

# Behavior existing between Web Usage Mining and Data Mining

Ankur G. Chavhan<sup>1</sup>

St. John College of Engineering and Management,  
Palghar (East)<sup>1</sup>  
[ankurgchavhan@gmail.com](mailto:ankurgchavhan@gmail.com)<sup>1</sup>

Chitra T. Wasnik<sup>2</sup>

Lokmanya Tilak College of Engineering,  
Koparkhairane, Navi Mumbai<sup>2</sup>  
[chitrawasnik@yahoo.com](mailto:chitrawasnik@yahoo.com)<sup>2</sup>

**Abstract**— This paper describes how the relationship between webs uses mining and data mining exists. Whenever we deals with web uses mining at that time we can consider e-Bank services as one of the example. Whenever we are dealing with e-business at that time we have to use following two techniques Web uses mining and 2) Data mining. Clickstream data is one of the organizations used to describe the nature in Banks e-services. Web uses mining give some additional result than we expecting result along with intermediate data that describe additional information. So that user can access important data as well as relatively consequent, due to this reason customer having much more information in hand. In data mining it contain huge data by avoiding redundancy effectively in a easy manner. Data mining is useful in banking sector for increasing their performance by avoiding fraud. Data mining increase customer product services level effectively.

---

**Keywords**— Web Mining; Clickstream; Data Mining; Data Mining Techniques; Banking Sector

---

## 1. INTRODUCTION

Data mining contain huge data stored in database also some data shows relationship between them. Many algorithms is available to help to determine particular information into a database automatically. Such techniques are as follows. And they are statistics, machine learning and pattern recognition. Data mining is also known as Knowledge-Discovery in Databases (KDD) [2, 3]. The data mining is used to determine prediction by seeming pattern and generate report capabilities. From prediction available with help of data mining, it gives knowledge about customer regarding his background than means it tells how the particular customer can be treating. [2]

In this paper, the relationships between the techniques and data mining Web usage are studied and a general framework is presented for fully integrating of domain Web usage mining and data mining techniques for processes at different stages. The rest of the paper is organized as follows: Section 2 briefly introduces the web data mining and the web usages mining process. Section 3 provides Web data mining types. Section 4 introduces data mining techniques. Section 5 data mining in banking sector. Section 6 contains the conclusion of the research.

## 2. WEB DATA MINING

The definition from Gartner Group seems to be most

comprehensive, as they define data mining as "the process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of Information and data stored in repositories and by using the different pattern" [5]. Data mining for higher education is a process of uncovering hidden trends and patterns that lend them to predicative modeling using a combination of explicit knowledge base, sophisticated analytical skills and academic domain knowledge [5, 6]. It is producing new observations from existing observations. Or, as explained by [6] "data mining is the process of automatically extracting useful information and relationships from immense quantities of Information and data. In its purest form, data mining doesn't involve looking for specific information. Data mining simply finds patterns that are already present in the existing data, rather than starting from a question or a hypothesis [6].

The increasing growth in online information combined with the unstructured Web data requires the development of powerful and computationally efficient Web data mining tools Web mining intends to find out useful facts and knowledge from Web pages, Web hyperlinks, page content and usage log. Web mining of e-bank service enables employees or a bank to support e-business, understanding marketing dynamics, new promotions suggesting on the Internet, etc. There is an increasing tendency among banking companies,

organizations and individuals to collect information through Web data mining to utilize that information in

their best interest and to gain business intelligence that helps companies make necessary business decisions [7].

