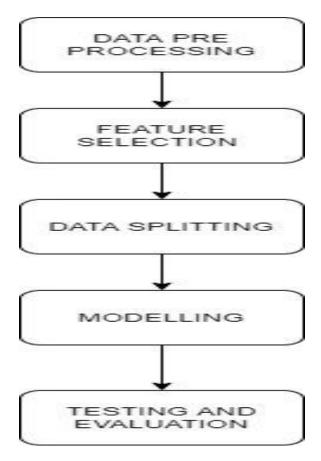
MACHINE LEARNING AND DATA MINING PROJECT VEHICLE LOAN DEFAULT PREDICTION

REGISTRATION NUMBER	NAME
BL.EN.U4CSE17072	SARATH CHANDRA .M
BL.EN.U4CSE17076	M.M. SHOBANDARI
BL.EN.U4CSE17121	SUNIL CHOWDARY .B

ABSTRACT

- Financial institutions incur significant losses due to the default of vehicle loans. This
 has led to the tightening up of vehicle loan underwriting and increased vehicle loan
 rejection rates.
- The need for a better credit risk scoring model is also raised by these institutions. This warrants a study to estimate the determinants of vehicle loan default.
- A financial institution has hired you to accurately predict the probability of loanee/borrower defaulting on a vehicle loan in the first EMI (Equated Monthly Installments) on the due date.
- Doing so will ensure that clients capable of repayment are not rejected and important determinants can be identified which can be further used for minimising the default rates.

IMPLEMENTATION



MODELS:

- RANDOM FORESTS
- LOGISTIC REGRESSION
- DECISION TREE
- NAIVE BAYES

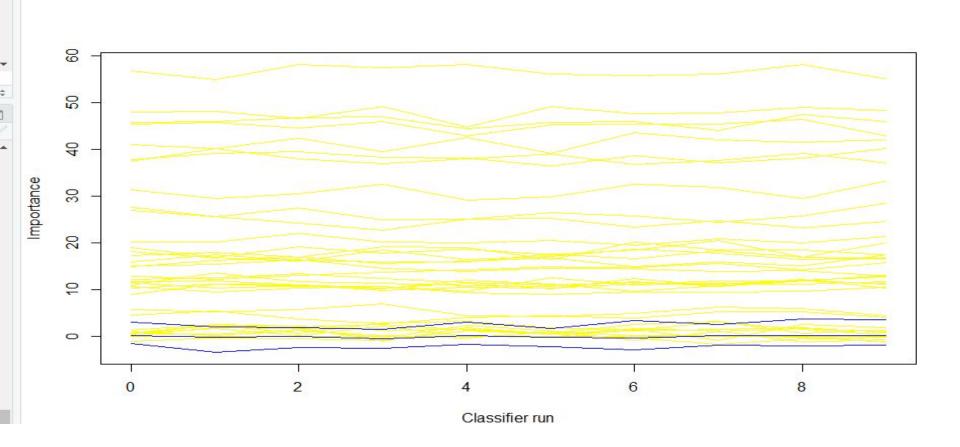
IMPORTED LIBRARIES

```
train 2 ×
         test_2_ × X0 ×
        ☐ Source on Save Q / •
    library(caret)
    library(Boruta)
    library(mlbench)
    library(randomForest)
     require(pROC)
    library(MLmetrics)
    library(e1071)
  8 library(pscl)
  9 library(rpart.plot)
 10
```

