AN ENHANCED APPROACH FOR CUSTOMER SEGMENTATION THROUGH PREDICTIVE MODELLING WITH COST-BENEFIT ANALYSIS

Dr. J.Meenakumari[1]; Venkatesha K H[2]; Vijayakumar S[3]
[1]*iNurture Education Solutions, meenakumari.j@inurture.co.in*
[2]*iNurture Education Solutions, venkatesha.k.hullumani@gmail.com*
[3]*iNurture Education Solutions, selvavijay1987@gmail.com*

Abstract: Segmentation, targeting and positioning (STP) is the most familiar strategic approach that is being followed for more than a decade for identifying credible customers in banking sector. However, identifying proper segmentation methodology remains as a great challenge due to heterogeneous factors attributing to it. In general, cluster-based approach was used to profile demographical aspects of customers till date. In this study, in-depth analysis is made to ascertain the behavioural aspects of the customers. A case analysis was performed using KS statistics to identify high response segments precisely from the total population. The significant interpretations of the analysis are documented in this paper together with cost-benefit approach. Analytical based ROI approach revealed optimal solution to obtain higher responsive segments at a lower cost. This approach will help the banking sector to further enhance their procedure to find potential customers.
Keywords: Customer segmentation, Clustering, Predictive modelling, Cost-benefit analysis, Behavioural pattern

1. Introduction

In modern marketing scenario segmentation, targeting and positioning is the most familiar strategic approach. It has been accepted as a popular approach in banking sector too. Banks get to know their customers better and manage to reach to their micro details through proper customer segmentation. However, identifying proper segmentation methodology remains a great challenge due to heterogeneous factors attributing to it. More than a decade customers have been segmented through traditional ways, but currently relying just on these categories is not going to yield many actionable insights. Again, the effectiveness of segmentation is not on the quantity of data but how efficiently it could be drilled down to know the customers precisely (Achim Machauer & Sebastian Morgner, 2001)

Usefulness of segmentation is associated with bell curve. If it is too broad, the results are less than insightful - too narrow the value of the insights gained will have minimal bottom-line impact. Hence to visualize customers better exploratory data analysis becomes mandatory. This analysis could also provide a lead in understanding the customers need for loans time to time. At present, banks incur good quantum of money to identify or market their products pertaining to loans. Knowing the customers appropriately well would help in providing good customization and personal attention and it would serve as a special ingredient for successful customer experience management. Loans are important assets of bank and the majority of revenue of the bank occurs through giving loans (Neda Shokrgozarl & Farzad Movahedi Sobhani, 2016) .Hence focussing on this particular most prospective activity is gaining momentum.

Many studies have revealed that majority of revenue occurs to the bank by giving loans however it is subjected to limited resources associated with the bank. Customer segmentation is of utmost importance here to prioritize the customers to make effective decision making. The purpose of segmentation is not all about identification of customers in a given market, but rather in search of deserving customers in most cost effective way .Customer segmentation is a complex activity and requires a good analysis of the market scenario for better outcome. Though lot of analysis has happened and still happening in this aspect there has been a continuous need for precise and more accurate analysis.

In this paper, it is intended to provide a cost effective predictive model which would benefit the banks to a greater extent. Predictive model is developed to segregate non-responsive and responsive customers in order to segment high respondents among the total base of customers. Segmenting is done using various machine learning techniques and validated through rank ordering. High responsive segments is chosen based on decile's the category of segments.

2. Theoretical Background

The main objective of performing customer analysis as part of a business plan is to examine the customer behaviour and segregate prospective customers. The outcome of this analysis also helps to understand the customer requirements better and in turn serves to promote overall business. Identifying and understanding of customers are two key activities in customer segmentation. Segmentation should always have an associated measurable and it should be well distinguishable too (Dagmar Recklies, 2015) Classification and clustering of customers are also significant tasks in the process.

Although demographic data is still the most common characteristics used by banks, behavioural segmentation has proved to achieve better results as it takes into account the heterogeneous patterns of services, their usage and consumption among individuals with the same age and income. Banks are the organizations by excellence mostly capable to apply this method, as it operates in the sector of activity with more organized and complete information about the desired group that is their customers. In general, the banks have an access to the consumption patterns as well as to information about customers' income and attitudes towards investment and saving products etc.,. The results of various recent analyses of commercial banking trends show that proper classification of borrowers is fundamental for better sustainability of the business. Hence customer segmentation and segment based classification becomes inevitable [7].

