

# Data Mining in E-Commerce: A CRM Platform

Lipsa Sadath  
Department of Computing  
Muscat College,  
Muscat, Oman

## ABSTRACT

Data is considered as a basic form of information that needs collection, management, mining and interpretation to create knowledge. Modern e-commerce is also vigorously developing that makes resources and services on the internet richly colorful. At the same time there are lots of fraudulent situations happening with people coming closer to the e-commerce system. This is an era where e-commerce is considered to be a killer-domain for successful mining data as it gives the apt ingredients from situation to situation. One of oldest things that e-commerce can do is customer relationship management (CRM). Businesses targeting customers has a direct link with the economy of a country as the current e-commerce system is used by people from lay man to business tycoons. The paper aims at a study on e-commerce with data mining proposing architectural model integrating an improved CRM system for handling business predictions and make strategies to enhance best customer relationship management.

## Keywords

Data Mining, e-Commerce, e-Business, CRM, Issues, Architecture, Business Strategies

## 1. INTRODUCTION

According to Ginni Rometty, CEO, IBM, New York, 'Big Data' services will let customers mine vast troves of information to make better decisions that will give firms and governments a competitive advantage [1]. Until early 90's electronic commerce was not web based. The term simply referred to the use of computers and telecommunication to automatically forward and process commercial documents such as invoices and inventory requests. Mountains of data are quickly and thoroughly explored using data mining techniques. Data mining aims at extraction of implicit, previously unknown and useful patterns and models from data sets. Data mining may be viewed as the extraction of patterns and models from observed data [2]. These technologies have aided science, business and engineering applications. In business, effective predictions can be made from prehistorically available data. Retail industry, financial, communication and marketing organizations which have strong consumer focus used data mining primarily for business predictions. These predictions enable companies determine relationships between internal and external factors. Internal factors may include economic indicators, competition and customer demographics. Data mining help users direct the future of their activities by delivering accurate and useful data because decisions here are made on sound database intelligence and not on instincts on emotions. Many researches are being done on CRM maintenance in relation with e-commerce and models developed. Since customers' play an important role in ecommerce; an effective fraudulent free, strategic CRM is needed for successful ecommerce. An attempt for the same is made here by proposing a new model

for tight integration of data warehouse with analysis components.

Section- I is an introduction about the paper, Section- II gives definition and concepts of DM and e-commerce systems , Section -III briefs on the applications of data mining in e-business, Section -IV is an architecture of a data mining and e-commerce system with XML explanations, Section-V is a detailed analysis on CRM including issues, Section –VI is explained with technology, a basic customer behavior model and then a powerful proposed model integrated with DM and e-commerce in a CRM platform, Section- VII talks about business strategies and an overview diagram on knowledge attained from raw business data and Section-VIII is the conclusion of the paper.

## 2. DATA MINING AND E-COMMERCE / E-BUSINESS – DEFINITIONS AND CONCEPT

Data mining tools aid the discovery of patterns in data [3], [4], and [5]. Gartner, the global leader in technology research and IT services define mining as the process of discovering meaningful correlations, patterns and trends by sifting through large amount of data stored in depositories [6] and ecommerce provides a perfect work bench for data mining [7] , [8] .

A wide verity of multimedia data such as images, videos, signals, and text that are available in electronic form with temporal and spatial characteristics shows variety of data types and structures. These are basic resources in the present generation data applications. Many people into business and research uses internet as the basic infrastructure in data mining to revolutionize business and scientific landscape. The web has declared itself as a powerful global connecting force. Internet established world merge to one community and technology created intranets and extranets creating communities within companies [9]. Portal concepts integrated massive amount of information and their functionality that brought flexibility and structure as a package tailoring the employee needs. In this internet economy, a start-up company can readily overtake an established business. This enterprising environment helps develop and deploy innovative new strategies and find new internet matrices that help succeed in e-Business. Since data continues to grow, it is crucial to ensure systems scalability and interactivity of data mining implementation in high performance internet based computing environments. Thus developing methodologies that scales to thousands of attributes and millions of transactions is a big challenge. A verity of techniques is used for the same. Various pre-processing and post processing operations like sampling, data reduction and transformation, feature selection, rule grouping, exploratory and interactive browsing, meta-level mining etc. have also enriched to support data mining applications.

