

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES ENHANCEMENT IN E-COMMERCE BASED RECOMMENDATION SYSTEM OF DATA MINING

Arvind Singh^{*1}, Prof. Manish Rai² & Prof. (Dr.) Mohit Gangwar³

^{*1}M.Tech. Research Scholar, CSED, RKDFCE, Bhopal

²RKDFCE, Bhopal

³BERI, Bhopal

ABSTRACT

the current expansion in Internet technology is extremely functional for the increase of enterprise. Almost every organization is on the internet and many of them have started their business through e-commerce. Organizations collect information about customers and their business transactions which are very beneficiary for their business growth. Data mining is an arrangement of procedures used to remove unknown pieces of data from the vast database repository. There are various DMT existing to remove expensive and functional information for enterprise. Data mining techniques help e-commerce business in many ways. The paper investigates manipulation of the e-commerce in markets positioned businesses face competitor by the web-established entrant in targeted option this paper, here proposed a technique which is based on recommendation system for the discounted items for the different procedures. The sale established if product sells rarely at a particular time which improves the overall system.

Keywords: *Data Mining Technique, Recommendation System, Collaborative Filtering*

I. INTRODUCTION

Data mining is exploiting to divide demonstrable and in advance unintelligible information from data. DM is the process which gives an initiative to illustrate in thought of customers as of high convenience of oversize compute of information and require changing above such information into important data. So, various communities use the time KDD for DM. Knowledge extraction or revelation is done in seven sequential strides utilized as a part of data mining: [1].

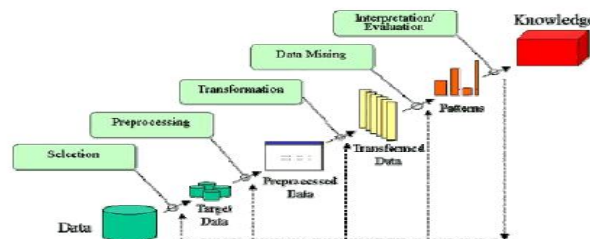


Fig. 1. Data mining Process

II. RECOMMENDATION SYSTEM (RS)

Recommender systems practice techniques and methodologies from any other neighboring areas-which include Human computer interplay (HCI) or Information recovery .though, limit of this system experience in their middle an algorithm that can be identify as a chosen illustration of a DM advance. The manner of data mining includes 3 steps, completed in succession:

- Data Preprocessing,
- Data Analysis and
- Result Interpretation.

Contain you still amused how the "People you may know" feature on Face book or LinkedIn? This feature suggest a list of people whom you capacity know, who are like to you based on your friends, friends of friends in your friend circle, current location or may be past location, skill sets, groups, liked pages, and so on. These recommendations are exact to you and vary from consumer to client. The initial RS was produced by Goldberg, Nichols, and Oki&Terry in 1992. A recommendation system is an approach to the issue like to provide suitable things to the customer despite of searching lots of items. Although People's tastes vary from one to another but they also follows some pattern. RS are software device and method that give proposal foundation on the character go through to find out new necessary contented for them like constructive yield on e-commerce sites like amazon.in, videos on YouTube, posts on the wall of the social media like Facebook, News recommendation on online news websites automatically. RS identify suggestions therefore to the consumers by consider preceding browsing history, the reaction allocate to the goods and unusual user's manners. RS more often than not deliver various proposals in one of the given procedures (Figure. 2) [2]

