OPEN ACCESS Freely available online

Journal of Information Technology & Software Engineering

Review Article

Provide a Web Recommender System to Improve Search Quality Using DBSCAN Clustering Algorithms and SVM Machine Tracking Method

Molae Fard *, **Mohammad**

Department of Computer Engineering, Azad University, Dezful Branch, Khuzestan, Iran

ABSTRACT

Today, due to the increasing growth of web pages, the existence of a system that can extract the information needed by users from the huge amount of data available on the web seems necessary. To do this, we need to customize the systems in question. One of the best ways to customize your system is to use recommendation systems. Recommender systems are systems that can provide appropriate suggestions to the user by obtaining limited information from the user. Recommender systems can predict a user's future requests and then generate a list of the user's favorite pages. In other words, an accurate index of user behavior can be obtained and a page can be predicted that the user will select in the next move, which can solve the problem of cold start system and improve the quality of the search. In this article, a new method is proposed to improve the recommender systems in the field of web, which uses the DBSCAN clustering algorithm for data clustering, which achieves a 99% efficiency score. Then, using the Page rank algorithm, the user's favorite pages are weighed. Then, using the SVM method, we categorize the data and give the user a hybrid recommender system to generate a forecast, which will eventually provide the recommender with a list of pages that the user may be interested in. Evaluation of the research results showed that using this proposed method can achieve a score of 95% in the call section and a 99% score in the accuracy section, which proves that this recommending system can achieve up to 90%. Identify user pages correctly and greatly reduce the weaknesses of other previous systems.

Keywords: Recommender system; Data mining; Web mining; DBSCAN algorithm; Machine learning

INTRODUCTION

With the rapid advancement of technology and the Internet and the increase of web resources, the existence of a mechanism that can predict the needs and desires of users seems necessary. To this end, recommending or proposing systems were created. Suggestion systems are systems that help users find and select the items they want. It is natural that these systems will not be able to offer without having sufficient and correct information about users and their desired items as well as user search history; therefore, one of their main goals is to collect various information about the tastes of users and items in the system. One of the methods used by referral systems is the use of user behaviors, activities, and records such as pages visited, user interests, or interactions with other users. One of the major problems with web browser systems is the cold start. This problem occurs when new users log in or new items are added to the directory. In such cases, neither the tastes of new users can be predicted no new items can be rated or purchased by users, which leads to inappropriate and less accurate offers. In this research, a recommending system consisting of clustering algorithms and machine learning techniques is presented that will be able to solve the cold start problem that reduces the efficiency of recommending systems. For this purpose, DBSCAN clustering algorithms are used to discover patterns in the data. DBSCAN is the basic algorithm of density-based clustering methods. This algorithm is capable of detecting clusters of different sizes from a large volume and is also resistant to noise. The advantage of the DBSCAN clustering algorithm is that it can cluster data with different and irregular shapes, which other algorithms such as K-MEANS and C-MEANS are not able to do. In addition, it is used using machine techniques to generate predictions and classify data. The method used is SVM or backup vector machine, which will perform the best classification and data separation by benchmarking backup vectors. This recommender system can predict the user's future requests and then generate a list of the user's favorite pages. In other words, an accurate index of user behavior can be obtained and a page can

Correspondence to: Molae Fard, Department of Computer Engineering, Azad University, Dezful Branch, Khuzestan, Iran, E- mail: molae4@gmail.com

Received: March 15, 2021, **Accepted:** March 29, 2021, **Published:** April 05, 2021

Citation: Fard M, Mosleh M (2021) Provide a Web Recommender System to Improve Search Quality Using DBSCAN Clustering Algorithms and SVM Machine Tracking Method. J Inform Tech Softw Eng.11:260.

be predicted that the user will select in the next move, which can solve the user's problems when faced with a cold start and improve the quality of the search.