### 3. APPLICATIONS

Today many companies are relying on internet based matrices such as site statistics, user demographics and audience measurement data [9]. These extensions open new possibilities and quality in data mining applications that include bridging business and engineering fields on one hand and business and social sciences on the other hand. A statistics on the analysis of data mining applications done in 2003 had revealed a 5 % of usage of the application in ecommerce / web industry. There is no doubt to what level this has increased after ten years. The reason why ecommerce can be a successful domain for data mining is because there are plenty of electronic, reliable data record collections where insight can be turned into actions easily and an investment return can be measured. An advantage from the system occurs when the data transformation bridges are used appropriately between transaction processing and data warehouse and vice versa. An integrated solution can also provide users with a uniform user interface and seamless access to metadata [10], [11]. Simple analytical tools utilized different hypothesis for data mining but knowing the best method to be applied according to the customers is very crucial to maintain the relationship.

### 4. ARCHITECTURE OF DATA MINING IN INTERNET PLATFORM

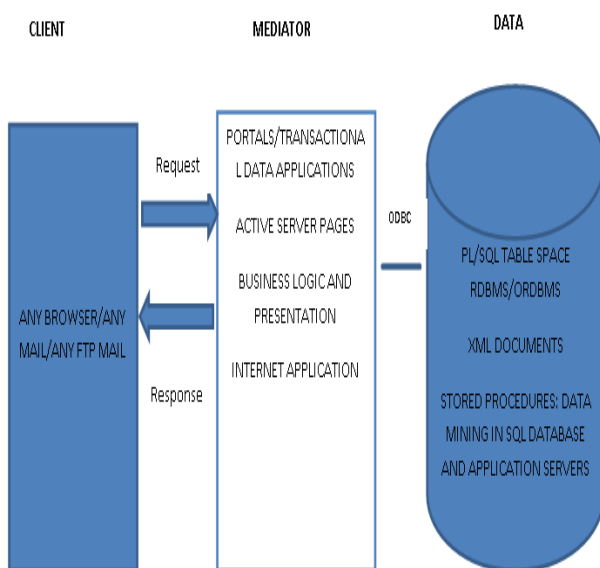


Fig 1: 3-tier architecture integrating data mining in internet platform

The architecture (Fig. 1.) tries to explain a multi-tier application which is divided into several modular tiers where each tier is located at different physical location [20]. The data tier contains data base information, which uses SQL server and stored procedures. The advantage of an object relational data base management system is that user defined data types and objects can be used, they are fully compatible relational database, they support multimedia and large objects and have high quality database features. The mediator or the middle tier takes care of the implementation of business and presentation logic to control clients (users) and data transactions. They use XML documents (extensible markup language) and Active Server pages. Logic of the mediator checks the authenticity of any client request first from active

directory of the respective domain to check the use ids and passwords. This retrieves the required data from data base and responds to client request in suitable representation. A user always interacts with the user interface in the client tier. It could be HTML or WML documents.

#### 4.1 XML-issues and challenges

Studies over the years have proved XML as flexible hierarchical model suitable to represent huge amounts of data with no absolute or fixed schema, a possible irregular and incomplete structure [12]. Unlike HTML, XML was created to structure, transport and store information. It is the most common tool used for data transformation between all sorts of applications. Hence data mining performed on XML queries can be very effective. There are two types of approaches to XML document access; key-word base search and query-answering. A key-word base search has no much advantage as search happens on textual contents of the document. As query-answering is concerned for the query formulation to be effective, users need to know the structure in advance which is often not the case. Dramatic outcomes of such situations are either problems like information overload where too much data are included because of the key-word used or information deprivation where incorrect answers are received due to inappropriate key-words. Though researches have had tough time implementing the best methods for data mining in XML either by using languages such as XQuery [13] and techniques developed in XML context or by implementing graph or tree based algorithms, recent studies have reached methods like Tree-Based Association Rules (TARs) as means to represent intentional knowledge in native XML.