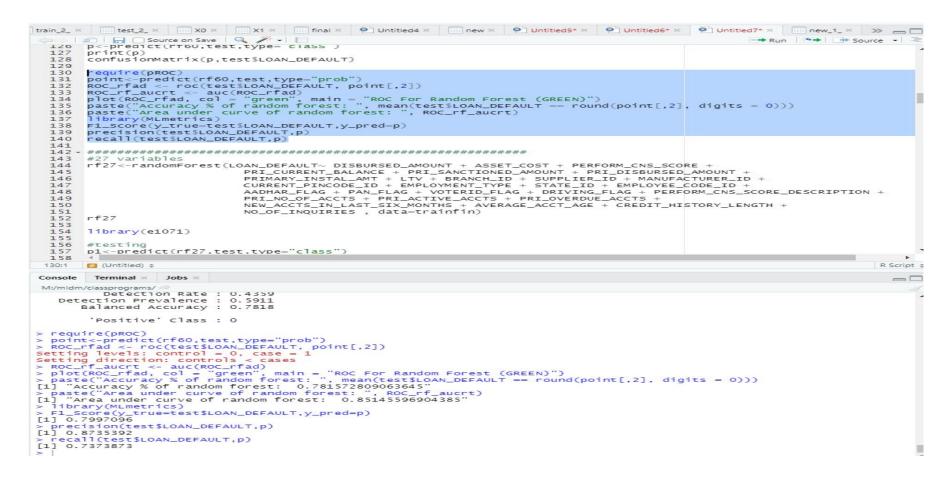
SELECTING FEATURES USING BORUTA

	test_2_ × X								new_1_ ×	
100	Ø□ □ Source on S	ive Q	- I L							ource -
109										
110	#random forest									
111										
112	# getNonRejectedFormula(select)									
113										
	fpart1\$LOAN_DEFA	ULT=as.fa	ctor(fpart:	1\$LOAN_DEFAU	LT)					
115	set.seed(222)									
116										
117	#data splitting									
	ind<-sample(2,nr),replace=	T,prob=c(0.7)	,0.3))					
	trainfin<-fpart									
	test<-fpart1[ind									
	******	******	****	<i>~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~</i>	****					
	#all variables									
123		/			>					
124		S L (LOAN_DI	EFAULT~.,	uaca=crainti	ri)					
			"-3"							
126 127		cest, type	= Class)							
128	print(p) confusionMatrix	n tostsio	AN DEFAULT							
129	Com us formatt IX	p, cescalo	AN_DEFAULT							
130	require(pROC)									
131	point<-predict(ffor test	type="prob							
132										
133				DOc [, 2]						
			en", main :	"ROC FOR R	andom For	est (GREEN)")			
134	paste("Accuracy	% of rand	en", main : om forest:	"ROC For R	andom For	est (GREEN)")	nd(point[.2].	digits = 0))))	
135 136	paste("Accuracy	% of rande	om forest:	", mean(tes	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	
135	paste("Accuracy	% of rander curve of	om forest:	", mean(tes	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	
135 136	paste("Accuracy paste("Area unde library(MLmetric	% of rander curve of (5)	om forest: f random f	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	
135 136 137 138 139	paste("Accuracy paste("Area unde library(MLmetric	% of rander curve of (s) test\$LOAN.	om forest: f random fo _DEFAULT,y.	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	
135 136 137 138 139 140	paste("Accuracy paste("Area unde library(MLmetric F1_Score(y_true=	% of rander curve of (s) test\$LOAN.	om forest: f random fo _DEFAULT,y.	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	.
135 136 137 138 139 140	paste("Accuracy paste("Area unde library(MLmetric F1_Score(y_true=	% of rander curve of (s) test\$LOAN.	om forest: f random fo _DEFAULT,y.	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	
135 136 137 138 139 140	paste("Accuracy paste("Area unde library(MLmetric F1_Score(y_true- precision(test\$)	% of rander curve of (s) test\$LOAN.	om forest: f random fo _DEFAULT,y.	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	R Script
135 136 137 138 139 140 112:32 Console	paste("Accuracy paste("Area und library(MLmetric F1_Score(y_true precision(test\$) (Top Level) : Terminal × Jobs ×	% of rander curve of (s) test\$LOAN.	om forest: f random fo _DEFAULT,y.	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	R Script
135 136 137 138 139 140 112:32 Console	paste("Accuracy paste("Area unde library(MLmetric F1_Score(y_true- precision(test\$1 4 (Top Level) \$	% of rander curve of sis) test\$LOAN.OAN_DEFAUL	om forest: f random fo _DEFAULT,y.	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	R Script
135 136 137 138 139 140 112:32 Console	paste("Accuracy paste("Area und library(MLmetric F1_Score(y_true precision(test\$) (Top Level) : Terminal × Jobs ×	% of rander curve of sis) test\$LOAN_OAN_DEFAUL	om forest: f random fo _DEFAULT,y. _T,p)	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))	>>	R Script