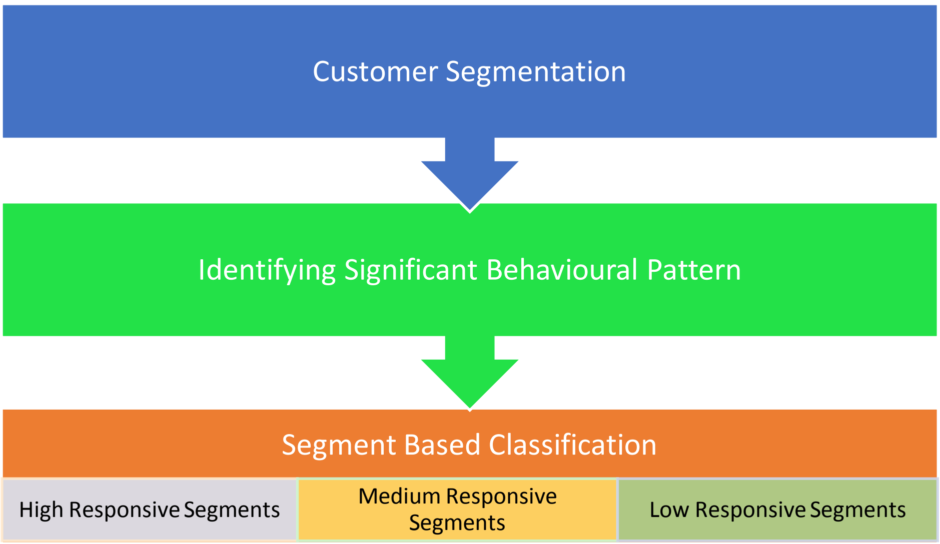

Figure 1: Process Overview of Customer Segmentation

Customer segmentation can be done through multiple ways .However there are certain steps that has to be followed: The initial step being preparation of well collated customer list together with assigning characteristic based score for each customer. The next important step is to define segmentation hypotheses followed by analyses and prioritization of the segments. By default, segmentation is expected to enhance a bank's ability to address the conflict between individual service and cost- saving standardisation (Asiedu, 2016). The Banking and Financial Environment has witnessed so many regulatory changes and competitive dynamics. This has made segmentation practices an imperative determinant in their service offering. Findings from the study indicate that, segmentation practices have immensely impacted on the performance of the selected banks in Colombia in reducing overall operation unit cost, expanding their market shares, retain their customers, better their communications, increase portability and focus. By evaluation of Customer segments, banking system will consider more efficient and crucial factors in decision process to estimate more accurate credential of each group of customers and will grant more appropriate types and amount of loan services to them therefore it is expected these solutions will reduce the risk of loan service in banks. The mortgage sector is one of the most profitable and secure of the banking system (Julia Kagan, 2014). Issuing a housing loan not only helps create long-term relationships with customers and expand existing cooperation in the future, but also minimizes credit risks due to the strong guarantees it entails(John Mylonakis, 2007) .Segmentation and targeting the various sub-segments using different methods each time has become indispensable.

3. Problem Formulation:

It is evident from past research that cluster–based customer segmentation is a prominent approach currently being followed. However there is a clear need for precise focus on high responsive segment prediction in order to reduce cost.

4. Objective:

To identify the profitable segments to target for cross-selling personal loans by adapting cost benefit approach there by reducing the inherent cost and increase success ratio.

5. Methodology:

The data set was fetched from open source repository available for research purpose. The data is related with direct marketing campaigns of a banking institution. The predictive model built on this data set consisted of 20,000 records out of which the responders where 2512 that accounted for 12.5% of the total. The stratified sampling technique was used. The data is extracted from the different subsystems of the bank to detect behavioural pattern of the customers. Several algorithms such as decision tree, Random Forest and Neural Networks were built on the data and random forest proved to give a best fit. The validation methods like K-S statistics, Confusion Matrix, Lift Curve and AUC curve were accessed to check the strength of the model and the results for the same is elaborated in the analysis section. Based on the K-S statistics the cost benefit approach were conducted to get higher success rate with minimal marketing cost.

6. Model Analysis & Validation

Initially prominent data mining techniques like CART, Clustering, and Random forest were explored on the data set. Amongst all random forest gave a good fit. A random forest is a meta estimator that fits a number of classified decision trees on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting.

Dr. J.Meenakumari *et al*, International Journal of Computer Science and Mobile Applications,
Vol.7 Issue. 7, July- 2019, pg. 1-6 ISSN: 2321-8363
Impact Factor: 5.515

Below table depicts the KS value of the Development model to be 59% and Hold-out model to be 46%.

	deciles	cnt	cnt_resp	cnt_non_resp	rrate	cum_resp	cum_non_resp	cum_rel_resp	cum_rel_non_resp	ks
1	10	1442	859	583	60%	859	583	49%	5%	0.44
2	9	1832	459	1373	25%	1318	1956	75%	16%	0.59
3	8	1117	147	970	13%	1465	2926	83%	24%	0.59
4	7	2356	157	2199	7%	1622	5125	92%	42%	0.50
5	6	7329	140	7189	2%	1762	12314	100%	100%	0.00

Table 1: Model Evaluated on Development Sample

	deciles	cnt	cnt_resp	cnt_non_resp	rrate	cum_resp	cum_non_resp	cum_rel_resp	cum_rel_non_resp	ks
1	10	631	304	327	48%	304	327	41%	6%	0.35
2	9	844	182	662	22%	486	989	65%	19%	0.46
3	8	538	68	470	13%	554	1459	74%	28%	0.46
4	7	1054	96	958	9%	650	2417	87%	47%	0.40
5	5	2857	100	2757	4%	750	5174	100%	100%	0.00

