WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Cluster Analysis of Climate Change Impact on Agriculture Production

Ravi I [1], Lakshana M [2]

[1]Assistant Professor, [2] II M.Sc., Computer Science,

[1, 2] K.R. College of Arts & Science, K.R. Nagar, Kovilpatti.

Email: 1ravi_cs@krcollege.net,2lakshana200307@gmail.com

DOI : https://doi.org/10.55248/gengpi.6.0325.1194

ABSTRACT:

Agricultural practice is affected by climate change because of its direct dependence on climate change. The primary aim of this research analyzed the impact of climate change on agriculture using clustering techniques. Specifically, the research seeks to understand the relationship between various climatic and agricultural variables, including crop yield, economic impact, and adaptation strategies, by performing unsupervised machine learning techniques on a dataset that includes features like temperature, precipitation, CO2 emissions, Irrigation access, soil health Index, Adaptation strategies and more. Various clustering algorithm including K-Means, DBSCAN, and Agglomerative Clustering to group the data into distinct clusters. These clusters will help to identify patterns in how different regions and agricultural practices are impacted by climate change and ultimately linking higher crop yields to greater economic gains. The results from each model were evaluated using silhouette scores to assess the quality of the clustering.

KEYWORDS: Economic impact, Crop yield, Climate change, Clustering, Precipitation.

1. INTRODUCTION:

Climate change poses a significant threat to agricultural productivity and food security worldwide. The climate change impacts on crop yields become increasingly pronounced, particularly in developing and low-income countries that rely heavily on agriculture for subsistence. These regions often lack the infrastructure necessary for effective adaptation, making them particularly vulnerable to climate-related disruptions. This study examines the intricate relationships between climate change and agriculture, highlighting how changes in temperature, precipitation patterns. This research primarily focused on examining the effects of climate change on agriculture through the application of clustering techniques. Specifically, it aimed to explore the relationships between various climatic and agricultural factors, such as crop yields, economic impacts, and adaptation strategies. Using unsupervised machine learning methods, the study analyzed a dataset containing features like temperature, precipitation, CO2 emissions, irrigation access, soil health index, and adaptation strategies. Clustering algorithms, including K-Means, DBSCAN, and Agglomerative Clustering, were employed to group the data into distinct clusters, revealing patterns in how different regions and agricultural practices are influenced by climate change. Additionally, the study highlighted the connection between higher crop yields and increased economic benefits. The clustering performance of each model was evaluated using silhouette scores to measure the quality of the clusters.

2. LITERATURE SURVEY:

Shams, Mahmoud Y., et al. (2023) [1] developed a machine learning model to predict temperature changes related to climate change using features like year, month, and atmospheric variables. They tested various regression models, including SVM, KNN, Random Forest, and Cat Boost Regressor, to analyze global temperature trends. The Cat Boost Regressor outperformed others with an MAE of 0.0036, RMSE of 0.054, MSE of 0.003, and R^2 of 92.40%. The study emphasizes the significance of accurate climate forecasting among rising global temperatures and extreme weather.

Rajesh (2011) [2] applied spatial data mining in agriculture. The k-means algorithm was used to select at random the number of k objects, which initially denote the cluster mean. And for each of other objects, another object is assigned to the cluster of the most similar object. From this, new mean for each cluster is computed; this process is performed until the criterion function converges. This method is applied in agriculture by supplying temperature an rainfall as the initial spatial data and analyzing the agricultural meteorology for the improvement of agricultural produce as well as reducing the losses of yield.

Novia Andini.,et al(2021) [3] This research focus on climate prediction using RNN LSTM. This prediction can provide results in the form of climate forecasts for the next day. The results of this prediction are in the form of climate types belonging to tropical climates, dry climates, temperate climates, continental climates and arctic climates. The selection of data in this study is based on the use of Koppen theory which will be used to categorize climate types. The data used in this study are rainfall data, average temperature, duration of sun exposure and humidity for the last 10 years from 2010 to 2020.It provide accurate category in predicting climate to estimate agricultural yields.

