International Journal of Research
Available at https://edupediapublications.org/journals

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

A Survey Study in Recommendation Systems for E-Commerce Websites

Venkatesh Y R
M-tech, Network and Internet Engineering, SJCE
Mysore, Karnataka, India

Dr. M N Jayaram
Dept. of ECE
SJCE, Mysore, Karnataka, India

Abstract— Electronic Commerce is process of doing business through computer networks. A person sitting on his chair in front of a computer can access all the facilities of the Internet to buy or sell the products. Today there is a big variety of different approaches and algorithms of data filtering and recommendation. In this paper we describe the recommendation system related research and then introduces various techniques and approaches used by the recommender system User-based approach, Item based approach, Hybrid recommendation approaches and related research in the recommender system. One way to do this is to target customers with the particular offers most likely to attract them back to the store and to spend more time and money on their next visit. Demographic market segmentation is an approach to segmenting markets. A company divides the larger market into groups based on several defined criteria. Age, gender, marital status, occupation, education, locality and income are among the commonly considered demographics segmentation criteria.

Keywords—Recommended system; Collaborated; Content; User; Item based filtering; Data mining

I. INTRODUCTION

E-Commercial is a transaction of user, over internet. Electronic commerce is trading in products or services using computer networks, such as Internet.

Today, most supermarkets record sales and collect customers' shopping details via a card dedicated to customer who holds customer's personal information (e.g. age, gender, job, income). Data mining helps using this huge amount of data in an efficient way and provides statistical information, thus predicting future customer behavior. One of the most important data mining methods is, also used in this study, association rule mining. The main purpose of this method is to determine correlations among the sales of items using a set of customer transactions on items. Association rule mining is also known market basket analysis. Market basket analysis helps to understand about the sets of items that are likely to be purchase together. In this paper, some questions were explored such as "Which products are commonly purchase together?" Generally it is being sought the dependency between two products X and Y. This information will be gathered by processing given a transaction database which is huge in size. This information will be used for store layout, promotions, discounts, catalog design, etc. These results will be analyzed whether they are related to customers' data (age, gender, income, marital status, locality etc.). According to the surveys made by the companies which use recommendation system, product recommendation is a milestone which helps the company to increase revenue by 15-40%. According to a Forrester study, 15% of the visitors declare to buying recommended products. According to a recent survey, 74% of consumers get annoyed when the websites. They visit feature content, offers, and ads that don't match their interests. The main

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

objective is meant for market segmentation that gives services according to the customer's interest/Profile. The prediction is based upon customer's age, gender, marital status, occupation, education and income.

In the search case, illustrated in Fig 1(a), users actively express their information needs by submitting queries to the search system (engine), and then the system tries to find the items (e.g., texts, images, videos, music) in the collection that best match the queries.

In the recommendation case, illustrated in Fig. 1(b), recommenders intend to provide people with suggestions of products they will appreciate, based upon their past preferences, history of purchase, or demographic information[25,26]

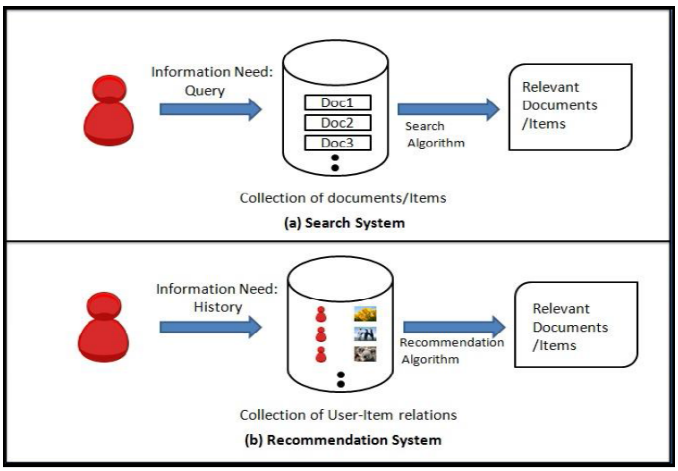

Figure 1 Types of Information Retrieval System

II. EASE OF USE

This part shows all steps (Fig1) from start to end of data mining process, including gathering raw data, normalizing, preparing raw data to be processed, as well as processing data with data mining software. After data is completely processed, results are shown in specific sub-part. Prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III. RELATED WORK

