Defection Model

Agenda

- Business Understanding
- Data Dictionary
- ERD
- Code Flowchart
- Model Function
- Prediction Function
- Most At-risk Customers

Business Understanding

Background information:

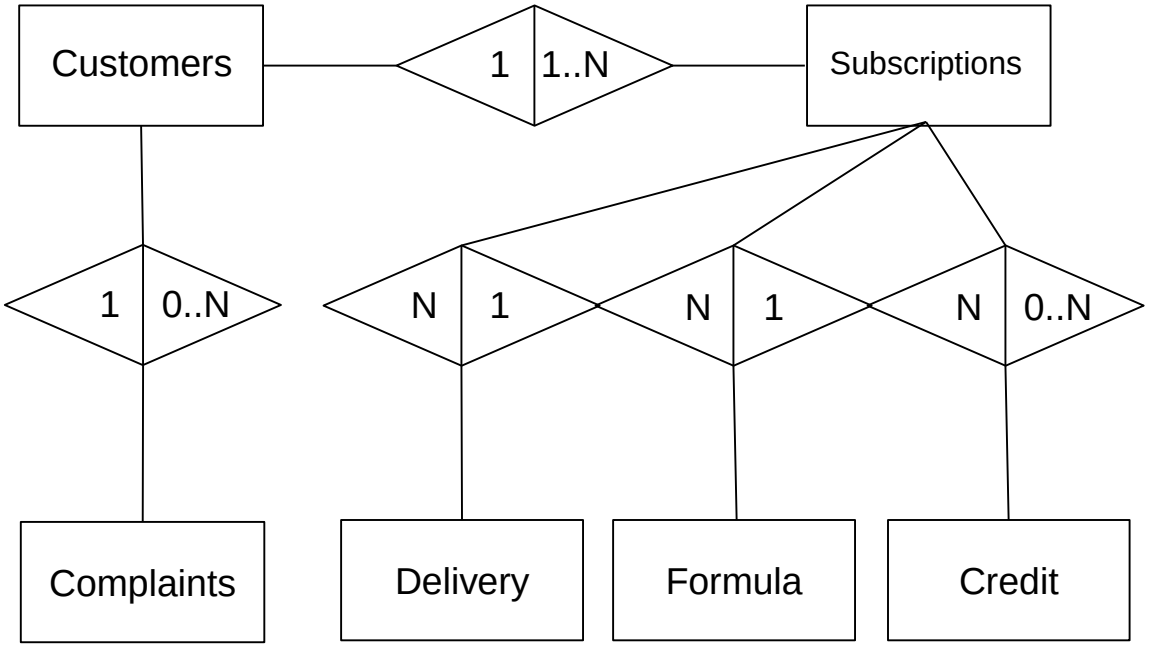
- NPC has been facing increasing churn rates for their newspaper subscriptions

Mission:

