Target Marketing On Banking Product

Presented By:-
Siddhesh Tiwari

Objective

Build a predictive model which can be used to target banking customers for a newly introduced Banking Product.

Product:- “Annuity”

Data Description

Dataset : Banking data

Source of Data: bigML

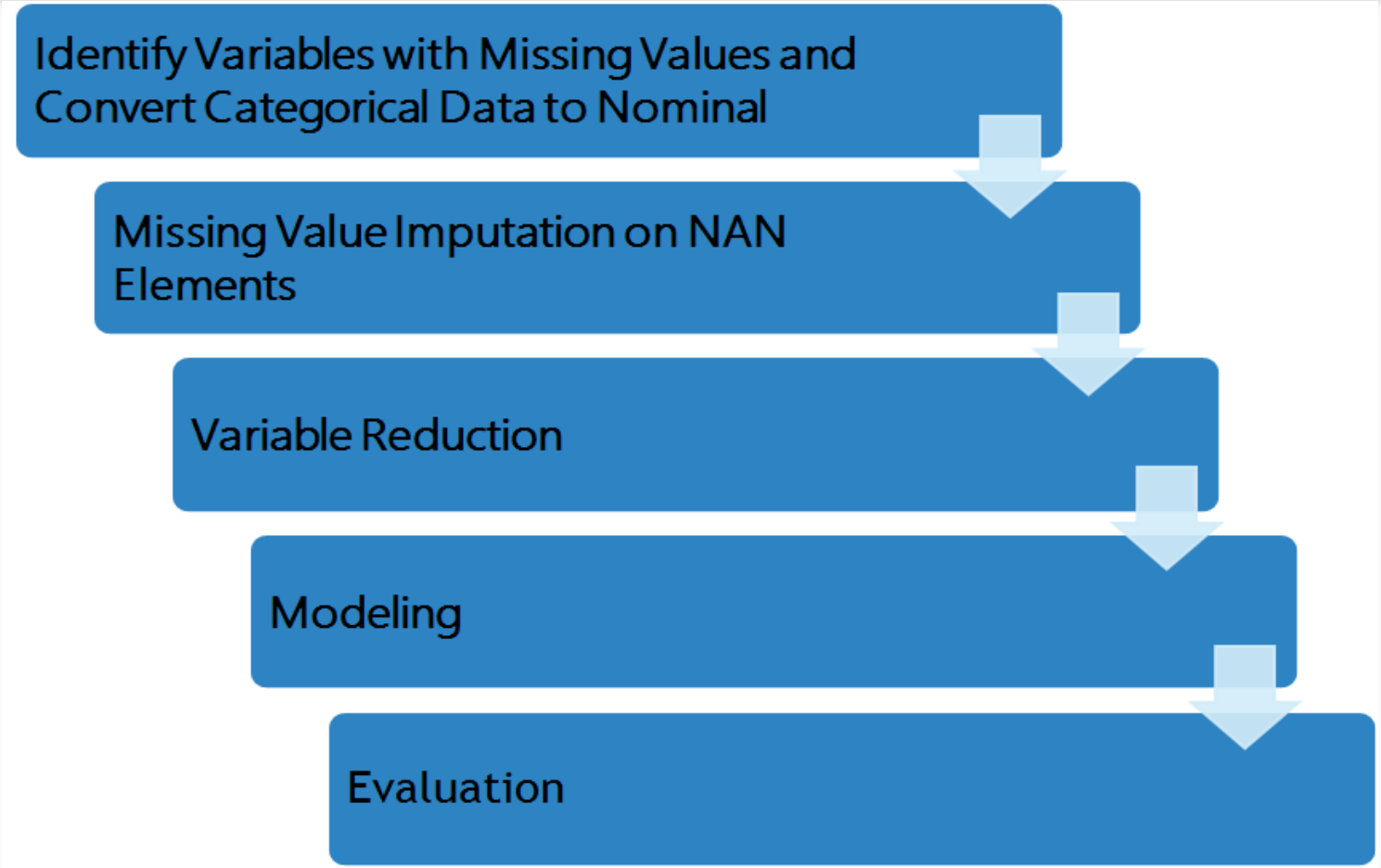
48 variables like Age of Oldest Account, Checking Balance, Amount Deposited, Direct Deposit , Number Insufficient Fund

