SALES ANALYSIS OF E-COMMERCE USING DATA MINING TECHNIQUES

Dr.K.Chitra, MCA.,M.phil.,Phd
Assistant Professor
Department of Computer Science
Mangayarkarasi College of arts and science for Women Madurai Dt
Paravai, Madurai

G.Sivabharathi M.Sc.,M.phil.,B.Ed.,(Phd)
Assistant Professor
Department of Computer Science
Government Arts College, Melur

ABSTRACT
In the emerging global economy, E-commerce is a strong method for financial development. The fast growth in usage of Internet and Web-based applications is decreasing operational costs of large enterprises, extending trading opportunities and lowering the financial barriers for active ecommerce participation. Many companies are reorganization their business strategies to attain maximum value in terms of profits as well as customer’s satisfaction. Data mining (DM) is used to attain knowledge from available information in order to help companies make weighted decisions. An organization needs to invest only on the group of products which are frequently purchased by its customers as well as price them appropriately in order to attain maximum customer satisfaction. The objective of this paper is to evaluate, propose and improve traditional pricing strategies by using web mining techniques to collect information from e-commerce websites and apply data mining methods to induce and extract useful information out of it.

Keywords: E-Commerce, Data Mining, ID3 Algorithm

1. INTRODUCTION
The Web is one of the most revolutionary technologies that changed the business environment and has a dramatic impact on the future of electronic commerce (EC). The future of EC will accelerate the shift of the power toward the consumer, which will lead to fundamental changes in the way companies relate to their customers and compete with one another. Previous studies in Information Science (IS) literature like The Consumer Behavior towards online shopping of electronics in Pakistan (Adil Bashir 2013), Online Consumer Behaviour (Dr. Bas Donkers 2013), Influencing the online consumer’s behavior: the Web experience (Efthymios Constantinides 2010), ) Post-purchase behavior (Dibb et al., 2004; Jobber, 2010; Boyd et al., 2012; Kotler, 2011; Brassington and Pettitt, 2013) have proposed various models explaining customer buying behavior. These research models typically derive hypotheses from a literature review. Based on this hypotheses, evaluation of a
multi-channel customer choice data can be done. Commerce networks involve buying and selling activities among individuals or organizations. [1] Getting a deeper understanding of e-commerce networks, such as the Flipkart market space, in terms of structure, interactions, trust and reputation has tremendous value in developing business strategies and building effective user applications. Nowadays, web data provides comparative advantages for mass merchants to analyze and reveal important parts of online consuming behavior [2]. This paper discusses examples of multi-channel strategies and designs a pricing model which focus on 4 P’s of Marketing mix. Based on the analysis of the retailer’s transaction data and a literature review, we derive hypotheses to explain consumer purchasing behavior.

2. BACKGROUND

The E-Commerce industry represents one of the largest industries worldwide. For example, in the United States, it is the second largest industry in terms of both the number of establishments and profits, with $3.8 trillion in sales annually. [3] In addition, this industry is facing similar trends to those affecting other sectors, for instance, the globalization of markets, aggressive competition, increasing cost pressures and the rise of customized demand with high product variants.

Manual capture of sales information increases transaction costs and can cause inventory inaccuracies.

This kind of processing involves numerous human interventions at different levels such as order taking, data entry, processing of the order, invoicing and forwarding. The accuracy of the model is questionable and may not be consider few important factors while developing it. To overcome this problem, data mining can be used to analyze big data and develop efficient marketing strategies. It is ideal because many of the ingredients required for successful data mining are easily satisfied: data records are plentiful, electronic collection provides reliable data, insight can easily be turned into action, and return on investment can be measured [4].

3. DATA MINING AND CONSUMER BEHAVIOR IN E-COMMERCE

In the past few years, the development of the World Wide Web exceeded all expectations. Retrieving data has become a very difficult task taking into consideration the impressive variety of the Web. Web consists of several types of data such as text data, images, audio or video, structured records such as lists or tables and hyperlinks. Web content mining can be used to mine text, graphs and pictures from a Web page and apply data mining algorithms to generate patterns used for knowledge discovery [5]. For a successful e-commerce site, reducing user-perceived latency is the second most important quality after good site-navigation quality.
The most successful approach towards reducing user-perceived latency has been the extraction of path traversal patterns from past users buying history to predict future user buying behavior and to fetch the required resources. [6] Vallamkondu & Gruenwald (2003) describe an approach to predict user behavior in e-commerce sites. The core of their approach involves extracting knowledge from integrated data of purchase and path traversal patterns of past users to develop a pricing model which focuses on profits as well as customer satisfaction. [7] Web sites are often used to establish a company’s image, to promote and sell goods and to provide customer support. The success of a web site directly affects the success of the company in an electronic market.

3.1 Product Strategy

A product is anything that can be offered to a market for attention, acquisition, use, or consumption that might satisfy a want or need (Kotler, 2001). In an e-commerce marketing strategy, it is important to remember that information is now its own viable product. In the physical world, a shopper who wants to buy something has to manually sift through the millions of choices. A complete search of all offerings would be extremely expensive, time-consuming and practically impossible. Instead consumers rely on product suppliers and retailers to aid them in the search. This allows the suppliers and providers to use the consumers’ cost-of search as a competitive advantage. However, on the Internet, consumers can search much more comprehensively and at virtually no cost[11].

