

# Case Studies On Data Mining In Market Analysis

Shravan Kumar Manthri<sup>1</sup> and Hari Priyanka Chilakalapudi<sup>2</sup>

Department of Computer Science and Engineering

<sup>1</sup>Bhoj Reddy Engineering College for Women,

Hyderabad-500059, Telangana, India

Department of Computer Science and Engineering

<sup>2</sup> Bhoj Reddy Engineering College for Women,

Hyderabad-500059, Telangana, India

**Abstract-** A huge chunk of data is generated each minute in enterprise business. Extracting information from piles of data helps in extracting patterns that can predict and guide future behaviour of the enterprise. Data mining techniques filter through large amounts of raw data and extract useful information that gives enterprise businesses a competitive edge in the market. Various cases on customer purchasing habits have been presented and also used in real problems. Data mining techniques are highly effective in analysing consumer behaviours. It helps enterprises to make informed business decisions, enhances business intelligence, thereby improving the company's revenue, detecting anomalies, fraud detection and reducing cost overheads. In this paper the author reviews various such techniques. Further, the application of these techniques in various scenarios is analyzed.

## I. Introduction

### Data mining

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers and develop more effective marketing strategies as well as increase sales and decrease costs. Data mining depends on effective data collection and warehousing as well as computer processing. To study a customer psychological mindset and converting this into statistical format and see that if there is any technical format by which we can analyze his buying behaviour.

### *BREAKING DOWN 'Data Mining in marketing'*

Grocery stores are well-known users of data mining techniques. Many supermarkets offer free loyalty cards to customers that give them access to reduced prices not available to non-members. The cards make it easy for stores to

track who is buying what, when they are buying it and at what price. The stores can then use this data, after analyzing it, for multiple purposes, such as offering customers coupons targeted to their buying habits and deciding when to put items on sale or when to sell them at full price. Data mining can be a cause for concern when only selected information, which is not representative of the overall sample group, is used to prove a certain hypothesis.

### *Power of hidden information in data:*

Enterprises generate terabytes of data each day that is stored in databases, data warehouses, or various other kinds of data repositories. Most of the valuable information may be hiding behind such data; the overwhelming data volume makes it difficult for human beings to extract them without the help of powerful tools and techniques. At the beginning of the last decade information was only available on papers and only at a specific time. In this age, information is now easily accessible, as content providers, content locators and powerful search engines have enabled access to huge amounts of information in no time. Natural Language Processing has allowed access to the same information in a language familiar to the person. From the time when people used to wait in huge queues to pay bills, taxes or for movie tickets and entertainment, we have reached a time where everything happens in a few clicks. All these changing factors and trends made a huge impact on the enterprises which are highly dependent on the consumer behaviour and basket trends. Every enterprise needs to be highly scalable and be able to predict the future behaviour of the customer to make profits in their business. This very needs lead to the Market Analysis. Earlier techniques like questioners, surveys, test marketing, previous

sales data and leading indicators were used to predict the future of a product. All these techniques although proved to be effective, were highly time consuming and required a lot of manpower. Data mining techniques have completely changed the scenario. Information which was once gathered by travelling miles is now available in few seconds.

## II. Data Mining Tasks

### Classification:

Classification is the process of finding a model that describes the data classes or concepts. The purpose is to be able to use this model to predict the class of objects whose class label is unknown. This derived model is based on the analysis of sets of training data.

Eg: 1.Assigning voters into known buckets by political parties .

2.Bucketing new customers into one of known customer groups.

### Regression:

In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. Eg:Predicting unemployment rate for next year. Estimating insurance premium.

### Anomaly detection:

Anomaly detection (also outlier detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset. Eg:Fraud transaction detection in credit cards.

### Time series

A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data.

Eg:Sales forecasting, production forecasting, virtually any growth phenomenon that needs to be extrapolated.

### Clustering

clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to

those in other groups (clusters).

eg: Finding customer segments in a company based on transaction, web and customer call data.

