



MODEL PERFORMANCE MEASURES:

Train data:

Train Specificity	99.697760
Train Sensitivity	05.74516
Train KS	18.15148
Train AUC	61.57517

Test data:

Test Specificity	0.9971407
Test Sensitivity	0.0503977
Test KS	19.76188
Test AUC	61.97537

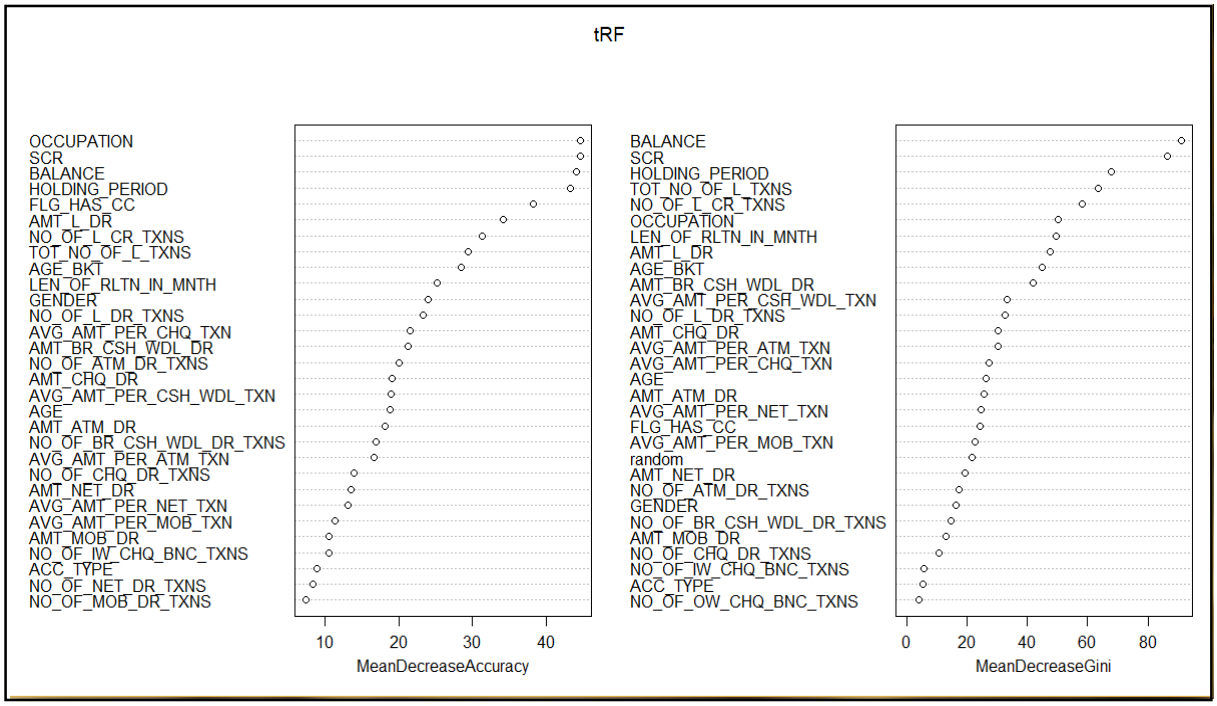
VARIABLE IMPORTANCE:

	ptree.variable.importance
TOT_NO_OF_L_TXNS	79.1823232
NO_OF_L_DR_TXNS	70.13617539
NO_OF_L_CR_TXNS	56.71210916
NO_OF_NET_DR_TXNS	39.54335536
OCCUPATION	33.68393952
NO_OF_CHQ_DR_TXNS	24.84896261
NO_OF_ATM_DR_TXNS	19.11253826
GENDER	16.99444209
BALANCE	16.71105477
NO_OF_BR_CSH_WDL_DR_TXNS	16.53269599
AMT_CHQ_DR	15.41556014
AVG_AMT_PER_MOB_TXN	10.34716544
ACC_TYPE	8.663985849
NO_OF_MOB_DR_TXNS	8.282253349
SCR	2.1309573
HOLDING_PERIOD	1.784756647
AGE	1.492967662
AMT_MOB_DR	0.718351174
AMT_MIN_BAL_NMC_CHGS	0.701796137
LEN_OF_RLTN_IN_MNTH	0.426562189
AVG_AMT_PER_CSH_WDL_TXN	0.131065913

From the above Variable measures of the model, the likelihood of a person taking a loan depends upon

- Total number of transactions
- Occupation
- Gender
- Balance

RF MODEL:



MODEL PERFORMANCE MEASURES:

Train Data:

Train Specificity	99.901976
Train Sensitivity	12.741751
Train KS	61.74217
Train AUC	96.54777

Test Data:

Test Specificity	99.885627
Test Sensitivity	7.9575596
Test KS	54.35961
Test AUC	89.40187

Clearly, from the variable importance measures it shows that the likelihood of a customer taking a loan depends mainly on

- Occupation
- Generic Marketing Score
- Balance
- Holding Period