CUSTOMER BEHAVIOR MODEL USING DATA MINING

Milan Patel[1], Srushti Karvekar[2], Zeal Mehta[3]

[1, 2, 3] *Institute Of Technology, Nirma University (India)*

ABSTRACT

Customer behavior models seek common behaviors among particular groups of customers in order to predict how a similar customer will behave under similar circumstances. This report shows the problem of customer relationship management (CRM) and how data mining tools are used to support the decision making. We describe the methods towards predicting customer's behavior, such as collection and preparation of data, segmentation and profiling modeling. This report covers discussion about web mining which can be considered as a separate section due to its current popularity e-commerce. Data mining technologies extract hidden information and knowledge from large data stored in databases or data warehouses, thereby supporting the decision making process of firms. Data mining helps marketing professionals improve their understanding of customer behavior. In turn, this better understanding allows them to target marketing campaigns more accurately and to align campaigns more closely with the needs, wants and attitudes of customers and prospects.

Keywords: ***Customers, CRM, Data Mining, E-Commerce, Profiling***

I. INTRODUCTION

Traditional method of conducting business and industrial operation have undergone a sea change due to globalization of business extensive use of internet and telecommunication networks and use of information technology . The economics of customer relationships are changing in fundamental ways, and companies are facing the need to implement new solutions and strategies that address these changes. Firms today are concerned with increasing customer value through analysis of the customer lifecycle and study the customer's psychological mindset and see that is there any technical format by which we can analyze his buying behavior. The decision makers are required to read quickly to mission critical needs due to rapidly changing volatile and competitive markets. The tools and technologies of data warehousing, data mining, and other customer relationship management (CRM) techniques afford new opportunities for businesses to act on the concepts of relationship marketing. The old model of "design-build-sell" (a product-oriented view) is being replaced by "sell-build-redesign" (a customer-oriented view). Data mining has quickly emerged as highly desirable tools for using current reporting capabilities to uncover and understand hidden pattern in vast database. These patterns are then used in models that predict individual behavior with high accuracy. The result data mining helps in decision making helps in Customer Relationship Management (CRM). The advent of the Internet has undoubtedly contributed to the shift of marketing focus. As on-line information becomes more accessible and abundant for consumers, it keeps them informed. Collecting customer demographics and behavior data helps in targeting easily. This kind of targeting also helps when devising an effective promotion plan to meet tough

competition or identifying prospective customers when new products appear. How customer relationship management and data mining help in decision making are discussed in further sections. Section 1 contains introduction of the paper. Information about data mining is given in section 2. Customer relationship management is discussed in section 3. How data mining and CRM are related with each other is discussed in section 4. Section 5 consists of various data mining techniques. Web mining which is part of data mining is discussed in section 6. And finally conclusion of this paper is given in section 7.

II. DATA MINING

Data mining is defined as a sophisticated data search capability that uses statistical algorithms to discover patterns and correlations in data. Data mining discovers patterns and relationships hidden in data, and is actually part of a larger process called "knowledge discovery" which describes the steps that must be taken to ensure meaningful results. In simple terms, data mining is another way to find meaning in data. Data mining does not find patterns and knowledge that can be trusted automatically without verification. Data mining helps business analysts to generate hypotheses, but it does not validate the hypotheses. For example, as an automobile manufacturer, it is surprising to know that a man with children tends to buy a sports car rather than a man with no children. Hence, this pattern in valuable. Data mining is primarily used day by day comprise with a strong consumer focus retail, financial, communication & marketing organizations. Data mining techniques are the result of a long research and product development process. The origin of data mining lies with the first storage of data on computers, continues with improvements in data access, until today technology allows users to navigate through data in real time.

III. CUSTOMER RELATIONSHIP MANAGEMENT

3.1 Definition

Customer relationship management (CRM) is a process that manages the interactions between a company and its Customers. The primary users of CRM software applications are aiming to automate the process of interacting with Customers. CRM comprises a set of processes and enabling systems supporting a business strategy to build long term, profitable relationships with specific customers. CRM is defined by four elements of a simple framework: Know, Target, Sell, and Service. CRM requires the firm to know and understand its markets and customers. In selling, firms use campaign management to increase the marketing department's effectiveness. Finally, CRM seeks to retain its customers through services such as call centers and help desks. CRM is essentially a two-stage concept. The task of the first stage is to master the basics of building customer focus. This means moving from a product orientation to a customer orientation and defining market strategy from outside-in and not from inside-out. Companies in the second stage push their development of customer orientation by integrating CRM across the entire customer experience chain, by leveraging technology to achieve real-time customer management, and by constantly innovating their value proposition to customers.

