DORMANCY PREDICTION MODEL IN A PREPAID PREDOMINANT MOBILE MARKET : A CUSTOMER VALUE MANAGEMENT APPROACH

Adeolu O. Dairo and Temitope Akinwumi

Customer Value Management Department, Segments and Strategy Division, Etisalat Nigeria

ABSTRACT

Previous studies have predicted customer churn in the mobile indutry especially the postpaid customer segment of the market. However, only few studies have been published on the prepaid segment that could be used and operationalised within the marketing team that are responsible for the management of incident of prepaid churn. This is the first identifiable literature where customer dormancy is predicted along the customer value segmentation. In this article, we use a popular data mining technique to predict when a customer will go dormant or stop performing revenue generating events in a prepaid predominant market. Our study is unique as we considered ~1,451 attributes derived from CDR and SIM registration database (previous studies only considered maximum of ~1,381 potential variables). We built 3 different models for Very High, High and Low value segments. We applied our models on the prepaid base and the output was later compared with the actual dormant customers. Very High segment has the highest accuracy and lift while Low segment has the least at the same threshold. We show that once the problem of prepaid churn is well defined, it can be predicted. We recommend a value segmentation dormancy prediction with decision tree for prepaid segment with a certain threshold. Our study shows that this approach can be easily adopted and operationalised by the campaign management team responsible for the management of prepaid churn in a mobile industry.

KEYWORDS

Dormancy Prediction Model (DPM), Revenue Generating Event (RGE), Retention, Call Detail Record (CDR)

1. INTRODUCTION

1.1. The Need for Dormancy Models

One of the major challenges of mobile operators in a prepaid predominant market is the rate at which customers go into inactivity state. This state is also known as a dormant state. Customers enter this state when they stop performing whatever actions the mobile operators define as a revenue generating event (RGE). Generally, revenue generating events are actions performed by a mobile customer which are seen as chargeable events by the operators. These events are capable of resetting the dormancy days of customers to zero depending on the set business rules of the operator. When a customer remains in the dormant state for a period, the customer will be

reported as a churner. In Nigeria, this period as defined by the regulator is 90 days. A customer is reported as a churner at 91st day of remaining in a dormant state.

Customer dormancy makes it difficult for the marketing team to reach such customers for retention interventions. As the days of inactivity increase, it becomes more difficult to reach the customers with retention activities. Hence, the need to know when the customer is likely to enter dormancy state rather than when the customer will be reported as a churner. At this stage, appropriate marketing activities can be targeted at these customers as they will still be reachable. The Dormancy Prediction Model (DPM) predicts the likelihood of becoming a dormant customer.

Churn prediction has been well researched regardless of the industry or sector; see for example Ngai, Xiu, and Chau (2009) for review. Also see Pendharkar (2009), Wei and Chiu (2002), Hung, Yen, and Wang (2006). All these papers looked critically at data collected for the postpaid segments which is not the market segment that we have considered in this study. In the postpaid segment, the dormancy is well defined as these customers have a contractual agreement with the operators. For churn model of prepaid customers, see Marcin (2010).

1.2 Prepaid Predominant Market and Prepaid Customers

In this article, we focus on the prepaid segment of the mobile customers in Nigeria. It is necessary and important to focus on this segment as this is the pain of all mobile operators in Nigeria and other prepaid predominant markets. In a market where switching cost is less than half a dollar or nearly free, the problem of churn is so great. A prepaid customer owes no explanation to the operator before stopping the usage of the SIM card. Unlike a postpaid customer that has a monthly commitment. A situation where a customer is carrying all the SIMs of the operators but with less number of handsets, inactivity of SIM cards become a serious issue to the operators. Also, the competition activities in Nigeria mobile market are very fierce and dormancy sets in when a customer picks a competition SIM due to an aggressive offer or a promotional activity. The problem of prepaid churn is well defined in Nigeria by the regulatory body. We described this problem in the next subsection based on the Nigeria mobile market.