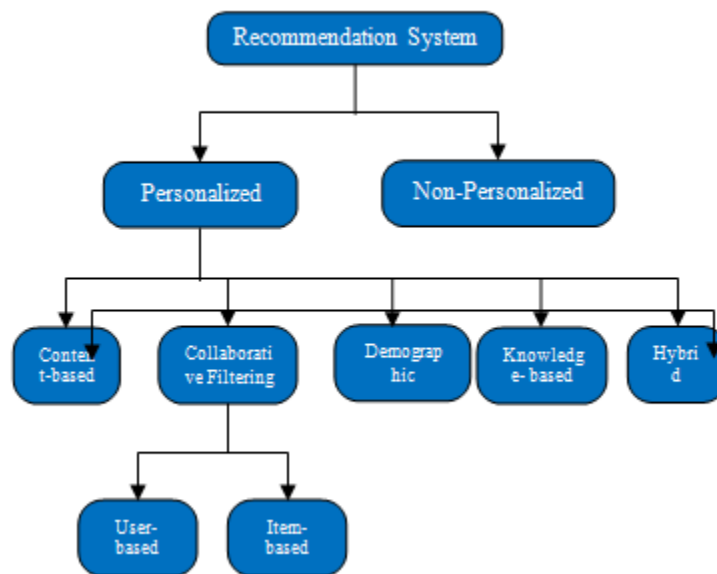


Fig.2. Types of Recommendation System

III. COLLABORATIVE FILTERING (CF)

CF is the manner of filtering records or patterns the use of strategies which involves collaboration amongst various agents, viewpoints, information sources, and so on. Applications of CF include very big statistics units. Collaborative filtering techniques have been carried out to many exclusive varieties of data including: financial data, consisting of financial provider establishments that incorporate many monetary foundation; monitor and sense information, comprehensive of in stone examination, environmental sense more colossal area or a combine of sensor; or in electronic commerce and internet programs wherein the focus is on user data, and so on. CF be able to be use for manufacture automatic prediction around the pastime of a user by collect preference or taste information from many user by using association.

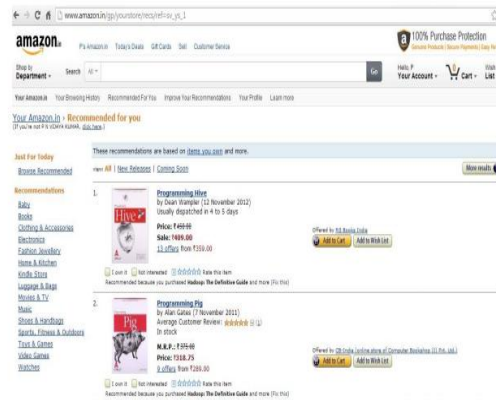


Fig.3 Amazon recommends products to customers by collaborative system

CF approach expect that if a man X has an indistinguishable sentiment from a man Y on an issue, X is extra expected to have Y's view on a unlike concern 'a' than to have the estimation on 'a' of a individual selected arbitrarily. For example, a CF recommendation system for processor taste have to create calculation about which processor product a consumer have to like known a limited register of that user tastes. characteristic workflow of a CF device is as under:

- A user articulates his or her alternative by evaluation substance (e.g. articles, videos or books) of the system. These ratings may be regarded as an approximate illustration of the user's interest within the corresponding area.
- The machine matches those users' rankings against other users and finds the human beings with most similar tastes.
- With similar users, the gadget recommends items that the similar users have rated relatively but no longer but being rated by means of this user.

A. Types of collaborative filtering

- Memory-based: these devices create use of customer evaluation data to subtract comparison between consumer or substance. This is used for making recommendations. This was the earlier mechanism and is used in many commercial systems. It is easy to implement and is effective. Regular cases of this system are neighborhood based absolutely CF and object-based based/person-based N recommendations. The Representative strategies are Neighbor- based CF (item-based / user -based CF algorithms with Pearson/vector cosine correlation) and Item-based/ user - based top-N recommendations.

Advantages:

- simple completion
- New data can be additional simply and incrementally
- require not think the satisfied substance of the substance being optional
- extent well with co-rated matter.