32000 Observations

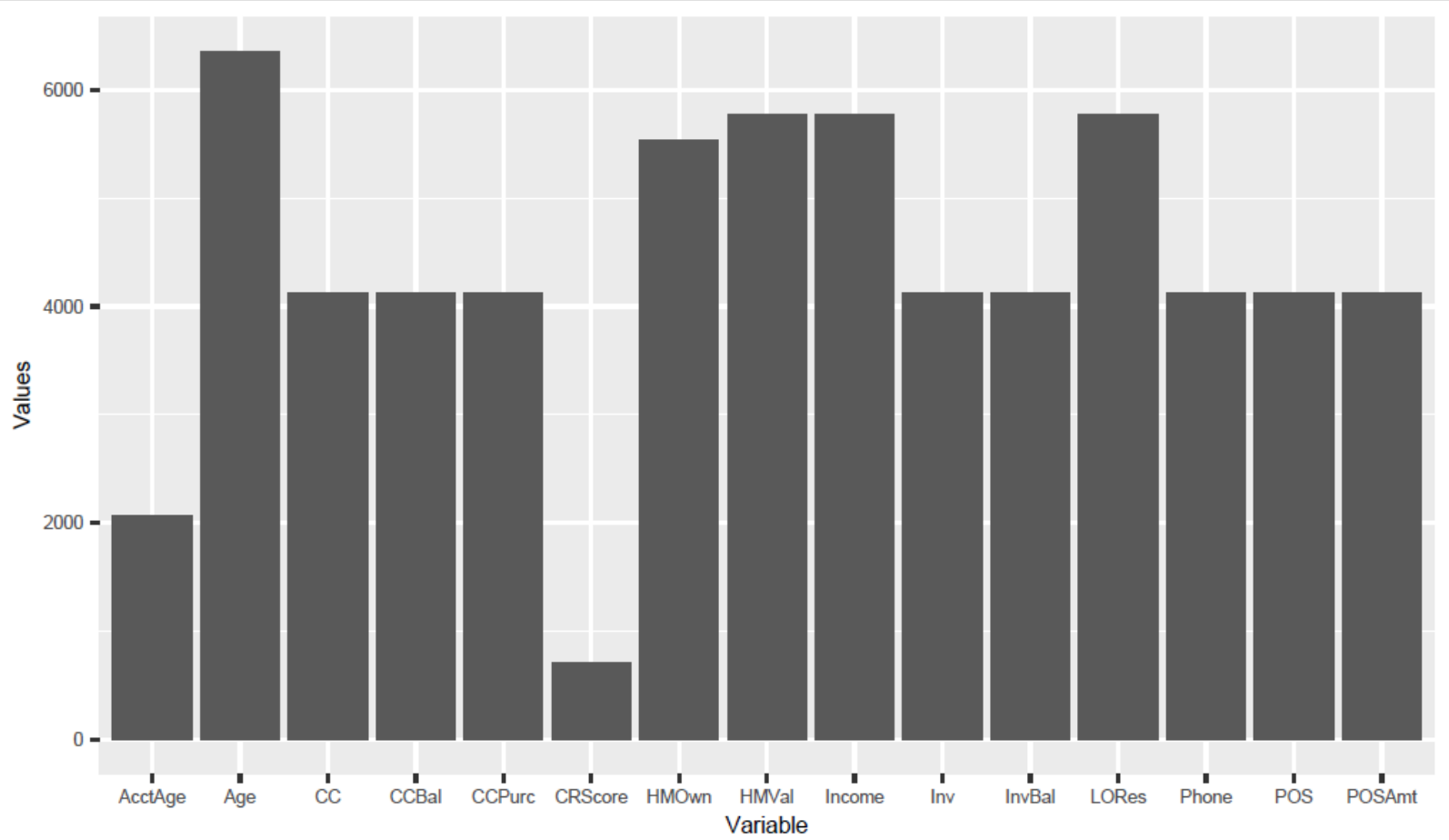
Goals :

- To improve the performance of model and increase the classification accuracy using various approaches.
- To improve class Recall as our focus is to predict the one’s who are interested in buying the product.

Implementation Flow



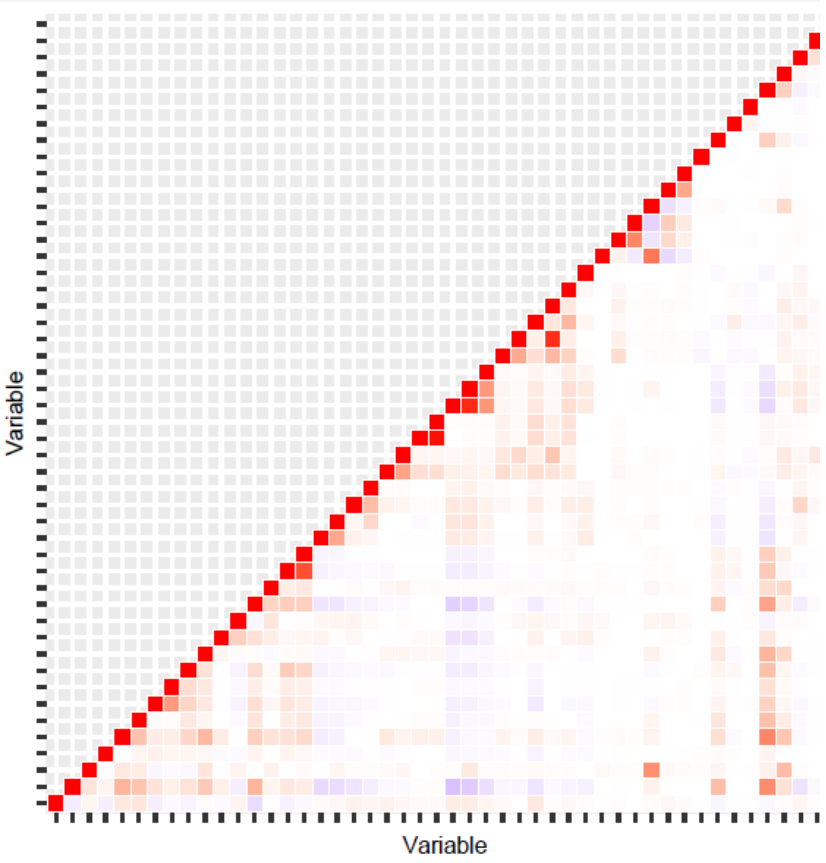
Step 1 : Identification of Variables with Missing Values

