IMPLEMENTATION OF NEW DATA MINING RECOMMENDATION MODEL USING MODIFIED K-MEANS ALGORITHM FOR SOCIAL MEDIA NETWORKS ON WORLD WIDE WEB

IRFAN AHMED[1], Dr. Shivani Chauhan[2]

[1]*Assistant Professor, Dept. of Computer Science, Sree Siddaganga College of Arts Science and Commerce for Women, Tumkur, irfanahmed.dba@gmail.com*
[2]*Associate Dean of Student Welfare & Assistant Professor, Bagawanth Institute of Technology, Muzaffarnagar, shivanichauhanbit@gmail.com*

Abstract

In web content mining large collection of web documents contents are used to extract useful information from web data. Web usage mining is used to analyze customer behavior and choice of interest. Web users on web clients may access all social media tools, with help of these tools they can interact with their friends, also possible make new friends by accepting friend requests, also possible to recommend or share data, audio, video files to friends. In a traditional web recommended system, first consider web user navigations on web browser, previous session search keywords, previous navigations of user interest, new set of web documents are recommended to users. Social media recommendations may be correct or sometimes may lead to wrong recommendations to attacker's site, due these users may lost money and also waste of time to spend time on unwanted sites, and recommendation of right documents or files is also challenging task. In this research paper, we are planning to integrate social media networks to data mining engine to recommend right and appropriate web documents to web users as and when they required or interested on recommendations.

Key words: *Data Mining, Social Media Networks, Recommender model, Rating, Learning, product rating,*

I. INTRODUCTION

Data mining is a process of extracting useful and previously unknown patterns from different kinds of dat. If web data is used in data mining then it is called web mining. Web mining can be used as discovery and analysis of useful information from the WWW [4]. Derived from the different emphasis and ways to obtain information, mining of web data can be separated into three major parts, one is web contents mining (WCM), web structure mining, and other one is web usage mining. WCM can be treated as the automatic search and retrieval of information and resources available from millions of sites and online databases though web spiders or search engines. Web Usage Mining can be described as the discovery and analysis of user access patterns through the mining of weblog files and associated data from a specific website [6]. Taxonomy of web mining is shown in figure 1 and which is broadly classified into three categories.

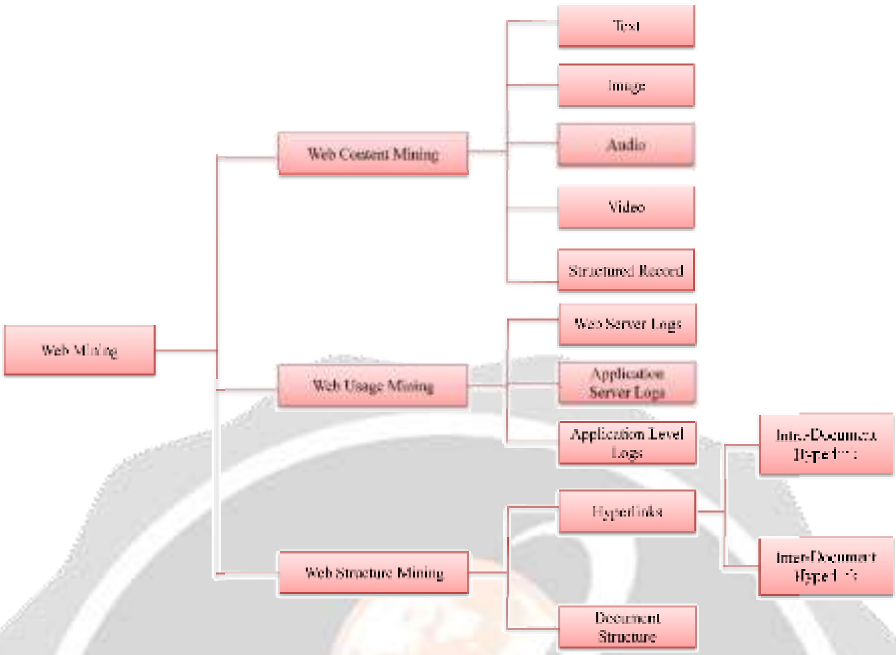

Figure 1: Taxonomy of web mining

In web content mining large collection of web documents contents are used to extract useful information from web data [5]. To analyze structure of web data hyperlinks web structure mining is used. Web usage mining is used to analyze customer behavior and choice of interest [7]. Web users on web clients may access all social media tools, with help of these tools they can interact with their friends, also possible make new friends by accepting friend requests, also possible to recommend or share data, audio, video files to friends, and same is shown in figure 2.

Figure 2: Web user interact with social media

In a traditional web recommended system, first consider web user navigations on web browser, previous session search keywords, previous navigations of user interest, new set of web documents are recommended to users [8], and a sample recommendations to users or friend suggestions are shown in figure 3. Day by day usage of social media is increased tremendously, social media recommendations may be correct or sometimes may lead to wrong recommendations to attacker's site, due these users may lost money and also waste of time to spend time on unwanted sites, and recommendation of right documents or files is also challenging task [9]. In this research paper, we are planning to integrate social media networks to data mining engine to recommend right and appropriate web documents to web users as and when they required or interested on recommendations.