A. *What is Recommender System?*

Recommender systems are used by E-commerce sites to suggest products to their customers and to provide consumers with information to help them decide which products to purchase. The products can be recommended based on the top overall sellers on a site, on the demographics of the consumer, or on an analysis of the past buying behavior of the consumer as a prediction for future buying behavior. [15] The forms of recommendation include suggesting products to the consumer, providing personalized product information, summarizing community opinion, and providing community critiques. Broadly, these recommendation techniques are part of personalization on a site because they help the site adapt itself to each customer. Personalization, to this extent, is one way to realize Pine's ideas on the Web. Mass customization originally referred to the physical modification of products and services to make them fit each consumer's needs *[2]*. More recently, mass customization has evolved to encompass a wide range of methods for customizing the *consumer experience [3]*. The consumer experience includes the physical products, which can be customized in function or in appearance, and the presentation of those products, which can be customized automatically or with help from the consumer. Under this broader definition, recommender systems serve to support a customization of the consumer experience in the presentation of the products sold on a Web site. In a sense, recommender systems enable the creation of a new store personally designed for each consumer. Of course, in the virtual world, all that changes is the selection of products shown to the consumer, not an underlying physical store.[4]

B. *What is Recommendation Algorithms?*

Recommendation algorithms are best known for their use on e-commerce Websites,[4,15] where they use input about a customer's interests to generate a list of recommended items. Many applications use only the items that customers purchase and explicitly rate to

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

represent their interests, but they can also use other attributes, including items viewed, demographic data, subject interests, and favorite artists. At Amazon.com, recommendation algorithms are used to personalize the online store for each customer. The store radically changes based on customer interests, showing programming titles to a software engineer and baby toys to a new mother. *The click-through and conversion rates* — two important measures of Web-based and email advertising effectiveness — vastly exceed those of untargeted content such as banner advertisements and top-seller lists. [5, 15]

C. *What are the challenges in Recommender System?*

E-commerce recommendation algorithms often operate in a challenging environment.[5]

For example:

• A large retailer might have huge amounts of data, tens of millions of customers and millions of distinct catalog items.

• Many applications require the results set to be returned in realtime, in no more than half a second, while still producing high-quality recommendations.

• New customers typically have extremely limited information, based on only a few purchases or product ratings.

• Older customers can have a glut of information, based on thousands of purchases and ratings.

• Customer data is volatile: Each interaction provides valuable customer data, and the algorithm must respond immediately to new information

a) Cold-start[8]

It's difficult to give recommendations to new users as his profile is almost empty and he hasn't rated any items yet so his taste is unknown to the system. This is called the cold start problem. In some recommender systems this problem is solved with survey when creating a profile. Items can also have a cold-start when they are new in the system and haven't been rated before. Both of these problems can be also solved with hybrid approaches.

b) Trust[8]

The voices of people with a short history may not be that relevant as the voices of those who have rich history in their profiles. The issue of trust arises towards evaluations of a certain customer. The problem could be solved by distribution of priorities to the users.

c) *Scalability*[8]

With the growth of numbers of users and items, the system needs more resources for processing information and forming recommendations. Majority of resources is consumed with the purpose of determining users with similar tastes, and goods with similar descriptions. This problem is also solved by the combination of various types of filters and physical improvement of systems. Parts of numerous computations may also be implemented offline in order to accelerate issuance of recommendations online.

d) Sparsity[8]

In online shops that have a huge amount of users and items there are almost always users that have rated just a few items. Using collaborative and other approaches recommender systems generally create neighborhoods of users using their profiles. If a user has evaluated just few items then it's pretty difficult to determine his taste and he/she could be related to the wrong neighborhood. Sparsity is the problem of lack of information. [14]

e) Privacy[8]

Privacy has been the most important problem. In order to receive the most accurate and correct recommendation, the system must acquire the most amount of information possible about the user, including demographic data, and data about the location of a particular user. Naturally, the question of reliability, security and confidentiality of the given information arises. Many online shops offer effective protection of privacy of the users by utilizing specialized algorithms and programs.

