International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August 2019

Crop Prediction Analysis in North western Zone of Tamilnadu using Artificial Bee Colony with Weighted based Fuzzy Clustering

P.Surya, I.Laurence Aroquiaraj

Abstract: ***Agriculture is most important part for living people. Crop prediction analysis is more useful analysis to choose the particular crop for cultivation in particular season. In this paper, a hybridization technique that is Artificial Bee Colony with weighted based Fuzzy Clustering algorithm was proposed to predict a district which produces a most high yield in the north western zone of Tamilnadu in that particular district. North western zone of Tamilnadu consist four districts such as Dharmapuri, Krishnagiri, Namakkal and Salem. In this research work the proposed algorithm shows that the district in the north western zone of Tamilnadu which yields high production of particular crop for that particular season. In north western zone of Tamilnadu, Crop prediction analysis consider as high yield area, moderate yield area and low yield area as district wise. The Experimental result shows that hybridized artificial bee colony with weighted based fuzzy clustering algorithm yields better performance than other clustering algorithm like k-means and k-medoids with high accuracy.***

Keywords: ***Data mining, Clustering, Fuzzy clustering, Artificial Bee Colony, Agriculture, Crop prediction analysis.***

I. INTRODUCTION

Agriculture is a mainstay for the people living. In India, Agriculture is backbone of our country [10]. Above 55% of people is dependent on the field of agriculture as per the recent research news. In this agriculture field, it gives above 60% employed people as farmers from total population. It is almost 17% of GDP [8]. In worldwide scenario, the farm output of India was scored second rank. So that compared to the other domains, in the field of agriculture produce more data day to day. With the knowledge of proper diagnosis and analysis using various techniques such as data mining, machine learning, and big data analytics and so on, for this domain will increase the productivity and profitability from the intelligence techniques. It is much more help to the farmers and as well as the government.

Clustering is an unsupervised learning algorithm to group the similar and dissimilar objects within the group. It is a technique to collect the similar group of data objects.

Another term of the cluster is a technique used to partitioning or segmenting the data in to groups. As that clustered group might or might not be disjoint.

Revised Manuscript Received on August 30, 2019.

* Correspondence Author

P.SURYA*, Ph.D research scholar, Deptarment of computer science, Periyar University, Salem, Tamilnadu, India. Email: suriyaa14@gmail.com

Dr. I. Laurence Aroquiaraj, Assitant Professor, Deptarment of computer science, Periyar University, Salem, Tamilnadu, India.

The high quality of the cluster should contain good clusters with high intra cluster similarity and low inter cluster similarity.

Here, in this research paper mainly focused to fuzzy clustering method with bio inspired algorithm to developing the intelligent recommendation model for crop analyzing particularly for Tamilnadu north western zone.

The main contribution of this research paper:

- To hybrid the weighted based fuzzy clustering with artificial bee colony algorithm for agriculture crop analysis using Tamilnadu north western zone dataset.
- The proposed algorithm hybrid artificial bee colony with weighted based fuzzy clustering.
- This algorithm used to analysis the crop dataset and provides the result for district wise crop prediction analysis.

This paper organized as follows: section II explains about the related work already done by previous researcher from this particular agriculture domain. This section helps to motivate and eye opener for a new researcher. In Section III, it describes what are all the materials and methods used for this proposed work. Novelty method will be discussed in the proposed approach that is in section IV. From that proposed approach technique, various performance metrics and measures are discussed in the section V that is Experimental analysis. Section VI elaborates about the discussion and result part. Finally Section VII conclusion describes about concludes of the work with possible future enhancement.

II. RELATED WORK

This section discusses about various related works already done in data mining clustering techniques and also agriculture dataset related work.

Anjana Gosain, Sonika Dahiya, observed the research study on various fuzzy clustering algorithms to find out the performance of those algorithms. Researchers particularly focused on FCM, PCM, PFCM, FCM-σ, DOFCM, T2FCM, KT2FCM, IFCM, KIFCM, IFCM-σ, KIFCM-σ, NC, CFCM fuzzy clustering algorithms [7]. They used standard data points with present of outliers and noise for better performance analysis. Researchers used standard dataset for implementation. From various fuzzy clustering algorithms, DOFCM yields the best performance for their dataset.

B M Sagar, Cauvery N K, observed the research study on crop yield prediction in agriculture. And also researchers designed the model for crop yield prediction through machine learning, data mining techniques and big data analytics [8].

D Ramesh, B Vishnu Vardhan, researchers build a prediction for crop yield using multiple linear regression technique and density based clustering technique [1] for the region of East Godavari district of Andhra pradesh in India. They used 40 years interval of sample data for that particular East Godavari district. The results achieved by multiple linear regression technique was -14% and +13% and the density based clustering technique produce -13% and +8% for 6 clusters.

T.Rajasekaran, P.Jayasheelan, K.R. Sri Preethaa focused on predictive analytics for crop production in agriculture using ZeroR algorithm [6]. By their research, they find out that prediction method should help to the farmers to improve the cop productivity and also gives an idea about too high amount of supplements of cultivation, and to decrease the usage of pesticides and fertilizers. In their prediction model, they used different data like weather condition, soil type, humidity, air quality, crop maturity, labor costs. Research work carried under the technique of machine learning.

D. Karaboga, B. Basturk, researchers compares the Artificial Bee Colony algorithm with differential evolution (DE), evolutionary algorithm (EA) and Particle Swarm Optimization (PSO) for multi-dimensional and multimodal numeric problems [5]. From the simulation results Artificial Bee Colony produce the best results than other algorithms such as DE, EA and PSO with high dimensionality to solve the multimodal engineering problems.

Shailesh Pandey, Sandeep Kumar proposed enhanced artificial bee colony and its application to travelling salesman problem [2]. Researchers added the linear crossover operator to the ABC and its application to travelling salesman problem. From their simulation and experimental analysis, compare than ABC without crossover, enhanced ABC with linear crossover produce a better accuracy.

P. Vinciya, A. Valarmathi researchers focused on crop yield prediction using multiple linear regressions [6] for high tech farming. Their research mainly focused on analyzing the organic farming and inorganic farming, plant cultivation time on farming and also they concentrate on Gross Domestic Product (GDP) and Terms of Trade (TOT) to identify the agriculture growth.

In this literature study shows that many clustering algorithm was proposed for various dataset. Many researchers focused on crop yield prediction in agriculture and they will not produce any district wise crop prediction analysis. To overcome this kind of problem the proposed algorithm works with agriculture dataset for district wise crop prediction analysis for north western zone of Tamilnadu.

III. METHODS AND MATERIALS

This section describes the methods and materials needed for this research work. In section A explains about fuzzy clustering and section B explains Artificial Bee Colony algorithm.

A. Fuzzy clustering

The other name of the Fuzzy clustering is known as soft clustering techniques. Fuzzy Clustering techniques can be applied to data that are quantitative that is numerical, qualitative that is categorical and or a mixture of both data (numerical and categorical). In this method, it allows simultaneously all the objects to belong to several clusters at the time. Compare to the hard clustering, fuzzy clustering is most natural.

Fuzzy C- Partition:

In ordinary (usual) C-partition (C≥2, C is an Integer) of a set S= $\{u_1,\ldots x_n\}$ represented as, $P(A_{1,}{}^{A2,\ldots An})$.

Definition:

$A_i \neq \phi \quad \forall_i = 1, 2, \ldots C$

$A_i \cap A_j \quad \forall_i \neq j$

$\cup_i^c Ai = s$

Fuzzy c-partition of s, represented by (U,S) where (U_{nxc}). U is an nxc matrix. U=$((U_{ij}))_{nxc}$ where, U_{ij} denotes membership value of ith point to the jth fuzzy set; naturally $1 \le i \le n$, $1 \le j \le c$.

Objective Function:

$U_{ij}^r V_i$=mean of A_i; i= 1,2, …c

$$\sum_{i=1}^{c} \sum ||x - v_i||^2$$

Let r>1 let $v_j = (\sum_{i=1}^{n} x_i u_{ij}^r)/(\sum_{i=1}^{n} u_{ij}^r)$; j=1,2,…. C.

$J_r(U,S,A) = \sum_{i=1}^{n} \sum_{j=1}^{n} u_{ij}^r (x_i - v_j)\, A_{mxn} (x_i - v_j)$; A is a positive definite matrix. Where r is a exponent term. J_r is a objective function.

Steps for Fuzzy c-means Algorithm:

1. Let assume S,A,r and the number of clusters C.
2. Start with a fuzzy c-partition U of S.
3. Compute v_j, j=1,2,…c
4. $U_{ij} = \left(\sum_{k=1}^{c} \left(\frac{d_{ij}^2}{d_{ik}^2} \right)^{\frac{1}{r-1}} \right)^{-1}$; $1 \le i \le n$, $1 \le j \le c$

   $d_{ij}^2 = (x_i - v_j)\, A (x_i - v_j)$
5. Go to 3 if the convergence criterion is not satisfied.

B. ABC Algorithm

Artificial Bee Colony (ABC) algorithm is wide used algorithm developed by Dervis Karaboga, in the year 2005. It is the bio inspired algorithm motivated by the intelligent behavior of the honey bees. Artificial bee colony as an optimization technique, it provides a population-based search procedure that is swarm (group) in which individuals called foods positions. Those are modified by the artificial bees with time. And the bee's aim is to find out the places of food sources with high nectar amount, finally the one with the highest nectar.

ABC algorithm consists of four main phases to develop. They are: Initialization phase, Employed Bee phase, Onlooker bee phase and Scout phase.

i). Initialization Phase

The initial food sources are randomly produced using this following expression x_{mi}

$$x_{mi} = l_i + rand(0,1) * (u_i - l_i) \text{ ---------- } (1)$$

in (1), where u_i is the upper bound and the l_i is the lower bound of the solution space of objective function and rand(0,1) is the random number range within [0,1].

ii) Employed Bee Phase

The following expression is used to calculate the neighbor food source υ_{mi} is determined

$$\upsilon_{mi}=x_{mi}+\phi_{mi}(x_{mi}-x_{ki}) \text{ -------- (2)}$$

In (2), where i and x_{ki} is correspondingly denotes as a randomly selected parameter index and randomly selected food source. ϕ_{mi} denotes the random number within the range of [-1,1].

Here, the following expression for Fitness function should be calculated after the processing of greedy selection is applied between x_{mi} and $v_{mi.}$

$$Fit_m\ (x_{mi})= \frac{1}{1+f_{mi}(x_{mi})},\ f_{mi}(x_{mi})>0 \text{ and}$$
$$Fit_m\ (x_{mi}) = 1+ |\ f_{mi}(x_{mi})|,\ f_{mi}(x_{mi})<0 \text{ ---------- (3)}$$

In (3), where, $f_{mi}(x_{mi})$ is the objective function value of $x_{mi.}$

iii) Onlooker Bee Phase

In ABC algorithm, Onlooker Bee Phase is used to identify the quantity of food source is evaluated by its profitability. The following expression

$$P_m=\frac{fit_m(x_{mi})}{\sum_{m=1}^{SN} fit_m((x_{mi})} \text{ ------- (4)}$$

In (4), where $fit_m(x_{mi})$ denotes the fitness of x_{mi}. In onlooker bee phase, the following expression of (5) is used for find the neighborhood of food sources.

$\upsilon_{mi}=x_{mi}+\phi_{mi}(x_{mi}-x_{ki})$ --------- (5)

iv) Scout Bee Phase

The job of the unemployed bees (i.e Scout bees) is randomly searching the new solutions. To discover the new solution x_m by the scout was expressed as follows,

$$X_m=l_i+rand(0,1)*(u_i-l_i) \text{ ---------(6)}$$

In (6), where rand(0,1) is denotes as a random number from the range of 0,1. And u_i is the upper bound and l_i is the lower bound of the solution space of objective function.

IV. PROPOSED APPROACH

This research paper, crop dataset was collected from government website portal. The raw structured dataset should be preprocessed for data analysis. Particle filtering technique was taken for preprocessing the data. This preprocessing technique mainly focused for the data redundancy. After the preprocessing, k-means and k-medoids clustering algorithm was implemented to get the district wise crop data. Here, the k value assigned as 2 for both algorithm. Then the artificial bee colony with weighted based fuzzy clustering algorithm implemented for the crop dataset to predict the district which produce the high production of a particular crop in particular season. The implementation work will be carried out through Matlab. The below figure 1 explains about implementation steps.

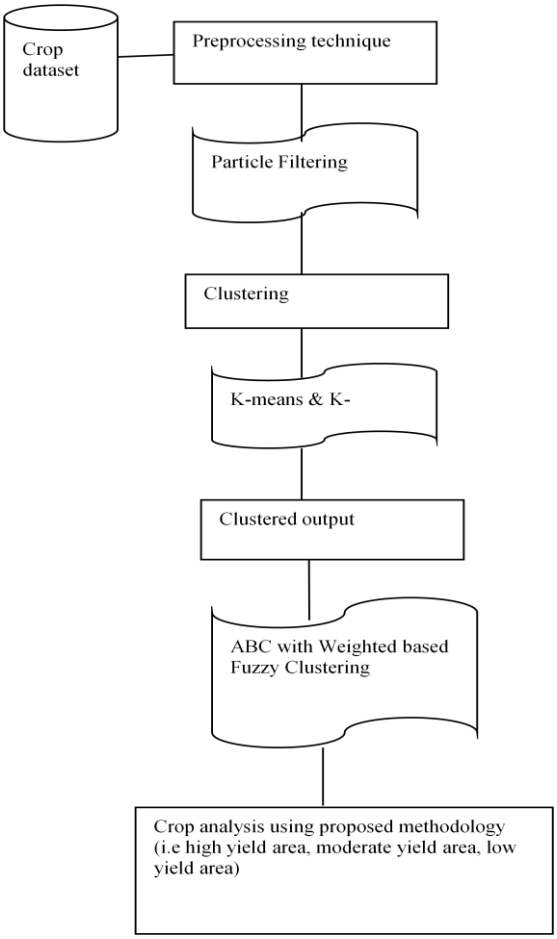

Fig: 1. Proposed approach

V. EXPERIMENTAL ANALYSIS

This section discuss about the experimental analysis of implemented algorithms of K-means, K-medoids, Artificial Bee Colony with Weighted based Fuzzy Clustering for agriculture crop dataset to analysis the district wise crop prediction (i.e High Yield Area, Moderate Yield Area, Low Yield Area) particularly for north western zone of Tamilnadu.

Dataset description:

The dataset was collected from India governmental agriculture portal. Dataset description was explained in below table 1 and attributes description was explained in below table 2.

Table: 1. Dataset description

Dataset name	Number of Attributes	Number of Instances
Agriculture crop dataset	7	13,457

Table: 2. Attribute description

Attributes	Description
Year	Contains data from the year of 1997 to 2015
State name	Contains State names in India
District name	Contains 32 District name in Tamilnadu
Season	Contains 3 types of seasons
Crop name	Contains 124 types of crops
Area	Contains range of cultivated area
Production	Contains production range of crops.

The following figures 2, 3, 4 correspondingly represent the result of K-means, K-medoids and artificial bee colony with weighted based fuzzy clustering algorithms works with this above agriculture crop dataset.

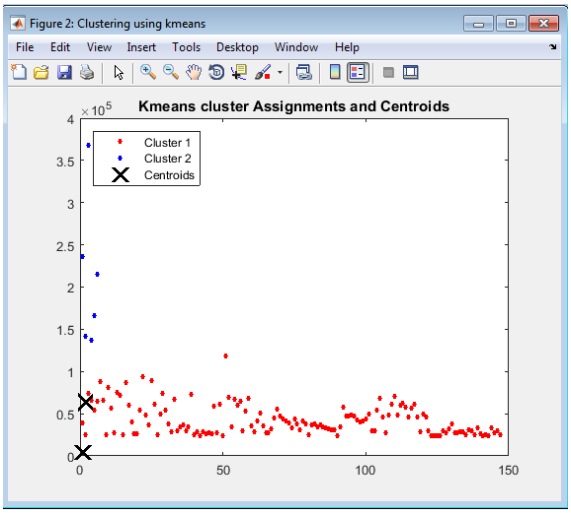

Fig: 2. Result of K-means

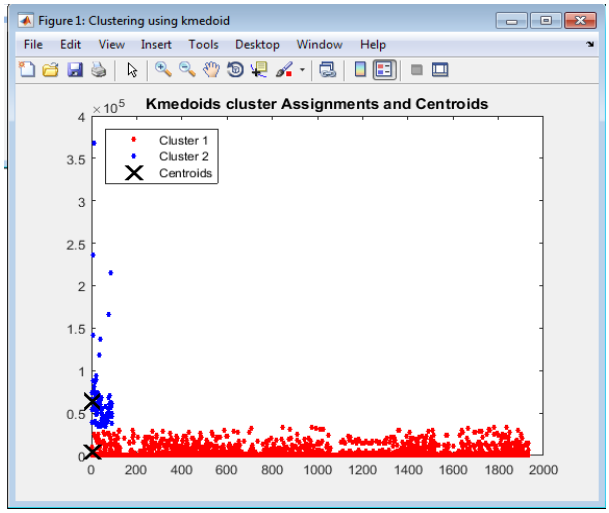

Fig: 3. Result of K-medoids

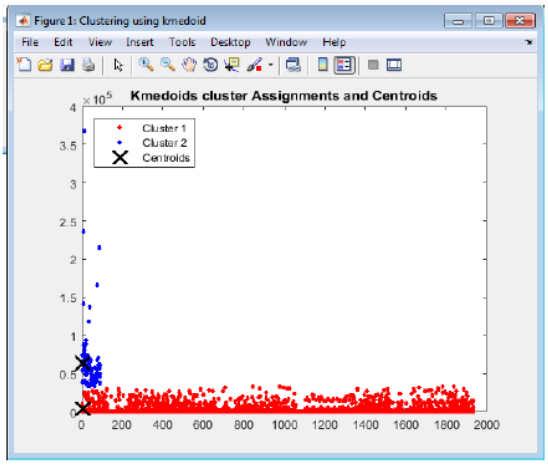

Fig: 4. Result of ABC with Weighted based Fuzzy Clustering

Table: 3. Performance metrics

Performance Metrics	Accuracy	Precision	Recall	MAE	RMSE
K-Means	83%	74%	74%	1.356	0.987
K-Medoids	86%	77%	75%	1.0298	0.896
ABC with Weighted based Fuzzy Clustering	95.87%	92%	90%	0.05678	0.035627

The above table represents the performance metrics for an algorithm of K-Means, K-Medoids and ABC with Weighted based Fuzzy Clustering works with agriculture crop dataset. The computational works referred with the performance metrics of accuracy, precision, recall and Mean Absolute Error (MAE), Root Mean Square Error (RMSE) for error rate. K-Means algorithm produces the results with 83% of accuracy, 74% of precision, 74% of recall and the error rate 1.356 of MAE, 0.987 of RMSE. K-Medoids algorithm produces the results with 86% of accuracy, 77% of precision, 75% of recall and the error rate 1.0298 of MAE, 0.896 of RMSE. Proposed ABC with Weighted based Fuzzy Clustering algorithm produces the results with 95.87%of accuracy, 92% of precision, 90% of recall and the error rate 0.05678of MAE, 0.035627of RMSE. The following fig 5 and 6 respectively represents the chart representation for accuracy, precision, recall and chart representation for Mean Absolute Error (MAE), Root Mean Square Error (RMSE).

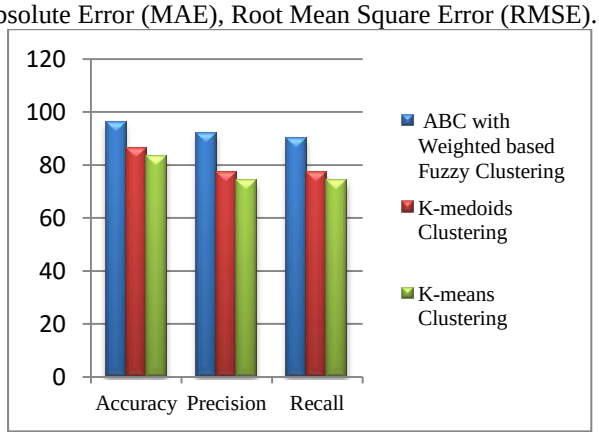

Fig: 5. Chart representation for Accuracy, Precision & Recall

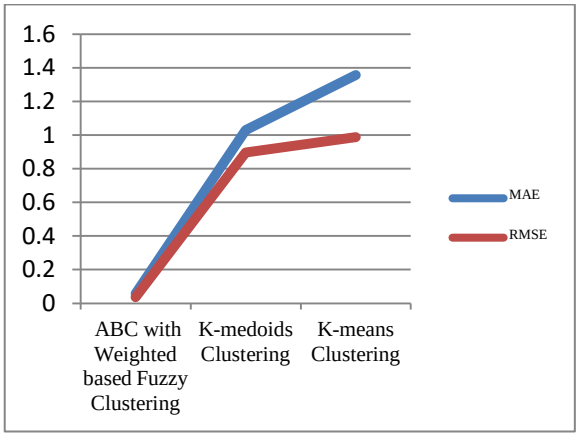

Fig: 6. Chart representation for Mean Absolute Error (MAE) & Root Mean Square Error (RMSE)

VI. RESULT AND DISCUSSION

In this result section discuss about the results of crop yield prediction analysis using ABC with weighted based fuzzy clustering based on agriculture crop dataset particularly for north western zone of Tamilnadu to predict the district wise crop yield. The following figures 7, 8, 9 are shown respectively season selection, crop selection and district wise crop yield prediction analysis in north western zone of Tamilnadu.

Fig: 7. Season selection

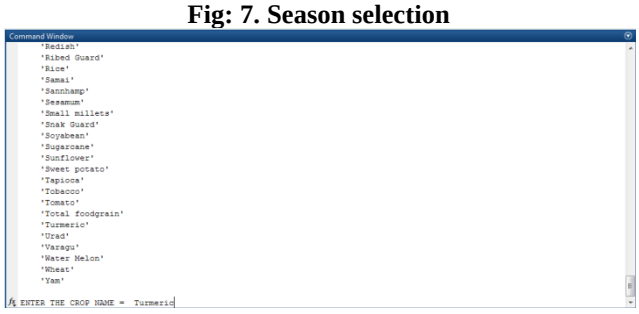

Fig: 8. Crop selection

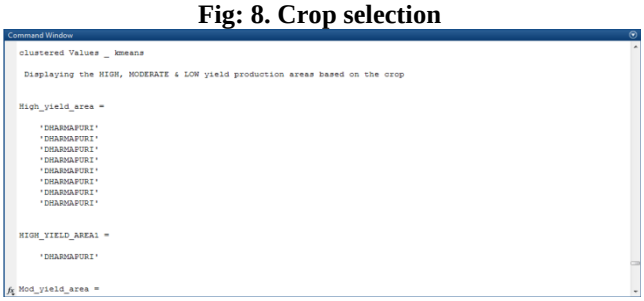

Fig: 9. District wise Crop prediction analysis
(i.e high yield area, moderate yield area, low yield area)

Table 4: comparison between K-Means, K-Medoids and ABC with Weighted based Fuzzy Clustering works with agriculture crop dataset

s.no	Parameters	K-means	K-medoids	ABC with Weighted based Fuzzy Clustering
1.	Accuracy	Low	Low	High
2.	Precision	Low	Low	High
3.	recall	Low	Low	High
4.	Mean Absolute Error (MAE)	High	High	Low
5.	Root Mean Square Error (RMSE)	High	High	Low

The above table 4 explains about the comparison between K-Means, K-Medoids and ABC with Weighted based Fuzzy Clustering works with agriculture crop dataset. By the experimental analysis, proposed algorithm produced high accuracy, precision, recall and less error rate.

VII. CONCLUSION

This paper dealt with Crop yield prediction analysis in North western zone of Tamilnadu. The research focused on Clustering techniques such as K-means, K-medoids and our proposed work is hybrid Artificial Bee Colony with weighted based Fuzzy Clustering. From the experimental analysis and results, proposed algorithm yields high accuracy, precision, recall and less error rate.

REFERENCES

1. D Ramesh , B Vishnu Vardhan, " Analysis of Crop Yield Prediction using Data Mining Techniques", Ijret: International Journal of Research in Engineering and Technology, eISSN: 2319-1163, pISSN: 2321-7308, Volume: 04 Issue: 01, Jan-2015.
2. Shailesh Pandey, Sandeep Kumar, "Enhanced Artificial Bee Colony Algorithm and its
a. Application to Travelling Salesman Problem", International Journal of Technology Innovations and Research, HCTL Open IJTIR, Volume 2, March 2013 e-ISSN: 2321-1814.
3. P.Surya, I.Laurence Aroquiaraj, "Crop Yield Prediction in Agriculture Using Data Mining Predictive Analytic Techniques", IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.5, Issue 4.
4. Karaboga , B. Basturk, "On the performance of artificial bee colony (ABC) algorithm", Elsevier, Applied soft computing, Jan-2007.
5. P. Vinciya, Dr. A. Valarmathi, "Agriculture Analysis for Next Generation High Tech Farming in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 6, Issue 5, May 2016.
6. T.Rajasekaran, P.Jayasheelan, K.R. Sri Preethaa, "Predictive Analysis in Agriculture to Improve the Crop Productivity using ZeroR algorithm", International Journal of Computer Science and Engineering Communications, Volume.4, Issue.2 (2016).
7. Anjana Gosain, Sonika Dahiya, "Performance Analysis of Various Fuzzy Clustering Algorithms: A Review", 7th International Conference on Communication, Computing and Virtualization 2016.
8. B M Sagar, Cauvery N K, "Agriculture Data Analytics in Crop Yield Estimation: A Critical Review", Indonesian Journal of Electrical Engineering and Computer Science Vol. 12, No. 3, December 2018, ISSN: 2502-4752.
9. P.Surya, Dr. I.Laurence Aroquiaraj, "Performance Analysis of K-Means and K-Medoid Clustering Algorithms Using Agriculture Dataset", Journal of Emerging Technologies and Innovative Research, January 2019, Volume 6, Issue 1

AUTHORS PROFILE

P.Surya is pursuing Ph.D research scholar in The Department of Computer Science, Periyar University, Salem. Her Educational Qualifications includes M.Sc., B.Ed., M.Phil in Computer Science. Her Area of research is Data Mining. She published two international journals and attended three international conferences.

Dr. I. Laurence Aroquiaraj is working as an Assistant Professor in The Department of Computer Science, Periyar University, Salem. His Educational Qualifications includesM.Sc., M.Tech., M.Phil., M.C.A., and Ph.D in Computer Science. He has fourteen years of Teaching and Research experience. He also has a three year of Industrial experience. His main area of Research is Medical Image Processing. He is also much interested in the areas including Data Mining, Internet Technologies and Networking. He attended fourteen International Conferences. A total of ten Conferences, Seminars and Workshops are organised by him. Overall forty three research papers are published by him in International Journals and Conference Proceedings. Total twenty three National Journals and Conference Proceedings are published by him. He is a member of Professional Bodies such as CSI (Life Member), IEEE, IAENG, IET Image Processing, Elsevier-ACI, IRG-IJCT and IJCII. He has extended his guidance to hundred and four M.C.A. and twenty M.Sc. (Computer Science) students. He also guided thirty-four M.Phil., (Computer Science) scholars. Eight Ph.D. (Computer Science) scholars are pursuing their research work under his guidance. He immensely worked on UGC funded project entitled, "Early Detection of Breast Cancer from Mammogram Using Soft Computing Techniques".