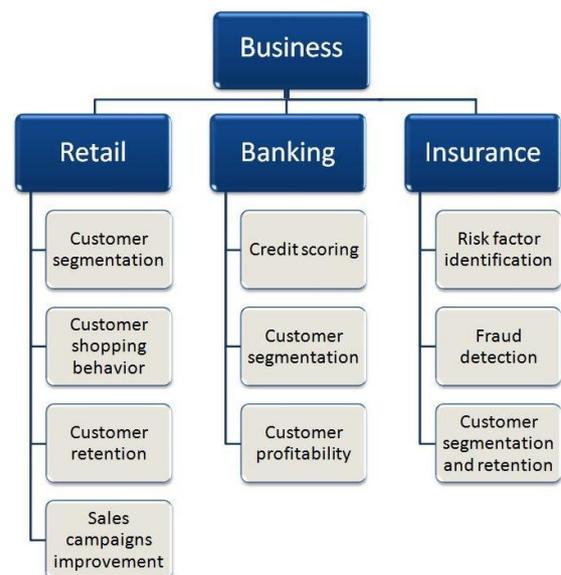
### Association analysis

Association is a data mining function that discovers the probability of the co-occurrence of items in a collection. The relationships between co-occurring items are expressed as association rules.

Eg: Find cross selling opportunities for a retailer based on transaction purchase history.

### Exploring and Expanding Business

Data mining is defined as a business process for exploring large amounts of data to discover meaningful patterns and rules [4]. Companies can apply data mining in order to improve their business and gain advantages over the competitors. The most important business areas that successfully apply data mining, presented in below figure.



Digital marketing is a method of marketing from the viewpoint of customers. Digital information is not only more easily integrated, sorted, and spread, but also enables providers and consumers to interact more quickly. In the past, it usually took a long time for marketing to analyze and achieve effectiveness. Today, digital marketing enables marketing promotion to have a higher synergistic effect.

With the rapid changes in the business environment, technology progress, and digital transmission, the marketing of businesses should be changed rapidly. Similarly, the

strategy of the market should be changed from the Red Sea Strategy to the Blue Ocean Strategy. With the environment changing, not only market space continues to expand, but also the market environment becomes more tightened. From traditional store sales, telephone marketing, and face-to-face marketing, to the development of Internet marketing, such as sale and purchase through the Website, keyword marketing, blog marketing, and so on, and with wireless network development, the ubiquitous Internet will be bringing the world. It makes the development of digital marketing communications seem increasingly important. The business is no longer limited by traditional ways such as time and space, but increases the opportunities of contact and interactions with customers.

#### *Apriori Algorithm*

In 1994, the Apriori algorithm was proposed by Agrawal and Srikant [3]. It is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of commerce website frequentation). As is common in association rule mining, given a set of itemsets (for instance, sets of retail transactions, each listing individual items purchased), the algorithm attempts to find subsets which are common to at least a minimum number of the item sets. Apriori uses a “bottom up” approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

The purpose of the Apriori Algorithm is to find associations between different sets of data. It is sometimes referred to as “Market Basket Analysis”. There are a number of items in each set, and is called a transaction. The output of Apriori is sets of rules that tell us how often items are contained in sets of data.

### III. Predicting the customer behaviour

Predicting the customer behaviour is the most important activity in enterprise business. All the previously described methods give enterprises huge amounts of useful information. The following section presents

some scenarios that have implemented the above methods.

#### *Case study 1 : Application of Association Rule mining in Recommender systems*

Recommender systems are hugely popular these days in variety of fields. To name a few Movies, music, books, research articles, search queries, social tags, etc. These systems help the enterprise by combining the ideas of intelligent systems, machine learning, information retrieval to predict the customer behaviour. There are two approaches in recommender systems, one is collaborative filtering and content-based filtering.

Collaborative filtering methods collect and analyze a large amount of information on users’ behaviors, activities or preferences and predict what users will like based on their similarity to other users. One of the approaches is to use the Apriori algorithm.