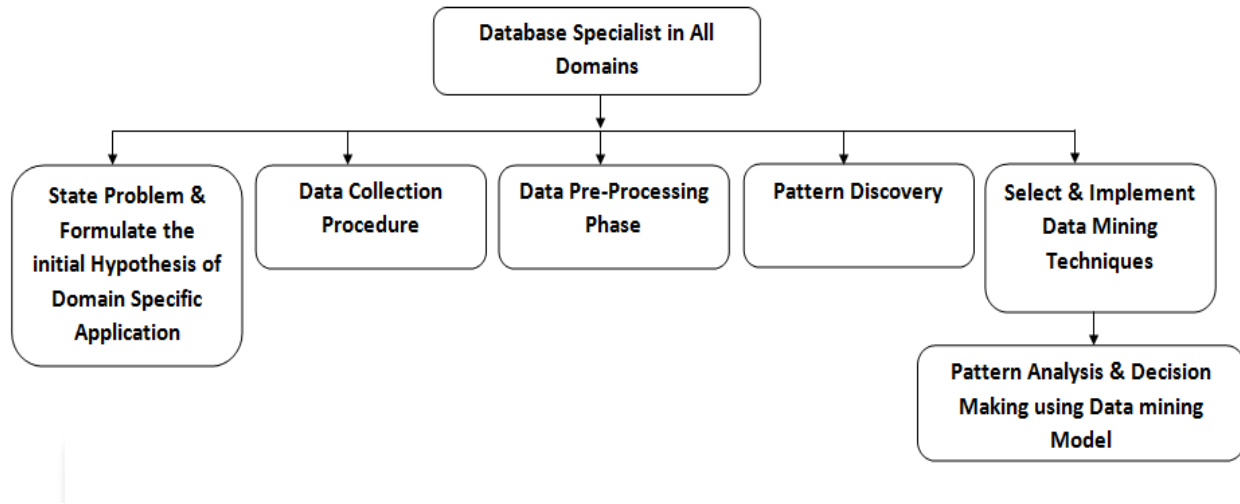


Fig.1. Web Usage mining process

### 3. WEB DATA MINING – TYPES

Web involves three various types of data, which are:

- a. Heterogeneous Information and data content on the WWW and Hyperlinks exist among Web pages within a site and across different sites.
- b. The web log data concerning the users who browsed the Web pages, and
- c. The Web structure Information and data (whether they are structured, semi-structured and unstructured Web pages).

Thus, the WWW data mining should focus on three issues; web structure mining, web content mining and web usage mining [8].

- a. **Web structure mining:** examines how the Web documents are structured, and attempts to discover the model underlying the link structures of the Web.
  - i. *Intra-page structure mining:* evaluates the arrangement of the various HTML or XML tags within a page.
  - ii. *Inter-page structure mining:* refers to hyper-links connecting one page to another [9].
- b. **Web usage mining (Clickstream Analysis):** involves the identification of patterns in user navigation through Web pages in a domain.

Processing, Pattern analysis, and Pattern discovery.

- c. **Web content mining:** used to discover what a Web page is about and how to uncover new knowledge from it [9]. It includes the Information and data from different server access logs, user registration or profiles, user sessions or transactions etc.

### 4. DATA MINING TECHNIQUES

There are several major data mining techniques have been developing and using in data mining projects recently including data mining Techniques. We will briefly examine those data mining techniques in the following sections.

#### A. Association

Association and correlation is one of the best known data mining technique. It is usually to find frequently used data items in the large Information and data sets. In association, a pattern is discovered based on a relationship between items in the same transaction. That's the reasons why association technique is also known as relation technique. The problem of mining association rules can be stated as follows: Let  $I = \{i_1, i_2, \dots, i_m\}$  be a set of items. Let  $T = (t_1, t_2, \dots, t_n)$  be a set of transactions (the database), where each transaction  $t_i$  is a set of items such that  $t_i \subseteq I$ . An *association rule* is an implication of the form:

$X \rightarrow Y$ , where  $X \subset I$ ,  $Y \subset I$ , and  $X \cap Y = X$  (or  $Y$ ) is a set of items, called an *item set*. (1)

The association technique is used in market basket analysis to identify a set of products that customers frequently purchase together. Retailers are using association technique to research customer's buying habits. Based on historical sale Information and data, retailers might find out that customers always buy crisps when they buy beers, and therefore they can put beers and crisps next to each other to save time for customer and increase sales [10, 11].