Crane -Droesch (2018) [4] utilized machine learning to estimate crop yields and assess the impact of climate change. To predict climate change's effects on agriculture, the author emphasizes crop yields' weather dependence. The article introduces a semi parametric deep neural network yield modeling method. The study also examines how climate change affects corn yield using climate model scenarios. Climate change reduces corn output, although less than classical statistical approaches predicted. In hottest places and scenarios, the proposed method appears less pessimistic.

Habib-ur-Rahman et al. (2022)[5] explored the impact of climate change on food-insecure agricultural production in Asia, highlighting threats like drought, heat waves, and unpredictable rainfall. The study projected a 14.1%-17.2% yield loss for rice and wheat by mid-century (2040-2069) and emphasized climate-smart agriculture for resilience. Strategies include legume-based rotations, agroforestry, climate-resilient crop varieties, decision support systems, and smart technologies for water, soil, and energy management. Crop simulation models (DSSAT and APSIM) were used with field data from 155 farms and historic weather data from Pakistan Meteorological Department. The findings underscore sustainable practices to mitigate climate risks and ensure food security.

Haq et al. (2022)[6] utilized AI and ML, specifically the Long Short-Term Memory (LSTM) model, to forecast key environmental factors using a Himachal Pradesh dataset (2001–2017) on temperature, snow cover, and vegetation index. The study highlights LSTM's ability to detect long-term dependencies, aiding hydrological modeling and forest expansion predictions. A coarse-to-fine approach was employed to analyze environmental parameters and improve time series forecasting accuracy. The proposed system enhances the assessment, adjustment, and optimization of environmental factors efficiently.

Basel Y.,et al (2022) [7] This research presents a significantly Climate forecasting model using a deep convolutional Long Short-Term Memory (LSTM) to forecast temperatures world widely.

3. PROPOSED METHODOLOGY:

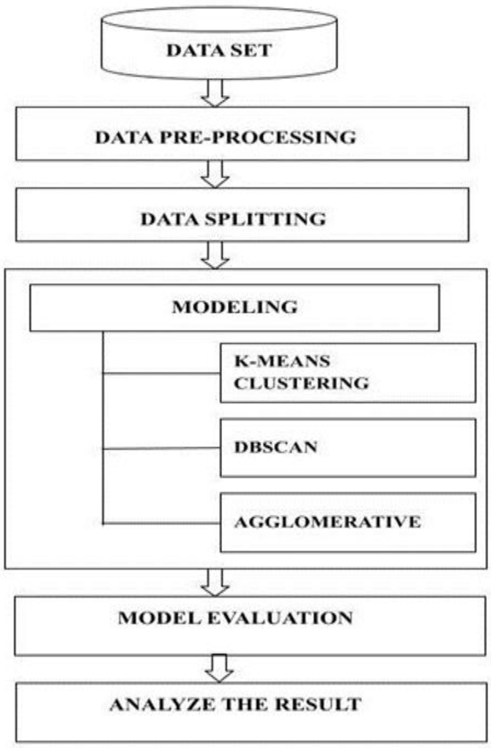

Fig.1.Methodology of the Research work

Clustering using K-Means:

K-Means clustering is an unsupervised machine learning algorithm that partitions a dataset into a specified number of distinct groups, or clusters, based on feature similarity. It works by initializing cluster centroids, assigning data points to the nearest centroid, and iteratively updating the centroids to minimize the variance within clusters. This method is widely used for pattern recognition, data segmentation, and grouping tasks.

