

2021

Profitable Retail Customer Identification Based on a Combined Prediction Strategy of Customer Lifetime Value

Yinglu Sun

National Academy of Innovation Strategy, CAST

Dong Cheng

Renmin University of China

Subir Bandyopadhyay

Indiana University Northwest

Wei Xue

Renmin University of China, xuewei@ruc.edu.cn

Follow this and additional works at: <https://scholar.valpo.edu/mssj>



Part of the [Anthropology Commons](#), [Business Commons](#), [Criminology Commons](#), [Economics Commons](#), [Environmental Studies Commons](#), [Gender and Sexuality Commons](#), [Geography Commons](#), [History Commons](#), [International and Area Studies Commons](#), [Political Science Commons](#), [Psychology Commons](#), and the [Urban Studies and Planning Commons](#)

Recommended Citation

Sun, Yinglu; Cheng, Dong; Bandyopadhyay, Subir; and Xue, Wei (2021) "Profitable Retail Customer Identification Based on a Combined Prediction Strategy of Customer Lifetime Value," *Midwest Social Sciences Journal*: Vol. 24: Iss. 1, Article 10.

DOI: <https://doi.org/10.22543/0796.241.1053>

Available at: <https://scholar.valpo.edu/mssj/vol24/iss1/10>

This Article is brought to you for free and open access by ValpoScholar. It has been accepted for inclusion in Midwest Social Sciences Journal by an authorized administrator of ValpoScholar. For more information, please contact a ValpoScholar staff member at scholar@valpo.edu.

***Profitable Retail Customer Identification Based on a Combined
Prediction Strategy of Customer Lifetime Value****

YINGLU SUN

National Academy of Innovation Strategy, CAST

DONG CHENG

Renmin University of China

SUBIR BANDYOPADHYAY

Indiana University Northwest

WEI XUE

Renmin University of China

ABSTRACT

As a fundamental concept of customer relationship management, customer lifetime value (CLV) serves as a crucial metric to identify profitable retail customers. Various methods are available to predict CLV in different contexts. With the development of consumer big data, modern statistics and machine learning algorithms have been gradually adopted in CLV modeling. We introduce two machine learning algorithms—the gradient boosting decision tree (GBDT) and the random forest (RF)—in retail customer CLV modeling and compare their predictive performance with two classical models—the Pareto/NBD (HB) and the Pareto/GGG. To ensure CLV prediction and customer identification robustness, we combined the predictions of the four models to determine which customers are the most—or least—profitable. Using 43 weeks of customer transaction data from a large retailer in China, we predicted customer value in the future 20 weeks. The results show that the predictive performance of GBDT and

* Correspondence concerning this article should be addressed to Wei Xue, *School of Statistics, Renmin University of China, Beijing 100872, China*; xuewei@ruc.edu.cn.

This study is supported by funds for building world-class universities (disciplines) of Renmin University of China. The third author is financially supported by the Byron Root Foundation Grant.

The authors express sincere thanks for the support of Dr. Dehong Yang, chairman of Miya Information Technology Inc.

RF is generally better than that of the Pareto/NBD (HB) and Pareto/GGG models. Because the predictions are not entirely consistent, we combine them to identify profitable and unprofitable customers.

KEY WORDS Customer Lifetime Value (CLV); Pareto/NBD (HB); Pareto/GGG; Gradient Boosting Decision Tree (GBDT); Random Forest (RF)

With the development of relationship marketing, customer relationship management (CRM) has been widely studied by academics and industry insiders. Many believe that the primary task of CRM is to identify, satisfy, and retain the most profitable customers to reduce costs and increase revenues. Customer profitability should not be judged by a customer's single transaction with the firm but rather by a series of transactions or potential transactions (i.e., a customer's lifetime income stream; Buttle 2004). As a result, customer lifetime value (CLV) becomes a fundamental CRM concept and a crucial metric in relationship marketing (Rust, Zeithaml, and Lemon 2000). CLV is defined as the net present value of all streams of contributions to profit resulting from a customer over his or her entire life of transactions with the firm (Jain and Singh 2002). Because not all customers are profitable and financially attractive to firms, CLV works as a metric to segment customers, allocate resources, and formulate related strategies (Zeithaml, Rust, and Lemon 2001). In the early 1990s, companies emphasized the importance of measuring and managing customer satisfaction and customer loyalty, but researchers found that customer satisfaction, customer profit, and the relationship between customer loyalty and customer profit are not as strong as anticipated (Reinartz and Kumar 2000, 2002). Satisfied customers and loyal customers are not always the most profitable customers (Kumar 2008). Many companies have seriously misallocated resources by taking customer satisfaction or customer loyalty as a simple proxy measure for customer profit. It is therefore essential to measure customer-level profitability, and CLV is ultimately required for making good marketing decisions. According to CLV, firms can allocate resources and establish long-term relationships with the "right" customers.

CLV works as a foundation for companies to make marketing strategies concerning customer acquisition, customer retention, and customer win-back. It is essential to accurately predict CLV and identify the most profitable customers. The misestimate of CLV may lead to the wasting of limited marketing resources and the mismanagement of customers. Marketing researchers have proposed various methods to predict CLV in different contexts, including the Pareto/NBD (negative binomial distribution) model (Schmittlein, Morrison, and Colombo 1987), the logit/probit model (Thomas 2001), the hazard rate model (Meyer-Waarden 2007), and the Markov chain model (Bandyopadhyay 2009; Pfeifer and Carraway 2000). Although improving prediction accuracy can never be overemphasized, it's still the main task for researchers in this field. Besides, with advanced data-analysis techniques, machine learning algorithms have been gradually adopted by researchers in customer behavior analytics. The prediction performance of these new algorithms deserves further exploration and study. In our study, we introduce two machine learning algorithms—the gradient boosting decision tree (GBDT) and the random forest

(RF)—in retail customer CLV modeling and compare their predictive performance with that of two classical probability models—the Pareto/NBD (HB) (Abe 2009; Ma and Liu 2007) and the Pareto/GGG (gamma-gamma-gamma; Platzer and Reutterer 2016). Each of the four algorithms has its pros and cons in CLV prediction (Table 1). Pareto/NBD (HB) is a hierarchical Bayes extension to the Pareto/NBD model that is well known for describing customer purchasing behavior in a noncontractual context. Pareto/GGG is another generalization of the Pareto/NBD model by considering the regularity of customer interpurchase timing. GBDT and RF have integrated learning models. They are representative prediction methods in machine learning, and both of them can effectively improve prediction accuracy. The purpose of our study is to establish a framework to identify profitable retail customers based on their CLV. To ensure the robustness of customer identification, we predict CLV by the four aforementioned models and combine their predictions to determine which customers are the most profitable for the firm.

The next section presents a literature review of CLV modeling approaches. Section 3 explains the basic logic of Pareto/NBD (HB), Pareto/GGG, GBDT, and RF. Section 4 outlines the empirical application based on customer transaction data of a retailer, and Section 5 presents our summary and conclusion.

MODELING CLV

CLV is a forward-looking metric that considers a customer's future behaviors and enables firms to treat individual customers differently according to their contributions (Kumar and Reinartz 2016). Researchers have developed various CLV models that, in general, can be divided into two different types: deterministic and stochastic. Deterministic CLV analysis adopts simplified calculations and uses formulas without any stochastic components, ignoring individual customers' heterogeneity (Estrella-Ramón et al. 2013). Jain and Singh (2002) described the basic CLV deterministic model as

$$CLV = \sum_{i=1}^n \frac{(R_i - C_i)}{(1+d)^{i-0.5}},$$

where i represents the period of cash flow from a customer transaction, R_i is the revenue from the customer in period i , C_i is the total cost of generating the revenue in period i , n is the total number of periods of projected life of the customer under consideration, and d is the discount rate. Deterministic CLV analysis is more basic and general and has fewer variations. It is often adopted in contractual settings, such as telecommunications or magazine subscriptions. Stochastic CLV modeling approaches view the observed customer behavior as realizing an underlying stochastic process, thus emphasizing customer heterogeneity. As a result, this type of model brings more precision to CLV estimation (Estrella-Ramón et al. 2013). Stochastic CLV analysis is usually adopted in noncontractual settings such as retailing. Generally, two types of stochastic CLV modeling methods are based on deductive reasoning and the other based on inductive reasoning.