### 5. CUSTOMER RELATIONSHIP MANAGEMENT (CRM)

Customer management is one of the really old things that ecommerce has done. Things have changed from the time small communities and shop keepers kept customer satisfaction to the e-commerce / e-business trends, which means relationship became impersonal. Apart from billing, contacts and retentions provided, data mining can be used to anticipate the kind of relationship specific to people.

Though clustering technique of data mining is ever considered a classic technique to segregate consumers, make predictions and future plans for example, clustering customers according to age would help to identify senior citizens who need special support and reminders, these data mining methods should be revised periodically to reflect important changes. According to situation and information available strong predictions may not be always easy with clusters. Basically three CRM tasks are to be considered.

#### 5.1 Acquiring a new customer

This is often done by statistics from mining information from address list, cookies and data warehouses. This is actually more expensive than retaining a customer.

#### 5.2 Please the customer

A process done by depending on the kind of customer and the sector to which he belongs to. This is also done by giving a 24 hour help desk support, website services and sharing small happiness like remembering birthdays of customers and sending e-personalization birthday cards etc.

### 5.3 Retain the customer

When competitors offer constant services and prices, it becomes very challenging to retain a customer. In most business data mining play a very important role in retaining customers mostly by good maintenance of customer detail files for excellent services.

### 5.4 Issues in CRM

In any CRM system for the above mentioned tasks to work well the customer's trust becomes the key. This is one main reason why businesses spend a lot on gaining customer trust. Main issues related to CRM trust are perception, professionalism and security.

Perception discusses situation where customers go behind standard, branded, large, rich and of course non scandal ridden companies. An established company logo, glitch free software and a convenient website assures the company's professionalism on e-Business. While talking about security, this was considered a major hurdle to early e-commerce when people didn't have confidence in credit card purchases and details about encryption certificate. Situations have changed now with people making more ecommerce transactions and does third party collection like pay pal. Intrusion detection methods using data mining is also undergoing much research which is essential to gain public trust to escape hacking and data misuse.

## 6. CUSTOMER BEHAVIOR WITH TECHNOLOGY

CRM systems most importantly use data mining to predict customer behavior [14]. Mostly this consists of capturing relevant customer information, behavior modeling, and value assessment.

### 6.1 Capturing the relevant customer information

Earlier customer information used to be distributed across the company through different departments. Then there was a need to harmonize them. This was essential from both technical and business point of view. A customer may interact through the web or through call centers. Thus all the data regarding the customer should stay updated for the employees to interact smoothly and capture new information. Thus the CRM solution integrates information from multiple sources to create consolidated customer view and then make this customer knowledge base available as source data for the numerous CRM analytical applications [14]. Thus every customer interaction can create new insights on customer behavior.

### 6.2 Customer Behavior Model (CBM)

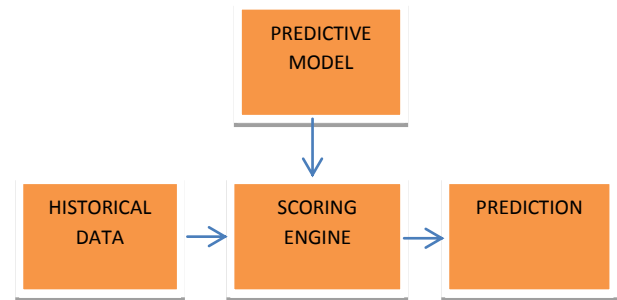


Fig 2: Customer Behavior Model [14]

The model (Fig. 2.) describes the system where the customer historical data are taken and a suitable predictive model is applied at the scoring engine. There are a number of business predictive models like statistics which is an old discipline and then using regression analysis, clustering- method by which like records are grouped together, nearest neighbor techniques, neural networks, rule induction etc.

