Personalized Nutrition Recommendation For Diabetic Patients Using Improved K-Means And Krill-Herd Optimization

K.Renuka Devi, J.Bhavithra, Dr.A.Saradha

Abstract: In the growing world of rapid technology, Recommender system (RS) plays one of the significant roles in making decisions to the appropriate users. To maintain blood glucose in balanced level, there is a need of recommender system to recommend appropriate nutrition to those diabetic patients. An optimization technique plays a significant role in refining the attributes of appropriate algorithm to produce more optimized results to the user. The usage of recommender systems is to analyze the individual patient profiles to recommend the specific nutrition by means of collaborative filtering. The Patient's profile will get generated by analyzing thirty features for each of them. The Improved Krill Herd based optimization with Improved K-Means (IKH-IKC) system clusters those profiles using improved k-means clustering algorithm. To enhance the accuracy of recommendations, Improved Krill-Herd optimization algorithm has been applied. To validate the IKH-IKC idea, the experiment was carried out with 150 patient profiles. The efficiency of IKH-IKC system was analyzed with different metrics like Precision, Recall, F-measure, Accuracy, Matthews correlation, Fallout rate, Miss rate, Root Mean Squared Error (RMSE). The Experimental evaluation conveys that the IKH-IKC idea generates better clustering and optimized results to the user with low error rate.

Index Terms: Recommender systems, collaborative filtering, Improved K-means clustering, Improved Krill-Herd Optimization, Diabetes.

1 INTRODUCTION

Data mining is the technique inquiring huge amount of data in databases to provide relevant information and results to the users. In the context of recommender systems, data mining plays significant part in making recommendations based on the user interests [1]. Those systems make use of the knowledge gained by analyzing several attributes such as the user profiles, ephemeral data etc., The knowledge, which is utilized by several machine learning algorithms to provide accurate and relevant results to the users [1]. The Machine learning algorithms which include clustering, classification, regression, association rule mining. Recommender systems are developed to suggest users the items that suits with their interests and preferences [2]. RS makes recommendations generally in two such ways: Collaborative filtering or content-based filtering. The combination of two such approaches is called as hybrid approach [3]. The process of RS is entirely based on the rating or user profile generated for the users. The overview of recommender systems is shown in Fig.1

1.1 Collaborative filtering

Collaborative filtering is the most common approach for recommendation. It gathers the user's information and their interests using ratings, either they are explicitly provided or implicitly computed [3].

- *K. Renuka Devi, is currently pursuing her Master's degree program in Department of computer Science, Dr.Mahalingam College of Engineering and Technology, Tamilnadu, India, krenukagiri@gmail.com, 8838088448.*
- *J.Bhavithra, Assistant Professor (SS), Department of computer Science, Dr.Mahalingam College of Engineering and Technology, Tamilnadu, India, bavi@drmcet.ac.in, 8946032023.*
- *Dr.A.Saradha, Professor and Head, Department of Computer Science and Engineering, Institute of Road and Transport Technology, Erode, saradha.irtt@gmail.com, 9443170753.*

This method focuses on collecting the user's data, processes them and predicts the items based on the similarity with other users. The Collaborative filtering is independent of operational contents but it is capable of recommending items to the users without the knowledge of items in the collection. It calculates the similarity based on the rating of items and its attributes. This method is mainly used to compare the active user with similar users to make recommendations.

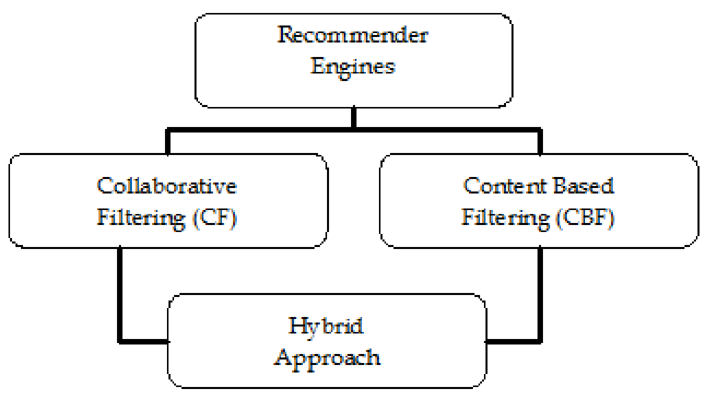

Fig 1.*Overview of Recommender Systems*

1.2 Content based filtering

Content based filtering is another approach to make recommendations. This method is used to recommend appropriate items to the users based on the user's preferred profile [4]. This type of filtering is mainly based on keywords. Those keywords were considered as the user's interest. Content based filtering make recommendations based on the items that are of user's interest or liked by them in the past. In the nutshell, this kind of filtering is used to recommend the items based on the ratings of user's interests by comparing with the items in the collection.

