

Data Mining as a Tool to Predict Churn Behavior of Customers

Vivek Bhambri

Research Scholar, Singhania University, Pachari Bari, Jhunjhunu, Rajasthan, India

Abstract: Customer is the heart and soul of any organization. The era of globalization and cut throat competition has changed the basic concept of marketing, now marketing is not confined to selling the products to the customers, but the objective is to reach to the hearts of the customers so that they feel belongingness towards the organizations and hence should remain the loyal customers. In the dynamic market scenarios where companies are coming up with varied options every now and then, customer retention is a critical area to ponder upon, as customers usually churn from one company to another quite often and this too is happening at an alarming rate and is becoming the most important issue in customer relationship management. So prediction of the customer behaviour and hence taking remedial actions before hand is the need of the hour. But the ever growing data bases make it difficult to analyze the data and to forecast the future trends. The solution lies in the use of Data Mining tools for predicting the churn behavior of the customers. This paper throws light on the underlying technology and the perspective applications of data mining in predicting the churn behavior of the customers and hence paving path for better customer relationship management.

Keywords: Churning, Customer Relationship Management, Data Mining, Globalization.

I INTRODUCTION:

Knowing your customer is the buzzword today in the industry. If you are unable to cater to the needs of the customer before your competitors that means you are dead. Reaching out to the right customer at the right time with a right offer is the rule for the survival. The basic concept of marketing which was the production of goods as per the requirements of the customers and then to sell the products to them through various channels, has been changed to touching the hearts of the customers and to create a sense of belongingness, so that they should remain the loyal customers to the organization. So the concept of CRM is evolving at a great pace.

CRM can be defined as the process of predicting customer behavior and selecting actions to influence that behavior for the benefit of the company.

But the irony of the situation is that with the liberalization and globalization, which provide an easy access to information, a wide range of offerings are available in the market due to which the customers are changing their demands at the blink of the eye. So with in no time the customers shift from more service

provider to the other giving birth to the problem of churning.

Churn is defined as the propensity of a customer to stop doing business with an organization and subsequently moving to some other company in a given time period.

The customer churn is becoming a major area of concern for the industry now a days. Identifying the churn before hand and taking necessary steps to retain the customers would increase the overall profitability of the organization. Losing customers not only leads to opportunity lost because of reduced sales, but also to an increased need for attracting new customers, which is five to six times more expensive than customer retention.. In the present day environment, a huge amount of electronic data is being maintained by organizations around the globe. The huge size of these data bases makes it impossible for the organizations to analyze these data bases and to retrieve useful information as per the need of the decision makers. The various commercial organizations are recognizing the need of generating relevant information out of the huge repositories of the data and they are trying to find out the ways and means to provide concise and crisp information as per the requirements. Working in this direction, the business intelligence systems have played a significant role in making the organizations capable of achieving their business objectives, which includes customer retention, profitability and increase in efficiency. Since 1980's the business houses are incorporating the concept of Management Information System, through which they are generating various kinds of reports, which are analyzed for the decision making with in the organization. Due to the vast expansion of the horizons of the data and its multivariate uses, the organizations and the individuals are feeling a need for some centralized data management and retrieval system. The centralization of the data is required basically for better processing and in turn facilitating the user access and analysis. With the rapid development in the hardware and software industry the terms like data warehouse and data mining are gaining importance, which basically equip the decision makers with the What if simulations. Data Mining enables companies to reach consumers with the right product and the right offer at the right time.

II REVIEW OF LITERATURE

According to the definition given by Bill Inmon- Data Warehouse is a subject-oriented, integrated, time variant and non-volatile collection of data in support of management's decision making process.

These data warehouses were able to bring in data from a range of different data sources, such as mainframe computers, minicomputers, as well as personal computers and office automation softwares such as spreadsheet, and integrate this information at a single place to facilitate the user to explore the benefits of operational data to its full extent.

Data mining is defined as the process of extracting previously unknown, valid, and actionable information from large databases and then using the information to make crucial business decisions – Cabena et al.

Data mining is described as the automated analysis of large amounts of data to find patterns and trends that may have otherwise gone undiscovered — Fabris.

The objective of data mining is to identify valid, novel, potentially useful, and understandable correlations and patterns in existing data — Chung and Grey

It is the process of using raw data to infer important business relationships. Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods. Thus, data mining is not only collecting and managing data; it also includes analysis and prediction. The raw data collected or stored in the organization's data base needs to be processed so that it can be converted into useful information and hence can be used for decision making. This process of discovering the hidden knowledge from raw data is called the process of knowledge discovery.