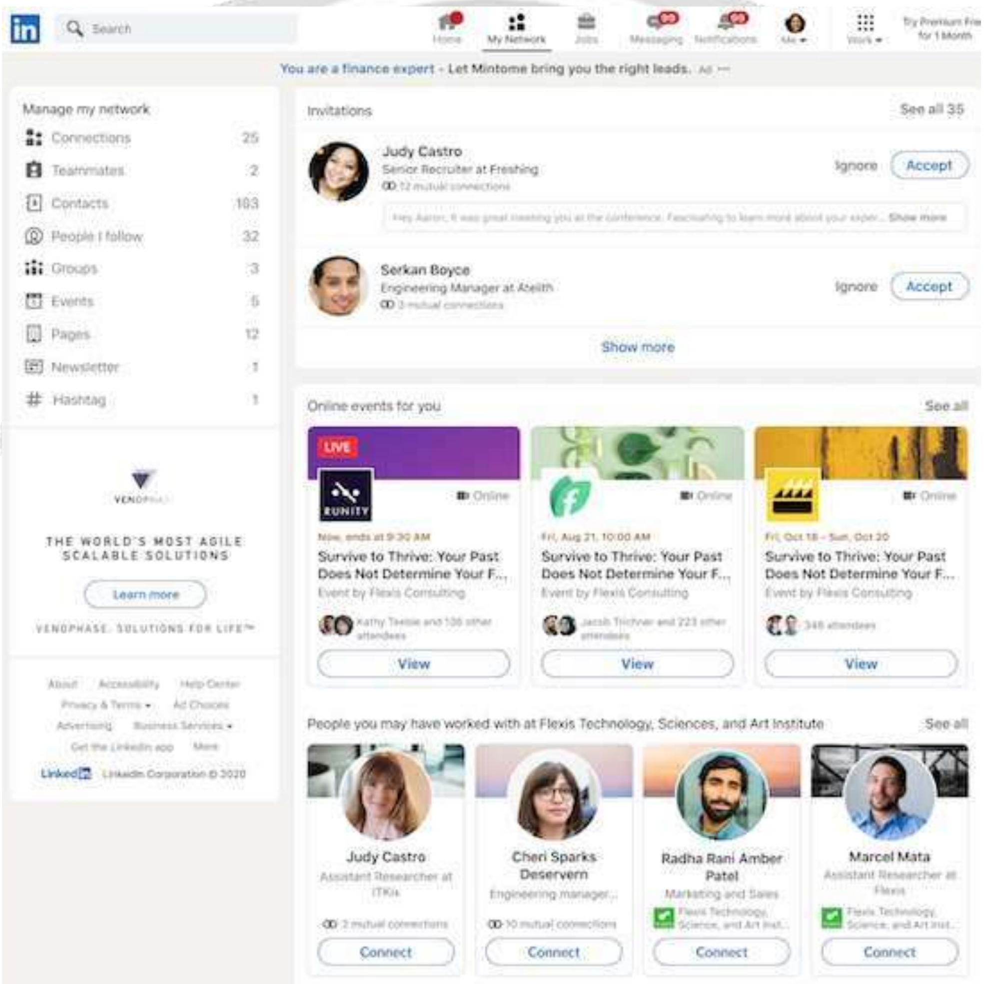

Figure 3: Friends, video, audio files recommendation on social media

II. RELATED WORK

Recommender frameworks can be characterized as projects which endeavor to suggest the most reasonable things (items or administrations) to specific clients (people or organizations) by foreseeing a client's advantage in a thing in light of related data about the things [20], the clients and the communications among things and clients [15]. The point of creating recommender frameworks is to decrease data over-burden by recovering the most important data and administrations from a colossal measure of information, in this manner offering customized types of assistance [14]. The main element of a recommender framework is its capacity to "surmise" a client's inclinations and interests by dissecting the way of behaving of this client or potentially the way of behaving of different clients to create customized suggestions [26].

E-administration personalization methods are embodied by recommender frameworks, which definitely stand out in the beyond couple of years [25]. Early examination in recommender frameworks outgrew data recovery and separating research [16], and recommender frameworks arose as an autonomous exploration region during the 2020 when scientists began to zero in on proposal issues that expressly depend on the rating structure [21]. Generally utilized proposal procedures incorporate cooperative separating, content-based and information based methods [16]. Every proposal approach enjoys benefits and impediments; for instance, CF has inadequacy, versatility and cold-start issues, while CB has overspecialized suggestions [24]. To tackle these issues, many high level suggestion approaches have been proposed, for example, informal organization based recommender frameworks, fluffy recommender frameworks, and setting mindfulness based recommender frameworks and gathering recommender frameworks [17].

