Analysis Of Agricultural Data To Maximize Crop Production Using Data Mining Technique In The State Of Jharkhand

Jayanti Kumari [a*] , Irfan Alam [b]

[a] Research Scholar, Jharkhand Rai University, Kamre, Ranchi, 835222, India
[b] Associate professor,Jharkhand Rai University, Kamre, Ranchi, 835222, India
*Corresponding author email id: jayantikumari@gmail.com

ABSTRACT

In agribusiness area where ranchers and horticulture-based organizations need to settle on endless choices consistently and elaborate intricacies comprise of the different elements impacting them. For agrarian arranging expectation a fundamental issue is the precise yield valuation for the plentiful harvests remembered for the preparation, useful and income increment. Data mining techniques are one of the required methodologies for accomplishing practical and real time solutions for tackling such type of problems. Agriculture has been considered as one of the noticeable objectives for big data. The information of Environmental conditions, soil inconsistency, input stages, prices of arrangements and commodity have made it more significant for farmers which help them for making critical farming decisions, and also increase the revenue and production in agriculture sector. This paper reviews different data mining tools like CLARA, PAM, DBSCAN etc. for the agricultural data to analyze optimal criteria for maximizing productivity in agricultural sector. Excavating the existing crop production, soil type and respective climate, as well as non-practical data help in optimization of the production which enables the agriculture too robust towards climatic variations.

Keywords: Big Data, PAM, CLARA and DBSCAN

1. Background

Nowadays, India is one of the horticultural-based nations and it positions second across the world in the ranch yield. In this fly pattern horticulture is one of the major financial areas that assumes an exceptional part in the comprehensive fiscal advancement of India. Horticulture is a procedure of harvest manufacturing that relies upon different climatic and monetary elements. The elements which impact horticulture are soil, pesticide, environment, development, water management system, manures, temperature, precipitation, gathering, and others. Crop yield data from the past is likewise radically considerable for the working inventory network of different organizations associated with agribusiness and united businesses. These horticultural and united enterprises utilize agrarian items as natural substances, food, and dairy cattle feed, compounds, compost, pesticides, seeds, and so forth. An inexact assessment of harvest creation and hazard helps these organizations in independent direction and arranging inventory network necessity like planning creation. Different organizations, Seed creation, manure ventures, agrochemical, and horticultural apparatus enterprises program their creation and key anticipating encouraging their items rely on the harvest creation gauges [1, 2]. The two significant advantages which help the ranchers and the public authority in direction are as per the following:

1. The chronicled crop reap record with an estimate helps in the danger the executives.
2. It helps the public authority in making strategies for crop protection and inventory network tasks.

Data mining techniques are essentially utilized for the huge information investigation. Data mining helps in deciding examples from enormous informational indexes utilizing mixture techniques by amalgamating various methodologies like man-made reasoning, AI, measurements, and data set administration framework and so on the two learning techniques in the information mining are

unaided (grouping) and administered (orders) learning. Bunching is the act of grabbing a "information" as a "groups" according to some distance measure. A definitive target is to have a little measurement distance between relevant informative elements in a similar group from each other. Information is being partitioned into very much framed gatherings utilizing Cluster investigation. These Well-framed bunches catch the regular type related information [3]. The research focuses on PAM, CLARA and DBSCAN grouping techniques and its effect in the farming information of Jharkhand.

2. Literature survey

Bunching is named one of the solo sorts of characterization [4]. Various bunching calculations possesses advanced with the energy for various purposes [5–6]. The bunching strategies can be described into dividing grouping, Hierarchical bunching, Density based techniques, Grid bunching and Model bunching. Different apportioning calculations utilized for grouping are specifically K-implies, PAM, CLARA and CLARANS which convert objects into k bunches, and for expanding the nature of bunching result it likewise iteratively redistributes the articles. In the progressive grouping calculations objects are arranged in the tree structure where the elements of a bunch can introduce the low-level groups [7]. Thickness put together bunching technique is based with respect to the possibility that in each neighborhood for a given unit distance, each mark of group contains a base number of focuses, for example the thickness of neighborhood achieves some edge level. [8]. Diverse anticipating systems have been planned and demonstrated by the specialists across the world for agrarian assessments are as per the following:

