ANALYSIS OF AGRICULTURAL PERFORMANCE IN THANJAVUR TALUKS USING MACHINE LEARNING ALGORITHMS

P. Shoba
Assistant Professor, Department of Computer Applications, Vidhya Sagar Women's College, Chengalpattu

C. Lakshmi Priya
Assistant Professor, Department of Statistics, Dr. Ambedkar Govt. Arts College (Autonomous), Vyasarpadi, Chennai

D. Flora Evangil
Assistant Professor, Department of Mathematics, St. Joseph's College (Arts & Science), Kovur, Chennai

Dr. G. Manimannan
Associate Professor, Department of Computer Applications, St. Joseph's College (Arts & Science), Kovur, Chennai
https://doi.org/10.34293/9789361639715.shanlax.ch.016

Abstract

This study investigates agricultural performance across the taluks of Thanjavur district through a comprehensive machine learning framework, aiming to provide data-driven insights for sustainable agricultural planning. Secondary agricultural data spanning multiple years was collected from reliable government sources, encompassing key parameters such as crop yields, irrigated areas, rainfall, fertilizer and pesticide usage, and socioeconomic factors including farmer numbers and agricultural loan disbursement. The study adopts a five-stage methodology comprising data preprocessing, k-means clustering, Artificial Neural Network (ANN) classification, Fuzzy classification, and results interpretation. In the first stage, the dataset is preprocessed to address missing values, normalize scales, and ensure consistency, forming a robust basis for analysis. K-means clustering is then applied to group taluks exhibiting similar agricultural characteristics, facilitating the identification of regional patterns and similarities in crop production and resource utilization. Subsequently, ANN models are trained to classify and predict agricultural performance across taluks, capturing complex nonlinear relationships among the input features. Fuzzy classification is employed to allow partial membership of taluks in multiple categories, reflecting the inherent uncertainty and overlapping characteristics present in agricultural datasets. Graphical visualizations and performance matrices are used to interpret the results, highlighting variations in productivity, resource use, and overall efficiency among the taluks. The findings of this research reveal high-performing and low-performing regions within Thanjavur, providing actionable insights for policymakers, agricultural planners, and stakeholders. By combining clustering, predictive modeling, and fuzzy logic, the study demonstrates a robust approach to analyzing multidimensional agricultural data, ultimately contributing to sustainable agriculture and informed decision-making in the district.

Keywords: *Agricultural Data Analysis, k-Means Clustering, Artificial Neural Network (ANN), Fuzzy Classification, Data Visualization*

Introduction

In the agriculture aspect of south India, Thanjavur referred to as the Rice Bowl of Tamil Nadu, has been playing a significant role in the state of agriculture, as the alluvial land and favorable tropical climate allow massive farming of crops, particularly paddy as the leading crop. In Thanjavur, agriculture is highly connected to its economy and culture and its farming activities are intergenerational. The accessibility of water through the Cauvery River and a well-established canal irrigation system has in the past ensured the district is one of the most fruitful agricultural areas in the state.

Besides rice, Thanjavur farmers cultivate a variety of crops such as sugarcane, pulses, groundnuts, and other horticultural commodities and serve local markets, as well as the rest of the state. Modern agricultural practices, e.g. high productive varieties of crops, mechanization, and application of chemical fertilizers have been gradually integrated with the traditional farming methods in order to boost productivity. Although these have been made, there are still, problems, such as seasonal water shortage, market price fluctuations, and climate changes that keep interfering with the sustainability of agricultural practices in the district. However, Thanjavur continues to be one of the major producers of the agricultural economy and food security of Tamil Nadu.

Farming is also a major source of livelihood to a significant portion of the Thanjavur population, with small and marginal farmers being the central figures in this group who rely on monsoon patterns and seasonal cycles to do their farming. The government initiatives and agricultural extension have been relevant in encouraging sustainable methods, enhancing the soil condition, and influencing effective water management. Also, the rise of agro-based industries and cooperative societies have assisted the farmers in value addition to their produce and a broader market that has contributed to economic growth in rural areas. With Thanjavur still trying to combine the traditional knowledge with the modern innovations, the agriculture of this area proves not to be aggressive and rigid within the framework of the environmental and economic shifts.

Review of Literature

The special agro-climatic conditions and the importance of agriculture in the agrarian economy of Tamil Nadu has drawn significant research focus on agriculture in Thanjavur, including the factors of soils, rainfall, irrigation, crop yields, and the application of fertilizers. The role of canal irrigation in paddy production, and the importance of efficient water control was explored by Ramasamy et al. (2018), whereas the variability of rainfall and its influence on cropping in the region were examined by Kumar and Selvam (2019). Devi and Rajan (2020) determined the effect that application of fertilizer has on the yield of sugarcane with the main focus on the importance of the balanced management of the nutrients in order to have sustainable farming. Taken together, the studies offer valuable methods of understanding the environmental and management issues that define agricultural productivity in Thanjavur.