With the advancement of suggestion approaches and strategies, increasingly more recommender frameworks (programming) have been executed and some genuine world recommender framework applications have been created [23]. It was brought up as of late that application study is the primary examination focal point of momentum recommender framework research, particularly in the flow time of large information [22]. The utilizations of recommender frameworks incorporate suggesting motion pictures, music, TV programs, books, archives, sites, meetings, the travel industry grand spots and learning materials, and include the areas of web based business, e-learning, e-library, e-government and e-business administration’s [19]. Consequently, to assist specialists with understanding the recommender framework improvement experience and to help engineers to endorse material frameworks advancement practically speaking, this paper surveys the most recent recommender frameworks that have been created involving grouped strategies in a scope of utilization fields [18]. We bunch recommender framework applications into eight primary areas: e-government, e-business, online business, e-library, e-learning, e-the travel industry, e-asset administrations and e-bunch exercises. The most average recommender frameworks in every application space are introduced and broke down, and the significant suggestion procedures utilized in the application space are distinguished.

A few review papers on recommender frameworks have been distributed over the most recent couple of years. In any case, these papers center on either suggestion procedures and approaches or a particular space of recommender framework improvement; none of these study papers centers around the extensive examination of recommender framework applications. For instance, the creator in [3] introduced an outline of content based, cooperative separating based, and half breed suggestion draws near. It portrays the different restrictions of these proposal draws near and talks about potential augmentations that could further develop suggestion abilities. In [1] explored principal proposal, assessment, social separating, and bunch suggestion procedures, as well as a few as of late evolved methods, for example, the area mindful and bio-motivated proposal strategies. In [13] assessed more papers on recommender framework regions and characterized them by the diary and year of distribution, their application fields, and their information mining methods. In [11] overviewed the scene of genuine and conceivable mixture recommender frameworks. The paper analyzes suggestion strategies and surveys hybridization techniques. In [10] explored suggestion calculations, zeroing in on a cautious clarification of how the most often involved calculations in recommender frameworks work. They additionally introduced the fundamental ideas of cooperative sifting and their assessment measurements, dimensionality decrease strategies, dispersion based techniques, social separating and Meta approaches. Moreover, there are recommender framework overview papers on unambiguous application spaces, for example, internet business recommender frameworks and e-learning recommender frameworks [2, 12].

III. WEB RECOMMENDATION SYSTEM

Friends on social media network interact with each other, eight friends and their relations are shown in figure 4, relations are represented using 8 X 8 matrix shown in figure 5, edges associated with social media network graph represent connections between web users in a same session, and in general it is different from traditional networks. Web users on social media may access may items digitally and they may rate items of interest from zero to five.

