New York University Stern School of Business

Group Assignment #3

Kelly Leong, Lucy Lian, Teddy Le Nguyen, Nancy Toussie
Data Mining For Business
Professor Fernandez
Spring 2020

1. Define the business problem and explain how to use predictive modeling to address the problem.
We want to prevent customer churn and must decide who to target with a $200 retention offer with the objective of minimizing the expected loss for the churn campaign.

To calculate the expected loss due to churn, we need:
1. A customer's annual value, V(x)
2. A customer's probability of churn, p(L|x, NT)

We assume that customers with different contracts (month-to-month, one year, or two years) are affected similarly by the discount offer. As the offer requires the customer to renew a contract for at least one year, we computed the customer's annual value V(x) by multiplying their monthly charges ("MonthlyCharges") by 12, to standardize specific contract terms.

We could then use historical data to build a predictive model to sort customers according to how likely they are to leave prior to the offer, p(L|x, NT).

Since this is a new retention offer, we don't have data to build a model to predict the probability of a customer staying if he or she were to be targeted. Thus, we assume a "perfect" offer, in other words, that a customer will stay if we send an offer. If the cost of churn is greater than the combined cost of the offer and contacting, then we should target that customer with the offer.

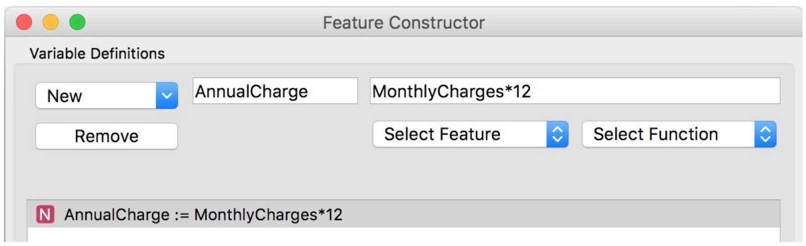
2. Describe the models you considered, the process by which you built and compared models, the evaluation metric(s) you used, and the model(s) you chose.

An instance is a customer of Telco. The target variable is the probability that the customer will leave upon contract expiration without receiving the offer p(L|x, NT). The features are gender (categorical), senior citizen (categorical), partner (categorical), dependents (categorical), tenure (numerical), multiple lines (categorical) internet service (categorical), online security (categorical), online backup (categorical), device protection (categorical), tech support (categorical), streaming TV (categorical), streaming movies (categorical), contract (categorical), paperless billing (categorical), payment method (categorical), and monthly charges (numerical). We should not include PhoneService (categorical) because this is leakage, we would not know this at the time of making the decision.

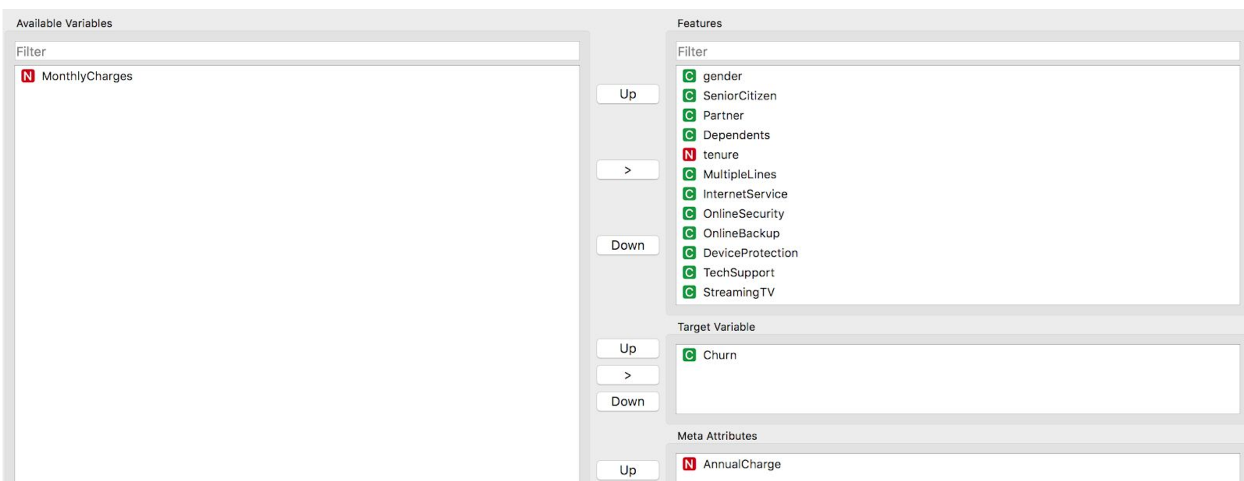
	Name	Type	Role	Values
1	gender	C categorical	feature	Female, Male
2	SeniorCitizen	C categorical	feature	0, 1
3	Partner	C categorical	feature	No, Yes
4	Dependents	C categorical	feature	No, Yes
5	tenure	N numeric	feature	
6	PhoneService	C categori...	skip	No, Yes
7	MultipleLines	C categorical	feature	No, No phone service, Yes
8	InternetServi...	C categorical	feature	DSL, Fiber optic, No
9	OnlineSecurity	C categorical	feature	No, No internet service, Yes
10	OnlineBackup	C categorical	feature	No, No internet service, Yes
11	DeviceProte...	C categorical	feature	No, No internet service, Yes
12	TechSupport	C categorical	feature	No, No internet service, Yes

