



CUSTOMER PROFILING AND SEGMENTATION IN RETAIL BANKS USING DATA MINING TECHNIQUES

M Mubasher Hassan
Sr. Assistant Professor, Dept. of ITE
Baba Ghulam Shah Badshah University
Rajouri (J&K)-India

Tabasum M
Lecturer, Dept. of Computer Science
School Education
Govt. of J&K-India

Abstract: The objective of achieving profitability is one of the main targets of any banking sector for longer sustainable existence. The customer satisfaction index determines the longevity in relation of customer-bank and thereby provides the idea of devising new policies and strategies for healthy connection of customers with the bank. Offering the services and products to the customer based on his choice and needs requires understanding the customer. The customer data available in the bank can provide the deep insights to the bank for designing the customized service and products. Deriving useful information from customer data using data mining techniques is of paramount importance in these days. Leveraging existing granular customer data can help banks gain deep actionable customer insights useful to understand the customers and to reveal and unlock opportunities for increasing profitability. Customer segmentation and profiling are vital in achieving two main objectives of CRM (Customer Relationship Management) i.e.; customer retention and customer development. The main aims of customer profiling and segmentation include expanding customer base, design of tailor made products, micro targeting of sales, aligning right channels for right products, increasing effectiveness of cross selling and up-selling, enhanced customer experience by focused customer relationship, prioritizing relationship with high value customers, effectively managing cost with low value customers based on the profiling and segmentation of customers. In this paper we are using data mining techniques i.e. Naïve Bayes classification algorithm for customer profiling and BIRCH clustering algorithm for customer segmentation

Keywords: Customer profiling, customer segmentation, retail

1. INTRODUCTION

banking, data mining, BIRCH algorithm Introduction Customer is the most important asset in banking business and banks around the world are trying to make their business customer centric i.e. based on deep understanding of customer needs with the help of analytics, customization of services and products to meet the requirements of different segments of customers. Providing outstanding service, flexibility in customer orientation, convenience orientation, pricing orientation and relationship orientation is vital for customer retention, churn prevention, deeper market penetration, preventing downward migration in terms of value as well as increasing profitability with existing customers by effective cross selling of products in current competitive business scenario. In present multichannel banking environment, social and demographic characteristics of customers are changing rapidly creating demand for dynamic customer management that can respond to customer needs in an adaptive manner. As the banking customer data is multidimensional, data mining techniques can be used for analysis of customer information useful for achieving goals form (Customer Relationship Management). Customer profiling and segmentation are vital tasks in CRM (Customer Relationship Management) and provides basis for managing trustworthy relationship with existing customers and customer development. CRM focuses on customer retention by enhancing experience of existing customers and increasing profitability by targeting new customers, deeper market penetration, effective cross selling and providing tailored offers to high value customers.

Customer profiling means classification of customers as per their factual and transactional attributes. Customer profiling is an important tooling CRM and data mining techniques can be used to increase accuracy of customer profiling methods as the customer data of banks is very sparse and complex. With the help of data mining tools customer behavior can be analyzed to derive patterns from huge customer records, this information can be used as predictive tool for futuristic

behavior of customers[1]. Retention of highly profitable customers is the key challenge in current highly competitive business scenario and can be achieved by continuous effort of up gradation of customer centric products/services. Understanding the customer is prerequisite for building strong relationship with the customer. When we properly understand

customer needs, product preferences, buying pattern, purchase history etc. we can improve or customize products/services suitable for them that will contribute to customer satisfaction and loyalty that will in turn result in increased profitability and customer retention[2].

Customer segmentation also known as consumer segmentation or client segmentation is key technique to understand customers, gain customer insight for decision making and strategy formulation. Customer Segmentation is an important task in Commit divides customer base into discrete, homogenous customer groups based on various attributes, having similar characteristics or buying preferences. Customer segmentation is defined as partitioning of markets into homogenous sub markets in terms of customer demand and characteristics resulting in identification of customer groups that are similar in nature[3].

