# Identifying Influencers of Consumer Activity: A Case Study in Predictive Modeling

**Angela D. Nalica and Joseph Ryan G. Lansangan**
*School of Statistics*
*University of the Philippines Diliman*

Marketing efforts usually entail universal blast of information to all consumers. Oftentimes, only a small proportion of the consumers react positively to such efforts, resulting to waste in marketing expenses. If a circle of influencers can be identified for certain events or phenomena, then strategic planning, allocation of resources, and implementation of activities may be optimized, focusing only on specific group of individuals and/or factors. In this study, a usage database is used to identify consumers who could initiate or influence the complex dynamics of consumer behaviour. The data mining processes of clustering, sampling, aggregation, modeling, and validation are used to mine such information from the database.

***Keywords**: logistic regression, segmentation, influencers, consumer behaviour, customer relationship management*

## 1. Introduction

The consumer products and service industry continues to change from a rapidly growing market into a state of intense competition and saturation. Companies are becoming more and more aware of the crucial importance of their existing client database as inputs for marketing strategies. Predictions of customer behaviour, customer value, customer satisfaction and customer loyalty are some of the information that maybe extracted from a company's database. Data from scanned information in the point of sales (POS) machines, automatic teller machines (ATM), credit card transactions, and other records of customer/ consumer interactions are some of the sources of a company's database.

Consumers are being bombarded by a lot of marketing activities, commonly through SMS or e-mail blasts as well as registered mailings and call-outs. However, not all consumers are responsive to these types of activities. Thus, including the entire customer base to these types of activities may prove to be a waste of resources and time.

The major motivation for this study is the Tipping Point Hypothesis by Gladwell (2002) which states that: (a) a circle of influencers can be identified for certain events or phenomena, and that (b) the search for explanations for the occurrence or onset of certain events or phenomena can be zoomed into a group of factors or individuals. Identifying influencers facilitates problem solving, mitigation of certain negative effects, or optimization of the outcomes (Gladwell, 2002). In the service sector or recurring purchasable products/services, the tipping point hypothesis may be translated into the following:

- There are circles or segments of customers responsible (or serving as influencers or catalysts) of some activities like acquisition/initiation, churn, patronage, and loyalty.
- A group of influencers may be identified so that resources for strategic optimization of outcomes of such events maybe allocated efficiently. Resources need not be spread thinly to the whole population of customers, and specific segments and/or factors may be focused on that can yield multiplier effects in fast-tracking acquisition, improvement of patronage and loyalty, and mitigation of churns. This simplifies marketing programs and activities targeted only on few customers.
- It is possible to identify from the existing database key customers who could initiate or influence the complex dynamics of consumer behaviour.

Given the tipping point hypothesis, this study wishes to derive some empirical evidence, particularly on on churn and loyalty of customers. Likewise, segments or circles of influencers will be identified and further determine their profiles. In addition, a roadmap on how a continuous consumer tracking system or customer relationship management (CRM) may be planned and implemented from the results of this study.

## 2. Customer Relationship Management

There are various definitions of customer relationship management. Chalmeta (2006) defines CRM as a set of business, marketing and communication strategies and technological infrastructure designed with the aim of building a lasting relationship with customers, through identifying, understanding and meeting their needs. Although most companies use a customer-focused strategy that necessitates CRM solutions, implementation usually fails. This is mainly due to the inadequacy of existing methodologies used to approach a CRM project such as failing to satisfactorily integrate and complement the strategic and technological aspects of CRM.

Coltman (2007) notes that a superior CRM capability can create positional advantage and subsequent improved business performance. However, CRM suffers when it is not properly understood and implemented. Therefore, in order to be successful, CRM should focus on the underlying or latent customer needs that emphasize a proactive market orientation. Coltman (2007) further demonstrates that by integrating three schools of thought—capabilities, market orientation and conversion feasibility—CRM programs can be successful.

