iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** I **Month of publication:** January 2018

DOI: http://doi.org/10.22214/ijraset.2018.1453

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 6 Issue I, January 2018- Available at www.ijraset.com

News Recommendation Solution using Big Data Analytics

Priyank Jain[1], Savita Sisodia[2]
[1]*Research Scholar –CSE Truba College of Engineering and Technology*
[2]*Head-ITTruba College of Engineering and Technol*

Abstract: ***With the rising technology and age of explosion of technique, massive amount of data is generating every day within every minute. Explosive amount of data is generating with the contribution of social networking sites, mobile devices, sensors, enterprise and scientific data. Their contribution has increased a vast volume of data in recent years. Massive amount of data is called Big data, the potential of Big data can benefits many organizations, business, public sectors with the contribution of economy in the development of sphere. HDFS is a Hadoop distributed file system, it was developed to support distributed designs of file system. It is based on java and is scalable and reliable for data storage. Low cost hardware's are required to design HDFS with high fault tolerant. Data is replicated in HDFS across clusters. In case of failure of single process, then also computation is continued without stopping process. No restrictions are there on storage of datasets in HDFS. Both structured and schema less data is supported by it.***

Keywords: Hadoop; Recommendation system; News recommendation system; TF-IDF

I. INTRODUCTION

Many technologies based on Hadoop is used including HDFS is used for distributed file system. Data warehouse application using Hadoop server is developed using Hive component. MapReduce is a Hadoop Programming model. Query language in Hadoop is used for Pig, that is similar to SQL languages. Data is uploaded in HDFS using Sqoop and MySql to Hive. Various technologies are developed using Hadoop environment for big data. The framework used in Hadoop is a MapReduce widely used by different companies so that massive amount of data can be processed, solving the issue of large data. For cost-effective environment Hadoop is integrated with Linux for computing. It is vulnerable to attack and is more complicated because of its distributed environment. Big data is a big problem, if it is implemented without any correct measures of security and encryption. Datasets are included in Big data without any identifiable information. Therefore, owners information is important to address and classify.

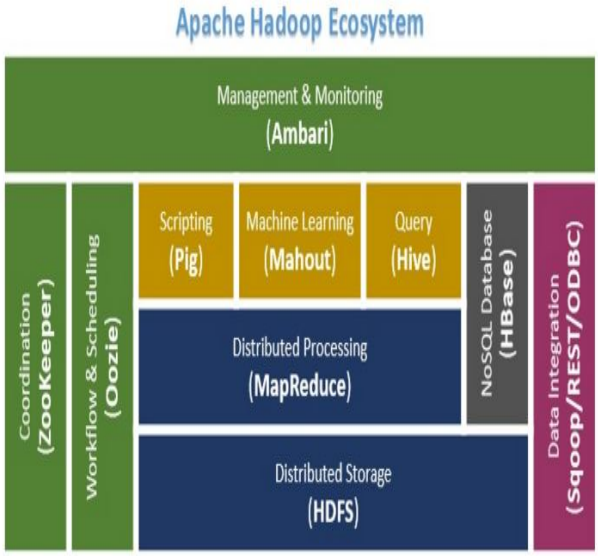


Figure 1. Block representation of Hadoop Ecosystem

A. Limitations of hadoop

1) *Security Concerns:* Hadoop can be challenging in managing complex application. Hadool security model is the classic example, by default it is disable due to complexities.
2) *Vulnerable By Nature:* Its framework is written in Java, which is the widely used programming language. But it is exploited by cybercriminals, which results in leaving this language by experts and getting more efficient and better alternative.
3) *Not Fit for Small Data:* Big data are not suitable to fulfill small data needs, as it is formed for big platform but not exclusive for it.
4) *Potential Stability Issues:* It is the open source platform, which summarizes that it is developed by the contribution of many experts so as to continue working on the platform. Issues like stability issues arises.
5) *Very Limited SQL Support:* In a distributed framework an open-source projects and programming frameworks are combined. It offers limited support for SQL and lacks in its basic functions like grouping, sub queries etc
6) *Inefficient Execution.:* Query optimizers are not considered in HDFS. That is the reason why query optimizer are not considered in HDFS. It also does not have any cost-efficient plan. Since, Hadoop has a large cluster size then the similar database needed.