Shortcoming:

- Are reliant on individual ranking
- concert reduce when data are thin
- Cannot suggest for new consumer and substance
- Have unnatural scalability for substantial datasets.

b) Model-based:

Representation be residential with DM, machine learning algorithms to stumble on example support on instruction information. These are used to make predictions for actual data. There are many model -based CF algorithms. This

technique has a more holistic aim to discover latent factors that specify found ratings. These incorporate Bayesian systems, clustering models, inactive semantic models, for example, singular value decomposition, probabilistic inert semantic analysis, Multiple increasing issue, Latent Dirichlet portion and markov option procedure based model.

Advantages:

- Higher address the sparsity, scalability and other issues
- Improve prediction overall performance
- Deliver an intuitive reason for recommendations.

Shortcomings:

- Expensive model –constructing
- Have trade -off among prediction overall performance and scalability
- Lose beneficial records for dimensionality reduction strategies.

c) Hybrid recommenders:

A number of applications integrate the memory-based and the model -based CF algorithms. These overcome over the restrictions of local CF methods. It improves the prediction performance. Significantly, it defeats the CF issues which incorporates sparsity and loss of data. Usually maximum of the commercial RS are hybrid, for example, Google information RS.

Advantages:

- Overcome limitations of CF and content based or diverse recommenders
- Improve expectation execution
- Overcome CF issues comprehensive of sparsity and grey sheep.

Shortcomings:

- Have augmented difficulty and expenditure for completion
- Need outside information that frequently not obtainable [3]

IV. LITERATURE SURVEY

Sonali R. Gandhi, et al. [2017] in our work we present a model that combines RS method such as CF with big data technique such as ARM. The major focus of this work is to give a scalable and healthy recommendation system that can give fine truth. In our work, we have proposed to conduct a personalized movie recommendation by considering user's past behavior [4].

Jingjing Cao et al. [2017] in this work, we initially build users medium as a substitute of customary user-item template, and use such medium to get the nearby neighbor consumer whose partiality is in agreement with goal consumer. Then a list of recommended substance is rank in arrange to choose top-N substance for proposal. Base on the future structure, we consider that this advance can improve the concert of recommender system with the help of sentimental features [5].

Ms. Shakila Shaikh et al. [2017] This paper also emphasizes the need for semantics in current recommendation system to recommend products accurately. This also describes various limitations that are present in the current recommendation methods and suggests possible solutions that can improve current recommendation system used in e-commerce websites. [6].

Luciano Rodrigo Ferretto et al. [2017] this paper aims to identify and analyze recommender systems developed for the health area, available in mobile applications. Therefore, literature was reviewed systematically from ACM, IEEE, Springer and Science Direct databases. 1006 studies were found, eight of which met the eligibility criteria.

From the chosen studies, only one was not applied to the areas of nutrition and physical activity. The recommender strategies embraced were CF and Content-based Filtering. [7].

Hao Li, et al. [2016] in this paper, our propose a multi-stream SGD (MSGD) approach, for which the refresh procedure is hypothetically merged. Scheduled with the intention of source, we propose calculate Unified Device Architecture, parallelization MSGD precede. CUMSGD can improve elevated parallelism and flexibility on explicit dispensation Units. On Tesla K20m and K40c GPUs, the trial comes about demonstrate that CUMSGD outflanks earlier works that quickened MF on shared memory frameworks, e.g., DSGD, FPSGD, Hogwild!, and CCD++. For extensive scale CF issues, we propose different GPUs (multi-GPU) CUMSGD (MCUMSGD). The untried outcome shows that MCUMSGD can recover MSGD concert extra. With a K20m GPU card, CUMSGD can be 5-10 times as fast compared with the state-of-the-art approaches on shared memory platform [8].

Anand Kishor Pandey et al. [2016] in this paper, recommendation for new user in system has been focused. For such system, demographic trend by finding similarity between old user and new user has been followed. The proposed work is based on movie recommendation and has been implemented using MovieLens Dataset [9].

Caifeng Zou et al. [2015] in this paper, the system is divided into the offline part and the online part. In the offline element, the formal context and the concept lattice are made from the transaction database, and the association rules based totally on concept lattice are extracted and saved in the rule of rule library. The new added data are used to replace the idea lattice and the rule library frequently [10].