### B. Classification (predictive)

Classification is a classic data mining technique

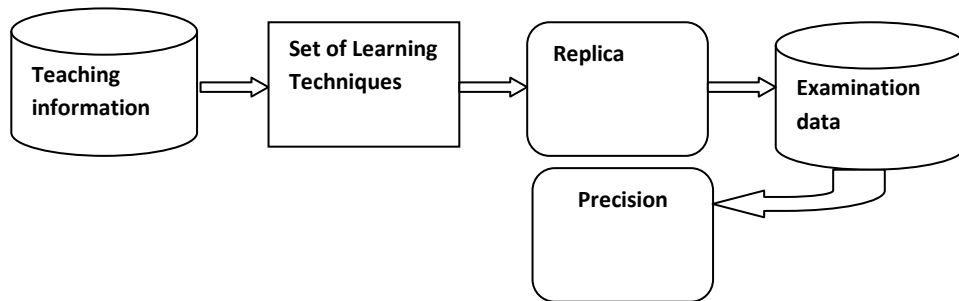


Fig. 2. The basic learning process: [1, 10, 11]

In classification, the Information and data items are classified into groups. For example, we can apply classification in application that “given all records of employees who left the company; predict who will probably leave the company in a future period.” In this case, the records of employees are divided into two groups that named “leave” and “stay”. And then data mining classifies the employees into separate groups [10, 11].

### C. Clustering (descriptive)

Clustering is a data mining technique that makes meaningful or useful cluster of objects which have similar characteristics using automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes. To make the concept clearer, we can take book management in library as an example. In a library, there is a wide range of books in various topics available. The challenge is how to keep those books in a way that readers can take several books in a particular topic without hassle. By using clustering technique, we can keep books that have

based on machine learning. Basically classification is used to classify each item in a set of Information and data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics [12].

The accuracy of a classification model on a test set is defined as:

$$\text{Precision} = \frac{\text{Number of correct Classifications}}{\text{Total Number of Examination Cases}} (2)$$

Where, a correct classification means that the learned model predicts the same class as the original class of the test case.

some kinds of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in that topic, they would only have to go to that shelf instead of looking for entire library [14].

### D. Prediction (regression)

The prediction, as it name implied, is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. For instance, the prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit Information and data, we can draw a fitted regression curve that is used for profit prediction [10, 11].

### E. Sequential Patterns

Sequential patterns analysis is one of data mining technique that seeks to discover or identify similar patterns, regular events or trends in transaction Information and data over a business period. In sales,

with historical transaction Information and data, businesses can identify a set of items that customers buy together different times in a year. Then, businesses can use this information to recommend customers buy it with better deals based on their purchasing frequency in the past [10, 11].

F. Classification Decision Trees

Decision tree is one of the most used data mining techniques because its model is easy to understand for users. In decision tree technique, the root of the

decision tree is a simple question or condition that has multiple answers. Each answer then leads to a set of questions or conditions that help us determine the Information and data so that we can make the final decision based on it [9]. Once the relationship is extracted, then one or more decision rules can be derived that describe the relationships between inputs and targets [15]. The general form of this modeling approach is illustrated in Figure 3. The data set can be divided into: a) Training set used to build the model. b) Test set used to determine the accuracy of the model [16]