LITERATURE REVIEW

Research background

In their 2020 article, Maazouzi et al. presented a new way to improve the recommendation of web pages to users. In their paper, the researchers presented an effective suggestion system based on user conversations and grouping them into different groups, then effectively advising a group that had similar characteristics. The authors used Pearson correlation coefficient method and user TED negotiations. They then used the pages using the K-Means clustering method to group users and then suggest to the target user. The results of the researchers' research, which was performed using RMSE, obtained better accuracy and recall than other available methods [1].

In his 2018 article, Chawla outlined a way to improve recommendation systems for personalizing web pages. In his paper, the researcher used a combination of genetic algorithms and trust in clicked and reliable URLs to recommend web pages. User-trusted web pages based on cluster query sessions were used for optimal ranking with GA to retrieve more relevant documents in the rankings and improve the accuracy of the results. Optimal ranking of trusted clicked URLs recommends relevant documents to web users for their search purpose and broadly meets the user's information needs [2].

In their 2018 article, Buorkoukou and Omar presented a recommendation system to improve results using users' previous search history. In their proposed approach, the researchers tried to suggest learning resources to the learner based on his or her preferences and his or her previous search history, a history extracted from log files. This approach combines the styles of learning and collaborative filtering methods to improve the quality of offers. The results of these researchers indicated the improvement of recommendation systems in the field of offering pages to users [3].

In their 2020 paper, Riyahi and Sohrabi proposed a way to improve web page recommendation using a hybrid recommender system and the use of data tagging. The researchers extracted the meaning of the tags using the WORDNET vocabulary database and then organized the tags into a hierarchical structure based on their semantic significance. The hierarchical structure was used to search for relevant tags in the content filtering section, and user requests were expanded using related semantic web and calculated in the collaborative filtering section. The result of combining these two sections was a hybrid suggestion system that could suggest pages to users [4].

In a 2020 paper, Sun et al. presented a way to improve web-based referral systems. The researchers believed that the growth of information on the Web would cause information to explode, resulting in inaccurate predictions. In the method proposed by these researchers, they tried to use the Internet of Things can be an effective technology for filtering useless information and can be used to recommend useful cases. The proposed mobile edge computer method is a new computing pattern through computing/resource pressure from remote cloud servers to network edge servers to provide smarter and more personalized services. The results of this study indicate the improvement of recommender systems with this proposed method [5].

In 2020, Idrissi and Zellou systematically reviewed recommender systems and discussed methods for dispersing information on the Web. The researchers proposed a new system by examining similar measures and different proposed methods. In the proposed method, criteria were provided to increase the accuracy of the recommendation in scattered data. The results show that scattered data are of great importance that can be collected by collecting this information and using this information by a recommending system can increase the accuracy of the results [6].

Waqh and Patil in 2019 use web mining techniques to personalize the web and recommend web pages. These techniques are used to find the relationship between web pages, the clustering stage, and the classification in data mining and data analysis methods [7]. The two researchers have modeled new measures for the relationship between pages, such as the distance matrix, the occurrence frequency matrix, and the relationship matrix. For the relationship between web pages, a virtual graph is created that fits the relationship matrix. By presenting an advanced search algorithm, they divide the virtual chart into different clusters, i.e infertility patterns. This method is a diagram-based algorithm. Using the LCS algorithm, the active user is classified into one of the clusters, and finally a threshold value is used to suggest only the optimal pages to the user.

Recommender system

Recommender systems have become very important in recent years. The goal of any offering system is for consumers to be able to find new goods or services, such as the web, books, music, restaurants, or even people, based on information about the consumer or the recommended item [8,9]. The recommender system is a system that, according to the user's preferences, recommends items jointly to a group of users [10]. Recommending systems are systems that help to find the user's interests in situations of over-information. Where the user's preferences are estimated based on the behavior observed in the past and can provide the user with a ranked list of suggestions [11].