Customer segmentation also known as consumer segmentation or client segmentation is key technique to understand customers, gain customer insight for decision making and strategy formulation. Segmentation divides customer base into discrete, homogenous customer groups based on various attributes, having similar characteristics or buying preferences. Customer segmentation is defined as partitioning of markets into homogenous sub markets in terms of customer demand

and characteristics resulting in identification of customer groups that are similar in nature.

Segmentation helps in identification of segments of particular interests to business depending upon business goals[4]. Customer data has many dimensions and depending upon business requirements and goals we can segment data on particular attributes describing particular customer behavior E.g. if main priority of business is customer retention, business may interested in one dimension. So, the first step is to define business goal and then go for segmentation.

Segmentation can be objective (supervised) or non-objective (unsupervised).objective segmentation is used to identify customers who respond to particular services or product offers

or for identification of high risk customers who will not repay loans etc. Non-objective segmentation is used to understand customers, for profiling of customers, to understand specific customer groups that exist within customer base for different marketing processes and channelizing of resources. In this paper non-objective segmentation will be done using cluster analysis techniques.

Segmentation process can be apriori i.e. when number and type of segments is known in advance or adhoc when number and type of segments are based on results of data analysis. We are using apriori segmentation

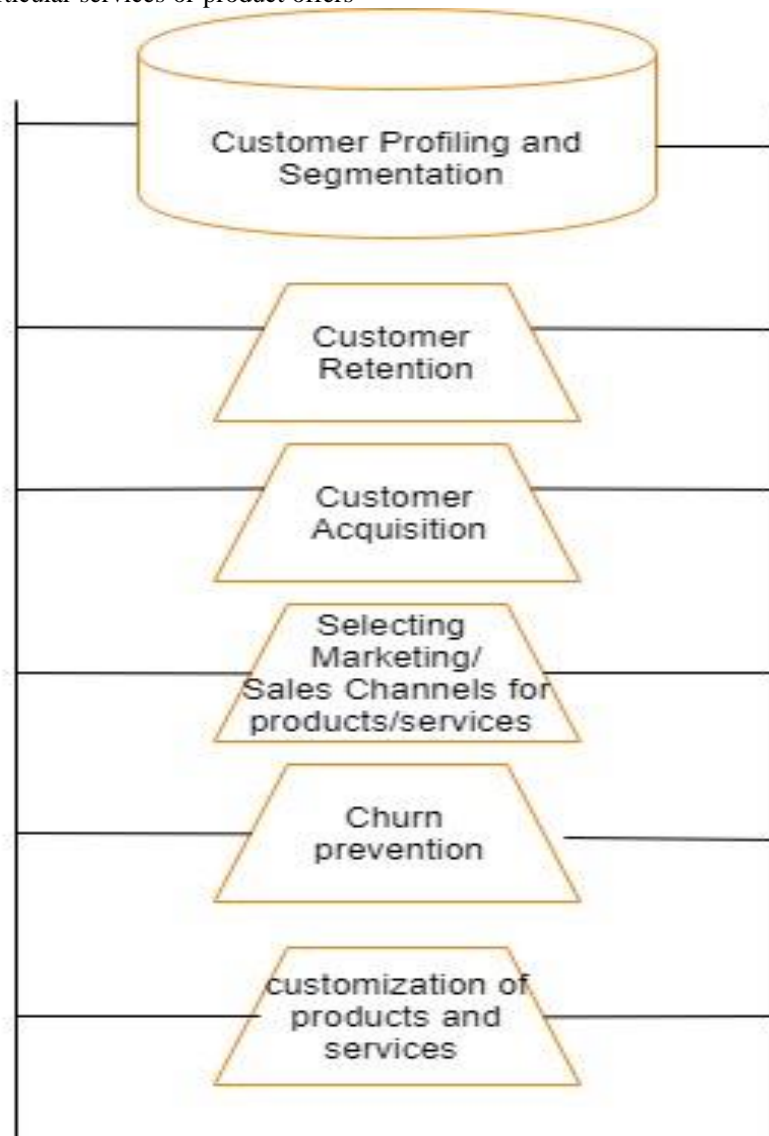


Fig.1. Schematic diagram of customer profiling and segmentation

II. CONCEPTUAL FRAMEWORK

Every customer has some attributes associated with him that comprises of demographic data like age, gender, education, occupation, income, location, psychographic characteristics, financial parameters etc. and transactional data associated with his banking transaction history i.e. buying preferences, purchase history, repayment pattern, churn history etc.

