

**An Integrated Approach to collaborate Web usage mining and Business  
Intelligence for recommendations in e-commerce**<sup>1</sup> Dimpal visani,<sup>1</sup>Research Scholar, Master in Computer Engineering (M.E.-C.E.), Darshan Engineering College, Rajkot-360001

**Abstract** — Web usage mining is the research area of web mining to discover usage patterns and behaviors from web data (clickstream, purchase information, customer information) in order to understand and serve e-commerce customers better and improve the online Business. The rapid e-commerce growth has made both Business community and customers face a new situation. Due to intense competition on the one hand and the customer's option to choose from several alternatives, the business community has realized the necessity of intelligent marketing Strategies and relationship management. This type of web mining allows for the collection of Web access information for Web pages. This usage data provides the paths leading to accessed Web pages. Web Usage mining allows companies to produce productive information pertaining to the future of their business function ability. The usage data that is gathered provides the companies with the ability to produce results more effective to their businesses and increasing of sales. Usage data can also be useful for developing marketing skills that will out-sell the competitors and promote the company's services or product on a higher level. It will also help-full to e-businesses whose business is based solely on the traffic provided through search engines.

**Keywords-** Web Mining; Web Usage Mining; Business Intelligence; E-Business, Internet Marketing.

**I. INTRODUCTION**

Internet facilities are day by day reaching to customer to all over the world without any physical market place and without any restriction with effective use of E-commerce. Due to which customers who use Internet to purchase product are increasing more. Every day millions of money transactions are done through Internet. The rapid e-commerce growth has made both Business community and customers face a new situation. Due to intense competition on the one hand and the customer's option to choose from several alternatives, the business community has realized the necessity of intelligent marketing Strategies and relationship management.[1].

Recommender systems have the effect of guiding users in a personalized way to interesting objects in a large space of possible options. To realize this, recommender systems use a number of different technologies. We can classify these systems into two broad groups:

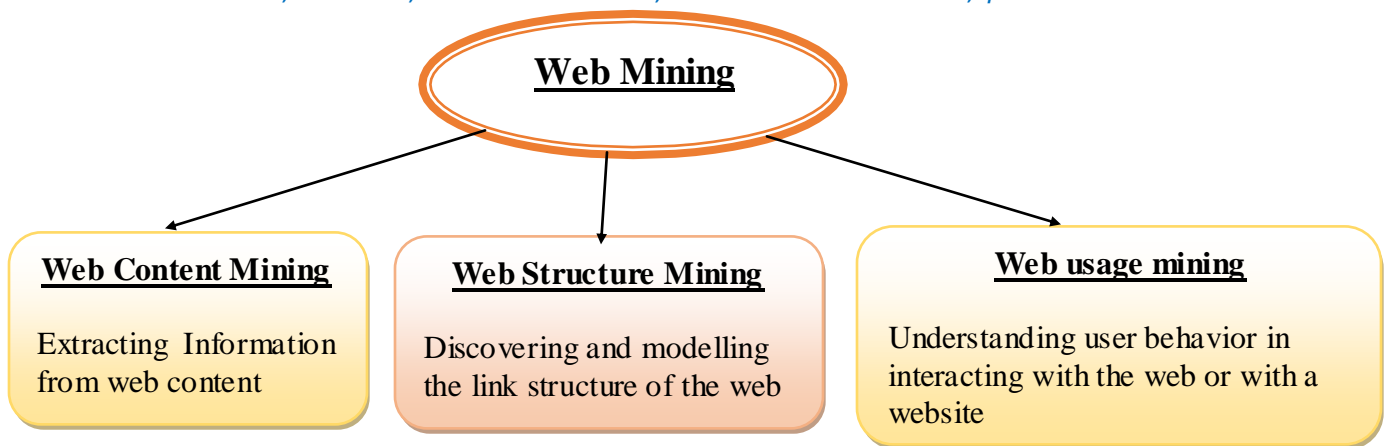
1. Content-based recommender's Content-based recommenders make recommendations by matching a description of an item (a general term for any kind of content or information, e.g. a book, video, or event) and a profile of the user's interests.
2. Collaborative filtering systems Collaborative filtering systems produce user specific recommendations of items based on patterns of ratings or usage (e.g. purchase) without need for exogenous information about items or users.