135 136 137 138 139 140 112:32 Console M:/mldm	paste("Accuracy paste("Area undo library(MLmetric F1_Score(y_true precision(test\$) { (Top Level) : Terminal × Jobs ×	% of rander curve of solutions	om forest: f random forest: perault,y. lt,p) eraii	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))))	R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFORI	paste("Accuracy paste("Area und library(MLmetric F1_Score(y_true- precision(test\$) (Top Level) \$ Terminal × Jobs × h/classprograms/ \$ M_CNS_SCORE	% of rander curve of solutions	om forest: f random for the following for the following for the following for the following for the forest for the forest for the forest fores	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))	>>	R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFORI	paste("Accuracy paste("Area undo library(MLmetric F1_Score(y_true precision(test\$)	% of rander curve of solutions	om forest: f random f _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.21	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0)))	R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFOR! LTV PERFOR! LTV PRI_SA	paste("Accuracy paste("Area und library(MLmetric F1_Score(y_true- precision(test\$) (Top Level) Terminal > Jobs > ** ** ** ** ** ** ** ** ** ** ** ** *	% of rander curve of solutions	om forest: f random for _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.21 28.24	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))	>>	R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFORI LTV DISBUR. PERI_SAI	paste("Accuracy paste("Area undo library(MLmetric F1_Score(y_true precision(test\$)	% of rander curve of solutions	om forest: f random f _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.21 28.24 27.47	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0)))	R Script
135 136 137 138 140 112:32 Console M:/midm PERFORI PERFORI DISBUR PRI_DI PRI_CU	paste("Accuracy paste("Area undd library(MLmetric F1_Score(y_true-precision(test\$) Terminal × Jobs × n/classprograms/ MCNS_SCORE MCNS_SCORE_DESCISED_AMOUNT NCTIONED_AMOUNT SBURSED_AMOUNT REENT_BALANCE	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT, y. LT, p) erall 00.00 98.84 46.31 28.24 28.24 27.47 24.56	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	ond(point[,2],	digits = 0))	>>	R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFOR! LV DISBUR. PRI_SA! PRI_DI	paste("Accuracy paste("Area undo library(MLmetric Fl.score(y_true precision(test\$)	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.21 28.24 27.47 24.56 22.29	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))		R Script
135 136 137 138 140 112:32 Console M:/midm PERFORI PERFORI DISBUR PRI_DI PRI_DI PRI_CU	paste("Accuracy paste("Area unde library(MLmetric F1_Score(y_true-precision(test\$1	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT, y. LT,p) erall 00.00 98.84 46.21 28.24 27.47 28.24 27.47 24.56 22.29 0.00	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	ond(point[,2],	digits = 0))		R Script
135 136 137 138 140 112:32 Console M:/midm PERFORI TV DISBUR PRI_DI PRI_SA PRI_DI STATE_	paste("Accuracy paste("Area undo library(MLmetric Fl. Score(y_true precision(test\$)	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.32 46.21 28.24 27.47 24.56 22.29 0.00	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour) nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFORI PERFORI PRI_DI PRI_CU STATE_ PRI_CU STATE_ PAN_FL.	paste("Accuracy paste("Area unde library(MLmetric F1_Score(y_true-precision(test\$1	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT, y. LT,p) erall erall 00.00 98.84 46.21 28.24 27.47 22.29 0.00 0.00	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	ond(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm ERFOR! TV ISBUR RI_DI RI_DI RI_CU TATE_ MPLOY! SSET_ SSET_ SSET_	paste("Accuracy paste("Area undo library(MLmetric Fl.Score(y_true-precision(testState))	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.21 28.24 27.47 24.56 22.29 0.00 0.00 0.00	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm DERFORI PRI_SAI PRI_CU STATE_ PRI_CU STATE_ DRI_CU STATE_ DRI_CU STATE_ DRI_VI DISTATE_ DRI_VI	paste("Accuracy paste("Area unde library(MLmetric Fl_Score(y_true- precision(test\$1 (Top Leve) := Terminal × Jobs × M_CNS_SCORE M_CNS_SCORE_DESCI SED_AMOUNT NCTIONED_AMOUNT SBURSED_AMOUNT RRENT_BALANCE ID MENT_TYPE AG COST G_FLAG —HISTORY_LENGTH	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT, y. LT,p) erall 00.00 98.84 46.21 28.24 27.47 22.29 0.00 0.00 0.00 0.00	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	ond(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm ERFOR! TV TV ISBUR RI_DI RI_CU ISBUR RI_DI RI_CU ISSET_ MPLOYU SSET_ RI_VIN REDIT. URREN	paste("Accuracy paste("Area undo library(MLmetric Fl.score(y_true-precision(tests) (Top Level) (Top Level)	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.21 28.24 27.47 24.56 22.29 0.00 0.00 0.00 0.00 0.00	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M://midm PERFORIPERFORI TV DISBUR PRI_CU STATE_ PRI_CU STATE_ DRIVINI CURREN TU CURREN TU CUT TU CURREN TU CUT TU T	paste("Accuracy paste("Area undid library(MLmetric Fl_Score(y_true-precision(test\$1	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT, y. LT,p) erall 00.00 98.84 46.21 28.24 27.47 22.29 0.00 0.00 0.00 0.00 0.00 0.00	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFOR! LTV PERFOR! LTV PERFOR! LTV PERFOR! LTV PERFOR! LTV PERFOR! LTV PERFOR! VICE PERFOR! VICE PER	paste("Accuracy paste("Area undo library(MLmetric Flascore(y_true-precision(tests) 4	% of rander control of rander curve of solutions of the control of	om forest: f random forest: f random forest: pressure forest: f random forest: DEFAULT, y. DEFAULT, y.	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm ERFORI ERFORI TV VISBUR RIL_SU ISBUR RIL_SU INTALE_ IMPLOY! AN_FL ISSET_ URREN' URREN' OTERI IO_OF_ ISSEC_OV	paste("Accuracy paste("Area undid library(MLmetric Fl.Score(y_true-precision(test\$1	% of rander control of rander curve of solutions of the control of	om forest: f random f f random f _DEFAULT, y. LT,p) erall 00.00 98.84 46.21 28.24 27.47 22.29 0.00 0.00 0.00 0.00 0.00 0.00 0.0	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFOR! -TV DISBUR: PRI_GU DISBUR: PRI_GU DRIVING CEMPLOY! CASSET_GU DRIVING CEMPLOY! COREVING COREV	paste("Accuracy paste("Area undo library(MLmetric Flascore(y_true-precision(test\$)] [Top Level] : Terminal × Jobs × Jobs ×	% of rander control of rander curve of solutions of the control of	om forest: f random f _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.32 46.21 28.24 27.47 24.56 22.29 0.00 0.00 0.00 0.00 0.00 0.00 0.0	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFORI TV DISBURA PRILDIP PRILDIP PRILDIP PRILDIP PRILDIP FRIL	paste("Accuracy paste("Area undid library(MLmetric Fl_Score(y_true-precision(test\$1	% of rander control of rander curve of solutions of the control of	om forest: f random f DEFAULT, y. LT,p) erall 00.00 98.84 46.21 28.24 27.47 22.29 0.00 0.00 0.00 0.00 0.00 0.00 0.0	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M:/midm PERFOR! LTV DISBUR. PRI_SA DISBUR. PRI_SA DISBUR. PRI_SU DISBUR.	paste("Accuracy paste("Area undo library(MLmetric F1_Score(y_true-precision(test\$)] [Terminal × Jobs × (Top Level) † (Top Level)	% of rander curve of solutions	om forest: f random f _DEFAULT,y. LT,p) erall 00.00 98.84 46.32 46.32 46.21 28.24 27.47 24.56 22.29 0.00 0.00 0.00 0.00 0.00 0.00 0.0	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script
135 136 137 138 139 140 112:32 Console M/midm M/midm PERFFOR PERFOR PERI_CU STATE_EMPLOYI PAN_FL ASSET_U DRIVIN CREDIT UOF SEC_OV ADDHAR PSEC_IN PORI_SEC_IN PARIL	paste("Accuracy paste("Area undid library(MLmetric Fl_Score(y_true-precision(test\$1	% of rander curve of s) Test\$Loan, OAN_DEFAUL OV. RIPTION	om forest: f random f DEFAULT, y. LT,p) erall 00.00 98.84 46.21 28.24 27.47 22.29 0.00 0.00 0.00 0.00 0.00 0.00 0.0	", mean(tes orest: ", RO	t\$LOAN_DE	FAULT == rour	nd(point[,2],	digits = 0))		R Script