These attributes can be used to build customer profiles that will act as descriptors of customers and will be suggestively used for customer assessment, marketing of suitable products, enhanced experience ,direct marketing, customizing of products/services, cross-selling, deep-selling ,up-selling to increase profitability, churn prevention, risk categorization, default prediction and considering customer’s eligibility for different banking products/services. Then, segmentation of customers can be done using clustering algorithms as per their profiles into

high value customers i.e. highly profitable, low risk customers, low value customers i.e. customers who are less profitable and pose high risk, medium value customers who are moderately profitable and risk level associated with them is also average and negative value customers who incurs more cost to bank than the profit generated and poses high risk also[5].

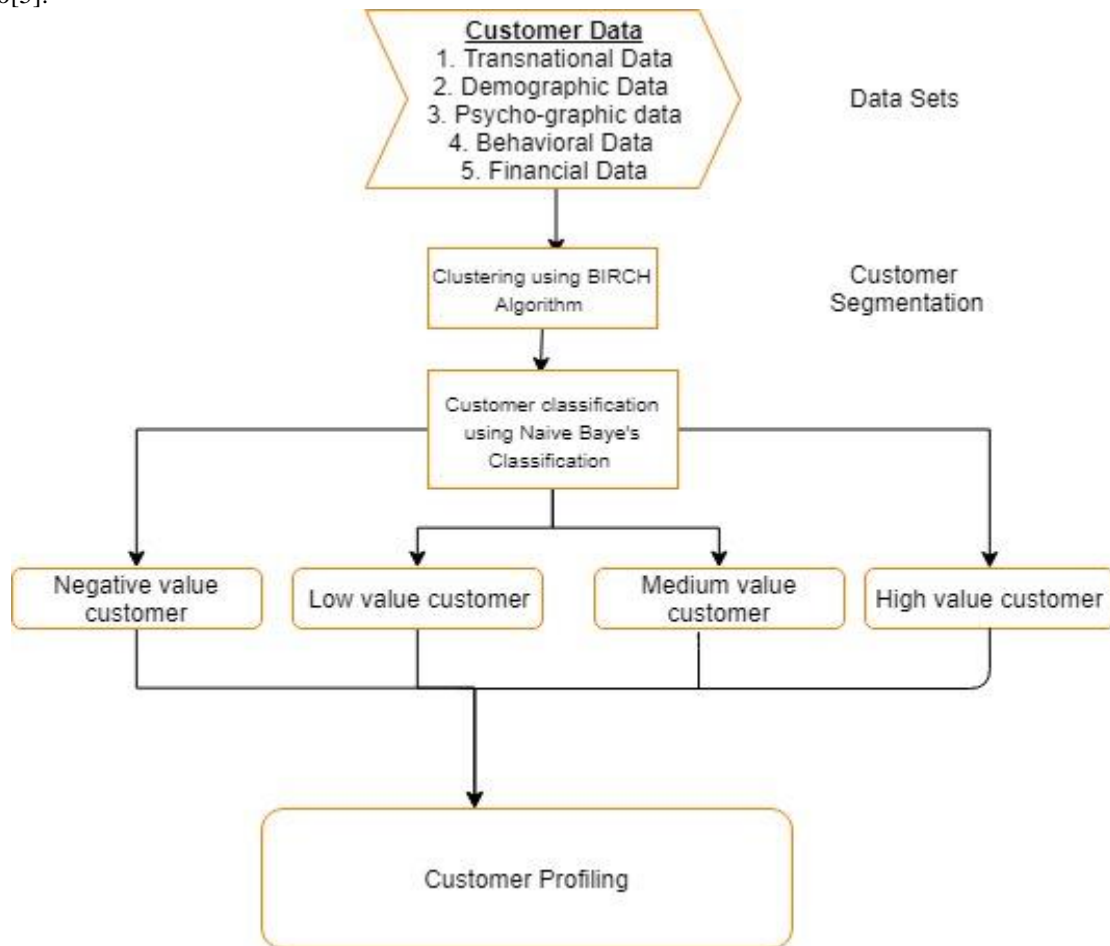


Fig.2. Block Diagram of Customer Profiling

III. Data Mining

Data mining is the process of knowledge discovery from large complex databases helpful to companies like banks to predict customer behavior and decision making from available data by analyzing it and extraction of patterns[6]. Data mining techniques helps banks to increase accuracy of customer profiling and segmentation by using

