Clustering Based Approach to Overcome Cold Start Problem in Intelligent e-Learning system

Gopal Sakarkar
Department of MCA, G. H. Raisoni College of Engineering, Nagpur,440016,India

Dr. S. P. Deshpande
Associate Professor & Head, P. G. Department of Computer Science and Technology
HVPM, Amravati. (MS) India

Abstract : Online learning and teaching is new methodology adapted by both learners as well as teachers. Recommendation is the recent demanding trend in every online services provider. Along with various e-business and e-commerce service providers, e-learning websites are willing to start and deliver customize and recommend based learning systems. Proper and accurate recommendation is very challenging and demanding research area in 21st century as a numbers of web sites are increasing dramatically every day. But to deliver the essential product to accurate user is basic obligation of good recommendation system. One of the most challenging task for developing intelligent e-learning system is to overcome the cold start problem that occurred for new learners. In this study , data generated from the learners was analyzed using K-Nearest Neighbour , K-means algorithm and Apriori algorithm . To develop a better recommendation systems , we are considering learners past educational data, parental information and his current technical knowledge. Result of data analysis reveals that socio-economic background and educational academic past data plays important role in recommendation system.

Keywords: Cold start problem, Recommendation, KNN, K-Means, Socioeconomics ,Personalization, e-learning .

I. INTRODUCTION

In 2020 , India will become highest literate country in world and having highest young IT (Indian Talent) people to use the computers and mobile apps. To provides constructive and appropriate digital information is new challenging and researching area in web based applications. On the other hand digitalization in studying, learning and teaching is a new trend adapted by new generation of leaner. Providing and recommending latest and appropriate study material to leaner is a Hercules task for e-learning providers.
Socioeconomic plays a vital role in their children's education. As per authors proposal [1] ,the parent education, occupation ,annual income also play important role in the development of learners educational carrier. It is the need of an hour , parental involvement and their role will help their children to have better education[2]. Kevin Majorbanks were focussed on two parent vs single parent households, divorce ,family size , number of sibling etc. The socio-economic status is a combination of parent's education ,occupational status and income level by comparing all aspects of these parameters , result shows that only those learners show good result in educational field those having a good socioeconomic status [3].

For any recommendation system , it is extremely difficult to provide an accurate recommendation to the user on his first attempt use of the system. The cold start issue for new users can lead to new user who decide to stop using the system due to the lack of accuracy in the recommendations received in that first stage[4].

Apart from the recommendation system , cold start problem have been presented in personalized web searching techniques as it has lacks of browsing history or any relevant information regarding use. [5].

Overindulgence of recommending the services are also one of the problems faced by users. They are getting irrelevant recommended suggestions, emails or SMS in mobile . This is due to lack of personal information of learner by service providers. Knowing learners complete profile is an important task for content recommendation in web personalization[6] .
One of the novel architecture for an e-Learning environment using an intelligent personalized context-aware recommendation system was proposed by Lantao Hu et.al. Its uses a Rules Engine to manage a set of rules for each user to achieve personalized recommendation. For recommendation they used Recommendation Algorithm and a Random Walk with Restart (RWR) that algorithm can be used in the diverse object network model and

the recommendations are generated according to the similarity and homophile of the objects .However , authors have neglected the user registration parameters for developing effective recommendation system [7].
Lakshmi Sunil and Dinesh K Saini , designed the recommender system that based on mapping a e-learning content ontology and learner profiles in the system, but again they have not focused on users profile building. Recommendation system has been developed using data mining techniques that use the course content , leaner's style and learner aptitude knowledge to construct a complete new course from existing learning content in a course repository[8].
Similar Profile Recommender System is one of the innovative recommender systems developed by Ye Xu et.al.[9]. It helps recruiters and hiring managers to discover other similar quality talent by pivoting of a model user profile. It models each member profile, by extracting a labelled bags of canonicalized keywords from profile fields such as summary, skills, companies worked at, schools attended, job titles etc. They used cosine similarity to find out similarity score from two members profile.
One of the very popular recommendations method used by researcher is tagging for personalization . It has been recognized that if system has lack of information about tagging then the recommendation is totally based on users fully design profile. Eventhough they proposed new dimension in folksonomy and use the Quadricons algorithm to mine quadratic concept of users ,tags, resource and profile, the conclusion of proposed work is focused on strong build up of users profile only[10].
It has been found that when the learner is new in a e-learning system , the system is enable to extract sufficient information from the learners profile that is required to start a recommendation, which has been popular by the name of Cold Start problem[11].It means that when the new learner in the system does not visit and not rate any learning content, the system does not claim the learners required goals and is unable to filter the starter or new recommendations.
Two important aspects have been mainly focused that increasing the profile length and make the profiling process smooth for the user by limiting the number of ratings[12]. They have inferred that if profile length is small then it affects the accuracy of recommendations and burden of the rating process. It has stronger effects on the perceived quality with recommendation systems. Also Several studies show that the risk in requiring users to rate many items is to annoy them or to have them give up the rating process[13].
Usefulness of user profile plays an vital role in CQA(community questioning and answering)system. Authors find out that the close relationship between user profile information and the quality of their answers underground truth that user information records the users behaviour and histories as a summary [14].

II. LITERATURE REVIEW

One of the major neglecting and critical issue in any recommendation systems is a 'Cold Start Problem'. The cold start is a problem , when any new user enter first time in online system and system doesn't have any past historical information about user. In this situation , it is very difficult to provide a proper recommendations or suggestions to the user. This problem was faced by News recommendation system developed by Lei Li et.al. They observed that traditional user profile can be used for keeping track of articles the user has read so far. However , simply representing these user's profile as a weighted topic distribution cannot effectively capture user's exact reading preference[15].
Authors projected a small amount of efficient methods based on ask-to-rate technique in which the profile of a new user is made by integrating information gained from a quick interview called as ask-to-rate technique . The ask-to-rate technique is the most direct way for obtaining some information about the new user and for learning the user's preferences that will be useful to overcome the problem of cold start[16].

Social networking data can be used to solve the cold start problem[17].To overcome the problem of cold start ,this framework consists of a different language model, whose mixture weights are estimated with a factor graph. The factor graph is used to incorporate prior knowledge and heuristics to identify the most appropriate weights.