B. recommendation System

Recommendation system is very essential and gets attention in different fields of recommendation. It provides suggestion on the basis of previous transactions or searched data. So, the research is based on improvement. In any research area it is very important and emerging technique. Recommendation system does not depends on any services or products, it maintains a crucial part in establishing a right recommendation. For example if a user wants to purchase a book, then there is a probability of recommending high rated books. Here, association rule can also be used with it. Therefore, a satisfactory association are linked for the better recommendation for buying associated items. Product recommendation is always not necessary to perform. Other items can also be recommended like movie, newspaper etc. for example if any comedy movie is searched, then all the related results will be displayed. Moreover, the accessed data history is important for the analysis and prediction of data in future by suggesting on the basis of past records.

II. RELATED WORK

Benefit of computerization on different stages and equipments are required in information mining method. Enormous databases are analyzed in minutes by the information preparing apparatuses squares, on multiprocessing framework. Faster process needs faster mechanism for exploring and suggesting clients for different deals. Due to massive amount of data rapid examining makes it more sensible for client. As big the database is, thus, require increased forecast to yield. Bahram Amini et. al. In[1] described about client's profile and method of separating client client's data, this all is constituted in recommender framework. From the large amount of data client's significant and inclination data is separated. All these features of cloud are significant in getting essential information to achieve accurate recommendation. A survey is proposed in this paper for recommendation framework. Personalization can be achieved in various ways. Gediminas Adomavicius et al. In[2] described about improving the framework of recommender and is some what critical to individuals and individual organizations to provide the proposal of customization. They analyze that majority of analyst have exact proposal, and also have the quality of suggestion have can be critical. For example the miscellaneous quality of suggestions will be customarily. More different suggestion are generated in this paper by investigating several things for positioning system Neeraj Raheja et al. In[3] proposed about web mining which is the better way to deal with huge web information and concentrated learning. Including structure, web content and utilization mining. Information can be concentrated using these records and also help in improvement of websites. Minsuk Kahng et al. In[4] described about the inclination of client which can be due to setting of imperative component. To enhance the behavior of suggestion framework enormous specialist are utilized. The reason behind the proposal of single thing inclination is critical component. Ch.Nagini et al. In[5] address about the questions inquired against the record brought by web mining. Without effort client does not get outcome. For the suppression of this problem recommender framework is used. They have stand on the expectation of client by fulfilling there need and opposing the numerous archives.

Table 2.1 Comparative Table

AUTHOR	TITLE	PROPOSED WORK
Bahram Amini	Discovering the impact of knowledge in recommender systems:a comparative study.	Client's profile and method of separating client's data, this all is constituted in recommender framework. From the large amount of data client's significant and inclination data is separated.
Gediminas Adomavicius	Improving Recommendation Diversity Using Ranking-Based Techniques.	Improving the framework of recommender and is some what critical to individuals and individual organizations to provide the proposal of customization.
Neeraj Raheja	Survey on ameliorate data extraction in web mining by Clustering the web log data.	Web mining which is the better way to deal with huge web information and concentrated learning. Including structure, web content and utilization mining.
Minsuk Kahng	Ranking in Context-Aware Recommender Systems.	Described about the inclination of client which can be due to setting of imperative component. To enhance the behavior of suggestion framework enormous specialist are utilized.
Ch.Nagini	An implementation view for news recommendation system	Address about the questions inquired against the record brought by web mining. Without effort client does not get outcome. For the suppression of this problem recommender framework is used.

III. PROBLEM DOMAIN

In the recent years many changes have been analyzed, people used to prefer reading news in computer or laptop or tablet or any other system. Every day new systems (portals) are designed and these portals try to search more quantity of user. Demand of having more business is enhancing. The issue arises at that moment where user fetching the news and gets dissimilar one. Our study focuses on the news recommendation based on need. For example if user is searching political news, then the appeared news should be relevant manner.