Suggested method

In the proposed method, a way to improve web recommender systems is presented. Thus, we must first collect user registration files. The user registration file contains information such as the client's IP address, request time, requested URL, type of operating system used, date visited, and so on. The first step is to pre-process the data to prepare the data because raw data cannot be injected into data mining algorithms. The data must then be cleared, the cleanup operation performed to ensure that not all of the data used is suitable, and the extra data must be cleared. Next we need to normalize the data. Normalization means moving data from one suffering to another. The goal of normalization is to eliminate redundancy and maintain dependencies between data. After normalizing the data, they should be clustered. To cluster the data, we use the DBSCAN clustering algorithm. Then we have to weigh the clustered data. Weighing is done in order to determine what rating the user has given to the desired goods. We then calculate the similarity of the data using the Euclidean algorithm. In the next step, the data is categorized using the machine learning method and the pages are recommended to the target user using the recommender system. Figure 1, shows an overview of the proposed method.

Data preparation

In the proposed system, the data must first be prepared because the collected web data is usually large, very heterogeneous and unstructured. This data must be converted into consistent and integrated data in order to be useful for the pattern discovery phase. In this module, event log files are first collected from the desired servers and stored in the database for future review.

Data preprocessing

In the first step of the proposed method, we must first perform the data preprocessing operation. If different data are pre-processed, the same reliable and efficient performance will occur in all data sets [12]. In the data mining process, such as classification and clustering, we need to prepare the data for the algorithm, because it is usually not possible to inject data raw into data mining and machine learning algorithms. To prepare the data, it is necessary to take it out of its original form and state and transform it into a form that is suitable for the algorithm. Data preprocessing consists of 4 main steps. These steps are:

Data cleansing: At this point we need to clear the existing data. Data cleansing is the process of eliminating errors and inconsistencies in the data and is in fact the stage of quality control before performing data analysis. Data from real-world sources are often inaccurate, incomplete, and inconsistent due to operational errors and systems implementation. Such data needs to be cleared first. This includes some basic operations such as normalizing, eliminating noise or clutter, and dealing with missing data, reducing redundancy, and clearing data.

The three main phases defined for the data cleansing process are as follows:

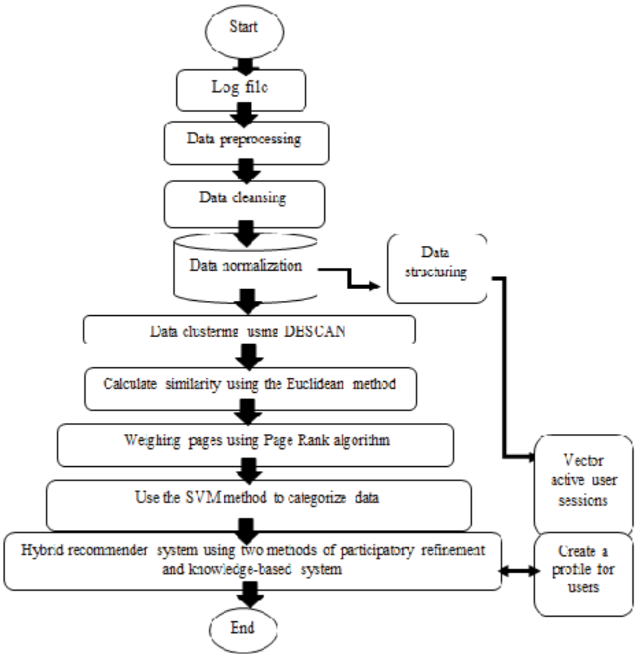

Figure 1: An overview of the proposed method.

- Define and determine the type of error
- Search and identify errors
- Correction of detected errors

Data integration: In the next step, we need to integrate our data. Data integration reduces integration costs and improves accuracy [13]. Integration is done due to the variety of sources and overlapping data in different sources, the variety of data storage and the ability to process different requests in different sources. In this step, the problems related to data conflict and redundancy are examined and resolved. Therefore, if the images are registered in different information sources, they need to be integrated.

Data reduction: Another step in data preprocessing is data reduction. The goal of data reduction is to achieve smaller volumes of data. One of the most important reasons for data loss is the high volume of data, which makes their analysis complex, time consuming and sometimes impossible. Extracting knowledge from large volumes of data takes a lot of time; therefore, it is necessary to use methods to reduce the size of the data. The purpose of the data reduction technique in data mining is to extract a small subset from a large volume of data while preserving the characteristics of the original data. This makes difficult or impossible data mining operations performed efficiently and effectively.

