



ISSN 2278 – 0211 (Online)

## Transforming E-Commerce through Data Mining

**Aparna Bulusu**

Lecturer, Department of Computer Science  
St. Ann's Degree College for Women, Mehdiptnam, Hyderabad, India

### **Abstract:**

*Data mining has drastically altered how businesses are being operated, especially in the online domain. This paper attempts to review the massive growth of data mining applications to improve the profitability of ecommerce ventures. As consumers increasingly tend to make their purchases online and with many companies investing heavily on improving their web presence, it is only natural that the whole process of collecting and mining web data has become instrumental in determining a company's bottom line. What started off as a simple analysis of user behavior and online transactional data has now transformed into a massive, multidisciplinary knowledge discovery process involving huge data sets, algorithms, statistical techniques and analytics to glean new information. The article focuses on some data mining applications unique to ecommerce ventures and how they are being utilized in the current scenario.*

**Keywords:** Data Mining, e-Commerce, Segmentation, Personalization

### **1. Introduction**

E-Commerce domain can provide all the right ingredients for successful data mining. It is a killer domain for data mining[1] Heavy usage of eCommerce websites by consumers has led to the generation of huge amounts of data like click streams, web logs, routing and server requests, web links and cookie information etc. Innovative analysis of all this information is helping businesses in customer segmentation, profiling, targeted advertising etc. Utilizing this kind of information about customers to boost sales is often classified as Customer Relationship management (CRM) and helps in boosting sales and improving customer loyalty [2]

E-commerce is an important domain for data mining [3], with massive amounts of click stream and transactional data that dwarf in size data warehouses from a few years ago [4]

One of the main challenges with datamining however is the inability to convert huge data masses into useful information because of problems with data collection, irrelevant data or difficulties in cleaning and transforming data. Some sources of data for E-Commerce ventures include: Information from customer signup/registration forms, demographics, web click streams, page views and session information, catalogs, email and direct mail campaigns, operational/transactional data from point of sales (POS) outlets, call records and transcripts from call centers , public opinions and reviews from social networking mediums etc. Gathering the right kind of data from business events is especially necessary for effectively mining E-Commerce data. Some of these include the kind of searches being made for products and the results being generated, various shopping cart activities like placing orders /cancelling products or abandoning carts, information generated through payment gateways etc.

### **2. Data Mining – Methodology and Techniques**

Data mining lifecycle stages are in the following natural order: Requirements gathering, data collection, data warehouse construction, business intelligence, and deployment [5] The entire data mining process can be briefly summed up in simple layman terms as follows: Large amounts of data – transactional, operational and other non traditional forms of data is all collected and stored in large databases typically referred to as data warehouses. This data needs to be preprocessed, cleaned and transformed before it can be mined. Mining involves discovery of novel, non trivial and useful patterns. The patterns so discovered give rise to domain specific knowledge which users can act upon.

An overview of commonly used data mining techniques includes Association analysis, clustering, classification, regression analysis, summarization and anomaly detection.

Association analysis involves making simple co-relations between two or more variables to identify patterns. A commonly cited example being the discovery of an association between purchase of beer and diapers obtained through market basket analysis.

Classification involves breaking down of available data into a set of classes based on some classification function. Predictive diagnostics make use of classification based on DNA patterns.

Regression analysis involves mapping a predictor to a predicting variable and determining whether strong co-relations exist between variables. Regression analysis is widely used to determine customer spending activity based on credit card and demographic information.

Clustering is another important technique where the entire data sets are represented as a set of finite clusters or categories. Clustering and its variants like nearest neighbor technique offer a high level view of the entire data set and help in determining the strategies for each separate cluster.

Summarization is another important and fairly simple technique which involves representing large data sets within a few variables through use of statistical models.

Anomaly or change detection tries to detect significant changes between temporal data sets to generate predictions with least probability of error. A related technique is outlier detection through clustering.

Data mining is heavily dependent on techniques from various allied fields like statistics, machine learning, neural networks, artificial intelligence etc. Specific techniques used often include decision trees, nearest neighbor classification, neural networks, rule induction, k-means clustering, genetic algorithms, Bayesian networks etc

### **3. Suitability of E-Commerce for Mining**

Ecommerce is exceptionally well suited to perform data mining since it satisfies most of the desiderata required for data mining. Ecommerce yields massive data records

Collected automatically and are hence very reliable. Also it is easy to convert any insight yielded through mining process into action, and finally return on investment can be measured. Ecommerce systems by their very nature and design allow a variety of questions to be answered

eg:

- Products with highest and lowest sales.
- Most viewed web pages.
- Frequently used search terms
- Most canceled shopping cart items
- Brands with highest turnover.
- User choice of web browsers.
- Products with highest profit margins
- Events currently trending.

