ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 2, February 2021

An Effective Agriculture cropping Analysis method using CLARANS Algorithm for Spatial Data Mining

A.Rajheshwari[1], Dr.R.Sankarasubramanian[2]
1(Ph.D Research Scholar, a.rajheswari@easc.ac.in, Department of Computer Applications, Erode Arts and Science College, Erode, Tamil Nadu, India)
2(Associate Professor, rsankarprofessor@gmail.com, Department of Computer Science, Erode Arts and Science College, Erode, Tamil Nadu, India)

Abstract: In agriculture sector where farmers as well as agricultural businesses have to formulate numerous decisions every day and complicated complexities involves the various factors influencing them. An essential issue for agricultural planning intentions is the accurate yield estimation for the numerous crops involved in the planning. Data mining techniques are essential move toward for accomplishing practical as well as valuable solutions intended for this difficulty. Environmental conditions, variabaility in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions. This paper focuses on the analysis of the agriculture data as well as decision most favorable parameters to capitalize on the crop production using data mining techniques like PAM, CLARA, DBSCAN and Multiple Linear Regression. Mining the large amount of existing crop, soil and climatic data, and analyzing new, non-experimental data optimizes the production and makes agriculture more resilient to climate change.

Keywords: Big Data, PAM, CLARA and DBSCAN

I. Introduction

India defenses moment international in the farm output. Agriculture is demographically the broadest economic sector as well as plays a important role in the overall socio economic fabric of India. Agriculture is a private business crop production which is needy on frequent surroundings as well as economy factors. Several of the factors on which agriculture is reliant are soil, climate, cultivation, irrigation, fertilizers, temperature, rainfall, harvesting, pesticide weeds. Historical crop yield in sequence is also very important for supply chain operation of companies engaged in industries. These industries utilize agricultural products as raw material, livestock, food, animal feed, chemical, poultry, fertilizer, pesticides, seed. A precise approximation of crop production and farmers risk helps these companies in planning supply chain decision like production scheduling. Business such as seed, fertilizer, agrochemical and agricultural machinery industries plan production and marketing activities based on crop production estimates [1, 2]. There are 2 factors which are helpful for the farmers and the government in decision making helps farmers in providing the historical crop yield record with a forecast reducing the risk management. It assist the management in production crop insurance policies in support supply chain operation. Data mining technique acting an imperative role in the exploration of data. Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database system. Unsupervised (clustering) and supervised (classifications) are two different types of learning methods in the data mining. Clustering is the procedure of exploratory a compilation of "data points," as well as alliance the data points into "clusters" according to some distance measure. The purpose is with the intention of data points in the same cluster have a miniature distance as of one another, while data points in different clusters are at a large distance from one another. Cluster analysis separate data into attractive groups. Well-formed clusters must confine the "natural" arrangement of the data [3]. These methods are used to categorize the different districts of Tamilnadu which are having similar crop production.

II. RELATED WORK

Clustering is considered as an unsupervised classification process [4]. A large number of clustering algorithms have been developed for different purposes [4–6]. Clustering techniques can be categorized into Partitioning clustering, Hierarchical clustering, Density- based methods, Grid-based methods and Model based clustering methods. Partitioning clustering algorithms,

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 2, February 2021

such as K-means, K-medoids PAM, CLARA and CLARANS assign objects into k (predefined cluster number) clusters, and iteratively reallocate objects to improve the quality of clustering results. Hierarchical clustering algorithms allocate substance in tree prepared clusters, i.e., a cluster preserve contain data point's legislature of low level clusters [7]. The idea of Density-based clustering methods is that for each point of a cluster the neighborhood of a given unit distance contains at least a minimum number of points, i.e. the compactness in the neighbourhood be supposed to reach some entry. The idea of the density-based clustering algorithm is that, for each point of a cluster, the neighbourhood of a given unit distance has to contain at least a minimum number of points [8]. There are different forecasting methodologies developed and evaluated by the researchers all over the world in the field of agriculture. Several of such studies are: Researchers like Ramesh and Vishnu Vardhan are analysed the agriculture data for the years 1965–2009 in the district East Godavari of Andhra Pradesh, India. Rain fall data is clustered into 4 clusters by adopting the K means clustering method. Multiple linear regressions (MLR) is the method used to model the linear relationship between a dependent variable and one or more independent variables. The needy variable is rainfall as well as self-governing variables are year, area of sowing, production. Rationale of this work is to find suitable data models that achieve high accuracy as well as a high simplification in terms of capitulate prediction capabilities [9]. 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. 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.

