International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

Whale Optimization Based Recommendation System

Bharti Sharma, Adeel Hashmi, Ankit Kumar

Abstract*: *Recommendation System is an information filtering system which seeks to predict the "liking" of a user for an item, with the aim to suggest the user those items which he/she is most likely to select/buy. The focus of this paper is on rating prediction whose main objective is to predict the ratings the current user is going to give to the items which are yet to be rated/viewed by him/her. This paper uses a collaborative filtering based approach for generating recommendation, and the model used is a clustering-based model. In this approach all the existing users are clustered using whale optimization technique, instead of traditional clustering approaches like k-means, EM algorithm, etc. The appropriate cluster is then identified for the active user, and the ratings of the active user are predicted based on ratings given by other users belonging to the same cluster. Different measures like MAE, SD, RMSE and t-value are used for performance analysis of the proposed method and the results obtained are found to be highly accurate.

Index Terms*: *Collaborative filtering, MAE, SD, RMSE, T-value, WCSS.

I. INTRODUCTION

The most popular approach used in recommendation systems [1] is the cluster-based collaborative filtering, in which the users are clustered based on the ratings provided by them in the past. The quality of the clusters generated has a significant impact on the recommendations generated. The k-means algorithm is the most commonly used algorithm for clustering but the drawback of this algorithm is that the authors have to provide the value of *k* (number of clusters) which usually leads to generation of poor quality clusters. Swarm intelligence algorithms have also been used in literature for generating high quality clusters [2]. Whale Optimization Algorithm (WOA) is one of the newly proposed algorithms belonging to the category of swarm intelligence. In this work, WOA has been used for generating the clusters for cluster-based recommendations. For clustering large datasets, tools like Apache Hadoop or Apache Spark can be utilized for developing a distributed version of the algorithm. So, WOA for clustering has been implemented on Apache Spark.

This paper is divided into the following sections: section-2 deals with the survey of the various approaches for generation of recommendations, section-3 presents the mathematical formulation of the whale optimization algorithm as well as the vanilla version of the algorithm, section-4 proposes a WOA-clustering based algorithm for generating recommendations, and section-5 compares the performance of the distributed version of the proposed algorithm with state-of-the-art.

II. RELATED WORK

The approach for generating recommendations can either be classified as content-based (feature-based) or collaborative filtering (ratings-based) as shown in Fig 1. A content-based approach identifies the features of items in the dataset (like genres of movies), and builds a user profile where weights are assigned to each feature based on his past interactions with the system, using machine learning algorithms like neural networks, decision tress, bayesian classifiers, etc. For collaborative filtering based movie recommendation, a "database of ratings" given to different movies by the users is maintained .To generate recommendations for the current user, the ratings given to different movies by this user are matched with the "database of ratings" to identify users with similar ratings pattern, and the movies which are given high ratings by identified users which are not yet viewed by the current user, are recommended.

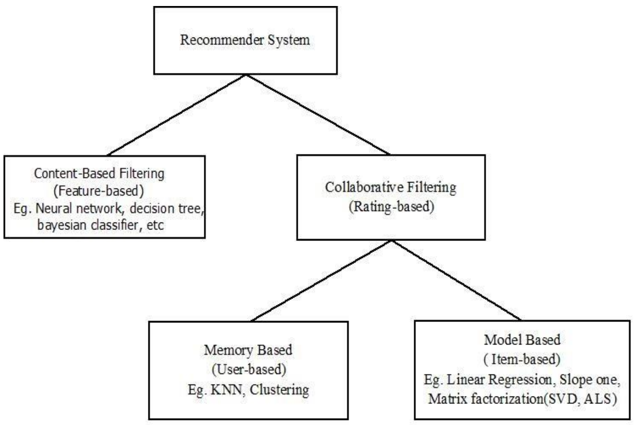

Fig.1: Recommendation Systems

The collaborative filtering approach can be classified as memory-based (user-user) or model-based (item-item). The user-user approach has been described earlier, where the users with similar rating pattern are identified for the current user. For calculating the similarity, a metric like pearson correlation coefficient or cosine similarity

(between the ratings of two users) may be utilized.

