

TARGET MARKETING USING DATA MINING

RITIKA PANDEY

Computer Science and Engineering, Bipin Tripathi Kumaon Institute of Technology, Dwarahat, Uttarakhand, 263653, India

Abstract - Target marketing, a method of attracting customers who are considered likely to buy a product. In this paper, we propose a theoretical framework for gathering consumer data based on needs and preference; we also provide methods for discovering relationships between customers' preferences. With rapid globalization and competition in the field of business, it is necessary to reinforce and redefine customer relationship. Today, customers are in charge so it is necessary that the business strategy is more customer-oriented. The main aim of this research paper is to customize the advertisement in such a way that it targets a segment of the total customers' base rather than the whole so as to convert potential buyers into customers. The idea is to use data mining techniques to identify the high-profit, high-value and low-risk customers.

Keywords - Data mining, knowledge discovery, clustering, market segmentation.

I. INTRODUCTION

In today's world of increasing demands and high competitiveness target marketing has become a necessity. It is very essential to make marketing strategies more customer oriented for product development. For successful business, identification of high-profit and low-risk customers, retaining those customers and introducing new customers is a vital task.

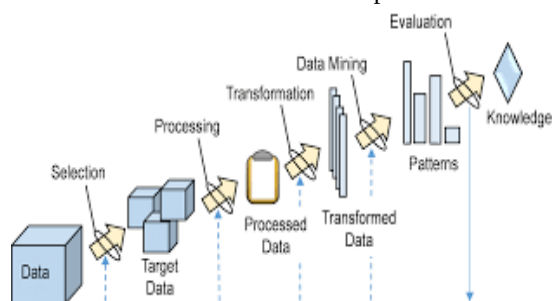
Customer analysis is crucial phase for companies in order to create new campaign for their existing customers. Companies are able to group or cluster certain customers which have similar features. This may assist companies to make better marketing strategies over certain customer groups.

Increasing the leads for the company is an important task for introducing new customers. Lead generation is the way in which company collects contact information from potential customers. Once leads have been generated, marketing and sales work together to convert leads into customers.

Using data mining techniques, we propose an approach through which we can discover patterns and identify the characteristics of customers so as to enhance customer satisfaction and formulate marketing strategies to increase profitability. Since there are many different kinds of customers with different kinds of needs and preferences, performing market segmentation is necessary: divide the total market, choose the best segments and design strategies for profitability, serving the chosen segment better than the company's competitors do.

Data mining technologies and techniques can be used for recognizing and tracking patterns within data to help business sift through layers of seemingly unrelated data for meaningful relationship. In this paper, we discuss a business and technological overview of data mining and outline how can we optimize Customer profitability through data mining application, along with sound business processes and

complement technologies, data mining can reinforce and redefine Customer relationship.



2. Data Mining Framework

Data mining, or KDD (Knowledge Discovery from Data), is the process of analyzing data and extracting useful information and knowledge from large amount of data. Some people say that We are living in the information age but We are actually living in the data age. Terabytes or petabytes of data is generated every day. So, to extract knowledge from this huge amount of data, data mining has become a necessity.

2.1 Domain Understanding

It is essential to understand the business goals of an application before starting any data mining application because blind application of data mining technique without the requisite domain knowledge may result in meaningless and irrelevant patterns. So, it is important for a business to identify target market so that it direct its marketing effort in order to sell their products and services. Target markets or target customers are clusters of consumers with similar or the same needs that most businesses target their marketing efforts in order to sell their products and services.

For instance, the target market for a jewellery store might be middle aged women who care about fashion, and their secondary target market could be middle aged men who may want to buy gifts for the women in their lives. Similarly, children are of course the targets for the chocolate products but if we pay

attention we see that many adults have a sweet tooth as well. This was proven by some research that adults consume over 55% of all the candy sold. Therefore, by demographics chocolate industries market the bite size chocolates to adults.

2.2 Dataset

We need a dataset containing data about your target market to perform mining and extract knowledge. This dataset or collection of data about the customers can be found by variety of ways. Survey research can prove to be a very beneficial method in collecting data. This method can provide information quicker and at a low cost as compared to manual processing. The method will consist of questionnaire from the customers. The questions are flexible and related to the choices that customer make and why they choose those choices. These questions are multiple choice and must consist of all the probable choices. The answers provided by the customers will be stored in the database and will act as a dataset.

Another method for data collection could be interviews. When you know your target audience, you could collect data using phone interviews, since getting mobile phone numbers of people is not a complicated task at all. Asking people about their preference and needs in an interview could provide huge amount of necessary information.

Furthermore, queries fired by customers can also act as an important data source. These are the items searched by customers. This data is usually generated by e-commerce web servers.

2.3 Data Preprocessing

Since the real-world database or dataset is susceptible to noise, inconsistency and missing values due to their huge size and origin from multiple and heterogeneous sources, data preprocessing is essential. Data preprocessing can be done through various ways i.e., data cleaning, data integration, data reduction.

2.3.1 Data Cleaning

Data containing missing values, noise and inconsistency must be cleansed. First step towards data cleaning is identifying data discrepancy. Data discrepancy can be due to faulty equipment, human errors, deliberate errors (not wanting to disclose information) or data decay (outdated data). After the discrepancy has been found, actual cleaning of data starts.

Missing values can be corrected by the following ways:

- Ignore the **tuple** containing missing attribute.
- Fill in the **missing values** with mean, media or mode.
- **Noise** is an error or variance in the measured value of the variable. It can be rectified using following methods:

- **Binning:** It is the method of smoothing sorted values considering its neighboring values. In smoothing by bin means, each value in a bin is replaced by the mean value of bin. In smoothing by bin medians, each value in a bin is replaced by bin median. In smoothing by bin boundaries, minimum and maximum value are identified in the bin and the values inside the bin is replaced by minimum and maximum value with respect to distance closeness.
- **Regression:** Linear regression is the method of finding the best line that fits between two attributes and predicting other attributes. Multiple regression is the method where multiple attributes are involved and data is fit into multidimensional surface.

2.3.2 Data Integration

Since data collection is done from various heterogeneous sources, data integration is necessary for data mining. Data integration is the process of merging from multiple data sources. This helps in avoiding inconsistency and redundancy. Data integration can be done by correlation analysis. Below is the correlation equation.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^n (x_i y_i) - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

where n is the number of tuples, x_i and y_i are the respective values of X and Y in tuple i, $\bar{x}$ and $\bar{y}$ are the respective mean values of X and Y, s_x and s_y are the respective standard deviations of X and Y.

2.3.3 Data Reduction

Data reduction is the process of reduced representation of dataset in much smaller volume and yet closely maintaining the integrity of the data. Data reduction techniques include dimensionality reduction, numerosity reduction and data compression.

- **Dimensionality reduction:** This is the process of reducing the number of variable or attributes.
- **Numerosity reduction:** This technique replaces original volume by smaller form of representation.
- **Data Compression:** This includes the compressed representation of the original data.

2.4 Extracting pattern

After the data has been preprocessed, market segmentation takes place. Since every individual is different from other in various aspects, their needs and preference also vary. For instance, each woman will have different preferences for jewellery. Some might be interested in heavy jewellery while others may be interested in light jewellery. Similarly, some children may like milk chocolate or sweet chocolate

while others may like dark chocolate. Segmentation is necessary to make market more customer centric. We perform market segmentation using clustering technique.

2.4.1 Clustering

Clustering is the process of grouping data items in a cluster in such a way that each item in a cluster has high similarity and dissimilar objects belong to different clusters. Clustering is an unsupervised learning because training set is unavailable. Clustering is a form of learning by observation. Customers are arranged within a group according to similarity. The customers which share similar characteristics are grouped into one cluster while customers which don't share any similarity are grouped into different clusters. So, a jewellery store will group customers such that customers that share predilection for heavy jewellery is clustered into one group while the one having predilection for light jewellery is grouped into another. Clustering can be done using various techniques.

- **Partitioning Method:** Given set of n objects, this method forms k partition where $k \leq n$ and each partition represents a cluster. Each group must have at least one object. Partitioning based methods are k -means (centroid based technique) and k -medoids (object based technique).
- **Hierarchical Method:** This method creates a hierarchical decomposition of given set of data items. It can be classified into agglomerative and divisive approach. The agglomerative approach, also called the bottom-up approach, starts with each object forming a separate group. It successively merges groups close to one another, until all the groups are merged into one, or a termination condition holds. The divisive approach, also called the top-down approach, starts with all the objects in the same cluster and a cluster is split into smaller clusters, until eventually each object is in one cluster, or a termination condition holds.
- **Density - based method:** This method creates clusters of arbitrary shapes. The main idea is to continue growing a given cluster until the density or the number of data items in the neighborhood exceeds the threshold.
- **Grid - based method:** This method quantizes the object space into a finite number of cells that form a grid structure. All the clustering operations are performed on the grid structure. The main advantage of this approach is its fast processing time,

which is typically independent of the number of data objects and dependent only on the number of cells in each dimension in the quantized space.

CONCLUSION AND FUTURE SCOPE

In this paper, we proposed a theoretical approach for data mining framework for target marketing. We provided a theoretical approach to gathering customer data and deriving relationship between the data so as to make marketing strategies effective. Marketing strategies are a result of this meticulous market analysis by analyzing the customer data and understanding the customers effectively and efficiently.

The proposed system will help the analysts and various companies on basis of historical data to predict who will respond to the new marketing campaigns. Through the results the marketers will have an appropriate approach to sell profitable products to targeted customers. Data mining brings a lot of benefits to retail companies in the same way as marketing. Through market based analysis, a store can have an appropriate production arrangement in a way that customers can purchase frequent buying products together. In addition, our proposed model will also help retail companies offer certain discounts for particular products that will attract more customers.

REFERENCES

- [1] Ms. Nazneenarannum S. H. Rizvi "A Systematic Overview on Data Mining: concepts and techniques" International Journal of Research in Computer & Information Technology (IJRCIT) Vol. 1, Special Issue 1, 2016
- [2] Heling Jiang, An Yang, Fengyun Yan and Hong Miao "Research on Pattern Analysis and Data Classification Methodology for Data Mining and Knowledge Discovery" International Journal of Hybrid Information Technology Vol.9, No.3 (2016)
- [3] Daniela Stan Raicu "A Data Mining Framework for Target Marketing"
- [4] Anshul Sinha "Data Mining Techniques for E - Business Service using" IJRCEMAS VOLUME 3, ISSUE 2, February-2016
- [5] Anjali Nair, Rohan Khasgiwala, Snigdha Mishra "Improving Customer Relationship Management Using Data Mining" International Journal of Scientific & Engineering Research, Volume 7, Issue 2, February-2016
- [6] Shreyas Karnick, Dr. Shivakumara Aradhya "Data Mining Applied to Customer Categorization Based on Load Profiling" International Journal of Research in Advent Technology, Vol.4, No.5, May 2016
- [7] Saroja T.V., Greeshma Nair, Kajal Nalawade, Yojana Prajapati "Data Mining Technique to Predict Missing Items and Find Optimal Customer for Beneficial Customer Relationship and Management" International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016