III. EXISTING WORK

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= {u1,…xn} represented as, P(A1, A2, … An).

Definition: Ai≠ ϕ ʉi = 1, 2, ... C Ai Aj ʉi≠ j =s Fuzzy c-partition of s, represented by (U,S) where (Unxc). U is an nxc matrix. U=((Uij))nxc where, Uij denotes membership value of ith point to the jth fuzzy set; naturally 1 i n, 1 j c. Objective Function: Uij r Vi=mean of Ai; i= 1,2, …c 2 Let r>1 let vj= ; j=1,2,…. C. Jr (U,S,A)= (xi-vj) Amxn(xi-vj); A is a positive definite matrix. Where r is a exponent term. Jr 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 vj , j=1,2,…c 4. Uij= ; 1 i n, 1 j c dij 2 = (xi-vj) A(xi-vj) 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

xmi xmi=li+rand(0,1)∗(ui−li)

in (1), where ui is the upper bound and the li 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 υmi is determined

υmi=xmi+ϕmi(xmi−xki)

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 2, February 2021

In (2), where i and xki is correspondingly denotes as a randomly selected parameter index and randomly selected food source. ϕmi 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 xmi and vmi.

3.1 K-Means Clustering

K-means Clustering is an unsupervised learning algorithm. The basic idea of this algorithm is to define k clusters using k centers i.e., one for each, respectively. There are certain steps for this algorithm that are as follows:

Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

Step 1: Randomly select 'c' cluster centers.

Step 2: Compute the remoteness among every data point and cluster centers.

Step 3: Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..

Step 4: Recalculate the new cluster center using:

where, 'ci' represents the number of data points in the ith cluster.

Step 5: Recalculate the distance between each data point and new obtained cluster centers.

Step 6: If no data point was reassigned then stop, otherwise repeat from step 3).

Advantages:

1) It is easier to understand, fast and robust,
2) It works just on numeric values,
3) The cluster have convex shapes [2], and
4) It gives the best result when dataset are distinct or well separated from each other[3].

Disadvantages:

1) The learning algorithm needs the prior information of number of cluster to be formed.
2) If there is an overlapping of data, K-means will not be able to solve as there are clusters formed at the same space.
3) This algorithm uses a Euclidean distance whose measures can unequally weight underlying factors.
4) It is also not able to handle outliers and noisy data,
5) This learning algorithm is not invariant to non-linear transformation i.e., with different representation of data we get different results,
6) This algorithm will not work for non-linear data,
7) If we choose the cluster center randomly, this will not lead to the actual result.

IV. METHODOLOGY

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

Table.4.1 Agricultural Cropping Production in District Wise Dataset

District Name	Crop Year	Season	Crop	Area	Production
Cuddalore	2012	Whole Year	Sun Flower	589.0	659.0
Dindugal	2012	Kharif	Dry Ginger	5	71.0
Dindugal	2013	Whole Year	Coconut	24785	4894.0
Kanchipurar	2013	Whole Year	Cashewn	336	165.0
Kanchipurar	2013	Whole Year	Dry Chillies	228	135.0
Thirunelveli	2015	Whole Year	Dry Chillies	2161	1822.0
Tuticorin	2008	Whole Year	Banana	9849	482959

4.1 K-Medoid Clustering Algorithm

K-Medoid Clustering algorithm is like the K-means clustering algorithm but the difference lies in the center point .Medoids is the center point in K-Medoids Clustering instead of means. In addition, K-Medoids use Manhattan while a remoteness metric as an option of Euclidean, which is utilized by K-Means. Outstanding to this, K-Medoids is additional vigorous to outliers as well as noises. K-Medoid Clustering Algorithm [3] is as follows:

Input: Ky: the number of clusters, Dy: a data set containing n objects.