IV. METHODOLOGY

Methodology explains about Data mining its techniques and data warehouse. Where, Data mining is used to extract knowledge from large database. Hidden patterns can be find out using the tool of data mining. These hidden patterns are popularly called as Knowledge extraction. Learning step and classification step are the two steps consisting in classification process. By using

classification algorithm data can be obtained and then is formed into training data. The design of a decision tree is like a tree structure consisting of collection of nodes, with some branches. Its structure is like binary tree. Bayesian classifiers are used to check the probability of tuple belonging to the class.

A. Algorithm

1) In our work we are analyzing on news so, D indicates the document which is news.
2) F is the frequency of word occurred in document and w represent the word.
3) In one document F is given ten and in other thirty, then it is the frequency of occurrence of word in document.
4) High frequency does not indicate relevancy in document, but it is directly proportional to the frequency of occurrence.
5) Document frequency is represented by DF, describing the word occur in number of documents.
6) Weight associated with words is calculated, log frequency can be given as:

$$W_{D,w} = \begin{cases} 1+\log 10(Fw,D) & \text{if } Fw,D>0 \\ 0 & \end{cases}$$

7) D→W:
0 → 6, 1 → 5, 2 → 90, 10 → 45, 1000 → 880, etc.
8) Using log intersection matching scores are calculated.
9) High score represent that most of the words are found in document. And null score represents dissimilarity between two documents.

IV. PROPOSED WORK

Recommendation system used in our approach helps to find desired news, filtering is used in this recommendation to provide with best outcome. On the basis of choice of developer algorithm can be designed for recommendation. To achieve the desired outcome collaborative filtering is used in our work approach. Aim of the project is to help accessing the desired news in a convenient manner. TF-IDF algorithm is implemented in the proposed work. On certain MB of datasets analysis is performed to released the recommendation system concept. The java and database technologies are used which is of Oracle.