Manuscript published on 30 August 2019.
***Correspondence Author(s)**
Bharti Sharma, Department of Computer Engineering, BMU Rohtak, Haryana, India.
Adeel Shiraz Hashmi, Department of Computer Engineering, Jamia Millia Islamia, Delhi, India..
Dr Ankit Kumar, Department of Computer Engineering, BMU university, Rohtak, Harayna, India.

In item-item approach [3], the similarity between ratings of every pair of item is calculated and the items most similar to the item given high rating by the current user are recommended. The item-item approach is a model-based approach which uses a machine learning algorithm to predict ratings which a user can give to an item which he has not rated yet. Slope One [4] is an example of item-item collaborative filtering algorithm, which instead of linear regression (ax+b) uses Slope One predictor (x+b) to generate recommendations from the current user's rating vector. Matrix Factorization [5] is model-based collaborative filtering technique which uses a matrix factorization algorithm like SVD (Singular Value Decomposition) and predicts rating which a user will give to an item. The simplest approach for recommendation generation is memory-based (user-user) collaborative filtering. Algorithms like kNN (k nearest neighbors) may be utilized to generate recommendations. However, finding similarity of current user with rest of the users (which may be in millions) is a time-consuming task. To solve this problem, clustering technique may be used where the whole dataset may be partitioned into clusters and similarity is found only with the users belonging to the nearest (to the current user) cluster.Many authors have merged content-based and collaborative filtering resulting in a hybrid recommendation system [6]. Other than these traditional techniques many new types of techniques have emerged like context-aware recommendation systems [7], knowledge-based [8], location-based/mobile [9], etc. Swarm intelligence algorithms have also been utilized in literature for generating recommendations. Particle Swarm Optimization was used for optimizing the feature weights for content-based recommendations [10]. Ant Colony Optimization was used for choosing optimal clusters (generated by k-means clustering algorithm) to generate recommendations [11].

III. WHALE OPTIMIZATION ALGORITHM

Whales are believed to be intelligent as well as emotional animals. Whale's brain contains spindle cells which are similar to cells in the human brain. If we compare the amount of cells then a whale has twice the number of cells than a human does; this enables them to judge, think, learn and communicate. Humpback whales feed on small fish's herds and krill fish. The humpback whales follow an exquisite hunting method. This feeding method is called "bubble net method" [12]. They prefer to forage near the surface, feeding on school of krill and fishes. They create different paths (generally a spiral path) and bubbles along the path to feed, as shown in Fig. 2. Two types of movements are observed which are named as 'upward-spirals' and 'double loops' [13].

A. *Mathematical Formulation*

In this section, the mathematical formulation of Whale Optimization Algorithm (WOA) [14] is presented.

Bubble-net attacking method (Exploitation phase)

Let (X,Y) be the current position of the whale and (X*,Y*) be the position of the target prey. In the coordinate system, if (X,Y) is the current position and (X*,Y*) is the next optimal position (i.e. the nearest point), then the rest of the possible positions can be represented as shown in the Fig. 3.