Output: A set of Ky clusters.

Algorithm:

Step 1: Randomly select Ky as the Medoids for 'n' data points.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 2, February 2021

Step 2: Find the closest Medoids by calculating the distance between data points n and Medoids k and map data objects to that.

Step 3: For each Medoids 'm' and each data point 'o' associated to 'm', do the following:

a) Swap 'm' and 'o' to compute the total cost of the configuration, next

b) Select the Medoids 'o's with the lowest cost of the configuration.

Step 4: If there is no change in the assignments, repeat steps 2 and 3 alternatively.

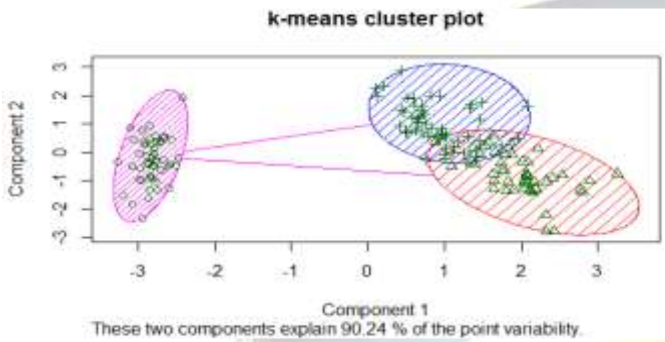

Fig.4.1 K-Means Cluster Plot Methods

Class centers are measured towards subsist qualified best dampness, virtual most favourable air temperature (intermediate cluster temperature cluster centre is considered to be relative optimal temperature), and qualified best air humidity In the data mining module, the K-means algorithm clustering every x hours of sensor data is clustered.

V. PROPOSED WORK

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 kmedoids 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. In proposed work, customized move toward of DBSCAN method is utilized to cluster the data based on districts which are having similar temperature, rain fall and soil type. PAM and CLARA are utilized to cluster the data based on the districts which are developing maximum crop production. Based on these analyses obtaining the most favorable parameters to construct the maximum crop production. Manifold linear regression method is utilized toward calculate the annual crop yield.

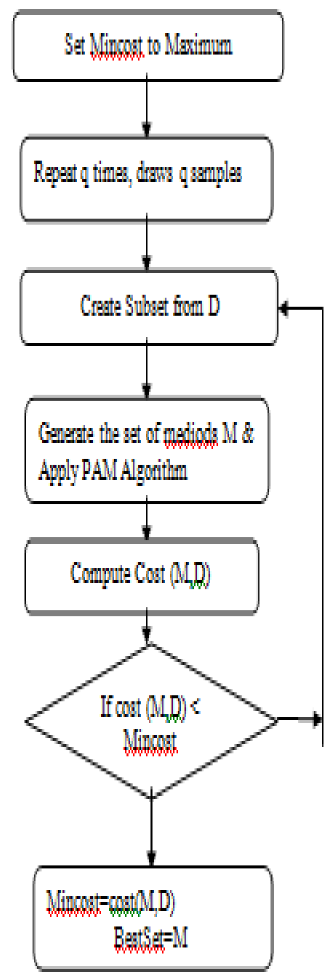

Fig. 5.1 Proposed Approach in CLARA Algorithm

Table 5.1 District Level Agricultural Area

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 2, February 2021

S.No	District	Agricultural Area
1	Dharmapuri	991139
2	Villupuram	875426
3	Tiruvannamalai	668226
4	Salem	649433
5	Cuddalore	596696

5.1 Modified approach of DBSCAN

DBSCAN is a base algorithm for density based clustering containing large amount of data which has noise and outliers. DBSCAN have double parameters explicitly Eps and MinPts. On the other hand, traditional DBSCAN cannot produce optimal Eps value [15]. Resolve of the optimal Eps value automatically is one of the mainly necessary modification for the DBSCAN. Modified DBSCAN proposes the method to find the minimum points as well as Epsilon (radius value) automatically. KNN plot is utilized to discover out the epsilon value where input to the KNN plot (K value) is user defined. To avoid the user define K value as input to the KNN plot, Batchelor Wilkins clustering algorithm is applied to the database and obtain the K value along with its respective cluster centres. This K value is given as input to the KNN Plot.

5.2 Determination of Eps and Minpts

The Epsilon (Eps) value can be found by drawing a "K-distance graph" for entire data- points in dataset for a given 'K', obtained by the Batchelor Wilkins Algorithm [16]. Originally, the distance of a point to every 'K' of its nearest-neighbours is calculated. KNN plot is plotted by enchanting the sorted standards of average distance values. When the graph is plotted, a knee point is determined in order to find the optimal Eps value [15].

Table 5.2 Crops Production by Seasonwise

S.No	Season	Crops	Production
1	Kharif	Arhar/Tur	421335.10
2	Kharif	Bajra	1642119.00
3	Kharif	Banana	2762324.00
4	Kharif	Castor seed	43901.53
5	Kharif	Cotton(lint)	2922520.00

5.3 Partition around medoids (PAM)

It is a partitioning based algorithm. It breaks the input data into number of groups. It finds a set of objects called medoids that are centrally located. The algorithm has two phases:

1. BUILD phase, a collection of k objects are selected for an initial set S.

- Arbitrarily choose k objects as the initial medoids.
- Until no change, do.
- (Re) assign each object to the cluster with the nearest medoid.

1. Improve the quality of the k-medoids (randomly select a non medoid object, O random, compute the total cost of swapping a medoid with O random).
2. SWAP phase, one tries to enlarge the superiority of the clustering by exchanging selected objects with unselected objects. Choose the minimum swapping cost.

For Example: For each medoid m1, for each non-medoid data point d; Swap m1 and d, recomputed the cost sum of distances of points to their medoid, if total cost of the con-figuration increased in the previous step, as well as undo the swap Fig. 5. 1 depicts the steps worried the PAM algorithms.

5.4 CLARA (clustering large applications)

Instead of sentence medoids for the entire data set, CLARA considers a minute illustration of the data with fixed size (*sampsize*) as well as applies the PAM algorithm to produce an optimal set of medoids for the sample. The quality of resulting medoids is calculated by the average dissimilarity between every object in the entire data set and the medoid of its cluster, defined as the cost function. CLARA repeats the sampling as well as clustering processes a pre-specified numeral of times in regulate to reduce the sampling bias. The finishing clustering consequences communicate to the set of medoids through the nominal cost.

It is designed by Kaufman and Rousseeuw to handle large datasets, CLARA (clustering large applications) relies on sampling [17, 18]. As an substitute of sentence delegate substance designed for the complete data set, CLARA draws a sample of the data set, applies PAM on the illustration, as well as finds the medoids of the section. CLARA draws numerous samples as well as gives the best clustering as the output. Here, for accuracy, the quality of the clustering is measured based on the average dissimilarity of all objects in the entire data set. **CLARA Algorithm**

The algorithm is as follow:

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 2, February 2021

1. Create randomly, from the original dataset, multiple subsets with fixed size (sampsize)
2. Compute PAM algorithm on each subset and choose the corresponding k representative objects (medoids). Assign each observation of the entire data set to the closest medoid.
3. Calculate the mean (or the sum) of the dissimilarities of the observations to their closest medoid. This is used as a measure of the goodness of the clustering.
4. Retain the sub-dataset for which the mean (or sum) is minimal. A further analysis is carried out on the final partition.