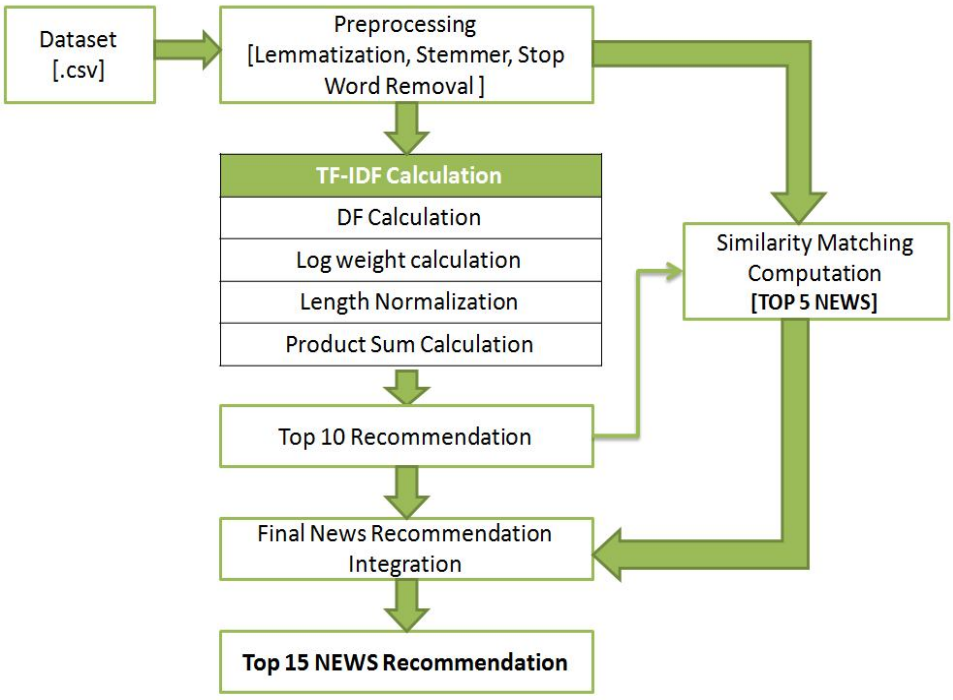


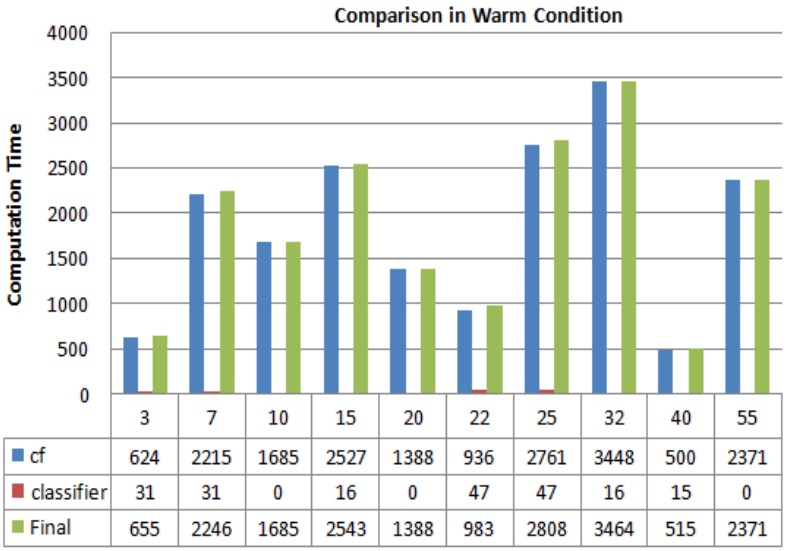
Figure Proposed Architecture

Tool used to collect user data and information to get available hidden patterns is defined as recommendation system. Recommendation can be possible broadly in wide range. In our research work we are dealing with recommendation system based on news. This recommendation can be possible when the users interest is in news. If user wants defence news then it will be difficult because of it availability is tough to find. On the basis of rating, ranking and weight age every news is determined and get popularity. Whose rating is more gets more chance for recommendation. News are collected from all the directions to get high rating and points assigning. News can be of any type it can be any sports news, politics, knowledge news, food, travel or bollywood news.

V. RESULT ANALYSIS

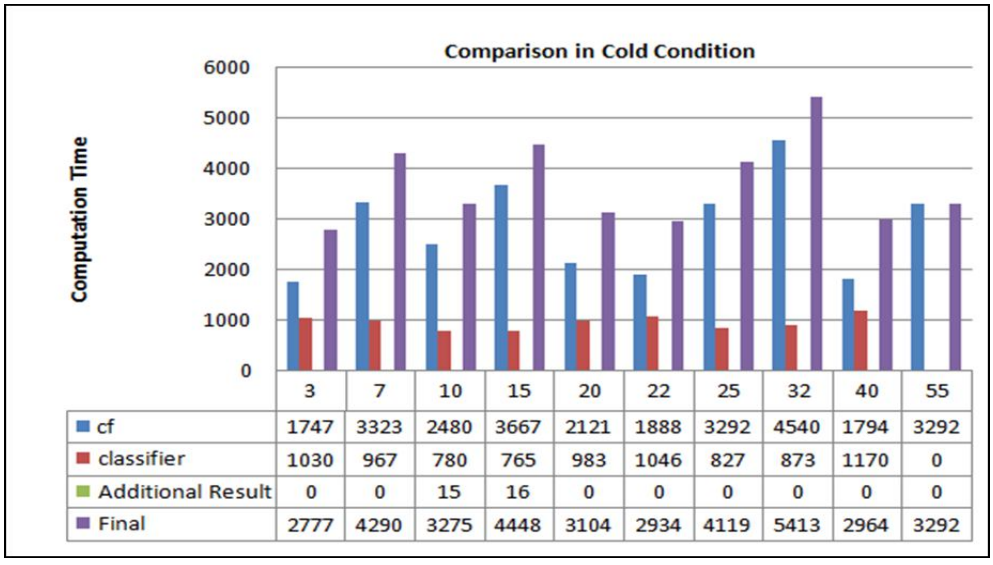
We cannot say any work to be completed till the complete execution is not achieved. In this paper depiction of the proposed arrangement is performed with the existing arrangement. Different clients evaluates proposal arrangement at different prospect. Examination of the result analysis in our work is based on two conditions, which are described below :