IV. TYPES OF RECOMMENDER

Two popular versions of these algorithms are *collaborative filtering* and *cluster models*. Other algorithms — including search-based methods and our own item-to-item collaborative filtering — focus on

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

finding similar items, not similar customers. For each of the user's purchased and rated items, the algorithm attempts to find similar items. It then aggregates the similar items and recommends them.[5,25,26]

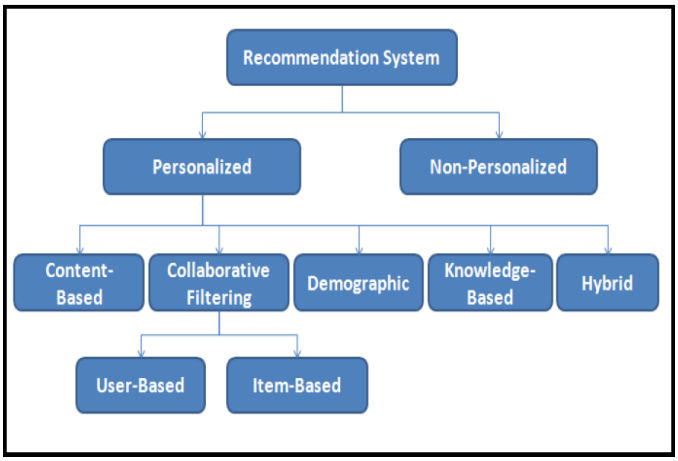

Figure 2 type of Recommed system

Personalized recommendation enables the online introduction insertion, customization or suggestion of data in any format that is relevant to each and every user, based on the user's implicit behavior and tastes and explicitly provided details Non-personalized recommendation system recommend items to consumers based on what other consumers have said about the product in an average. That is, the recommendations are independent of the customer, so all customers gets the same recommendation.

a. Collaborative filtering based recommender systems.

Tapestry [6] is one of the earliest implementations of collaborative filtering based recommender systems. This system relied on the explicit opinions of people from a close-knit community, such as an offices workgroup. However, recommender system for large communities cannot depend on each person knowing the others. Later on several ratings-based automated recommenders systems were developed. The GroupLens research system [7,8] provides a pseudonymous collaborative filtering solution for Usenet news and movies. Ringo [9] and Video Recommender [10] are email and web-based systems that generate recommendation on music and movies respectively. A special issue of Communications of the ACM [11] presents a number of different recommender systems. Although these systems have been successful in the past, their widespread use has exposed some of their limitations such as the problems of sparsity in the data set , problems associated with high dimensionality and so on . Sparsity problem in recommender system has been addressed in [12,13].

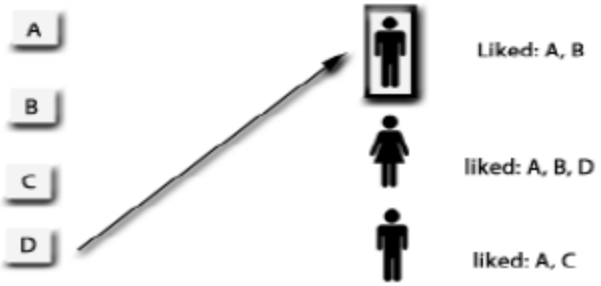

Figure 3 Collaborated filtering

b. User-based collaborative Filtering

For each user, compute correlation with others users. For each item, aggregate the ratings of the users highly correlated with each user. Problem: sparsity of data, bad correlation, easy to attack.

Figure 4 User based

c. Item-based collaborative Filtering

For each item, compute correlation with others items. For each user, aggregate her ratings of the items highly correlated with each item. For items, sparsity of data is less important, less problems with cold start and attacks

Figure 5 Item based

Advantages:

Collaboration Filtering does not need a representation of items in terms of features but it is based only on the judgment of participating user community. Example websites, movies, songs, books, jokes, etc. Scalability of

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

the items database is large because it does not require any human involvement. They can have cross type recommendations for predictions which are different to users and does not require any domain knowledge which saves time. The recommendations can be improved over a period of time.