### **4. Applications Specific to E-Commerce**

#### *4.1. Sentiment Analysis*

Sentiment analysis is a popular research topic in NLP, text mining, and Web mining in recent years [6]. Sentiment analysis or opinion mining has been defined as the computational study of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, etc., expressed in the form of text[7]. A rich collection of sources is now available for performing sentiment analysis: like personal experiences posted on twitter, face book, blogs, forums, online groups, comments, internal data of an organization like customer feedback emails, articles and news on opinions and commentaries etc.

Applications of sentiment analysis: Very helpful in determining the reason behind a products success or failure. Helps in understanding user opinions on a wide range of features like pricing specifications, especially for books, movies, art, music etc.

#### *4.2. Web Usage Mining*

Web usage mining is the automatic discovery of patterns in click streams and associated data collected or generated as a result of user interactions with one or more Web sites. Web mining is used to analyze the behavioral patterns and profiles of users interacting with a Web site [8]. Web usage mining has emerged as the essential tool for realizing more personalized user-friendly and business-optimal Web services. Some practical uses of web mining include keeping track of actual number of buyers as opposed to just browsers, detailed product views, the product click through i.e. the order of the exact set of pages being clicked on by users etc. Web mining is being extensively used to broaden and deepen customer relationships, build customer loyalty, proactively market products to consumers through up-sell/ cross-sell.

#### *4.3. Recommender Systems*

Recommender systems are being used by an ever-increasing number of E-commerce sites to help consumers find products to purchase. Recommender systems use product knowledge—either hand-coded knowledge provided by experts or “mined” knowledge learned from the behavior of consumers—to guide consumers through the often-overwhelming task of locating products they will like [9].

Consumers face an information overload in terms of number of products being offered on ecommerce websites that choosing the right product has become quite cumbersome. Recommender systems offer mass customization of ecommerce sites to users by using information collected from various sources like best selling product lists, consumer demographics, opinion and sentiment analysis collected about those products etc. Recommender systems enhance ecommerce sales in multiple ways: They help website visitors in finding products they wish to purchase and convert browsers into buyers. They suggest additional products and increase cross selling of products and ultimately help in building consumer loyalty. Customizing the whole purchasing experience leads to better returns in terms of products sold as well as improves customer base and customer loyalty. Customizing the whole purchasing experience leads to better returns in terms of products sold as well as improves consumer base and consumer loyalty.

#### *4.4. Consumer Profiling and Personalization*

An important and rich source of information about customer is his/ her profile, which is used to make several important marketing decisions. A customer profile is a model of the customer, based on which the marketer decides on the right strategies and tactics to meet the needs of that customer. Ecommerce sites use information like frequency of purchases, size of purchases, recency of purchases to identify customer groups, determine lifetime of customers, and determine prospective customers through past behavior and measure success or failure of marketing programs [10]

Consumer profiling is being done in multiple ways: Aggregating geo demographic information, transaction based profiling based on a consumers purchase history, explicit profiling through sign up and registration forms, contextual profiling based on the kind of content being consumed by a user, behavioral profiling that uses past tracked activity and look alike profiling that helps in identifying new customers through modelling the behavior of actual customers [11].

Personalization is defined as the ability to provide content and services tailored to individuals on the basis of knowledge about their preferences and behavior [12] Personalization begins with collecting customer data from various sources. This data might include histories of customers' Web purchasing and browsing activities as well as demographic and psychographic information.

Customer profiling and personalization are being utilized by E-commerce companies in multiple ways. Companies email messages to inactive consumers to solicit business, send notifications regarding available products which are usually displayed on request, and also try to cross sell additional products based on purchases already made.

Some other common applications of data mining include market basket analysis that utilizes data mining techniques to come up with association rules to identify cross selling opportunities as well as figure out product affinities

Fraud detection or Compliance Monitoring for Anomaly Detection includes a primary monitoring system that keeps checking against a set of predetermined conditions against the actual data .Any variation and anomaly detected is reported as an alert or exception. Other uses include sales and merchandise forecasting, personalized promotional campaigns and creating customized products.