Ramesh and Vishnu Vardhan have investigated the horticultural information from 1965–2009 in the different region of Andhra Pradesh district. Rainfall information has been characterized into four unique gatherings by applying the K means grouping technique. Numerous straight relapses (MLR) have been applied to foster the direct connection between one ward variable and somewhere around one free factor. In that the reliant variable is precipitation while year, space of planting and creation are considered as autonomous variable. The essential point is to plan a suitable information model that can ready to accomplish high accuracy for yield forecast capacities [9]. In Bangladesh various sorts of rice are created in various trimming season. For this an earlier investigation of various climatic conditions dependent on temperature and precipitation and its impact on rural creation of rice has been done in Bangladesh. Further this review was investigated through relapse examination with temperature and precipitation. Temperature adversely affects the creation of harvest. The information has been gathered through the report of "Bangladesh Agricultural Research Council (BARC)" in the course of recent years with seven unmistakable traits in particular: "precipitation", "temperature", "daylight", "wind speed", "dampness" and "fog announcing". In this Pre-handling, the entire dataset was commonly separated into three stages March-June, July-October, and November-February. This pre-handling system has been applied for every assortment of rice. Further, the pre-handled tables have been examined to recognize the gathering of districts having comparable climate attribute [10].

Data mining procedures have been utilized for Soil portrayal. In a review the k-implies bunching has been utilized for making the group of soils using GPS technologies [11]. Researchers named Alberto Gonzalez-Sanchez, Juan Frausto-Solis & Waldo Ojeda-Bustamante have concentrated on a prescient fitness of AI strategies like Decision Tree, Artificial neural network, support vector machine and k-nearest neighbor in order to crop production [12]. Wheat crop production conjecture utilizing AI and progressed detecting strategies has been finished by Pantazi, Dimitrios Moshou, Thomas Alexandridis and Abdul Mounem-Mouazen. The goal of the exploration is to distinguish the varieties in wheat products which depends on multi-facet soil related information and satellite pictures of harvest development features [13]. S. Veenadhari, B. Misra and CD Singh have developed a software tool 'Crop Advisor' which is a handy web page to prognosis the impact of climate for effective crop production [14]. C4.5 algorithm has been used to identify different parameters of climate for production of different crops in districts of Madhya Pradesh.

Value Chain Study for the production of tomato in Ranchi, Jharkhand' for the year 2017-18

3. Methods

To break down the agribusiness information utilizing information mining strategies is the target of arranged work. Farming information has been gathered from following sources in arranged work:
Dataset in agrarian area [https://data.gov.in/, http://raitamitra.kar.nic.in/statistics], Insightful agribusiness information [html://CROPWISE_NORMAL_AREA], agricultural details of different districts[http://14.139.94.101/fertimeter/Distkar.aspx],
http://raitamitra.kar.nic.in/ENG/statistics.asp],
Horticulture information are being dependent on climate, temperature, and moisture [http://dmc. kar.nic.in/trg.pdf].

Input datasets includes the long term information with boundaries that are: year, State-Karnataka (28 locale), District, crop (cotton, rice, groundnut, jowar, rice, wheat etc), seasons such as kharif, rabi, & summer, region (in hectares), creation, Normal temperature (°C) & precipitation (mm), soil, PH esteem, soil type, significant composts, nitrogen, phosphorus, Potassium, least precipitation required, least temperature required. Further developed methodology of DBSCAN is utilized to bundle the information focused on location which put forward the comparable rainfall, temperature and soil type in arranged work. To hitch the information dependent on the areas which are creating most extreme harvest creation, PAM and CLARA have been utilized (Wheat crop is considered as model in arranged work). We are getting the ideal boundaries to deliver the most extreme harvest creation dependent on these examinations. To conjecture the yearly edit yield, various straight relapse strategies is utilized.