Faryal Ali, et al. [2015] this work illustrates separate recommendation systems that discover connection among consumer relations on community system sites and ideal substance. The intended Web support recommendation system gathers and change data as of communal system situate concern satisfied and mutual filter method. For instance, to recommend books, the system treats data from Facebook. The experimental results show that these relations improve the value of future suggestion [11].

V. PROPOSED WORK

Data Pre-processing is mainly useful for converting raw data into useful way for making it more reliable for users. It contains various steps in pre-processing like data cleaning which performs the cleaning over the web data. For pattern discovery in recommendation system, cleaned data is mainly used. In the proposed work, there are various tasks performed over the data like the file is loaded for user and the perform data acquisition. There are many details available in log of client, server, and agent for further processing.

For performing Recommendation System, there are two types of user such as registered and unregistered users. Then we perform the counting of all products on the basis of sessions and form the three lists for different types of users. Wish list, purchased list and not purchased list are form for a particular session. Then we offered the discount on each product separately according to the condition for individual user. The sale on the items is also scheduled for the rare time of sale which increases the user attraction for that particular product and increases the selling of the product for that time.

Proposed Algorithm:

1. For every user, load the web log details in the form of file
2. Perform data mining techniques on that data like data acquisition, pre-processing and then cleaning
3. Then this data applied in the recommender system
4. At each session, divide the users into registered and unregistered users
5. Create three lists like wishlist (WL), purchased list (PL) and not purchased list (NPL)
6. Count the products for all users at each session
7. Generate these lists by estimating the frequency of visiting the products
8. For registered users, the frequency of the visiting is 1 and for unregistered users, the frequency should be 3 or more

9. Create new lists with providing 10% discount on each items
10. Now plan sale on the products which sale rarely on particular time
11. End

VI. RESULT ANALYSIS

In the figure below, there is a menu screen which has three options like load database, propose paper and exit button. This menu is used for RS. For database loading, first option is selected.

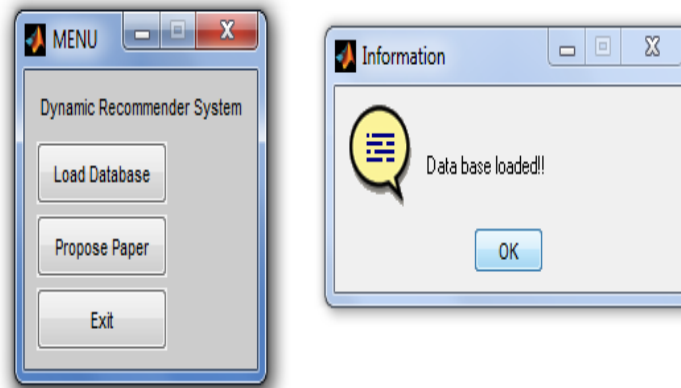


Fig. 4 Interface for loading Database

Base Result

```

Final recommendation list
-----
 2      5      6      7      8
11     12     13     14     15
16     17     18     19     20
21     22     23     24     25
26     27     28     29     30
31     33     35     36     37
40     41     42     43     44
46     47     48     50     9
38

Recall  0.869565
Precision  0.714286
Accuracy  0.780000
    
```

Proposed Result

Now there are multiple browsers used by the user which is shown in the figure below. It shows that apple safari is used by the maximum users. Pie chart is used to demonstrate the results.