In this case study Apriori algorithm is used to extract association rules from user profiles. As an example PVT system is used. PVT system is a recommender program that suggests TV channels to users based on their viewing habits. This system maintains both positively and negatively rated TV channels. Treating user profiles as transactions and the program ratings therein as itemsets, the Apriori algorithm can be used to derive a set of rules and associated confidence levels between programs. The confidence values are taken as similarity scores and used to fill in a program similarity matrix. The procedure goes as follows, the relationship between programs is identified beyond a simple overlap. Like for example a person who watches Reality shows like Rodies and Big Boss may not be interested in shows like KBC and Indian Idol. But if a relation between Rodies and Indian Idol can be established then it can provide a basis for pattern matching. The relation can be identified by finding the support and confidence values. In this case study, the confidence values are taken as the similarity scores and recommended to the user. Using direct program similarity we can derive rules and further by chaining these rules together we can get new results.

#### *Case Study 2: Classification model for Target selection in direct marketing*

Using historical purchase data, a predictive response model with data mining techniques

was developed to predict a probability that a customer in Ebedi Microfinance bank (Nigeria) will respond to a promotion or an offer.[2] To achieve this purpose, a predictive response model using customers' historical purchase data was built with data mining techniques. The data were stored in a data warehouse to serve as management decision support system. The response model was built from customers' historic purchases and demographic dataset. The purchase behaviour variables used in the model development are as follow. Recency: This is the number of months since the last purchase and first purchase. It is typically the most powerful of the three characteristics for predicting response to a subsequent offer. This seems quite logical. It says that if you have recently purchased something from a company, you are more likely to make another purchase than someone who did not recently make a purchase. Frequency: This is the number of purchases. It can be the total of purchases within a specific time frame or include all purchases. This characteristic is second to recency in predictive power for response. Again, it is quite intuitive as to why it relates to future purchases. Monetary value: This is the total amount. Similar to frequency, it can be within a specific time frame or include all purchases. Of the three, this characteristic is the least powerful when it comes to predicting response. But when used in combination, it can add another dimension of understanding. Demographic information includes customers' personal characteristics and information such as age, sex, address, profession etc Bayesian algorithm precisely Naïve Bayes algorithm was employed in constructing the classifier system. Both filter and wrapper feature selection techniques were employed in determining inputs to the model. The results obtained shows that Ebedi Microfinance bank can plan effective marketing of their products and services by obtaining a guiding report on the status of their customers which will go a long way in assisting management in saving significant amount of money that could have been spent on wasteful promotional campaigns.

#### IV. Conclusions

In today's business world, grabbing a customer attention initiates an important role.

Since every business produces many products, standing out in a competitive market is a great issue to be solved for businesses. The advantage of operating businesses in addition to their innovative products, services, brands, quality, etc., how to combine marketing techniques with products, make everyone know about the product, and increase customer's purchase intention and interests is one of the key success factors of operating a business. Overall, the digital marketing will suggest a proper communication method with consumers based on analyzing the marketing information of the product, history records, and purchase behavior of consumers. Among them, the product information includes the type, price, place, and promotion of the product. History records contain the previous marketing strategies, practices and market reactions to estimate, consult, and explore the potential or unknown influence factors of marketing. In terms of consumer behavior, feedbacks can be received and suggesting them for the products of interest. Different characteristics of consumer behavior, the marketing strategy should be ingratiated with the different of consumer's age, sex, occupation, income, lifestyle. In other words, product information, history record, and consumer purchase behavior in terms of products association will affect the business to formulate the marketing strategy for different products.

#### REFERENCES

- [1] Barry Smyth, Kevin McCarthy, James Reilly, Derry O'Sullivan and Lorraine McGinty, "Case-Studies in Association Rule Mining for Recommender Systems". Science Foundation Ireland (SFI) under Grant No. 03/IN.3/I361
- [2] Eniafe Festus Ayetiran; "A Data Mining-Based Response Model for Target Selection in Direct Marketing", I.J. Information Technology and Computer Science, 2012, 1, 9-18
- [3] Shu-Ching Wang, Shun-Sheng Wang, Chih-Ming Chang, "Systematic Approach for Digital Marketing Strategy through Data Mining Technology". Department of Information Management, Chaoyang University of

Technology Taichung 409, Taiwan,2014,4-1.

Economic Studies, Bucharest, Romania.  
Database Systems Journal vol.IV, no.4/2013

[4] Ruxandra PETRE,” Data Mining Solutions  
for the Business Environment”. University of