The steps involved in Knowledge discovery are 10:

1. Data Selection The data relevant to the analysis is decided and retrieved from the various data locations.
2. Data Preprocessing: In this stage the process of data cleaning and data integration is done
3. Data Cleaning: It is also known as data cleansing; in this phase noise data and irrelevant data are removed from the collected data.
4. Data Integration: In this stage, multiple data sources, often heterogeneous, are combined in a common source.
5. Data Transformation: In this phase the selected data is transformed into forms appropriate for the mining procedure.
6. Data Mining: It is the crucial step in which clever techniques are applied to extract potentially useful patterns. The decision is made about the data mining technique to be used.
7. Interpretation and Evaluation: In this step, interesting patterns representing knowledge are identified based on

given measures. The discovered knowledge is visually presented to the user. This essential step uses visualization techniques to help users understand.

The Various data mining softwares available in the market are:

M Purna Chander[6] emphasized that

acquiring new customers is more costlier process than retaining existing customers. They emphasized on modeling churn behavior of bank customers in Indian scenario. Many characteristics of customers like demographic details, psychographic, product purchase details, customer perception details are vital in modeling the churn behavior of bank customers. In Indian banking system, the banks have not arranged their data earlier as per the requirements for prediction of their churn behavior, so they highlighted the method for converting raw customer data into meaningful data that suits modeling churn behavior. Then, this meaningful data can be converted into knowledge; predictive data mining techniques are used.

Hanjiewei, Michelin Kamber[13] opines that data mining techniques help companies' particularly banking, telecommunication, insurance and retailing to build accurate customer profile based on customer behavior. Thus it is becoming a necessity in this competitive environment to analyze the data from data warehouse containing hundreds of gigabytes or terabytes of data.

U. Devi Prasad [12] explored that the use of data mining tools to predict the churn behaviour in the banking sector. The raw data from collected from banks can be converted into meaningful useful information and then into knowledge for which predictive data mining techniques are used. They experimented with 2 classification techniques namely CART, and C 5.0. The prediction success rate of Churn class by CART is quite high but C 5.0 had shown poor results in predicting churn customers. However, the prediction success rate of Active class by C 5.0 is more effective than the other technique. They have considered a dynamic time period, which differs for each customer. In this analysis, they have experimented with 2 classification techniques namely CART and C 5.0 on 1,484 samples of bank customers in which 1,163 were active customers and 311 were churn customers. While CART yielded 95.01 per cent classification rate on training data and 91.22 per cent on test data, C5.0 yielded 69.3 per cent classification rate on training data and 68.9 per cent on test data. Thus, a model with higher prediction success rate of Churn class (i.e., CART) has to be chosen for reaping higher benefits. The study predicts the future churn of banking customers that can be checked, by formulating intervention strategies based on churn prediction to reduce the lost revenue by increasing

customer retention. It is expected that, with a better understanding of these characteristics, bank managers can develop a customized approach to customer retention activities within the context of their Customer Relationship Management efforts.

Madan Lal[15]¹⁵ opines that data mining can help in targeting ‘new’ customers for products and services and in discovering customer’s previous purchasing patterns so that the bank will be able to retain existing customers by offering incentives that are individually tailored to each customer’s needs. When Chase Manhattan Bank in New York began to lose customers to competitors, it began using data mining to analyze customer accounts and made changes in its account requirements, thereby allowing the bank to retain its profitable customers. Data Mining is also being used by Fleet Bank, Boston, to identify the best candidates for mutual fund offerings. The bank mines customer demographics and account data along different product lines to determine which customers may be likely to invest in a mutual fund, and this information is used to target those customers. Bank of America’s West Coast customer service call centre has its representatives ready with customer profiles gathered from data mining to pitch new products and services that are the most relevant to each individual caller.

To avoid and prevent churning out of the customers, the organization must know:-

1. What is the profile, tastes, preferences and purchasing behavior of the customer?
2. What is the transaction behaviour of various customers?
3. Which products are often purchased together by the customers of which particular profile?
4. What services and benefits would current customers likely desire?
5. Identifying the customers who are getting all types of services from your company?

III APPLICATION OF VARIOUS TECHNIQUES OF DATA MINING IN THE PREDICTION OF CHURN BEHAVIOUR OF CUSTOMERS

A. Association: Association and correlation is used to find frequently used data items in the large data sets. It is the technique of finding patterns where one event is connected to another event. These type of findings help businesses to make certain decisions regarding pricing, selling and to design the strategies for marketing, such

as catalogue design, cross marketing and customer shopping behavior analysis. However the number of possible association rules for a given dataset is generally very large and a high proportion of the rules are usually of little value. The association may be direct or in direct. Direct such as purchasing a pen and paper, That means when the customer buys paper then he/she will buy a pen also, this association will help the organization in designing the layout of store, by placing these two products adjacent to each other, which will lead to convenience to the customer and organization. As these results can be used for designing the pricing decision and offers can be given based on this study. The organization can find that which customer buys which product most of the times together and hence can provide discounts based on the results. It will increase the satisfaction level of the customers and hence churning of customers will reduce. Mathematically,

$$\text{Churning of Customers} \propto \frac{1}{\text{Satisfaction Level of Customers}}$$

As Satisfaction level increases, churning will decrease.