1.3 Hybrid Approach

The research in the field of Recommender systems results in a new approach called hybrid approach. The hybrid approach which combines the two techniques called Collaborative Filtering (CF) and Content-Based Filtering (CBF). This new approach can be implemented by making the predictions of two filtering and then combining them and then applying the functionalities of content based to collaborative method and vice versa or by combining them into a single model [4]. The hybrid approach performs better and also provides accurate recommendations than the pure collaborative and content-based method which was revealed by several studies. The common problems such as cold start and the sparsity problem were overcome by the hybrid approach.

2 NEED FOR NUTRITION RECOMMENDATION

Diabetes Mellitus occurs in humans due to high level of sugar or glucose in the blood. When the glucose level in blood increases (i.e.) after a meal, it initiates the secretion of hormone insulin from pancreas. This insulin triggers muscle and fat cells to remove glucose from blood, thus causing sugar level to be maintained in normal [4]. For the people with diabetes, their glucose level in blood is always high. This is due to the insulin produced may not be sufficient or it is produced in excess amounts. Two forms of diabetes which prevails (i) Type-1diabetes (ii) Type-2 diabetes. The former type is related to autoimmune disorder. The later type is associated with obesity. The efficient nutrition therapy plays a major role in improvising the health conditions of diabetic patients. The intake of adequate nutrition will lead to the prevention of diabetes, maintaining the blood glucose level, reduces the uneven secretion of blood insulin. Nutrition therapy associates with diabetic patients by proper intake of nutrition can help to reduce the blood glucose level, blood pressure and cholesterol [5]. Recommendation of nutrition in appropriate amount for diabetic patients results to (i) Attain and maintain optimal metabolic outcomes (ii) Prevention of chronic complication of diabetes (iii) Improves health through healthy food choices (iv) Facilitate by moderate weight loss or to prevent from weight gain.

3 RELATED WORK

Machine learning (ML) algorithms play a vital role in the recommender systems as it has the potential to improve the systems in terms of capacity. There are variety of algorithms prevails but the designer who makes decisions on the appropriate selections on those Machine learning (ML) algorithms as per the study by (Gunawardana & Shani, 2009) [5]. According to (Portugal et al.2015) ML is about computers which gains knowledge from the real world [6]. Based on that knowledge it can able to improve the accuracy and performance of the recommendations. The ML algorithms were classified into supervised, Semi supervised, unsupervised and reinforcement learning based on the learning type (Portugal et al.2015) [6]. Classification can be done on the training data which is few with several models. Unsupervised learning algorithms can be applied on those training data where it doesn't require any classified labels. The author (Feng Zhang et al. 2009) claims that unsupervised learning would be the better option for pattern extraction in web page recommendations [7]. To achieve better clustering quality, unsupervised learning plays a significant role in this aspect. Moreover, the author claims that it can increase the learning accuracy and reduces the memory overhead. DUE TO increase in the scaling of users and the items, which raises the problem of low efficiency and inaccurate recommendations while using collaborative filtering recommendation algorithms. The author (Xiaofeng Li et al.2019) IKH-IKC an improved hybrid collaborative filtering recommendation algorithm based on K-means which improves the efficiency of the recommendations [3]. Due to high scalability and efficiency, Partitional clustering algorithms raises an issue in pattern recognition. A new method called Manhattan Frequency k-means based on partitions was detailed by (Semeh Ben Salem et al.2018) which converts categorical data into numerical values [8]. K-means algorithm IKH-IKC by (Rui Máximo Esteves et al.2013) which addresses the problem of accuracy over large datasets. The study by (Zahra Nazari et al.2015) reveals that hierarchical clustering algorithm based on Euclidean distance which produces highly accurate recommendations.

From the study of (Ayon Dey 2016), the following conclusions can be drawn:

- Unsupervised learning would be better for the clustering and feature selection since it predicts from previously recognized class [9].
- K-means clustering would be suitable for large datasets and produces optimal clustering for making better recommendations [9].

Optimization of algorithm plays a vital role in improving the accuracy, efficiency, quality, performance and to reduce the computational time and cost. It is used to solve the problem in a most effective way. It is used to produce best results with least resources. The goal of optimization techniques is used to maximize the productivity and minimize the waste. The study by (Deepak Rai et al.2013) presents a new era of optimization techniques called Bio-inspired optimization techniques [10]. The behaviour of biological societies acts as a technique for resolving the optimization problems. As per the study of (Deepak Rai et al.2013), Krill Herd (KH) algorithm is based on processing of behavior of krill individuals [10]. It is based on the objective function of its krill movement. To make better krill behavior, the author claims for the inclusion of crossover and mutation. In order to solve complex problems, the author (Laith Mohammad Abualigah et al.2017) conveys that KH algorithm plays significant role [13]. To improve the search ability of the optimization problem, the IKH-IKC an algorithm with objective function called Improved Krill-Herd algorithm. This algorithm plays major role in clustering domain. The study of Laith Mohammad Abualigah (2018) shows that Improved Krill herd performs robust search to handle high dimensional data [13].