Table 2: Model Evaluated on Hold-out Sample

7. Gini Coefficient

AUC ROC number helps to derive Gini coefficient. Gini is ratio of ROC curve and the diagonal line. For a good model, the value of Gini has to be above 60%. Hence, the model built by Random Forest technique can be accepted.

a) AUC value using Random Forest: 0.867

Gini = 2*AUC – 1

= 2*0.867 – 1

= 0.734

8. Kolmogorov Smirnov Chart(KS)

K-S or Kolmogorov-Smirnov chart measures performance of classification models. More accurately, K-S is a measure of the degree of separation between the responders and non-responders distributions. The KS value for the model to be accepted has to be between 45-60%.

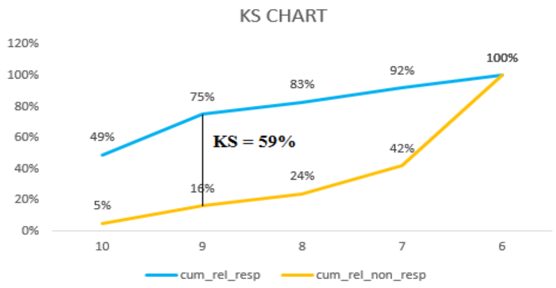

Figure 2: KS Chart of Development Sample

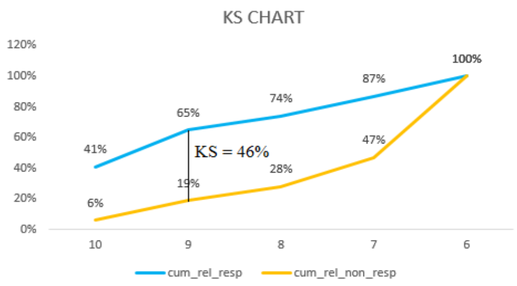

Figure 3: KS Chart of Hold-out Sample

9. Validation of the Model:

Model Validation is an integral part of the model development process. Validation helps to identify and deploy the best model that represents the given dataset and also how good it will be in futuristic .The above model replicates the stability of the model for which the KS value difference among the development and Hold-out should not be more than 15%.

10. Cost Benefit Analysis:

It gives the return on investment approach of implementing the model and depicts a correct understanding of the ROI's of different segments obtained altogether. The below table depicts the sample ROI's obtained for the model developed and same can be implemented with the following assumption note stated below.

Assumption: Costs of targeting per customer: INR 10/- and Expected revenue per convert: INR 2500/-

Segment	Customer	Conversion Rate (%)	Total conversions	Cost of Targeting	Expected Revenue	Profit	ROI (%)
High Response Segment	250000	1.3	3250	2.5 Mn	8.125Mn	5.625Mn	225
Medium Response Segment	250000	0.4	1000	2.5 Mn	2.5 Mn	0	0
Low Response Segment	500000	0.15	750	5 Mn	1.875Mn	-3.125Mn	-ve
Total	1000000	0.5%	5000	10 Mn	12.5Mn	2.5Mn	25

Table 3: Cost Benefit Analysis of the different segments

11. Conclusion:

Based on the in- depth analysis, high responsive segments were classified among the total population. The theoretical model built as indicated in Fig- was validated using KS statistics. The outcome revealed good stability of the proposed model. This model when followed would help the banks to segregate the high responsive segments precisely thereby increasing the profit of the banks by reducing the cost incurred in the process. The analytical –based ROI approach proved to be efficient in deriving optimal succession rate.

References

1. Achim Machauer and Sebastian Morgner (2001) "Segmentation of bank customers by expected benefits and attitudes", International Journal of Bank Marketing, Vol. 19 No. 1, pp. 6-18.
2. Asiedu E (2016) A Study of use and Impact of Market Segmentation Practices on Bank Performance: With Special Reference to Commercial Banks in Colombia. J Bus Fin Aff 5: 162.
3. Dagmar Recklies(2015)," Market Segmentation – Why is it Important?", www.themanager.org/2015/02/market-segmentation.
4. Julia Kagan, 2014," Mortgage-Backed Security (MBS),reviewed ,www. investopedia.com/terms.
5. John Mylonakis (2007). A Bank Customer Analysis and Mortgage Services Evaluation: Implications of Market Segmentation Policies. Banks and Bank Systems, 2(3).

ISSN: 2321-8363
Impact Factor: 5.515

6. Neda Shokrgozar1 & Farzad Movahedi Sobhani,2016, Customer Segmentation of Bank Based on Discovering of Their Transactional Relation by Using Data Mining Algorithms, Modern Applied Science; Vol. 10, No. 10; 2016 ISSN 1913-1844 E-ISSN 1913-1852Published by Canadian Center of Science and Education.
7. https://www.csp.com/customer-segmentation-value.