A. Cold condition
B. Warm condition



	3	7	10	15	20	22	25	32	40	55
cf	624	2215	1685	2527	1388	936	2761	3448	500	2371
classifier	31	31	0	16	0	47	47	16	15	0
Final	655	2246	1685	2543	1388	983	2808	3464	515	2371

Figure 5.1: Warm condition



	3	7	10	15	20	22	25	32	40	55
cf	1747	3323	2480	3667	2121	1888	3292	4540	1794	3292
classifier	1030	967	780	765	983	1046	827	873	1170	0
Additional Result	0	0	15	16	0	0	0	0	0	0
Final	2777	4290	3275	4448	3104	2934	4119	5413	2964	3292

Figure 5.2: Cold condition

"Impact of Information Security Measures on the Velocity of Big Data Infrastructure" is the base paper and comparison of it is done with our proposed work. Time required is represented in millisecond. Our work proves that less processing time is required in our work in comparison to the existing work.

VI. CONCLUSION

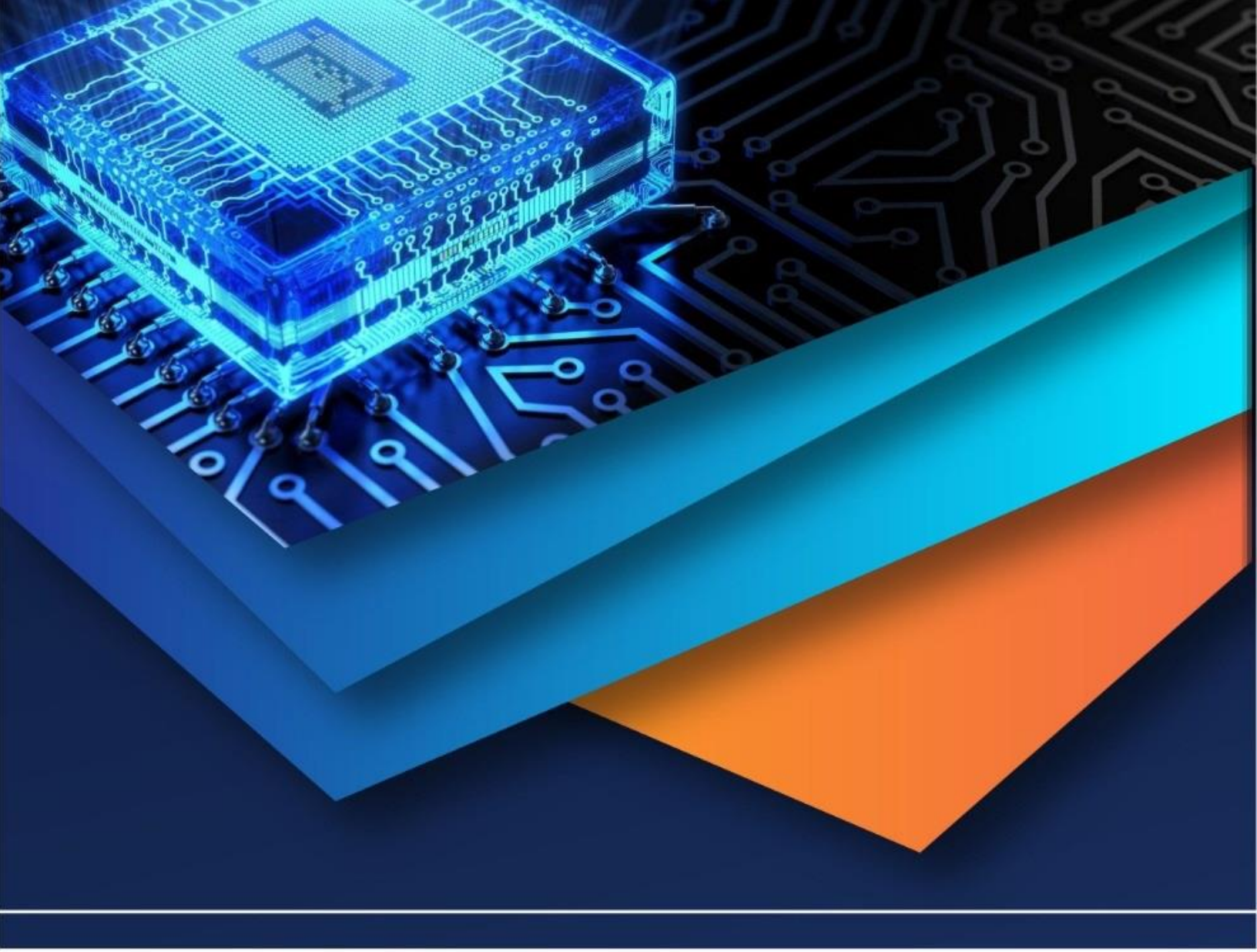
The conclusion says that the piece of work to be performed is completed and is performed well. Genuine executions can be handled by the makers and posses that sooner the approach of News Recommendation will become more effective and efficient. Technique used in our work is generalized technique, thus best and exact possible outcomes can be concluded. While comparing the outcome of our approach with the existing approach then concluded result states that our outcome is better then the existing work. 35% better result in comparison to existing technique is analyzed. Thus the result indicates that have performed some contribution in our work.