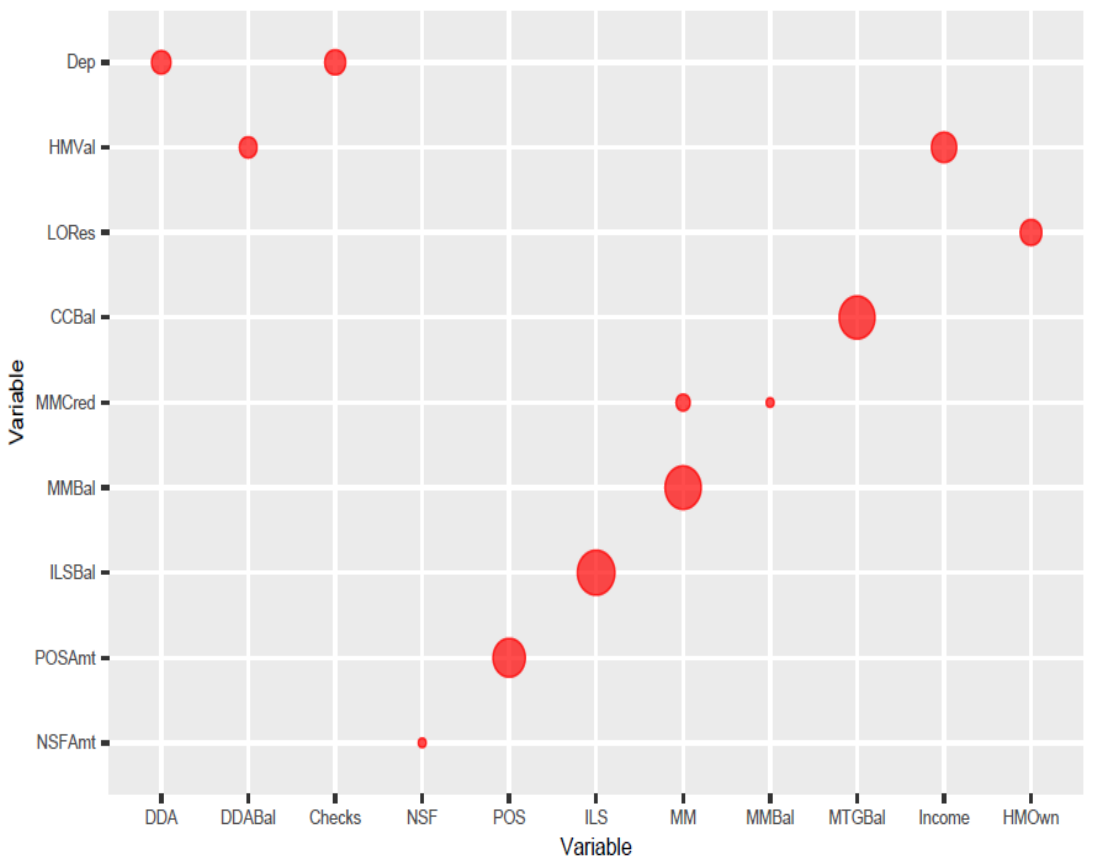


Step 2 : Handling of Missing values

- Imputed missing values using Median
- Regression was not very feasible as target variable was dichotomous
- After missing value imputation we converted 2 categorical variables into nominal
- Finally, we normalized the data using z-transformation