Final recommendation list

2	5	6	7	8
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30
31	33	35	36	37
40	41	42	43	44
46	47	48	50	9
38				

Discounted Wish List

Product Id	Actual Price	Discounted Price
1	500.000	450.000
4	65.000	58.500
6	234.000	210.600
20	1000.000	900.000
37	200.000	180.000
43	657.000	591.300
45	321.000	288.900
46	123.000	110.700
50	300.000	270.000

Discounted Product Near Expiry

Product Id	Actual Price	Discounted Price
1	500.000	450.000
4	65.000	58.500
5	198.000	178.200
39	560.000	504.000
40	670.000	603.000
44	876.000	788.400
45	321.000	288.900
47	234.000	210.600
48	432.000	388.800
50	300.000	270.000

Country Based Recommendation List

Product Id	Country Name
3	Brazil
7	Brazil
13	Brazil
17	Brazil
18	Brazil
30	Brazil
33	Brazil
42	Brazil
5	Canada
25	Colombia
27	Egypt
2	India
9	India
14	India

41	India
44	India
47	India
48	India
49	India
8	Israel
19	Malaysia
28	Malaysia
34	Portugal
38	South Africa
31	Thailand
10	United States
11	United States
12	United States
15	United States
16	United States

Schedule Sale for rarely available products

Schedule Time is 13 hour

Product Id	Product Price
7	550.000
21	239.000
22	342.000
31	987.000
35	432.000

Recall 0.892857
Precision 0.892857
Accuracy 0.880000

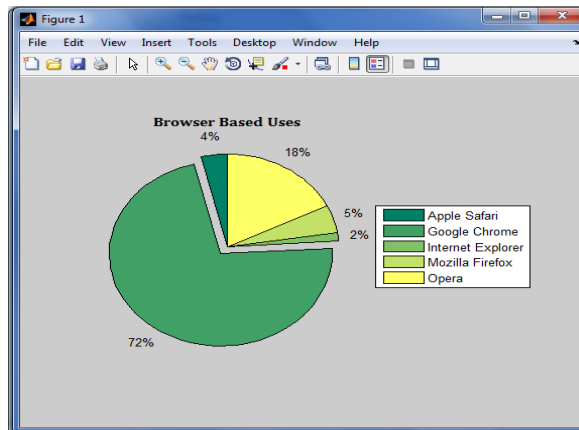


Fig.5.Browsers Used

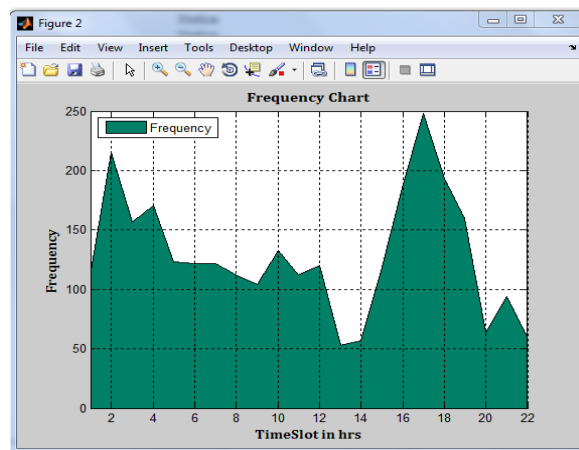


Fig.6. Frequency of page for various time period

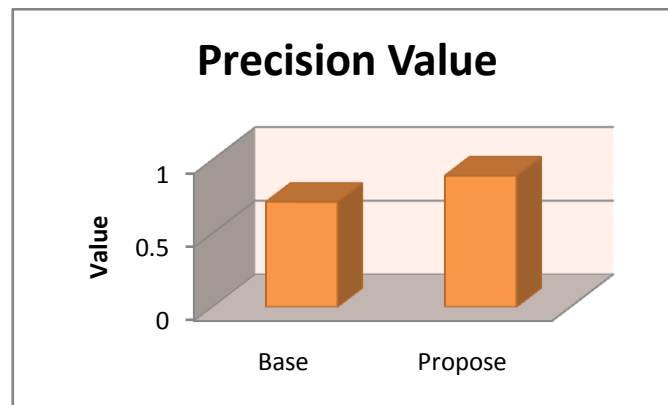


Fig.8. Precision Value

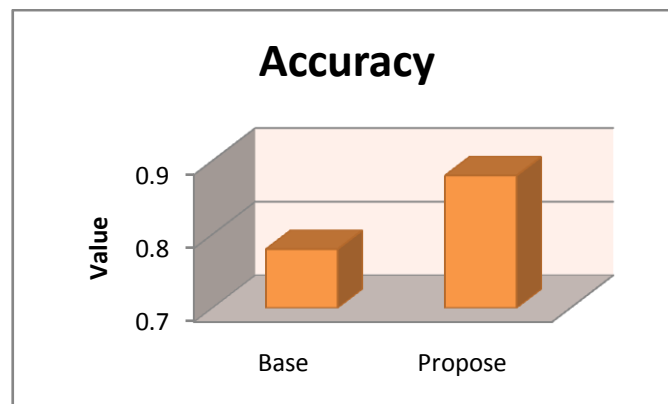


Fig.9. Accuracy Value

VII. CONCLUSION

The term Data Mining refers to the discovery of appropriate and valuable information from large databases. Data mining or KDD is the non-trivial origin of concealed, earlier unidentified and greatly precious information from the enormous quantity of databases. The proposed technique is concerned with the study and analysis of collaborative filtering to improve the efficiency and usefulness of the solution for recommending items utilizing CF system. The developed system is predictable to provide precious support to the customers in decision making at the time of choosing products at the E-commerce sites. The system recommends produced for the customer's presentation is discovered in paper. The system produced is of hugely consequence as it gives user's a possibility to obtain the possessions that they are attracted in. The money off and the factors that affect the user's ideas thus lead to an improved system. This paper gives a new proposed system with the new approaches stand on customer's apprehension in the field of e-commerce and increase the selling of the products on rare time.