Data conversion: Operations such as data normalization, data modification and conversion are performed in this step, because the properties stored in information sources are raw data. That is, the features are designed and maintained to suit a specific area of work or the result of a specific system. This data is not suitable for processing and needs to be standardized to fit the project.

Data normalization: At this point we need to normalize the data, Change the data so that data normalization for them to be given such a narrow range between -1 and 1 are mapped. The goal of normalization is to eliminate data redundancy and maintain dependencies between related data. This process often results in more tables, but reduces database measurement and ensures improved performance. There are several methods for data normalization, the most famous of which is the Min-Max Normalization method. In this method, each data can be converted to an arbitrary interval. The general formula of this method for converting data to a range between 0 and 1 is as follows:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Data structuring: In this step, we have to perform the data structuring operation on the data obtained from the previous steps. By converting the normalized data from the log files into user sessions. User sessions represent the interests and behaviors of users that are used in the referral system to extract user behavior patterns. If a specific time is used to identify the sessions, the pages in which the specified time is displayed are considered as one of the user sessions.

Vector active user sessions: In order to cluster user information, we first convert it into a vector using techniques. The sum of these vectors forms a matrix. Each row of the matrix is a user, and in each column a page that the user has viewed is displayed with a number that indicates the number of times that page has been visited by the user. S_i The i-th session is a user defined as:

$$Si = < w(p1, si), w(p2, si, \ldots, \mathrm{w(pk, si)}, \ldots \mathrm{w(pn, si)} > \quad (2)$$

Is the number of web pages visited in all user access sessions? The k plane and represent the weight and number that should be in the k-m vector of the . This weight is calculated according to the frequency criterion. Frequency is the number of visits to a web page. Higher frequency pages are supposed to be more popular with users, which seems natural. The frequency of each page in each session is calculated using the following equation:

$$Frequency(\mathrm{page}) = \frac{numberofvisit\,\mathrm{sin}\,thesession(page)}{\sum numberofvisit\,\mathrm{sin}\,thesession(page)} \quad (3)$$

The face of this fraction represents the number of times a user visits a page in a given session. Its denominator represents the total number of web page views in the same session. Finally, all the vectors related to user access sessions are put together, and the next $m*n$ matrix consists of the weights of the web pages (m is the total number of user sessions). The rows of this matrix represent user sessions, and its columns represent pages that have been visited in various web sessions.

Using the DBSCAN clustering algorithm

In order to understand the DBSCAN algorithm, it is necessary to first introduce some of the definitions used in this algorithm:

The DBSCAN algorithm requires the determination of two parameters, Minpts and Eps. These two parameters are used to determine the minimum density of a cluster.

Definition 1: Neighbors of the radius Eps of a point: Neighbors in the radius of Eps A point like p, denoted by are a set of points whose distance from p is less than the radius of Eps, ie:

$$NEPS = S(\mathrm{P}) = \left\{ q \in \frac{D}{Dist(p,q)} \leq EPs \right\} \quad (4)$$

Definition 2: A central object is an object that has at least the number of Minpts of an object in the neighborhood of its Eps radius.

Definition 3: Direct density accessible is the point p direct density accessibility is from the point q if, firstly, p is part of the neighbors of the Eps radius of the object q and secondly, object q is an object q is a central object.

Definition 4: Density accessible, point p is density accessibility from point q if a chain of points is directly density accessible from pi.

Definition 5: Connected to Density, Point p is connected to density from point q.

Definition 6: Cluster: Assume that D is a database of points. Cluster C is a non-empty subset of D such that it satisfies the following conditions:

- For all pairs of points p and q if $p \in C$, i.e p is a member of cluster C (maximum condition).
- For all pairs of point's p and q, a density of q must be connected (connection condition).

Definition 7: Noise: Assume that $C_1, C_2, C_3, \ldots, C_n$ are clusters found in the D database. A set of points that exist in base D but do not belong to any of the clusters found are called Cie/k $Noise = \{ p \in D / \forall i : p \in Ci \}$

Definition 8: Marginal Object: A marginal object is an object that is not a central object but is densely accessible from another central object.