Disadvantages:

The item cannot be recommended to any user until the item is either rated by another user(s) or correlated with other similar items of the database. In practice, in many e-commerce applications the active users rate only few items in spite of having a large item database which leads to very sparse results. Due to diversity of opinions of large number of participants make this approach expensive and time consuming. It is common those individuals do not agree or disagree with other individuals. It recommends only popular items but cannot recommend unique tastes of a user.

d. Content Based

Content-based filtering methods are based on a description of the item and a profile of the user's preference [18]. These algorithms try to recommend items that are similar to those that a user liked in the past or is examining in the present. In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research. Basically, these methods use an item profile characterizing the item within the system. The system creates a content-based profile of users based on a weighted vector of item features. The weights denote the importance of each feature to the user and can be computed from individually rated content vectors using a variety of techniques. Simple approaches use the average values of the rated item vector while other sophisticated methods use machine learning techniques such as Bayesian Classifiers, cluster analysis, decision trees, and artificial neural networks in order to estimate the probability that the user is going to like the item [19].

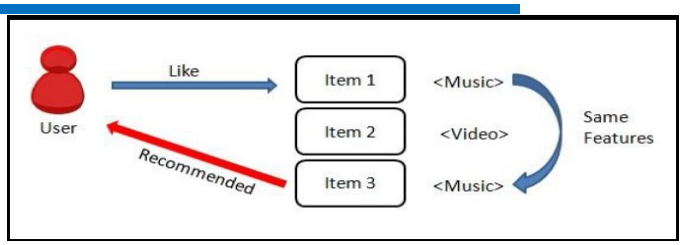

Figure 6 Content based

e. Demographic filtering

The demographic recommendation systems aims to classify the user based upon personal attributes and then give suggestions based on demographic classes. The users' responses are compared with a collection of manually collected user stereotypes. The representation of demographic information in a user model can vary greatly. Demographic techniques form "*people-to-people*" association like collaborative ones, but use different data. The major advantage of demographic technique is that it may not need a history of user ratings of the type as required by collaborative and content-based techniques.

V. DATA MINING PROCESSES

Data mining is defined as a process of discovering hidden valuable knowledge by analyzing large amounts of data, which is stored in databases or data warehouse, using various data mining techniques such as machine learning, artificial intelligence (AI) and statistical [28].

Data understanding: Review the data that you have, documented it; identify data management and data quality issues.

Data preparation: Get your data ready to use for modeling

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

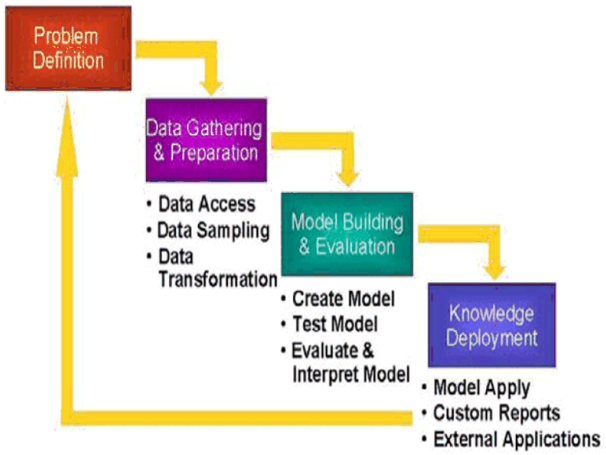

Figure 7. Data mining Process

Modelling: Use mathematical techniques to identify patterns within your data.

Evaluation: Review the patterns you have discovered and assess their potential for business use.

Deployment: Put your discoveries to work in everyday business.

VI. APRIORI ALGORITHM

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

Apriori is designed to operate on database containing transactions (for example, collections of items bought by customers, or details of a website frequentation).

Frequent Item sets: All the sets which contain the tem with the minimum support (denoted by L_i for i^{th} item set). Apriori Property: Any subset of frequent item set must be frequent. Join Operation: To find L_{k+1}, a set of candidate k-item sets is generated by joining L_k with itself. Is explained in Fig 4.

