

# International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: [www.ijarcsms.com](http://www.ijarcsms.com)

## *Application of Data Mining Techniques for Web Personalization*

**Jijnasa Patil<sup>1</sup>**Computer Department  
VJTI College, Mumbai, India**Pranoti Nage<sup>2</sup>**Computer Department  
VJTI College, Mumbai, India**Gagandeep Chhabda<sup>3</sup>**Computer Department  
VJTI College, Mumbai, India**Shreyansh Jain<sup>4</sup>**Computer Department  
VJTI College, Mumbai, India

*Abstract: The web has expanded much more than expected in past few years. Further, the advent of new secure technologies, the online shopping trend has increased. People are bored of searching products online page by page. So they prefer websites which provide a quick access to the products along with recommending new products based on their preferences. It gives a personal feel that they are being valued and also helps in retaining of the customers. So we need to personalize the web. We have used a simple concept of using cookies to store the searches and then applying Apriori algorithm to recommend products to the customer. The idea of using Data Mining techniques and availability of vast data online will help in successfully recommending the products to the customers. So, customers can save their time and can also get a value for their money.*

*Keywords: Web Personalization, Apriori, Cookie, E-Commerce.*

### I. INTRODUCTION

In past years, with the advent of online transaction facilities, E-Commerce websites, increase in reliability, availability and speed of internet, the trust of customers in online shopping, and competitive prices in retail and online markets, a lot of transactional data is available online. A large number of products are available online, and everyone in their busy schedule has no time to go to retail markets and search for shops and stand in queues to make payments, so people prefer online shopping. Different products like jewellery, electronic goods, furniture, clothing, cooking utensils, etc. are being purchased online. So, data for different types of transaction is available. Also, people now a days prefer websites where they can easily search their products, or websites which recommend them products based on their preferences, current trends, or their past searches and past purchases. So, a strong recommendation algorithm is necessary for the E-Commerce website to reach out to people.

Web is a collection of hyperlinked documents on one or more Web servers. Web mining is data mining techniques used to extract knowledge from web. Mining is a helpful tool in the process of transforming human understandable content in to machine understandable semantics. Data mining is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

Personalising Web is a method used for improving customer satisfaction with the website and also helps in increasing the customer trust and thus helps to maintain customers. Personalization is the practice of dynamically tailoring your site to the

wants and needs of each user. Personalizing will give visitors a more relevant and engaging experience and is one of the best ways to boost loyalty and conversion. Web pages can be personalized based on the characteristics (interests, social category, context, etc.), actions (click on button, open a link, etc.), intent (make a purchase, check status of an entity), or any other parameter that can be identified and associated with an individual, therefore providing them with a tailored user experience. There are several categories of web personalization:

- Behavioural
- Contextual
- Technical
- Historic data
- Collaboratively filtered

This paper proposes the use of historic searches of the user and his past transactions, his preferences and the transactions of other users, to recommend products of related interest, which have a higher probability of being purchased by the customer and also representing the current market trends to the user, while he is on the website. The data of past searches is stored in cookies in the web browser and the transaction data is obtained from the database of the website.

## II. RELATED WORK

The Internet grew from 2000-2009 at an estimated rate of 380% [1]. This exponential growth of information resources and services available on www is making web site exploration difficult, therefore new information services like Web personalization are in high demand and distinguished personalization schemes have been suggested in the recent decade.

According to Kobsa (Kobsa, 2001), Web Personalization is the ability to adapt to the user's preferences, knowledge, capabilities and to the environmental factors. Systems that take individual characteristics of the users into account and adapt their behaviour accordingly have been empirically shown to benefit users in many domains. Examples of adaptation include customized content (e.g. personalized finance pages or news collections), customized recommendations or advertisements based on past purchase behaviour, customized (preferred) pricing, tailored email alerts, express transactions (Kobsa, 2001).

Letizia [2] is considered to be the first system that records the user's navigation behaviour and gives interesting recommendations to the user. WebWatcher [3] is web content based system that provides navigation hints to the user, based on a knowledge acquired about user's interests, the location and relevance of the items in the site, and the way in which other users interacted with the collection in the past and it performs well from the perspective of the end user who is searching the Web for information but it is less useful in E-commerce applications. WebWatcher [3] and Letizia [1] generally rely on personal profiles and content similarity of Web documents to these profiles to make recommendations.

Analog [4] was one of the leading personalization systems based on the web usage mining methodology. The analysis of log data discovers valuable web usage patterns [5]. In 2000 Mobasher [6] proposed the web usage-based Web personalization system called Web Personalizer for recommending Web pages on Server-Side to users. The Web Personaliser provides a personalization framework based on web log mining and using data mining techniques for extraction of knowledge for generating the recommendations to current users based on their browsing navigational history.