The massive adoption of the Web as an e-commerce platform has led to a fundamental change in the way that businesses of all sizes interact with their customers. Whereas potential access to a larger, more diverse customer base is generally viewed as an opportunity, this can also represent increased competition. The stakes are high and businesses have to develop sophisticated strategies to attract and retain customers.

**I.1 WEB MINING AND IT'S TAXONOMY**

Web mining is the application of Data Mining technique to discover pattern from WWW (World Wide Web). Web mining can be further divided into three categories,

- 1) Web content mining
- 2) Web structure mining
- 3) Web usage mining



1) **Web content mining :**

Web content mining is the mining, extraction and integration of useful data, information and knowledge from Web page content. It deals with extracting valuable information from Web page contents which is well beyond using keywords in a search engine. Web content mining mainly focuses on the Web Page content rather than the links. Web content is a very rich information resource consisting of many types of information, for example unstructured free text, images, audio, video and metadata as well as hyperlinks. The heterogeneity and the lack of structure that permits much of the ever-expanding information sources on the World Wide Web, such as hypertext documents, makes automated discovery, organization, and search and indexing tools of the Internet and the World Wide Web but they do not generally provide structural information nor categorize, filter, or interpret documents. In recent years these factors have prompted researchers to develop more intelligent tools for information retrieval, such as intelligent web agents, as well as to extend database and data mining techniques to provide a higher level of organization for semi-structured data available on the web.

2) **Web structure mining :**

Web structure mining deals with discovering and modelling the link structure of the Web. Work has been carried out to model the Web based on the topology of the hyperlinks. It is the process of using graph theory to analyze the node and connection structure of a web site.

3) **Web usage mining :**

Web usage mining is the application of data mining techniques to discover interesting usage patterns from web data in order to understand and better serve the needs of web-based application. It deals with understanding user behavior in interacting with the Web or with a Web site. The mined data often contains data logs of users' interactions with Web. The logs include the Web server logs, proxy server logs, and browser logs. The logs include information about the referring pages, user identification, time a user spends at a site and the sequence of pages visited. Information is also collected via cookie files [2]

Web Usage mining allows companies to produce productive information pertaining to the future of their business function ability. Some of this information can be derived from the collective information of lifetime user value, product cross marketing strategies and promotional campaign effectiveness. The usage data that is gathered provides the companies with the ability to produce results more effective to their businesses and increasing of sales. Usage data can also be useful for developing marketing skills that will out-sell the competitors and promote the company's services or product on a higher level.

Web Usage mining is valuable not only to businesses using online marketing, but also to e-businesses whose business is based solely on the traffic provided through search engines. The use of this type of web mining helps to gather the important information from customers visiting the site. This enables an in-depth log to complete analysis of a company's productivity flow. E-businesses depend on this information to direct the company to the most effective Web server for promotion of their product or service.

This web mining also enables Web based businesses to provide the best access routes to services or other advertisements. When a company advertises for services provided by other companies, the usage mining data allows for the most effective access paths to these portals. In addition, there are typically three main uses for mining in this fashion.

The first is usage processing, used to complete pattern discovery. This first use is also the most difficult because only bits of information like IP addresses, user information, and site clicks are available. With this minimal amount of information available, it is harder to track the user through a site, being that it does not follow the user throughout the pages of the site.

The second use is content processing, consisting of the conversion of Web information like text, images, scripts and others into useful forms. This helps with the clustering and categorization of Web page information based on the titles, specific content and images available.

Finally, the third use is structure processing. This consists of analysis of the structure of each page contained in a Web site. This structure process can prove to be difficult if resulting in a new structure having to be performed for each page.

Analysis of this usage data will provide the companies with the information needed to provide an effective presence to their customers. This collection of information may include user registration, access logs and information leading to better Web site structure, proving to be most valuable to company online marketing. These present some of the benefits for external marketing of the company's products, services and overall management.

Internally, usage mining effectively provides information to improvement of communication through intranet communications. Developing strategies through this type of mining will allow for intranet based company databases to be more effective through the provision of easier access paths. The projection of these paths helps to log the user registration information giving commonly used paths the forefront to its access.