1.3 Dormancy and Churn Definition in Nigeria Mobile Market

The Nigerian Communication Commission (NCC) is very clear in its definition of an active mobile prepaid customer. A customer is said to be active if the customer has performed a revenue generating event (RGE) within the last three months. Every operator defines the activities that make up its own rge. Events such as recharges, outgoing calls, incoming calls, data usage, incoming and outgoing SMS are some of the rge for all the operators. A prepaid customer is reported as a churner if for three months, the subscriber did not perform any of the activities defined as RGE. This paper focuses on predicting when a customer will stop to perform any of the defined RGE and become a dormant customer.

1.4 Data Mart

Marcin (2010) considered about 1,381 variables in his churn prediction model. This was the highest variables that have been reported so far in the study of churn model. Pendharkar (2009), Wei and Chiu (2002), Hung et al. (2006), considered small number of exploratory variables. In our study, we have considered 1,451 variables. While a considerable amount of the variables was

from the CDR, variables were also collected from the SIM registration server. Our variables from the CDR are rich and detailed. From customer balance information to social network information and customer experience information. Hundreds of variables were gathered on minutes of use. Minutes of use from time of the day, day of the week, local, international and weekends. CDR Information on recharge behaviour, VAS, SMS, roaming, on-net and off-net were collected. We gathered our data from 2012 to 2013.

2. DORMANCY PREDICTION MODEL (DPM)

In our study, we used decision tree algorithm. We actually needed a model that can be easily interpreted and could be adopted as part of the daily activities of the marketing team of an operator. A model that senior management and decision makers can easily follow. Decision trees have clear interpretation which can be expressed in terms of what-if rules. Our data mart is a robust Analytical Data Store (ADS) which will be further discussed in subsection 2.2.

2.1 Before the DPM Model

We analysed 15 hypotheses and carried out a detailed CDR analytics of customer data to validate the correlations with overall customer base. The pre-analytic phase revealed the main behaviour levers and their influence on customer dormancy. In order to understand causes of churn, we later tested the hypothesis based on behaviours and validated these hypothesis with overall customer base performance using the ADS. We later associated relevance to the hypothesis results which formed the basis for the DPM. We classified the associated relevance of the hypothesis result into (i) strong correlation with dormancy (ii) mild correlation with dormancy and (iii) No correlation with dormancy. From the validated hypothesis, below are 7 critical highlights:

1. Customers with lower tenure have higher churn tendency;
2. In certain regions/locations, churn rates are marginally higher than their contribution to the entire customer base;
3. Customers with lower end handsets have higher churn rate;
4. Regular and frequent SMS users are less prone to churn;
5. Non voice usage is a strong indicator of potential churn
6. Customers with more than 2 revenue streams have lower tendency to churn;
7. Customers with data as favourite revenue stream are more risky.

2.2 Data

Our ADS consists of various CDR attributes and demographic variables which are all considered for the training of our models. The ADS was further transformed in order to expand the attributes by computing averages and sums on the defined attributes. This process generates another dataset which we call the Transformed Analytical Data Store (T-ADS). The model learned the pattern exhibited by the T-ADS and generated the final dormancy scores based on the learnt patterns.

2.3 Models

The insights derived from the hypothesis testing became fundamental input in order to design the proper models. A key insight from the pre-analytics phase identified the need to segment the