Good literature of the web usage mining field has been made available by Eirinaki [7], Koutri [8], Pierrakos [9]. Web usage mining, the main component of a web personalization system, is generally, a three step process, consisting of data preparation, pattern discovery, and pattern analysis. The web usage mining extensively focus on discovering the interesting patterns in order to understand the users' navigational behaviour. The navigational behaviour may play an important role in the decisions concerning site restructuring or modification [10, 11].

## III. PROPOSED SYSTEM

The aim of our system to be developed is that the customer should get the products he is interested in purchasing and he should not waste time in searching the products. So this will give the customer a personalized feel while he is shopping. Initially, when the customer comes for the first time, he has to fill the preferences and the system will check the transactions of other people with similar preferences and recommend some products. But preferences may keep on changing, so as the customer is shopping, he will get the products based on transactions of other people who purchased the similar products. So the recommendation will be dynamic. We basically attempt to combine the traditional Apriori algorithm with the server side data analysis through cookies to create a recommendation engine. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time.

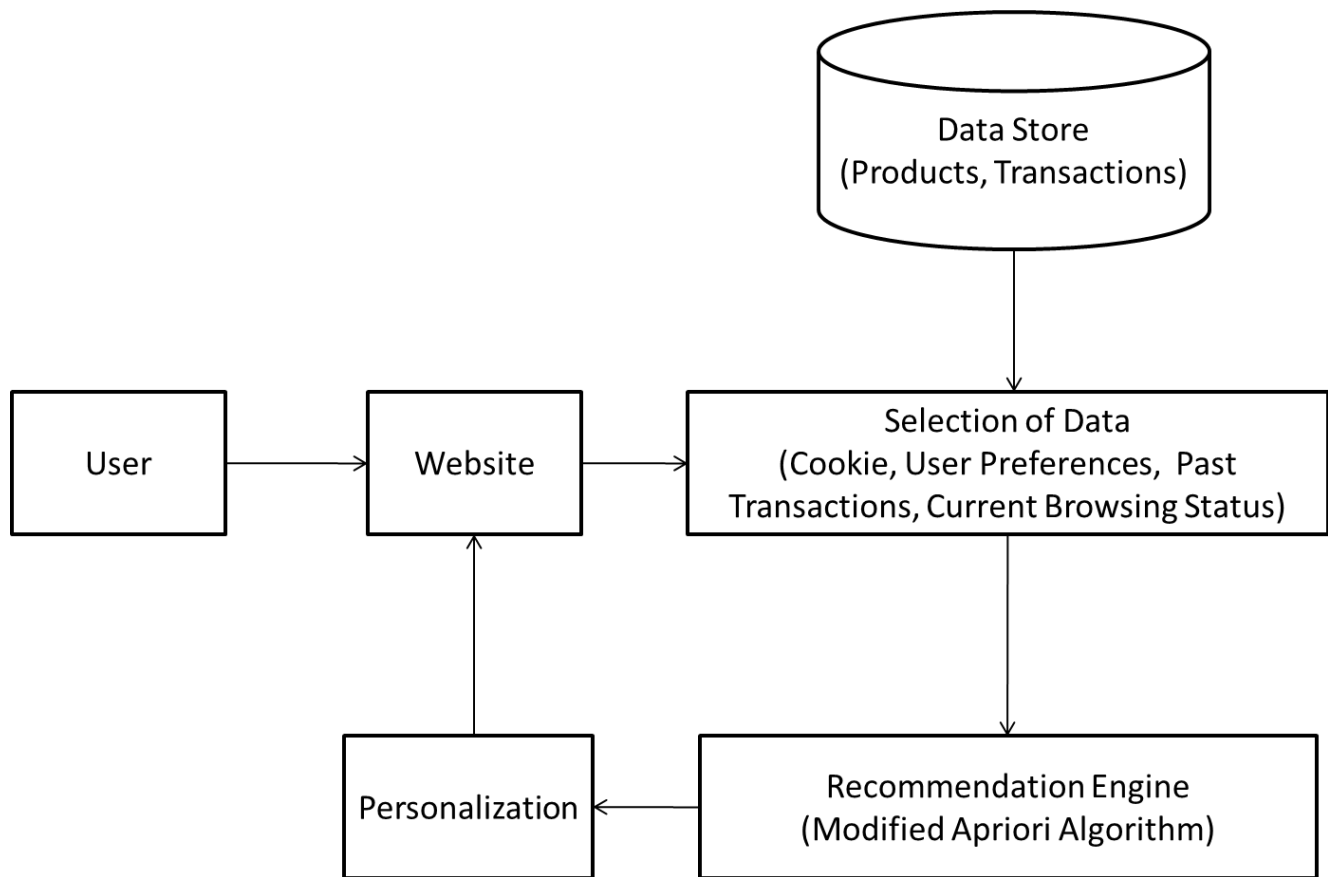


Fig. 1: Proposed System

The steps followed within the system can be defined as:

*A. Customer enters his preferences:*

A customer has logged in the system for the first time. So this time there is no data known about customer interests, searches, purchases, etc. So this time he is just shown a list of products which are currently being purchased, i.e., the current market trend. The customer has to enter his preferences and personal details so that the system can now recommend products.