Therefore, it is easily determined that usage mining has valuable uses to the marketing of businesses and a direct impact to the success of their promotional strategies and internet traffic. This information is gathered on a daily basis and continues to be analyzed consistently. Analysis of this pertinent information will help companies to develop promotions that are more effective, internet accessibility, inter-company communication and structure, and productive marketing skills through web usage mining.

## **I.2 Business Intelligence**

Business Intelligence (BI) refers to the ability to collect and analyze huge amount of data pertaining to the customers, vendors, markets, internal processes, and the business environment. Business Intelligence (BI) provides historical, current and predictive views of business operations with the help of some technologies that include reporting, online analytical processing, analytics, data mining, process mining, complex event processing, business performance management, benchmarking, text mining, predictive analytics and prescriptive analytics [3].

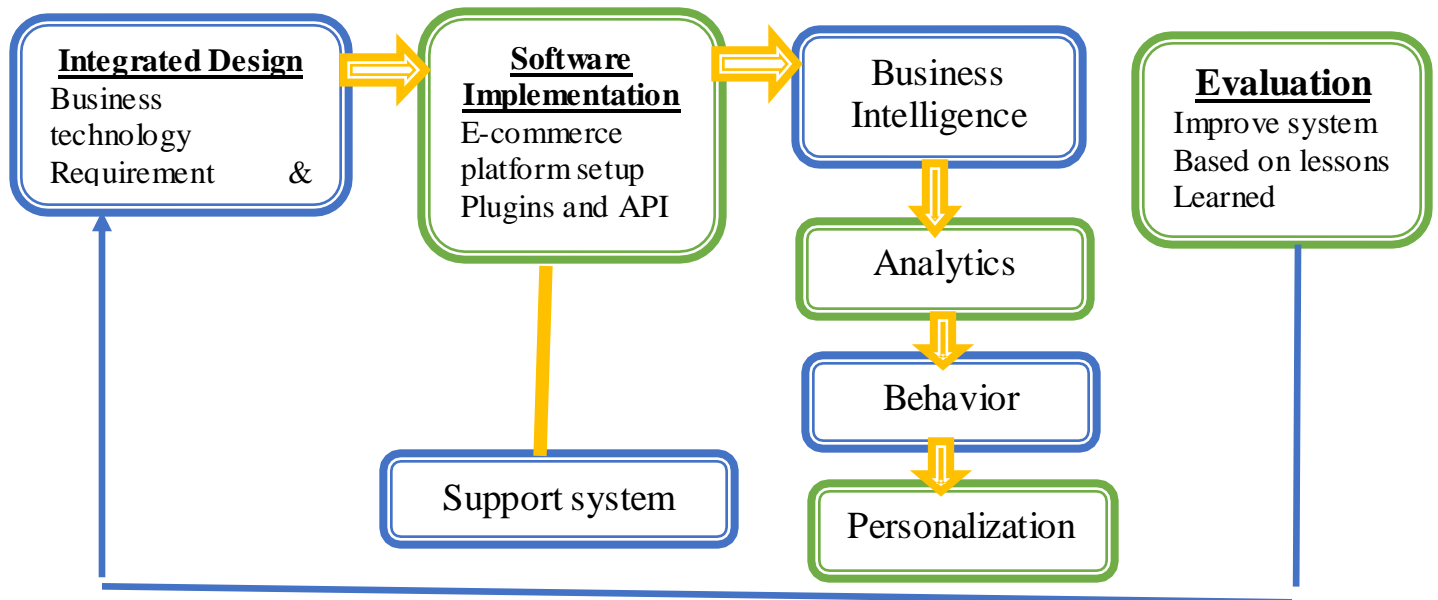
Business intelligence (BI) is a set of theories, methodologies, architectures, and technologies that transform raw data into meaningful and useful information for business purposes. BI can handle enormous amounts of unstructured data to help identify, develop and otherwise create new opportunities. BI technologies provide historical, current and predictive views of business operations. Common functions of business intelligence technologies include reporting, online analytical processing, analytics, data mining, process mining, complex event processing, and business performance management, benchmarking, text mining, predictive analytics and prescriptive analytics. When planning for business data and business intelligence requirements, it is always advisable to consider specific scenarios that apply to a Particular organization, and then select the business intelligence features best suited for the scenario.