Base calculation for thickness-based grouping containing enormous measure of information which has clamor and exceptions are a changed methodology of DBSCAN. Eps and MinPts are the two boundaries of DBSCAN.

However, ideal Eps esteem [15] can't be created by customary DBSCAN. One of the most essential revisions   for the DBSCAN is the assurance of the ideal Eps esteem consequently. Adjusted methodology of the DBSCAN strategy is informed by fig. [1].

The strategy to track down the base focuses and Epsilon (sweep esteem) consequently is proposed by changed DBSCAN. To examine the epsilon esteem where contribution to the KNN plot (K worth) is client characterized, KNN plot is utilized. Batchelor Wilkins gathering calculation is applied to the information base and secure the K worth alongside its separate bunch places to dodge the client characterize K worth as contribution to the KNN plot. K worth is given as contribution to the KNN Plot.

3.1 Determination of Minpts and Eps:-

By preparing a "K-distance chart" for whole data points in dataset for a assumed 'K', found by the Batchelor Wilkins Algorithm [16], Epsilon (Eps) worth can be found. Right off the bat, the distance of a highlight each 'K' of its closest neighbor is determined. KNN scope is planned by holding the coordinated upsides of customary distance esteems. A knee fact is chased to catch the ideal Eps esteem [15] when the chart is planned.

4. Partition around medoids (PAM)

PAM is calculation method grounded on the apportioning. PAM breaks the information into figure of bunches. PAM finds a bunch of articles named medoids that are midway situated. Closest information focuses can be determined and made it as groups with the medoids. There two stages in calculation:
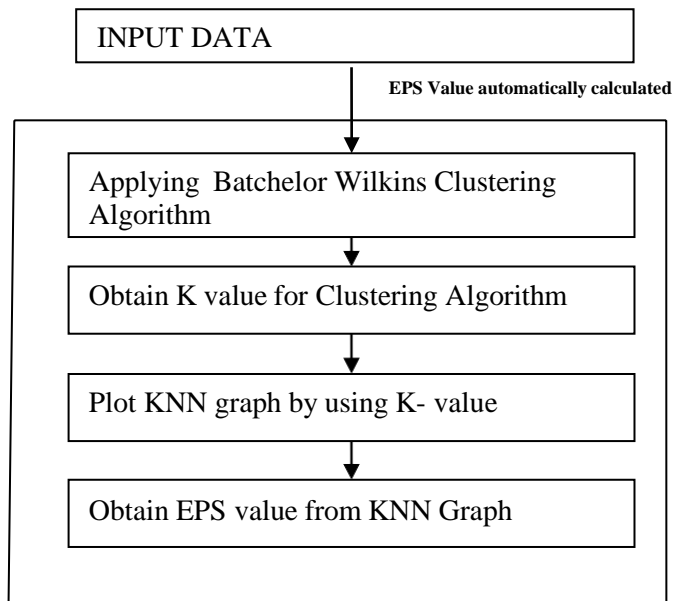
INPUT DATA

EPS Value automatically calculated

Applying Batchelor Wilkins Clustering Algorithm

Obtain K value for Clustering Algorithm

Plot KNN graph by using K- value

Obtain EPS value from KNN Graph

Fig.1 EPS Calculated

Select K representative object automatically

Compute TC_{ih} for all pairs (O_iO_h)