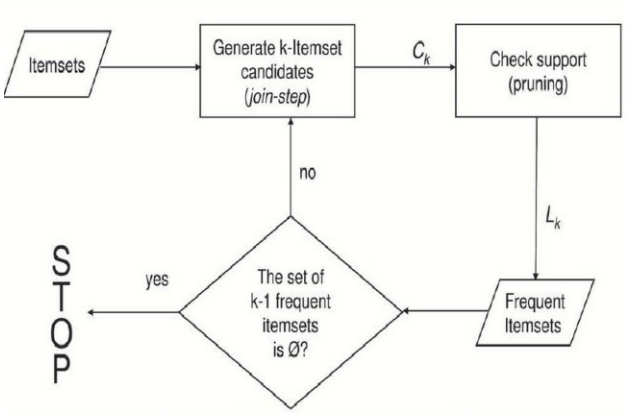

Figure 8: Algorithm flowchart

The Apriori Algorithm(Fig 3) : Pseudo Code

1.Join Step: C_k is generated by joining L_{k-1} with itself

2.Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset

3.Pseudo-code :C_k : Candidate itemset of size k

L_k : frequent itemset of size k

4.L_1 = {frequent items};

for (k = 1; L_k !=∅; k++) **do begin**

C_{k+1} = candidates generated from L_k;

for each transaction t in database do

increment the count of all candidates in C_{k+1} that are contained in t

L_{k+1} = candidates in C_{k+1} with min_support

end

return $\cup_k L_k$;

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

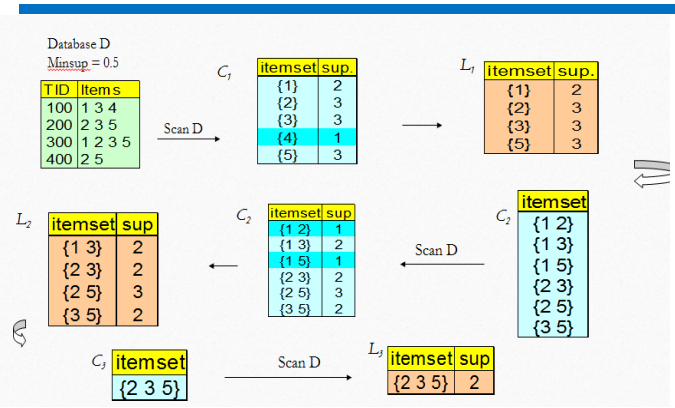

Figure 9 :Algorithm example

Conclusion:

In common, today's commerce must continuously struggle to generate the next best thing that consumers need since consumers endure to wish their products, services etc. to constantly be better, faster, and cheaper. In this world of novel technology, businesses want to provide accommodations to the novel types of consumer requirements and developments because it will prove to be dynamic to their business' success and survival. The system "An integrated package for market segmentation in E Commerce" is developed and tested successfully and satisfies all the requirement of the client.

The goals that have been achieved by the developed system are: Maximize the Business Profit. Increase the customer satisfaction by not giving unrelated recommendation. The system gives services according to the customer interest/profile. It reduces the recommender system challenges like sparsity, scalability and cold start problem. As we use only last one month transaction for recommendation it helps to recommend product for change in customers' interest.