13	StreamingTV	C	categorical	feature	No, No internet service, Yes
14	StreamingM...	C	categorical	feature	No, No internet service, Yes
15	Contract	C	categorical	feature	Month-to-month, One year, Two year
16	PaperlessBill...	C	categorical	feature	No, Yes
17	PaymentMet...	C	categorical	feature	Bank transfer (automatic), Credit car...
18	MonthlyChar...	N	numeric	feature	
19	**Churn**	C	**categori...**	**target**	**No, Yes**

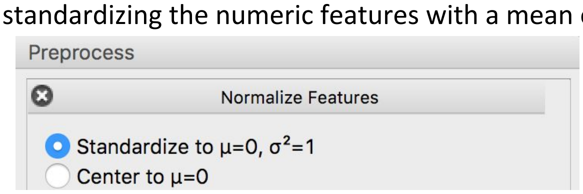
In preparation of the data, we used the Feature Constructor to create a feature called “AnnualCharge” (numerical) by multiplying the MonthlyCharges values by 12.



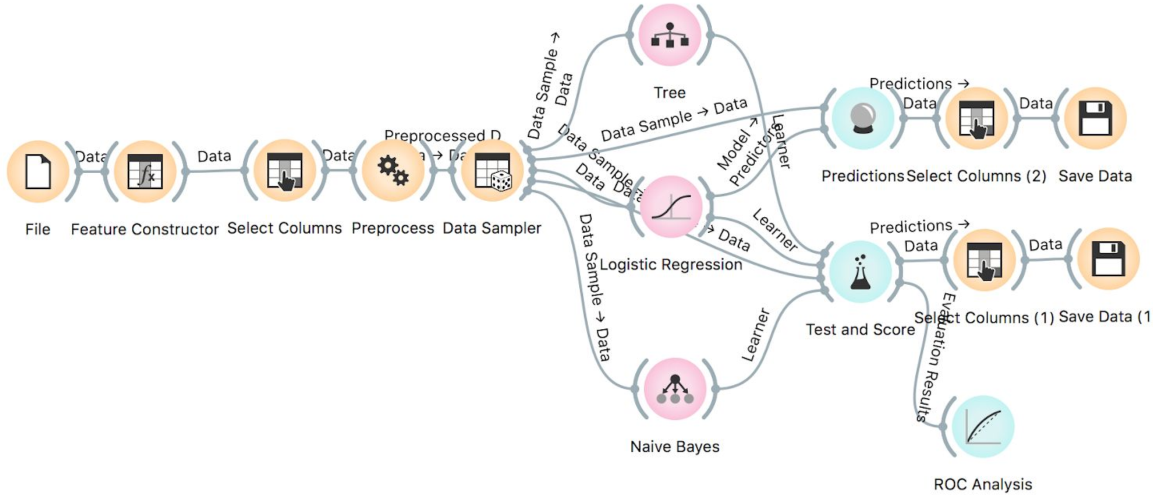
We then used Select Columns to remove MonthlyCharge from the analysis since it's now redundant. We defined AnnualCharge as a meta attribute such that the values will not be normalized by preprocessing.