Clustering using DBSCAN:

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is an unsupervised clustering algorithm that groups data points based on density, identifying clusters of high density separated by areas of low density. It defines clusters using two parameters: **eps** (maximum distance for neighboring points) and **minPts** (minimum points to form a dense region), labeling outliers as noise. DBSCAN is effective for irregularly shaped clusters and robust to noise but requires careful tuning of parameters

Clustering using Agglomerative:

Agglomerative clustering is a hierarchical clustering method that builds clusters by iteratively merging smaller clusters based on similarity or distance until all data points form a single cluster or a predefined number of clusters are achieved. It starts with each data point as its own cluster and uses a linkage criterion (e.g., single, complete, or average) to decide which clusters to merge. This approach is effective for visualizing data structure but can be computationally expensive for large datasets. Certainly, Here are the steps for the clustering algorithm in a concise format.

3.1 Pseudo Code for Proposed Methodology

Step 1: Collect dataset and preprocess it, including handling missing data. Data transformation using label encoder and view the correlation of all classes

Step 2: Split the data.

Step 3: Applying K-Means, DBSCAN, Agglomerative. Evaluate the model using Silhouette scores.

Step 4: Visualize and analyze the result for country, crop type, Adaptation strategies and Economic growth.

4. DATASET DESCRIPTION:

The dataset used for this study contains a comprehensive collection of various attributes such as Crop type, Average temperature in Celsius, Total Precipitation in millimeters, carbon dioxide emissions from agricultural activities, measured in metric tons. Below is a detailed description of the dataset features and then type.

The year of data were recorded. The country where the data was collected or where the agricultural activity occurred. The specific geographical region or area within the country. The type of crop being studied (e.g., wheat, rice). The average temperature in Celsius over a given period, typically a year. The total amount of rainfall or precipitation in millimeters during a specific period. The total carbon dioxide emissions from agricultural activities, measured in metric tons. The yield of a crop in metric tons per hectare of land. The number of extreme weather events, such as droughts or floods, recorded during the period. The percentage of agricultural land with access to irrigation systems. The amount of pesticide applied per hectare of agricultural land in kilograms. The amount of fertilizer used per hectare of land in kilograms. A numerical index representing the overall health and fertility of the soil. The strategies used by farmers to adapt to environmental or climatic challenges. The economic impact of agricultural factors, measured in millions of U.S. dollars.

Table i: Dataset of the research work

S.NO	Attribute Name	Type
1	Year	Int
2	Country	Object
3	Region	Object
4	Crop_Type	Object
5	Average_Temperature_C	Float
6	Total_Precipitation_mm	Float
7	CO2_Emissions_MT	Float
8	Crop_Yield_MT_per_HA	Float
9	Extreme_Weather_Events	Int
10	Irrigation_Access_%	Float
11	Pesticide_Use_KG_per_HA	Float
12	Fertilizer_Use_KG_per_HA	Float
13	Soil_Health_Index	Float

14	Adaptation_Strategies	Object
15	Economic_Impact_Million_USD	Float

5. EXPERIMENTAL RESULTS:

Silhouette scores are calculated using a formula that measures how similar a data point is to its own cluster compared to other clusters. The formula for the silhouette score s(i) for a data point I is:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Where:

- a(i): The average distance of i to all other points in the same cluster (intra-cluster distance).
- b(i): The average distance of i to all points in the nearest neighboring cluster (inter-cluster distance).
- $s(i) \in [-1,1]$

The **overall silhouette score** for a dataset is the mean of s(i) values for all points, providing a measure of clustering quality. The results from each model were evaluated using silhouette scores to assess the quality of the clustering.

Table ii. Comparison of three algorithms

S.NO	**Model**	**Silhouette Scores**
1	K-means	0.57
2	DBSCAN	0.46
3	Agglomerative	0.52

K-Means has the highest score (**0.57**), indicating that it provides the best clustering quality among the three models in this evaluation.

Agglomerative Clustering comes second with a score of **(0.52)**, which is still decent but not as good as K-Means.

DBSCAN has the lowest score (**0.46**), suggesting it may not be as effective for the given dataset compared to the other two methods.

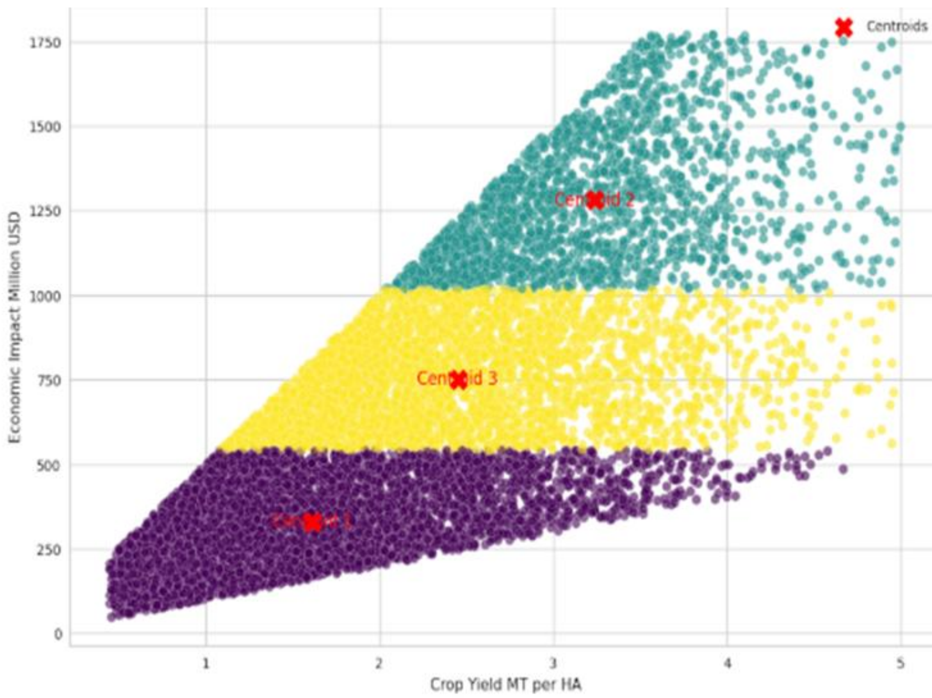

Fig. 2. Cluster visualization with centroids

Value Centroids:

Centroid 1: Crop Yield = 1.61, Economic Impact = 330.23

Centroid 2: Crop Yield = 3.24, Economic Impact = 1281.80

Centroid 3: Crop Yield = 2.45, Economic Impact = 751.39

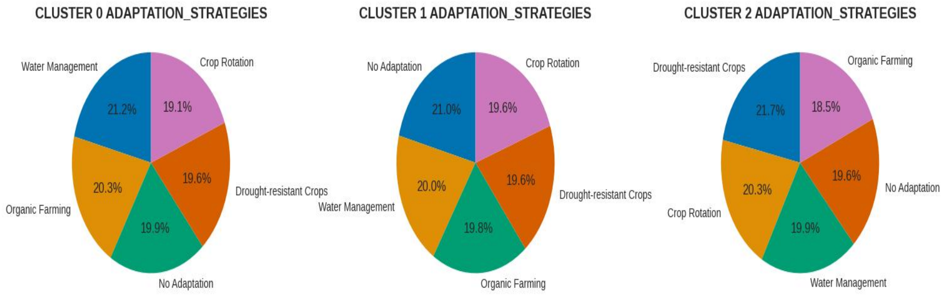

Fig. 3. Distribution of adaptation strategies across cluster 0,1 and 2.

Cluster 0: This cluster has the highest levels of CO2 emissions, extreme weather events, and pesticide use. However, it also has the lowest average temperature, crop yield, and the highest soil health index, resulting in the lowest economic impact compared to the other clusters.

Cluster 1: While Cluster 1 experiences higher average temperatures, more CO2 emissions, more extreme weather events, and a lower soil health index than Cluster 0, it has a higher economic impact. This could be due to factors such as lower total rainfall and higher crop yields compared to Cluster 0, or perhaps better access to irrigation, which could reduce the economic impact when compared to Cluster 2.