REFERENCES

1. Dr. Sankar Rajagopal,*" Customer Data Clustering Using Data Mining Technique"*, International Journal of Database Management Systems (IJDMS) Vol.3, No.4, November 2011.
2. Pine II, B.J. 1993. Mass Customization. Harvard Business School Press. Boston, Massachusetts.
3. Pine II, B.J. and Gilmore, J.H. 1999. The Experience Economy. Harvard Business School Press. Boston, Massachusetts.
4. J. Schafer, J. Konstan, and J. Riedl. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5(1-2):115-153, 2001.
5. Greg Linden, Brent Smith, and Jeremy York , *"Amazon.com Recommendations" February 2003 , pp. 76-79.*
6. Dr.K.Duraiswamy & B.Jayanthi, "*A Novel preprocessing Algorithm for Frequent Pattern Mining in Multidatasets"*, International Journal of Data Engineering (IJDE), Volume (2) : Issue (3) : 2011.
7. Priyanka Asthana,Anju Singh, Diwakar Singh, " *A Survey on Association Rule Mining Using Apriori Based Algorithm and Hash Based Methods"*, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013.
8. *Cover, T., and Hart, P., Nearest neighbor pattern classification. Information Theory, IEEE Transactionn on, 13910:21-27,1967*
9. Shardanand, U., and Maes, P. (1995). Social Information Filtering: Algorithms for Automating 'Words of Mouth'. In Proceeding of CHI '95. Denver, CO.
10. Hill, W., Stead , L., Rosenstein , M., and Furnas, G. (1995). Recommending and Evaluating Choices in a Virtual Community of Use. In Proceedings of CHI '95.
11. Resnick, P., and Varian, H. R. (1997). Recommender Systems. Special issue of communications of the ACM. 40(3).
12. Sarwar, B. , M., Konstan , J. A. , Borchers, A. , Herlockers, J. , Miller, B. , and Riedl , J. (1998). Using Filtering Agents To Improve Prediction Quality in the GroupLens Research Collaborative Filtering System In Proceedings of CSCW '98, Seattle, WA.
13. Jeetesh Kumar Jain, Nirupama Tiwari, Manoj Ramaiya, *"A Survey: On Association Rule Mining"*, International Journal of Engineering Research and Applications (IJERA) www.ijera.com Vol. 3, Issue 1, January -February 2013.
14. T. Karthikeyan and N. Ravikumar,*"A Survey on Association Rule Mining"*, International Journal of Advanced Research in Computer and Communication Engineering,Vol. 3, Issue 1, January 2014.
15. Suyash R, "Recommendation system in E-commerce webside", International Compuetr science research Vol 4.
16. Jochen Hipp,Ulrich G¨untzer,Gholamreza Nakhaeizadeh," *Algorithms for Association Rule Mining – A General Survey and Comparison"* ,ACM SIGKDD,Vol 3,Issue 1,July 2000.
17. Masoumeh Mohammadnezhad1 and Mehregan Mahdavi, "*An Effective Model For Improving The Quality Of Recommender Systems In Mobile E-Tourism",* International

Journal of Computer Science & Information Technology (IJCSIT) Vol 4, No 1, Feb 2012.

18. *Kyoung-jae Kim, Hyunchul Ahn, and Sangwon Jeong,"Context-aware Recommender Systems using Data Mining Technique"s, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering Vol:4, No:4, 2010.*
19. *Pronk, V., Verhaegh, W., Proidl, A., and Tiemann, M., Incorporating user control into recommender systems based on naive bayesian classification. In RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems, pages 73–80, 2007*
20. V.L. Miguéis , A.S. Camanho, João Falcão e Cunha, *"Customer data mining for lifestyle segmentation"*, Elsevier,2012.
21. David Stern,Ralf Herbrich,Thore Graepel, *"Matchbox: Large Scale Online Bayesian Recommendations*", International World Wide Web Conference Committee (IW3C2), April 20–24, 2009.
22. Alípio Jorge, Mário Amado Alves, Paulo Azevedo,"*RECOMMENDATION WITH ASSOCIATION RULES:A WEB MINING APPLICATION",* Department of de Informatics, University do Minho, Portugal.
23. Narimel Bendakir and Esma Aımeur, "*Using Association Rules for Course Recommendation"*, American Association for Artificial Intelligence (www.aaai.org), 2006.
24. Hasan Ziafat *, Majid Shakeri, "*Using Data Mining Techniques in Customer Segmentation"*, Hasan Ziafat Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 9(Version 3), September 2014.
25. Rakesh Agrawal, Ramakrishnan Srikant, "*Fast Algorithms for Mining Association Rules"*, Proceedings of the 20th VLDB Conference Santiago, Chile, 1994.
26. Mukta kohar,Chhavi Rana, "Survey paper on recommendation system"IJCSIT, Vol. 3 (2) , 2012,3460-3462
27. A survey on Recommendation system, IJIRAE,2349-2163
28. MS. Ashwini M, Ms Kavya Ravishankar; "A integrated package for market segmentation in E-commerce "; International journal of research; Vol 3 issues 9.