Step 3 : Variable Reduction





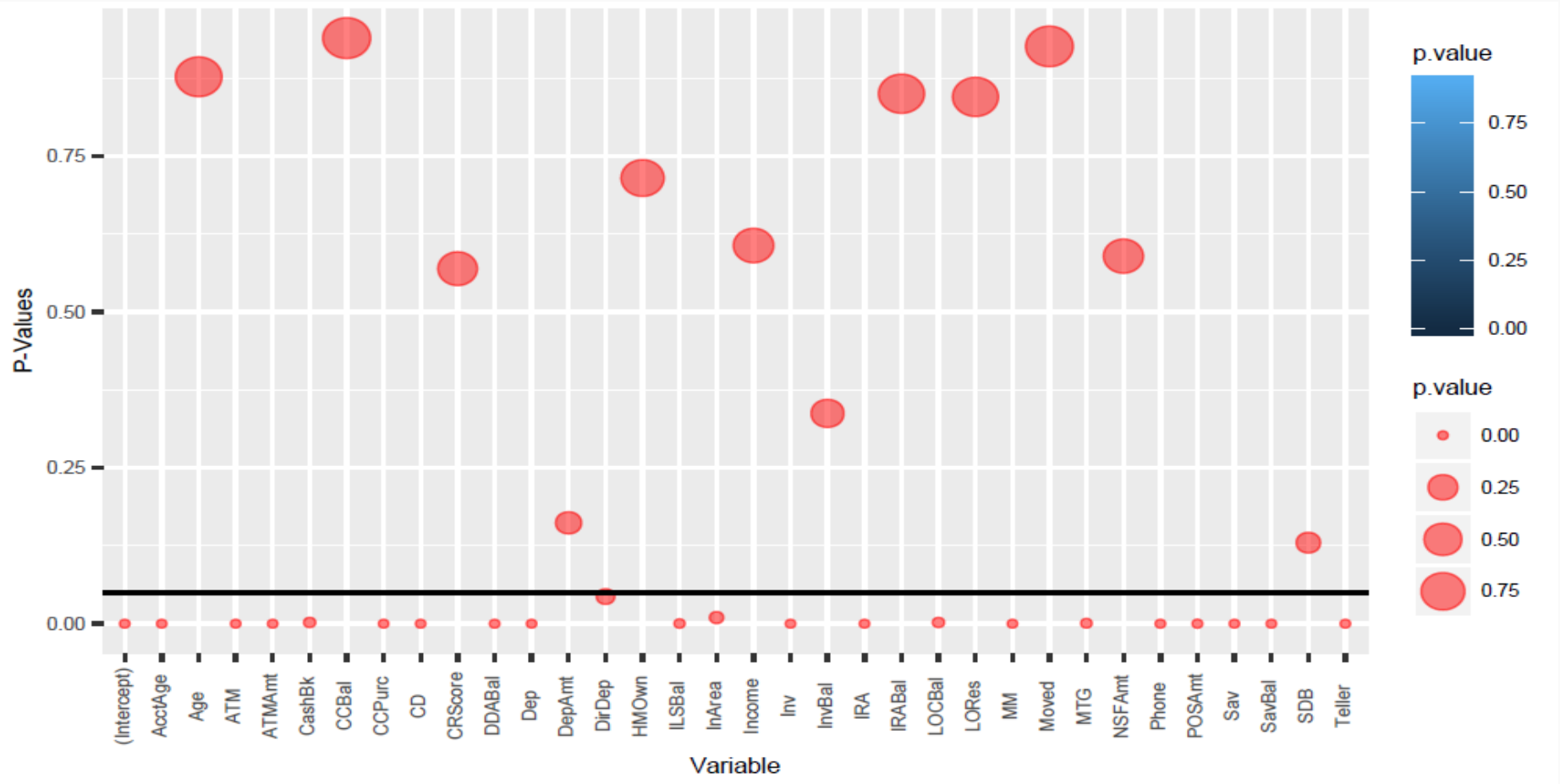
Split Data

Original Distribution of 0 and 1 in Target variable		
0	21089	65.36%
1	11175	35%

Split Data Using Random Sampling

	Training	Distribution	Validation	Distribution
0	14028	65%	7061	66%
1	7483	35%	3692	34%

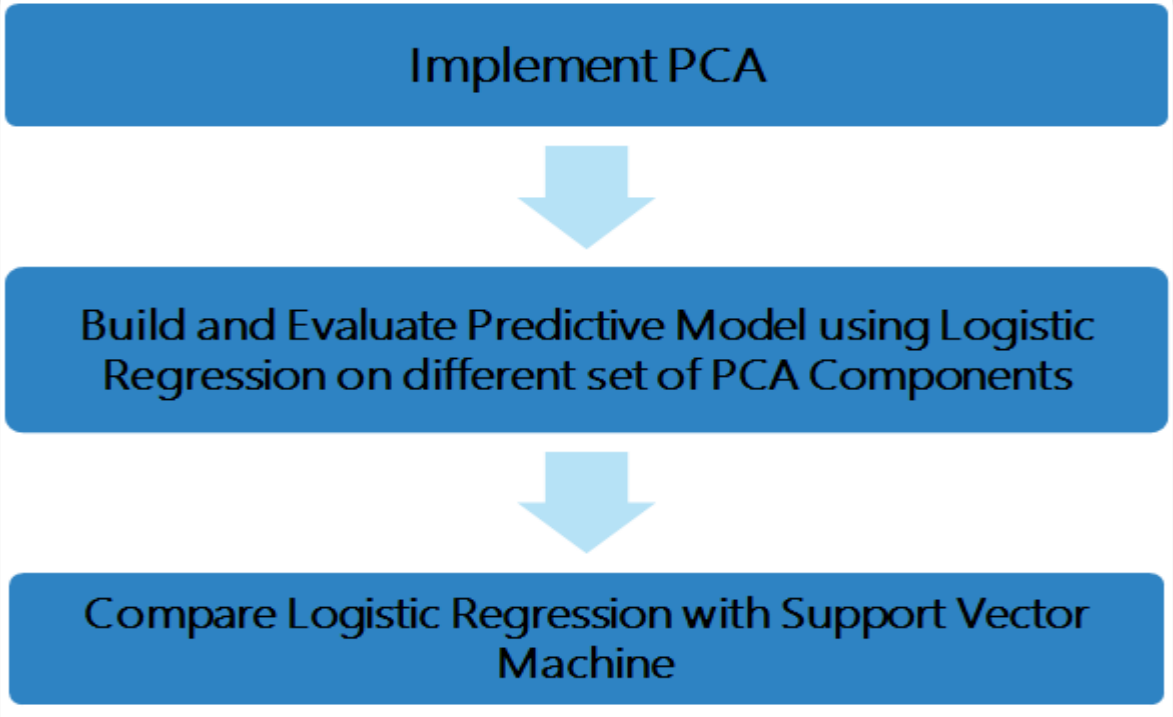
Fitted Logistic Regression to find significant variables