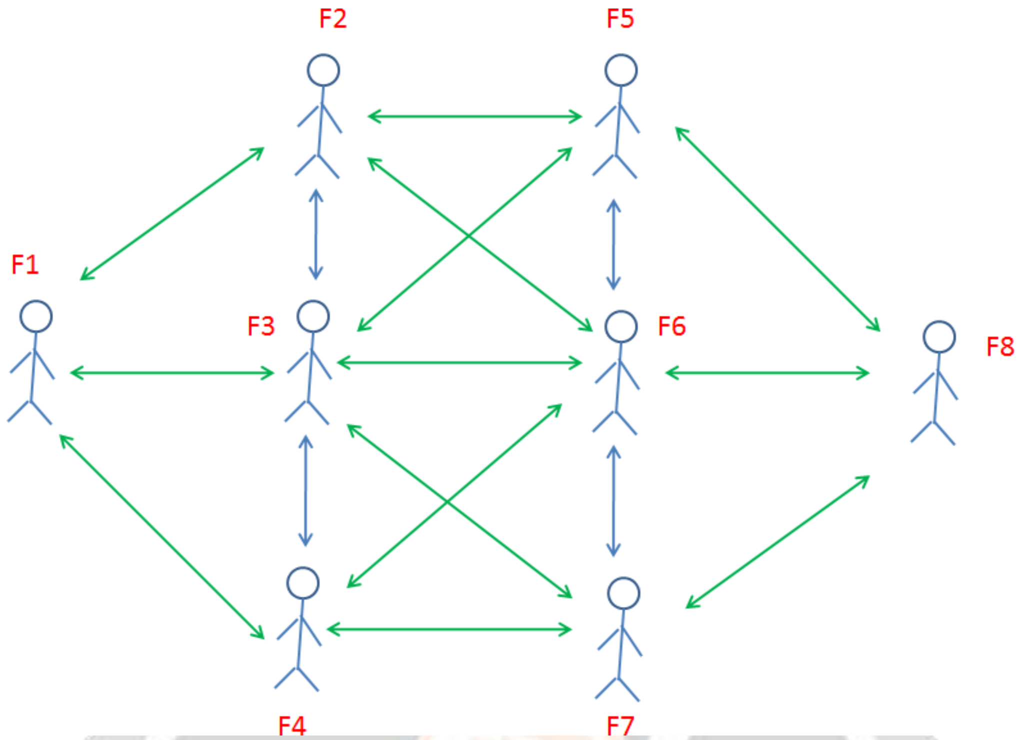

Figure 4: Social network friend relation graph

$$\begin{matrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{matrix}$$

Figure 5: Social network friendship matrix list

Large collection of documents, online products, and items are available World Wide Web (WWW). The web users can review products and items on five point grade scale. Each user has personal data, friends list, new friend suggestions, already used products review comments, and provision to give comments on products or items. In figure 6 web user $f_1$has friends list f_2, f_3, f_4 , and review rating on products $e_1, e_4, e_7, e_9, e_{12}$, web user $f_2$has friends list f_1, f_3, f_5, f_6 , and review rating on products $e_2, e_3, e_4, e_9, e_{11}$, web user f_3 has friends list $f_1, f_2, f_4, f_5, f_6, f_7$, and review rating on products $e_1, e_3, e_4, e_6, e_7, e_9, e_{11}, e_{12}$, and similarly all web users has friends and review comments, and same is listed in Table 1.

Web user	Friends List	Review on products or rating											
		e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}	e_{12}
f_1	f_2,f_3,f_4	5	-	-	4	-	-	4	-	3	-	-	5
f_2	f_1,f_3,f_5,f_6	-	5	4	3	-	-	-	-	5	-	3	-
f_3	f_1,f_2,f_4,f_5,f_6,f_7	4	-	5	4	-	4	1	-	2	-	3	5
f_4	f_1,f_3,f_6,f_7	-	2	-	5	-	3	-	4	-	4	-	3
f_5	f_2,f_3,f_6,f_8	5	-	3	-	4	-	5	-	2	-	2	2
f_6	f_2,f_3,f_4,f_5,f_7,f_8	-	5	3	5	4	5	-	4	2	-	3	2
f_7	f_3,f_4,f_6,f_8	-	4	4	-	3	3	-	3	1	3	4	-
f_8	f_5,f_6,f_7	3	-	4	-	4	-	3	-	2	-	4	4

Table 1: Web users friends list and review rating on products

Web users and reviews on product rating is fit into two dimensional surface, and Modified K-Means (MKM) algorithm is used to group web users. In 2D surface similarity between users are calculated, and if both are close then place them into same cluster or otherwise place them into different clusters. Repeat this procedure until no new clusters are formed.

Algorithm MKM ( )

Definitions:

$\propto$, is a similarity between f_i and f_j , both users are friends and are in the same cluster.

þ , is a similarity between f_i and f_j , both users are not friends and are in different cluster.

y , is a probability of supporting same products by user f_i and f_j.

f^+, denotes user ‘i’ and ‘j’ are in same cluster

f^- , denotes user ‘i’ and ‘j’ are in different clusters

N, is a number of products rated by both users f_i and f_j.

ð, number of products and list of users on same cluster rated product P.

8, number of products rated by both users f_i and f_j.

S, Similarity measure

P is a set of products, $\{e_1, e_2, e_3, \ldots\ldots, e_n\}$

WU is a set of web users, $\{f_1, f_2, f_3, \ldots\ldots, f_n\}$

L is a list of clusters

Input: web users-product review matrix, M

Output: List of clusters with users

Step 1: Start

Step 2: initialize, $L = \{\emptyset\}$

Step 3: Repeat

Step 4: for (WU, P) in WU X P do

Step 5: find similarity between web users f_i and f_j

Step 6: $S(f_i, f_j) = y * \alpha(f_i, f_j^{+}) + (1 - y) * þ(f_i, f_j^{-})$

Step 7: $\alpha(f_i, f_j^{+}) = \frac{\sum_{i=1}^{ð} Coc\,(P_{f_i}, P_{f_j})}{N}$

Step 8: $þ(f_i, f_j^{+}) = \frac{\sum_{i=1}^{8} Coc\,(P_{f_i}, P_{f_j})}{N}$

Step 9: Apply K-means algorithm on web users and identify group of users in a Cluster (C).

Step 10: $L = L + C$

Step 11: End for

Step 12: until no new clusters are added

Step 13: End Algorithm

IV. PERFORMANCE EVALUATION

Example of recommendation rating by various users on Microsoft word and pages is shown in figure 6. These to software tools are compared with list of quality metrics like is it meat requirements of users, easy to use, easy to setup, administration of software, quality support, and customer support. We implemented proposed recommendation model and also compared proposed system results with [1], [5], [8], [9], and [15]. Performance of proposed model is evaluated using two metrics and prediction is defined as number of products that web user f_i rated is proportional to overall recommendation of products to other web users. Recall denoted as number of products rated by user f_i and reflect the proportional to other products, and which in turn web user f_i may rated to be recommend to other web users.

$$\text{Precision} = \frac{|q \cap r|}{|q|}$$

$$\text{Recall} = \frac{|q \cap r|}{|r|}$$

Where q denotes web user may be rate products and r denotes web user already rated products. Product (e_1, e_2, e_3, e_4) rating of various web users are analyzed with different recommendation methods are evaluated and its precision values are shown in figure 7. Product (e_5, e_6, e_7, e_8) rating of various web users are analyzed with different recommendation methods are evaluated and its precision values are shown in figure 8. Product (e_9, e_{10}, e_{11}, e_{12}) rating of various web users are analyzed with different recommendation methods are evaluated and its precision values are shown in figure 9. Product (e_2, e_4, e_7, e_{12}) rating of various web users are analyzed with different recommendation methods are evaluated and its recall values are shown in figure 10.