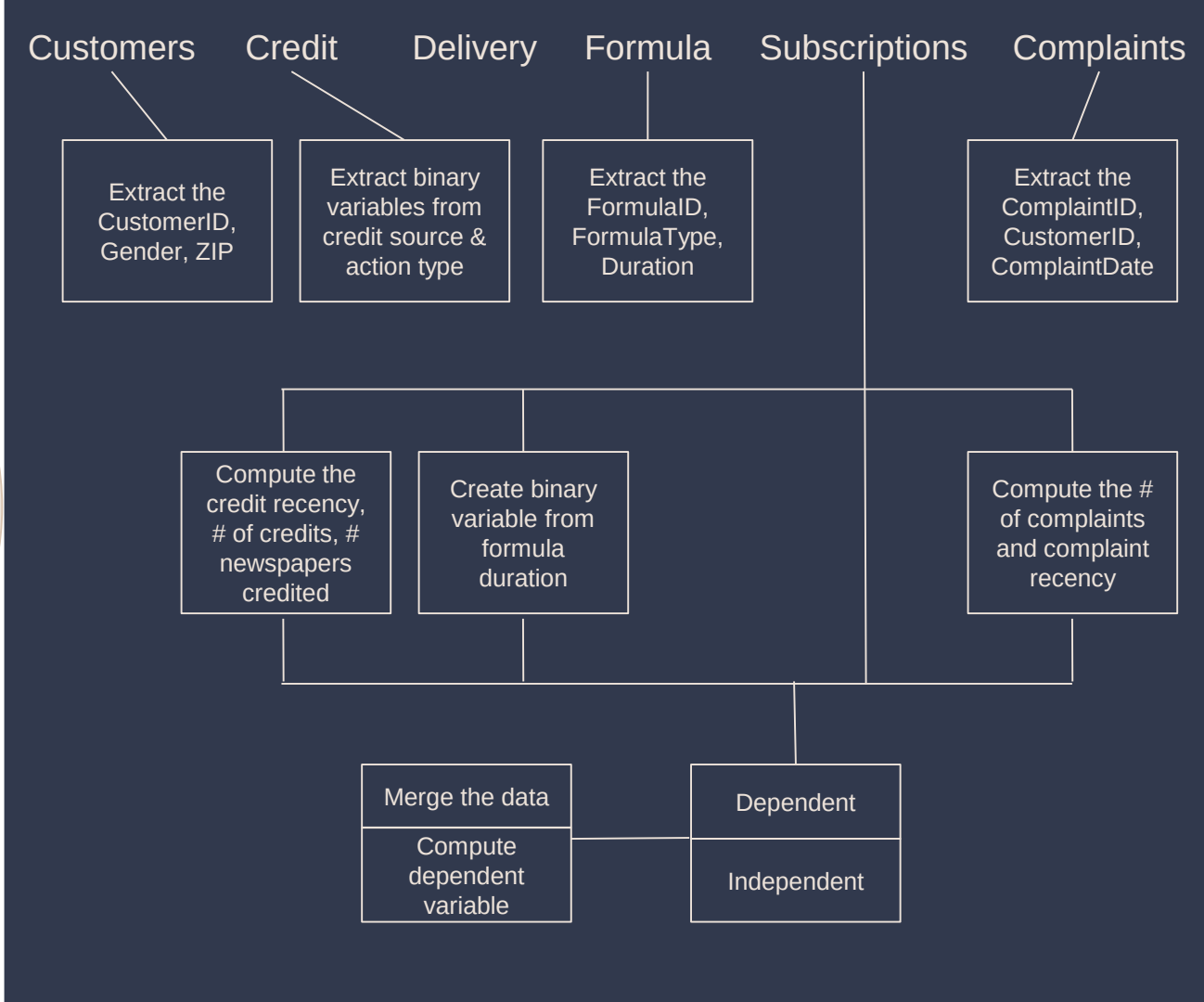
- Create a model to predict which customers are most likely to churn
- Rank customers by churn propensity so NPC can contact the most at-risk customers and try to improve retention

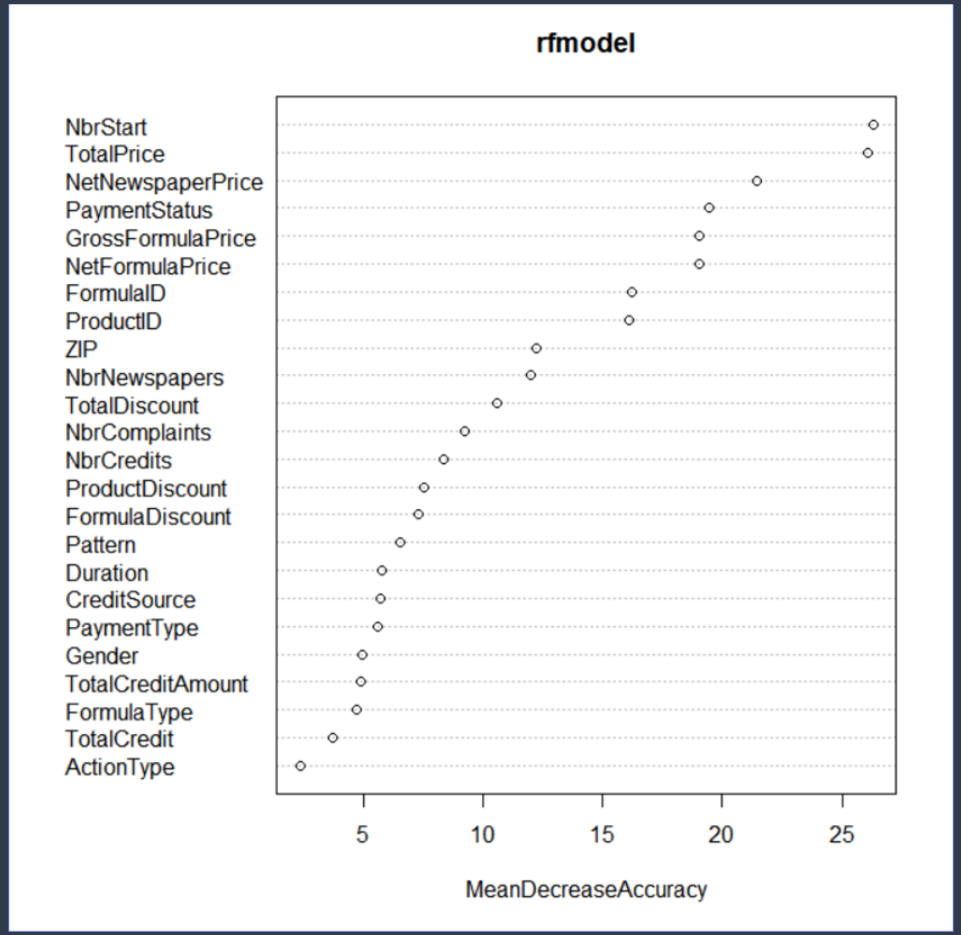
Table	Variable	Definition
Subscriptions	NbrCredits	Shows the total number of credits that are associated with a particular subscriptionID
	TotalCreditAmount	The combined amount credited to the customer for that particular subscriptionID
	CreditRecency	The number of days between when the credit was applied and the max credit date
	ActionType	Factor variable indicating the type of action that was taken by the company to address a credit
	CreditSource	Factor variable the source of the credit
	FormulaType	Shows whether that subscriptionID was a CAM=campaign formula or REG=regular formula
	Duration	Binary variable showing whether the subscriptions duration was >=3 months (1) or <3 months (0)
	Gender	Binary variable showing whether customer associated with the subscription was M=male or F=female
	ZIP	ZIP code of customer delivery site (First 2 digits)
	NbrComplaints	Shows the total number of complaints associated with a particular customerID