churning base by value bands. We created three separate models for Very High (VH), High (H) and Low (L) segment. The rest of the strong correlation identified via the hypothesis testing drove the iteration process of the modelling phase.
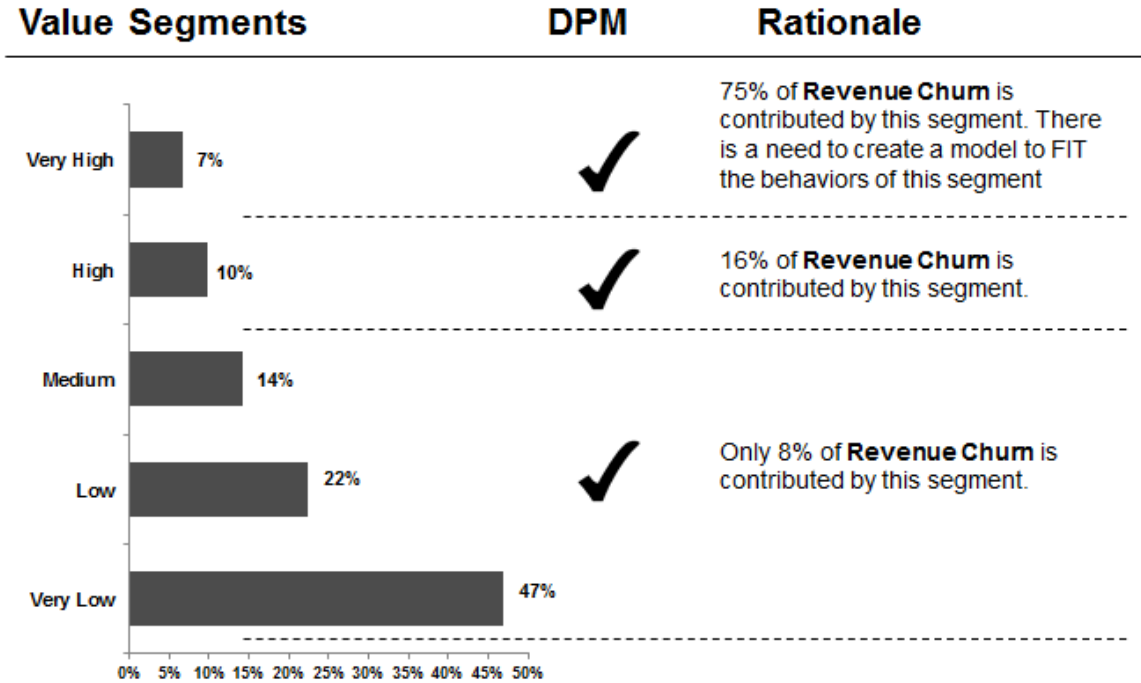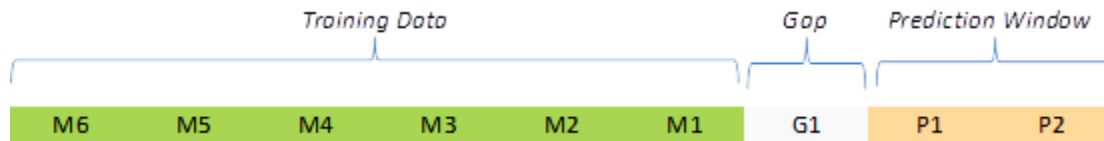

Figure 1. Churning base segmentation

Figure 2. Gap and prediction chart

Training data consists of customers who were active in the last month of the Training Period [M1] and customers who have been on the network for >= 60 days. Gap [1 month] is the period for which no information is shared with the model. This will be intervention period for marketing initiatives. Prediction Window or target refers to customers who will be dormant for 2 month [P1 & P2]. T-ADS was required to carry out the modelling phase. We populated the model with 6 months history of customers. One time exercise of Target Flag creation was done and value bands [VH, H, L] were created. 3 tables were created; each has a separate model built on it. Attributes were tested with Oracle Data Miner (ODM) using decision tree technique.

2.4 Results

We tested our models using the back testing approach. This served as a form of cross-validation on sample dormant customers. We used this approach on the entire prepaid base in December