Table 1. Advantages and Disadvantages of the Pareto/NBD (HB), Pareto/GGG, GBDT, and RF Algorithms

Algorithm	Similarity	Characteristic	Advantage	Disadvantage
Pareto/NBD (HB)	Probability models	Combines Pareto and NBD (HB) models to predict customer churn and purchase behavior	Very stable prediction performance	Not sensitive to extreme data
Pareto/GGG		Considers regularity of customer interpurchase timing and uses gamma distribution to describe this regularity	Very stable prediction performance; considering regularity of customer interpurchase timing can effectively improve prediction accuracy	Not sensitive to extreme data
GBDT	Combination forecasting model, also known as integrated learning model	For the loss function, finds the current optimal tree through continuous iteration	Can effectively improve prediction accuracy and has fast convergence speed	May lead to overfitting model
RF		For the loss function, the diversity enhancement strategy is used to build multiple unrelated trees	Can effectively improve prediction accuracy and reduce prediction variance	Convergence speed is relatively slow

Stochastic CLV Models Based on Deduction

Deduction and induction, which can be traced to Greek antiquity, are two reasoning patterns for scientific inquiry. In the research field of marketing and customer behavior, deductive reasoning techniques have been dominant, with researchers building and testing hypotheses to find answers (Lawson 2005). The study of CLV is no exception, as many stochastic CLV modeling methods were built on existing knowledge and theories. These methods include probability, econometric, and persistence models (Estrella-Ramón et al. 2013; Gupta et al. 2006). Probability models adopt probability distributions to model observed customer behaviors such as purchase frequency and contribution margin. Two widely recognized probability models are Pareto/NBD (Schmittlein, Morrison, and

Colombo 1987) and BG/NBD (beta geometric/negative binomial distribution; Fader, Hardie, and Lee 2005a). Many econometric models share the same underlying logic as probability models (Gupta et al. 2006). Unlike probability models using a probability distribution to describe customer behaviors, econometric models focus on explaining different customer responses as a function of covariates. Typical econometric models include a simple regression model (Venkatesan and Kumar 2004), a logit/probit model (Thomas 2001), and survival analysis (Meyer-Waarden 2007). When the customer data time series is long enough, persistence models are suitable for CLV estimation. Similar to the econometric models, persistence models emphasize CLV drivers; however, these covariates are considered part of a dynamic system, and their movements over time affect CLV in the long run. Persistence models are based on the development of multivariate time series analysis, such as VAR models (Bandyopadhyay 2009), unit roots, and cointegration (Gupta et al. 2006).

Stochastic CLV Models Based on Induction

Although deductive reasoning has been widely used in scientific inquiry (Lawson 2005), an increasing number of researchers believe that deductive reasoning techniques have limitations in analyzing big data (Erevelles, Fukawa, and Swayne 2016; Lycett 2013). According to Erevelles, Fukawa, and Swayne (2016:900), knowledge-based deductive reasoning “result[s] in considerable linear growth in understanding marketing phenomena about which much is already known, at the expense of nonlinear advances in understanding marketing phenomena about which little or nothing is known,” thus hindering the search for new information and insight. On the contrary, ignorance-based inductive reasoning enables researchers to observe a phenomenon before forming any hypotheses and to mathematically identify the hidden patterns in customer big data (Lycett 2013). Inductive CLV modeling approaches are usually based on computer science development, especially data mining, machine learning, and nonparametric statistics. These models include the GBDT, the generalized additive model (GAM), the RF, the support vector machine (SVM), and the neural network model, among others. Compared with deductive models based on theory and easy to interpret, inductive models based on computer science often have better predictive abilities (Gupta et al. 2006). Studies show that SVM, GAM, and a multivariate decision tree all provide more accurate predictions than a logit model (Coussement, Benoit, and Van den Poel 2010; Cui and Curry 2005). According to Gupta and colleagues (2006), these inductive machine learning models need further exploration in the field of CLV prediction, especially in the age of consumer big data.

Although various CLV models have been developed, no well-accepted prediction model suits all situations, despite many researchers' comparisons. For example, Vafeiadis et al. (2015) compared an artificial neural network, a support vector machine, a decision tree, a naïve Bayes, and a logistic regression for customer churn prediction and found that SVM performed best. Martínez and colleagues (2018) used machine learning algorithms including logistic lasso regression, extreme learning machine, and gradient tree boosting to predict customer purchases, and gradient tree boosting performed best. Current research suggests that the prediction performance of different models depends on different situations

and datasets and that no well-recognized CLV model performs best under all circumstances. Customer identification based on one CLV prediction model may therefore be biased. In this paper, we combine the prediction results of four CLV models, two probability models—Pareto/NBD (HB) and Pareto/GGG—and two machine learning models—GBDT and RF—to ensure the robustness of profitable customer identification.

MODEL SPECIFICATION

Retailing is a typical noncontractual context, and the relationship between retail customers and the firm is uncertain. Modeling retail customer CLV can be very challenging because customer defection is not observable; therefore, the key point of retail customer CLV estimation is to predict customers' future purchase behaviors. It is strongly suggested that the three variables of the RFM model—recency, frequency, and monetary—are sufficient to describe an individual customer's purchase history and that customers' past purchases act as good predictors of their future purchases (Bandyopadhyay and Martell 2007; Estrella-Ramón et al. 2013; Fader, Hardie, and Lee 2005b). RFM can therefore provide a solid foundation for CLV modeling.

This paper emphasizes the prediction of a customer's future purchase frequency based on Pareto/NBD (HB), Pareto/GGG, GBDT, and RF. We compare their predictive performance and recommend combining the four models to predict purchase frequency and further identify the most profitable customers to ensure the robustness of the results. The monetary value of a customer's future purchases was estimated based on a normal distribution (Schmittlein and Peterson 1994). We assume that all the retail customers have the same acquisition cost and direct cost. Without considering the discount rate, the calculation of CLV can be simplified as the product of a customer's future purchase frequency and monetary value. We believe that a retail customer CLV prediction strategy combining classical probability models with machine learning approaches will be a future research direction for CLV and CLV-based customer identification. Below, we briefly introduce Pareto/NBD (HB), Pareto/GGG, GBDT, and RF before describing our empirical study.

Pareto/NBD (HB)

Pareto/NBD (HB) is a hierarchical Bayes extension to the Pareto/NBD model (Abe 2009; Ma and Liu 2007). Developed by Schmittlein, Morrison, and Colombo (1987), the Pareto/NBD is a well-recognized model that describes customer purchasing behavior in a noncontractual context. Pareto/NBD (HB) extends the Pareto/NBD model using a hierarchical Bayesian (HB) framework. The Pareto/NBD (HB) model obtains the posterior value based on the prior parameters and data likelihood. A hierarchical Bayesian version of the NBD model of transactions x is

$$\pi(r, \alpha, \{\lambda_i\} | \{x_i\}) \propto \prod_i p(x_i | \lambda_i) g(\lambda_i | r, \alpha) \pi(r) \pi(\alpha),$$

where λ_i is the purchase rate of customer i and $\pi(r)$ and $\pi(\alpha)$ are prior distributions on the parameters of the gamma distribution of λ_i (Jen, Chou, and Allenby 2003). The left side is posterior, the last three factors on the right side form the prior distribution, and $p(x_i | \lambda_i)$ is the data likelihood. Thus can we estimate not only the parameters of the purchase rate of the customer cohorts but also the individual purchase rate by integrating the joint posterior density:

$$\pi(\lambda_j | \{x_i\}) = \int \dots \int \pi(r, \alpha, \{\lambda_i\} | \{x_i\}) dr d\alpha d\lambda_{-j},$$

where $-j$ denotes all customers except customer j . Pareto/NBD (HB) can be estimated by the Markov chain Monte Carlo (MCMC) method.

Pareto/GGG

Pareto/GGG (Platzer and Reutterer 2016) is another generalization of the Pareto/NBD model. It notes that the regularity of customer interpurchase timing can effectively improve prediction accuracy. Pareto/GGG assumes that the intertransaction timing $\nabla t_j = t_j - t_{j-1}$ follows the gamma distribution with shape parameter k and rate parameter $k\lambda$ —that is, $\nabla t_j \sim \text{Gamma}(k, k\lambda)$ —when the customer remains alive. Here, λ determines the frequency and k determines the regularity of intertransaction timings. There are also differences in individual intertransaction timing, which follows the gamma distribution. This model is called the Pareto/GGG because the individual-level parameters of the purchase process follow three gamma distributions: $k \sim \text{Gamma}(t, \gamma)$; $\lambda \sim \text{Gamma}(r, \alpha)$; $\mu \sim \text{Gamma}(s, \beta)$. Pareto/GGG parameters can also be estimated by the MCMC method.