How the algorithm works?

How this algorithm works is that DBSCAN starts with a desired starting point that has not been visited. The range of this point is extracted using the epsilon distance (all points at distance are points of the same group or neighbors). It should be noted that the algorithm uses the Euclidean distance to find neighbors in a two-dimensional and three-dimensional space. Neighborhood is thus defined by the smallest distance from the principal point. If there are enough mini points in this range, the clustering process starts (border point) and the current data point becomes the first point of the cluster in the new cluster, otherwise the point is considered as noise (later this noise point may become part of the cluster). In both cases the point is marked as visited. For this first point in the new cluster, the points in the range are also part of a cluster. This method is used to construct all points in the group belonging to the same cluster and then is repeated for all new points that are only added to the cluster group. This process is repeated in steps 2 and 3 until all points in the clusters are entered, i.e all points in the ε range of the clusters are visited and labeled [14,15] (Figure 2).

The code of DBSCAN clustering algorithm is as follows:

Use the Euclidean method to calculate the similarity between clusters

Clustering is an automation process in which objects are divided into categories whose members are similar in terms of desired characteristics, so distance measurement is used to measure the similarity between data objects. There are different methods for measuring the distance between two objects, the Euclidean distance is the most famous and most widely used type of distance, which is calculated as the following relation:

$$d = \sqrt{\left(X_{i1} - X_{c1}\right)^2 + \left(X_{i2} - X_{c2}\right)^2 + \ldots + \left(X_{in} - X_{cn}\right)^2} \quad (6)$$

Weighing pages

At this point we need to weigh the resulting web pages. We do weighting using the page rank algorithm. In this method, it assigns a rating to each web document once and uses this rating, with or without considering a criterion according to the user's query to rank the documents. This algorithm ranks each page by assigning weight to the link given to that page. The amount of this weight depends on the quality of the page where the link is placed. In this case, the links of the more important pages gain more weight. To determine the quality of the referenced pages, the Page Rank uses the rank of that page, which is determined recursively and its initial value is optional. If n documents are available, the initial value of the document rank can be considered equal to $1/n$ .The rank of each page, such as P, is calculated according to the following formula: B_p is the set of all pages pointing to P. The output links are on the Q page. The rank of step j on page P_i is calculated according to the following formula:

$$r_j = \frac{\varepsilon}{n} + (1-\varepsilon)\sum_{Q \in B_{PI}} \frac{r_{j-1}(\mathrm{Q})}{Out\,\mathrm{dgree(Q)}} \quad (7)$$

Support vector machine classification method

In this step we have to categorize our data usingsvm. Support vector machine has been widely used in real applications due to its efficient performance in data classification [16]. The way svm works is that we assume that we have a set of data points and we want them to have two classes is a p-dimension vector of real numbers, which are essentially the same variables that represent software behavior. Linear classification methods try to separate data by constructing a superficial (which is a linear equation). The backup vector machine classification method, which is one of the linear classification methods, finds the best surface cloud that separates the data related to the two classes with maximum margin. In order to better understand the content Figure 3 shows an image of a data set belonging to two classes, in which the backup vector machine method selects the best super-surface to separate them.

In the Support vector machine method, we map the input vectors to a multidimensional space. After that, a surface cloud will be created that will separate the input vectors with the maximum possible distance. This super-surface is called the super-surface with the maximum separating boundary. As shown in Figure 3, two parallel surface clouds will be constructed on either side of the surface with a maximum separating boundary that will separate the data for the two classes in such a way that no data at the boundary between these two surface clouds. A surface cloud with a maximum separating boundary is a surface cloud that maximizes the distance between two parallel surface clouds. It is assumed that the greater the separator boundary or, in fact, the distance between two parallel surface clouds, the smaller the classification error.