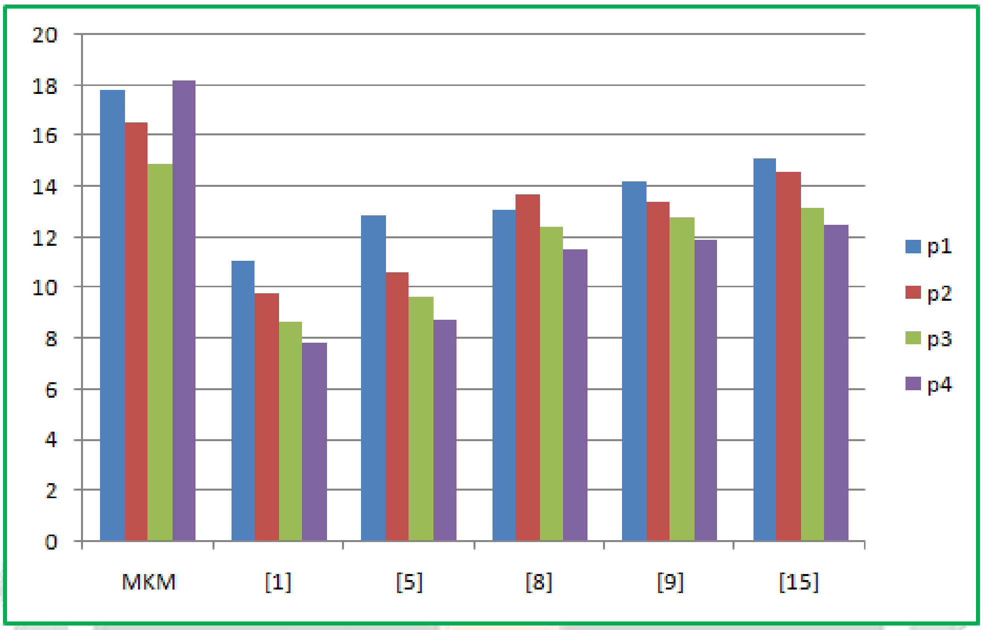

Figure 7: Comparison precision values of products p1, p2, p3, p4

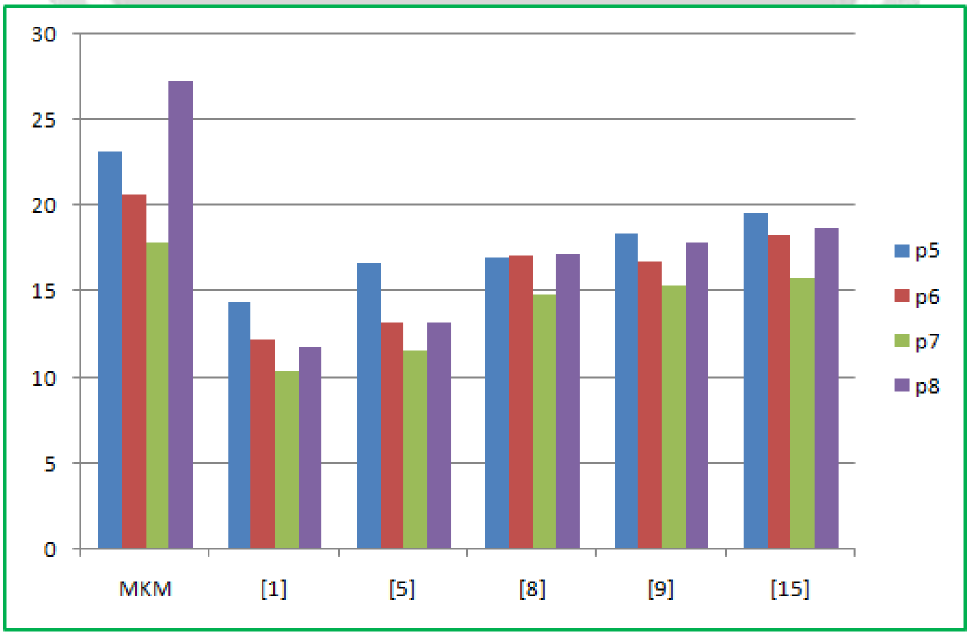

Figure 8: Comparison precision values of products p5, p6, p7, p8

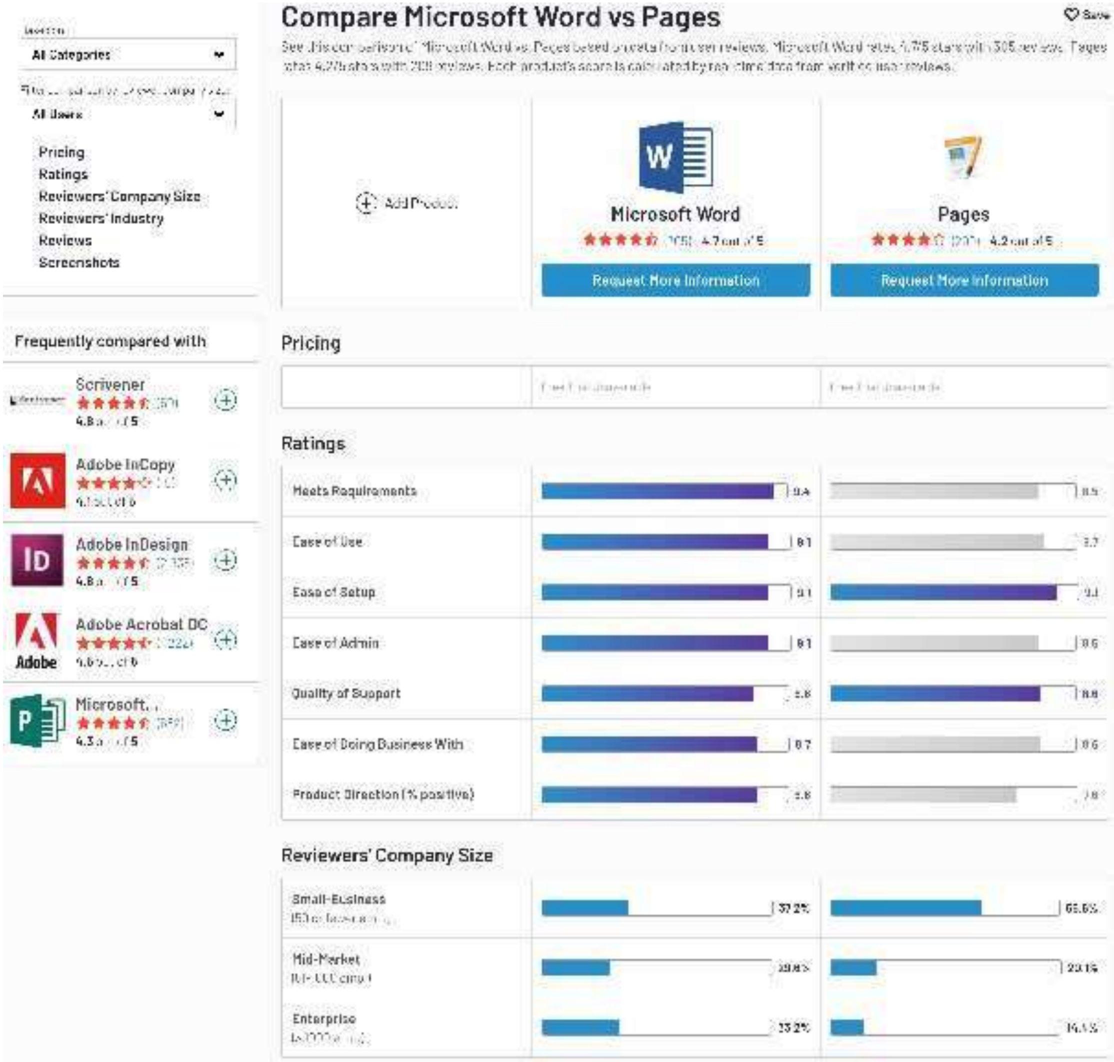

Figure 6: Comparison of software tools Microsoft word Vs pages

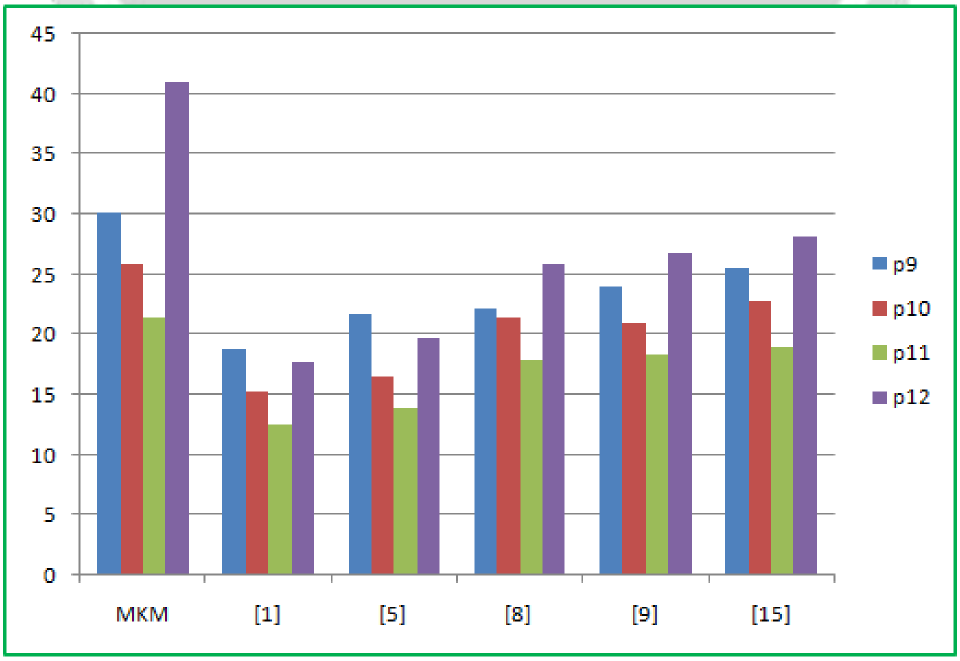

Figure 9: Comparison precision values of products p9, p10, p11, p12

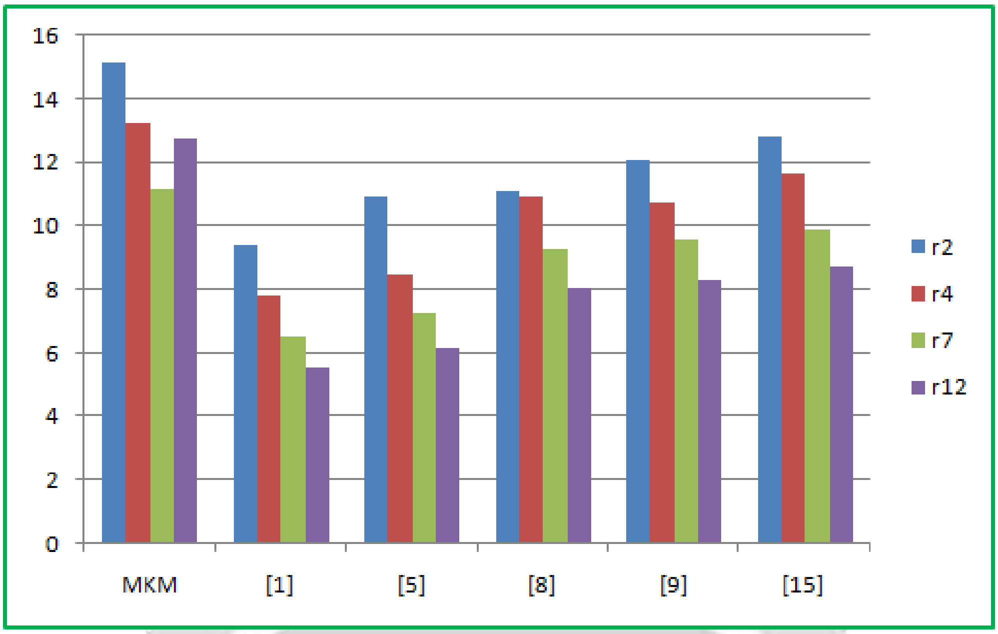

Figure 10: Comparison recall values of products p2, p4, p7, p12

V. CONCLUSION

Web users on web clients may access all social media tools, with help of these tools they can interact with their friends, also possible make new friends by accepting friend requests, also possible to recommend or share data, audio, video files to friends. In a traditional web recommended system, first consider web user navigations on web browser, previous