Gradient Boosting Decision Tree

A GBDT is one of the representative prediction methods in machine learning and has an outstanding performance in combination forecasting. The basic learner of GBDT is usually the classification and regression tree (CART; Breiman et al. 1984). CART avoids the linear assumption in traditional statistical models and can find the nonlinear relationship between the dependent and independent variables, thus effectively improving prediction accuracy. The CART (including a classification tree and regression tree used later in this paper) algorithm includes two processes: tree growth and tree pruning. Tree growth is a multi-iteration grouping process for training datasets. The “excessive” growth of trees can be limited by pre-pruning strategies that specify the maximum depth of trees and the sample size of tree nodes or by minimal cost-complexity pruning strategies after the tree grows. The principle of minimum test error determines the optimal tree. Generally, the test errors (out-of-bag errors) can be estimated by an N-folds cross-validation method.

Random Forest

RF is also a combination forecasting strategy consisting of several CARTs with high accuracy and weak correlation or even irrelevance. Forecasting is achieved by a tree's voting or averaging. The randomness of RFs is reflected in the two aspects (sample randomness and variable randomness). Using the strategy of bagging (bootstrap aggregating), multiple trees are built based on independent random samples. Independent samples are obtained by a resampling bootstrap method. A random sample (called a bootstrapping sample) with a sample size of n is obtained by repeated sampling B times with playback from the train dataset with an n of the same size. In the process of tree building, a few input variables are selected randomly to form a subset of variables Θ . Only the explanatory variables entering the subset Θ have the chance to become bin-variables to prevent multiple CARTS from being highly correlated.

We can see that the basic starting point of the classic customer behavior prediction model represented by Pareto/NBD (HB) and Pareto/GGG is that the distribution of customers' historical purchase behavior will remain unchanged in the future; therefore, the assumptions of customer purchase behavior and its distribution form are crucial for establishing the deductive models. As long as the distribution parameters of customers' previous purchases are obtained, the future behavior can be predicted based on the distribution function. Only a few variables—including the historical number of purchases (F), recent purchases (R), monetary value of purchases (M), and observation period (T)—are needed to identify distribution parameters and develop distribution function. Although many scholars (Abe 2009; Fader, Hardie, and Shang 2010; Ma and Büschken 2011) have proposed a variety of improved models based on the revision of the assumptions, there is no significant change in the basic modeling framework. These models belong to unsupervised learning in terms of modern statistics (i.e., the parameters of the prediction model are estimated without supervision of the customers' future purchase behavior). Pareto/NBD (HB) and Pareto/GGG are therefore suitable to describe stable purchase behavior without much consideration of purchase fluctuations.

Machine learning methods bring us new thoughts and ideas regarding customer behavior prediction, however, and they can work as an important supplement to the classic deduction-based customer behavior prediction models. Machine learning algorithms such as GBDT and RF are supervised methods, meaning that parameter estimation is carried out based on customers' historical behavior. These models directly reflect the nonlinear relationship between customers' purchase histories and future purchases with no need for the assumptions of customer purchase behavior, which is suitable to capture unstable and unconventional purchases. Besides, we can introduce more related variables into the explanatory variable set; therefore, modeling customer behavior is no longer limited to a few variables such as R, F, T, and M. More variables describing the characteristics of customer behavior can be introduced into the model, but it is critical to decide which

variables should be contained in the model's explanatory variable set. In our study, we used four groups of explanatory variables in GBDT and RF.

First, we gained insights from Pareto/NBD (HB) and Pareto/GGG that customers' historical purchase behavior would determine the number of their future purchases. We therefore kept the classic variables that describe customers' historical purchases—including F, R, M, and T—in the GBDT and RF models.

Second, we added variables that described customers' recent historical purchases. Because remote purchase history is less useful for predicting the future, we introduced the more recent purchases—including monetary value and purchase intervals—into the models. These variables can reflect whether the customer has frequently purchased or has stocked up recently because of a sales promotion. If so, the possibility of frequent purchases in the near future will be reduced.

The level of historical purchasing power was also an important factor in determining future purchases; therefore, we next introduced variables that reflected customers' purchasing power during a certain time. One of the most important variables was the accumulated monetary value of purchases in a given time. If two customers had the same F and R for a period, the customer with the higher accumulated monetary value of purchases should have a higher purchasing power.

Finally, we believe that the number of future purchases is closely related to intertransaction timing (i.e., purchase intervals), which can depict customers' regularity and effectively improve prediction accuracy. Customers with shorter time intervals between purchases must have a different number of purchases from customers with longer time intervals between purchases in the same future period. We thus introduced variables into the model that described the purchase intervals.

All of the explanatory variables of the GBDT and RF algorithms are shown in Table 2.

EMPIRICAL APPLICATION

Dataset

The data we used in this empirical study came from a large retailer in China. We used 114,973 pieces of transactional data from 25,800 customers in a 43-week period from July 25, 2017, to May 20, 2018. The preliminary data processing was done in the following steps.

Step 1: We reorganized the data by combining all of a customer's transactions in the same week and then calculating the variables in Table 2.

Step 2: We divided the dataset into two time periods. The first period was July 25, 2017, through January 1, 2018 (23 weeks), and the second was January 2, 2018, through May 20, 2018 (20 weeks). For Pareto/NBD (HB) and Pareto/GGG, we used the data from the first period to estimate model parameters and the monetary value of an individual customer's future purchases. Furthermore, based on the estimated model parameters and monetary value of the first period, we predicted a customer's number of purchases and then calculated the CLV of the second period. For GBDT and RF, we used variables listed in Table 2 from the first-period dataset as explanatory variables and the number of purchases

from the second-period dataset as the explained variable to build the models and estimate the parameters. Then we predicted a customer’s number of purchases and calculated CLV for the second period.

Step 3: After data cleaning, we randomly selected a portion of the customers who survived to the 43rd week and observed the variable distribution.

Table 2. Explanatory Variables of the GBDT and RF Algorithms

Variables	Variable Description	Classification
F	Number of purchases that describes customers’ historical purchase frequency	Classical variables
R	Latest purchase that describes the time period since customers’ last purchases	
M	Monetary value of purchases that describes customers’ purchasing power	
T	Observation period that describes customers’ survival time	
R.1	Interval (weeks) between customers’ last and third-to-last purchases	Variables describing the purchase intervals of recent historical purchases
R.2	Interval (weeks) between customers’ last and second-to-last purchases	
sale.1	Monetary value of customers’ third-to-last purchases	Variables describing the monetary value of recent historical purchases
sale.2	Monetary value of customers’ second-to-last purchases	
sale.3	Monetary value of customers’ last purchases	
sale.sum.1	Accumulated monetary value of customers’ last three purchases	Variables describing the purchasing power of customers during a certain time
sale.sum.2	Accumulated monetary value of customers’ last two purchases	
timespace.mean	Mean of the intertransaction timing that reflects the average level of customer purchase intervals	Variables describing customers’ purchase time intervals
timespace.sd	Standard deviation of the intertransaction timing that reflects the fluctuation of customer purchase intervals	
timespace.max	Maximum value of the intertransaction timing that reflects the extreme case	
litt	Sum of logarithmic intertransaction timing that reflects the overall level of customer purchase intervals	

The line chart in Figure 1 shows changes in the number of customer purchases during the 43 weeks; the solid black line is the total number of purchases, and the red dashed line is the number of repeat purchases. The number of purchases reached its peak in weeks 22 to 25, when the retailer frequently promoted at the end of the year, and dropped to the lowest point in May 2018. The boxplot in Figure 2 shows customer intertransaction timing for the first period. The median for the third-to-last time interval—the time between the third-to-last purchase and the last purchase—was about four weeks. The median for the second-to-last time interval—the time between the second-to-last purchase and the previous purchase—was about two weeks. The variance for the third-to-last time interval was relatively larger than that of the second-to-last time interval. The distributions of both time intervals were right-skewed, with fewer customers having longer purchase-time intervals. The last time interval shows the time interval since the last purchase. The median was about three weeks, with relatively large variance, and the number of customers who had not purchased for a long time was relatively small.