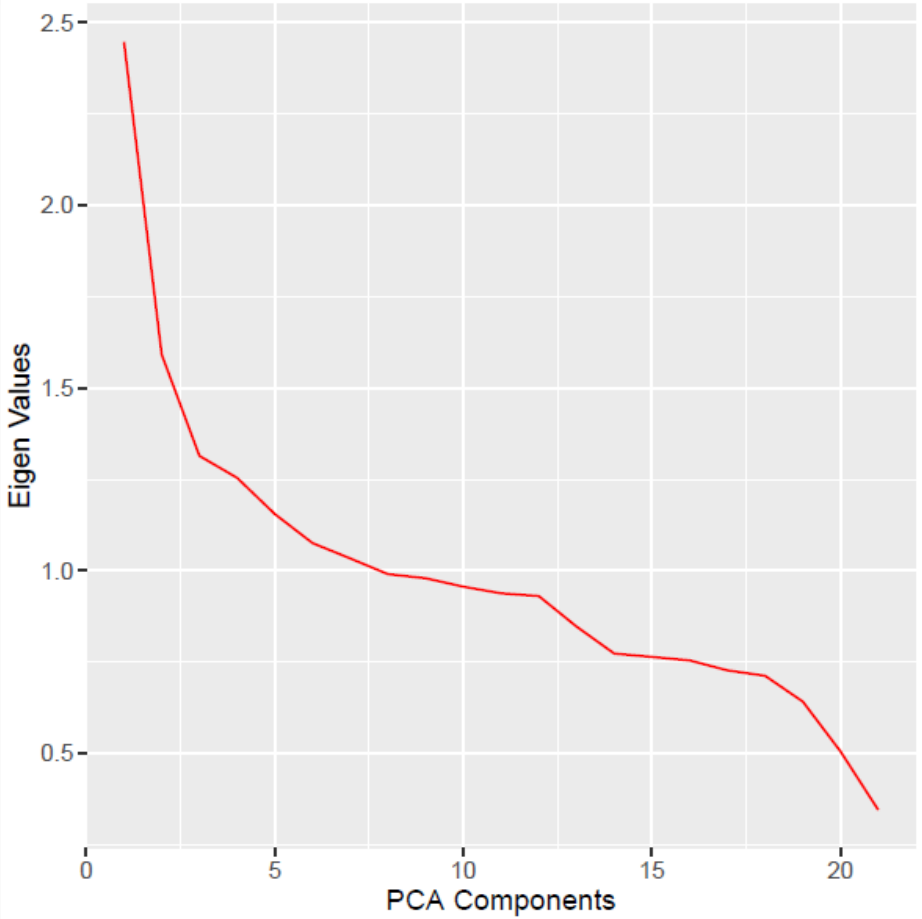
Explanatory Model after Variable Reduction

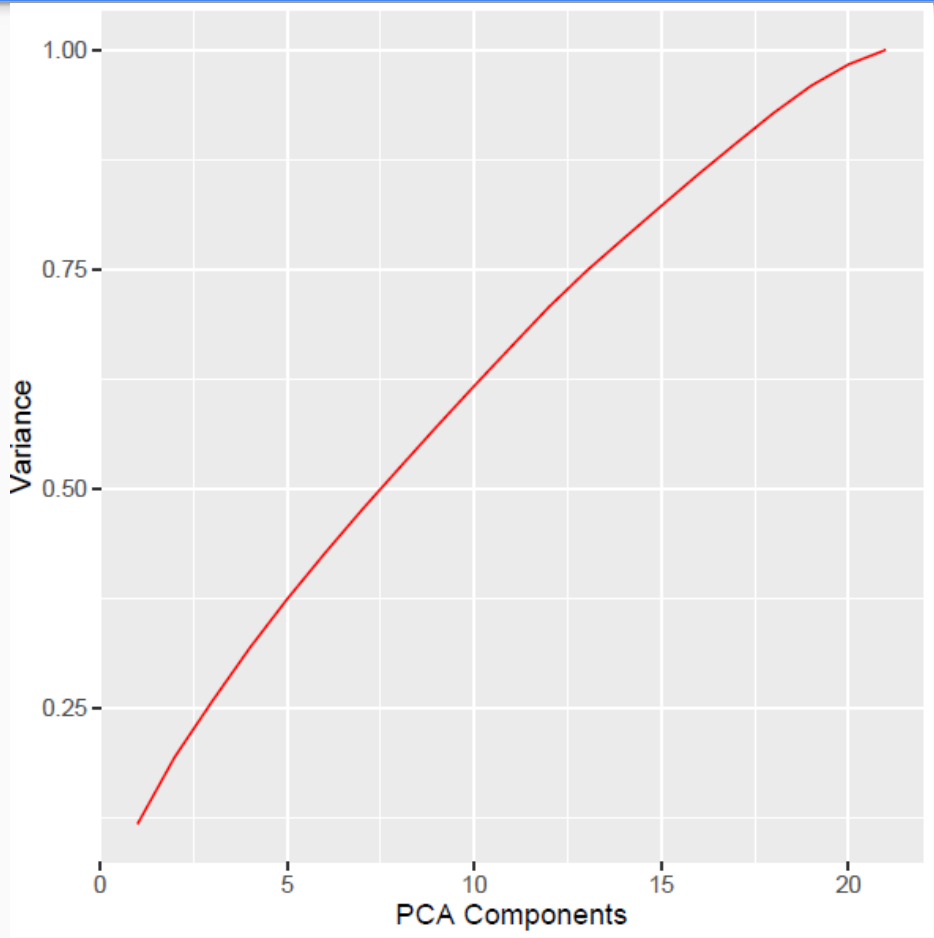
Variables	Coeff	Variables	Coeff
Teller	0.16272	CashBk	-0.05754
MM	0.2591	IRA	0.108798
ILSBal	-0.05934	AcctAge	-0.0896
LOCBal	-0.05267	SavBal	0.732304
POSAmt	0.068936	DDABal	0.36221
CD	0.318465	InArea	-0.04063
CCPurc	0.086557	Sav	0.228052
ATMAmt	0.263584	Phone	-0.0903
Inv	0.104159	MTG	-0.06
Dep	-0.29566	DirDep	-0.0355
		ATM	-0.14261