classification and clustering algorithms[7]. This can be done by personal data and transactional data of customers[8]. Based on this customer data, clustering is the way of identification of segments and split customers into distinct groups. Then, customer profiling is followed to label these segments based on their characteristics. Customer segmentation is done so that customer belongs to one of the following segments:-

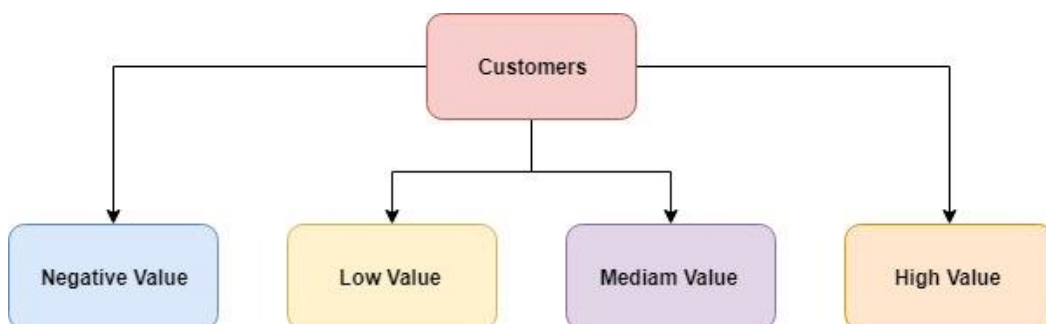


Fig.3. Customer Segmentation

A. High value

These are low risk customers having high net worth, large deposits, loans with the bank and have major contribution towards banks profitability. These customers should be provided best quality customer service, on-priority grievance response and timely offers and incentives to

ensure retention. There should be close monitoring of churn risk and use of maximum resources to prevent churn. Communication should be done through preferred channel.

B. Medium Value

These can be medium risk customers who have maximum of their business with our bank and have scope of up gradation to high value customers by gaining their trust and providing better service and offers than competitors. They provide significant profit to bank and this profit can be increased by cross selling of products/services. Or these are the customers falling into mediocre income group where focus should be on providing feasible products/services they are eligible for and moderate efforts should be made for retention. Communication should be done through preferred but cost effective channel

C. Low Value

These are customers who fall into low income group or high risk customer group or low interest or need of banking products/services. They contribute to little profit of the bank

and there is little scope for upward migration of such customers. Or they can be customers who have maximum portion of their business with other financial organizations and can be migrated to medium value group by making suitable efforts. Limited resources should be allocated for churn prevention. Communication should be done through lowest cost channel.

D. Negative Value

These are the high risk customers who incur more costs to banks in terms of maintenance, operational costs, revenue etc than the profit generated by them. These can be customers with NPA loans or non operational accounts. Efforts should be made for upward migration and reducing costs to serve and to drive up revenue generated.

After value of customer is identified efforts can be made to avoid downward migration of customers and increase upward migration of customers from low or negative value to high value to increase profitability of banking business[9].