Figure 3 shows the distribution of T and F. As shown in Figure 3A, during the 23 weeks of the first period, only 6 percent of customers bought for the first time in the latest four weeks (i.e., survival time less than four weeks) and only 5 percent of customers survived more than 23 weeks. Most of the customers had a survival time of about 10–20 weeks. Figure 3B shows that during the first period, 40 percent of customers purchased in this retail store fewer than five times, 38 percent purchased six to ten times, and only 3 percent purchased more than fifteen times.

Figure 1. Customer Purchase Frequency

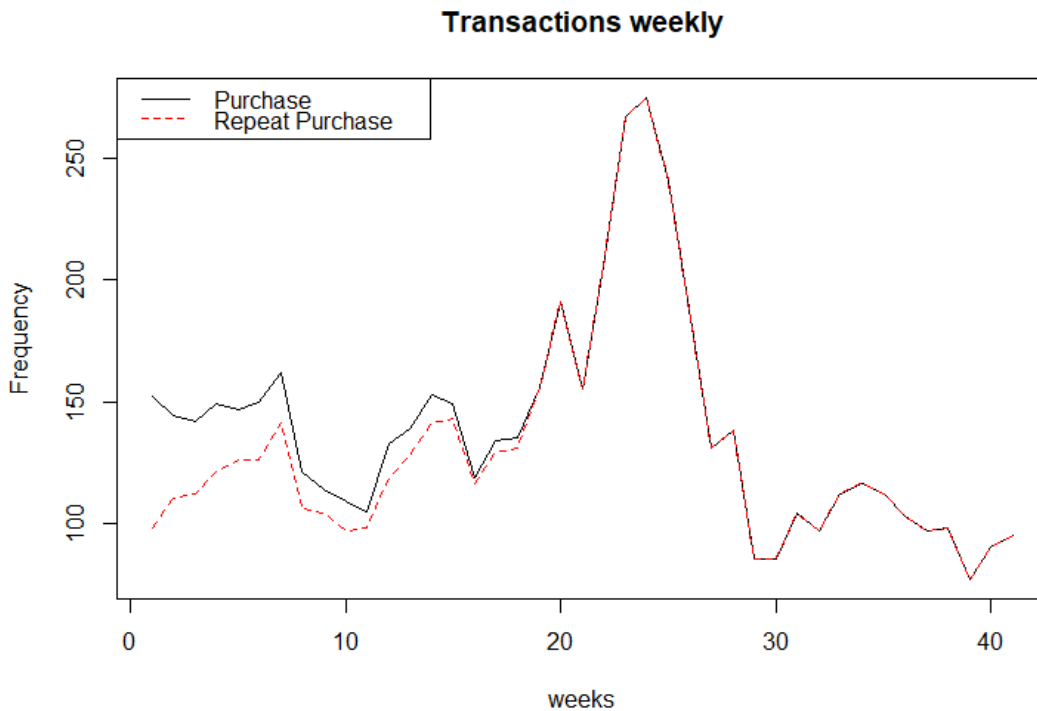


Figure 2. Intertransaction Timing of the First Period

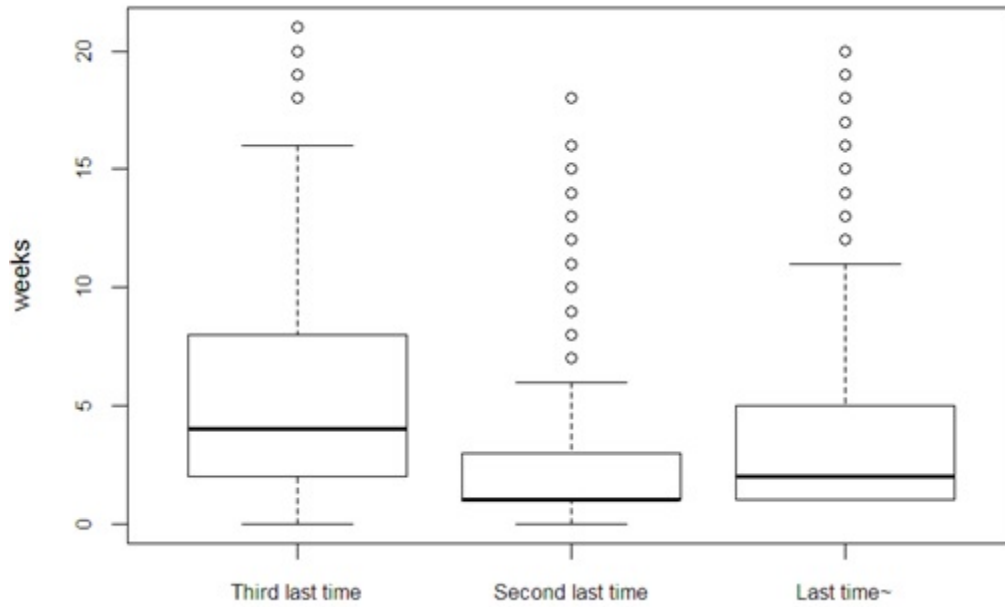
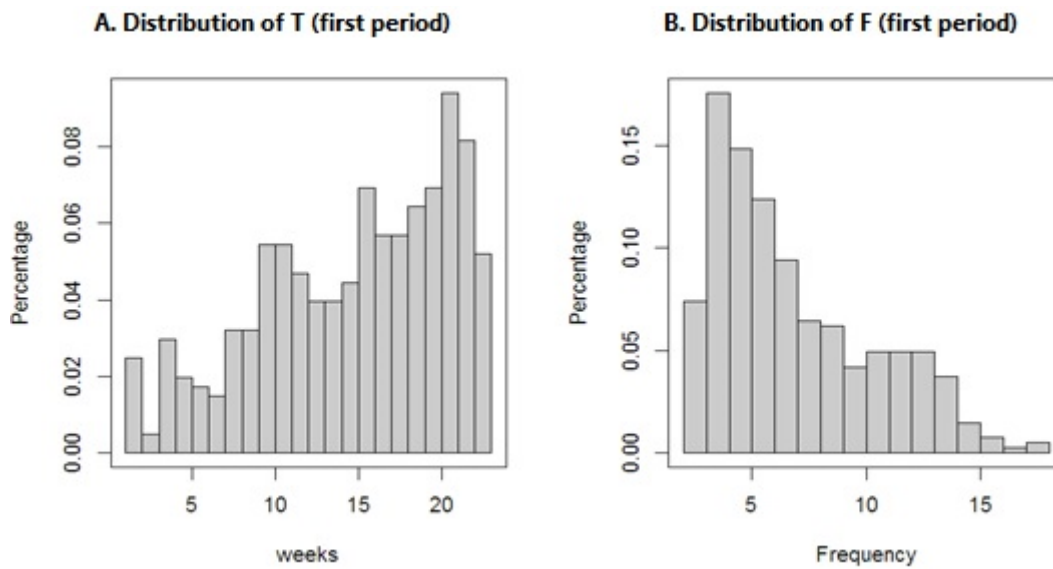


Figure 3. Distribution for the first period. (A) Time. (B) Frequency.



Estimating Frequency of Future Purchase Using Pareto/NBD (HB) and Pareto/GGG

We adopted the MCMC method to estimate the parameters of Pareto/NBD (HB) and Pareto/GGG based on the first-period dataset. MCMC estimation is a simulation-based estimation procedure in which random draws are recursively simulated from the model's full conditional distributions and are used as conditioning arguments in subsequent draws. Upon convergence, these draws form the true posterior. We estimated the model using 3,000 iterations of the Markov chain. The first 2,500 iterations were discarded, and the last 500 iterations were used to form estimates of the posterior distribution of model parameters. A time series plot of the draws indicated the convergence of two chains from multiple initial values.

We calculated the average of the draws as the estimated values of the parameters. The estimated parameter values of Pareto/NBD (HB) on an aggregated level were $\hat{\alpha} = 15.44$; $\hat{\tau} = 7.01$; $\hat{\delta} = 0.36$; $\hat{\beta} = 41.23$, which means the number of purchases per week was 0.45 and the dropout rate was 0.008. Meanwhile, the parameter estimation results of Pareto/GGG were $\hat{\tau} = 116.82$; $\hat{\gamma} = 131.56$; $\hat{\alpha} = 20.64$; $\hat{\tau} = 9.09$; $\hat{\delta} = 0.50$; $\hat{\beta} = 69.83$, meaning that the average number of purchases per week was 0.44, the dropout rate was 0.007, and the intertransaction timing was 0.89 weeks when the customer remained alive.