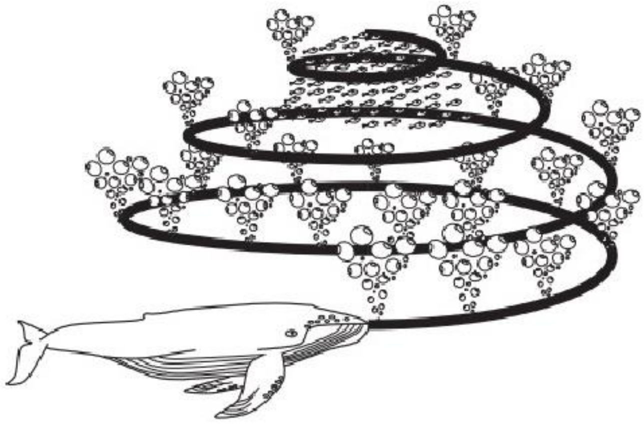

Fig. 2: Bubble-Net Movement Of Whale

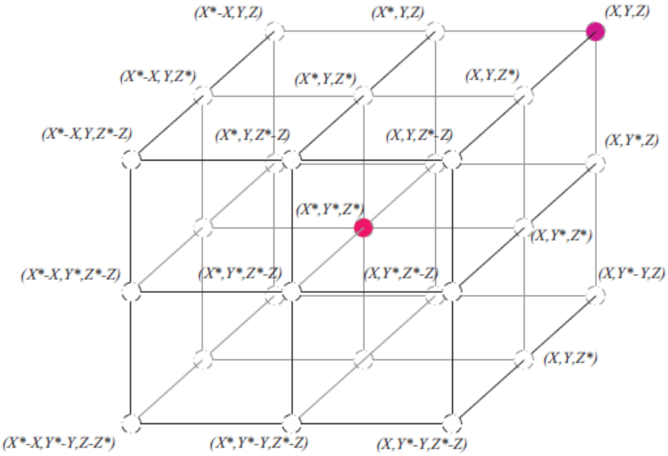

Fig. 3: 3D position vectors and their next possible locations (X* is the best solution obtained so far)

These positions can be modeled through the equation

$$X(t+1) = X^*(t) - \left|2r \cdot X^*(t) - X(t)\right| \quad , \qquad (1)$$
$$r \in [0,1]$$

Now this needs to be gradually shrunk to obtain encircling mechanism, for this variable A is introduced,

$$A = 2ar - a \qquad (2)$$

where, *a* decreases gradually from 2 to 0 and $r \epsilon [0,1]$, such that A gradually decreases with (a,r) from +2 to -2.
So the encircling equation becomes,

$$X(t+1) = X^*(t) - A \cdot \left|2r \cdot X^*(t) - X(t)\right| =$$
$$X^*(t) - (2ar - a) \cdot \left|2r \cdot X^*(t) - X(t)\right| \qquad (3)$$

The spiral movement of the whale is modeled by the equation

$$X(t+1) = X(t) + D' \cdot e^{br} \cdot \cos(2\pi r) \qquad (4)$$

where,
$D' = |X^*(t) - X(t)|$

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

indicates the distance of the i[th] whale to the prey (best solution found so far), r is a random number in [-1,1], and b is a constant for defining the shape of the logarithmic spiral.

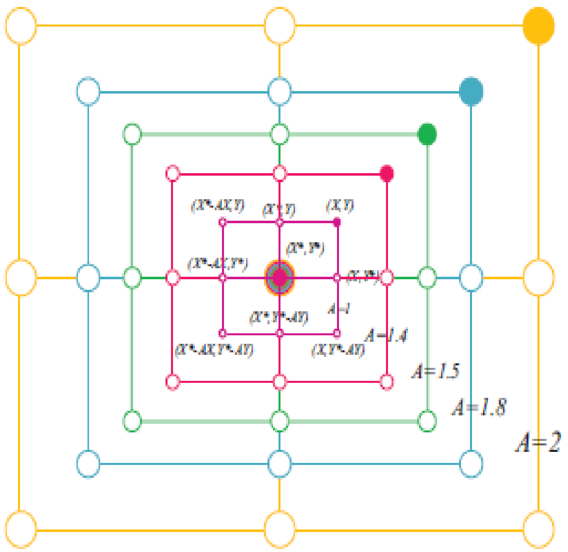

Fig. 4: Exploration Mechanism Implemented In WOA

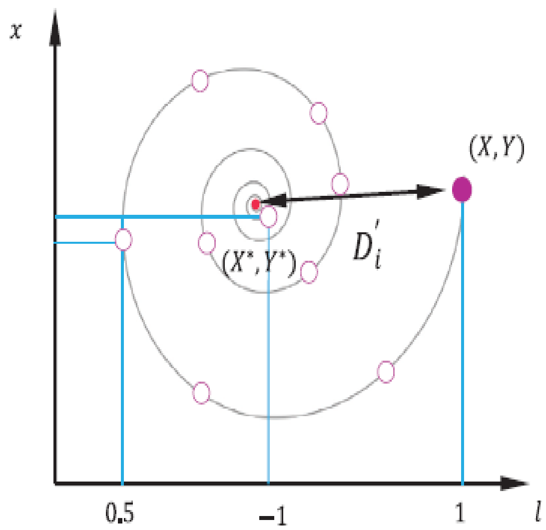

Fig. 5: Spiral Updation Of Position.