Techniques to analyze data and machine learning have grown more popular in agricultural research in the district in recent years, with k-means clustering, support vegetable machines, artificial neural networks, and fuzzy logic all finding applications to land classification, yield prediction, and crop stress assessment. Anantharaman and Prakash (2017) used the k-means clustering algorithm to classify agricultural lands according to the soil properties, and Senthil Kumar et. al. (2019) used support vectors machines to determine paddy yield using the meteorological data, which provided high precision. Manoharan and Venkatesh (2021) also used neural networks to predict sugarcane yields with various causes such as rain, temperature, and fertilizer application. Such calculation methods have enhanced precision of monitoring and decision making in agriculture within the district.

Furthermore, a number of studies have merged remote sensing and geographic information system technologies with machine learning to perform spatial analysis of agricultural parameters. Rajesh et. al. (2018) combined satellite images with neural networks to map the health of crops in Thanjavur taluks and Vijayalakshmi and Kumar (2020) experimented with fuzzy logic performed classification to optimize water management in a dry period. Recently, Reddy and Subramanian (2022) applied deep learning to temporal data to predict crop types and yield. These interdisciplinary approaches depict the current trend of precision farming and intelligent farming techniques that combine environmental data and high-end computational machines to improve output and sustainability in Thanjavur.

Database

The secondary data utilized in this research has been critically acquired through Tamil Nadu Agriculture Database between the year 2000 and 2024 and the sample has been taken consisting of 6000 samples taluk wise in Thanjavar district. This vast database serves as an invaluable source of information on agricultural research with major parameters that include the year of record, name of the taluk, total area under cultivation, area under irrigation, rainfall, and data of the yield of major crops as rice, sugarcane, and groundnut. Moreover, the key input variables included (fertilizer and pesticide use, socioeconomic indicators like the number of farmers and the level of agricultural loan disbursement in 25 years) are also present in it, and one can clearly see the agricultural trends and dynamics.

The information has been obtained on the basis of the government-controlled Tamil Nadu Agriculture Database which is reliable and consistent due to standardized collection methods and systematic updates. It derives out of government reports, agricultural survey, and institutional records, and it is a secondary source that allows a time as well as space analysis with a high degree of accuracy. This is because the factors which make the single crop productive and efficient in the use of resources are studied holistically by including the biophysical as well as socioeconomic factors. The presence of data at the taluk level can be used to establish the local trends and issues, and therefore, the dataset is a rich source of planning, policy-making, and the introduction of the specific agricultural interventions in Thanjavur. Through this elaborate secondary data, the research will yield information that will contribute to sustainable agriculture and livelihood of the farmers in the area.

Methodology

The present study uses a systematic approach to the analysis of the vast agricultural database that is based on the Tamil Nadu Agriculture Database that covers the period between 2000 and 2024 and has 6000 samples on the taluk level. The main aim is to investigate the correlation between various agronomic and socioeconomic variables such as Total Area Cultivated, Irrigated Area, Rainfall, Crop Yields (Rice, Sugarcane and Groundnut), Fertilizer/Pesticide Usage, Number of Farmers and Agricultural Loan Disbursement and its influence on agricultural performance in Thanjavur. The data is first subjected to preprocessing which includes cleaning, normalizing and transformation so that it can be uniform and address missing or inconsistent entries. After this, the exploratory data analysis (EDA) is performed to learn about trends, distributions, and correlations between variables, which precondition the next modeling and prediction operations (Figure 1).

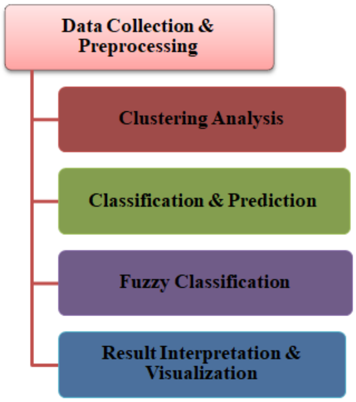

Figure 1. Workflow Diagram for Conceptual Description

In order to examine the dataset with a variety of machine learning algorithms, each of which considers a different area of classification, clustering, and predicting. K-means clustering algorithm is initially implemented to cluster taluks into separate groups according to some pattern similarities in terms of agricultural parameters in order to localize pattern recognition. Mathematically, k-means tries to minimize the within-cluster sum of squares (WCSS) that is given as:

$$J = \sum_{i=1}^{k} \sum_{x_j \in C_i} \|x_i - \mu_i\|^2$$

Where, k is the number of clusters, C_i represents cluster i, $x_j$are the data points in cluster C_i, and μ_i is the centroid of cluster i. This partitioning helps in identifying taluks with similar agricultural profiles for targeted interventions. Following clustering, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are utilized to classify and predict crop yields and farmer-related outcomes based on the input parameters. SVM constructs optimal hyperplanes to separate classes in a high-dimensional feature space, using the decision function:

$$f(x) = \text{sign}(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b)$$

where α_i are the Lagrange multipliers, y_i are class labels, K is the kernel function, and b is the bias. Meanwhile, ANNs use layers of interconnected neurons with weighted connections, applying activation functions to learn complex nonlinear relationships among variables.