session search keywords, previous navigations of user interest, new set of web documents are recommended to users. In this research paper, we are planning to integrate social media networks to data mining engine to recommend right and appropriate web documents to web users as and when they required or interested on recommendations. We implemented proposed recommendation model and also compared proposed system results with [1], [5], [8], [9], and [15]. Performance of proposed model is evaluated using two metrics precision and recall. From our experimental results we come to know that proposed modified k means algorithm shows better results than other recommendation systems.

VI. REFERENCES

[1]. Rui, C., Qingyi, H., Yan Shuo, C., Bo, W., Lei, Z., Xiangjie, K., "A Survey of Collaborative Filtering-Based Recommender Systems: From Traditional Methods to Hybrid Methods Based on Social Networks", IEEE Access, 6, 64301–64320, 2018.

[2]. Chen, X., "A real time anti-spamming system in crowd sourcing platform", 7th IEEE International Conference on Software Engineering and Service Science, pp. 981–984, 2016.

[3]. Massimo, D., Ricci, F., “Clustering Users’ POIs Visit Trajectories for Next POI Recommendation”, in Information and Communication Technologies in Tourism 2019, Proceedings of the International Conference in Nicosia, Cyprus, Springer: Cham, Switzerland, pp. 3–14, 2019.

[4]. da Silva Neubert, P., Rodrigues, R. S., Fachin, G. R. B., “Use of Web 2.0 tools by Ibero”, American journals on Transformation, vol. 28, no. 2, pp. 127-141, 2016.

[5]. Athanasiadis, C., Hortal, E., Koutsoukos, D., Lens, C.Z., Asteriadis, S., “Personalized, Affect and Performance-Driven Computer-Based Learning”, In Proceedings of the 9th International Conference on Computer Supported Education, pp. 132–139, 2017.