### **5. Practical Applications**

A variety of online sources [13,14,15] contain glowing reports of ecommerce ventures successfully implementing various Data Mining techniques to increase their revenues, expand their customer bases and decrease promotional and marketing costs. A few commonly cited examples of data mining success stories are compiled below:

- Amazon makes use of its proprietary recommender engine based on predictive modelling techniques to suggest products using the "you may like" feature. It also makes extensive use of customer demographics, purchase history and preferences to increase average purchase size by generating "you may also like" prompts.
- Customer segmentation and clustering is being used heavily by telecom companies to cluster consumers based on their call patterns. This information is utilized in suggesting different call plans with a variety of features to attract and retain consumers.
- Tesco uses its loyalty program to generate tremendous amounts of customer data that is mined to inform customers regarding promotional events.
- High end retailer Neiman Marcus uses behavioural segmentation to target affluent consumers thereby increasing purchases of big margin products..
- Hannah's a group of US hotels and casinos uses detailed consumer profiles to target marketing campaigns and increase customer loyalty.
- Sites like eBay mine their click stream data to determine optimal layout and make site navigation easy.
- Most e-retailers are creating different versions of their catalogues and promotional campaigns according to the behaviour and preferences of different groups of customers and send out targeted emails.
- Over 94% of media companies believe that personalization is critical to current and future success. Most media websites are offering services like alerts, suggestions and customized apps to improve user experience and capture a bigger consumer base.
- Top digital companies like Face book, Twitter etc have high valuation primarily due to their huge data assets which can be used to extract useful information.

## 6. Conclusion

An analysis of sales across domains and industries has revealed that more than 50% of retail sales will be online or will be influenced by online channels by 2014. All kinds of businesses and ecommerce ventures in particular are making extensive use of data mining applications in a variety of ways to achieve a multitude of benefits. Effective use of patterns and information gained from data mining processes is helping businesses in increasing revenues generated through cross sell and upsell offers, increased customer satisfaction scores, improvement in customer loyalty and retention, reduced marketing campaign costs, servicing customers proactively through social listening and using customer profiles to personalize advertising campaigns. Data mining has led to effective use of existing resources by increasing advertising revenues, improved conversion ratios and improvement in customer experiences. It has also led to increased efficiencies as companies are able to target multiple consumers with reduced advertising costs.

In order to become market leaders and maximize their Return on Investments (ROI) ecommerce ventures should aggressively use data mining to generate web intelligence and that will provide them with a better insight into their customers. Data mining will help companies in measuring all important metrics and thereby plan their strategies for better performance.

## 7. References

1. Suhail Ansari, Ron Kohavi, Llew Mason, and Zijian Zheng, Integrating E-Commerce and Data Mining: Architecture and Challenges, IEEE Proceedings International Conference on Data Mining, 2001
2. McGarry, Ken, Andrew Martin, and Dale Addison. "Data Mining and User Profiling for an E-Commerce System." *Classification and Clustering for Knowledge Discovery*. Springer Berlin Heidelberg, 2005. 175-189.
3. Kohavi, R. & Provost, F. (2001). Applications of data mining to electronic commerce. In *Data Mining and Knowledge Discovery*, 5(1/2). Kluwer Academic. <http://robotics.Stanford.EDU/users/ronnyk/ecommerce-dm>
4. U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledgediscovery: an overview. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthursamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 1–34. AAAI-Press, 1996
5. Ron Kohavi, Llew Mason ,Rajesh Parekh ,Zijian Zheng, Lessons and Challenges from Mining Retail E-Commerce Data, *Machine Learning Journal, Special Issue on Data Mining Lessons Learned*, 2004
6. Kimball, R. & Merz, R. (2000). *The Data Webhouse Toolkit: Building the Web-Enabled Data Warehouse*. John Wiley & Sons
7. Bing Liu. *Sentiment Analysis and Opinion Mining* Morgan & Claypool Publishers, May 2012.
8. Bing Liu, *web usage mining*
9. J. Ben Schafer, Joseph A. Konstan , John Riedl, *E-Commerce Recommendation Applications*, *Data Mining and Knowledge Discovery*, 5, 115–153, 2001
10. Michael J. Shaw a,b,c, Chandrasekar Subramaniam a, Gek Woo Tan a, Michael E. Welge b, *Knowledge management and data mining for marketing*, *Decision Support Systems* 31 Ž2001. 127–137
11. URL:<http://www.dtcperspectives.com/wpcontent/uploads/2012/10/Neuner-ConsumerProfiling.pdf> , Chris Neuner, *Future of customer profiling is Here*.
12. P. Hagen, "Smart Personalization," *The Forrester Report*, Forrester Research, Cambridge, Mass., July 1999.
13. *Econsultancy Report*, Digital Media Exchange, NGData Webinar, 2014
14. Roopam Upadhyay, *Customer Segmentation and Cluster Analysis*, Telecom Case Study, URL: [ucanalytics.com/blogs/tag/roopam-upadhyay](http://ucanalytics.com/blogs/tag/roopam-upadhyay)
15. Mckinsey Global Institute, MGI Research report on big data