#### 6.2.1 Clustering

Though a number of techniques are being established as new generation techniques, a base technique that could be used in business strategies in e-commerce environment could be clustering. The technique segments records on a database into subsets based on a set of attributes. Clustering is an automated grouping of related records together. Records having similar values for the attributes are grouped together [15].

#### 6.2.2 Scoring, Scoring engine and scoring process

The process of using any predictive model that is created by data mining on historical data for predictions is called scoring and the prediction obtained is called a score. Scoring engine is a software application to evaluate mathematical functions on a set of data inputs. In the scoring process a marketing user may identify customers with similar interest or attributes and group them from the database to separate views or table space. These selected groups are then scored by predictive models and sorted according to their scored values. The top 'X' percent are chosen for promotion.

### 6.3 Proposed Model for Value Assessment

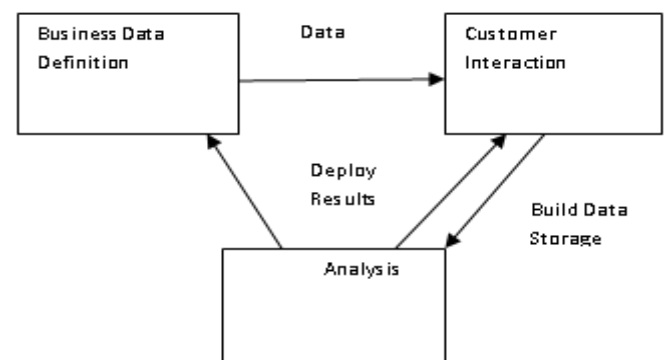
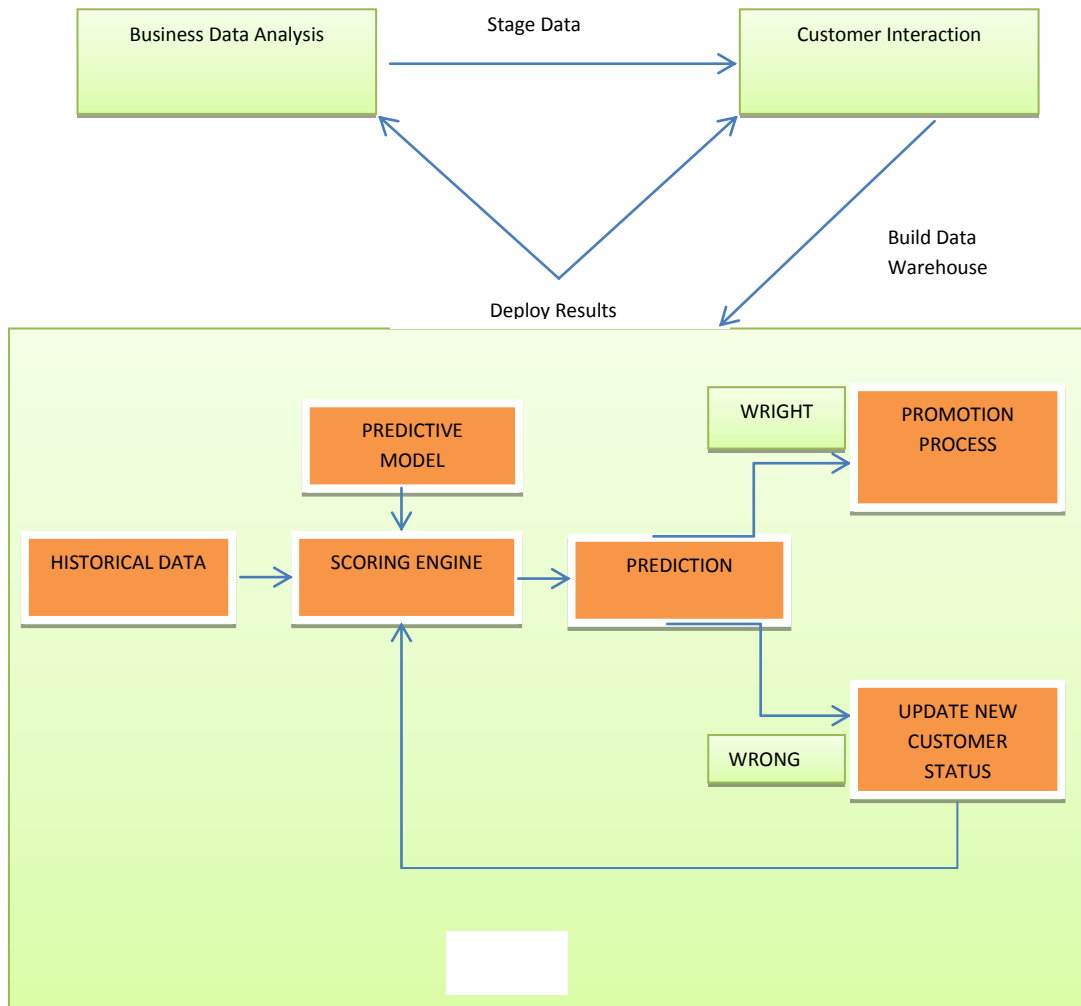