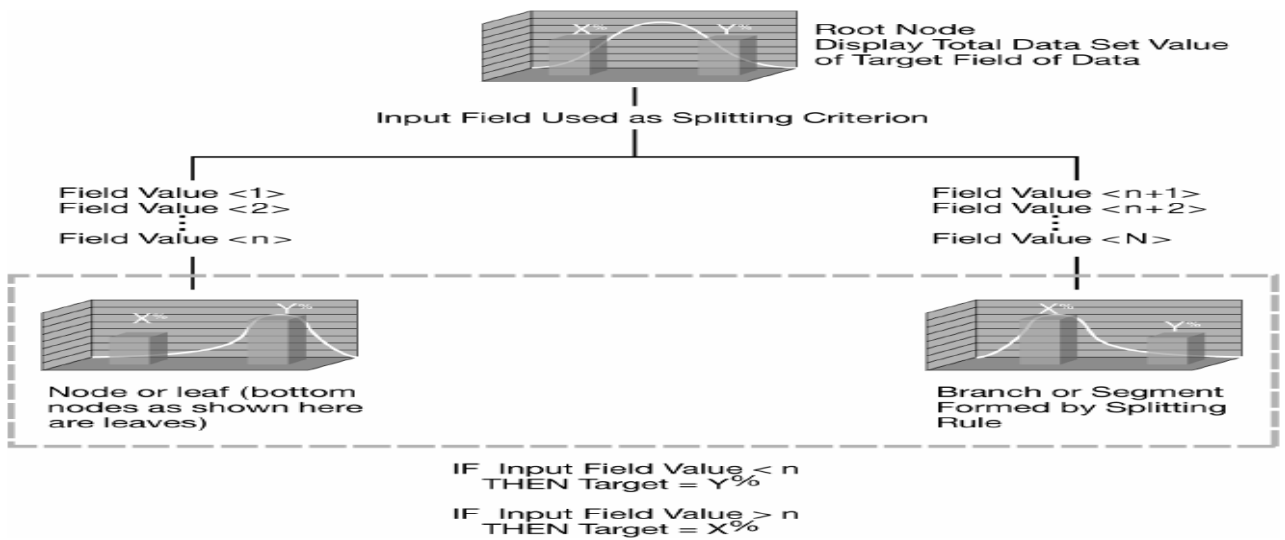


Fig.3. Decision trees [15]

A record or observation from the Information and data set is assigned by each rule to a node in a branch based on the value of one of the fields or columns in the Information and data set [15]. Decision rules can predict the values of new observations that contain values for the inputs, but might not contain values for the targets. Rules are applied one after another. The nested hierarchy of branches is called a *decision tree*. For each leaf, the decision rule provides a unique path for Information and data to enter the class that is defined as the leaf. Framework of the integrating between Web usage Mining and data mining techniques is shown in figure 4. Although the boundaries between prediction and description are not sharp (some of the predictive models can be descriptive, to the degree that they are understandable, and vice versa). The goals of prediction and description can be achieved using a variety of particular data-mining methods. Classification is learning a function that maps

(classifies) a Information and data item into one of several predefined classes.

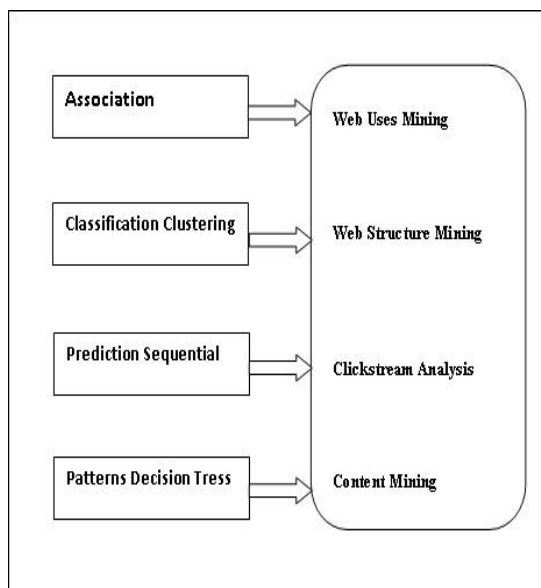


Fig. 4. Research replica alternative: Research Adapted [1]

## 5. DATA MINING IN BANKING SECTOR

In banking e-services, data mining can possibly answer these questions [10, 18, and 17], about what customer Information and data the banking sector needs to explore?

- [1] What transactions does a customer do before shifting to a competitor bank?
- [2] What is the profile of an ATM customer and what type product's type is he likely to buy?
- [3] Which bank products are often served together by which groups of customers? (Target Marketing).
- [4] What are credit transactions patterns that lead to fraud?
- [5] What is the outline of a high-risk borrower? (To prevent defaults and bad loans).
- [6] What services and benefits would current customers request?
- [7] What is worldwide just-in-time availability of information that improves business performance of financial institutions?
- [8] How big is a need for continuous analysis of resulting Information and data?
- [9] What is the transaction behavior of the bank customers that the managers may need to know?