ERD


Customers
1
1..N
Subscriptions
1
0..N
N
1
N
1
N
0..N
Complaints
Delivery
Formula
Credit

Code Flowchart




rfmodel
NbrStart
TotalPrice
NetNewspaperPrice
PaymentStatus
GrossFormulaPrice
NetFormulaPrice
FormulaID
ProductID
ZIP
NbrNewspapers
TotalDiscount
NbrComplaints
NbrCredits
ProductDiscount
FormulaDiscount
Pattern
Duration
CreditSource
PaymentType
Gender
TotalCreditAmount
FormulaType
TotalCredit
ActionType
5
10
15
20
25
MeanDecreaseAccuracy

Method: Random Forest

Model Function:

- About 220 lines with comments/spaces
- About 6 seconds to run
- X containing 24 variables
- Y equals customer churn

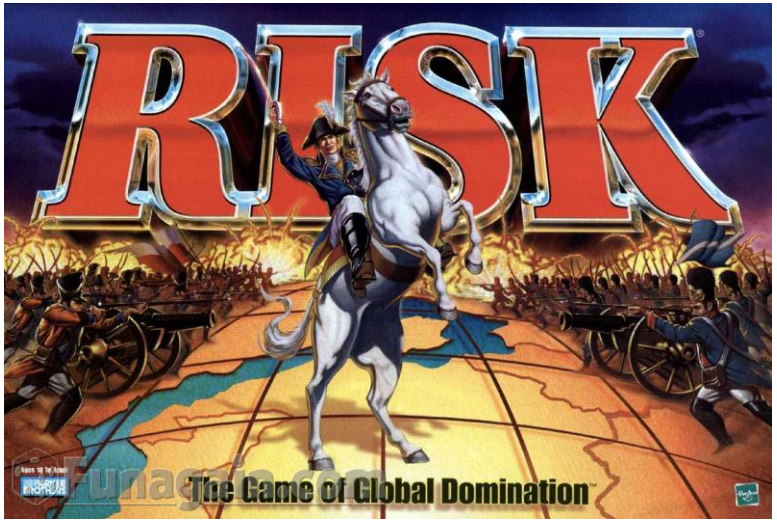
Predict Function:

- About 30 lines with comments/spaces
- About 1 second to run

Performance Measures

AUC: 0.98
Top Decile Lift: 1.52

Most At-risk Customers



Number	CustomerID	Probability
1	674586	0.99
2	659050	0.99
3	659050	0.99
4	477669	0.99
5	659050	0.99
6	1026514	0.99
7	674586	0.99
8	674586	0.99
9	813975	0.99
10	674586	0.99

Questions?