B. Clustering: - Clustering can be said as identification of similar classes of objects. This is the technique of combining the transactions with similar behavior into one group, or the customers with same set of queries or transactions into one group. Classification approach can also be used as effective mean of distinguishing groups. For Example: The customers of a given geographic location and of a particular job profile demand a particular set of services, like in banking sector the customers from the service class always demand for the policy which ensures more security as they are not intending to take risks, like wise the same set of service class people in rural areas have a the preferences for some particular brands which may differ from their counterparts in urban areas. This information will help the organization in cross-selling their products, Instead of mass pitching a certain “hot” product, the bank’s customer service representatives can be equipped with customer profiles enriched by data mining that help them to identify which products and services are most relevant to callers. The different groups will be offered customized services. The customers can be clustered in to groups according to their transaction behavior and hence dissatisfied customers can be dealt accordingly.

C. Forecasting

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more dependent or independent variables. In data mining independent variables are attributes already known and response

variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction. For instance, sales volumes, stock prices, and product failure rates are all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. These techniques of data mining will help in discovering patterns from which one can make reasonable predictions regarding the churning of customers.

D. Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to this type of analysis. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data is analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable, the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fraudulent and valid activities determined on a record-by-record basis.

IV Challenges in the Prediction Process

The data mining technology will be a boon for the organization in predicting the customer behavior but while implementing this concept following challenges will occur, which needs to taken care while implementing this technology:

1. To predict the target customers who are likely to churn in the near future, the customer behavior in the recent past is to be analyzed. The problem is that how to fix a specific period of time to analyze the customer's activities and moreover fixing a particular time period for all the customers will not be appropriate⁶. For example a model is designed using 100 customers out of which 70 are active i.e. they are continuing business with the organization and the rest 30 are churn customers i.e. who have stopped doing business with the organization. So there is a need to scan the activities done by churned out customers in last three months. Let us say from March, 2011 to May 2011. Now let us consider that out of 30 churned customers, say 20 of them have churned away in March itself. This means, the model will not be fully trained with the behavior of churn customers before churning as only one month's

activity is analyzed. This problem occurred because of fixing the timeline before hand as shown in figure 1. So it would be better if we have a dynamic time period, which differs for each customer. This concept would be better explained by continuing the above example. If a customer has churned away in March 2011, from that point of time, the past 3 months activity is to be considered i.e., transaction activities done in January to March, 2011 should be considered. That means the variable time slot for each customer should be designed so that behavior can be scanned dynamically. This can be seen in figure 2.

2. Undoubtedly a huge amount of data is being collected but in spite of that sometimes relevant data required for the current analysis is not available. For example, if we are finding out the reasons for a customer closing an account, there may be many reasons other than the performance issues related to the bank like a person creates an account for a specific purpose and closes it immediately after the purpose is solved or a person is relocated and has to move to another place and hence closes all the accounts or a customer may stop transacting with the bank just because of the unavailability of bank's ATMs at important places and hence close his/her account. The problem here is that, in real world scenario, the bank does not always capture this kind of feedback data. Hence, no further analysis can be done and this type of churning behaviors could not be stopped. This leaves us in a situation where we need to think which kind of churn patterns are possible to identify. Hence prediction of churn becomes a very challenging problem in banking sector.

3. Sometimes to predict the customer behavior, the data from various sources is to be explored and is to be integrated to derive a particular result. And moreover this data is heterogeneous in nature and while transferring the data, large amount of noisy data is also added which creates problem. So data is to be preprocessed by filtering the noisy data. This step is usually time consuming.

4. It is rightly said that everybody want to be understood but nobody wants to be known. Customers are really in a fix while giving their information to the companies. As the organizations are using the personal information provided by the customers for their vested interests. There is war like situation between the data miner and the subjects, as while mining the data the privacy of the subject is at the stake. So customers refrain from providing their strategic information and hence posing a great hindrance to the process of predicting the churn behavior of the customers.