[6]. Zhang, Z., Lin, H., Li, P., Wang, H., Lu, D., “Improving semi-supervised text classification by using Wikipedia knowledge in Web-Age Information Management”, Berlin Germany: Springer-Verlag, pp. 25-36, 2013.

[7]. Junior, A.F., Medeiros, F., Calado, I., “An Evaluation of Recommendation Algorithms for Tourist Attractions”, In Proceedings of the Thirty Second International Conference on Software Engineering and Knowledge Engineering, Pittsburgh, PA, USA, 2020.

[8]. Pham, X. H., Jung, J. J., Park, S.B, “Exploiting social contexts for movie recommendation”, Malaysian Journal of Computer Science, vol. 27, no. 1, pp. 68-79, 2014.

[9]. Chai, Z., Li, Y.,L., Han, Y.M., Zhu, S. F., “Recommendation System Based on Singular Value Decomposition and Multi-Objective Immune Optimization”, IEEE Access, 7, 6060–6071, 2018.

[10]. Karydi, E., Margaritis, K., “Parallel and distributed collaborative filtering: A survey”, ACM Computing Survey., vol. 49, no. 2, p. 37, 2014.

[11]. Yuan, X., Han, L., Qian, S., Xu, G., Yan, H., “Singular value decomposition based recommendation using imputed data” Knowledge Based Systems, 163, 485–494, 2019.

[12]. Martinez-Cruz, C., Porcel, C., Bernabé-Moreno, J., Herrera-Viedma, E., “A model to represent users trust in recommender systems using ontology’s and fuzzy linguistic modeling”, Information Science, vol. 31, pp. 102-118, 2015.

[13]. Pujahari, A., Sisodia, D.S., “Model-Based Collaborative Filtering for Recommender Systems: An Empirical Survey”, in Proceedings of the 1st International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, pp. 443–447, 2020.

[14]. M. Giri, S.Jyothi, “Big Data Collection and Correlation Analysis of Wireless Sensor Networks Yielding to Target Detection and Classification”, Springer Lecture Notes on Data Engineering and Communications Technologies, Springer Singapore, Vol. 9, ISSN: 2367-4512, pp 201-213, December 2017.

[15]. Xue, F., He, X., Wang, X., Xu, J., Liu, K., Hong, R., "Deep Item-Based Collaborative Filtering for Top-N Recommendation", ACM Transactions on Information Systems, 1–25, 2019.

[16] Beel, J., Dinesh, S., "Real-world recommender systems for academia: The pain and gain in building, operating, and researching them", presented at the 5th Int. Workshop Biblio metric-Enhanced Information Retrieval, 2017.

[17]. Yin, C., Shi, L., Sun, R., Wang, J., "Improved Collaborative Filtering Recommendation Algorithm Based on Differential Privacy Protection", Journal of Super computing, 76, 5161–5174, 2020.

[18]. Harper, F.M., Konstan, J.A., "The movielens datasets: History and context", ACM Transactions on Interactive Intelligent Systems, vol. 5, no. 19, 2016.

[19]. Wakil, K., Bakhtyar, R., Ali, K., Alaadin, K., "Improving Web movie recommender system based on emotions", International Journal of Advanced Computer Science Applications, vol. 6, no. 2, pp. 218-226, 2015.

[20]. Musto, C., Narducci, F., Semeraro, G., Lops, P., de Gemmis, M., "A framework for explaining recommendations based on the linked open data cloud", 10th ACM Conference on Recommender Systems, pp. 151-154, 2016.

[21] Niemann, K., Wolpers, M., "Creating usage context-based object similarities to boost recommender systems in technology enhanced learning", IEEE Transactions on Learning Technologies, vol. 8, no. 3, pp. 274-285, 2015.

[22]. Zheng, X. L., Chen, C. C., Hung, J. L., He, W., Hong, F. X., Lin, Z., "A hybrid trust-based recommender system for online communities of practice", IEEE Transactions on Learning Technologies, vol. 8, no. 4, pp. 345-356, 2015.

[23] Colombo-Mendoza, L.O., Valencia-García, R., Rodríguez-González, A., Alor-Hernández, G., Samper-Zapater, J.J., "RecomMetz: A context aware knowledge-based mobile recommender system for movie show times", Expert Systems Applications, vol. 42, no. 3, pp. 1202-1222, 2015.

[24] Sarne, G.M.L., "A novel hybrid approach improving effectiveness of recommender systems", Journal of Intelligent Information Systems, vol. 44, no. 3, pp. 397-414, 2015.

[25] Tejeda Lorente, A., Porcel, C., Bernabe-Moreno, J., Herrera-Viedma, E., "REFORE: A recommender system for researchers based on biblio metrics", Applications Soft Computing, vol. 30, pp. 778-791, 2015.

[26]. Zou, H., Gong, Z., Zhang, N., Zhao, W., Guo, J., "Trust Rank: A cold start tolerant recommender system", Enterprise Information Systems, vol. 9, no. 2, pp. 117-138, 2015.