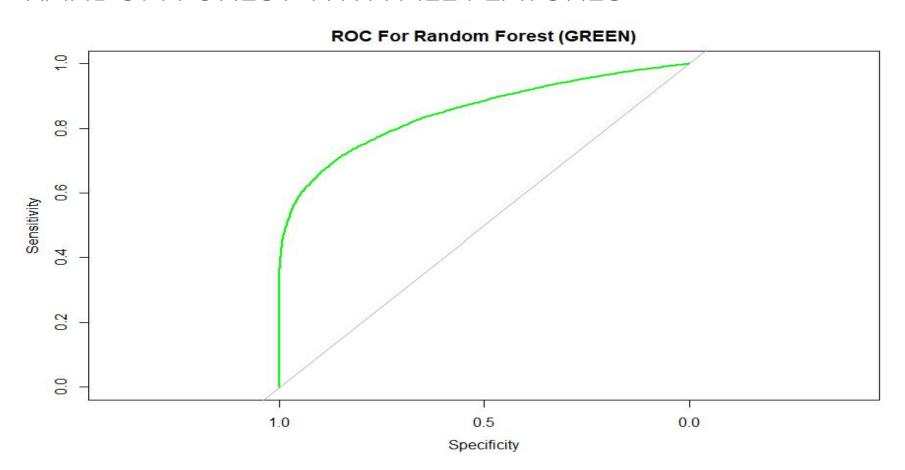
SELECTING FEATURES USING BORUTA



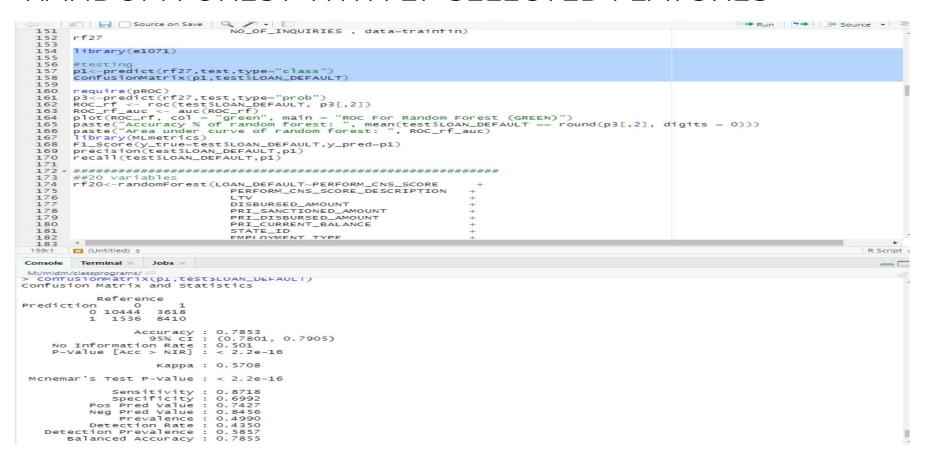
RANDOM FOREST WITH ALL FEATURES



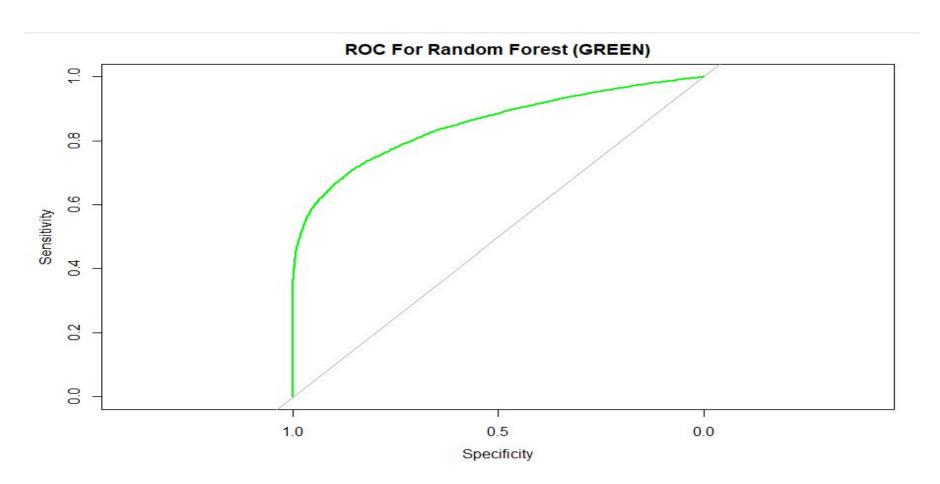
RANDOM FOREST WITH ALL FEATURES



RANDOM FOREST WITH 27 SELECTED FEATURES



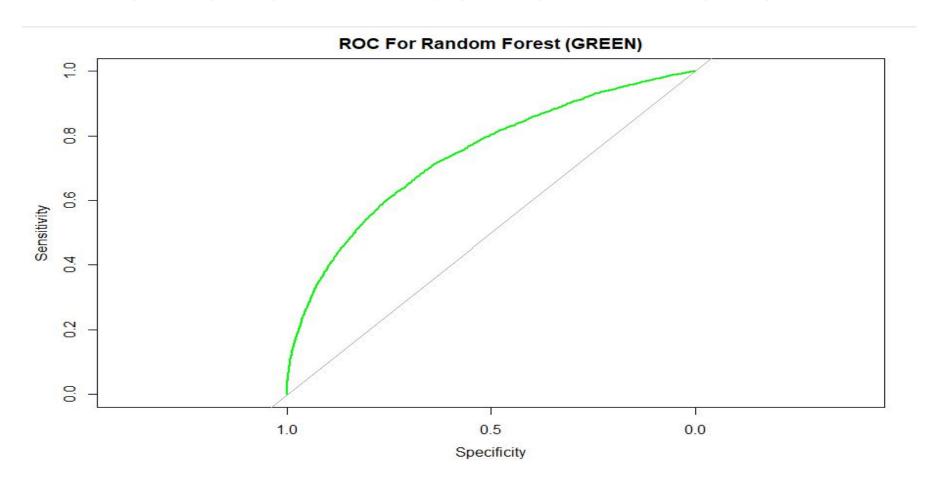
RANDOM FOREST WITH 27 SELECTED FEATURES



RANDOM FOREST WITH 20 SELECTED FEATURES

train_2_ ×	test_2_ × X0 ×	X1 ×	final ×	Untitled4 ×	new × C	Untitled5* ×	Untitled6* ×	Untitled7* ×	new_1_ × >>>
< 1		10/						→ Ru	n 🍑 🕩 Source 🕶
192 193			TIVE_ACCT	s	+ , data=tr	ainfin)			
	rf20								
195	#testing								
197	p2<-predict(rf20,								
	confusionMatrix(p	2,test\$LO	AN_DEFAULT)					
199	library(ROCR)								
201	, is any choche								
202	#metrics								
203	require(pROC) point2<-predict(r	F20 +as+ 1	type-"prob	***					
205	ROC_rfas <- roc(t								
206	ROC_rf_aucas <- a	uc(ROC_rfa	as)						
207 208	plot(ROC_rfas, co paste("Accuracy %	I = "green	n", main =	"ROC For R	andom Fores	t (GREEN)"	od(noint35 3	l dioite o	222
209	paste("Area under	curve of	random fo	rest: ". Ro	C_rf_aucas)	OLI == I OUI	na (pornez [, 2	j, digits = 0	
210	library(MLmetrics)							
211	F1_Score(y_true=t			pred-p2)					
213	precision(test\$LO. recall(test\$LOAN_								
214	1								
215 216									
217									
218 -	###############	########	++++++++	<i>++++++++++</i>	##########	#########	#####		
219 220									
221	#logistic regress	ion							
222									
223 +	4								
	(Untitled) \$								R Scrip
224	1				. A.F. A.F. A.F. A.F. A.F. A.F. A.F. A.				R Scrig
214:1 Console	(Untitled) \$				42 05 4F 4F 4F 4F 4F 4F 4F 4F 4F				
214:1 Console M:/midm > libr Warnin packag >	Terminal × Jobs × //classprograms/ ary (ROCR) g message: e 'ROCR' was built								
214:1 Console M:/mldm > libr warnin packag > #met	Terminal × Jobs × //classprograms/ ** ary(ROCR) g message: e 'ROCR' was built								
214:1 Console M:/midm > libr Warnin packag > > #met > requ	Terminal × Jobs × //classprograms/ ary (ROCR) g message: e 'ROCR' was built rics ire(pROC)	under R	version 4.						
214:1 Console M:/midm > libr warnin packag > / met > requ > poin > ROC_	Terminal × Jobs × //classprograms/ ary (ROCR) g message: e 'ROCR' was built rics ire(pROC) t2<-predict(rf20,t rfas <- roc(test\$L	under R · est,type= oan_defaul	version 4. "prob") LT, point2	0.3					