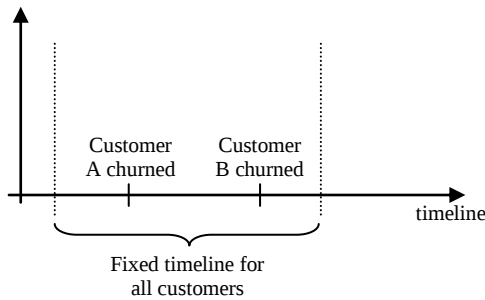


Fig 1: Fixed timeline for all

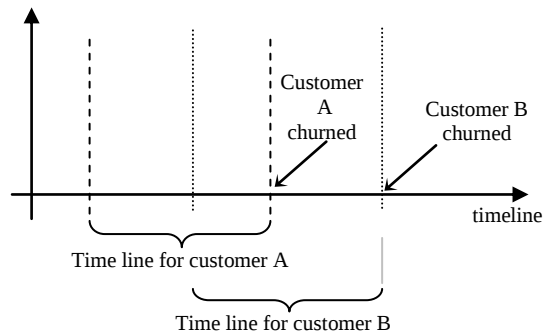


Fig 2 Dynamic timeline

5. The results obtained from the various data mining tools are subject to validation, as these are to be tested before acceptance.

6. While studying the customer behavior usually the purchasing or transactional behavior is considered, but this analysis needs deep study of customers and their circumstances. The transactional behavior of the customer is based on his social, physiological, attitudinal and lot many other non-quantifiable factors, which should be included in the study..

7. The data captured during various transactions done by the customer does not give sufficient evidence to predict the churn behavior of the customers. So the need is to convert the raw data available into knowledge through which the required inferences can be obtained

V Conclusion

Data mining techniques can be of immense help to the organization in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases. Organizations can use these techniques for acquiring new customers, fraud detection, providing segment based products for better targeting the customers, analysis of the customers' purchase patterns over time for better retention and relationship, detection of emerging trends to take proactive approach in a highly competitive market and hence converging all these activities for predicting customer churn behavior before hand. The organizations usually face various implementation problems, which should be addressed before going for the final execution. But by judiciously selecting the data mining techniques and their proper implementation can prove to be a boon for the organization and they will be able to avoid churning of customers that will lead to long term relationship of the customers with the organization.

References:

- [1] Bharati M. Ramageri [2010]-Data Mining Techniques and Applications- Indian Journal of Computer Science and Engineering- Vol. 1 No. 4 301-305.
- [2] Berry, M.J.A. and Linoff, G., Mastering Data Mining: The Art and Science of Customer Relationship Management, Wiley Computer Publishing, New York, NY, 2000.
- [3] David Hand, Heikki Mannila, Padhraic Smyth. (2001). Principles of Data Mining.
- [4] Devendra Kumar Tiwary- Application of Data Mining In Customer Relationship Management (CRM)-, Advances in Computational Sciences and Technology Volume 3 Number 4 (2010) pp. 527-540.
- [5] M. Purna Chandar, Arijit Laha and P. Radha Krishna -Modeling churn behavior of bank customers using predictive data mining techniques", National Conference on Soft Computing Techniques for Engineering Applications.
- [6] Naveeta Mehta, Ruchi Mittal [2011]- Segmenting and Profiling Customers of A Retail Store Using Data Mining Approach- International Journal of Research in IT & Management-Volume 1, Issue 6.
- [7] Rajanish Dass [2006], Data Mining In Banking And Finance: A Note For Bankers- Technical note, Note No.: CISG88., April 2006
- [8] R.K. Mittal, Rajeev Kumra[2001], E-CRM In Indian Banks-An Overview, - Delhi Business Review Vol. 2, No. 1.
- [9] S. P. Deshpande and V. M. Thakare[2010]-Data Mining System And Applications: A Review-International Journal of Distributed and Parallel systems (IJDPs) Vol.1, No.1, September 2010.
- [10] Suresh Chandra Bihari[2011]- Technology In The Banking Sector In India- How Profitable it is for The Customer- Asian Journal of Business and Management Sciences, Vol. 1 No.2 [56-76] 2011 .
- [11] U. Devi Prasad, S. Madhavi Prediction Of Churn Behavior Of Bank Customers Using Data Mining Tools- Business Intelligence Journal - January, 2012 Vol.5 No.1.
- [12] Hanjiewei, (2001) Michel Kamber, Data mining concepts & Techniques: San Diego.
- [13] Yogita Narang, Atul Narang, Shalini Nigam[2011]- Gaining the Competitive Edge through CRM – A Study on Private Sector Banks- IJRFM Volume 1, Issue 3 (July, 2011)
- [14] Madan Lal Bhasin (2006), "Data Mining: A Competitive Tool in the Banking and Retail Industries", The Chartered Accountant October, 2006.