Fig: 3 High Level System architecture integrating DM and e-commerce [7]

The above Customer Behavior Model may not be recommendable for a situation where predictions rarely go wrong and where fraudulent can occur in e-commerce area especially in financial sector. While integrating the redefined CRM model with Ansari et.al [7] (Fig. 3.), a new clear model can be developed where the whole CRM redefining takes place at the analysis stage.



**Fig 4: Proposed Model for Value Assessment of CBM integrating the High Level System Architecture**

Here an example of a bank can be taken which has both resident and non-resident customers. In the present e-commerce scenario, the bank sends e-mails and automated messages for customers according to their status that are taken from historical data of the bank. Prompt customers may be updating bank about their status. Now if the bank uses the scoring engine and predicts a customer who has already returned to his home country for good to be a non-resident and gives his account the benefits for the same, then the system is at fault whereas technically the system is correct. Such systems should have a double check as to when the bank realizes no operations happening in the customer account the bank should check the customer of the new status, correcting historical data to be processed again in the scoring engine. Thus fraudulent cases can be restricted. Sometimes manual checking also results in fraud where customers who do not close their account suffer from. Rarely corruptions are reported where banks absorb interest rates of un-operated accounts. Although human intervention is required from case to case, a successful data mining with correct predictive models can prevent frauds in e-commerce. The above model also is a very good architecture portrayed. The same is applicable in any area factual from lay man depended on the web/internet medium to any industrialist.

In the proposed model (Fig. 4.), the analysis stage in the figure provides an integrated atmosphere for decision support using data transformation, reporting, data mining algorithms, scoring and visualizing. The Stage Data bridges the Business Data Definition component to the Customer Interaction component. Here the data or metadata is transferred through the bridges. The Build Data Warehouse bridges transfer of data collected within Customer Interaction to the Analysis stage. In Customer Interaction the data collector can be implemented using On-Line Transaction Process (OLTP) system designed using entity relation modeling technique. This is because of the reason that large updates can be made with OLTP using short queries.