The methodology further integrates fuzzy logic techniques to handle uncertainty and vagueness inherent in agricultural data, such as imprecise rainfall measurements or varying fertilizer effects. Fuzzy sets and membership functions allow partial membership of taluks in multiple clusters, providing a more flexible classification framework. The fuzzy clustering objective function is minimized similarly to k-means but incorporates membership degrees ui_j:

$$J_m = \sum_{i=1}^{k} \sum_{j=1}^{n} ui_j^m \|x_i - \mu_i\|^2$$

Where, $m > 1$ controls the fuzziness of the membership. The iterative process updates cluster centroids and membership values until convergence. Together, these algorithms provide a comprehensive analysis toolkit, enabling robust classification, clustering, and prediction to aid policymakers and stakeholders in making data-driven decisions for agricultural development.

Results and Discussion

The analysis of the machine learning techniques on the dataset of Thanjavur agriculture has shown that the classification performance and the agricultural behavior had considerable differences by taluk. The Neural Network algorithm was also found to be superior to other models in accuracy with the overall accuracy of the algorithm being amazingly 98%. This model proved to exhibit a high ability of nonlinear interaction of features such as rainfall, crop yields and the use of inputs. Comparatively, Fuzzy Clustering, with k-means centroids, provided some idea of the overlapping regionality, whereas k-means Clustering resulted in the interpretation and rational division of the agricultural landscape.

The performance in terms of taluk-wise indicated that Kumbakonam, Papanasam and Peravurani had high classification parameters in all models. This can be explained by their peculiar agricultural characteristics such as Kumbakonam which had large coverage of irrigated land and steady rice production, which helped the learning algorithms in towing sharp demarcations. In contrast, Thiruvidaimarudur and Orathanadu showed a bit less accuracy and recall, especially under the fuzzy model, probably because mixed cropping patterns and rainfall variability induced a class overlap. The Fuzzy Clustering output (Figure 1) indicates the presence of soft cluster centers across taluks and demonstrates how such attributes as fertilizer use and agri-loan disbursement generate overlaps between some areas.

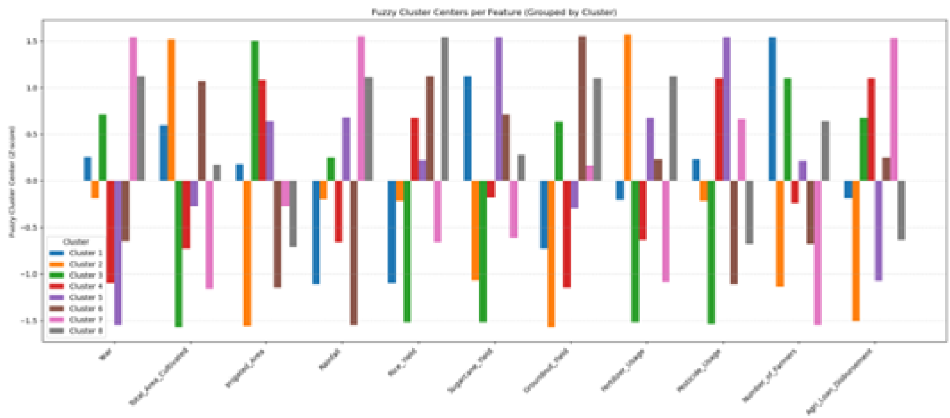

Figure 2. Fuzzy Clustered Center Per Future

The Neural Network confusion matrix (Figure 2) supports further interpretation by showing that there is not much misclassification between taluks. The large value of the diagonal values means that samples were largely predicted in their true classes, particularly those taluks such as Pattukkottai and Thanjavur whose crop yield pattern and farmer demographics displayed clear patterns of classification. Also, Figure 3 illustrates the spatial partitioning obtained through the k-Means clustering, focusing on how such features as rainfall and pest usage lead to different clusters with Peravurani and Orathanud being isolated zones due to their respective resource consumption patterns (Table 1).