Step 4 : Modeling & Evaluation

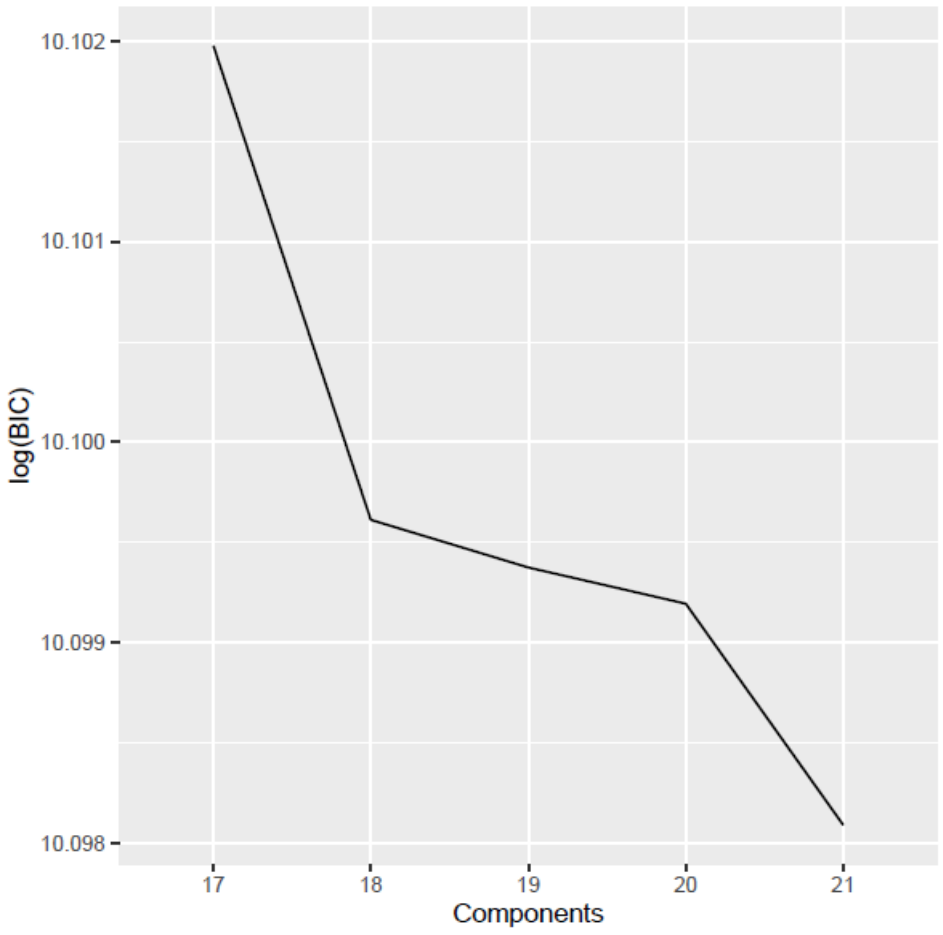


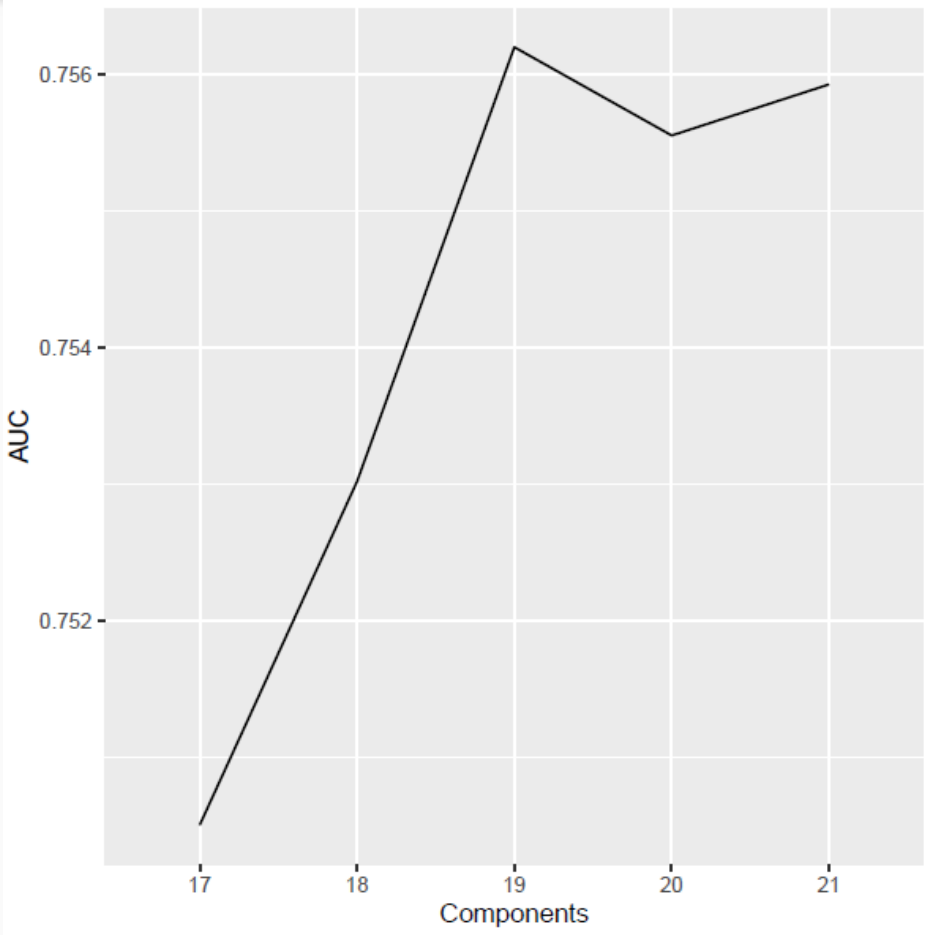
Principle Component Analysis



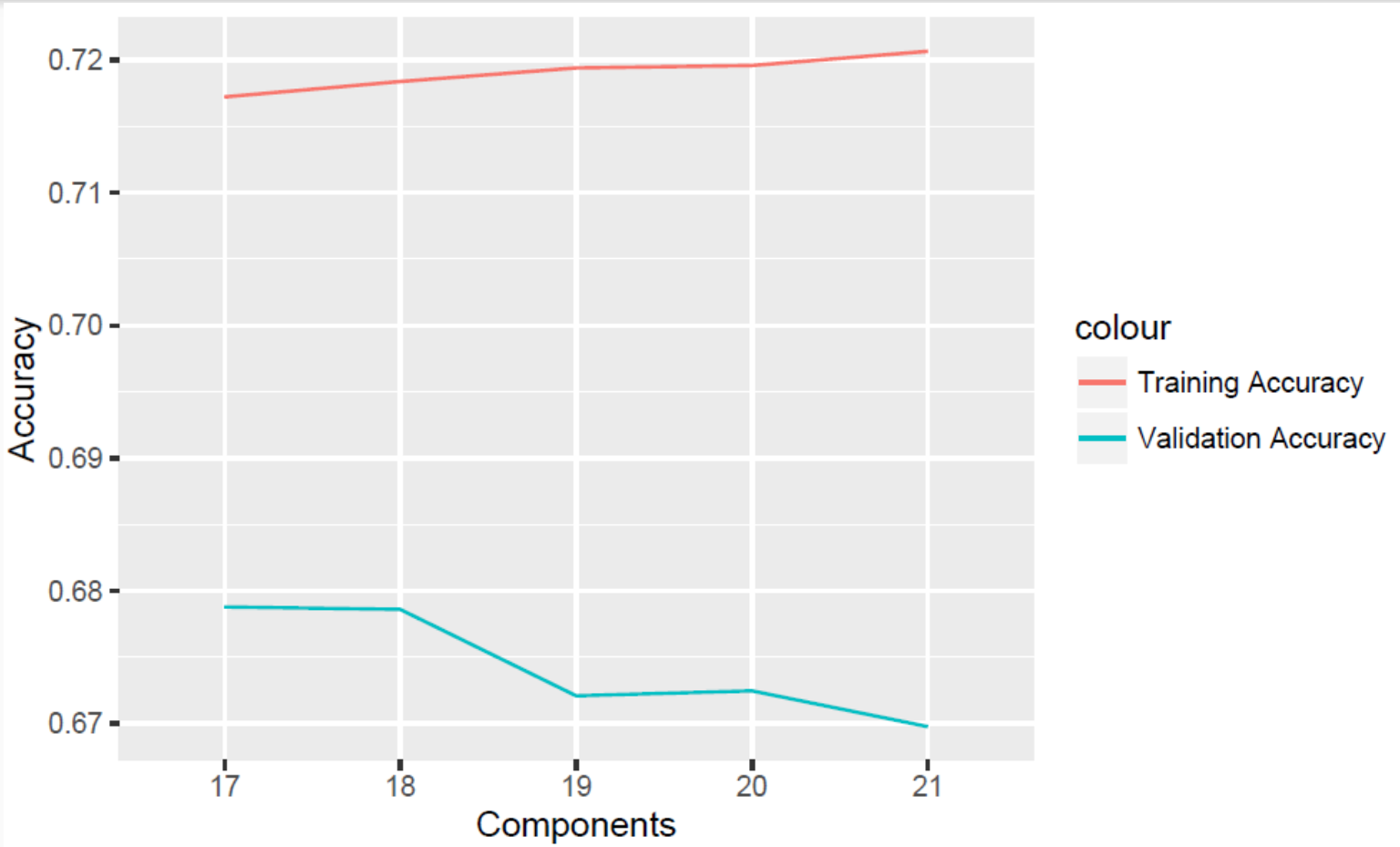


Model Evaluation

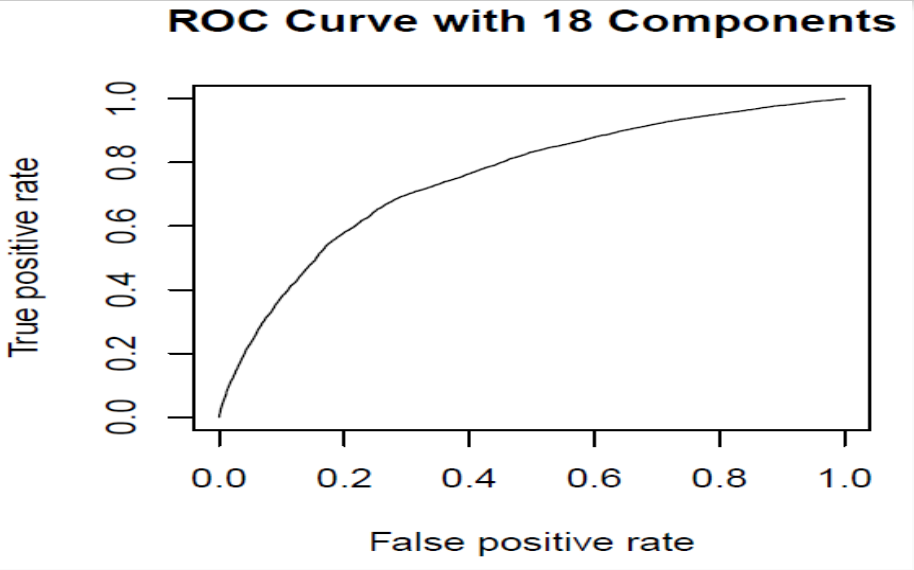