Some methods are also proposed to model consumer acceptance probabilities. Thomas et al. (2006) introduce techniques which may be used to build models for the likelihood that a particular consumer accept different variants of a generic borrowing product such as credit card. They emphasize that the probability acceptance models have become increasingly vital as the consumer lending markets mature and become more consumer-centered. The models satisfy the marketing credo of tailoring the product to the customer.

Because of saturated markets and fierce competition, the focus is now being shifted from building a large customer base to keeping the existing ones – i.e., preventing customers from churning. It is much cheaper to retain customers than to gain new ones. Thus, attention has been directed on coming up with churn models.

Glady et al. (2009) propose a new framework for the churner detection process using the earnings a customer brings to the company. They define customer lifetime value (CLV) as the discounted value of future marginal earnings based on the customer's activity. They assess the performance of several classifiers for churn prediction such as logistic regression, decision tree and neural networks together with cost-sensitive classifiers such as AdaCost and cost-sensitive decision tree. Glady et al. (2009) further emphasize that aside from good overall classification, it is also important to correctly classify potentially profitable churners.

Hadden et al. (2005) review the main trends for developing models to predict consumer behaviour specifically those used in the development of customer churn management platform. These include traditional methods such as decision trees, regression analysis, Markov models, and genetic algorithms (GA), among others. Most modeling efforts focus mainly on feature extraction rather than feature selection problem.

Popović and Bašić (2009) illustrate how data mining methods based on the fuzzy logic could be successfully applied in the retail banking analysis and, moreover, that the fuzzy c-means clustering performed better than the classical clustering algorithms in the problem of churn prediction. They use discriminant analysis to reveal variables that provide maximal separation between clusters of churners and nonchurners.

Padmanabhan and Tuzhilin (2003) further emphasize that automated CRM problems of customer analysis, customer interactions, and optimization performance methods maybe better analyzed with the simultaneous use of data mining and optimization methods.

## 3. Methodology

The data mining strategy of modeling and automation was used to provide empirical evidence on the tipping point hypothesis. The database came from a service company with a nationwide coverage. With customer information and detailed transactions covered in the database, data on usage and frequency of certain services were analyzed. A 1-year worth of time series data was used to characterize customer behaviour. Cluster analysis was employed to generate simple usage groupings based on usage profile and revenue indicators. Once groupings have been identified, a sample of customers by cluster was obtained and information on detailed activities was gathered. This was used in the identification of possible influencers.

To generate company-level (macro-level) benchmark of various activities, several indicators were aggregated on a weekly basis over a period of 53 weeks. Together with some customer/account level variables, these were used in coming up with transfer function models (TFM). A TFM with macro-level activities (number of churns and number of continuing accounts) as outputs and consumer activities as inputs were fitted for each of the potential influencer from the segmentation of customers.

For churn, a customer's life cycle that exhibits significant cross-correlation between the number of churns and the occurrence of its own churn was tagged as influencer and the delay (in weeks) was identified from the same cross-correlations. This resulted in a dichotomy that either confirmed or negated the postulation that the customer is an influencer of churn.

For loyalty, a customer's life cycle that exhibits significant cross-correlation between the number of continuing accounts and the number of distinct transactions made by the customer was tagged as influencer and the delay (in weeks) was identified from the same cross-correlations. This resulted in a dichotomy that either confirmed or negated the postulate that the customer is an influencer of loyalty.

The results of the TFM were then used in logistic regression modeling that helped automate the determination of influencers. Outputs from the logistic regression and the principal components analysis were then used to come up with influencer scoring equation. For each activity (churn and loyalty), customers were considered as possible influencer or non-influencer based on some pre-defined characteristics. As an example, the number of distinct relations (for different services) was counted for modeling churn.

Assessment of the soundness and validity of the results were done by considering training and validation data sets in the analyses. A 90-10 split was used for training and validation, respectively.

The Flow Chart of the methodology is summarized in Figure 1.