3.2 Components of Customer Relationship Management

Customer relationship management is a combination of several components. Before the process can begin, the firm must first possess customer information. There are several sources of internal data:

- summary tables that describe customers (e.g., billing records)
- customer surveys of a subset of customers who answer detailed questions
- behavioral data contained in transactions systems (web logs, credit card records, etc).

Most firms have massive databases that contain marketing, HR, and financial information.

Next, the CRM system must analyze the data using statistical tools, OLAP, and data mining. The firm should employ data mining analysts who will be involved but will also make sure the firm does not misuse the data retrieved. Thus, having the right people who are trained to extract information with these tools is also important. The end result is segmentation of the market, and individual decisions are made regarding which segments are attractive. The last component of a CRM system is campaign execution and tracking. These are the processes and systems that allow the user to develop and deliver targeted messages in a test-and-learn environment. There are software programs that help marketing departments handle the feedback procedure. Campaign management software manages and monitors customer communications across multiple touch points, such as direct mail, telemarketing, customer service, point-of-sale, e-mail, and the Web.

IV. DATA MINING AND CUSTOMER RELATIONSHIP MANAGEMENT

The application of data mining tools in CRM is an emerging trend in the global economy. Analyzing and understanding customer behavior and characteristics serves as the foundation for the development of a competitive CRM strategy which helps to acquire and retain potential customers and maximize customer value. Appropriate data mining tools, which are good at extracting and identifying useful information and knowledge from enormous customer databases. These databases serve as the best supporting tool for CRM. As such, the application of data mining techniques in CRM is worth pursuing in a customer-centric economy.

For a customer-centric economy, we need a framework for understanding customer behavior.

In general, there are four key stages in the customer lifecycle:

1. ***Prospects***—people who are not yet customers but are in the target market
2. ***Responders***—prospects who show an interest in a product or service
3. ***Active Customers***—people who are currently using the product or service
4. ***Former Customers***—may be "bad" customers who did not pay their bills or who incurred high costs; those who are not appropriate customers because they are no longer part of the target market; or those who may have shifted their purchases to competing products.

The customer lifecycle provides a good framework for applying data mining to CRM.

On the "input" side of data mining, the customer lifecycle tells what information is available. On the "output" side, the customer lifecycle tells what is likely to be interesting.

Data mining can predict the profitability of prospects as they become active customers, how long they will be active customers, and how likely they are to leave. It will help the organization identify patterns in their customer data that are predictive. For example, a company can concentrate its efforts on prospects that are predicted to have a high likelihood of responding to an offer rather than contacting any random prospect. Data clustering can also be used to automatically discover the segments or groups within a Customer data set. Rather

than one model to predict which Customers will churn, a business could build a separate model for each region and Customer type. Then instead of sending an offer to all people that are likely to churn, it may only want to send offers to Customers that will likely take to offer. And finally, it may also want to determine which Customers are going to be profitable over a window of time and only send the offers to those that are likely to be profitable. .Businesses employing data mining may see a return on investment, but also they recognize that the number of predictive models can quickly become very large.

V. DATA MINING TECHNIQUES

CRM consists of four dimensions:
(1) Customer Identification
(2) Customer Attraction
(3) Customer Retention
(4) Customer Development

CUSTOMER IDENTIFICATION: Elements for customer identification include target customer analysis which involves seeking the profitable segments of customers through analysis of customers' underlying characteristics, and customer segmentation which involves the subdivision of an entire customer base into smaller customer groups or segments.

CUSTOMER ATTRACTION: After identifying the segments of potential customers, organizations can direct effort and resources into attracting the target customer segments. This could be done by direct marketing so as to promote the customers to place orders through various channels. (e.g. direct mail or coupon distribution).

CUSTOMER RETENTION: Central concern of CRM is customer satisfaction which refers to the comparison of customer's expectations with his or her perception of being satisfied, is the essential condition for retaining customers.