Data mining can contribute to solve business problems in banking and financial institutions by finding patterns, and correlations in business

information and market prices that are not immediately clear to managers because of the very large volume of Information and data. Trends can be analyzed and predicted with the availability of historical Information and data and the data warehouse assures that everyone is using the same Information and data at the same level of extraction, which eliminates conflicting analytical results and arguments over the source and quality of Information and data used for analysis [10].

The banking sector consists of public sector, private sector and foreign banks. In the market, various IT-based banking products, services and solutions are available. The most common of them are Phone Banking; ATM facility; Credit, Debit and Smart Cards; Internet Banking and Mobile Banking. The banking sector is on the border of revolutionary change in the way it functions and delivers its services to customers. Customer Relationship Management (CRM) Encompasses the data mining activities that a service provider undertakes to understand its customers [19]. With the increasing economic globalization and improvements in information technology, large amounts of financial Information and data are being generated and stored. Data mining techniques are used to discover hidden knowledge, predicted patterns and new rules from large data set and obtain predictions for trends in the future and the behavior of the financial markets [20].

A bank has Information and data about clients to whom the bank gave credits in the past. The client Information and data include personal Information , financial status description data and the financial behavior before and at the time the client was given the credit. The clients are divided into four classes. a) The first class includes all the clients who paid back the credit without any problems; b) the second class includes the clients who paid back with little; c) the third class includes the clients who get a credit after detailed checks because considerable problems of payback happened in the past; and d) the forth class includes all the clients who did not pay back at all. A data mining prediction model can be used for new clients to predict the probability for each class. The combinations of attributes; which have a high probability, of clients will be identified by the prediction model. Data Mining can help banks to better predict the credit worthiness of customers [21].

As banking is in the service industry, the task of maintaining a strong and effective CRM is a critical issue. To do this, banks need to invest their resources to better understand their existing and prospective customers. Offering new customers credit cards, extending existing customers lines of credit, and

approving loans can be risky decisions for banks if they do not know anything about their customers. Data mining can reduce the risk of banks that issue credit cards [19]. Using characteristics such as credit history, length of employment and length of residency, Data mining can also derive the credit behavior of individual borrowers with installment, mortgage and credit card loans. With the help of data mining, more fraud pattern identification in banking industries are being detected and reported. The bank might want to use the classification regions to automatically decide whether future loan applicants will be given a loan or not. In Primary goals of data mining, the researchers intrude a description of the methods used to address these goals. Most data-mining methods are based on tried and tested techniques from machine learning, pattern recognition, and statistics: classification, clustering, regression, and so on. The array of different algorithms under each of these headings can often be amazing to both the novice and the experienced data analyst. Many data-mining methods publicized in the literature, there are really only a few fundamental techniques. The actual underlying model representation being used by a particular method typically comes from a composition of a small number of well-known options: polynomials, splines, kernel and basis.

## 6. CONCLUSION

Data mining contain tool used to access essential data from present data and enable better decision-making process and to increase the better performance. Many Data mining techniques have used for developing mining projects. Web documents are structured, and attempt to find out the model primary the link structures of the Web. There are relationships between the data mining technique and web usage. The paper provide a framework for fully integrate domain Web usage mining and data mining technique for process at different stage, including the preprocessing and pattern discovery phases. Data mining techniques can be very helpful to the banks for better increasing performance, acquiring new customers, fraud detection in actual time, providing information related to the product, and analysis for the customer product.