To evaluate the two models' prediction performance, we first grouped customers based on their number of purchases in the first period. The lowest category was two or fewer times, and the highest category was fifteen or more times. We then predicted the average number of purchases (i.e., the conditional expectation) of the corresponding groups in the second period according to Pareto/NBD (HB) and Pareto/GGG. We then compared the prediction results of the two models with the second period's actual values, as shown in Figure 4A. This figure also shows an overall positive correlation between the numbers of purchases in the first and second periods; however, the number of purchases in the second period decreased for the customers who purchased 10 times or 13 or more times, meaning that purchasing was unstable. The high purchase frequency in the first period may have been caused by factors such as seasonal fluctuation, holidays and festivals, sales and promotions, and similar. Figure 4A illustrates that neither the Pareto/NBD (HB) nor the Pareto/GGG method captured these fluctuations. The predictions of these two models were relatively stable. Figure 4B shows the cumulative number of purchases. The prediction values of Pareto/NBD (HB) and Pareto/GGG were very close, while both of them were overestimated with mean square error (MSE) terms of 3.05 and 3.23, respectively.

Furthermore, we grouped customers with a different observation time T , including one or fewer months—up to six months or more—in the first period and predicted the average number of purchases (i.e., the conditional expectation) of the corresponding groups in the second period, according to Pareto/NBD (HB) and Pareto/GGG. We then compared the two models' prediction results with the actual number of purchases of the second period, as shown in Figure 5. This figure shows that the average customer purchases' actual values did not increase monotonically with the duration of the observation time. The prediction values made by Pareto/NBD (HB) and Pareto/GGG were very close. Both models predicted well for groups 1 and 2, overvalued

for groups 3–5, and slightly undervalued for group 6. The MSE terms of the two models were 6.21 and 6.37, respectively.

Figure 4. Model Evaluation of Pareto/NBD (HB) and Pareto/GGG Based on Bins for F

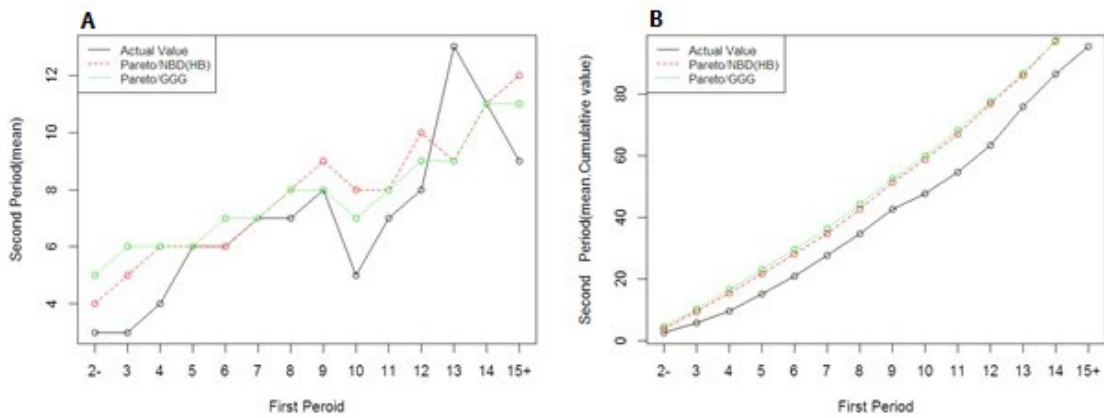
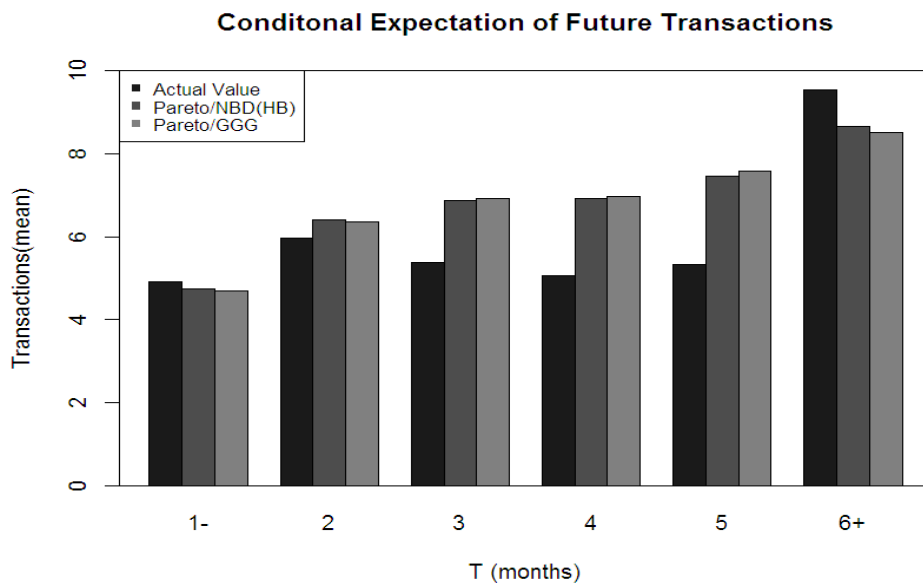


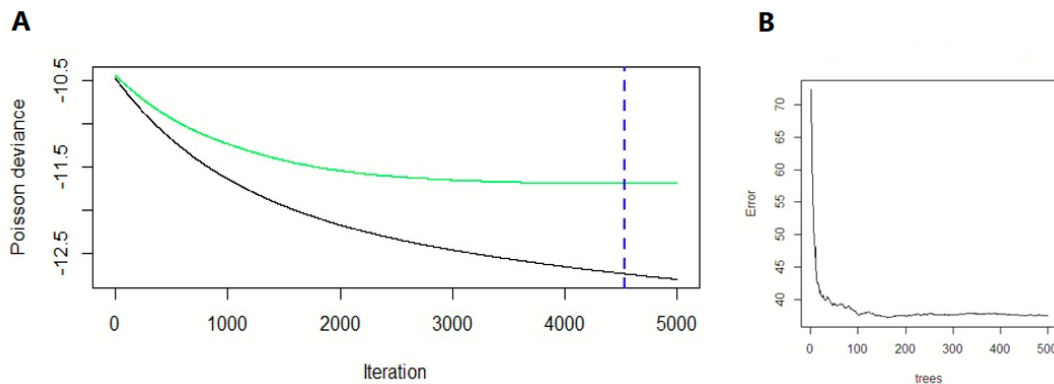
Figure 5. Model Evaluation of Pareto/NBD (HB) and Pareto/GGG Based on Bins for T



Estimating Frequency of Future Purchase Using GBDT and RF

For GBDT, we set the shrinkage parameter as 0.001, and it grew 5,000 trees. A 10-fold cross-validation method was used to prevent model overfitting. We investigated the influence of tree numbers on the training error and the test error as shown in Figure 6A. The green (upper) curve in Figure 6A represents the test error, and the black (lower) curve represents the training error. With the increase in the number of trees, the training error decreased monotonically, while the test error began to increase after reaching the minimum when the tree number was 4,532. This indicates that the model started to overfit when the tree number exceeded 4,532. We therefore chose the GBDT model when the number of trees equaled 4,532.

Figure 6. Influence of Tree Numbers of Training and Test Errors. (A) Training error (black) and test error (green) of GBDT. (B) Test error of RF.