CUSTOMER DEVELOPMENT: Customer development includes customer lifetime value analysis, up/cross selling and market basket analysis. Customer lifetime value analysis means the prediction of the total net income a company can expect from a customer. Up/Cross selling refers to promotion activities which aim at augmenting the number of associated or closely related services that a customer uses within a firm. Market basket analysis aims at maximizing the customer transaction intensity and value by revealing regularities in the purchase behavior of customers.

These four dimensions can be known as a closed cycle of a customer management system. They share the common goal of creating a deeper understanding of customers to maximize customer value to the organization in the long term. Data mining techniques can help to accomplish such a goal by extracting hidden customer characteristics and behavior from large databases.

The generative aspect of data mining consists of the building of a model from data. Each data mining technique can perform one or more of the following types of data modelling:
(1) Association;
(2) Classification;
(3) Clustering;
(4) Forecasting;

(5) Regression;

(6) Sequence discovery;

(7) Visualization.

Choices of data mining techniques should be based on the data characteristics and business requirements.

Association: Association aims to establishing relationships between items which exist together in a given database. It is intended to identify strong rules discovered in database using different measures.

Classification: Classification is one of the most common learning models in data mining. It aims at building a model to predict future customer behavior through classifying database records into a number of predefined classes based on certain criteria. It represents the largest part of problems to which data mining is applied today creating models to predict class membership.

Clustering: Clustering is the task of grouping a set of objects in a such a way that objects in the same group(cluster) are similar to each other than to those in other clusters.

Forecasting: Forecasting estimates the future value based on a record's patterns. It deals with continuously valued outcomes. It relates to modelling and the logical relationships of the model at some time in the future.

Regression: Regression is a kind of statistical estimation technique used to map each data object to a real value provide prediction value. The regression functions are used to determine the relationship between the dependent variable and one or more on independent variable. Regression process succeeds when a significant relationship between variable and dependent variables is a tested one.

Visualization: Visualization refers to the presentation of data so that users can view complex patterns. It is used in conjunction with other data mining models to provide a clearer understanding of the discovered patterns or relationships.

Segmentation: Segmentation is a process of identifying finite sets of data clusters. For example, Customer can be clustered using following clustering criterion: Buying behavior, Value of purchase, preference for high value, Preference for discount/ bargain purchase.

Link analysis: Link analysis is a process of finding the links between two sets of variables the link relationship may be of following types:

1 Lang and lead- sale of umbrella lags the rainfall

2 Moving together- Bread & Butter

Configured link- .drinks, chips and soda.

VI. WEB MINING

Internet plays major roles in today's business. It offers huge business opportunities. So this leads us to Web mining. It is the process of discovering information from the WWW and analysing that information for business purposes. Web mining consists of two parts : Web Content mining and Web Usage mining. Web Content mining includes discovering and organizing Web-based information. Web Usage mining includes analysing the behavioural patterns from data gathered about internet users. So it is more related to customers profiling. Because of this when we refer web mining it is same as web usage mining.

6.1 Internet Marketing

For e-commerce traders, Web mining techniques are needed for maintaining customer relationship management strategies. These techniques discover relationship, patterns and rules within Web data for 3 marketing actions:

1) Finding association rules for customer attraction
2) Finding patterns for customer retention
3) Finding classification rules and data clusters for cross-selling

6.1.1 Customer attraction

The two important parts of attraction are the selection of new customers and the acquisition of the selected customers. One strategy to perform this task is to find common characteristics in already existing visitors' information and behavior of profitable and non-profitable customers. Then customers are given labels like 'no customers', 'visitor once', 'visitor regular' on the basis of their visiting behavior. Depending on the labels, a dynamically created web-page is displayed which has contents depending on found associations between offered products and browser information.

6.1.2 Customer retention

Customer retention is the task of trying to keep the online buyer as loyal as possible. This is the difficult task in e-commerce. Here web-page is displayed with offers by considering associations across time which is known as sequential patterns.

6.1.3 Cross sales

The purpose of cross-sales is to horizontally and/or vertically distinguishing selling activities to an current customer base. For discovering potential customers, characteristic rules of current cross-sellers are discovered, which is done by the application of attribute oriented induction. In attribute-oriented induction, a simple Web-page is replaced by its corresponding general page based on the page hierarchy. The entire set of these rules can be used as the model to be applied on incoming requests from current customers.