Business intelligence includes the processes, methods, and technologies adopted by organizations to answer complex business questions and build comprehensive decision support systems. Business intelligence enables all users in a retail organization to answer questions about the business, for example:

- How do actual sales this period compare to the current plan?
- What is the retail value of inventory on hand, and how does it compare to the same period last year?
- What are the best-selling items in a division or department?
- How effective was the last promotion?

The answers to these questions and others are embedded in the enormous volume of sales and returns, price changes, receipts, and other transactions generated by your retail organization. These transactions are the raw material for business intelligence. Transaction-level data must be converted to information to support decisions in a retail enterprise.

These systems help organizations in maintaining secure, conformed, and highly available data for all levels of users, from top-level executives who make decisions based on corporate-level information to managers and analysts who analyze their areas and take actions based on their findings. Business intelligence is built using several processes, and applications that maintain these processes, using the latest tools and technologies. One of the main components of business intelligence is a data warehouse. A data warehouse is the repository that stores data extracted from multiple source systems, modeled to perform for both data loading, reporting, and ad hoc analysis needs.



## II. PROBLEM DEFINITION

- 1) Sales analytics using Current data trends in e-Commerce for e-businesses whose business is based solely on the traffic provided through search engines.
- 2) The customer's option to choose from several alternatives or recommendations, the business community has realized the necessity of intelligent marketing strategies and relationship management.
- 3) To increase sales of product to large market , society and help people to get right product

## III MOTIVATION

- 1) Due to intense competition on the one hand and the customer's option to choose from several alternatives, the business community has realized the necessity of intelligent marketing strategies and relationship management.

Millions of users access to Websites in all over the world. When they access a Websites, a large amount of data generated in log files which is very important because many times user repeatedly access the same type of Web Pages and the record is maintained in log files. These series can be considered as a Web access pattern which is helpful to find out the user behavior in order to design recommender systems. Through this behavior information, we can find out the accurate user next request prediction that can reduce the browsing time of Web pages.

In recent years, there has been an increasing number of research works done with regard to WUM. The main Motivation of this survey is to know what research has been done on WUM in future request prediction. Also, how we can use various WUM techniques to develop efficient and effective recommendation systems.

- 2) Analyze how far shoppers get in the shopping funnel and where they drop off.
- 3) Understand which products are viewed most, which are frequently abandoned in cart and which ones convert well.
- 4) Upload rich product metadata to slice and dice your data.
- 5) Create rich user segments to delve deeper into your users' shopping behavior and the products they interact with.

## IV. PROPOSED ARCHITECTURE



Having access to statistical information from all areas of your online marketing and sales activities gives you an advantage over competitors that do not have this information. Understanding trends and which marketing channels are no longer profitable allows you to maneuver as a business before damage is done to your bottom line. And, understanding shifts in consumer behavior gives you insights into the demands of your market. Knowing these things enables you to drop certain products or make strategic changes in your pricing that will result in big gains or, at the very least, limit damage to your profits

One of the basic concepts of web analytics is funnels. All marketing activities can and should be seen in terms of funnels. The idea of funnel analytics is that your target audience will go through a step-by-step flow or funnel until they make a purchase on your site. A typical marketing funnel may look like this:

1. A fan on your business's Facebook page sees one of your posts.
2. The fan clicks on the post.
3. The fan arrives on a landing page advertising a specific product and clicks on "add to cart."
4. The fan clicks on checkout.
5. The fan enters their personal information and finalizes the purchase.

At each step of the process, a certain percentage of people will drop out of the funnel. Knowing these percentages will help you determine the barriers and psychology behind your customers.

#### **Description:**

##### **1. Collecting Data**

Collect the data from web APIs and google analytics for analyzing and improving ecommerce and business. You can collect data from Paid/Open source apis and use web log files from web server.

##### **2. Cleaning Data:**

Clean this web log files in a desired format and a bad data to good data which would give a meaning full output out of this system. Validate data this data against business rules so that a data would have authenticity. Remove/alter the bad data to a meaning full information.

##### **3. Populate data mart:**

Design a physical data model using the data and kip's required for analytics. Follow the Kimball's dimension modelling to design a datamart. Data mart should be in the form of a star or snowflake model. Once datamart is designed populate the cleaned data to a model using data integration services.