VII. FUTURE WORK

Without the help of accuracy percent no research can be possible. Every research has the possibility of future work. Security can be implemented in the architecture with the focus of recommendation. Performance comparison can be performed and the language used in architecture can be changed. To make the architecture simple, easy and better some enhancement can be required to perform.

REFERENCES

[1] Bahram amini, roliana ibrahim, mohd shahizan othman, "Discovering the impact of knowledge in recommender systems:a comparative study", International Journal of Computer Science & Engineering Survey,vol 2,pp-3,2011
[2] Gediminas Adomavicius, Young, Kwon Improving Recommendation Diversity Using Ranking-Based Techniques, pp-1-33.
[3] Neeraj raheja, v.k.katiyar, "survey on ameliorate data extraction in web mining by Clustering the web log data" International Journal of Computer Science Issues" Vol. 11, pp-2, 2014.
[4] Minsuk Kahng, Sangkeun Lee, Sang-goo Lee, Ranking in Context-Aware Recommender Systems,pp-65-66, 2011.
[5] Ch.Nagini, M.Srinivasa Rao, Dr. R.v.krishnaiah, "An implementation view for news recommendation system" international journal of engineering research & technology, vol. 2,pp-701-704,2013.
[6] Michal Kompan, M_aria Bielikov, Content-based News Recommendation,pp-1-12.
[7] Getoor L, Sahami M. Using probabilistic relational models for collaborative filtering[C]. Proc.Workshop Web Usage Analysis and User Profiling (WEBKDD·99). 1999.
[8] J. Bobadilla, F. Ort ega, A. Hernando, A. Gutierrez, "Recommender Syst ems survey", Kn owledge Based System , July 2013, p p.109 -1 32.
[9] D Park, H. Kim, Y. Choi and J. Kim, "A Literature review and classificat ion of recommender syst ems research," Int ernat ional Conference on Expert Systems with Applications, Sept 2012, pp. 10059-10072.
[10] Lipi Shah, Het al Gaudani and Prem Balani, "A Survey on Recommendat ion System",International Journal of Computer Applications (IJCA),March 2016, pp. 43-49.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY