Agri Articles
(e-Magazine for Agricultural Articles)
Volume: 06, Issue: 02 (MAR-APR, 2026)
Available online at http://www.agriarticles.com
©Agri Articles, ISSN: 2582-9882

Cluster Analysis in Agriculture

*Meet Patel[1], Dr. Vishva Patel[2], Jay Upadhyay[1], Vashisth Patel[1] and Suyash Bhosale[1]
[1]Postgraduate Scholar (Agriculture Analytics), KKIASR, FAAST, Ganpat University, Mehsana, Gujarat-384012
[2]Assistant Professor, KKIASR, FAAST, Ganpat University, Mehsana, Gujarat-384012
*Corresponding Author's email: meetvpatel2911@gmail.com

A cluster is a set of data objects that are similar to one another within the cluster and dissimilar to the objects in other clusters. It is a technique used to classify data into multiple groups, called clusters such that objects in the same cluster are more similar than those in other clusters. Cluster analysis is a technique for grouping observations together such that:

- Within each group, individuals share homogeneous or compact characteristics – i.e., observations within the same group are similar.
- The groups should be dissimilar to other groups regarding the features i.e., Observations of one group should not resemble the observations of another group.

Many domains such as life science, medicine, engineering, agriculture, social science etc where the need of cluster analysis arises in a natural way. Cluster analysis is the analytical method used in the banking industry to identify potential customer data to lend money. For instance, looking into portfolio and CIBIL scores, one finds an idea whether to disburse loan or not. In marketing a cluster analysis is used to segment those individuals into groups based on similar purchasing behavior.

Methods of Clustering

1. Hierarchical Method

Hierarchical clustering is a type of cluster analysis where the clusters are arranged in nested tree structure. One is either merging the smaller clusters into larger ones (agglomerative approach) or separating the larger clusters from the smaller ones (divisive approach). The output is a dendrogram that illustrates the relationship between the objects or groups at various similarity levels. This is convenient, since it requires no pre-specification of the number of clusters. Hierarchical clustering is commonly used in disciplines such as biology (for instance, plant taxonomy and genetics) and soil classification where natural hierarchical structures may be present.

Hierarchical clustering is of two types mainly:

A. Agglomerative hierarchical clustering - Each object is a cluster in the beginning. Then, clusters are successively merged until the desired cluster structure is achieved.
B. Divisive hierarchical clustering - Starts with one cluster of all objects. Then the cluster gets divided into sub-clusters and then those are iteratively topsy-turvily clusters into their own set of sub-clusters. This process is repeated until the required structure of clusters is formed.

Hierarchical method has the following main advantages:

- Flexibility is built in as to the level of granularity.
- Use with any type of similarity or distance metric.
- Applicability to any attribute types.

Hierarchical method has the following main disadvantages:

- Vagueness of termination criteria.
- In hierarchical algorithms that do not revisit once.
- Computationally expensive and not suitable for a large data set.

2. Partitioning Method

In this method, k partitions of the data are constructed such as each partition represents a cluster ($k \leq n$). This means that it segments the data into k groups in which each group must have at least one object. A partitioning method generates an initial partitioning. It then employs an iterative refitting algorithm that seeks to improve the grouping by relocating objects from one group to another. The basic criterion of a good partitioning is that the closer (similar) objects must be in the same cluster; but far apart (dissimilar) objects should belong to different clusters.

A very simple and popular algorithm is also using a squared error criterion; we call it the K-means. The data is divided into K groups (C1, C2,…,CK) using this algorithm, which use their centers or means to represent. The mean of all instances in a cluster is considered to be the center of each individual cluster.

k-means algorithm steps:

1. Choose K initial centroids.
2. Calculate distance of each data object to the centroid Replace each point in the cluster with its closest centroid.
3. Calculate now the new centroid of each cluster.
4. Continue this process until there is no change in the centroid.

Partitioning method has the following main advantages:

- Relatively scalable and simple.
- Provides an optimal result when datasets are unique and well separated.

Partitioning method has the following main disadvantages:

- Poor cluster descriptors.
- It requires the user to specify the number of clusters beforehand.
- Highly sensitive to initialization phase, noise and outliers, Frequent trap into the local optima.
- Cannot handle non convex clusters of different shape and density.

Review of Literature

Meena et al. (2017) used data on area, production and productivity of 23 crops in Rajasthan for cluster analysis to analyze crop similarity. The data were grouped into pre (1980-1995) and post-World Trade Organisation (WTO) (1996-2014) periods, and Ward's hierarchical clustering approach was applied. Crops such as wheat, mustard and rapeseed, gram, cotton and bajra exhibited similarity in area and production across districts in the pre-WTO period while during the post-WTO phase similarity was mainly confined to wheat and bajra. Limited similarity was observed with regard to productivity in the first period (potato and wheat) while more crops including coriander, potato, wheat, garlic and pea showed similarity in the second period. Cluster analysis proved to be a tool for grouping different commodities in terms of similarity and understanding the cropping patterns over time would help policymakers better use resources and formulate future strategies.

Ravi and Lakshana (2025) used clustering techniques to study the effects of climate change on agricultural production. The study looked at crop yield, economic impact and farmers' adaptation in different regions. Three clustering techniques were employed, evaluated by silhouette scores as K-Means, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and Agglomerative clustering. Based on the evaluation, K-Means performed the best with a score of 0.57. The results of the study indicate that clustering techniques can detect patterns in agriculture data affected by climate, which may assist better informed decision-making for increasing agricultural productivity.

Sharma et al. (2025) used Mahalanobis D^2 statistics and Tocher's clustering method to study genetic diversity in forage pearl millet (Pennisetum glaucum L.) genotypes in North

Gujarat condition. All the 13 fodder associated traits were evaluated in a total of 30 forage pearl millet genotypes. Eight clusters were observed; 10 genotypes (54.7%) in Cluster V, six (32.4%) in Cluster III, five each representing two clusters (Clusters II and VII) while single genotype clusters comprised the remaining three. The highest intra-cluster distances were observed in clusters III (78.19) and II (75.26), depicting more variation among the traits under consideration within each cluster. The highest inter-cluster distance was observed between cluster I and VI (575.03) followed by cluster I and Cluster VII (553.99), which indicates maximum genetic divergence. Cluster V and Cluster VIII had the lowest inter-cluster distance (101.93), suggesting similarity. According to the study's findings, crossing genotypes from highly divergent clusters can increase variability and enhance breeding initiatives.

Conclusion

Cluster analysis can group similar agricultural data with same characteristics. Hierarchical and partitioning methods are alternative clustering techniques that allow the analyzing of more complex datasets. Based on the available literature, we also find that cluster analysis has been applied in agriculture for studies of crop similarity, genetic diversity as well as climate change effects on agricultural production. Overall, cluster analysis is a useful and effective tool for agricultural data analysis, supporting better decision-making, crop planning and improvement of agricultural productivity.

References

1. Han, J., Kamber, M. and Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
2. Karangale, T. and Bhoir, S. (2015). Cluster analysis: Preliminaries and techniques. *International Journal of Computer Applications*, 129(14): 36–40.
3. Meena, L.K., Sen, C. and Kushwaha, S. (2017). Cluster analysis to form similarity for major selected crops in Rajasthan, India. *International Journal of Current Microbiology and Applied Sciences*, *6*(4): 2673–2682.
4. Ravi, I. and Lakshana, M. (2025). Cluster Analysis of Climate Change Impact on Agriculture Production. *International Journal of Research Publication and Reviews*, 6(3): 2626–2631.
5. Rokach, L. and Maimon, O. (2005). Clustering methods. In Data mining and knowledge discovery handbook. Springer.
6. Sarkar, S.K. (2023). Cluster analysis. In Statistical procedures for analysing agricultural data using R. IASRI, New Delhi.
7. Sharma, R., Viradiya, Y. A., Jajoriya, R., Patel, D. R., Vekariya, R.G. and Patel, P.J. (2025). Study on genetic diversity in forage pearl millet (*Pennisetum glaucum* (L.) R. Br.) genotypes under North Gujarat condition. *Plant Archives*, 25(Supplement 1): 415–423.