Select pair (O_i,O_h) with Min $TC_{ih} (O_i,O_h)$

Replace O_i with O_h

Yes

$TC_{ih} < O$
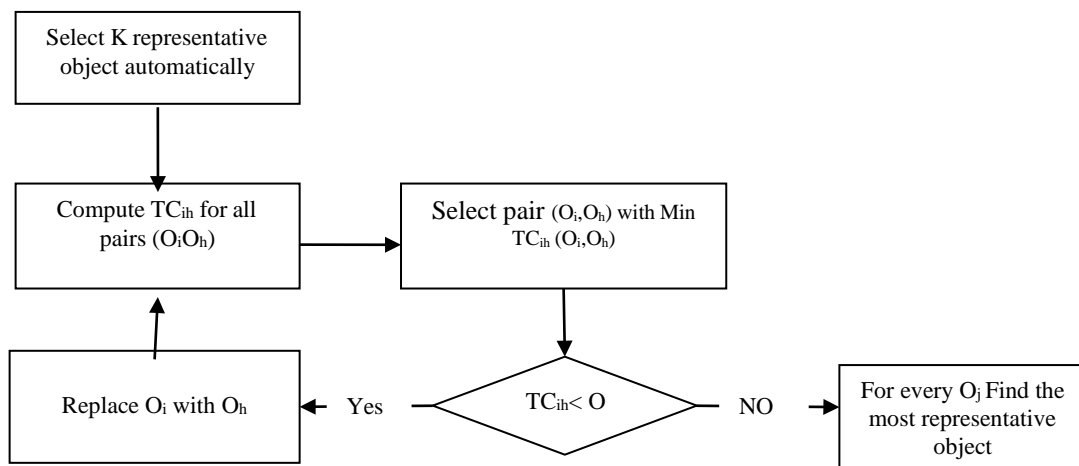
NO

For every O_j Find the most representative object

Fig. 1: The step 2 consisting to steps of PAM algorithm

Construct stage, an assortment of k articles are chosen for an underlying set S.
•• Self-assertively pick k items as the underlying medoids.
•• Till no variation, do.
— (Re) allot every entity to the bunch with the closest medoid.
— Repair the component of the k-medoids (whimsically handpick a non medoid object, O irregular, figure the whole expense of replacement a medoid with O arbitrary.

Trade stage, one endeavors to repair the prominence of the packaging by supplanting particular articles with non-selected items. Pick the base trading price.
Model: for every non-medoid information point d; for every medoid m1, Swap m1 and d, recompute

the charge (whole of spaces of focuses to their medoid), on the off chance that all out cost of the adaptation expanded in the past advance, unfasten the trade Fig. 3 depicts the means tangled in the PAM calculations.

CLARA (grouping enormous applications)

CLARA is planned by Rousseeuw and Kaufman to grasp tremendous datasets, CLARA (bunching enormous applications) relies upon examining [17, 18]., CLARA draws an example of the informational collection applies PAM on the example, and finds the medoids of the tester in its place of observing agent objects for the comprehensive informational index. CLARA draws different examples and gives the best bunching as the result to think of better approximations. The nature of the grouping is noble dependent on the normal variety of all items in the entire informational index for precision. Fig. 3 commands approximately the means tangled in the CLARA Algorithm. Numerous straight relapses to gauge the harvest, Multiple direct relapses are a variety of "straight relapse" examination. This model is designed to initiate the connection that uses one dependent inconstant and at least two freed factors [19]. The given dataset are used where x_1 to x_k are autonomous factors and Y is a reliant on variable, the complex straight hysterics the dataset to the model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i$$
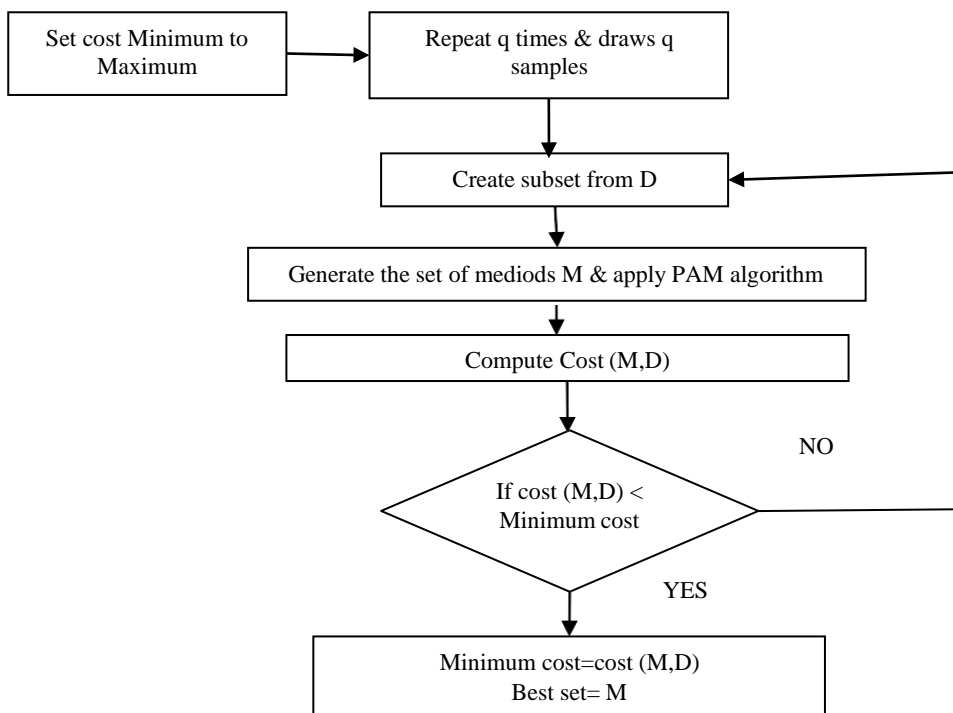