This is critical for running an e-commerce business, but is not appropriate for analysis [18, 19], which usually requires full scans of several very large tables and a star schema design which business users can understand. Dimensional modeling techniques can be used to build data warehouse for the purpose of data mining. Data transfers to and from OLTP systems could be complex and time-consuming. But the complexity can be reduced making the data warehouse the integral part of the system. However the bridges support the import and integrations of data from both external systems and data providers. The analytical results are actionable at the last bridge, Deploy Results. It provides the ability to transfer models, scores, results and new attributes constructed back into the Business Data Definition and Customer Interaction components for personalization in business rules. For instance, a customer can have a site personalized for the score he gained for his inclination to accept a cross- sell. This could be categorically tough to get implemented in a non-integrated system. However the results are directly reflected on the company's e-commerce business when the metadata are shared across the three components deploying the best results in a CRM environment.

## 7. BUSINESS STRATEGIES

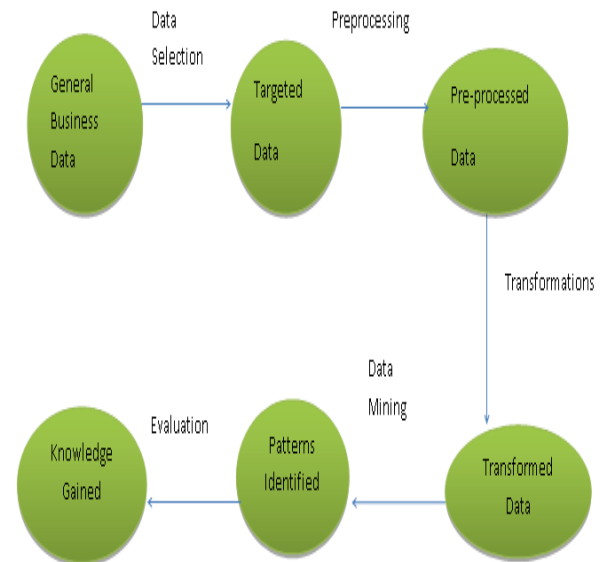
Many business practices can benefit from mining e-commerce information even if they are not directly using it to promote new services for customer satisfaction. Information collected

by e-commerce transactions can also inform businesses that practice no e-commerce at all [16].

Data mining agents like kd nuggets analyze data and submit statistics. Such knowledge reflects both marketing strategy and the e-commerce system. These mining results help companies doing international business get constant guidance on the ever changing business regulations. Some people are needed on board to plan strategies. Over the years, customer care and relationships have been established in the form of call centres with people answering customers' on-line as part of consistent e-business development plans. Thus data mining has created astonishingly impressive sales force in e-commerce. Some standard application summarizes:

- Data mining in transaction indicate the need of stores.
- A clustering technique helps systems make predictions on characteristic features of items.
- Strategic layouts help companies do great business like give discounts on one product but compensate it on another.

Marketing efficiency may be what businesses should fear most [16]. It has already come to a stage where unknowingly customers mine authentic data in e-commerce. This could possibly drive down profit margins of businesses forming tight competition when people have this ability to do intelligent shopping, churning and endlessly demand best values. Thus according to (Fig. 5.) an overview model of data mining process of business perspective [17], the ultimate aim of data mining in e-commerce is to generate knowledge about customers from raw business data, selecting targeted data, pre-processing and transforming them to identify patterns and evaluate knowledge.



**Fig 5: An Overview of Data Mining –A Business Perspective**

Economy is the back bone of any country, whether it is an under developed, developing or developed country. Every government is keen on its growth and economic status. When timely predictions are not made by statisticians on the economy and the financial plans for the future the country may have to face large consequences including recessions. People generally have tendency to migrate to countries with

steady financial background since such countries will always give their citizens the best service. So a close integration between the e-commerce system and the governments also play a very important role in the economy and prosperity of any country. With the need for more growth, e-commerce provides an unusual setting for statisticians, analysts and decision-makers. In such an environment data mining is strongly recommended to improve strategies, prediction and decision making for better customer relations. A tight integration between data mining and e-commerce will help best mining methods and their analytics help prevent cybercrime and predict the next attack as well. Data mining and analytics help track every employee performance, giving best customer treatment on personal preference leading to death of the average and zero blanket assumptions. Customers can mine vast troves of information to make the best decisions giving companies and governments a competitive advantage through 'big data' services thus making data the next natural resource.