Create a user profile

The recommender system is designed and implemented to anticipate interests and provide advice to users. In each of these systems, according to the scope of work and goals, a set of techniques for creating, updating and extracting data has been used, but the main axis in all these systems is the user profile. How to build the profile that will be used to build the recommendations, the default system profile for users, how to update the profile information and the source of this update are factors that play an important role in designing a recommender system. Suppose $S_1, S_2, S_3, \ldots S_k$ are a set of sessions related to the $i-th(ui)$ user. To create a user profile, the average Sui vector for the *ui* user is calculated as a representative and is actually a representation of the user's favorite pages. The weight of each web page in the vector is the average weight of that web page in all user sessions $S_1, S_2, S_3, \ldots S_k$.

DBSCAN Algorithm

```
D_unprocessed ← D          // D: data-set
no_of_clusters ← 0
while D_unprocessed ≠ ∅ do
    Arbitrarily select a p∈D_unprocessed
    if p is a non-core point, then
        ▪ Mark p as noise point
        ▪ D_unprocessed ← D_unprocessed - {p}
    else //p is a core point
        ▪ no_of_clusters ++
        ▪ D_DR(p) ← Determine all Density-Reachable points in D from p
          //note: The border points that may have been marked as noise, now belong to D_DR(p).
        ▪ Cluster_no_of_clusters ← {p} + D_DR(p)
        ▪ D_unprocessed ← D_unprocessed - Cluster_no_of_clusters
    end-if
end-while
```

Figure 2: DBSCAN algorithm code.

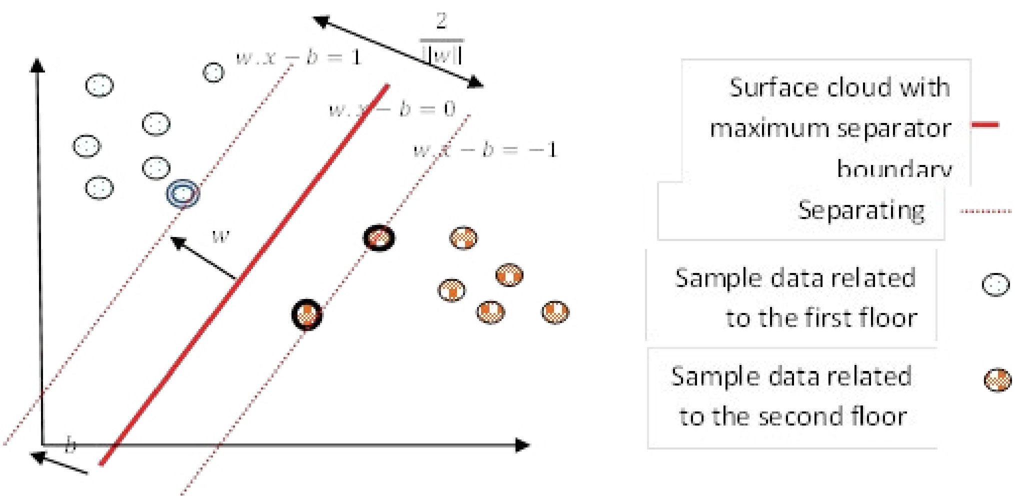

Figure 3: View of the SVM method.

Generate a list of predictions and suggestions to the user using a hybrid recommender system

In hybrid recommender systems, different techniques are combined based on a specific strategy to achieve the highest efficiency [17]. We combine the two algorithms of participatory refinement and knowledge-based refinement, the result will be a system that can overcome the cold start of the participatory refinement algorithm through a knowledge-based booklet and despite the component-based refinement component and its high power, finding similar priorities for users can make recommendations that no recommendation system will be able to build. Given that participatory and content-based refinement algorithms somehow face the problem of cold start, this problem occurs both due to insufficient data of items and due to insufficient data of users or both On the other hand, new items cannot be added to the recommended list until users rate them. On the other hand, new users who have not purchased and rated the items also have the problem that this problem can be eliminated by combining different recommender algorithms. However, the problem of cold start prevents new users from taking full advantage of content-based participatory algorithms. Systems based on participatory and content-based refinement algorithms are among the best systems for dedicated users who want to adapt and personalize the system over time based on their preferences and interests. Knowledge-based systems have less of a problem in this regard, because such systems do not pay attention to the user's background, and from their point of view, a new user is not much different from a user whose history of activities and purchases is registered in the system. Combined recommender systems are those systems that use a combination of one or more algorithms to achieve the highest performance. Research shows that hybrid systems are very successful systems.