One of the solution for this cold start problem was proposed in [18] by learning domain concepts derived from the profiles can form a rich resource for augmenting the limited learner model to overcome the classic 'cold start' problem in user adaptive environments.
According to a study conducted among a sample of American social media users, social media plays an important role as 93% of companies be present social media websites, and about 85% said that a company is required to not only have presence in social media but also interact with customers through social media [19].
Cross-domain user profile modelling is an innovative idea as addressed in [20] . They analyzed user's Facebook profile, and expanded it by linking it to Flicker in order to recommend socially relevant photos. They have find out that Flicker provide photos with more detailed metadata, and at a much higher resolution than

photos on Facebook, but at the same time they provide much less information about the user. Thus the two social networks can be used to complement each other.

Online dating is a rising of new world in social networking community , but it also facing a problem of 'cold start' problem because , finding suitable matching partner for dating , each user needs a person which is comfortable with him/her. Finding such comfortable partner ,they need to match some of the basic characteristics like lifestyle ,career ,education, hobby etc.[21]. User's detailed profile plays an important role in finding suitable dating partner , authors are emphasised on designing a detail user profile ,so that they can recommended with perfectly matching dating partner.

Shehab A. Gamalel-Din [22] suggests that the most effective means for teaching is through one-on-one interactions with learners. The better learning results would be achieved by adapting the e-tutor interaction with its individual learner user. To teach in better way to an individual learner system has to know in depth knowledge about profiling of user ,so that it will suggest an interesting and useful study material to learners.

Developing an effective web personalization system is to build and model dynamic user profiles. A proposed multi-agent based system [23] for building a dynamic user profile that is effectively capable of learning and adapting to user behaviour. On the basis of users browsing behaviour it will be possible to find out short-term and long-term users interests. An Ontology concept was used to predicate user's interests by mapping web pages visited by a them. Experiments demonstrate that developed system is able to effectively model a dynamic user profile that is capable of learning and adapting to user behaviour, as result performance of personalize system is more than non-personalized system.

João C. Prates and Sean S. M. Siqueira proposed information extraction techniques that applied to educational resources to expand the queries done by learners and adding contextual information in the search and finally recovering more appropriate educational resources. As a result it can can be used in an educational environment to improve the search of educational resources. The educational resources made available by this work was again not provided to precise learner because of system lacks a complete information about users profile [24].

Long-term interest profile was used in news recommendation system. It used the semantically annotated news items and the defined ontology-based user profiles help to provided users interested news items [25]. Learning user profiles of Web browsing behaviour can co-operate to personalize search results and identify persons of interest. This profile is generated at run time ,based on the users visited websites .From this auto generated profile , other web systems not get an exact personality of user and hence it may provide irrelevant type of information that will be not useful for user[26].
The ''online shopping reference model'' indicated that gender and age will affect users' intentions to shop online. The survey revealed that 91.9% of the younger group had online shopping experience and as compared with 24.4% of the older group[27]. In our study ,we had find out number of web services provider do not give an importance to age or DOB factor while registration process.

Blogs, micro-blogs, social networks, wikis, forums, and content communities are new ways of communication and information sharing. Eventhough these are very popular web tools , complete user profiling was not available in any one of these services [28].One of the interesting thing is that U.S. companies spend $3.08 billion to advertise on social networking sites in 2011, a 55% rise from the previous year (eMarketer, 2011). If we have such large financial platform to deliver web services to intended users, the web system must have completed information about user , so that they will recommend the suitable and literal web product to user , that's rejection of such services by user will be reduced and such huge amount spent on advertising will get beneficiated.

III. RECOMMENDATION TECHNIQUES

To provide a proper recommendation and personalization as per interest and using basic profile of users is a very challenging and demanding work in current Web 2.0 technology world. A Rule engine is one of the ways to achieve recommendation. It is basically a graph- based collaborative filtering that is used to identify the relationship between various users[7]. Collaborative filtering algorithm is one of the very popular recommendation system algorithm that use the rating concept given by a set of similar users[8], while a Semantic context-aware recommendation is one of the new paradigm used for recommendation purpose. Semantic context is a background topics under which user access activities occur within a given particular unit of time[25].

One of the traditional recommender method is RMSE i.e. Root Mean Squared Error ,which works in a system where users rate items in a certain separate range(scale from one to five)[29].
Web 2.0 tool is used for recommendation systems. Anita and Eugenijus developed Web 2.0 tools that analyses learning activities . They used a novel method of integrating Web 2.0 tools into personalised learning activities according to learners learning styles and developed a prototype of the recommender system[30].
To recommend a medical document according to user profile is an innovative idea proposed by Kleanthi Lakiotaki et.al., they first be labelled documents as either expert or consumer document. A linear transformation method is used to categorise provided utility score for every document [31].

IV.RESEARCH DESIGN AND METHODOLOGY

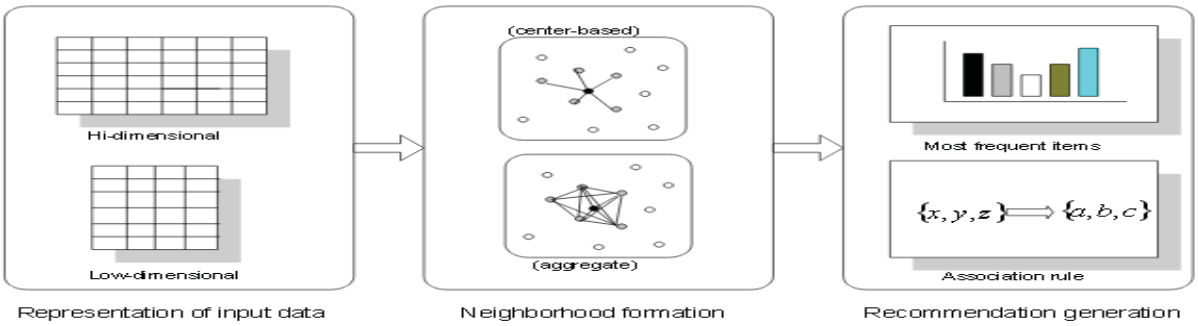
Collaborative filtering is one of the most widely used techniques, which plays a vital role in designing the recommendation systems. The collaborative filtering technique based recommender system may suffer with cold start problem . In this paper recommendation system generates suggestions for user by combining his personal information, educational information and parental information . This is one of the new approach we are proposing to overcome the cold start problem.
In collaborative filtering ,various algorithms are used to make automatic predictions about a leaner's interests by compiling preferences from several co-learners. CF system recommended the educational websites to a target learner based on the opinion of other co-learners present in his group. Once the neighbourhood is formed , system recommended the useful knowledge data set to the leaner.