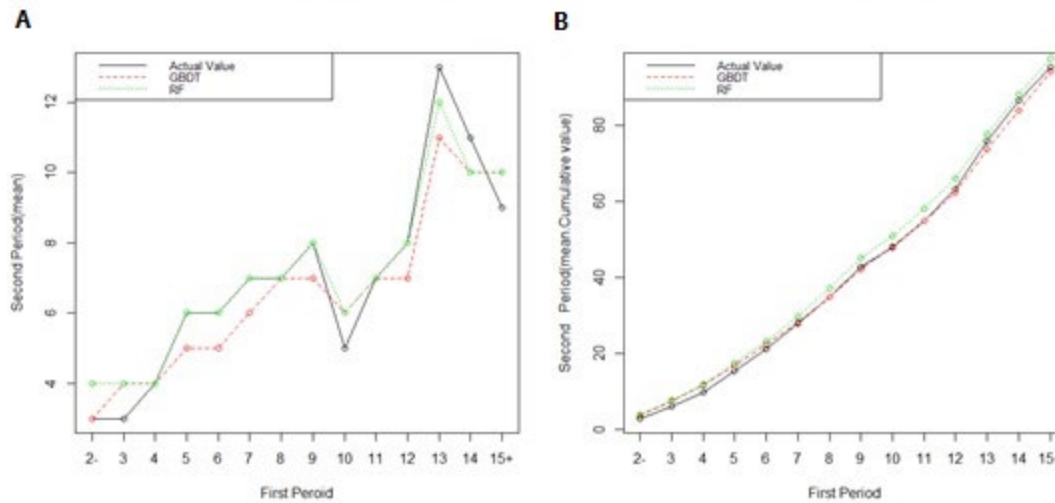


For RF, we set the variable subsets for each tree to include $p/3$ explanatory variables (p was the number of explanatory variables) to grow 500 trees. To prevent model overfitting, the test error curve (i.e. the OOB curve) was drawn as shown in Figure 6B. With the increase in the number of trees, the curve declined sharply at the beginning and then fluctuated up and down, which indicated model overfitting. The minimum test error was attained when the tree number reached 165. We therefore chose the RF model when the tree number equaled 165.

To evaluate the two models' prediction performance, we grouped customers according to the number of purchases—from two or fewer times to fifteen or more times—in the first period. We then predicted the average number of purchases (i.e., the conditional expectation) of the corresponding groups in the second period according to GBDT and RF. We then compared the two models' prediction results with the second period's actual value, as shown in Figure 7A. Compared with the prediction performance of the Pareto/NBD (HB) model and Pareto/GGG model as shown in Figure 4A, GBDT and RF had an excellent ability to track the sharp fluctuations in the dataset. The predictions for customers who purchased 10–13 times in the first period were significantly better than those of the Pareto/NBD (HB) and the Pareto/GGG models. We also compared the models' fitting effect based on the cumulative number of purchases, as shown in Figure 7B. The predicted

value curves of the GBDT and RF models coincided with the actual value curve, and the prediction results of the GBDT were slightly lower than those of the RF. Both GBDT and RF displayed excellent prediction ability, with MSE terms of 0.65 and 0.34, respectively.

Figure 7. Model Evaluation of GBDT and RF Based on Bins for F



Furthermore, similar to Figure 5, we grouped customers with different observation time T in the first period and predicted the average number of purchases (i.e., the conditional expectation) of the corresponding groups in the second-period GBDT and RF. We then compared the two models' prediction results with the actual number of purchases of the second period, as shown in Figure 8. Compared with the Pareto/NBD (HB) prediction performance and the Pareto/GGG models (Figure 5), the prediction values given by GBDT and RF were closer to the actual values. The prediction values of GBDT were slightly lower than those of RF. The MSE was 1.90 for GBDT and 0.64 for RF.

Both GBDT and RF can sort the importance of the explanatory variables. For example, as shown in Figure 9A, the RF model illustrates each explanatory variable's contribution to the reduction of the test error. Figure 9B shows each explanatory variable's contribution to the decrease in value heterogeneity of the tree nodes' explained variable. The greater the contribution value, the more important the explanatory variable. According to Figure 9A, T was the most important explanatory variable, followed by R , $R.1$, etc., while according to Figure 9B, $R.1$ was most important, followed by T , F , etc. The seven most important explanatory variables based on their contribution to reducing the test error are listed in Table 3.

Among the top seven variables, six were important for both GBDT and RF. Furthermore, the remaining variables (in **bold** in Table 3) for GBDT (average purchase interval) and RF (sum of logarithmic purchase intervals) were both functions of the purchase-time intervals. Number of purchases, purchase intervals, observation period, and

monetary value of the purchases were decisive factors for the number of purchases in the future (20 weeks in our study). This indicates that the number of purchases and the monetary value of purchases were not always independent of each other as the Pareto/NBD (HB) model assumed. It also shows the necessity of introducing variables that describe purchase-time intervals into the Pareto/GGG model.

Figure 8. Model Evaluation of GBDT and RF Based on Bins for T

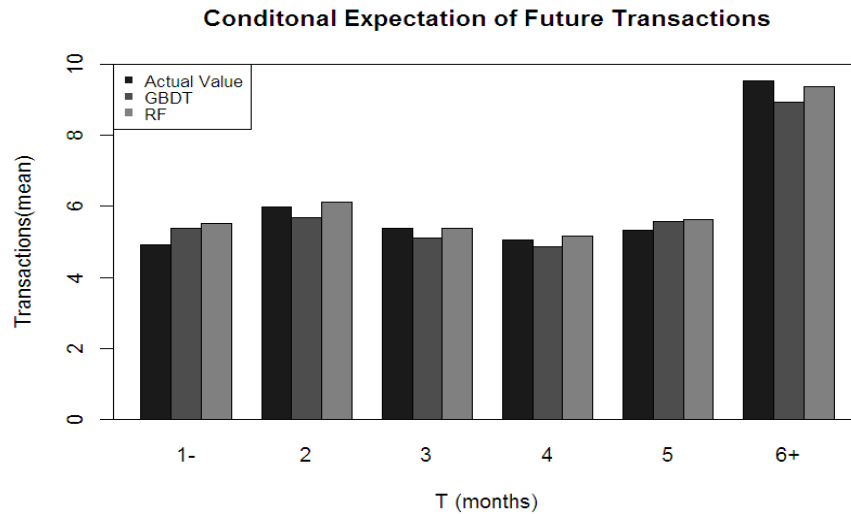


Figure 9. Variable Importance of RF to Reduction of Test Error (A) and Decrease in Value Heterogeneity (B)

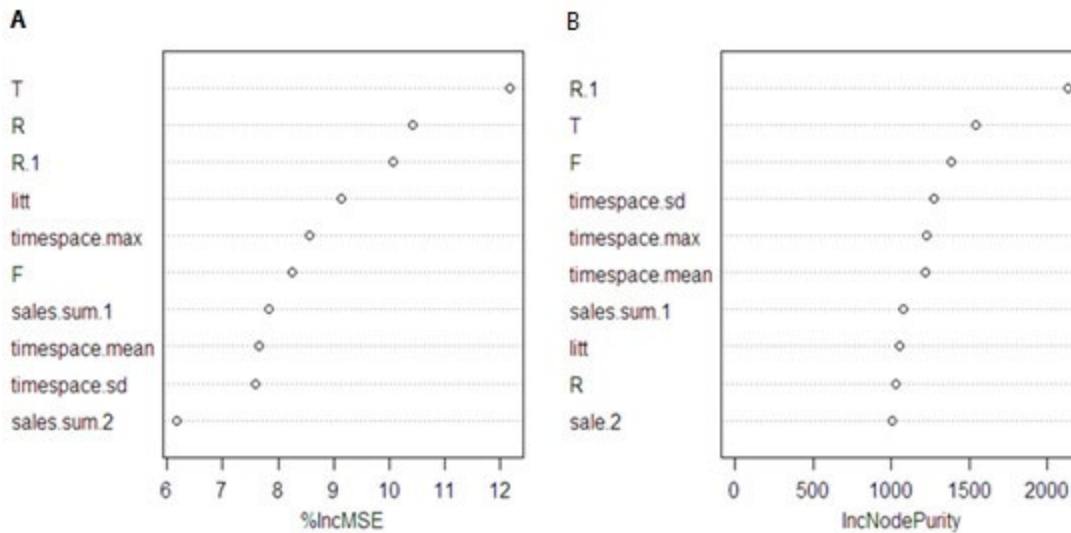
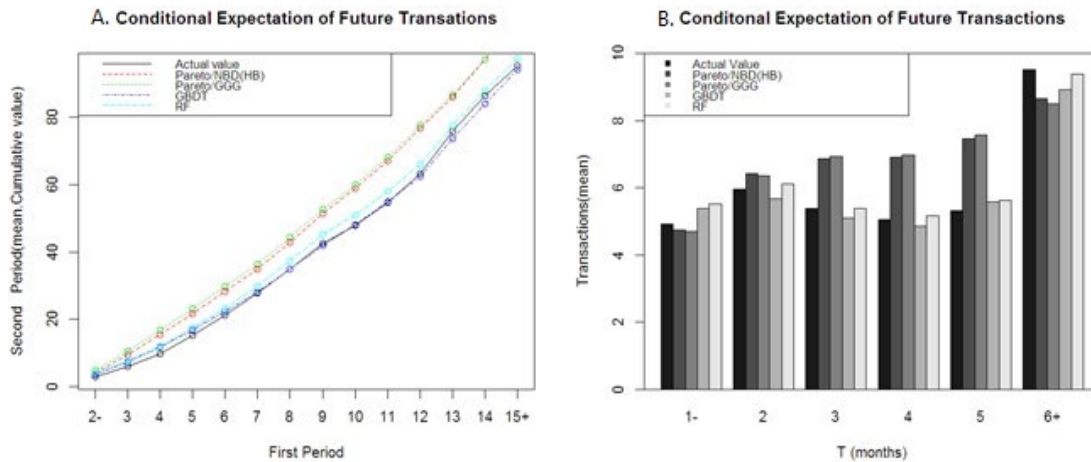