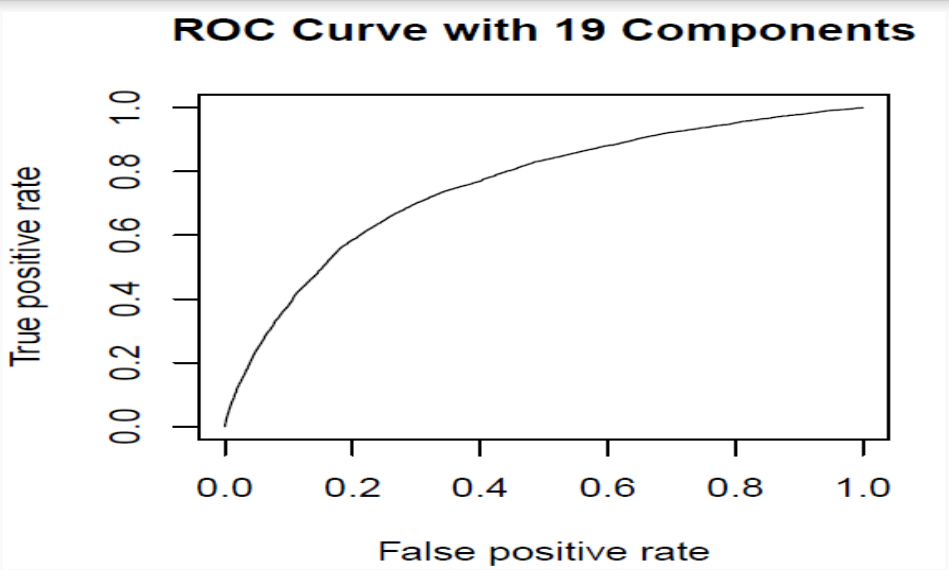


Training & Testing Accuracy



ROC Curve & Confusion Matrix of Best Models





	FALSE	TRUE	class.error
FALSE	12725	4712	0.0956
TRUE	1346	2728	0.6333

	FALSE	TRUE	class.error
FALSE	12760	4725	0.0931
TRUE	1311	2715	0.635

Comparison of Accuracy Between Logistic Regression and SVM

LOGISTIC REGRESSION ACCURACY = **67.86%**

SVM ACCURACY = **59.23%**

Thanks for listening!

Any Questions?

No?

SUPER!