4 THRESHOLD RANKING WITH K-MEANS CLUSTERING (TR-KC)

4.1 Clustering Using K-Means Algorithm

K-means algorithm is the simplest and unsupervised Machine learning algorithms that plays a major role in solving the clustering problems. It is as simple as well as easiest way to categorize the given data through the defined number of clusters (K clusters).The core idea is to define the k-centers, for each cluster identified. It is used to find groups which have not been explicitly labeled (i.e.) unsupervised learning algorithm [15]. It is widely used for customer profiling, market segmentation, computer vision, geo-statistics, and astronomy. Initially, the data from dataset has been collected and preprocessed. Then neighborhood of the most similar users in the form of clusters has been calculated by applying K-means clustering Algorithm. The active user is classified based on the similarity between the particular user and a cluster center. To start with k-means algorithm, first cluster centroids (K) have been initialized. K-means is an iterative algorithm and it includes two steps such as (1) Cluster Assignment (2) Move centroid step.

K-means clustering algorithm:

1. Input: k (the number of clusters), D (a set of lift ratios)
2. Output: a set of k clusters
3. Method:
4. Arbitrarily choose k objects from D as the initial cluster centers;
5. Repeat:
6. (re)assign each object to the cluster to which the object is the most similar,
7. Based on the mean value of the objects in the cluster;
8. Update the cluster means, i.e., calculate
9. the mean value of the objects for each cluster
10. Until no change;

4.2 Recommendation Based on Threshold Ranking Algorithm

Feature selection techniques are utilized by the researchers to improve the execution of high dimensional data. The objective of feature selection is to remove the irrelevant or redundant features, which can then be discarded from the analysis [16]. Reducing the number of features in a data set can lead to faster model training and improved classifier performance. The two general categories for feature selection are filters and wrappers. Threshold based feature selection is a ranking technique based on filter which is an extension of FAST algorithm [16]. This technique selects the attribute subset which has significant predictive power. The algorithm starts with the normalization of attribute values ranging between 0 and 1 by mapping Fj to F^j. Those normalized values are considered as posterior probabilities. The attributes which are independent has paired with the class attribute and the two attribute data set collection. Then the maxima and minima values computed from the normalized values are extracted [16]. These values are given as input to the following function to calculate the final threshold of the input (glucose values) for the user from equation (1) [16].

$$th_j = \frac{F_j - \min(F_j)}{\max(F_j) - \min(F_j)} \tag{1}$$

Where,
thj = threshold.
Fj = input (glucose) values.
min(Fj) = minimum value of feature.
max(Fj) = maximum value of feature.

The users with minimum threshold will be selected and recommends the nutrition to the active user. The threshold based algorithm is good as any other algorithm because it produces optimal results based on the user's inputs.

Threshold ranking algorithm:

1. input:
2. Data set D with features Fj, j = 1,…, m;
3. Each instance x $\in$ D is assigned to one of two classes c(x) $\in$ {fp, nfp};
4. The value of attribute Fj for instance x is denoted $Fj(x)$;
5. Threshold-based feature ranking technique
6. $\omega \in$ {BFM, OR, PO, PR, GI, MI, KS, DV, BGM, AUC, PRC};
7. A predefined threshold: number (or percentage) of the features to be selected.
8. output:
9. Creates feature ranking R using $\omega_i(F^{\wedge j})\forall j$.
10. Selects features according to feature ranking R and a predefined threshold.

5 IMPROVED KRILL HERD BASED OPTIMIZATION WITH IMPROVED K-MEANS CLUSTERING (IKH-IKC)

The patient data from the dataset are collected and applied preprocessing so that individual patient records can be extracted. This includes the profile for each diabetic patients. The Improved K-means clustering technique clusters the patient's profile based on different attributes such as age, height, blood glucose levels etc., The similarity between active user query and other patients has been calculated using Improved Krill-Herd algorithm. Finally the recommendation results include the optimized results (nutrition). The IKH-IKC system follows the process of collecting patient's data from the dataset, processing them followed by clustering based on Improved k-means algorithm and finally generating recommendations based on Improved krill-herd algorithm. The IKH-IKC system has been shown in Fig.2