The proposed IMSAA systems is tested with very widely used standard MovieLen data set[32]. MovieLens data sets were collected by the GroupLens Research Project at the University of Minnesota. This data set consists of 100,000 ratings (1-5) from 943 users on 1682 movies. Each user has rated at least 20 movies.The data was collected through the MovieLens web site (movielens.umn.edu) during the seven-month period from September 19th, 1997 through April 22nd, 1998.
This system employ statistical techniques to find a set of learners as neighbours , that have a history of agreeing with the target learners by either match in socioeconomic level or in knowledge level.

First group created by matching learners socioeconomic information called as G1 , and second group is created by combing learner's basic technical knowledge (by taking online C test) and selected skill set called as G2 . Final group is G1 □ G2 = G3 , that result contained common learners in G1 and G2, which is a final list of co-learners recommendation to a learner.

On the basis of these co-learners groups, our IMSAA system provided a useful technical website links recommendations , even though system does not have any past historical information of login leaner .

The proposed CF-based recommendation system generation divided into three part, 'representation' , 'neighbourhood formation' and ' recommendation generation'. The 'representation' task deals with the scheme used to model list of web links that have already been used by other co-learners and forming a group of co-learners on the basis of socioeconomic status . The 'neighbourhood formation' task focused on the problem of how to identify the other neighbouring learners on the basis of C test result and selected skills set. Finally, 'recommendation generation' task focuses on the problem of finding the top-N recommendation web links from the neighbourhood of learners.



a. Experiment Setup

The collection of data for the experiment was done through a online developed application IMSAA(www.imsaa.in).

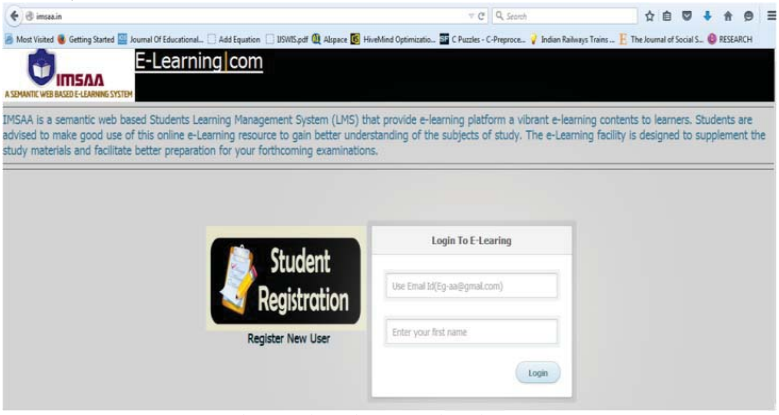


Fig.1 Login and New Registration Page

New learner or already registered learner has to first appear for the online C test, mean while system has classify the new learner in one of the group (A, B, C, D) using K-means algorithm (where K=4) and K-NN algorithm is used for the learner's profile that compute the similarity between the learner and each one of the other co-learners in the system(Say G1).

Fig.2 Online C Test

After attending the 20 questions C test with in 20 minutes , next leaner need to choose at least 3 skills set of his interest out of 15 skills.

Fig3. Selection of Skill set

On the basis of test result and interested skills , system generate second group (Say G2), which is a combination of learner's knowledge level (beginner, fresher, experience and expert) and chosen skills.

Finally ,system has to form the final group of co-learners (G1 G2=G3) for providing matching co-learners as a recommendation.

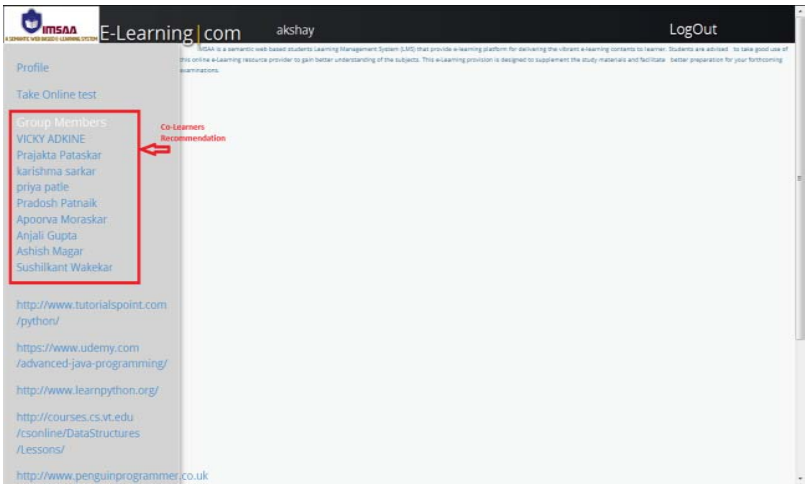


Fig.4 Result of G1 ⊓G2=G3

As our new learner used this system first time and system does not have his any past information , at this stage 'Cold Start Problem' occurred ,i.e. system don't provided proper recommendation of web site links , as leaner not used the system from long period of time.

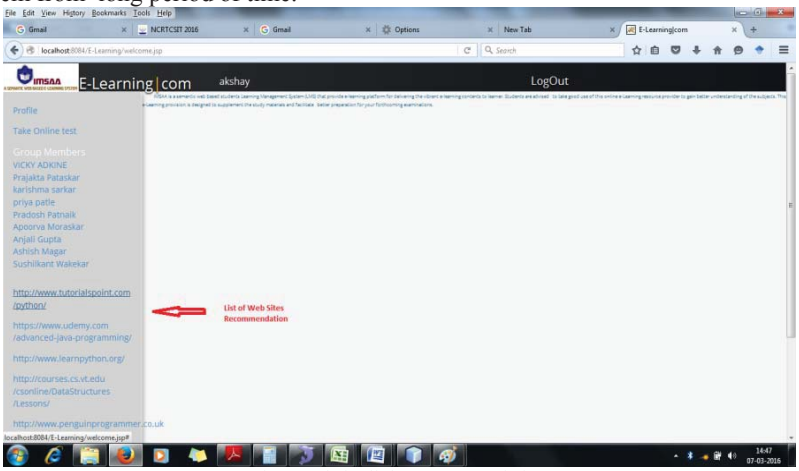


Fig.5 Website links recommendation

To overcome the this cold start problem, we are introducing an innovative approach of recommendation by combining the socioeconomic status of learners and his knowledge i.e. the group G1,which is a collection of users location , hobby , mother tongue, family members and sibling , father income and occupation , his educational information from secondary school to higher degree level.

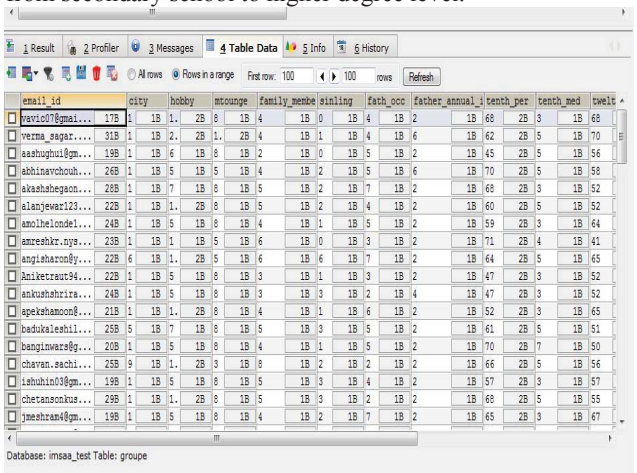
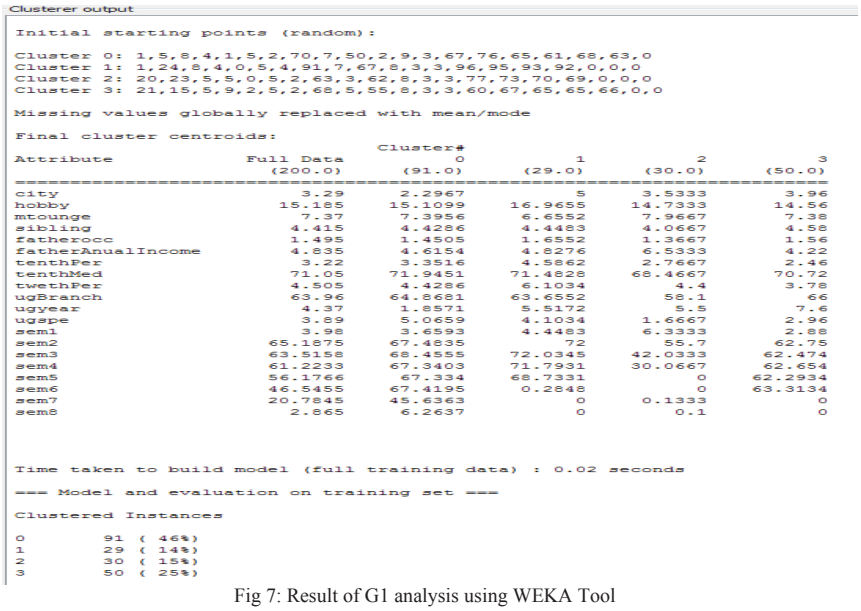


Fig6. G1 Result Dataset

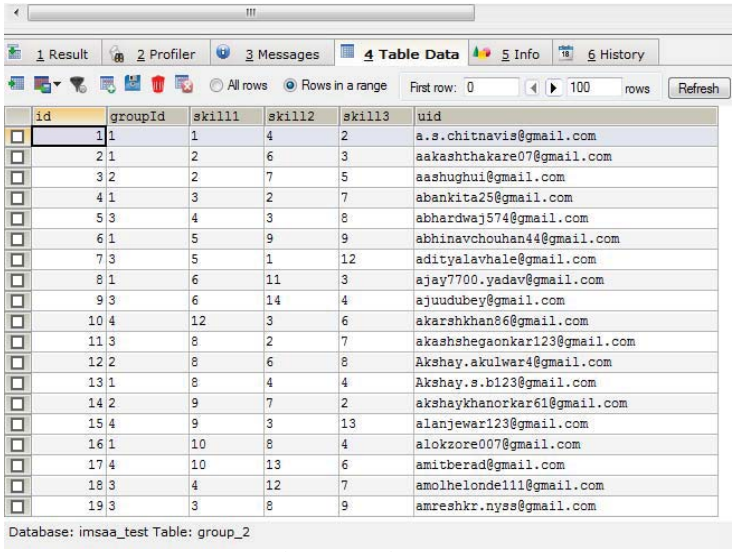
WEKA tool is used to analysis the k-means clustering algorithms ,which moved the existing data among sets of cluster until the desired set is reached. Here we used Euclidean distance to find the K0, K1,K2 and K3 clustered as shown in fig7.



```
Clusterer output

Initial starting points (random):

Cluster 0: 1,5,8,4,1,5,2,70,7,50,2,9,3,67,76,65,61,68,63,0
Cluster 1: 1,24,8,4,0,5,4,91,7,67,8,3,3,96,95,93,92,0,0,0
Cluster 2: 20,23,5,5,0,5,2,63,3,62,8,3,3,77,73,70,69,0,0,0
Cluster 3: 21,15,5,9,2,5,2,68,5,55,8,3,3,60,67,65,65,66,0,0

Missing values globally replaced with mean/mode

Final cluster centroids:
                                Cluster#
Attribute               Full Data       0          1          2          3
                          (200.0)    (91.0)     (29.0)     (30.0)     (50.0)
=============================================================================
city                        3.29     2.2967          5     3.5333       3.96
hobby                      15.185    15.1099    16.9655    14.7333      14.56
mtounge                      7.37     7.3956     6.6552     7.9667       7.38
sibling                     4.415     4.4286     4.4483     4.0667       4.58
fatherocc                   1.495     1.4505     1.6552     1.3667       1.56
fatherAnualIncome           4.835     4.6154     4.8276     6.5333       4.22
tenthPer                     3.22     3.3516     4.5862     2.7667       2.46
tenthMed                    71.05    71.9451    71.4828    68.4667      70.72
twethPer                    4.505     4.4286     6.1034        4.4       3.78
ugBranch                    63.96    64.8681    63.6552       58.1         66
ugyear                       4.37     1.8571     5.5172        5.5        7.6
ugspe                        3.89     5.0659     4.1034     1.6667       2.96
sem1                         3.98     3.6593     4.4483     6.3333       2.88
sem2                       65.1875    67.4835         72       55.7      62.75
sem3                       63.5158    68.4555    72.0345    42.0333      62.474
sem4                       61.2233    67.3403    71.7931    30.0667      62.654
sem5                       56.1766     67.334    68.7331          0      62.2934
sem6                       46.5455    67.4195     0.2848          0      63.3134
sem7                       20.7845    45.6363          0     0.1333            0
sem8                         2.865     6.2637          0        0.1            0


Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       91 ( 46%)
1       29 ( 14%)
2       30 ( 15%)
3       50 ( 25%)
```

Fig 7: Result of G1 analysis using WEKA Tool

The K-NN classification algorithm is used to form a group G2 which is based on the result of C test and his selected skill set .The KNN technique assumes that the entire training set include not only the data in the set but also the desired classification for each item .When this classification is to be made for new learner , its distance to each co-learners in the training set must be determine. Only k closest entries in the training set are consider further. The new learner is then placed in the class that contains the most co-learner from this set of k-closest co-learner.