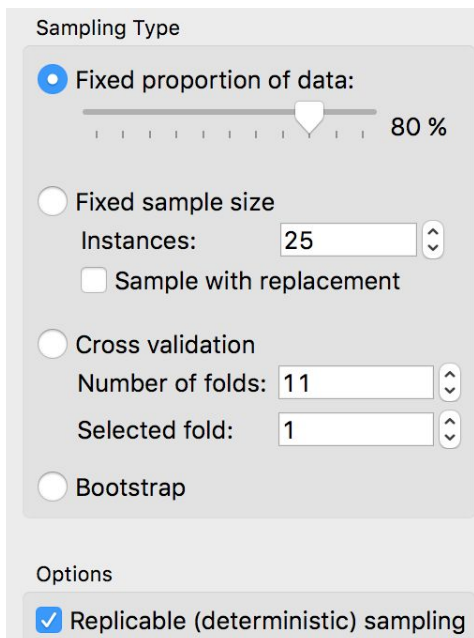
We preprocessed the data, standardizing the numeric features with a mean of 0 and var of 1.



We considered decision trees, logistic regression, and Naive Bayes as our potential models.

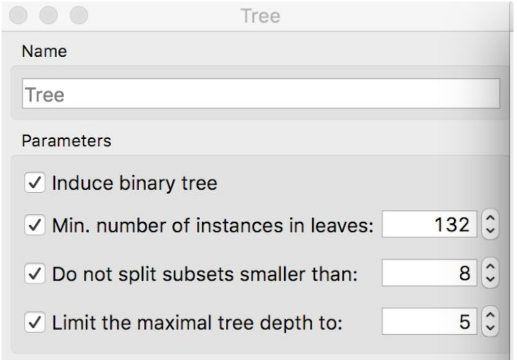


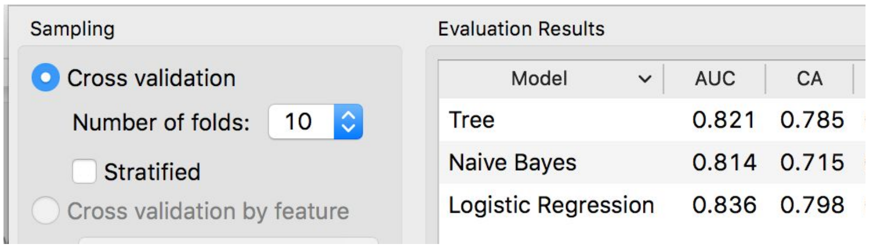
We sampled 80% of the dataset to use for training the models, and the other 20% to test the model. We selected replicable sampling to maintain consistent answers.



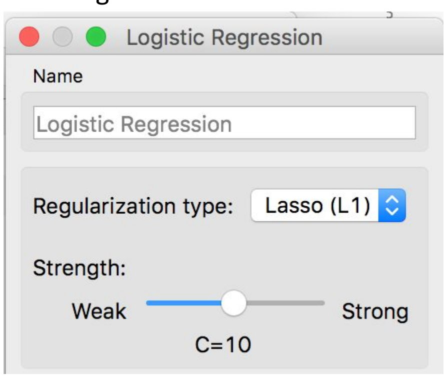
We used cross-validation with 10 folds to compare models with AUC.

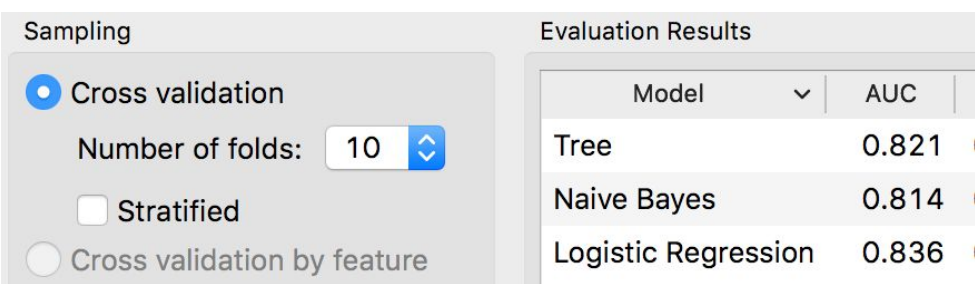
We adjusted in the tree model the minimum number of instances in leaves doubling starting from 2 to 264. We increased the maximal tree depth by 1 incrementally. We found the optimal number was 132 instances and tree depth of 5, yielding the AUC value of 0.821.