Cluster 2: It is the best in terms of economic impact, as it demonstrates the highest crop yields and overall positive outcomes despite its higher environmental challenges. Its stronger economic performance outweighs the increased use of pesticides, fertilizers, and extreme weather events.

6. CONCLUSION AND FUTURE SCOPE:

This research employs machine learning techniques and compares their effectiveness in solving real-world problems. The Kaggle datasets are used to assess the performance of three clustering algorithms: K-Means, DBSCAN and Agglomerative. K-Means emerged as the most effective and superior performance compared to other methods. The research findings reveal how different regions are impacted by climate change in terms of crop yield and economic impact, highlighting clusters of countries that face more severe challenges. The profiling of these clusters provides valuable insights for policymakers and farmers, offering a better understanding of which regions require more urgent adaptation strategies, such as irrigation access, pesticide use, and fertilizer management. The results from each model were evaluated using silhouette scores to assess the quality of clustering. This research provides better understanding and identifies patterns in how different regions and agricultural practices are impacted by climate change and ultimately linking higher crop yields to greater economic gains.

REFERENCES:

[1] Shams, M. Y., Tarek, Z., Elshewey, A. M., Hany, M., Darwish, A., & Hassanien, A. E. (2023). A machine learning-based model for predicting temperature under the effects of climate change. In The Power of Data: Driving Climate Change with Data Science and Artificial Intelligence Innovations (pp. 61-81). Cham: Springer Nature Switzerland. [Google Scholar]

[2] Rajesh, D. (2011). Application of spatial data mining for agriculture. International Journal of Computer Applications, *15*(2), 7-9.[Google Scholar]

[3] Andini, N., & Utomo, W. H. (2021). Climate Prediction Using RNN LSTM to Estimate Agricultural Products Based on Koppen Classification. *JISA* (Jurnal Informatika dan Sains), *4*(2), 96-99. [Google Scholar]

[4] Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environmental Research Letters, 13(11), 114003. [Google Scholar]

[5] Habib-ur-Rahman, M., Ahmad, A., Raza, A., Hasnain, M. U., Alharby, H. F., Alzahrani, Y. M., ... & El Sabagh, A. (2022). Impact of climate change on agricultural production; Issues, challenges, and opportunities in Asia. Frontiers in Plant Science, 13, 925548. [Google Scholar]

[6] Haq, M. A., Ahmed, A., Khan, I., Gyani, J., Mohamed, A., Attia, E. A., ... & Pandi, D. (2022). Analysis of environmental factors using AI and ML methods. Scientific Reports, 12(1), 13267. [Google Scholar]

[7] El-Habil, B. Y., & Abu-Naser, S. S. (2022). Global climate prediction using deep learning. Journal of Theoretical and Applied Information Technology, 100(24), 4824-4838. [Google Scholar]

[8] Adedeji, O., Reuben, O., & Olatoye, O. (2014). Global climate change. Journal of Geoscience and Environment Protection, *2*(2), 114-122. [Google Scholar]

[9] Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018).Present and future Köppen-Geiger climate classification mapsat1-km resolution. Scientific data, *5*(1), 1-12. [Google Scholar]

[10] Feng,Q.Y.,Vasile,R.,Segond,M.,Gozolchiani,A.,Wang,Y.,Abel,M.,...&Dijkstra,H.A. (2016). ClimateLearn: A machine-learning approach for climate prediction using network measures. Geoscientific Model Development Discussions, *2016*, 1-18. [Google Scholar]

[11] Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., ... & Zhu, Y. (2015). Rising temperatures reduce global wheat production. Nature climate change,*5*(2), 143- 147. [Google Scholar]

[12] Asseng, S., Martre, P., Maiorano, A., Rötter, R. P., O'Leary, G. J., Fitzgerald, G. J., ... & Ewert, F. (2019). Climate change impact and adaptation for wheat protein. Global change biology, *25*(1), 155-173. [Google Scholar]