Customer Propensity Models Explained

Written by
Rhonda Carraway Petty
Marketing Insights Data Scientist
rpetty@mathnetix.com

Customer Propensity Models are equations that predict the relative likelihood a customer will act in a specific way. Marketers use these models to direct marketing resources toward the right customers at the right time.

This paper describes how propensity models are used, what they look like and how they are created. Written in a conversational style, this paper takes the mystery out of predictive modeling for junior analysts and business executives who seek more information on an otherwise very technical topic. Readers interested in the statistical underpinnings of the methodology should seek a statistics book covering Logistic Regression.

WHAT IS A CUSTOMER PROPENSITY MODEL?

A Customer Propensity Model is an equation that predicts the odds a customer will behave in a specific way. The equation is used to calculate a numeric 'score' for each customer where the higher the customer's score, the more likely the customer will engage in the modeled behavior.

Customer Propensity Models are used to predict a future behavior when the future behavior has only two possible outcomes---a 'yes, she did' or a 'no, he did not'. For example:

- Suppose a retailer wants to know who will repeat within a fixed duration of time. Some customers will repeat (the 'yes' group). Other customers will not repeat (the 'no' group). There is no in-between state. Either the customer will repeat or not.

- Suppose a telecommunications company wants to know who will be delinquent on their bill 60 days after the due date. Some customers will become delinquent (the 'yes' group). Other customers will not become delinquent (the 'no' group). There is no in-between state. Either the customer will become delinquent or he will not.

- Suppose a beauty services company wants to know who will try their new manicure service. Some customers will try the new service (the 'yes' group). Other customers will not try the new service (the 'no' group). There is no in-between state. Either the customer will try the new manicure service or she will not.

Because the customer's future behavior will be either a 'yes(1)' or a 'no(0)', the output of a propensity model is a numeric score that falls somewhere between a 0 and a 1. If the customer's score is closer to 0, then the customer is less likely to exhibit the modeled behavior. If the customer's score is closer to 1, then the customer is more likely to exhibit the modeled behavior.

TYPES OF CUSTOMER PROPENSITY MODELS

There are many things Marketers would like to know about the future behavior of individual customers that can be depicted as a 'yes/no' binary outcome. Marketers would like to know things like:

- Who will repeat?
- Who will unsubscribe or cancel service?
- Who will become delinquent or default on a payment?
- Who will stop purchasing?

- Who will respond to a promotional offer?
- Who will try a new product or service?
- Who will upgrade?
- Who will refer a friend?
- Who will adopt new technology?

If a Marketer knows which customers will engage in these activities, then the Marketer will customize strategies specific to the anticipated behavior of individual customers. These strategies have the end goal of encouraging greater customer loyalty and financially-desirable customer behaviors.

WHAT DO PROPENSITY MODELS LOOK LIKE?

At the heart of a Propensity Model is a linear equation that consists of three components:

- Dependent Variable — This is the unknown behavior the Marketer wants to predict
- Independent Variables — These are variables used to predict the unknown behavior
- Parameters — These are numeric values that act as weights on the Independent Variables

Figure 1 illustrates a linear equation with only two independent variables X1 and X2. This figure shows the simplicity of the equation's structure which resembles a weighted average. This equation says the Dependent Variable is a weighted average of two other variables, plus an adjustment factor.

Figure 1: Linear Equation of a Two-Variable Propensity Model

Dependent Variable = Adjustment Factor + Weight $_1$ (X1) + Weight $_2$ (X2)

Suppose a telecommunications company wants to know the likelihood a customer will become delinquent as predicted from other known behaviors and customer attributes. The model development process yields the following linear equation:

Delinquency = 1.25 + 0.83(Average Bill Amount) – 0.20(Number of Months in Good Standing)

The positive weight (0.83) placed on the average bill amount says that as the average bill amount increases, so will delinquency. The negative weight (-0.20) placed on the number of months the customer has been in good standing says that as the good standing period lengthens, delinquency gets smaller. It makes intuitive sense that persons with larger bills are, on average, more likely to become delinquent. It makes intuitive sense that persons without recent payment issues will be, on average, less likely to become delinquent.

The Dependent Variable's value is ultimately transformed into a numeric score that rests between "No/0" and "Yes/1". The score depicts the customer's propensity to exhibit the dependent behavior. The transformation of the Linear Equation into a Probability Score flows as follows:

The odds of a behavior = Probability (behavior) / Probability (no behavior) = Odds Ratio
If the probability of success is the same as the probability of failure, then the odds are equal and the odds ratio = 1.0
If the probability of success is greater than the probability of failure, then the odds are positive and the odds ratio > 1.0

The odds of a behavior = e to the power of the Linear Equation = Odds Ratio
If the Linear Equation calculates to 0, then the odds of the behavior are equal because $e^0 = 1$
If the Linear Equation calculates to 1.5, then the odds of the behavior are positive because $e^{1.5} = 4.48$
The higher the odds, the more likely the behavior

The Odds Ratio is transformed into a propensity score = Odds Ratio / (1 + Odds Ratio)
If the odds ratio = 1.0, then the propensity score = 1 / 2 = 50% or break-even
If the odds ratio = 4.48, then the propensity score = 4.48 / 5.48 = 81.8% or high odds
The higher the odds ratio, the higher the propensity score. But, the propensity score will never fall below 0 nor above 1.