Every sub-data set is required to hold the medoids obtained as of the most excellent sub-data locate awaiting then. Randomly drawn explanations are accompanying to this set until sample size has been reached. CLARA (CLustering LARge Applications) relies on the sampling approach to handle large data sets. As an alternative of sentence medoids used for the complete data set, CLARA draws a tiny illustration as of the data set as well as apply the PAM algorithm to produce an most favorable set of medoids designed for the sample. The quality of resulting medoids is measured by the average dissimilarity between every object in the entire data set D and the medoid of its cluster, defined as the following cost function:

$$Cost\ (M, D) = \frac{\sum_{i=1}^{n} dissimilarity\,(Oi, rep(M, Oi))}{n}$$

where M is a set of selected medoids, dissimilarity(Oi, Oj) is the dissimilarity between objects Oi and Oj, and rep(M, Oi) returns a medoid in M which is closest to Oi. To improve variety bias, CLARA repeats the sampling as well as clustering process a predefined quantity of times as well as consequently selects as the concluding clustering effect the set of medoids by means of the minimal cost. Imagine q to be the number of samplings. The CLARA algorithm. Since CLARA adopts a sampling approach, the quality of its clustering results depends greatly on the size of the sample. When the sample size is small, CLARA's efficiency in clustering large data sets comes at the cost of clustering quality.

Plusplot of Clara Clustering on Iris Dataset Using Two Components Petal.Length and Petal.Width. The cluster plot of both the clustering techniques. K-means algorithm with Euclidian distance and CLARA clustering with Manhattan distances. From both the figures we conclude that CLARA clustering has more power to detect outliers and noise than K-Means clustering as the point of variability is 100% for CLARA and approx. 90% for K-Means.

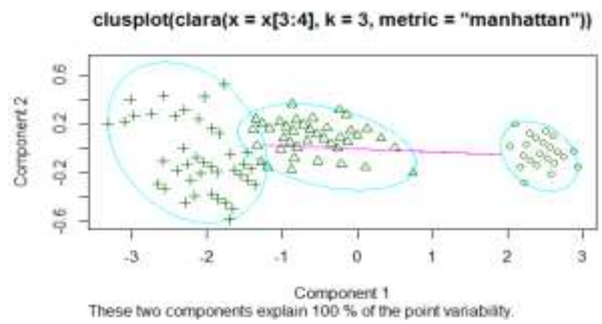

Figure 5.2. CLARA Algorithm

In Fig 5.2 proves the CLARA clustering to be more robust as compared to K-means algorithm. To Compare CLARANs with these works, we make the general following observations

- Many of the aforementioned techniques require some tree or grid structures to facilitate the clustering. Consequently, these techniques do not scale up well with increasing dimensionality of the datasets. While it is true that the material is predominantly 2-D, the CLARANs algorithm works the same way for higher dimensional datasets. Because CLARANs is based on randomized search, and does not use any auxiliary structure, CLARANs is much less affected by increasing dimensionality.
- Several of the aforementioned techniques presume with the intention of the distance function is Euclidean. CLARANS, being a local search technique, makes no requirement on the nature of the distance function.
- Many of the aforementioned techniques deal with point objects, CLARANs is more general and supports polygonal objects. A considerable portion is helps to handling polygonal objects effectively.

CLARANs is a main memory clustering techniques, while many of the aforementioned techniques are designed for out of core clustering applications. Clustering 1,000,000 objects would require slightly more than 16 Mbytes of main memory. This is an amount easily affordable by a personal computer, let alone for computers for data mining. The very low cost of RAM, main memory clustering algorithms, such as CLARANs are not completely dominated by out of core algorithms such as CLARANs are not completely dominated by out of core algorithms for many applications. Indeed, many partitioning methods have been developed, some based on k-means, some on k-medoid, some on fuzzy analysis, etc., K-medoid methods as the basis of this algorithm for the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 2, February 2021

following reasons. First, unlike many other partitioning methods, the k-medoid methods are very robust to the existence of outliers. Second, clusters found by k-medoid methods donot depend on the order in which the objects are examined.