Fig. 2 : CALRA algorithm steps

Where β_0 is the y-capture and $\beta_1, \beta_2, \ldots, \beta_k$ boundaries are known as the incomplete coefficients. It's important to know the critical traits from the data set in framework structure prior to applying the different straight relapse to gauge the harvest yield. Every one of the properties utilized in the data set won't be essential or moving the worth of these characteristics won't trouble anything on the dependent factors. That sort of qualities can be overlooked. P esteem test is executed on the data set to observe the vital properties and numerous straight relapse is spread distinctly on the noteworthy qualities to figure the harvest gather.

5. Evaluation Method

Data mining calculations effort with dissimilar belief systems, having the option to be biased by changed sorts of relationship on information. This work tracks down the ideal bunching strategy for agribusiness

information investigation to guarantee more pleasant conditions in assessment. Planned work accepts the fringe eminence measurements [3] like Consistency, Entirety, Precision, V Measure, Rand Index, Purity, Recall and F ration to analyze the PAM, CLARA and DBSCAN grouping strategies.

$$Y = XB + E$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \quad B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \quad E = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

It's important to know the huge traits from the information base prior to applying the various straight relapse to estimate the harvest yield. Every one of the qualities utilized in the data set won't be significant or changing the worth of these traits won't resentful anything on the reliant factors. Such ascribes can be overlooked. To observe the significant properties and various direct relapse is spread distinctly on the critical qualities to figure the harvest yield.

6. Experimental results

Altered methodology of DBSCAN Client desires to decide the Minpts and Eps esteems prior to applying DBSCAN calculation on the dataset. The Batchelor Wilkins calculation is spread on the dataset to oversee the K worth (Numeral of bunches) unavoidably. K worth acquired since the Batchelor Wilkins is 7 with subsequent locale as bunch habitats for the dataset utilized in the planned work. Complete areas are assembled into 3 bunches utilizing PAM grouping strategy [20].

7. Comparing and Comparison

Grouping correlation has done utilizing four execution quality measurements as referenced before. Table 6 displays the difference of CLARA, PAM and DBSCAN strategies for grouping the regions which are taking the same harvest efficiency.

Table 1 and Fig. 4 depict the differentiation of CLARA, PAM and DBSCAN gathering strategies. More excellent measurement esteems determine better grouping quality. Investigation of the quality measurements injuries for various DBSCAN gives the better bunching superiority than CLARA and PAM, CLARA gives the preferred grouping quality done the PAM.

7.1 Discussion

The harvests are routinely designated by its financial values. Though, the rural system demands a vintage guess of various harvests. Five harvests were selected for the proposed work utilizing the information accessibility as the vital scale in this sense.