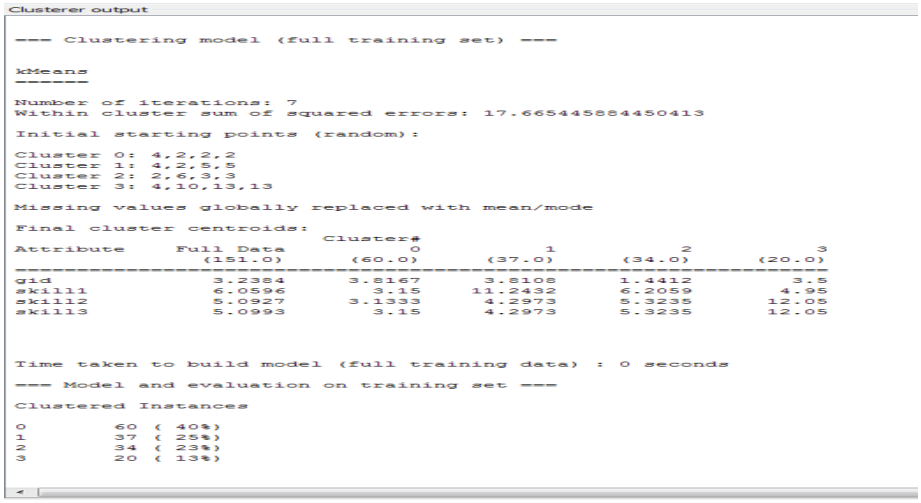
1 Result | 2 Profiler | 3 Messages | 4 Table Data | 5 Info | 6 History

All rows | Rows in a range | First row: 0 | 100 rows | Refresh

	id	groupId	skill1	skill2	skill3	uid
	1	1	1	4	2	a.s.chitnavis@gmail.com
	2	1	2	6	3	aakashthakare07@gmail.com
	3	2	2	7	5	aashughui@gmail.com
	4	1	3	2	7	abankita25@gmail.com
	5	3	4	3	8	abhardwaj574@gmail.com
	6	1	5	9	9	abhinavchouhan44@gmail.com
	7	3	5	1	12	adityalavhale@gmail.com
	8	1	6	11	3	ajay7700.yadav@gmail.com
	9	3	6	14	4	ajuudubey@gmail.com
	10	4	12	3	6	akarshkhan86@gmail.com
	11	3	8	2	7	akashshegaonkar123@gmail.com
	12	2	8	6	8	Akshay.akulwar4@gmail.com
	13	1	8	4	4	Akshay.s.b123@gmail.com
	14	2	9	7	2	akshaykhanorkar61@gmail.com
	15	4	9	3	13	alanjewar123@gmail.com
	16	1	10	8	4	alokzore007@gmail.com
	17	4	10	13	6	amitberad@gmail.com
	18	3	4	12	7	amolhelonde111@gmail.com
	19	3	3	8	9	amreshkr.nyss@gmail.com

Database: imsaa_test Table: group_2

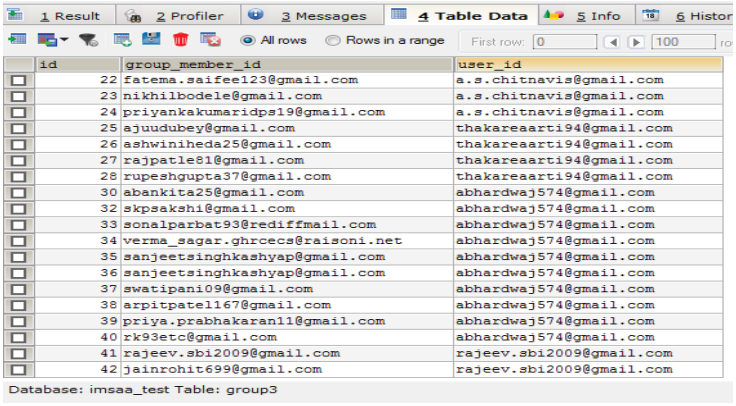
Fig8. G2 Result Dataset



```
Clusterer output

=== Clustering model (full training set) ===

kMeans
======

Number of iterations: 7
Within cluster sum of squared errors: 17.665445854450413

Initial starting points (random):

Cluster 0: 4,2,2,2
Cluster 1: 4,2,5,5
Cluster 2: 2,6,3,3
Cluster 3: 4,10,13,13

Missing values globally replaced with mean/mode

Final cluster centroids:
                        Cluster#
Attribute     Full Data         0          1          2          3
                (151.0)    (60.0)    (37.0)    (34.0)    (20.0)
=============================================================================
gid                3.2384    3.8167     3.8108     1.4412        3.5
skill1             6.0596      3.15    11.2432     6.2059       4.95
skill2             5.0927    3.1333     4.2973     5.3235      12.05
skill3             5.0993      3.15     4.2973     5.3235      12.05


Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0         60  ( 40%)
1         37  ( 25%)
2         34  ( 23%)
3         20  ( 13%)
```

Fig9. Result of G2 using WEKA Tool

Whereas group G3 is a intersection of G1 and G2 which contain the most matching co- learners for a new learner.