##### **4. Data Mining:**

Once the mart is populate apply the mining algorithm to analyses the data.  
Generate results and KPI's affecting business.  
Control the things moving in wrong directions.  
Apply the algorithm and get result sets.

##### **5. Data Visualization :**

Once data is mined apply suitable visualizations on this data using protovis,D3 and other graphics editors.  
Develop dashboards and reports to represent results.

## **VII. CONCLUSION**

Having a strong data collection setup and access to detailed reports in services like KISS metrics, Google Analytics, Facebook, and others will provide you with the information you need to make tough, but critical business decisions. By integrating Web usage mining with Business intelligence will provide a integrated frame work to the E-commerce base business who goes slowly and have less profit.

## **REFERENCES**

- [1] Ajith Abraham , “Business Intelligence from Web Usage Mining ” , Journal of Information & Knowledge Management, Vol. 2, No. 4 (2003) 375{390 c iKMS & World Scientific Publishing Co.
- [2] Prakul Gupta<sup>1</sup> Amit Sharma<sup>2</sup> Dr. Sunil Kr Singh<sup>3</sup> , “ Data Harvesting through Web Mining: A Survey ” , The International Journal Of Engineering And Science (IJES) || Volume || 3 || Issue || 4 || Pages || 21 -27 || 2014 || ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805.
- [3] Ranak Ghosh, Sujay Halder, Soumya Sen,” An Integrated Approach to Deploy Data Warehouse in Business Intelligence Environment”, 978-1-4799-4445-3/15/\$31.00 ©2015 IEEE.
- [4] Ajith Abraham, Department of Computer Science, Oklahoma State University,” Business Intelligence from Web Usage mining” , 8 September 2015.
- [5] Abdul-Aziz Rashid Al-Azmi , “Data, text, and web mining for business Intelligence: a survey ” , International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol.3, No.2, March 2013.
- [6] Konda Sreenu , Dept. of CSE, Sir CRR College of Engineering, Eluru, AP, India, “Web Data Mining Based Business Intelligence and Its Applications”, IJCST Vol. 4, Issue Spl - 4, Oct - Dec 2013.
- [7] Mehak , Mukesh Kumar , Naveen Aggarwal , “Web Usage Mining: An Analysis” , JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 5, NO. 3, AUGUST 2013.
- [8] Prakul Gupta<sup>1</sup> Amit Sharma<sup>2</sup> Dr. S nil Kr Singh<sup>3</sup> ,” Data Harvesting through Web Mining: A Survey ” , The International Journal Of Engineering And Science (IJES) || Volume || 3 || Issue || 4 || Pages || 21 -27 || 2014 || ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805 .
- [9] Jesus Angel Fernandez Canelas<sup>1</sup>, Quintin Martin Martin<sup>2</sup>, Juan Manuel Corchado Rodriguez<sup>3</sup> ,” Business Intelligence Expert System on SOX Compliance over the Purchase Orders Creation Process , *Intelligent Information Management*, 2013, 5, 49-72 , Published Online May 2013 .
- [10] Mr. Akshay A. Adsod, Prof. Nitin R. Chopde, ” A Review on: Web Mining Techniques ” , International Journal of Engineering Trends and Technology (IJETT) – Volume 10 Number 3 - Apr 2014.
- [11] P.Prabhu<sup>1</sup> and N.Anbazhagan,” A NEW HYBRID ALGORITHM FOR BUSINESS INTELLIGENCE RECOMMENDER SYSTEM”, International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.2, March 2014
- [12] Anupama Prasanth, “Web Usage Mining – Its Application in E-Services “International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 2, February 2013.
- [13] P Salman Raju, Dr. V Rama Bai, G Krishna Chaitanya, “Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries”, International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 1, January 2014.
- [14] A. Venkata Krishna Prasad<sup>1</sup>, Prof. S.Ramakrishna<sup>2</sup> “Mining for Web Engineering “ International Journal of Computer Trends and Technology- volume2Issue2- 2011 .
- [15]Jayantha ranjan, “business intelligence: concepts, components, techniques and benefits ” , Journal of Theoretical and Applied Information Technology © 2005 - 2009 JATIT. All rights reserved