REFERENCES

1. Aarti Sharma, Rahul Sharma, Vivek Kr. Sharma, Vishal Shrivatava "Application of Data Mining – A Survey Paper" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2023-2025.
2. Debashis Das, Laxman Sahoo and Sujoy Datta "A Survey on Recommendation System" International Journal of Computer Applications (0975 – 8887) Volume 160 – No 7, February 2017.

3. P. N. Vijaya Kumar, Dr. V. Raghunatha Reddy “A Survey on Recommender Systems (RSS) and Its Applications” *International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 8, August 2014*
4. Sonali R. Gandhi, Prof. Jaydeep Gheewala “A Survey on Recommendation System with Collaborative Filtering using Big Data” *International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2017)*.
5. Jingjing Cao, Wenfeng Li “Sentimental Feature based Collaborative Filtering Recommendation” 978-1-5090-3015-6/17/\$31.00 ©2017 IEEE.
6. Ms. Shakila Shaikh, Dr. Sheetal Rathi, Asst Prof. Prachi Janrao “Recommendation system in E-commerce websites: A Graph Based Approach” 2017 IEEE 7th International Advance Computing Conference.
7. Luciano Rodrigo Ferretto, Cristiano Roberto Cervi, Ana Carolina Bertolotti de Marchi “Recommender Systems in Mobile Apps for Health A Systematic Review” 2017 IEEE.
8. Hao Li, Kenli Li, Senior Member, IEEE, Jiyao An, Member, IEEE, Keqin Li, Fellow, IEEE “MSGD: A Novel Matrix Factorization Approach for Large-scale Collaborative Filtering Recommender Systems on GPUs” 1045-9219 (c) 2016 IEEE.
9. Anand Kishor Pandey, Dharmveer Singh Rajpoot “Resolving Cold Start problem in recommendation system using demographic approach” 978-1-5090-2684-5/16/\$31.00 ©2016 IEEE.
10. Caifeng Zou, Daqiang Zhang, Jiafu Wan, Mohammad Mehedi Hassan, and Jaime Lloret, “Using Concept Lattice for Personalized Recommendation System Design” 1932-8184 © 2015 IEEE.
11. Faryal Ali, Tauqir Ahmad, Martinez-Enriquez A.M. “Data Mining Based Recommendation System using Social Websites” 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.