Console M:/midm > libr warnin packag > requ > poin > ROC_ Settin	Terminal × Jobs × //classprograms/ ** ary(ROCR) g message: e 'ROCR' was built rics ire(pROC) t2<-predict(rf20,t rfas <- roc(test\$L g levels: control	under R · est,type= oAN_DEFAUL = 0, case	version 4. "prob") LT, point2	0.3					
Console M:/midm > libr warnin packag > requ > poin > ROC_ Settin > ROC_ Settin	Terminal × Jobs × //classprograms/ ** ary(ROCR) g message: e 'ROCR' was built rics ire(pROC) t2<-predict(rf20,t rfas <- roc(test\$L g levels: control g direction: control g direction: control rf_aucas <- auc(RO	under R v est,type= OAN_DEFAU = 0, case ols < case c_rfas)	version 4. "prob") LT, point2 = 1	0.3					
Console M:/midm > libr warnin packag > requ > poin > ROC_ Settin > ROC_ Settin	Terminal × Jobs × //classprograms/ ** ary(ROCR) g message: e 'ROCR' was built rics ire(pROC) t2<-predict(rf20,t rfas <- roc(test\$L g levels: control g direction: control g direction: control rf_aucas <- auc(RO	under R v est,type= OAN_DEFAU = 0, case ols < case c_rfas)	version 4. "prob") LT, point2 = 1	0.3					
Console M:/midm > libr warnin packag > #met > requ > poin > ROC_ Settin > ROC_ > plot	Terminal × Jobs × //classprograms/ ** ary (ROCR) g message: e 'ROCR' was built rics ince (pROC) t2<-predict(rf20,t rfas <- roc(test§L g direction: control g direction: contro	under R * est,type= OAN_DEFAUL OAN_Case ols < case C_rfas) green", m. andom for	version 4. "prob") LT, point2 = 1 es ain = "ROCest: ", me	0.3 [,2]) For Random an(test\$LOA				gits = 0)))	
Console M:/midm Warnin packag / / requ /	Terminal × Jobs × Terminal × Jobs × J	est, type= DAN_DEFAU! 0 case ols < case c_rfas) green", mandom for- m forest: e of rand	version 4. "prob") LT, point2 es ain = "ROC est: ", me 0.677149 om forest:	0.3 [,2]) For Random an(test\$LoAs 283572143" ", ROC_rf_",	n Forest (GR NN_DEFAULT = .aucas)			g1ts = 0)))	
Console M:/midm > libr warnin packag > #met > requ > poin > ROC_ Settin > ROC_ > plot > past [I] "A > past [I] "A	Terminal × Jobs × Acclassprograms/ ary(ROCR) g message: e'ROCR' was built rics ire(pROC) t2<-predict(rf20,t rfas <- roc(test\$t g levels: control g direction: contr rf_aucas <- auc(ROC (ROC_rfas, col = " e("Accuracy % of r curacy % of r and e("Area under curve rea under curve of	est, type= DAN_DEFAU! 0 case ols < case c_rfas) green", mandom for- m forest: e of rand	version 4. "prob") LT, point2 es ain = "ROC est: ", me 0.677149 om forest:	0.3 [,2]) For Random an(test\$LoAs 283572143" ", ROC_rf_",	n Forest (GR NN_DEFAULT = .aucas)			g1ts = 0)))	