Table 3. Variables Ranked by Importance

GBDT		RF	
Variable	Importance score	Variable	Importance score
Interval between last purchase and third-to-last purchase (R.1)	28.72	Observation period (T)	12.17
Observation period (T)	14.44	Recency (R)	10.41
Maximum purchase interval (timespace.max)	9.59	Interval between last purchase and third-to-last purchase (R.1)	10.06
Number of purchases (F)	9.35	Sum of logarithmic purchase intervals (litt)	9.13
Average purchase interval (timespace.mean)	7.97	Maximum purchase interval (timespace.max)	8.56
Recency (R)	7.17	Number of purchases (F)	8.25
Accumulated monetary value of last three purchases (sale.sum.1)	5.75	Accumulated monetary value of last three purchases (sale.sum.1)	7.82

Figure 10. Comparison of Model Performance



Comparing the Prediction Performance of the Four Models

We compared prediction results of Pareto/NBD (HB), Pareto/GGG, GBDT, and RF. As shown in Figure 10, the prediction values of the two probability models, Pareto/NBD (HB) and Pareto/GGG, were similar. The prediction values of the two machine learning models, GBDT and RF, were also quite similar. Table 4 compares the MSE of the four models. The MSE terms of GBDT and RF were significantly lower than those of the Pareto/NBD (HB) and Pareto/GGG models, suggesting that machine learning algorithms had better predictive abilities.

Table 4. Comparison of Conditional Expectation MSE

Prediction Models	Conditional Expectation MSE (1)	Conditional Expectation MSE (2)
Pareto/NBD (HB)	3.05	6.21
Pareto/GGG	3.23	6.37
GBDT	0.65	1.90
RF	0.34	0.64

Estimating the Monetary Value of Future Purchases

Schmittlein and Peterson (1994) noted that the monetary value of a customer’s future single purchase can be estimated based on a normal distribution. In our study, θ denoted the mean and σ_A^2 denoted the variance of purchase value per transaction for all customers, θ_i denoted the expectation, and σ_{wi}^2 denoted the variance of the future purchase value per transaction for the i th customer. Under large sample \bar{Z}_i , the average purchase value for X_i times (sample mean) was normally distributed with mean θ_i and variance σ_{wi}^2/X_i . The expected future purchase value per transaction was therefore

$$\theta_i = \left(\frac{X_i \sigma_A^2}{X_i \sigma_A^2 + \sigma_{wi}^2} \right) \bar{Z}_i + \left(\frac{\sigma_{wi}^2}{X_i \sigma_A^2 + \sigma_{wi}^2} \right) \theta.$$

According to the equation, we can estimate each customer's future purchase value per transaction, denoted by \hat{M} .

Identifying Profitable Customers Based on CLV

After predicting the frequency and monetary value of each customer’s future purchase, we combined the two to calculate CLV ($CLV = \hat{M} \times \hat{F}$). Here, \hat{F} was the predicted purchase frequency and \hat{M} was the predicted purchase value for the future 20 weeks.

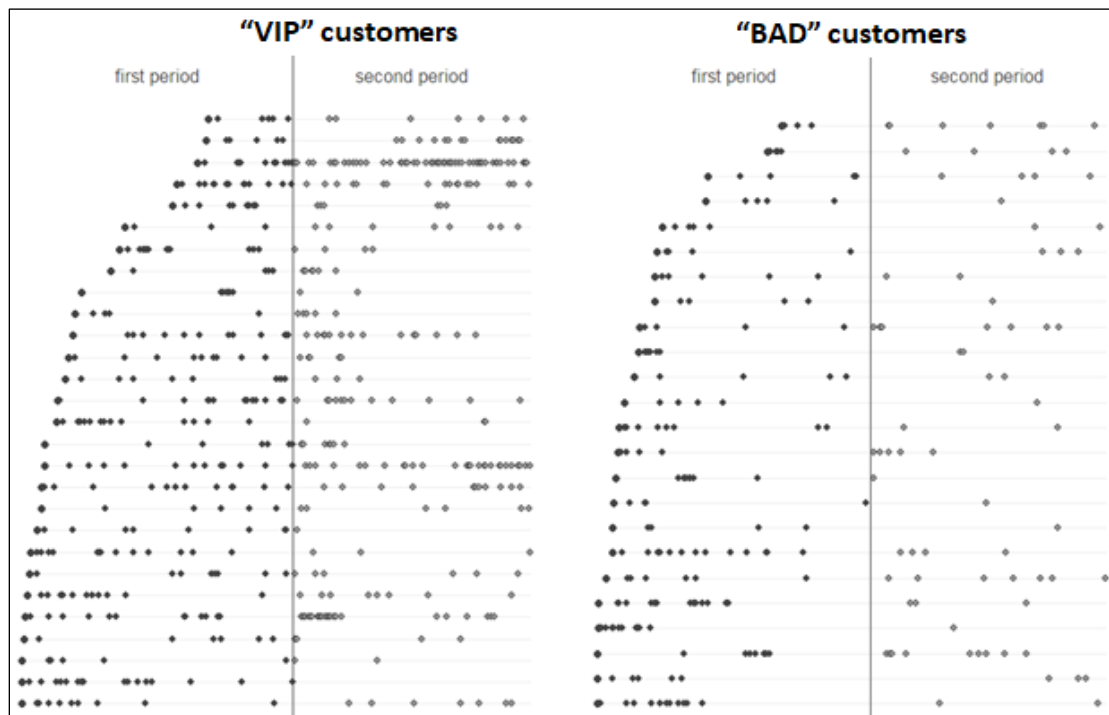
We identified the most- and least-profitable customers based on the combined CLV prediction model. First, we listed the 15 percent of customers with the highest

CLV as “VIP” customers (N total customers) and the 15 percent with the lowest CLV as “BAD” customers (N total customers), resulting in four VIP customer lists and four BAD customer lists. Next, we used the four prediction models to count the number of times each customer appeared on the VIP customer lists or the BAD customer lists. Finally, we identified a customer as profitable or unprofitable based on his or her scores. When a customer received a score of at least 3 for VIP, he or she was identified as a profitable customer; similarly, a score of at least 3 for BAD caused a customer to be identified as unprofitable.

After customer identification, we randomly selected 25 profitable and 25 unprofitable customers to further investigate with regard to purchasing behaviors. Figure 11 reflects the purchasing behaviors of both types of customers, where the horizontal axis is the week, the vertical axis is the customer ID, and the dots indicate that the customer made purchases in that week. A profitable customer could be an old customer with a longer history of purchases, an old customer with a larger number of purchases and relatively regular purchase cycles, or a new customer with shorter purchase-time intervals but a larger number of purchases within a rather short time. It should be noted that the value of purchases also has a great influence on the calculation of CLV but Figure 11 depicts only number of purchases, time intervals, and survival time of customers without considering the value of customer purchases. This might be an important reason why customers with fewer purchases were still identified as profitable customers. Compared with profitable customers, unprofitable customers usually have no regular purchase cycle and have increasingly longer purchase-time intervals despite the possibility of frequent purchases earlier in the observation period.

Based on the customer-identification results, we believe that it is necessary to identify customers based on a combined prediction strategy of CLV. Because only customers who received scores of at least 3 were identified as profitable or unprofitable, we can say that there should be approximately N customers identified as profitable and N as unprofitable when the four models give relatively consistent CLV values. Otherwise, the number of profitable or unprofitable customers should be lower than N . The prediction results for the four models were not completely consistent because of different method design and modeling approaches. This was reflected in that only 74 percent of the customers on the VIP customer list were identified as profitable and only 72 percent of customers on the BAD customer list were identified as unprofitable; that is, the overall agreement rates of the four models were 74 percent and 72 percent, respectively, for the identification of profitable and unprofitable customers. Moreover, the agreement rates of the Pareto/NBD (HB) and Pareto/GGG models were 92 percent and 88 percent, respectively, which were significantly higher than the overall agreement rate. The agreement rates of GBDT and RF were 78 percent and 76 percent, which was also slightly higher. These indicate that different models might get different prediction results. The customer identification based on a certain model might thus be biased. We therefore believe that a combined CLV prediction strategy is an effective way to improve the robustness of customer identification.