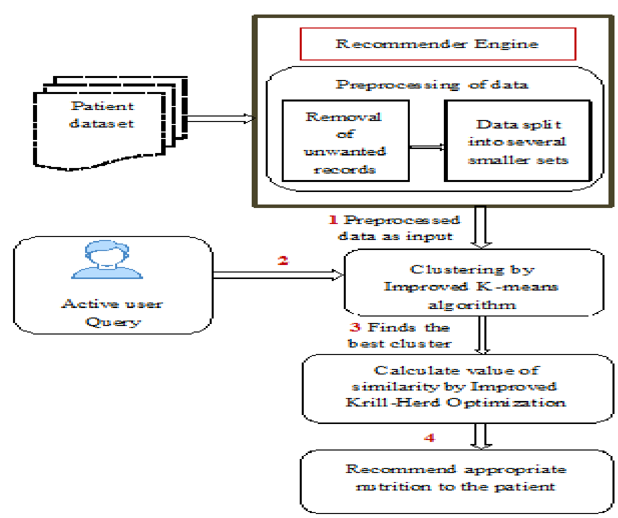

***Fig 2.** Improved Krill herd based optimization with Improved K-Means clustering (IKH-IKC)*

6 PREPROCESSING OF PATIENTS DATA

The appropriate profile for each patient has been generated based on the glucose level from the past history. The patient's record consists of thirty features where the first ten features indicate the generic characteristics of a patient. The succeeding twenty features associated with the glucose level in blood. Based on the similarity between individual user profiles, the significant nutrition acquired by the best match and accurate patient profile will be recommended [17]. To verify the correctness and effectiveness of IKH-IKC recommendation, different test cases were applied and the results have been obtained. The pre-processing was implemented by analyzing individual patient records. Those records were obtained from UCI Machine Learning Repository. The preprocessing has been done by the following steps:

1. The patient's data from dataset has been uploaded.
   - The data which includes the following details {date, time, code, value}.
   - They are cleaned and following steps are applied to pre-process them.
2. Change code with feature name.
3. Store each code in a file.
4. Extract the recent glucose level. From the preceding process, segregated user profiles generated.

7 IMPROVED K-MEANS CLUSTERING

The concept of clustering aims to combine the similar entities to a single group. This unsupervised learning methodology analyzes the data and discover the similarity between each data points and group the similar data into a single entity. This process of clustering is used for creating user profiles of different attributes which belongs to different groups. The clustering will minimize the dimensionality of data when the user deals with large number of attributes [17]. The clustering technique called Improved K-means which falls under the category of point assignment algorithms which is of unsupervised learning. This methodology aims to find the number of clusters dynamically. The initial partitions (centroids) have been calculated in a more significant way rather than random selection. This results in reducing the number of iterations. Thus, reducing the number of iterations increases the cluster quality but decrease in the number of empty clusters. The pre-processed data which includes individual patients profile that shows the blood glucose levels. Those processed data will be provided as input to the clustering technique called Improved K-means. The algorithm mainly consists of three phases. The first phase of algorithm finds the number of clusters (K value) dynamically [20]. The second phase of algorithm aims to calculate the initial partition which divides the data points into sub-arrays. The third phase thus calculates the final clusters. The distance between each data points will be computed and it is compared against the initial centroid, if the result of distance is equal or lesser, then the data remains in the same cluster or else it would be migrated to the most appropriate cluster [21]. This process will be repeated until there is no change in the mean for each cluster.

8 RECOMMENDATION THROUGH IMPROVED KRILL HERD OPTIMIZATION

Improved Krill-Herd (IKH) is a nature inspired algorithm which behaves by the nature of motion of krill individuals. It is one of the efficient algorithms as it solves different problems of optimization which is of different context. It is known as swarm intelligence based search algorithm because it imitates the herding behavior of krills. The IKH consists of specific objective function. This function computes the minimum distance of individual krill from the food and the density of the krill herd [27].

To model the behavior IKH algorithm, the following three operational processes will be considered,

1. Motion induced by other krill individuals.
2. Foraging movement.
3. Random physical diffusion.

The IKH algorithm is considered as a robust search methodology because it contains both exploration and exploitation methods and the movement generated by other krills. When the active user accesses the system, the appropriate cluster for active user has been identified [28]. To maintain the glucose level for active user, their features will be compared with each neighbors present within the cluster. The objective function [8] of IKH algorithm selects the appropriate user for each feature using equation (2).

$$\alpha_{target}^{i} = c^{best}\,\hat{k}_{i,best}\,\hat{x}_{i,best} \quad (2)$$

$\hat{k}_{i,best}$ = best objective function.

$\hat{x}_{i,best}$ = best position value.