## 9. REFERENCES

- [1] Gini Rometty, CEO, IBM, New York, Speech, Council on Foreign Relations in New York, Newspaper, Times of Oman, March 11, 2013.
- [2] Berry, M. J. A., & Linnof, G, Data mining Techniques, New York: Wiley, (1997).
- [3] Agrawal, R., Imielinski, T., and Swami, A., 1993. Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD International Conference on Management of Data (ACM SIGMOD '93), pages 207 – 216, Washington, USA
- [4] Agrawal, R. and Srikant, R., 1994. "Fast algorithms for mining association rules" in Proceedings of the 20th International Conference on Very Large Databases (VLDB '94), Santiago, Chile
- [5] Agrawal, R. and Shim, K., 1996. Developing tightly coupled data mining applications on a relational database system in Proceedings of the 2nd International Conference on Knowledge Discovery in Databases and Data Mining (KDD '96), Portland, Oregon, USA
- [6] Data mining definition available at <http://www.gartner.com/it-glossary/data-mining>
- [7] Ansari, S., Kohavi, R., Mason, L., and Zheng, Z., 2001. Integrating E-Commerce and Data Mining, Blue Martini Software Technical, Report, 2001, Available from the articles section of <http://developer.bluemartini.com>
- [8] Kohavi, R., 2001. Mining e-commerce data: The good, the bad, and the ugly (invited industrial track talk). In Foster Provost and Ramakrishnan Srikant, editors, Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2001. <http://robotics.Stanford.EDU/users/ronnyk/goodBadUglyKDDItrack.pdf>
- [9] Mehmed M.Kantardzic, Jozef Zurada, Next Generation of Data Mining Applications, Wiley Interscience, IEEE, ISBN 0-471-65605-4
- [10] Cooley, R., Mobashar, B., and Shrivastava, J., 1999. Data Preparation for Mining World Wide Web Browsing Patterns, Knowledge and Information Systems, 1
- [11] Berendt, B., Mobasher, B., Spiliopoulou, M., and Wiltshire, J., 2001. Measuring the Accuracy of Sessionizers for Web Usage Analysis, Workshop on Web Mining at the First SIAM International Conference on Data Mining
- [12] Mirjana Mazuran, Elisa Qunitarelli, and Latizia Tanca, IEEE Transactions on knowledge and Data Engineering, Vol.24, No.8, August 2012.
- [13] World Wide Web Consortium, XQuery 1.0: An XML Query language, <http://www.w3c.org/TR/xquery,2007>
- [14] Arumuga Perumal, Integrating E-Commerce and CRM with Data Mining: A New Era, Journal of Internet Banking and Commerce, December 2005, Vol.10, no.3, <http://www.arraydev.com/commerce/jibc/>
- [15] Ed Colet Approaches -DStar Clustering and Classification -Data Mining.
- [16] David L. Banks and Yasmin H. Said, Data Mining in Electronic Commerce, Statistical Science 2006, Vol. 21, No. 2, 234-246,
- [17] Fayyad, U., Piatesky -Shapiro, G., and Smyth, P "From Data Mining To Knowledge Discovery in Databases", AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0-26256097-6. MIT 1996
- [18] Ralph Kimball, The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, John Wiley & Sons, 1996.
- [19] Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite, The Data Warehouse Lifecycle Toolkit : Expert Methods for Designing, Developing, and Deploying Data Warehouses, John Wiley & Sons, 1998.
- [20] Wand Frank, Na Helian, Yau Jim Yip , An Ecommerce System Integrating Data Mining Functionalities, IADIS International Conference, 2003.