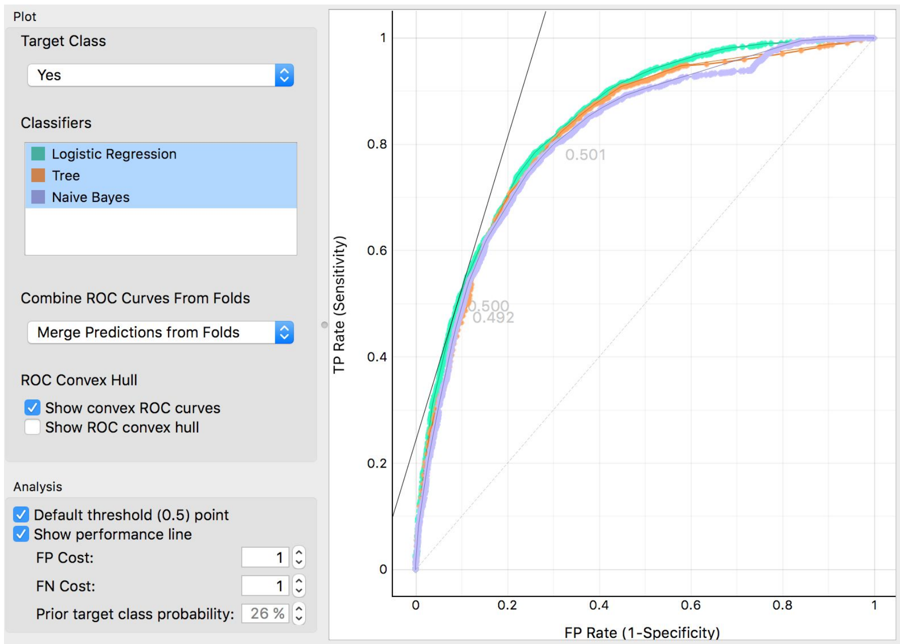


After trying both Ridge and Lasso regularization and a complexity constant of 500, 100, 10, and .013, the max AUC of the logistic regression model of 0.836 ultimately outperforms the other models at C=10 and Lasso regularization.





In the ROC analysis, we can see for the most part the three models closely follow each other. Using the iso performance line, we can see the Logistic Regression again outperforms the other models where it is tangent.



We, therefore, decided to proceed with the Logistic Regression model.

3. Explain how you would use the model to choose who to target with a retention offer.

We use the expected value framework to evaluate our model. Given the cost of the offer (C_o) is \$200 and the cost to contact (C_c) is \$5, we created a cost-benefit matrix.

	Stay (S)	Leave (L)
Target (T)	$V(x) - C_c - C_o = V(x) - 205$	$-C_c = -5$
Not Target (NT)	$V(x)$	0

We want to target a customer with the offer when the value of contacting them EV(T(x)) is greater than the value of not contacting them, EV(NT(x)).

$$EV(T(x)) = (p(S|x, T)*(V(x)-Co-Cc) + (p(L|x, T))*(-Cc)$$
$$EV(NT(x)) = (p(S|x, NT))*(V(x)) + (p(L|x, NT))*(0)$$

Because $(p(L|x, T) = 1 - (p(S|x, T)$:

$$EV(T(x)) = (p(S|x, T)*(V(x)-Co-Cc) + (1-p(S|x, T))*(-Cc)$$
$$EV(NT(x)) = (p(S|x, NT))*(V(x)) + (1-p(S|x, NT))*(0)$$

We want to contact if $EV(T(x)) > EV(NT(x))$:

$$(p(S|x, T)*(V(x)-Co-Cc) + (1-p(S|x, T))*(-Cc) > (p(S|x, NT))*(V(x)) + (1-p(S|x, NT))*(0)$$
$$(p(S|x, T)*(V(x)-Co-Cc) + (1-p(S|x, T))*(-Cc) > (p(S|x, NT))*(V(x))$$

Rearrange the left side to represent the effect of the offer*expected revenue generation:

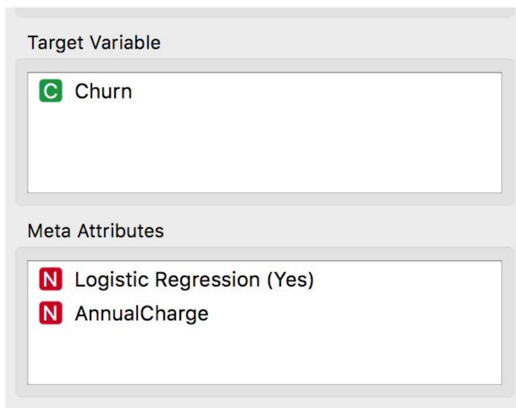
$(p(S|x, T)*(V(x) - (p(S|x, T)*Co - (p(S|x, T)*Cc -Cc + p(S|x, T))*Cc > (p(S|x, NT))*(V(x))$
$[(p(S|x, T)-(p(S|x, NT))]*(V(x)) > (p(S|x, T)Co+Cc$