The appropriate nutrition will be recommended to the active user by comparing the selected user's chemical composition with the default value.

9 RESULTS AND DISCUSSIONS

9.1 Data Set

Diabetes Data set has been collected from UCI Machine Learning Repository [9]. The data which includes several weeks to months' worth of outpatient care on 150 patients. It consists of Date, Time, Code and Value. Those data has been pre-processed based on value of each patient record. The current glucose level of each feature has extracted. For nutrition recommendation, the data has been collected from USDA National Nutrient Database [10]. The value of each record consists of nutrition and that nutrition contains the appropriate chemical composition.

- The data which includes the following details {date, time code, value}

i. The date is in the format MM-DD-YYYY.
ii. The time corresponds to XX.YY format.
iii. Code denotes numeric value assigned to each insulin measurements.
iv. Value denotes the corresponding glucose level for particular patient.

9.2 Experiment

To make recommendations, the experiment was carried out for the IKH-IKC system and it is compared with the TR-KC system. The TR-KC system utilizes K-means algorithm for clustering and threshold ranking algorithm for recommendation. The results obtained from IKH-IKC system perform better than the TR-KC system in terms of generating clusters and in making appropriate recommendations. The IKH-IKC system utilizes Improved K-means algorithm for clustering and Improved Krill-Herd algorithm for optimization. The utilization of optimization technique tends to produce more accurate results to the users.

9.3 Performance Evaluation

Several evaluation metrics has been carried out to measure the quality and performance of recommendation system. Through different kinds of users, the threshold ranking algorithm and Improved Krill-Herd optimization algorithm was compared. The evaluation metrics like Precision, Recall, F-measure, Accuracy, Matthews correlation are found to have higher values whereas Fallout rate, Miss rate, Root Mean Squared Error (RMSE) are found to be minimum for the IKH-IKC system. The performance of the system has been calculated using the following factors shown in Table 1.

Table 1. *Evaluation factors*

Evaluation Factors	Description
True Positive (TP)	The number of correctly recommended nutrition
True Negative (TN)	The number of correctly not recommended nutrition
False Positive (FP)	The number of wrongly recommended nutrition
False Negative (FN)	The number of missed correct nutrition

1) **Precision [15]:** The ratio of correctly recommended nutrition to the sum of correctly recommended nutrition and wrongly recommended nutrition from equation (3).

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

2) **Recall [15]:** The ratio of correctly recommended nutrition to the sum of missed correct nutrition and number of correctly recommended nutrition from equation (4).

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4}$$

3) **F-Measure [15]:** The harmonic mean of recall and precision. This is approximately the average of two metrics when they are close together. F – Score will be high only when the precision and recall is high from equation (5).

$$Fmeasure = \frac{2(precision * recall)}{precision + recall} \tag{5}$$

4) **Accuracy [15]:** The ratio of sum of correctly recommended nutrition to the sum of all the possibilities from equation (6).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

5) **Matthewscorrelation [15]:** The Matthews Correlation combines the informedness and markedness measures into a single metric by calculating their geometric mean from equation (7).

$$\text{Matthews correlation} = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FN)*(FP+TN)*(TP+FP)*(FN+TN)}} \tag{7}$$

6) Fallout rate [15]: Fallout or false positive rate is calculated as the ratio of recommended nutrition that are irrelevant to the total number of irrelevant nutrition from equation (8).

$$\text{Fallout rate} = \frac{FP}{FP+TN} \tag{8}$$

7) Miss rate [15]: The probability that a relevant nutrition is not recommended. Miss rate or false negative rate is calculated as the ratio of items not recommended but actually relevant to the total number of relevant nutrition from equation (9).

$$\text{Miss rate} = \frac{FN}{TP+FN} \tag{9}$$

8) Root Mean Squared Error (RMSE) [15]: It's the square root of the average of squared differences between prediction and actual observation from equation (10).

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^{\wedge}2} \tag{10}$$

Where,
n= total number of items
y_j = prediction value
$\hat{y}_j$ = Actual observation