Evaluate the results of the proposed method

It is often used to validate recommender systems, such as the efficiency of the clustering method, system accuracy, and calling. In this research, these criteria have been used to evaluate the system. Accuracy and recall in recommender systems are calculated using the following two equations.

Accuracy is calculated using the following equation:

The accuracy is equal to the number of correct system detections on the number of retrieved sets:

$$\Pr ecision = \frac{\left|\{\mathrm{re}levantpages\} \cap \{retrivedpages\}\right|}{\left|\{retrivedpages\}\right|} \quad (8)$$

The call is calculated using the following equation:

The call is equal to the ratio of the number of correct system diagnoses to the total number of standard system sets:

$$\mathrm{Re}call = \frac{\left|\{\mathrm{re}levantpages\} \cap \{retrivedpages\}\right|}{\left|\{\mathrm{re}levantpages\}\right|} \quad (9)$$

To evaluate the efficiency of DBSCAN algorithm, a comparison was made between this method and k-means algorithm. The results of performance evaluation indicate that the efficiency of the proposed clustering algorithm is about 0.99, while the K-MEANS clustering algorithm it obtained a performance score of 0.76, which indicates the efficiency of the DBSCAN algorithm compared to other clustering algorithms (Figure 4).

Also, in order to evaluate the accuracy and recall of the proposed method with other existing methods, a comparison was made between the proposed method and Knn, Rbf, Mlp algorithms Earn higher (Table 1) (Figures 5 and 6).

According to the obtained results, we conclude that the proposed method has obtained better results both from the clustering algorithm and from the two basic factors of accuracy and calling from other existing algorithms.

Figure 4: Performance diagram of DBSCAN and K-MEANS clustering algorithm.

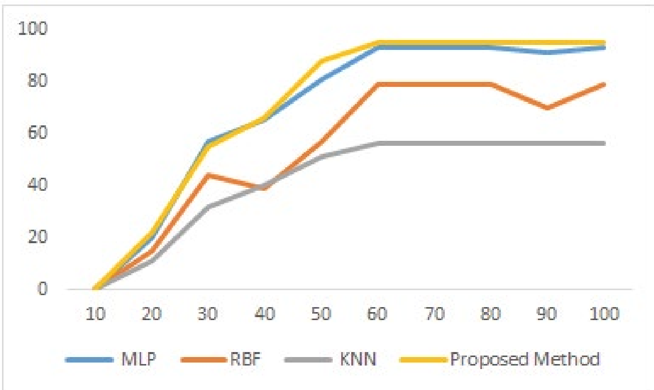

Figure 5: Comparison diagram of the proposed method call with other methods.

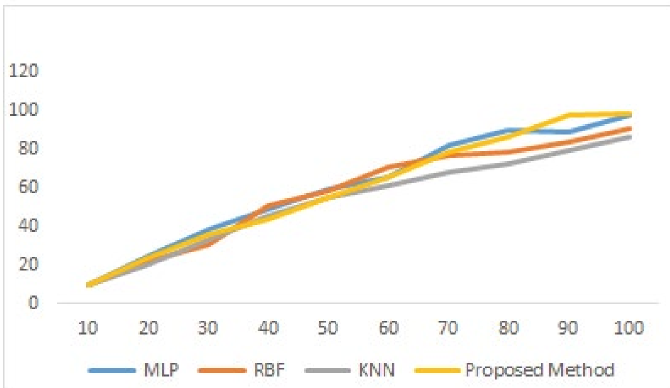

Figure 6: Diagram comparing the accuracy of the proposed method with other methods.