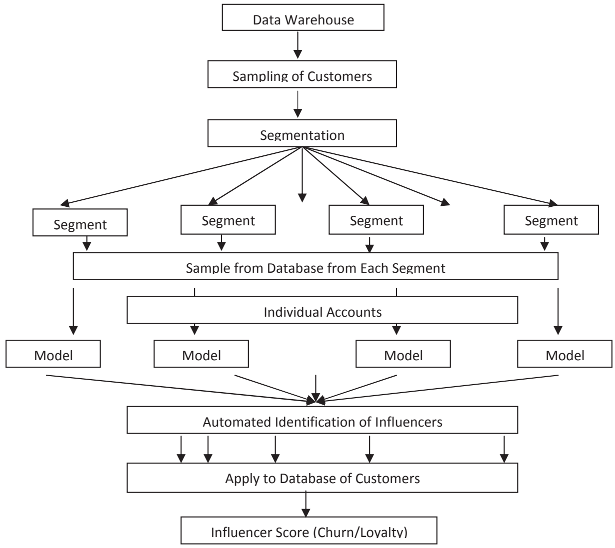


**Figure 1. Data Mining Procedures**

## 4. Results and Discussion

For purposes of identification of influencers from the history of usage, a vital assumption is that influence takes effect when the customer uses the services provided by the company. The segments then should be able to aptly represent various dynamics of customer usage of services provided by the company.

Results show that 7% or a total of 91 sample customers are potential influencers and the remaining 93% or 1,214 sample customers are non-influencers whoexhibited low usage on practically all services (i.e. low linkages to other customers thereby reducing the opportunity to influence acquisition, churn, loyalty or patronage).

### *4.1 Churn model*

A sample of 205 churners is stratified by week of churning from week 14 to week 40. About 72 or 35% of the sample churners are not influencers and more than one-half influence churning on the same week they churned. More than half of the potential influencers of churn has an effect on the churn counts within the same week the customers churned (contemporaneous), over 3% influence churning one week after they churned, 8% influence 2 weeks after, and a very few would influence churning after 3-8 weeks (see Table 1).

**Table 1. Churn Influencers By Number of Weeks After They Churned**

| Weeks After Influence Is Measured | Frequency | Percent |
|---|---|---|
| 0 | 103 | 50.49 |
| 1 | 7 | 3.43 |
| 2 | 16 | 7.84 |
| 3 | 1 | 0.49 |
| 4 | 1 | 0.49 |
| 6 | 1 | 0.49 |
| 7 | 2 | 0.98 |
| 8 | 1 | 0.49 |
| Not Significant | 72 | 35.29 |

Among the non-influencers, 95% do not have any activity in the last six months. While among those who can influence, 41% still have activities in the last six months prior to churning.

There is a problem, however, on the use of logistic regression to automate identification of influencers since there is minimal usage months prior to churning. A possible solution is to use churn prediction model among those who are probable churners but still have activities and maybe considered probable influencers.

The aggregate extent of influence is impossible to compute because the week of churning is a variable used as a basis in assessing the impact of influencers. There are several customers that could churn in that same week as other customers churn. This is an empirical assessment of causation between the week they churned and the number of churners. Also, once the customer has churned, no more activity can be observed, thus making it impossible to verify the extent of influence.

The best way to validate churning influence is to look at the activities before churning, whether the customer relationships/circles (i.e. linkages to other customers within the same service/s) include accounts that also churned. However, six months prior to churning, the customers hardly had any activities, and the data warehouse keeps data only for 13 months. It is impossible then to trace back many months before churning to understand the usage dynamics of churners. One possible solution is to retrieve data from archives – which is disregarded at the time of the study because it is deemed to take a longer period.

### *4.2 Loyalty model*

The study also aimed to identify influencers of continuing customers, i.e., model on loyalty. From a sample of 1,305 loyal customers, the number of distinct transactions/activities is determined per account. It is assumed that the more distinct transactions a customer has made, the more loyal customers associated with the customer is observed. Results show that 172 or 13.4% are significant influencers based on transaction A, 10 or 0.78% are influencers based on Transaction B. Overall, 177 or 13.8% are influencers based on the distinct Transactions A/B. This supports the earlier observation that influence is exhibited more for Transaction A and less for Transaction B.