As the whale swims around the prey within a shrinking circle as well as along a spiral-shaped path simultaneously, we assume that there is a 50% probability to choose between the shrinking encircling mechanism or the spiral model.

$$x(t+1)=\begin{cases} X^*(t)-A\cdot D & \text{if } p<0.5 \\ D'\cdot e^{bl}\cdot\cos(2\pi l)+X^*(t) & \text{if } p\geq 0.5 \end{cases} \quad (5)$$

where p is a random number in [0,1].

Search prey (Exploration)

For the exploration phase, a random walk mechanism like Levy flight can be utilized. Levy flight is a special case of random walk (in which the step-lengths have a probability distribution that is heavy-tailed). Mantegna's algorithm can be utilized to generate a distribution having step length s that have the same behaviour of the Levy flights given by:

$$s=\frac{u}{|v|^{1/\beta}} \quad (6)$$

where, $u\sim N(0,\sigma_u^2)$
$v\sim N(0,\sigma_v^2)$

and $\sigma_u=\left\{\frac{\Gamma(1+\beta)\sin(\pi\beta/2)}{\Gamma[(1+\beta)/2]\beta 2^{(\beta-1)/2}}\right\}^{1/\beta}, \sigma_v=1$

B. *Whale algorithm*

The whale optimization algorithm is presented below:

Initialize the whale population, X_i (i=1,2,.....,n)
Calculate the fitness of each search agent
X^*= the best search agent
while(t<maximum number of iterations)
for each search agent
if1(p<0.5)
Update a ,A ,C ,I and p
if2 (|2ar-a|<1)
Update the position of the current search agent by the Eq. 1
else if2 (|2ar-a|>=1)
Select a random search agent (X_{rand})
Update the position of the current search agent by Eq. 6
end if2
else if1(p>=0.5)
Update the position of the current search by Eq. 4
end if1
end for
Check if any search agent goes beyond the search space and amend it
Calculate the fitness of each search agent
Update X^* if there is a better solution
t=t+1
end while
return X^*

IV. PROPOSED WORK

The authors propose to implement the whale optimization algorithm in Spark for providing fast recommendations through cluster-based collaborative filtering. The benefits of using Spark is that it is known to be faster than Hadoop MapReduce and it also supports stream processing out-of-the-box. The Spark implementation is expected to reduce the time complexity considerably with increasing cluster size.

Whale algorithm for generating optimal clusters

Initialize the whale population X_i (i=1,2,.....,n) where each whale consists of "k" cluster-heads. Generate clusters in each whale by k-means.
Calculate the fitness (within cluster sum of squared errors - WCSS) of each whale.
X^*= the fittest whale (i.e. minimum WCSS)
while(t<maximum number of iterations)
for each search agent
Update a, A ,C ,I and p
If1(p<0.5)
If2 (|2ar-a|<1)
Update the position of the current search agent by the Eq. 1
Else if2 (|2ar-a|>=1)
Select a random search agent (X_{rand})
Update the position of the current search agent by Eq. 6
End if2
Else if1(p>=0.5)
Update the position of the current search by Eq. 4
End if1
End for
Check if any search agent goes beyond the search space and amend it
Calculate the fitness of each search agent
Update X^* if there is a better solution
t=t+1
end while
return X^*

V. EXPERIMENTS AND RESULTS

The results are gathered after performing experiments on four datasets and evaluation using various metrics. The details of the datasets used are as follows:

A. MovieLens 100,000

A dataset obtained from the movie lens website composed in 7 months from September 97' to April 98'. The dataset as the name suggests consists 100k ratings from 943 users and for 1682 movies. The rating scale ranges from 1-5, 1 being the lowest and 5 being the highest.

B. MovieLens 1 million

This dataset comprises of four features which include userID, MovieID, Rating, and Timestamp. The user base is 6040 who have rated around 3900 movies with the total values being present reaching to 1000,209. Rating scale remains the same and every user has rated at least 20 movies.

C. Jester

In the Jester dataset, the items are jokes which are rated by the users. There are about 59132 users who in total have provided 1.7 million ratings for about 150 jokes. The rating scale though is different from others and ranges from -10 to +10. The feature contains user id, item id, and rating.

D. Epinion

This is an open source dataset made collected from the Epinion.com. It includes user id, item id, and rating. The rating scale remains 1-5. This dataset contains 664,824 ratings for 139,738 items from about 49,000 users.