2012 [M1] that met the defined rules in subsection 2.3. Our models were applied on this sample [M1] to predict customers that went dormant between February and March [M1 & M2]. The output was later compared with the actual customers that went dormant between February and March to determine the accuracy of the models.
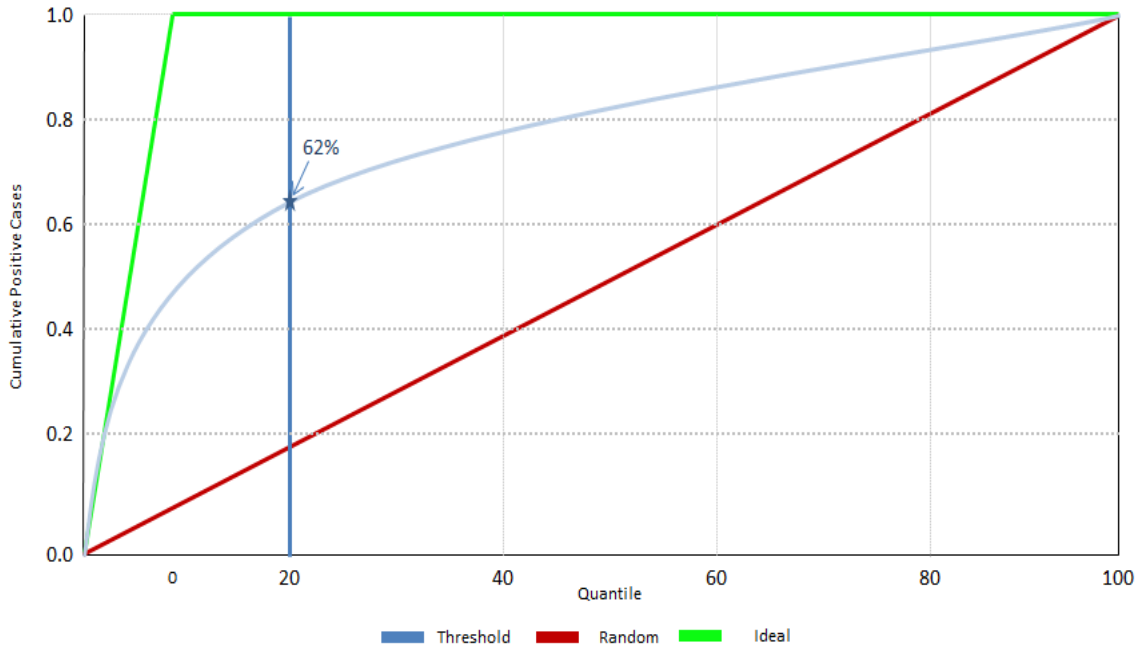