9.4 Results

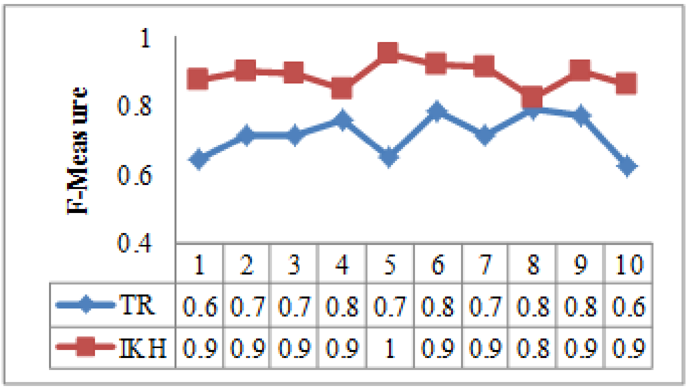

***Fig 3**. Comparison of F-measure values*

The experiment was carried out for 10 different users based on different attributes such as height, weight etc., The F-measure of the IKH-IKC system was observed to be higher than the TR-KC system from Fig.3 since the values of precision and recall are observed to be higher. It is due to the fact that it recommends more number of relevant nutrition to the user. As a result the performance metrics such as precision and recall have attained a higher value. The Fig.4 Shows the Accuracy graph for threshold ranking algorithm and Improved Krill-Herd algorithm. The result of Accuracy depends on all of the four evaluation factors such as TP, TN, FP and FN. From Fig.4 it is observed that the Improved Krill-Herd of IKH-IKC system has higher order of Accuracy values than the threshold ranking of TR-KC system.

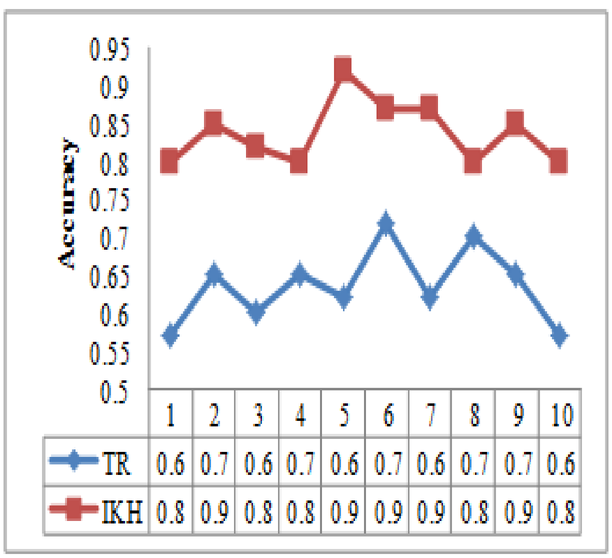

***Fig 4.**Comparison of Accuracy values*

The Fig.5 shows the Mathews correlation between the TR-KC and IKH-IKC system. Since this evaluation metric is based on Informedness and Markedness, the trustworthiness of positives and negatives should be higher. From the graph it is inferred that the IKH-IKC system has higher value than the TR-KC system. So, the IKH-IKC system is efficient for making recommendations.

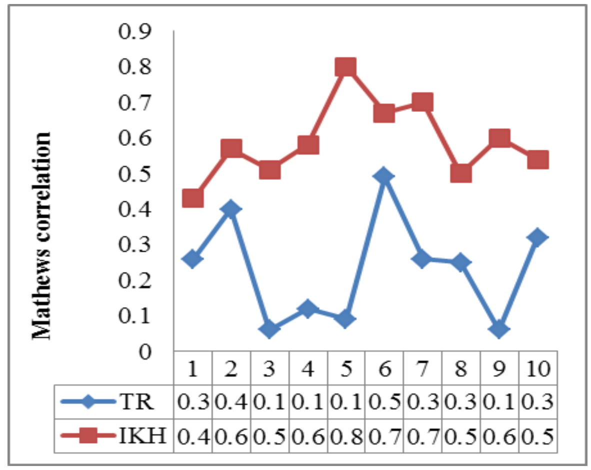

***Fig 5.**Comparison of Matthew's correlation*

The Fig.6 and Fig.7 shows the comparison of fallout rate and miss rate for the Threshold ranking and Improved Krill-Herd algorithm. From the graph it is inferred that the error rate such as fallout and miss rate found to be minimum for the threshold ranking and maximum for the Improved krill-herd algorithm. This shows that the IKH-IKC system has minimum false recommendations than the TR-KC system.