The accuracy and quality of metrics provided is the only way we can estimate the performance of a recommendation system. To estimate the quality of predictions, recall and precision values are important factors. The values of these two metrics signify the number of relevant recommendations that have been provided successfully and the number of relevant information found in the number of retrieved recommendation. The metrics are described as follows:

$$\textbf{Recall}= \frac{Correctly\ Recommended\ items}{Relevant\ items} \tag{7}$$

$$\textbf{Precision} = \frac{Correctly\ Recommended\ items}{Total\ Recommended\ items} \tag{8}$$

Mean absolute error (MAE) metric is used to measure accuracy of recommendations, expressed as

$$MAE = \frac{1}{N}\sum_{I=1}^{n} | p_i - r_i | \tag{9}$$

It is used to calculate the deviation between predicted and actual ratings. In the above formula, p_i stands for predicted ratings and r_i stands for actual ratings. The effect of cluster-size on WOA is shown in Table 1.

A comparison between the proposed method and other clustering based recommendation methods is done in Tables 2-5. The methods chosen are Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Cuckoo Search Optimization (CSO), and Firefly Algorithm (FA).

Table 1: Performance of WOA based on cluster size

S. No	No. of clusters (k)	MAE	SD	RMSE	t-value	Recall	Precision
1	10	0.80	0.18	1.30	3.39	5.40	4.08
2	20	0.77	0.13	1.28	3.15	5.20	3.94
3	30	0.76	0.13	1.27	2.91	5.02	3.86
4	40	0.74	0.13	1.26	2.87	4.90	3.63
5	50	0.72	0.12	1.25	2.84	4.80	3.42
6	60	0.71	0.11	1.24	2.81	4.80	3.34
7	70	0.70	0.11	1.22	2.81	4.7	3.36

Table 2: Performance comparison with other algorithms (k=70) on MovieLens 100,000

	k-means	PSO	ACO	Firefly	Cuckoo	Whale
MAE	0.69	0.7	0.68	0.69	0.7	0.69
SD	0.11	0.113	0.112	0.112	0.113	0.112
RMSE	1.23	1.23	1.22	1.23	1.22	1.22
t-value	2.81	2.81	2.81	2.81	2.81	2.8
Recall	5.4	5.4	5	4.6	4.7	4.75
Precision	4.2	4.0	4.2	3.3	3.3	3.3

Retrieval Number: J93730881019/19©BEIESP
DOI: 10.35940/ijitee.J9373.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table 3: Performance comparison with other algorithms (k=70) on MovieLens 1 million

	k-means	PSO	ACO	Firefly	Cuckooo	Whale
MAE	0.69	0.8	0.7	0.69	0.7	0.69
SD	0.113	0.11	0.112	0.112	0.11	0.1
RMSE	1.23	1.2	1.22	1.23	1.19	1.22
t-value	2.81	2.75	2.81	2.81	2.75	2.8
Recall	5.4	5.1	5.3	5.6	5.5	5.6
Precisionn	4.0	3.8	4.8	4.0	4.1	4.1

Table 4: Performance comparison with other algorithms (k=70) on Jester DataSet

	k-mean	PSO	ACO	Firefly	Cuckoo	Whale
MAE	0.69	0.7	0.7	0.69	0.7	0.69
SD	0.113	0.113	0.112	0.112	0.11	0.1
RMSE	1.23	1.23	1.22	1.23	1.19	1.22
t-value	2.81	2.81	2.81	2.81	2.75	2.8
Recall	5.4	5.4	5.8	5.6	5.5	5.6
Precision	4.0	4.0	4.8	4.0	4.1	4.1

Table 5: Performance comparison with other algorithms (k=70) on EpinionDataSet

	k-me ans	PSO	ACO	Firefly	Cucko o	Whale
MAE	0.69	0.7	0.7	0.69	0.7	0.69
SD	0.11	0.11	0.11	0.11	0.11	0.11
RMSE	1.23	1.23	1.22	1.23	1.19	1.22
t-value	2.81	2.81	2.81	2.81	2.75	2.8
Recall	5.4	5.4	5.7	5.6	5.5	5.6
Precision	4.1	4.1	4.2	4.0	4.1	4.1

VI. CONCLUSION AND FUTURE WORK

This work proposes whale clustered algorithm for recommendation systems. The whale optimization algorithm was used to obtain optimal clusters. The optimal clusters can then be used for providing fast and relevant recommendations based on the choice of other users in the cluster. The performance of the proposed algorithm is compared with state-of-the-art algorithms using statistical measures like Mean Absolute Error, Standard Deviation, Root Mean Squared Error and t-value. The results obtained indicate that whale algorithm approach provides highly relevant recommendations. As far as future work is concerned, other nature-inspired algorithms can be used like multi-objective met heuristic algorithm in place of single objective algorithm.

