

Priya Patel

MIS 245

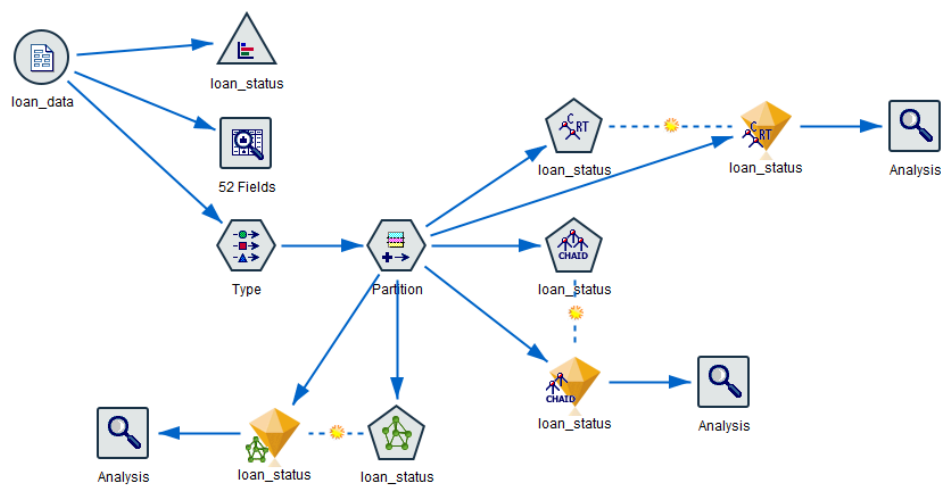
Professor Dantong Yu

April 30, 2021

IBM Predictive Modeler Project

Introduction:

The main goal of this project is to use models to predict whether a customer will default on his loan or not. I have built a stream in SPSS Modeler to help me predict the outcome for customers defaulting their loans. I did a partition of data by doing an 80/20 train test split, which in theory can make the results more accurate for new customers as well as old ones, instead of only the existing ones. The loan status was the designated target value for this project.



Data Description:

The data consists of customers details from the lending clubs database regarding loans and additional categories that would be included in determining their loan status. It includes the customer's loan amount, interest rate, their home ownership status, their fico range, their term, the state they reside in, the amount they paid, the total late fees, the recoveries, their public record bankruptcies, tax liens and more. There are more than 4000 records along with 140 features in the loan records.

Methods:

The models I used for this project totaled in three models and they consisted of:

Rule Induction:

I used two rule induction models which were the CHAID and the C&R Tree.

Black Box:

I used one black box model which was the neural net.

Results:

Model 1 – CHAID

Results for output field loan_status				
Comparing \$R-loan_status with loan_status				
'Partition'	1_Training		2_Testing	
Correct	31,156	99.85%	7,842	99.86%
Wrong	46	0.15%	11	0.14%
Total	31,202		7,853	
Coincidence Matrix for \$R-loan_status (rows show actuals)				
'Partition' = 1_Training		Charged Off	Fully Paid	
Charged Off		4,106	46	
Fully Paid		0	27,050	
'Partition' = 2_Testing		Charged Off	Fully Paid	
Charged Off		1,048	11	
Fully Paid		0	6,794	

Model 2 – C&R Tree

Results for output field loan_status

Comparing \$R-loan_status with loan_status

'Partition'	1_Training		2_Testing	
Correct	31,156	99.85%	7,842	99.86%
Wrong	46	0.15%	11	0.14%
Total	31,202		7,853	

Coincidence Matrix for \$R-loan_status (rows show actuals)

'Partition' = 1_Training		Charged Off	Fully Paid
Charged Off		4,106	46
Fully Paid		0	27,050
'Partition' = 2_Testing		Charged Off	Fully Paid
Charged Off		1,048	11
Fully Paid		0	6,794

Model 3 – Neural Net

Results for output field loan_status

Comparing \$N-loan_status with loan_status

'Partition'	1_Training		2_Testing	
Correct	10,325	33.09%	2,644	33.67%
Wrong	20,877	66.91%	5,209	66.33%
Total	31,202		7,853	

Coincidence Matrix for \$N-loan_status (rows show actuals)

'Partition' = 1_Training		Charged Off	Fully Paid	\$null\$
Charged Off		1,433	91	2,628
Fully Paid		21	8,892	18,137
'Partition' = 2_Testing		Charged Off	Fully Paid	\$null\$
Charged Off		356	20	683
Fully Paid		12	2,288	4,494