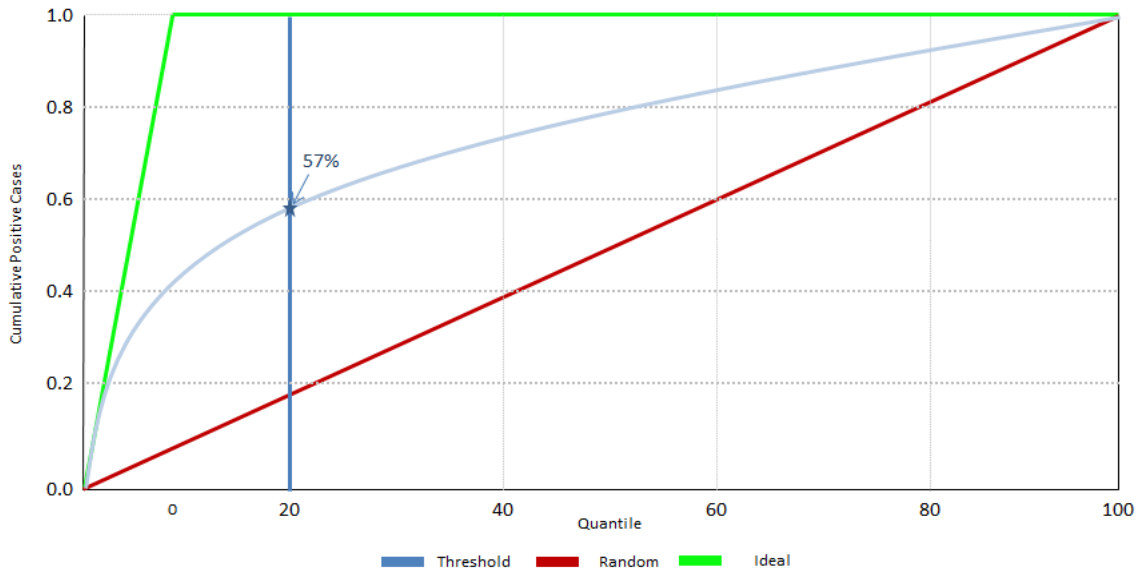
Figure3. Lift curves for Very High Segment
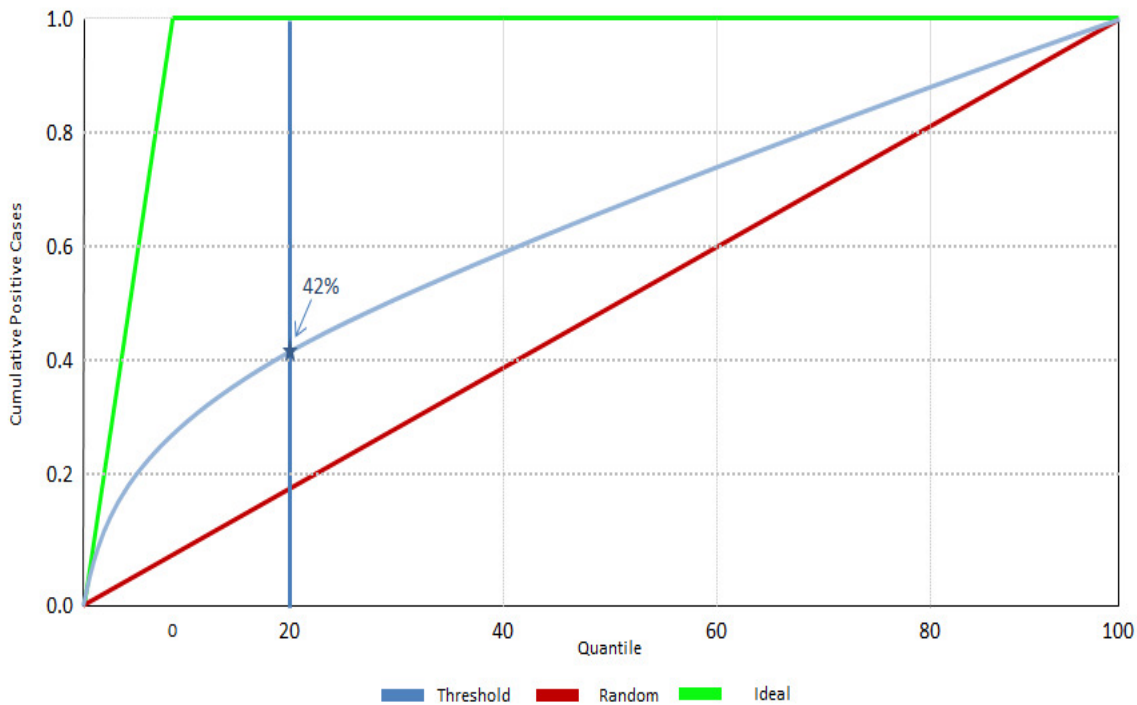

Fig4. Lift curves for High Segment

Fig5. Lift curves for Low Segment

We applied lift curves to our model output so as to determine cumulative gain (%) at set threshold (%). This set threshold helped to identify the percentage of customers from the model output that would yield the optimum lift for retention activities. The model for the [VH] value segment has an accuracy of 79% and a lift of 62% at a 20% threshold. The [H] value segment model gave an accuracy of 76% and a lift of 57% at a 20% threshold. A 68% of accuracy and a lift of 42% at 20% threshold were achieved in the [L] value segment. We may observe that at 20% threshold, the accuracy of the model declined as we moved from [VH] to [L] value segment. For threshold above 20%, the accuracy and the lift decline. At 20% threshold, the models are most useful for the marketing retention activities.