6.2. Web Data Collection

Internet provides a variety of tools to gather information about Internet users and actual customers. Data that can be collected are:

- Http-protocols which contain information about the users' OS, browser and browser version.
- Cookies contain information about the Internet user. A cookie is a tiny file which contains information about what an user does on a web site. Thus it is the most efficient way to discover Internet users.
- Queries to a web server e.g. online search for products.
- Number of hits is related to how often web site elements are requested on the server.
- Page view is the number of requests of a whole web-page.

6.3. Web Data Processing

Before any important discovery takes place from collected Web data, the data goes through a pre-processing phase to filter the data from irrelevant or redundant entries. Then the data is organized appropriately according to the application (Association Rules and Sequential Patterns require the input data to be in different forms).

6.4. Discovering Association Rules

In Web mining, discovering association rules turns into finding the correlations among accesses to various files available on the server by a given client. Since usually transaction-based databases contain very large amounts of data, current association rule discovery techniques try to prune the search space according to support for items under consideration. In Web mining, the hierarchical organization of the files can be also used for pruning.

6.5. Discovering Sequential Patterns and Time Sequences

Generally, transaction-based databases collect data over some period of time, and the time-stamp for every transaction is explicitly available. Given such a database of transactions, the problem of discovering sequential patterns is to find inter-transaction patterns such that the presence of a set of items is followed by another item in the time-stamp ordered transaction set. Usually, analyses are made using data taken within a certain time gap. In Web server transaction logs, every visit by a client is stored over a period of time. The time-stamp attached with a transaction in this case will be a time interval which is determined and attached to the transaction during the data cleansing process. The techniques used in sequential pattern discovery are similar to those used in association rule discovery or classification, except in this case the discovered rules have to be further classified based on the time stamp information. For better decision making, non-typical patterns have to be discarded. To do so, less frequent sequences are removed based on a minimum support threshold. The sequence is said frequent if its support is higher than the threshold. In order to find frequent sequences, one needs to find all data sequences satisfying the minimum support.

6.6. Classification and Clustering

After the discovery of hidden common patterns among data items, classification is used to develop a profile for items belonging to a particular group according to their common attributes. This profile can then be used to classify new data items that are added to the database. In Web mining, a profile is built for clients who access particular server files based on demographic information about those clients. In some cases, valuable information about the customers can be collected automatically from the client's browsers by the server. This includes information available on the client side in the cookie files, history files etc. For clustering other algorithms are used.

VII. CONCLUSION

The main contribution of this paper lies in the focusing important issues to improve decision making to optimize your relationships with Customer in highly Customer based business. The data mining system is useful to Business house to find out the association of the customers with different products and how customers are shifting from one brand to another brand of product to satisfy their need. It is being used increasingly in business applications for understanding and then predicting valuable data, like Customer buying actions and buying tendency, profiles of customers, industry analysis, etc. Application of data mining techniques in CRM is an emerging trend in the industry. It aims to give a research summary on the application of data mining in the CRM domain and techniques which are most often used. Data mining represents the link from the data stored over many years through various interactions with customers in diverse situations, and the knowledge necessary

to be successful in relationship marketing concepts. As customers and businesses interact more frequently, businesses will have to leverage on CRM and related technologies to capture and analyze massive amounts of customer information. Businesses that use customer data and personal information resources effectively will have an advantage in becoming successful. In this paper, we have shown that data mining can be integrated into customer relationship management and enhanced the process of CRM with betterment. So we can conclude that since CRM and data mining are very useful for decision making.

REFERENCES

[1] M-S Chen, J. Han and P. S. Yu. "Data Mining: An Overview from Database Perspective". IEEE Trans. On Knowledge And Data Engineering, vol. 8, pp. 866-883, 1996.

[2] H. P. Crowder, J. Dinkelacker, M. Hsu. "Predictive Customer Relationship Management: Gaining Insights About Customers in the Electronic Economy", in DM Review in February 2001

[3] B. Mobasher and N. Jain and E. Han and J. Srivastava, "Web mining: Pattern discovery from world wide web transactions", Technical Report TR-96050, Dep. of Computer Science, University of Minnesota, M inneapolis, 1996.

[4] Masseglia, P. Poncelet, and M. Teisseire. "Web Usage Mining: How to Efficiently Manage New transactions and New Customers". research report of LIRMM, Feb. 2000. Short version in Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'00), Lyon, France, September 2000.

[5] P. Holmes, "Customer Profiling and Modeling", in DMG Direct, Direct marketing Association,