Influencers maybe identified based on quality of transaction (higher revenue) together with the quantity (no. of transactions). To model the probability of being an influencer of continuing accounts, logistic regression model is used and results show that factors (mostly quantity measures) are significant. It shows that the odds of being influencer are higher when the proportion of incoming transactions from another account from a competitor is high. Although some indicators that are included in the model are not significant, they are not removed from the model since they contribute towards the stability of the model. A typical loyalty influencer model is given by:

$$P(\text{Inf. Loyalty}) = \frac{1}{1+\exp\left(1.58+(1.27*A)-(1.41*B)-(1.03*C)-(0.0005*D)+(0.52*E)\right)}$$

where

A = proportion of incoming transactions coming from another customer of the same company to total incoming transaction

B = proportion of usage during off peak periods from another customer of the same company to total usage during off peak periods

C = proportion of usage during peak periods from another customer of the same company to total usage during peak periods

D = total number of outgoing transactions

E = proportion of incoming transactions during off peak period from another customer of the same company to all incoming transactions during off peak periods

### *4.3 Business implications*

*Churns*

Results show that some churners may potentially influence other churners. Generally, churners have minimal usage details six months prior to churning, making it difficult to validate influencers of churn. However, there is an indication that potential influencers still exhibit usage/transactions a few months prior to churning. Since churners may be classified as either voluntary or involuntary, it is thus recommended that the churn prediction model be used in predicting churners. Subsequently, current activities can be assessed and those who exhibit relatively higher usage level maybe considered as potential influencers.

*Loyalty*

The output from the transfer function model shows the extent of the influence of each customer on the number of continuing customers. Results show an average impact of influencers. Some customers influence immediately, while others may have lags in actual increase of the number of continuing customers of up to 4 weeks. If incentives are given for 1,071 customers (<1% of actual base, on the assumption that they are the only influencers) every time their customer relationships/circles increase, they may encourage from 98-180 weekly loyal customers. In the most optimistic case, they may bring in almost one loyal customer each on a weekly basis.

## 5. Conclusions

It is illustrated that influencers of certain activities such as churn and loyalty can be identified from the customer database. Expenditures aimed to activate certain activities need not be spread across all customers. By focusing on some fewer influencers only, activation may become more efficient and higher returns can be expected.

This study identifies the influencers through a purposive selection of samples from possible influencers and noninfluencers. Purposive selection is intended to increase the likelihood of selecting influencers (hypothesized to be smaller in number compared to the total customer database). The individual transaction patterns are analyzed to confirm the customer being an influencer or not. Identification of influencers is automated through a logistic regression with monthly aggregate measures as determinants.

Results maybe used so that activation or promotions can focus on certain influencers and their reaction in terms of usage, likewise, transactions with internal and/or external circles maybe analyzed. Such data may be used not only to validate a customer being influencer, as well as in studying effectiveness and efficiency of certain activation strategies.

## REFERENCES

CHALMETA, R., 2006, Methodology for customer relationship management, *The Journal of Systems and Software* 79:1015-1024.

COLTMAN, T., 2007, Why build a customer relationship management capability?, *Journal of Strategic Information Systems* 16: 301-320.

GLADWELL, M., 2002, The tipping point: How little things can make a big difference, New York: Little Brown and Company.

GLADY, N., BAESENS, B., CROUX, C., 2009, Modeling churn using customer lifetime value, *European Journal of Operational Research* 197: 402-411.

HADDEN, J., TIWARI, A., ROY, R., and RUTA, R., 2005, Computer assisted customer churn management: State-of-the-art and future trends, *Computers and Operations Research* 34:2902-2917.

PADMANABHAN, B. and TUZHILIN, A., 2003, On the use of optimization for data mining: Theoretical interactions and eCRM opportunities, *Management Science* 49: 1327-1343.

POPOVIĆ, D. and BAŠIĆ, B., 2009, Churn prediction model in retail banking using fuzzy C-Means, *Informatica* 33:243-247.

THOMAS, L., JUNG, K., THOMAS, S., and WU, Y., 2006, Modeling consumer acceptance probabilities, *Expert Systems with Applications* 30:499-406.