<table>
<thead>
<tr>
<th>Table 1. Rating v/s Type Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
</tr>
<tr>
<td><strong>Plot Chart</strong></td>
</tr>
<tr>
<td><strong>Rating</strong></td>
</tr>
<tr>
<td>1 star</td>
</tr>
<tr>
<td>2 star</td>
</tr>
<tr>
<td>3 star</td>
</tr>
<tr>
<td>4 star</td>
</tr>
<tr>
<td>5 star</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

By using the direct access to consumers enabled by the Internet, companies can collect information, identify target consumers, and better introduce products or services to meet consumers’ needs. If a customer finds all the desired product type it will directly affect the customer satisfaction index (refer to table 1.1). After analyzing the training set with 1185 instances, It was found that the 60% customers were satisfied with phone and TV had least customer satisfaction terms of online product rating.

3.2 Price Strategy

In the earliest days of Internet commerce, many economists and media observers predicted that competition among Internet retailers would quickly resemble perfect competition. After all, the Internet already reduces search costs relative to visiting physical stores and comparison sites could be expected to lower search costs still further. The question of how pricing impacts consumer purchasing behavior is interesting. In this paper, we
discuss one such application, measuring the potential magnitude of bias in the consumer price index arising from underweighting Internet commerce. Price is the only element of the marketing mix to generate revenues. Internet pricing decisions will be just as crucial as they traditionally have been.

![Table 2. Correlation between price and online rating](image)

The Internet will lead to increased price competition and the standardization of prices. Also, the ability to compare prices across all suppliers using the Internet and online shopping services will lead to increased price competition. Finally, the price of providing Internet-based services often contains little or no marginal costs. Organizations will have to employ new pricing models when selling over the Internet.

3.3 Productivity Concept

Business profits can be increased by increasing revenue through stronger sales and/or by decreasing the costs associated with constant sales. One of the major factors in customer satisfaction is the availability and timeliness of the delivery of products. If a customer has to wait to receive a product, it can be detrimental to their feeling of satisfaction. [13] With that in mind, avoiding back orders should be a major goal of any business.
Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861
International Conference on Emerging Trends in IOT & Machine Learning, 2018

Fig 1.1. Application of ID3 algorithm on data extracted from e-commerce website

Fig 1.2. Partial Decision Tree after applying ID3 Algorithm
4. MODEL EVALUATION

The confusion matrix is a useful tool for analyzing how well your classifier can recognize tuples of different classes. TP and TN tell us when the classifier is getting things right, while FP and FN tell us when the classifier is getting things wrong mislabeling). Given m classes (where m≥2), a confusion matrix is a table of at least size m by m. An entry, CM_{i,j} in the first m rows and m columns indicates the number of tuples of class i that were labelled by the classifier as class j. For a classifier to have good accuracy, ideally most of the tuples would be represented along the diagonal of the confusion matrix, from entry CM_{1,1} to entry CM_{m,m}, with the rest of the entries being zero or close to zero. In this case we m=4 so we have a 4x4 matrix. After applying ID3 algorithm this model has 86.4780% accuracy (i.e: out of every 100 test cases it has correctly predicted 87 test cases [16].

<table>
<thead>
<tr>
<th>Confusion Matrix</th>
<th>PREDICTED CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>ACTUAL CLASS</td>
<td></td>
</tr>
<tr>
<td>A = 2 star</td>
<td>141</td>
</tr>
<tr>
<td>B = 3 star</td>
<td>14</td>
</tr>
<tr>
<td>C = 4 star</td>
<td>21</td>
</tr>
<tr>
<td>D = 5 star</td>
<td>7</td>
</tr>
</tbody>
</table>

5. CONCLUSION

In this paper, a detailed study based on data mining techniques was conducted in order to extract knowledge in a data set with information about user’s history associated to an e-commerce website. These datasets are directly mined from Flipkart.com using an online software which converts html documents to data tables. The main purpose to web mine data is to apply a set of descriptive data mining techniques to induce rules that allow data analyst working at e-commerce companies make strategic decisions to boost their sales as well as provide effective customer service. Techniques used to discover patterns are web mining and decision tree algorithms. In the future, this study can be used to analyze ecommerce websites and obtain interesting knowledge to further the companies’ profits.

Many of the e-commerce strategy frameworks offer a unique contribution to strategic planning but with limited solution.
This model based on web mining integrates the McCarthy’s 4Ps to provide a complete analysis of e-business strategies. Thus managers can use an organized and precise process to make more successful and effective decisions. An aggressive competition has been observed in market space among the companies, thus accelerating the consumer dynamics. E-commerce will lead to increased price competition and this web application will provide an efficient way to price a particular product. It was found that price, product and production had an impact on online customer ratings. This model considers these three attributes which are correlated to customer satisfaction and help the marketer make an informed decision.

8. REFERENCES