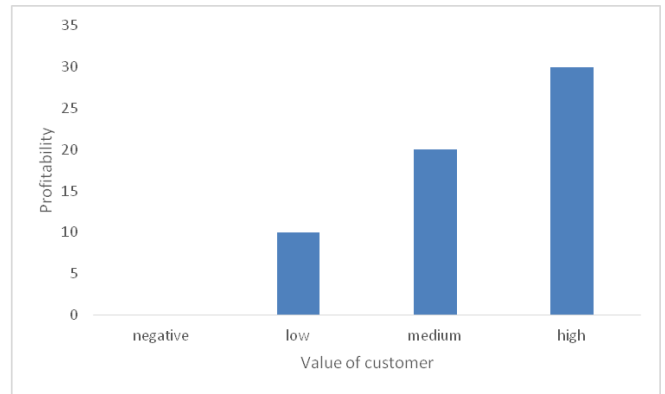


Fig. 4. Segmentation based on customer value

IV. Customer Profiling

This personal data in combination with transactional data can be used to build customer profiles and these customer profiles can be segmented to fall into one of the above mentioned segments using classification and clustering techniques [10].we are using Naïve Bayes algorithm for customer classification.NB classification algorithm is powerful probabilistic algorithm used for predictive modeling and classification problems .Based on Bayes theorem, it is particularly useful with large and high dimensional data sets. Bayesian classifiers are used to predict class membership probabilities such as the probability that a given record belongs to a particular class[11]. Naive Bayes is a supervised classification algorithm for binary i.e. two class and multi-class classification problems. It is easy to use and works well with real time and multi class prediction



Fig.5. Customer Profiling

A. Personal data of every customer consists of following attributes:-

- Age
- Education
- Gender
- No. of dependents
- Occupation
- Marital status
- Income
- Demographic location
- Lifestyle
- Social class
- Creditworthiness
- Relationship time with bank
- Assets

- Liabilities
- Contingent liabilities
- Deposits with bank
- Risk level
- Threshold level

best suitable to different segments of customers, acquiring new customers, design and development of customer tailored products/services, selecting proper marketing and sales channels for products/services[12]

Customers can be segmented according to value, behavior or other characteristics. Segmentation is necessary to prioritize customer handling for customer retention, identify most and least profitable customers, and develop products which are

Table.1. Attributes of Customer profile

DEMOGRAPHIC VARIABLES						
age	gender	marital status	income	education	occupation	no of dependents
PSYCHOGRAPHIC VARIABLES						
lifestyle	Social class	demographic location				
FINANCIAL VARIABLES						
creditworthiness	liabilities	contingent liabilities	risk level	threshold level	deposits with bank	assets
BEHAVIOUR DATA						
relationship with bank	customer behavior	buying preferences	repayment pattern	transaction history		

Clustering techniques use data mining algorithms to analyze data and identify clusters, the clusters identified form basis for segments[14]. Clustering algorithm will find correlation among attributes to identify association rules. Clusters should change dynamically to ensure accuracy in reflecting proper state of customer data[15]

We are using BIRCH (Balanced iterative reducing and clustering using hierarchies) clustering method. BIRCH is unsupervised data mining clustering algorithm to perform hierarchical clustering over large data sets. BIRCH is one of the fastest algorithms and mostly requires single scan on data is space and time efficient i.e.; has less memory and time constraints, reduces I/O cost involved in clustering and is well suited for multidimensional databases.[13] It can be used in spatial databases with noise. BIRCH is scalable clustering method and can be used for concurrent or parallel clustering .BIRCH works by building a dendrogram called (Clustering Feature) tree while scanning data set.CF tree is in memory structure, a height balanced tree that stores the clustering features for a hierarchical clustering. Each entry in CF tree represents cluster of objects or cluster of data points is represented by triple of numbers(N, LS, and SS)

Where N=no of items in sub cluster

LS=Linear sum of points

SS= Sum of the squared of points

The first step builds CF tree out of data points. Clustering features are organized in a CF tree, with two parameters

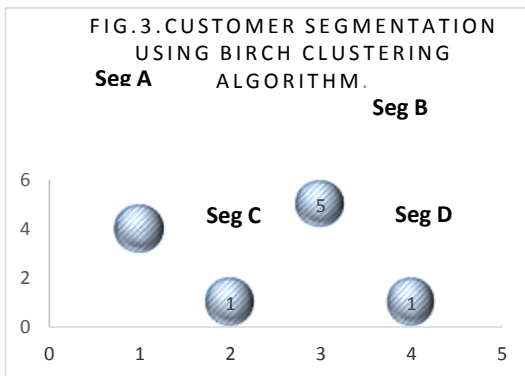
branching factor B i.e. maximum no of children in non leaf node and threshold T i.e. upper limit to radius of cluster in a leaf node, L is no. of leaf node entries.