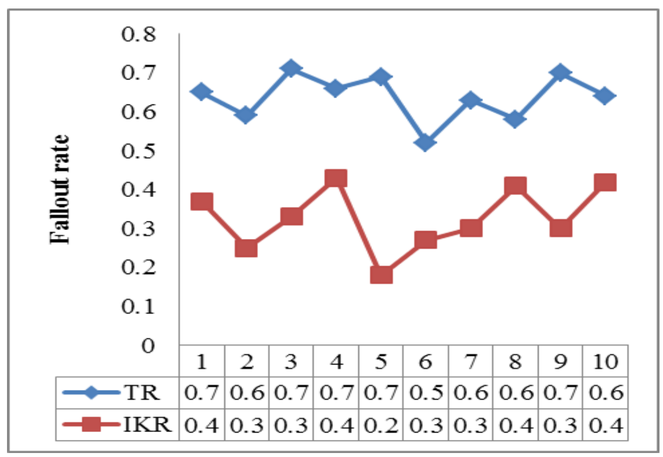

***Fig 6**. Comparison of Fallout rate*

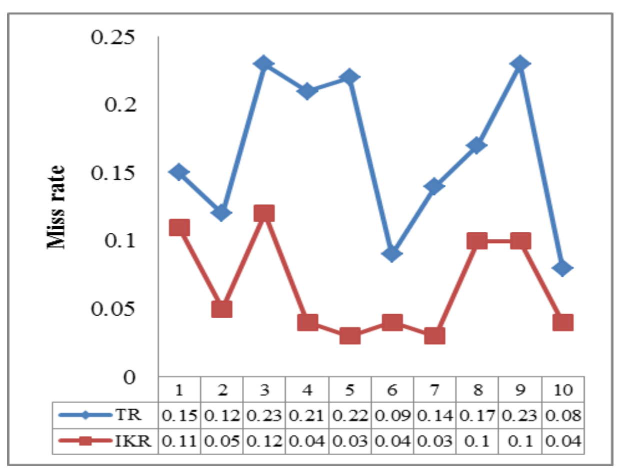

***Fig 7.** Comparison of Miss rate*

The Fig.8 shows the Root Mean Squared Error (RMSE) between Threshold ranking and Improved Krill-Herd algorithm. From Fig.8 it is clear that the RMSE of IKH-IKC system was lesser when compared with the TR-KC system. RMSE is the evaluation of error rate between actual observation and number of observations in the system. So, the IKH-IKC system has better observations to produce accurate results.

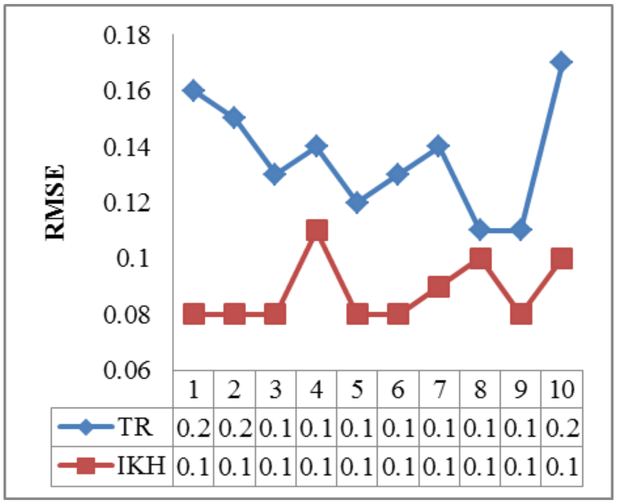

***Fig 8**. Comparison of RMSE*

10. CONCLUSION

In this paper, the recommendation system was developed for diabetic patients which utilize Improved K-means for clustering the data items and Improved Krill-Herd algorithm for optimization to produce more accurate results to the users. The quality and performance of the system was evaluated using various measures like Precision, Recall, F-measure, Accuracy, Matthews correlation, Fallout rate, Miss rate, Root Mean Squared Error (RMSE). From the results it is inferred that F-measure, Accuracy, Matthews correlation are found to be higher whereas the error values such as Fallout rate, Miss rate, Root Mean Squared Error (RMSE) are found to be minimum for the IKH-IKC system. This shows that the recommendations made by Improved Krill-Herd algorithm gives accurate results and also performs better than the threshold ranking algorithm.

11 REFERENCES

[1] Y. Koren, R.M. Bell, "Advances in Collaborative Filtering in Recommender Systems Handbook", Springer, vol. 1, issue. 3, pp. 45–186, 2011.

[2] Bharti KK, Singh PK, "Hybrid Dimension Reduction by Integrating Feature Selection with Feature Extraction Method for Text Clustering", Springer: Expert System Applications, vol. 42, issue. 6, pp. 3105–3114, 2015.

[3] Xiaofeng Li et al., "Improved Hybrid Collaborative Filtering Algorithm Based on K-Means", Elsevier: Expert Systems with Applications, vol. 842, pp. 928–934, 2019.

[4] Kunal Shah et al., "Unsupervised Recommender Systems: An Overview of Different Approaches to Recommendations", In the Proceedings of the 2017 IEEE International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-4, 2017.

[5] Shani G., Gunawardana A., "Evaluating Recommendation Systems", Recommender Systems Handbook. Springer, vol. 6, issue. 10, pp. 257-297, 2009.

[6] Portugal, I., Alencar, P.S., & Cowan, D.D, "The use of machine learning algorithms in recommender systems: A systematic review", Expert Systems Applications, vol. 97, pp. 205-227, 2015.