Finally, we assume that we have a perfect offer so (p(S|x, T) = 1. Therefore:
$(1-p(S|x,NT))*(V(x)) > Co + Cc$
$p\^(L|x,NT)*(V(x)) > Co + Cc$

These values we obtain from the model's output, p(L|x, NT), and the calculation of customer's annual charge, V(x). The cost to target, Co + Cc, totals to $200 + $5 = $205. We can compute the expected loss due to a customer churning and target those whose expected loss is more than our total targeting cost.

4. Measure the potential economic impact of your solution.

We connect the Logistic Regression model and the remaining test data from Data Sampler to the Prediction widget. We exported the variables that we need: Logistic Regression (Yes), Churn, and Annual Charge.



According to the model's predictions on the test data, Telco would send the discount offer to 575 customers whose expected loss is greater than the total targeting cost of $205. The total net benefit would be $176,498.74, resulting in a value of $306.95 per targeting decision made compared to a baseline (targeting everyone) of $39.77. Assuming that the number of customers with the discount offer per year is 1000, the economic impact would be $267,179.60.

Churn	AnnualCharg	Logistic Regr	Expected Los	More than Cost?	Value Saved	Net Benefit				
0	753.6	0.28389585	213.94391	1	$213.94	$8.94	=IF(F2>0,F2-J2,0)	Cost	$205.00	
0	966.6	0.51662395	499.368713	1	$499.37	$294.37		Total Benefit	$176,498.74	=SUM(G2:G4000)
0	1269	0.15556033	197.406054	0	$0.00	$0.00		# of People	1406	=COUNT(E2:E4000)
1	1078.2	0.65143126	702.373179	1	$702.37	$497.37		# of Offers Sent	575	=SUM(E2:E4000)
0	891.6	0.01755119	15.648643	0	$0.00	$0.00				
0	616.2	0.24822836	152.958315	0	$0.00	$0.00		Value per Decision	$306.95	=J3/J5
1	1010.4	0.34076852	344.312509	1	$344.31	$139.31		Baseline	$39.77	=((SUM(D2:D4000))-J2*J4)/J4
0	865.2	0.03405668	29.4658429	0	$0.00	$0.00				
0	1140	0.25638368	292.277392	1	$292.28	$87.28		People per Year	1000	
1	958.8	0.70739325	678.248649	1	$678.25	$473.25		Impact per Year	$267,179.60	=(J7-J8)*J10
0	851.4	0.06274743	53.4231615	0	$0.00	$0.00				
0	904.8	0.41084288	371.730639	1	$371.73	$166.73				
1	1065.6	0.5383721	573.689306	1	$573.69	$368.69				
0	294	0.045084	13.2546947	0	$0.00	$0.00				
0	790.8	0.20926694	165.488293	0	$0.00	$0.00				
0	232.2	0.06400627	14.8622566	0	$0.00	$0.00				

5. Describe the potential reasons (if any) why the proposed solution might not work.

1. We assumed a perfect offer, which is unlikely in real-life applications. The decision is whether to offer a discount, if the customer is leaving for a different reason than price, the offer would likely not be able to convince them to stay.
2. We annualized customer value when the two-year contract term customers may opt to renew for two years, which would mean that our model underestimated the customer value for those with two-year term contracts.
3. We assumed that customers of different payment contracts (month-to-month, annual, two-year) would not react differently to the one-year discount offer. Realistically, a customer currently on a monthly contract would be more likely to change their decision and not churn, or react more positively to the offer than a customer in a two-year contract.
4. We set the AnnualCharge variable as the meta attribute in our model in order to keep its value from being normalized by regularization. In the small chances that such value repeats, we may lose a few instances in our dataset.
5. The Telecom industry is not stable- the criteria that predict churn may have changed from when this historical data was collected (for example, new competitors and new offerings arise), hindering the accuracy of our predictions.
6. Customers who do not receive the offer may hear about it and get upset, leading them to churn.