3. SUMMARY, CONCLUSIONS AND DIRECTION OF FUTURE WORK

In this article, we predicted dormancy of prepaid mobile customers in a predominantly prepaid market. We predicted dormancy of three different value segments of prepaid customer base using decision trees approach. The accuracy of the models depends on the value churn not the volume churn. Interestingly, we showed that dormancy can be effectively predicted in a prepaid predominant market with a rich and robust analytical data store. We showed that Dormancy Prediction Model along the value and volume churn can be easily adopted and operationalized within the prepaid predominant mobile market. Our article showed that value can be protected with a focus on the top 40% revenue generating customers with high propensity to churn through VHV and HV DPM when the model output are proactively addressed with appropriate retention offers. It is important that mobile operators in a prepaid predominant market focus on prepaid customer dormancy or churn along their value and volume. This can be easily measured when

churn rate is measured along the value segmentation of the customer base. Targeted campaigns based on customer profiling and risk score with a focus on value churn [VH & H] and then volume churn [L] is an appropriate way to approach the management of churn in a market that is dominated with prepaid customers.

For future work, it would be interesting to know how the marketing team of mobile operators can effectively use the output of these models to reduce dormancy in a prepaid predominant market and the impact on the value churn and volume churn on the overall bottom line when the customer base is segmented along the customer spend. Another area of interest would be the stability of the model in the midst of various competition and promotional activities that are associated with prepaid market.

REFERENCES

[1] Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. Expert Systems with Applications 31, 515-524.

[2] Marcin. O. (2010). Churn models for prepaid customers in the cellular telecommunication industry using large data marts. Expert System with Applications, 37, 4710-4712.

[3] Ngai, E. W. T., Xiu, L., & Chau, D. C. k. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36, 2592–2602.

[4] Pendharkar, P. C. (2009). Genetic algorithm based neutral network approaches for predicting churn in cellular wireless network services. Expert Systems with Applications, 36, 6714-6720.

[5] Wei, C. P., & Chiu, I. T. (2002). Turning telecommunications call details to churn prediction: A data mining approach. Expert Systems with Applications, 23, 103-112

[6] Han & Kamber. Data Mining: Concepts and Techniques. Second Morgan Kaufman Publisher, 2006, pp.383-407.

[7] Kun Chang Lee & Nam Yong Jo. Bayesian Network Approach to Predict Mobile Churn Motivations: Emphasis on General Bayesian Network, Markov Blanket, and What – If Simulation. Lecture Notes in Computer Science, vol. 6405, 2010, pp. 304-313

[8] Bong-Horng Chu, Ming-Shian Tsai & Cheng-Seen Ho. Toward a hybrid data mining model for customer retention, Knowledge-Based Systems, vol. 20, 2007, pp.703-718

[9] Jonathan Burez & Dirk Van Poel. CRM at a pay-TV company: using analytical models to reduce customer attrition by targeted marketing for subscription services. Expert ystems with Applications, vol. 32, 2007, pp. 277-288

[10] Rupesh K. Gopal & Saroj K. Meher. Customer churn time prediction in mobile telecommunication industry using ordinal regression. PAKDD'08 Proceedings of the 12th Pacific-Asia conference on Advances in knowledge discovery and data mining: Springer-Verlag Berlin, Heidelberg, 2008, pp.884-889

[11] Wai-Ho Au, Keith C. C. Chan & Xin Yao. A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction. IEEE Transactions on Evolutionary Computation, vol. 7, No. 6, Dec 2003, pp. 532-545

[12] Xia Guo-en & JIN W. Model of Customer Churn Prediction on Support Vector Machine. Systems Engineering Theory & Practice, vol. 28, 2008, pp. 71 - 77

[13] Hee-Su K. & Choong-Han Y. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommunications Policy, vol. 28, 2004, pp. 636-647

[14] Lima, Mues & Baesens. Domain knowledge integration in data mining using decision tables: case studies in churn prediction. Data Mining and Operational Research, vol. 60, 2009, pp.1096 – 1106

[15] Luo Bin, Shao Peiji & Liu Juan. Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System. Service Systems and Service Management, 2007 International Conference on, 9 – 11 June 2007, pp. 1 – 5.