DISCUSSION AND CONCLUSION

Today, due to the increasing growth of web pages and the problems that arise for users when using the web, the existence of a system that can reduce the problems of users, it seems necessary. To reduce these problems, we need to customize the systems in question. There are many ways to customize search engines, but the most practical way to do this is to use referrer or recommender systems. Recommending systems are systems that, by taking limited information from the user, can make good suggestions to the user and help the user when faced with a cold start. By identifying users' behavior, these systems can discover their interests and use this information to offer appropriate suggestions to the user that may be of interest to him. In this research, a new method was proposed to improve the recommender systems in the field of web and it was tried to provide a way to improve the recommender systems by covering the problems and shortcomings of the previous systems. By comparing the proposed and studied methods and reviewing the results obtained with previous methods, the system has a more acceptable performance. Studies show that clustering using DBSCAN clustering algorithm, due to its features compared to other clustering algorithms such as K_MEANS and C_MEANS , can perform clustering with better performance to the extent that DBSCAN clustering algorithm achieves 99% but the clustering algorithm scored 76%. Also, according to the performance of the system and the results obtained, it was found that the hybrid recommender system used in the research had a stronger performance than other recommender systems that are used individually in terms of accuracy and recall. In the forecasting section, the use of the machine learning method discussed in the research SVM can be used in an acceptable way to predict the pages to the user. The general results of the experiment also indicate that the system used in the research in clustering, accuracy and suggestion of pages has a better and acceptable performance compared to other previous methods.

REFERENCES

1. Maazouzi F, Zarzour H, Jararweh Y. An effective recommender system based on clustering technique for ted talks. IJITWE. 2020;15(1):35-51.
2. Chawla S. Web page recommender system using hybrid of genetic algorithm and trust for personalized web search. J Infor Tech Res. 2018;*11*(2):110-127.
3. Bourkoukou O, Achbarou O. Weighting based approach for learning resources recommendations. JOIV: Int J Inform Visualization. 2018;2(3):104-109.
4. Riyahi M, Sohrabi MK. Providing effective recommendations in discussion groups using a new hybrid recommender system based on implicit ratings and semantic similarity. Electron Commer Res Appl. 2020; 40:100938.
5. Sun C, Li H, Li X, Wen J, Xiong Q, Zhou W. Convergence of recommender systems and edge computing: A comprehensive survey. IEEE Access. 2020; 8:47118-47132.
6. Idrissi N, Zellou A. A systematic literature review of sparsity issues in recommender systems. Soci Net Analysis and Mining. 2020;10(1):1-23.
7. Wagh R, Patil J. A novel web page recommender system for anonymous users based on clustering of web pages. AJCT. 2019;5(1).
8. Alashkar T, Jiang S, Wang S, Fu Y. Examples-rules guided deep neural network for makeup recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence. 2017;31(1) :941–947.
9. Gupta KD. A survey on recommender system. Int J Appli Eng Res. 2019;14(14):3274-3277.
10. Dara S, Chowdary CR, Kumar C. A survey on group recommender systems. J Intell Inf Syst. 2020;54(2):271-295.
11. Jannach D, Manzoor A, Cai W, Chen L. A survey on conversational recommender systems. arXiv preprint. 2020.
12. Alexandropoulos SA, Kotsiantis SB, Vrahatis MN. Data preprocessing in predictive data mining. Knowl Eng Rev. 2019;34.
13. De Giacomo G, Lembo D, Lenzerini M, Poggi A, Rosati R. Using ontologies for semantic data integration. In a comprehensive guide through the Italian database research over the last 25 Years. 2018;187-202.
14. El Bahi H, Zatni A. Document text detection in video frames acquired by a smartphone based on line segment detector and dbscan clustering. J Eng Sci Technol. 2018;13(2):540-557.
15. Siddharth Agrawal. Machine learning-DBSCAN. Toward Data Science-2019.
16. Hu R, Zhu X, Zhu Y, Gan J. Robust SVM with adaptive graph learning. World Wide Web. 2020;23(3):1945-1968.
17. Riyahi M, Sohrabi MK. Providing effective recommendations in discussion groups using a new hybrid recommender system based on implicit ratings and semantic similarity. Electron Commer Res Appl. 2020;40:100938.