Console M:/midm warnin packag / metale packag / metale packag p	Terminal × Jobs × Terminal × Jobs × J	est,type= OAN_DEFAU! 0	version 4. "prob") LT, point2 es ain = "ROC est: ", me 0.677149 om forest: orest: 0.	0.3 [,2]) For Random an(test\$LoA283572143" ", ROC_rf_73974056014	n Forest (GR NN_DEFAULT = .aucas)			gits = 0)))	
Console M:/midm > libr warnin packag > met > requ > poin > Rettin > plot plot plot plot libr > F1_5 [1] 0.	Terminal × Jobs × Terminal × Jobs × J	est,type= OAN_DEFAU OIO, Case OIO, Case OF as Green", mandom for e of rand random for	version 4. "prob") LT, point2 es ain = "ROC est: ", me 0.677149 om forest: orest: 0.	0.3 [,2]) For Random an(test\$LoA283572143" ", ROC_rf_73974056014	n Forest (GR NN_DEFAULT = .aucas)			g1ts = 0)))	
Console M:/midm > libr warnin packag > #met > requ > poin > ROC_ > plot > pl	Terminal × Jobs × Jobs × J	est,type= OAN_DEFAU OIO, Case OIO, Case OF as Green", mandom for e of rand random for	version 4. "prob") LT, point2 es ain = "ROC est: ", me 0.677149 om forest: orest: 0.	0.3 [,2]) For Random an(test\$LoA283572143" ", ROC_rf_73974056014	n Forest (GR NN_DEFAULT = .aucas)			g1ts = 0)))	
Console M:/midm > libr warnin packag > met > requ > poin > ROC Settin > ROC past libr > past libr > F1_S libr > F1_S lipr > prec ligr o.	Terminal × Jobs × Terminal × Jobs × J	under R ' est,type= OAN_DEFAUI -10, Case -10, Case -fas; green', mandom forest: e of rand random for	version 4. "prob") LT, point2 es ain = "ROC est: ", me 0.677149 om forest: orest: 0.	0.3 [,2]) For Random an(test\$LoA283572143" ", ROC_rf_73974056014	n Forest (GR NN_DEFAULT = .aucas)			gits = 0)))	
Console M:/midm > libr warnin packag > #met > requ > poin > ROC_ > plot > pl	Terminal × Jobs × Jobs × J	under R ' est,type= OAN_DEFAUI -10, Case -10, Case -fas; green', mandom forest: e of rand random for	version 4. "prob") LT, point2 es ain = "ROC est: ", me 0.677149 om forest: orest: 0.	0.3 [,2]) For Random an(test\$LoA283572143" ", ROC_rf_73974056014	n Forest (GR NN_DEFAULT = .aucas)			g1ts = 0)))	