VI. EXPERIMENTAL RESULTS

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

Table 6.1 Experimental Analysis of K-means, K-medoids DBSCAN Clustering Methods

Number of Clusters K= 3	PAM	CLARA	DBSCAN
Purity	0.578947	0.631578	0.708512
Homogeneity	0.853526	0.879624	0.895275
Completeness	0.758264	0.782356	0.786854
V-measure	0.814447	0.805181	0.83757
Precision	0.40369	0.415365	0.42152
Recall	0.24856	0.25634	0.25655
F-measure	0.307677	0.317028	0.318966

6.2 Performance Metrics for CLARA Algorithm

Performance Metrics	Accuracy	Precision	Recall	MAE	RMSE
K-Mean	86%	76%	77%	1.356	0.991
K-Medoids	87%	78%	76%	1.0298	0.896
ABCwith weighted based CLARA Algorithm	96.71%	92.67%	91.23%	0.05678	0.036357

VII. CONCLUSION

K-means clustering is the clustering technique which is utilized to construct a numeral of clusters of the explanation. Hence the cluster's center point is the mean of that cluster and the others points are the observations that are nearest to the mean value. The clustering Large Applications (CLARA) clustering, medoids are used as their center points for cluster, and rest of the observations in a cluster are near to that center point. Actually the clustering will work on large datasets as compared to K-Medoids and K-Means clustering algorithm helps to select the random observations from the dataset and perform partitioning Aroung Medoids (PAM) algorithm on it. This proposed method helps to find CLARA clustering gives better result. This paper dealt with Crop yield prediction analysis in south western zone of Tamilnadu. The research focused on Clustering techniques such as K-means, K-medoids and our proposed work is CLARA Algorithm. 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.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 2, February 2021

[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

[10]. 10.L.SathishKumar, A.Padmapriya, 2012 "ID3Algorithm Performance of Diagnosis For Common Disease", Department of Comp.sci & Engg,Alagappa University ,Karaikudi.

[11]. S. Veenadhari Dr. Bharat Mishra Dr.CD Singh, Aug 2011, "Soybean Productivity Modeling using DecisionTree Algorithms", Research Scholar MGCGV,Chitrakoot.

[12]. GAO Yi-yang, Ren Nan-ping, 2009, "Data Mining And Analysis Of Our Agriculture Based On The DecisionTree".

[13]. V. K. Somvanshi, et al., April 2006, "Modeling and prediction of Rainfall using artificial neural network and ARI."

[14]. K. Verheyen, D. Adriaens, M. Hermy, and S. Deckers, 2001, "High resolution continuous Soil classification using morphological soil profile descriptions", Laboratory for Forest, Nature and Landscape Research, Catholic Uniersity of Leuen, V. Decosterstraat 102, 3000 Leuen, Belgium

[15]. B. Rajagopalan and U. Lal, Oct 1999, " A K-nearest neighbor Simulator for daily Precipitation and other Weather variable", Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York.

[16]. Narendra Sharma, Aman Bajpai, Mr. Ratnesh Litoriya, May 2012, "Comparison the various clustering algorithms of weka tools", Department of computer science, Jaypee University of Engg. & Technology.

[17]. Prajwala T R, Sangeeta V I, January 2014, " Comparative Analysis of EM Clustering Algorithm and Density Based Clustering Algorithm Using WEKA tool", PESIT, Bangalore.

[18]. Wei Peng, Juhua Chen and Haiping Zhou, "An Implementation of ID3 --- Decision Tree Learning Algorithm".

AUTHORS PROFILE

A.RAJHESHWARI ,
Ph.D Research Scholar,
Department of Computer Science,
Erode Arts and Science College, Erode – 9.

Dr.R.SANKARASUBRAMANIAN
Associate Professor,
Department of Computer Science,
Erode Arts College, Erode – 9