[7] Feng Zhang et al., "Unsupervised Feature Selection Based on Feature Relevance", In the Proceedings of IEEE Eighth International Conference on Machine Learning and Cybernetics, Baoding, pp.487-492, 2009.

[8] Semeh Ben Salema et al., "A Fast and Effective Partitional Clustering Algorithm for Large Categorical Datasets using a K-means Based Approach", Elsevier-Computers & Electrical Engineering, vol. 37, issue. 9, pp. 463-483, 2018.

[9] Ayon Dey, "Machine Learning Algorithms: A Review", International Journal of Computer Science and Information Technologies, vol. 7, issue. 3, pp. 174-1179, 2016.

[10] Deepak Rai, Kirti Tyagi, "Bio-inspired optimization techniques: a critical comparative study." ACM SIGSOFT Software Engineering Notes, vol. 38, pp. 1-7, 2013.

[11] Jinlong Hu et al., "A User Similarity - Based Top-N Recommendation Approach for Mobile In-Application Advertising", Elsevier-Expert Systems with Applications, vol. 111, pp. 51-60, 2018.

[12] Anand Khandare and A.S. Alvi, "Clustering Algorithms: Experiment and Improvements", Springer-Computing and Network Sustainability, vol. 25, issue.12, pp. 263-271, 2017.

[13] Abualigah, Laith Mohammad, Ahamad Tajudin Khader, Essam Said Hanandeh, and Amir H. Gandomi, "A Novel Hybridization Strategy for Krill-Herd Algorithm Applied to Clustering Techniques", Elsevier-Applied Soft Computing, vol. 60, pp. 423-435, 2017.

[14] Laith Mohammad Abualigah, Ahamad Tajudin Khader, Essam Said Hanandeh, "Hybrid Clustering Analysis using Improved Krill-Herd Algorithm", Springer-Applied Intelligence, vol. 48, issue. 11, pp. 4047–4071, 2018.

[15] M.V.B.T.Santhi et al., "Enhancing K-Means Clustering Algorithm", International Journal on Computer Science and Technology (IJCST), vol. 2, issue. 4, pp. 235-246, 2011.

[16] Huanjing Wang et al., "A Comparative Study of Threshold-based Feature Selection Techniques", In the Proceedings of 2010 IEEE International Conference on Granular Computing, vol 10, pp. 499-504, 2010.

[17] Dataset 1: Dua, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[18] Dataset2:http://www.ars.usda.gov/Services/docs.htm?docid=8964.
[19] Rao RV, Rai DP, Balic J, “A multi-objective algorithm for optimization of modern machining processes”, Engineering Applications of Artificial Intelligence, vol.61, pp.103–125, 2017.
[20] X. Chen et al., “K means: Automated two – level variable weigting Clustering Algorithm for Multiview Data”, IEEE transactions on knowledge and Data Engineering, vol. 25, pp. 932-944, 2013.
[21] G.Y. Zhang et al., “K-means: Two Level Weighted Collaborative K-means for Multiview Clustering”, IEEE: Knowledge Based Systems, vol. 150, pp. 127-138, 2018.
[22] Divya Pandove et al., “A Comprehensive Study on Clustering Approaches for Big Data Mining”, In the proceedings of IEEE Sponsored 2nd International Conference on Electronics and Communication System, vol.6, pp. 1333-1338, 2015.
[23] Jonathan Jeffrey et al., “Optimizing Web Structures Using Web Mining Techniques”, Springer: Intelligent Data Engineering and Automated Learning, vol. 4881, pp. 653–662, 2007.
[24] Daniel Kluver et al., “Rating-Based Collaborative Filtering: Algorithms and Evaluation”, Springer - Social Information Access, vol. 42, issue. 8, pp. 344-390, 2018.
[25] Gunnar Schröder, Maik Thiele, and Wolfgang Lehner, “Setting Goals and Choosing Metrics for Recommender System Evaluations”, In UCERSTI2 Workshop at the 5th ACM conference on recommender systems, Chicago, USA, vol. 23, pp. 53-61, 2011.
[26] X.T. Zhang et al., “Multi-Task Multi-View Clustering”, In the proceedings of IEEE Transactions on Knowledge & Data Engineering, vol. 28, issue. 4, pp. 3324-3338, 2016.
[27] Abualigah LM et al., “A Novel Hybridization Strategy for Krill-Herd Algorithm Applied to Clustering Techniques”, Springer: Applications of Soft Computing, vol. 60, pp. 423–435, 2017.
[28] J. Bhavithra, A. Saradha, “Personalized web page recommendation using case-based clustering and weighted association rule mining”, Springer, vol. 22, issue. 3, pp. 6991–7002, 2018.