1 Result | 2 Profiler | 3 Messages | 4 Table Data | 5 Info | 6 Histor

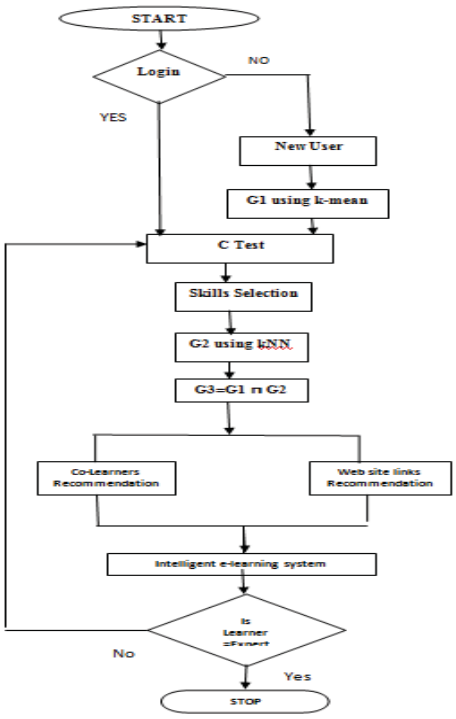
All rows | Rows in a range | First row: 0 | 100

	id	group_member_id	user_id
	22	fatema.saifee123@gmail.com	a.s.chitnavis@gmail.com
	23	nikhilbodele@gmail.com	a.s.chitnavis@gmail.com
	24	priyankakumaridps19@gmail.com	a.s.chitnavis@gmail.com
	25	ajuudubey@gmail.com	thakareaarti94@gmail.com
	26	ashwiniheda25@gmail.com	thakareaarti94@gmail.com
	27	rajpatle81@gmail.com	thakareaarti94@gmail.com
	28	rupeshgupta37@gmail.com	thakareaarti94@gmail.com
	30	abankita25@gmail.com	abhardwaj574@gmail.com
	32	skpsakshi@gmail.com	abhardwaj574@gmail.com
	33	sonalparbat93@rediffmail.com	abhardwaj574@gmail.com
	34	verma_sagar.ghrcecs@raisoni.net	abhardwaj574@gmail.com
	35	sanjeetsinghkashyap@gmail.com	abhardwaj574@gmail.com
	36	sanjeetsinghkashyap@gmail.com	abhardwaj574@gmail.com
	37	swatipani09@gmail.com	abhardwaj574@gmail.com
	38	arpitpatel167@gmail.com	abhardwaj574@gmail.com
	39	priya.prabhakaran11@gmail.com	abhardwaj574@gmail.com
	40	rk93etc@gmail.com	abhardwaj574@gmail.com
	41	rajeev.sbi2009@gmail.com	rajeev.sbi2009@gmail.com
	42	jainrohit699@gmail.com	rajeev.sbi2009@gmail.com

Database: imsaa_test Table: group3

Fig10.G3 result

b. System Work Flow



c. Experiment Evaluation

Data sets

1.MovieLen Data Set: The samples data used for evaluation was taken from the MovieLens data set, which provides 100,000 ratings (1-5) from 943 users on 1682 movies dataset. The data set was downloaded from the Grouplens website [GroupLens Research. http://www.grouplens.org/node/12.].

For this experiment , we are considering the sparsity level data set, which is find out by 1- nonzero entry/no. of cloumes A table that is 10% dense has 10% of its cells populated with non-zero values. It is therefore 90% sparse – meaning that 90% of its cells are either not filled with data or are zeros. :

2.IMSAA Data Set : The IMSAA is the online data set that generated data set real time . For collecting this data we had develop online web application www.imsaa.in and collected 200+ learners information from various colleges in India. This data set contain 245 learners data at tbl_personal_info having personal information of each learners , tbl_parent_details has details information of learners parent, while tbl_education_details has all educational information of each learners started from his middle school to higher studies.

Evaluation Metrics

To retrieve top-N recommendation , we used two metrics that widely used for Information Retrieve(IR) i.e. recall and precision[33]

Precision : Precision takes all retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n.

$$\boldsymbol{Precision} = \frac{|Relevent\ Data| \sqcap |Retrive\ Data|}{|Retrive\ Data|}$$

for this recommendation system experiment we define precision as

$$\mathbf{Precision} = 1 - \frac{Relevent\ Data}{Total\ Retrive\ Data}$$

Recall : Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\boldsymbol{Recall} = \frac{|Relevent\ Data| \sqcap |Retrive\ Data|}{|Relevent\ Data|}$$

for our recommendation system experiment we are considering recall is always 100%.

We used standard F1 metric(Accuracy of System) that gives equal weight to both recall and precision as it is computed by

$$F1 = \frac{2 * \text{recall} * \text{precision}}{Recall + Rrecision}$$

d. Result Analysis

To check our propose methodology , we are tested it on the IMSAA and MovieLens database.

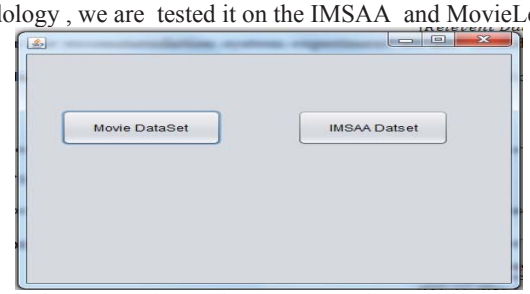


Fig 11.Comparing MovieLen Data set and IMSAA data set

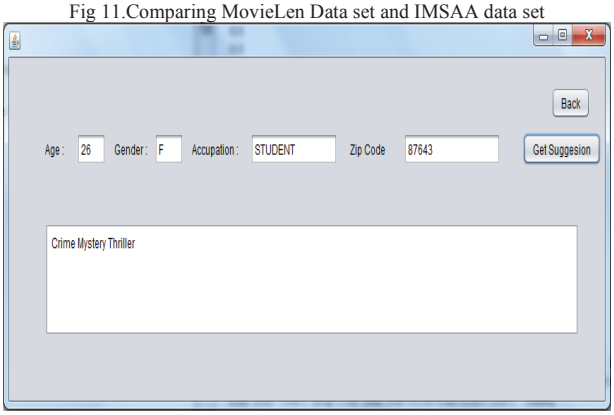


Fig12. MovieLen Recommendation

For MovieLen data set ,we are considering its 18 different categories of movie like Action, Crime, Mystery etc.