Comparison of Clustering methods		
number of clusters k=3		
	PAM	CLRA	DBSCAN
Purity	0.578947	0.631578	0.708512
Homogeneity	0.853526	0.879624	0.895775
Completeness	0.758264	0.782356	0.786854
V-measure	0.814447	0.805181	0.83757
Prcision	0.40369	0.415365	0.42152
Recall	0.24856	0.25634	0.25655
F-measure	0.307677	0.317028	0.318966
Rand index	0.785364	0.796352	0.814561

Table 1: Correlation table of CLARA, PAM and DBSCAN grouping strategies, where k=3 [data taken from Jharkhand Government websites].

Research is by and large restricted to the 5 yields those are wheat, cotton, ground nut, jowar and rice. Model wheat yield investigation is given in this red-top in current works.

This research shields the CLARA, PAM, Improved DBSCAN gathering tools and numerous straight relapse techniques. CLARA and PAM are the standard packaging methods whereas DBSCAN is elevated by acquainting the Batchelor Wilkins gathering technique for controlling the 'k worth and KNN technique to manage the base sweep and focuses esteem routinely. Crop informational index is not really settled the ideal boundaries for the wheat crop manufacturing utilizes these techniques. Numerous direct relapses are utilized to find the significant qualities and custom the proportionality for the vintage gauge.

A few works measure the component of the grouping strategies utilizing in-house quality measurements [21], some different uses the fringe quality measurements. Though, in these mechanisms, investigation is tight to the fringe eminence measurements which are blend of various measurements those are [22]: set coordinating with measurements, measurements focused on counting supports and measurements grounded on Entropy. The eminence measurements were reviewed, from the standout to the foulest, giving to uniformity, immaculateness, fulfillment, v measure, accuracy, review and rand list consequences, in the accompanying request: DBSCAN, CLARA and PAM.
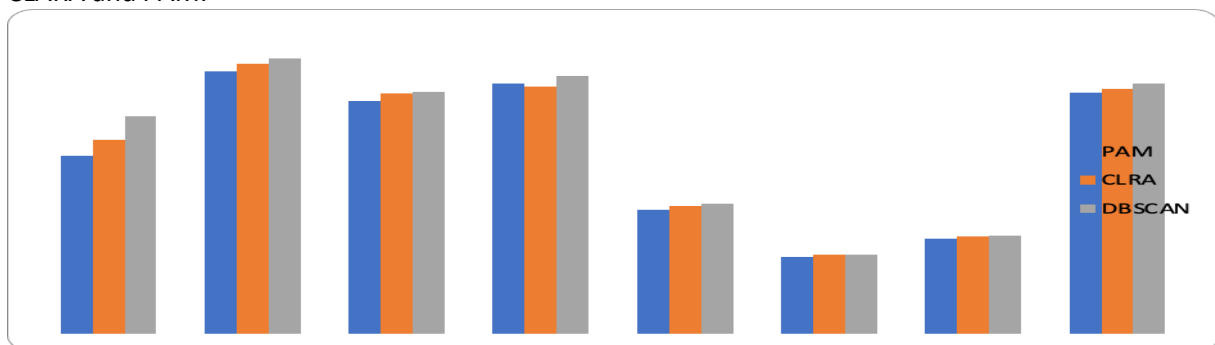

Fig. 3 : Comparison graph of different clustering methods

7.2 Maximization of crop production

The following two areas are engrossed by state government for exploiting the crop production in Jharkhand:

1. Water development

The necessity of enlightening water proficiency and yield has been progressively stressed. The situation can be attained by proficient practice of rain water harvesting, available water, micro irrigation and integrated harvesting systems supervision. The water optimistic savings have to be supplemented with farming systems savings would be mounted up in rain fed parts on maintenance scheme. The subventions on irrigation should be connected with sustainability and communal fairness worry. This will promote to circumvent further diminution of groundwater modality. The rainwater use proficiency desires to be enlarged to harness most of the latent. The Federations would labor out a widespread program on partaking groundwater supervision and famine- proofing. These savings want to be finished in a structure of general reassuring watering with concerted consumption of aquifer and superficial water forms. Assessments signpost normal venture proportion of Rs. 2 lakh/ha (of which Rs. 1.5 lakh stood by the Centre and Rs. 0.5 lakh by the Federations) in the dampened knack progress via Z in rain fed parts was solitary Rs. 12,000 apiece hectare in savannahs to Rs. 15,000 apiece hectare in the alpine expanses through Integrated water.