## REFERENCES

- [1] Omer Adel Nassar , Dr. Nedhal A. Al Saiyd: "The Integrating Between Web Usage Mining and Data Mining Techniques"
- [2] Murthy, I.,K., Data Mining- Statistics Applications: A Key to Managerial Decision Making, SOCIO 2010, available at: <http://www.indiastat.com/article/16/krishna/fulltext.pdf>
- [3] Zack, M.H. " Developing a knowledge strategy: epilogue" Available at: [http:// web .cba.neu.edu/-mzack /articles](http://web.cba.neu.edu/~mzack/articles). Y-Shapiro, P. Smyth, and R. Uthurusamy, 569–588. Menlo Park, Calif.: AAAI Press, 2001.
- [4] Hand, D., Mannila, H., Smyth, P., Principles of Data Mining, The MIT Press, 2001. Available at: [ftp://gamma.sbin.org/pub/doc/books/Principles\\_of Data Mining.pdf](ftp://gamma.sbin.org/pub/doc/books/Principles_of_Data_Mining.pdf)
- [5] Luan, J. Data Mining Applications in Higher Education, Executive report. Available at: [http://www.spss.ch/upload/1122641492 Data%20mining%20application s%20in%20higher%20education.pdf](http://www.spss.ch/upload/1122641492_Data%20mining%20application%20in%20higher%20education.pdf)
- [6] Ranjan, J., Ranjan,R., Application Of Data Mining Techniques In Higher Education In India, International Conference On Innovation In Redefining Business Horizons, Ghaziabad, India, 18 - 19 December, 2008.
- [7] Mitharam, M., D., Preprocessing in Web Usage mining, International Journal of Scientific & Engineering Research, Volume 3, Issue 2, February-2012. Available at: <http://www.ijser.org/researchpaper%5CPreprocessing-in-Web-Usage-mining.pdf>.
- [8] Fernandez, I. B. Gonzalez.,A. Sabherwal, R., knowledge management challenges, solutions, and technologies : Pearson Education Inc., New jersey, United states of America, 2004 .
- [9] Kolek, S., and Kirmaci , O., Web Mining & Clickstream Analysis. Available at: [http://diuf.unifr.ch/is/studentprojects/pdf/reports/CRM\\_SS06\\_Web\\_Mining\\_And\\_Clickstream\\_Analysis\\_ \(StefanKolek \\_OezkanKirmaci\).pdf](http://diuf.unifr.ch/is/studentprojects/pdf/reports/CRM_SS06_Web_Mining_And_Clickstream_Analysis_(StefanKolek_OezkanKirmaci).pdf)
- [10] Bhambri, V., Application of Data Mining in Banking Sector, IJCST Vol. 2, Issue 2, PP. 199-202, June 2011. Available at: <http://www.ijcst.com/vol22/1/vivek.pdf>
- [11] Laudon,K, Laudon. J., Management Information System, International Edition, Eighth Edition, Pearson Prentice Hall.2004.
- [12] Cheeseman, P., and Stutz, J. 'Bayesian Classification (AUTOCLASS): Theory and Results In Advances in Knowledge Discovery and Data Mining, eds. Articles FALL.1996 .PP. 51. [http://www.imn.htwk-leipzig.de/~kudrass/Lehrmaterial/Oberseminar/2010/09-Web\\_Mining-Khourivogel.pdf](http://www.imn.htwk-leipzig.de/~kudrass/Lehrmaterial/Oberseminar/2010/09-Web_Mining-Khourivogel.pdf)
- [13] Rubenking, N. Hidden Messages. PC Magazine. May 22, 2001
- [14] <http://www.rightpoint.com/community/blogs/viewpoint/archive/2011/11/08/data-mining-in-banks-and-financial-institutions.aspx>
- [15] Decision Trees available at: <http://support.sas.com/publishing/pubcat/chaps/57587.pdf>
- [16]

- [17] Decision Trees. Retail Industries [http://icai.org/resource\\_file/9935588-594.pdf](http://icai.org/resource_file/9935588-594.pdf)  
<http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf>
- [18] <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.94.4020> [21] Dass, R., Data Mining In Banking And Finance: A Note For Bankers. Available at: <http://iimahd.ernet.in/publications/data/Note%20on%20Data%20Mining%20%26%20BI%20in%20Banking%20Sector.pdf>
- [19] Abiteboul, R, et al., Foundation of Database, Addison –Wesley.1995
- [20] Data Mining: A Competitive Tool in the Banking and