Sr. No	Age	Sex	Occupation	Zipcode	Precision	Recall	F1
1	26	F	LEARNER	87643	0.833	1	0.909
2	39	M	Entertainment	43211	0.888	1	0.941
3	70	M	Executive	23567	0.888	1	0.941
4	28	F	librarian	76543	0.888	1	0.941
5	26	M	LEARNER	87655	0.833	1	0.909
6	56	M	Lawyer	65432	0.888	1	0.941
7	36	F	Writer	76543	0.833	1	0.909
8	18	F	OTHER	87665	0.833	1	0.909
9	47	M	EDUCATOR	65432	0.777	1	0.875
10	51	F	SCIENTEST	78644	0.888	1	0.941
11	32	F	Writer	23456	0.833	1	0.909
12	41	M	programmer	34567	0.833	1	0.909
13	29	M	programmer	54358	0.888	1	0.941
14	27	F	marketing	43543	0.833	1	0.909
15	53	M	marketing	76789	0.944	1	0.971
16	45	M	administrator	23567	0.888	1	0.941
17	23	F	learner	42489	0.777	1	0.875
18	21	M	writer	34432	0.833	1	0.909
19	28	M	educator	23442	0.833	1	0.909
20	18	F	learner	43134	0.888	1	0.941
21	26	M	programmer	43551	0.777	1	0.875
22	22	M	executive	42144	0.833	1	0.909
23	37	M	programmer	54232	0.888	1	0.941
24	25	M	librarian	54674	0.777	1	0.875
25	16	M	writer	34343	0.777	1	0.875
26	27	M	programmer	23546	0.833	1	0.909
27	49	M	educator	53454	0.944	1	0.971
28	50	M	healthcare	54350	0.833	1	0.909
29	36	M	engineer	54354	0.833	1	0.909
30	27	F	administrator	54534	0.888	1	0.941

Table 1: MovieLen Result

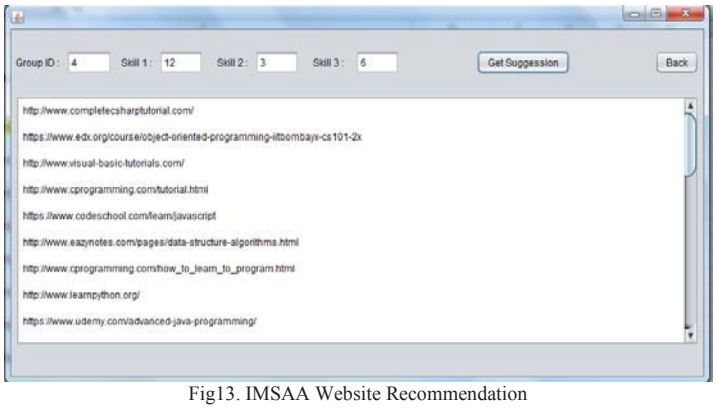


Fig13. IMSAA Website Recommendation

Sr.No	Group Id	Skill1	Skill2	Skill3	Precision	Recall	F1
1	2	2	7	5	0.666	1	0.800
2	1	3	2	7	0.533	1	0.696
3	1	6	11	3	0.729	1	0.843
4	3	6	14	4	0.621	1	0.766
5	4	12	8	14	0.836	1	0.911
6	4	12	4	9	0.734	1	0.847
7	3	4	3	8	0.600	1	0.750
8	3	4	12	7	0.571	1	0.727
9	4	10	3	2	0.666	1	0.800
10	2	9	7	2	0.600	1	0.750
11	4	8	8	4	0.700	1	0.824
12	2	2	5	6	0.900	1	0.947
13	2	14	8	3	0.730	1	0.844
14	3	11	9	12	0.450	1	0.621
15	4	12	4	9	0.450	1	0.621
16	3	3	8	9	0.570	1	0.726
17	2	9	7	2	0.730	1	0.844
18	4	12	3	6	0.730	1	0.844
19	3	8	2	7	0.810	1	0.895
20	4	6	4	13	0.720	1	0.837
21	1	7	3	12	0.540	1	0.701
22	4	8	6	8	0.900	1	0.947
23	4	6	9	12	0.540	1	0.701
24	1	8	10	6	0.850	1	0.919
25	4	1	13	4	0.570	1	0.726
26	1	4	2	3	0.550	1	0.710
27	2	13	3	6	0.733	1	0.846
28	4	14	2	7	0.733	1	0.846
29	2	9	9	5	0.700	1	0.824
30	3	11	2	8	0.800	1	0.889

Table 2: IMSAA Result

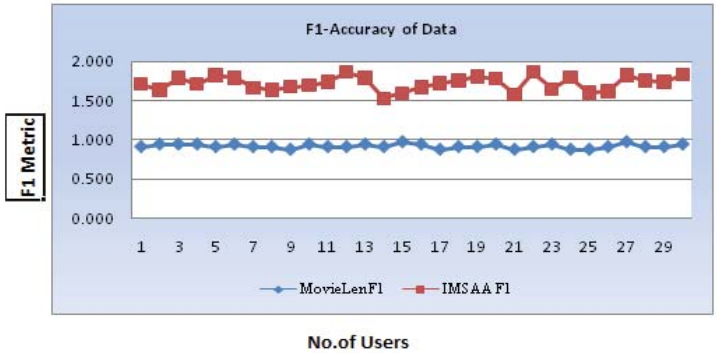


Fig 14: Impact of neighborhood size on recommendation quality .

d. RESEARCH IMPLICATIONS

In the proposed work, IMSAA dataset have been compared with MovieLen dataset. In table1 , we considered four parameters (age ,gender, occupation , location) of Movielen data to get the accurate recommendation of movies to user. We then calculated each individual precision and recall functions and finally F1 of individual users. The same methodology we implemented for IMSAA dataset to find out the accurate socioeconomic status and current knowledge level. In table2, we considered learners current group and his first 3 skills set interest . After comparing result of IMSAA and MovileLen [fig14] we had find out that the impact of neighborhood size works same in both system . Hence our proposed methodology of socioeconomic status of learners and current knowledge level are useful to solve the Cold Start Problem in e-learning system.

VI.CONCLUSION

Implemented collaborative filtering generate recommendations based on co-learners similarity.
Since the system does not have any data about the new learner preferences, it could not provide any personalized recommendation for him/her.In this paper we have reviewed several methods for dealing with the new learner problem via his socioeconomically status and knowledge level.
The results obtained by proposed technique is good and we can use that. From the above proposed methods, the results are reported in the form of graph. From that we can conclude that in CF eventhough system does not have enough information about new learners , it can provided sufficient recommendation to leaner so that they can start initial working with an e-learning system and overcome the problem of cold start.