REFERENCES

1. Resnick, Paul, and Hal R. Varian, "Recommender systems", Communications of the ACM 40.3 , 1997, pp. 56-58.
2. Zhongzhi, Shi, "A clustering algorithm based on swarm intelligence", Info-tech and Info-net, Proceedings, ICII 2001-Beijing, International Conferences, Vol. 3, IEEE, 2001.
3. Linden, Gregory D., Jennifer A. Jacobi, and Eric A. Benson," Collaborative recommendations using item-to-item similarity mappings", U.S. Patent No. 6,266,649 ,2001.
4. Lemire, Daniel, and Anna Maclachlan, "Slope one predictors for online rating-based collaborative filtering" ,Proceedings of the 2005 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, 2005.
5. Koren, Yehuda, Robert Bell, and Chris Volinsky, "Matrix factorization techniques for recommender systems" ,Computer 8, 2009, pp. 30-37.
6. Burke, Robin, "Hybrid web recommender systems" ,The adaptive web. Springer, Berlin, Heidelberg, 2007, pp. 377-408.
7. Adomavicius, Gediminas, and Alexander Tuzhilin, "Context-aware recommender systems", Recommender systems handbook, Springer, Boston, MA , 2011, pp. 217-253.
8. Aggarwal, CharuC.,"Knowledge-based recommender systems",Recommender systems. Springer, 2016, pp. 167-197.
9. Ye, Mao," Exploiting geographical influence for collaborative point-of-interest recommendation" ,Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, ACM, 2011.
10. Ujjin, Supiya, and Peter J. Bentley, "Particle swarm optimization recommender system", Swarm Intelligence Symposium, 2003, SIS'03, Proceedings of the IEEE, 2003.
11. Bedi, P., R. Sharma, and H. Kaur, "Recommender system based on collaborative behavior of ants" ,Journal of artificial intelligence 2.2 , 2009, pp. 40-55.
12. Watkins, William A., and William E. Schevill, "Aerial observation of feeding behavior in four baleen whales: Eubalaenaglacialis, Balaenoptera borealis, Megapteranovaeangliae, and Balaenopteraphysalus" ,Journal of Mammalogy 60.1 ,1979,pp. 155-163.
13. Goldbogen, Jeremy A ,"Integrative approaches to the study of baleen whale diving behavior, feeding performance, and foraging ecology", BioScience 63.2 ,2013,pp. 90-100.
14. Mirjalili, Seyedali, and Andrew Lewis, "The whale optimization algorithm", Advances in Engineering Software 95, 2016, pp. 51-67.

AUTHORS PROFILE

Bharti Sharma is a research scholar in Deparment of Computer Engineering of BMU Rohtak,Haryana, India. She has done his BTech. And M.Tech from Punjabi University Patiala, Punjab. She has over 9 years of teaching experience. Her areas of interest are Recommender System, Soft Computing, Swarm Intelligence, Data Mining.

Adeel Shiraz Hashmi is a research scholar in Deparment of Computer Engineering of Jamia Millia Islamia, New Delhi, India. He has done his BTech. And M.Tech. from GGSIPU, Delhi. He has over 7 years of teaching experience. His areas of interest are Artificial Intelligence, Soft Computing, Machine Learning, Swarm Intelligence, Data Mining, and Big Data. He has multiple publications in Scopus indexed conferences and journals.

Dr Ankit Kumar, Associate Professor in scholar in Deparment of Computer Engineering of BMU Rohtak,Haryana, India. He has 14 years teacher Experience. His areas of interest are Swarm Intelligence, Data Mining, Machine learning.

Retrieval Number: J93730881019/19©BEIESP
DOI: 10.35940/ijitee.J9373.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)