RANDOM FOREST WITH 20 SELECTED FEATURES



RESULTS OF ALL IMPLEMENTATIONS

ACCURACY:

MODEL	ALL FEATURES	27 FEATURES	20 FEATURES
RANDOM FOREST	78.17	78.53	67.69
LOGISTIC REGRESSION	66.05	66.05	61.72
DECISION TREE	67.04	67.04	62.45
NAIVE BAYES	64.51	58.16	58.18

RESULTS OF ALL IMPLEMENTATIONS

AUC:

MODEL	ALL FEATURES	27 FEATURES	20 FEATURES
RANDOM FOREST	85.14	85.42	73.97
LOGISTIC REGRESSION	66.05	66.05	61.72
DECISION TREE	69.71	67.90	63.58
NAIVE BAYES	70.30	70.99	62.21

RESULTS OF ALL IMPLEMENTATIONS

F1 SCORE:

MODEL	ALL FEATURES	27 FEATURES	20 FEATURES
RANDOM FOREST	0.7997	0.8020	0.6800
LOGISTIC REGRESSION	0.6679	0.6678	0.6104
DECISION TREE	0.7305	0.7305	0.6091
NAIVE BAYES	0.6494	0.6967	0.6971

REFERENCES

- https://rpubs.com/Monesh_23992/Bank_Loan_Default Reference model for the whole project
- https://www.analyticsvidhya.com/blog/2016/03/select-important-variables-boru
 ta-package/ A description for boruta feature selection
- <u>https://towardsdatascience.com/random-forest-in-r-f66adf80ec9</u> Used for implementing random forest.