2. Management programme (IWMP).

Estimations advise that the entire venture constraint aimed at solving the Latent of rain fed Farming can be Rs. 50.000/Ha or else further. This highlights that the venture proportion in the rain fed zones wants to stand heightened significantly in cooperation by Centre and the Federations to bind the massive latent of these expanses. Remark: Augmentation of Water usage proficiency concerning water improvement the plots of Jharkhand state is surging in addition to in asymmetrical slant in

environment. Shower drizzle is almost 1400mm intimate the state, obtainable of which 70 percent precipitation of entire precipitation rolling extensively over stippy slants unusable with stunning appreciated and lush loam (Silt) specks then lastly amasses in Extensive Watercourses. The overhead proportion of Stream water might be resourcefully castoff by certain mechanical organization thru loam & water cure programs, which certify water usage proficiency and give rise to accomplishments of adequately & justifiable outcome in Farming fabrication and yield in forthcoming.

8. Seeds

Development in Agriculture manufacture throughout XI blueprint stood catalyzed ascribable to advanced venture in seed. Acceptable obtainability of specialized seeds remains a main encounter in the eastern as well as north eastern expanse due to underprivileged seed structure in addition to recognized existence. Secluded seed Corporations have started building a bearing in the seed consequence nevertheless their emphasis rests on cross changes where income verge is great. Subsequently, value of seeds accessible to fringe and deprived ranchers stands identical tall. Datum that the civic division seed manufacture remains not proportionate by means of the necessities of the ranchers additional enhances to their desolation. Preserving solid native seed scheme that is fit connected to R&D scheme remains essential. Several of nearby modified seeds do not catch a habitation in recognized seed souks as they are not lucrative. Drizzled parts remain susceptible to great barometrical perils in addition to the seed schemes have to retort to conference deficiency owing to the risk. The wear and tear of seed due to protracted parched monsoon curses instantly next propagating is precise public. The Federations need to guarantee the accessibility of an additional set of seeds aimed at recap seeding, uncertainty the first sowing flops. Silage seeds remain constantly a rare source intended for which the distinct strategy wants to be established. RKVY coffers want to be consumed intended for making seed setup like treating, storing etc. The seed progressing strategy necessity to lay in place by Federations. NARS requisite to advance worthy amalgams intended for great capacity harvests to normalize charge of amalgam seeds. Remark: Eminence seed is utmost precarious intended for improving farming creation. The federation is reliant on external interventions used for seed as here is no solid systematized scheme of seed making. Sustaining sturdy native seed scheme through relation to R&D scheme in the academia or other associations is essential [24]. Illustrating awake of efficient plan would be the fundamental to accomplishment.
1. To set up State Seed Firm in the Crucial essential.
2. To invite public in addition to private segments in seed Manufacture program.
3. Inaugurating seed standby at State / District level.
4. Resilient care scheme in Elating & Manufacture of breeder and establishment seeds in order to certify systematic Strategy. 8
5. SAU must yield essential magnitude of Breeder and Establishment Seeds of altogether capable variations
6. State Government must improve an apparatus for fabrication of specialized seed of hopeful changes through conception of satisfactory sum of seed villages/ seed manufacturer civilizations.
7. State Government must safeguard obtaining of all-inclusive capacity of Establishment seed formed by the SAU as well as of proficient seed manufactured via seed parishes/ seed fabricator societies.
8. The stratagem must be to twofold the existing seed spare proportion (SRR) appraised by way of 10% for turf harvests and 20% of vegetal yields.