Figure 11. Customer Purchasing Behavior



MANAGERIAL CONTRIBUTION

Our study findings offer several managerial benefits. First, we have developed a new metric to measure CLV. Given the distinct benefits provided by CLV, a marketer should closely monitor this metric in the pursuit of growing its business. Buoyed by the technological advancements of analytics and a customer data platform, it delivers all the information a company needs to predict CLV. Second, CLV gives a company a closer look at the health of a business by taking a longer timeframe into account. CLV can help a company identify its best customers. Data about customers let a company spot those who spend the most. Taking advantage of this information enables a company to promote certain products. Third, a company can invite its customers to special events and can offer deals specially tailored for high-value customers. Finally, the company can take better care of its most valuable customers by providing them with individual assistants or advisers (Jain and Singh 2002; Kumar and Reinartz 2016).

SUMMARY AND CONCLUSION

Researchers have explored different CLV modeling methods, but most of the studies have focused on deductive approaches, such as probability and econometric and persistence models, because of their emphasis on parametric setup and easy interpretability in the

marketing literature. Compared with deductive approaches, inductive approaches based on modern statistics and machine learning algorithms have not received as much attention. With the development of data analysis technology, inductive methods are playing a more and more important role in marketing research. In our study, we introduced two inductive models, GBDT and RF, to predict CLV. Based on the empirical analysis of a Chinese retailer, we found that the predictive performance of modern statistics and machine learning algorithms was generally better than that of the methods based on probability distribution (i.e., Pareto/NBD (HB) and Pareto/GGG). To identify the most- and least-profitable customers for the firm, we first used the four aforementioned models separately to predict CLV and then combined the prediction results to ensure the robustness of customer identification.

Because the inductive methods have excellent predictive ability, we believe that CLV models based on modern statistics and machine learning should be further explored. Future research can introduce more modern statistics and machine learning algorithms in CLV modeling and can focus on the comparison of their predictive ability. With the continuous enrichment of CLV modeling approaches, we propose combining more different CLV methods as a way to ensure robustness in customer identification.

REFERENCES

- Abe, M. 2009. "Counting Your Customers One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model." *Marketing Science* 28(3):541–53.
- Bandyopadhyay, S. 2009. "A Dynamic Model of Cross-Category Competition: Theory, Tests and Applications." *Journal of Retailing* 85(4):468–79.
- Bandyopadhyay, S., and M. Martell. 2007. "Does Attitudinal Loyalty Influence Behavioral Loyalty? A Theoretical and Empirical Study." *Journal of Retailing and Consumer Services* 14(2007):35–44.
- Breiman L., J. Friedman, R. Olshen, and C. Stone. 1984. *Classification and Regression Trees*. Boca Raton, FL: Chapman & Hall.
- Buttle, F. 2004. *Customer Relationship Management: Concepts and Tools*. Burlington, MA: Elsevier Butterworth-Heinemann.
- Coussement K., D. F. Benoit, and D. Van den Poel. 2010. "Improved Marketing Decision Making in a Customer Churn Prediction Context Using Generalized Additive Models." *Expert Systems with Applications* 37(3):2132–43.
- Cui D., and D. Curry. 2005. "Prediction in Marketing Using the Support Vector Machine." *Marketing Science* 24(4):595–615.
- Erevelles S., N. Fukawa, and L. Swayne. 2016. "Big Data Consumer Analytics and the Transformation of Marketing." *Journal of Business Research* 69(2):897–904.
- Estrella-Ramón, A. M., M. Sanchez-Perez, G. Swinnen, and K. VanHoof. 2013. "A Marketing View of the Customer Value: Customer Lifetime Value and Customer Equity." *South African Journal of Business Management* 44(4):47–64.
- Fader, P. S., B. G. S. Hardie, and K. L. Lee. 2005a. "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model." *Marketing Science* 24(2):275–84.

- Fader, P. S., B. G. S. Hardie, and K. L. Lee. 2005b. "RFM and CLV: Using Iso-value Curves for Customer Base Analysis." *Journal of Marketing Research* 42(4):415–30.
- Fader, P. S., B. G. S. Hardie, and J. Shang. 2010. "Customer-Base Analysis in a Discrete-Time Noncontractual Setting." *Marketing Science* 29(6):1086–1108.
- Gupta S., D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, and S. Sriram. 2006. "Modeling Customer Lifetime Value." *Journal of Service Research* 9(2):139–55.
- Jain, D., and S. S. Singh. 2002. "Customer Lifetime Value Research in Marketing: A Review and Future Directions." *Journal of Interactive Marketing* 16(2):34–46.
- Jen, L., C-H. Chou, and G. M. Allenby. 2003. "A Bayesian Approach to Modeling Purchase Frequency." *Journal of Marketing Letters* 14(1):5–20.
- Kumar, V. 2008. *Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty*. Philadelphia: Wharton School Publishing.
- Kumar, V., and W. Reinartz. 2016. "Creating Enduring Customer Value." *Journal of Marketing* 80(5):36–68.
- Lawson, A. E. 2005. "What Is the Role of Induction and Deduction in Reasoning and Scientific Inquiry?" *Journal of Research in Science Teaching* 42(6):716–40.
- Lycett, M. 2013. "'Datafication': Making Sense of (Big) Data in a Complex World." *European Journal of Information Systems* 22(4):381–86.
- Ma, S-H., and J. Büschken. 2011. "Counting Your Customers from an 'Always a Share' Perspective." *Marketing Letters* 22(3):243–57.
- Ma, S-H., and J-L. Liu. 2007. "The MCMC Approach for Solving the Pareto/NBD Model and Possible Extensions." in *Third International Conference on Natural Computation*.
- Martínez, A., C. Schmuck, S. Pereverzyev Jr., C. Pirker, and M. Haltmeier. 2018. "A Machine Learning Framework for Customer Purchase Prediction in the Noncontractual Setting." *European Journal of Operational Research* 281(3):588–96.
- Meyer-Waarden, L. 2007. "The Effects of Loyalty Programs on Customer Lifetime Duration and Share of Wallet." *Journal of Retailing* 83(2):223–36.
- Pfeifer, P. E., and R. L. Carraway. 2000. "Modeling Customer Relationships as Markov Chains." *Journal of Interactive Marketing* 14(2):43–55.
- Platzer, M., and R. Reutterer. 2016. "Ticking Away the Moments: Timing Regularity Helps to Better Predict Customer Activity." *Marketing Science* 35(5):693–829.
- Reinartz, W. J., and V. Kumar. 2000. "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing." *Journal of Marketing* 64(4):17–35.
- Reinartz, W. J., and V. Kumar. 2002. "The Mismanagement of Customer Loyalty." *Harvard Business Review* 80(7):86–94.
- Rust, R. T., V. Zeithaml, and K. L. Lemon. 2000. *Driving Customer Equity: How Customer Lifetime Value Is Reshaping Corporate Strategy*. New York: The Free Press.
- Schmittlein, D. C., D. G. Morrison, and R. Colombo. 1987. "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science* 33(1):1–24.

- Schmittlein, D. C., and R. A. Peterson. 1994. "Customer Base Analysis: An Industrial Purchase Process Application." *Marketing Science* 13(1):41–67.
- Thomas, J. S. 2001. "A Methodology for Linking Customer Acquisition to Customer Retention." *Journal of Marketing Research* 38(2):262–68.
- Vafeiadis T., K. I. Diamantaras, G. Sarigiannidis, and K. Chatzisavvas. 2015. "A Comparison of Machine Learning Techniques for Customer Churn Prediction." *Simulation Modelling Practice and Theory* 55(6):1–9.
- Venkatesan, R., and V. Kumar. 2004. "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation." *Journal of Marketing* 68(4):106–25.
- Zeithaml, V. A., R. T. Rust, and K. N. Lemon. 2001. "The Customer Pyramid: Creating and Serving Profitable Customers." *California Management Review* 42(4):118–42.