*B. The searches are stored in cookie:*

Cookies are used to locally store the data on the user's system to perform efficient acquisition of the data relative to the user. As the customer searches the products, it shows his interest in the products. So this interest has to be stored on the system, so that similar products can be shown in future. Cookies are used because transferring so much of data over the network will be costly, and then again, data for applying the algorithm will have to be transferred to the local machine from remote server. So it is better to store the data in cookies on the local machine itself. This ultimately will reduce the execution overload on the system.

C. Association rules are mined by applying Apriori algorithm to the transactions.

Once the data is available, this data is used for finding similar transactions and recommending the products to the customers. Apriori algorithm is used for finding the knowledge to recommend the products based on user and product relations. The Apriori algorithm is used to generate a set of rules to relate the user preferences with the product features. First, the customer searches are fetched from cookie. Then these products are searched in the database transactions and other products in the same transactions are extracted. The products that cross the minimum threshold and support are the ones that have highest probability of getting purchased.

D. Personalization of Web Page.

This would act as the recommendation engine. The products that are fetched from that database in the past step are compared with the customer preferences and they are recommended to the customer. These products will be the ones which have the highest relative strength of getting purchased by the user.

#### IV. CONCLUSION

In this paper, we have proposed a model to personalize an e-commerce system based on the user interactions. We have presented the significance of mining using the client side data collection technique cookies and a modified version of Apriori algorithm for recommending products. Our major aim is to customize the behaviour of the user interaction with the system based on his past interactions. The major advantage of our system is that execution load is reduced on the system since we are using cookies which reduce the data transfer between the remote server and local machine. However, use of cookie may be limited by the size of cookie and even due to fact that they may be blocked by some browsers.

Our future efforts are to enhance the recommendation system by collaborating the server side data and client side user details. We can include data sources like server side logs, site maps, and navigational history through data logs and so on.

#### References

1. Miniwatts Marketing Group, World Internet Users and population statistics, (Online) [www.internetworldstats.com](http://www.internetworldstats.com), Accessed: 27.11.2009, 2009.
2. H. Lieberman, Letizia: An agent that assists web browsing, in: "Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence", 1995, pp. 924–929.
3. T. Joachims, D. Freitag, and T. Mitchell. Webwatcher: "A tour guide for the World Wide Web. In the 15th International Conference on Artificial Intelligence", Nagoya, Japan, 1997.
4. Nasraoui, O., Soliman, M., Saka, E., Badia, A., & Germain, R. (2008). "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites", IEEE Transactions on Knowledge and Data Engineering, 20 (2), 202-215.
5. Abdurrahman et al., "Web usage mining for analysing unique behaviour of web users, Proc. International Conference on Electrical Engineering and Informatics", 2007, pp. 356- 359.
6. B. Mobasher, R. Cooley, and J. Srivastava. "Automatic personalization based on web usage mining", Commun. ACM, 43, 142-151, August, 2000.
7. Eirinaki, M. and Vazirgiannis, M., "Web Mining for Web Personalization", ACM Transactions on Internet Technology, Vol. 3, Issue 1, 1-27, Feb2003 Vol.3, No.1, 1-27.
8. M. Koutri, N. Avouris, and S. Daskalaki, "A Survey on Web Usage Mining Techniques for Web-Based Adaptive Hypermedia Systems", in S. Y. Chen and G. D. Magoulas (ed), Adaptable and Adaptive Hypermedia Systems, IRM Press, pp. 125-149, Hershey, 2005
9. Pierrakos, D., Paliouras, G., Papatheodorou, C., and Spyropoulos, C. D., "Web Usage Mining as a Tool for Personalization: A Survey", User Modeling and User- Adapted Interaction, Vol. 13, No. 4, pp. 311-372, Nov. 2003.
10. B. Berendt, "Understanding Web usage at different levels of abstraction: coarsening and visualizing sequences, in Proc. of the Mining Log Data across All Customer TouchPoints", Workshop (WEBKDD'01), San Francisco, CA, August 2001.
11. B. Berendt, "Web usage mining, site semantics, and the support of navigation, in Proc. of the Web Mining for E-Commerce - Challenges and Opportunities", Workshop (WEBKDD'00), Boston, MA, August 2000.