9. Conclusion

A few data mining abilities are analyzed on the material to assess the finest presentation yielding procedure. The present work consumed data mining stratagems CLARA, PAM and DBSCAN to study the best environment need of wheat like idyllic latitude of top temperature, most despicable temperature and shower to achieve progressed assembling of wheat crop. Gathering methods are perceived at using eminence capacities.

DBSCAN gives the preferred bunching superiority over PAM and CLARA and CLARA gives the ideal grouping eminence over the PAM, presenting to investigate of collection quality measurements. The arranged work can likewise be delayed to investigate the grime and dissimilar issues for the harvest and to upsurge the yield fabricate below unique climatic circumstances and flood the pay. Thus, the paper concluded that with the application of data mining techniques the crop production can be maximized which will be helpful in improving farmers financial condition as well as economy of the agricultural sector.

References

1. Veenadhari S, Misra B, Singh CD (2011). Data mining techniques for predicting crop productivity—A review article. In: IJCST.
2. Gleaso CP. (1982) Large area yield estimation/forecasting using plant process models. Paper presentation at the winter meeting American society of agricultural engineer's palmer house, Chicago, Illinois.
3. Majumdar J, Ankalaki S. (2016) Comparison of clustering algorithms using quality metrics with in-variant features extracted from plant leaves. In: Paper presented at international conference on computational science and engineering.
4. Jain A, Murty MN, Flynn PJ. (1999) Data clustering: a review. ACM Comput Surv.
5. Jain AK, Dubes RC. (1988) Algorithms for clustering data. New Jersey: Prentice Hall.
6. Berkhin P. (2006) A survey of clustering data mining technique. In: Kogan J, Nicholas C, Teboulle M, editors. Grouping multidimensional data. Berlin: Springer.
7. Han J, Kamber M. (2001) Data mining: concepts and techniques. Massachusetts: Morgan Kaufmann Publishers.
8. Ester M, Kriegel HP, Sander J, Xu X. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Paper presented at International conference on knowledge discovery and data mining.
9. Ramesh D, Vishnu Vardhan B. (2013) Data mining techniques and applications to agricultural yield Data. International journal of advanced research in computer and communication engineering.
10. MotiurRahman M, Haq N, Rahman RM. (2014) Application of data mining tools for rice yield prediction clustered regions of Bangladesh. IEEE.
11. Verheyen K, Adrianens M, Hermy S Deckers(2001). High resolution continuous soil classifcation using morphological soil profle descriptions. Geoderma.
12. Pantazi XE, Moshou D, Alexandridis T, Mouazen AM (2016). Wheat yield prediction using machine learning and advanced sensing techniques. Comput Electron Agric.
13. Veenadhari S, Misra B, Singh D. (2014) Machine learning approach for forecasting crop yield based on climatic parameters. In: Paper presented at international conference on computer communication and informatics (ICCCI-2014), Coimbatore.
14. Rahmah N, Sitanggang IS.(2016) Determination of optimal epsilon (Eps) value on DBSCAN algorithm to clustering data on peatland hotspots in Sumatra. IOP conference series: earth and environmental. Science.
15. Forbes G.(2002) The automatic detection of patterns in people's movements. Dissertation, University of Cape Town.
16. Ng RT, Han J. CLARANS,(2002) A Method for Clustering Objects for Spatial Data Mining. In: IEEE Transactions Knowledge and Data Engineering.
17. Kaufman L, Rousseeuw PJ.(1990) Finding groups in data: an introduction to cluster analysis. Wiley. doi:10.1002/9780470316801.
18. Multiple linear regression (2017) http://www.originlab.com/doc/Origin-Help/Multi-Regression-Algorithm.
19. Elbatta MNT (2012). An improvement for DBSCAN algorithm for best results in varied densities. Dissertation Islamic University of Gaza.
20. Kirkl O, De La Iglesia B. (2013) Experimental evaluation of cluster quality measures. 978-1-4799-1568-2/13 IEEE.
21. Meila M (2003) Comparing clustering. In: Proceedings of COLT. Report: Jharkhand state Government.