HOW PROPENSITY MODELS ARE DEVELOPED

Step 1: Define the Dependent Variable. Determine the behavior to predict. Be specific. For example:

The Propensity of a Customer To….	
Good	**Better**
Repeat	Repeat Within First 3 Months
Unsubscribe	Unsubscribe to Email In Next 12 Months
Become Delinquent	Become Delinquent In First Year
Refer a Friend	Refer at Least 2 Friends in Next 12 Months

Step 2: Assemble the Development & Validation Samples. Start with a past customer population. Randomly set 30% of the customers aside as a Validation Sample. Assemble the Development Sample from the remaining 70%, so that 50% of the customers in the Development Sample exhibit the behavior and 50% of the customers do not. You will most likely not use the entire 70%. Although the 50/50 split is not a hard requirement and can certainly be relaxed if it produces too small of a Development Sample, the 50/50 split provides an ample number of Yes's and No's, proportionately, from which to perform discriminant analysis.

Step 3. Derive the Candidate Independent Variables. For each customer in the Development and Validation samples (which are rows in the Analytical Data File), derive variables that describe the customer's behavior through the historical point in time where the prediction will be calculated (preserving enough history after this point for comparing your prediction with what actually happened). Be creative. Include variables that intuitively correlate with the behavior to predict. Include behaviors like historical frequency, spend, spend-per-trip and elapsed days since the last purchase. Include behaviors that depict the customer's productivity by product and by purchase channel. Include customer attributes like binary permission flags and tenure. All derived variables must be numeric. So, transform categorical variables such as preferred market and preferred store-type into numeric dummy variables. If a rich supply of behavioral variables are not available, append the customer file with third-party preferences, attitudes and demographics. All said, deriving the Candidate Independent Variables is the most time-consuming step of the entire model development process. But, it is also the most fruitful step. The more variables passed to the model calibration process, the better the model will be, generally speaking. Hint: Build an initial model using CHAID analysis. Use the results to offer suggestions on how to combine levels of a numeric field into new variables and interactions.

Step 4. Transform the Independent Variables. Use the Development Sample to explore the relationship between each independent variable and the dependent variable. See if the two variables are correlated. See if they share a linear relationship. If the relationship is not linear, transform the independent variable to see if this will create a linear relationship or a stronger linear relationship to the dependent variable. The stronger the linear association between the dependent variable and the Candidate Independent Variables, the more likely a good model will emerge. Popular transformations of the independent variable X include $1/X$, $1/X^2$, X^2 and $\ln(X)$.

Step 5. Standardize the Candidate Independent Variables. Modeled output will be easier to interpret if the Candidate Independent Variables are placed on a standard normal scale prior to calibrating the model. A standard normal transformation scales each independent variable so that the mean of its distribution is 0 and the standard deviation is 1. When the Candidate Independent Variables are standardized, the weights placed on each variable by the model are easier to compare relationally. Higher absolute weights reveal variables of greater influence while lower absolute weights reveal variables of lower influence. Without standardization, the weights can vary wildly, just as the unstandardized variables vary wildly, making it difficult to judge the relative significance of a variable from the weight alone. For example, household income is a much larger number than age. Unstandardized, the values for these variables vary wildly just as their unstandardized weights will vary wildly.

Step 6. Calibrate a Logistic Regression Model. The calibration process refers to the Logistic Regression statistical iterations that culminate into a predictive equation. The equation will identify which of the Candidate Independent Variables predict the Dependent Behavior and to what degree. Most statistical software packages have built-in procedures for Logistic Regression, giving the Marketing Scientist options for controlling various elements of the model calibration process. The difficult calculations are performed by the imbedded Logistic Regression procedure. But, the Marketing Scientist must navigate the results and experiment in search of the best possible model. After a series of iterations and experiments, the Marketing Scientist will conclude with a final equation for predicting behavioral propensity. This equation highlights the independent variables in the model and the weight each variable retains.

Step 7. Validate the Model. Score the customers in the Validation Sample using the equation generated by the model from the Development Sample. Sort the customers in the Validation Sample so that those with the highest predicted scores are at the top of the file. Then, produce a Gains Chart to depict the model's ability to successfully shift customers with the true behavior to the top of the file. This is how model success is measured---in its ability to successfully sort the customer file--shifting the yes customers to the top of the file and the no customers to the bottom of the file.

In Figure 2, the customers of the Validation Sample have been sorted so that customers with the highest *predicted* scores are at the top of the file. Of all of the customers in the entire Validation Sample with the true desired behavior, 20.8% are found in the top predicted decile. This compares to 10% that could have been expected in this decile under a 'no-model' random-sort scenario. On a cumulative basis, the model most outperforms a random sort through the 4th decile where the differential between the modeled sort and a random sort is the highest at 17.2%. If the example of Figure 2 was predicting the customer's likelihood to respond to a promotion, the Marketer would target the top 40% most likely to respond because this is where the lift of the model over a random-sort is maximized. Figure 3 illustrates this outcome graphically. The gap between the red random line and the green predicted line is maximized at the 4th predicted decile.

Figure 2. Gains Chart Table

	Predicted Decile	No Model	With Model	Cumulative Lift	Differential
Most Likely	1	10.0%	20.8%	208	10.8%
	2	20.0%	35.5%	188	15.5%
	3	**30.0%**	**47.2%**	**172**	**17.2%**
	4	**40.0%**	**57.2%**	**161**	**17.2%**
	5	**50.0%**	**66.5%**	**151**	**16.5%**
	6	60.0%	74.9%	144	14.9%
	7	70.0%	82.6%	137	12.6%
	8	80.0%	89.5%	132	9.5%
	9	90.0%	95.5%	127	5.5%
Least Likely	10	100.0%	100.0%	122	0.0%

Figure 3: Gains Chart Graph