Second step is optional and condenses initial CF tree into smaller CF trees. The algorithm scans all the leaf entries in the initial CF tree to rebuild a smaller CF tree

In third step, apply existing algorithm on leaf nodes of CF tree to combine these sub clusters into clusters. Optionally refine these clusters.

A single scan yields clustering results and additional scans can be used to refine the clustering results to yield better clusters.

Birch algorithm yields clusters that can be seen as segments of customers e.g. segment a, segment Basement C and segment D in our case.



V. CONCLUSION

The main objective of this paper is to help banks in achieving goals of CRM by customer segmentation and profiling with the help of data mining algorithms. This has been done by classification and identification of customers segments by clustering of customer data and then profiling of customers to label these segments by analyzing behavioral, transactional, psychographic and demographic data of customers. Segmentation and profiling helps in identification of different customer typologies that helps banks in understanding customers to serve them better, design of suitable market strategies, customer retention and customer development.

REFERENCES

- [1] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Expert Systems with Applications Application of data mining techniques in customer relationship management : A literature review and classification," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2592–2602, 2009.
- [2] P. T. Upadhyay, "Customer Profiling and Segmentation using Data Mining Techniques."
- [3] N. Shokrgozar and F. M. Sobhani, "Customer Segmentation of Bank Based on Discovering of Their Transactional Relation by Using Data Mining Algorithms," vol. 10, no. 10, pp. 283–288, 2016.
- [4] M. J. A. B. Gordon S. Linoff, *Data Mining Techniques : For Marketing, Sales, and Customer Relationship Management - 2nd edition*, 2nd editio. John Wiley & Sons, Inc., 2004.
- [5] T. C. Services, "CUSTOMER DATA CLUSTERING USING DATA," vol. 3, no. 4, 2011.
- [6] M. J. A. Berry and G. S. Linoff, *AM*.
- [7] H. Ziafat and M. Shakeri, "Using Data Mining Techniques in Customer Segmentation," vol. 4, no. 9, pp. 70–79, 2014.
- [8] R. Baradaran, K. Zadeh, A.-A. Highway, and A. Faraahi, "Profiling bank customers behaviour using cluster analysis for profitability," pp. 458–467, 2011.
- [9] K. Tsipitsis and A. Chorianopoulos, *Data Mining Techniques in CRM: Inside Customer Segmentation*.
- [10] J. Doe, "Using Data Mining," 2001.
- [11] L. Hou, J. A. Johnson, and S. Wang, "Radio frequency heating for postharvest control of pests in agricultural products: A review," *Postharvest Biol. Technol.*, vol. 113, pp. 106–118, 2016.
- [12] Y. Xi and M. Chen, "Application of Data Mining Technology in CRM System of Commercial Banks," no. Eeta, pp. 370–373, 2017.
- [13] D. Mining and K. Discovery, "BIRCH: A New Data Clustering Algorithm and Its Applications," vol. 182, pp. 141–182, 1997.
- [14] D. Bhardwaj, "Building Data Mining Application for Customer Relationship Management," vol. 3, no. 1, pp. 33–37.
- [15] G. Punj, "Cluster analysis in marketing research : Review and suggestions for application," no. March, 2014.

AUTHORS PROFILE



Malik Mubasher Hassan: Received his B.Tech degree from University of Jammu followed by M.Tech degree from NIT Srinagar in 2007. He is presently working as faculty in the Department of Information

Technology and Engineering (ITE) at Baba Ghulam Shah Badshah University Rajouri (J&K), India-185234. His specialization is in wireless communication, optical wireless, computer Networks, Data Mining and Cloud Computing.



Tabasum Mirza: She has received Masters Degree in Computer Applications from University of Kashmir in 2008. She is presently working as Lecturer in the Department of

Computer Science School Education, Government of Jammu and Kashmir, India. She has a 6.5 years experience of working in JK Bank Pvt. Ltd. Her specialization are software Engineering, Java Programming and Data Mining.