REFERENCES

[1] Faisal Ibrahim Mohammad Al-Matalka, “The Influence of Parental Socioeconomic Status on Their Involvement at Home”, International Journal of Humanities and Social Science, Vol. 4 No. 5; March 2014,pp146-154
[2] IGBO J. N, OKAFOR, RITA A & EZE, J. U , “THE ROLE OF SOCIO-ECONOMIC BACKGROUND ON SELF-CONCEPT AND ACADEMIC ACHIEVEMENT OF IN-SCHOOL ADOLESCENTS IN NIGERIA”, International Journal of Research in Humanities, Arts and Literature , ISSN(E): 2321-8878; ISSN(P): 2347-4564 Vol. 2, Issue 2, Feb 2014, 1-10
[3] Kevin Majorbanks , “Family Learning Environment and Students Outcomes : A Review ” Journal of Comparative Family Studies, Vol. 2, Issue 2 ,pp 373-394
[4] Angira Bodhit, Kiran Amin , “Possible Solutions for Cold-Start Problem in Web Personalization”, ELSEVIER publication
[5] Darshana Gupta, Vatika Tayal, Amit Thakkar, Kamlesh Makvana, “Cold Start Problem in Personalized Web Search”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 1, January 2016
[6] Mohammed Amine ALIMAM, Hamid SEGHIOUER, “Personalized E-Learning Environment based on Ontology to Improve Learners’ School Levels” , International Journal of Computer Applications (0975 8887), Volume 84 - No. 12, December 2013
[7] Lantao Hu, Zhao Du, Qiuli Tong, Yongqi Liu , “Context-Aware Recommendation of Learning Resources Using Rules Engine”, IEEE 13th International Conference on Advanced Learning Technologies,2013,pp.181-183
[8] Lakshmi Sunil, Dinesh K Saini, “Design of a Recommender System for Web Based Learning”, Proceedings of the World Congress on Engineering 2013 Vol I,, July 3 - 5, 2013, London, U.K
[9] Ye Xu, Zang Li, Abhishek Gupta, Ahmet Bugdayci, Anmol Bhasin, “Modeling Professional Similarity by mining Professional Career Trajectories”, KDD’14, August 24–27, 2014, New York, NY, USA
[10] Mohamed Nader Jelassi, Sadok Ben Yahia, Engelbert Mephu Nguifo,” A Personalized Recommender System Based on Users’ Information In Folksonomies”, International World Wide Web Conference Committee (IW3C2), May 13–17, 2013, Rio de Janeiro, Brazil.
[11] Muhammad Waseem Chughtai,Ali Selamat, Imran Ghani,Jason J. Jung , “E-Learning Recommender Systems Based on Goal-Based Hybrid Filtering”, International Journal of Distributed Sensor Networks Volume 2014
[12] Paolo Cremonesi, Francesco Epifania, Franca Garzotto, “ User Profiling vs. Accuracy in Recommender System User Experience”, AVI ‘12, May 21-25, 2012, Capri Island, Italy
[13] Lekakos, G., Giaglis, G. M. “ A hybrid approach for improving predictive accuracy of collaborative filtering algorithms”. User Modeling and User-Adapted Interaction 17, 2007, 5–40
[14] Zhi-Min Zhou, Man Lan, Zheng-Yu Niu, Yue Lu “Exploiting User Profile Information for Answer Ranking in cQA”, International World Wide Web Conference Committee (IW3C2), April 16–20, 2012, Lyon, France.
[15] Lei Li, Dingding Wang, Tao Li, Daniel Knox, Balaji Padmanabhan, “SCENE : A Scalable Two-Stage Personalized News Recommendation System”, SIGIR’11, July 24–28, 2011, Beijing, China.
[16] Mohammad-Hossein Nadimi-Shahraki and Mozhde, “Cold-start Problem in Collaborative Recommender Systems: Efficient Methods Based on Ask-to-rate Technique”, Journal of Computing and Information Technology - CIT 22, 2014, 2, 105–113
[17] Yu-Yang Huang, Rui Yan, Tsung-Ting Kuo, Shou-De Lin , “Enriching Cold Start Personalized Language Model Using Social Network Information”, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 611–617, Baltimore, Maryland, USA, June 23-25 2014.
[18] Ahmad Ammari, Lydia Lau, Vania Dimitrova, “Deriving Group Profiles from Social Media to Facilitate the Design of Simulated Environments for Learning”, LAK’12, 29 April – 2 May 2012, Vancouver, BC, Canada

[19] Larrumbide, A., "Cone finds that Americans expect companies to have a presence in social media" , Cone Business in Social Media Research. October 10, 2012
[20] Najeeb Elahi, Randi Karlsen, Einar J. Holsb, "Personalized Photo Recommendation By Leveraging User Modeling On Social Network", iiWAS2013, 2-4 December, 2013, Vienna, Austria.
[21] Lin Chen, Richi Nayak, "Social Network Analysis of an Online Dating Network", C&T'11, 29 June – 2 July 2011, QUT, Brisbane, Australia
[22] Shehab A. Gamalel-Din, "An Intelligent eTutor-Learner Adaptive Interaction Framework", Interaccion'12, Oct 3–5, 2012, Elche, Alicante, Spain
[23] Ahmad Hawalah, Maria Fasli , "A multi-agent System Using Ontological User Profiles for Dynamic User Modelling," IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology,2011,pp-430-437
[24] João C. Prates, Sean S. M. Siqueira, "Using educational resources to improve the efficiency of Web searches for additional learning material", 11th IEEE International Conference on Advanced Learning Technologies,2011,pp.563-567
[25] Iván Cantador, Alejandro Bellogín, Pablo Castells, "Ontology-based Personalised and Context-aware Recommendations of News Items", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology,2008,pp.562-565
[26] Myriam Abramson, "Learning Temporal User Profiles of Web Browsing Behaviour", ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conference, Stanford University, May 27-31, 2014.
[27] Chang, M. K., Cheung, W., & Lai, V. S. (2005). Literature derived reference models for the adoption of online shopping. Information & Management, 42(4), 543–559
[28] Sylvia M. Chan-Olmsted, Moonhee Cho, Sangwon Lee, " User Perceptions of Social Media: A Comparative Study of Perceived Characteristics and User Profiles by Social Media", Online Journal of Communication and Media Technologies Volume: 3 – Issue: 4 – October – 2013
[29] Jeff David, Samir Bajaj, Cherif Jazra, "A Facebook Profile-Based TV Recommender System".
[30] Anita JUŠKEVIČIENĖ1, Eugenijus KURILOVAS, " On Recommending Web 2.0 Tools to Personalise Learning", Informatics in Education, 2014, Vol. 13, No. 1, 17–31
[31] Kleanthi Lakiotaki, Angelos Hliaoutakis, Serafim Koutsos and Euripides G.M. Petrakis, "Recommending medical documents by user profile".
[32] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets:History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4, Article 19 (December 2015), 19 pages
[33] Kowalski,G.,1997 "Information Retrieval Systems: Theory and Implementation" , Kluwer Academic Publisher,Norwell,MA .