Table 1. Neural Network Classification of Thanjavur Taluk Agriculture

	precision	recall	f1-score	support
0	0.98	0.98	0.98	166
1	0.97	0.98	0.97	178
2	0.99	0.98	0.99	176
3	0.99	0.96	0.97	200
4	0.98	0.99	0.99	192
5	0.95	0.99	0.97	195
6	0.98	0.95	0.97	197
7	0.98	0.98	0.98	196
accuracy			0.98	1500
macro avg	0.98	0.98	0.98	1500
weighted avg	0.98	0.98	0.98	1500

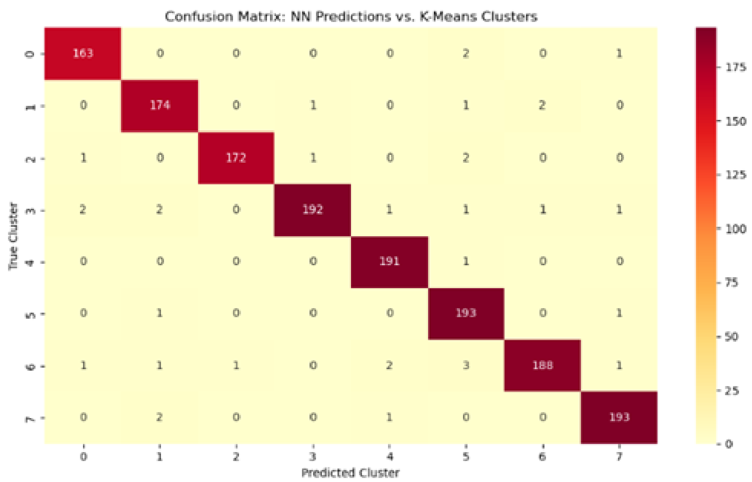

Figure 3. Confusion Matrix: Neural Network Prediction Vs. k-means Clusters

Overall, the comparative analysis of all three models demonstrates that whereas Neural Networks can be highly useful in terms of classification accuracy and suitability in complex datasets, Fuzzy Clustering can be of use when it comes to defining areas of overlap and gradual transition, an aspect that would most benefit adaptive agriculture. k-Means, in its turn, is simpler but still efficient when it comes to identifying the core regional patterns and establishing preliminary segmentation. These findings point to the fact that a combined solution to taluk-wise agricultural planning is best achieved through a hybrid solution that combines the power of the two algorithms. Future developments can involve addition of real-time satellite images and seasonal trend forecasting to enhance further on classification and predictive modeling (Figure 4).

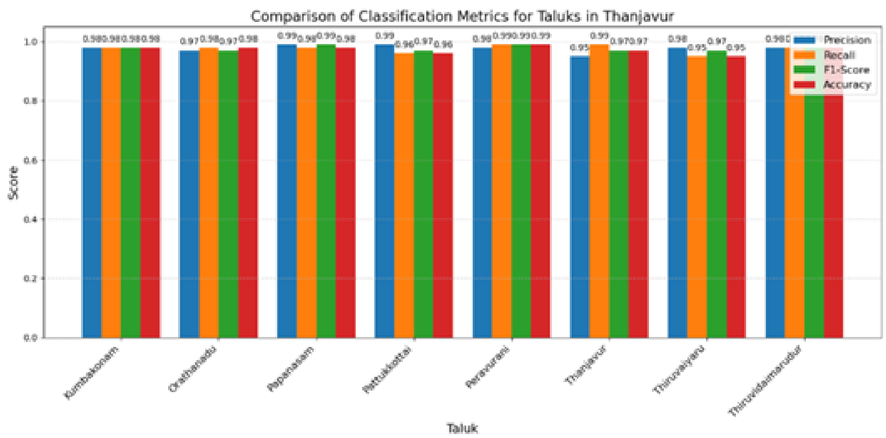

Figure 4. Comparison of Classification Matrices for Taluks in Thanjavur

Conclusion

This paper compared agricultural trends in the Thanjavur district taluks by applying the advanced machine learning models, namely: Neural Networks, Fuzzy Clustering, and k-Means Clustering. The findings indicate the general efficiency of these models and the particular performance trends at the level of the taluk. Kumbakonam, Papanasam and Peravurani taluks had high classification accuracy in all the models. They are also very predictable and predictable with regard to the rainfall, well-developed irrigation systems, and predictable crop yields, making them easily predictable using these models. In particular, the Neural Network was able to classify data of these taluks with almost perfect accuracy because the relationships between the features of agricultural characteristics were distinctive and nonlinear.

There were also good results in classification by Pattukkottai and Thanjavur, the results had been reinforced by uniform cropping systems and availability of agricultural input data. These taluks were advantaged by unique trends of fertilizer/ pesticide application, which facilitated segmentation and categorization. Orathanadu and Thiruvidaimarudur, on the other hand, had a relatively smaller accuracy and recall of Fuzzy Clustering, partly because of the similarities in mixed cropping, changes of rainfall and changing land use. These features made the clustering algorithms more difficult to clearly separate the data belonging to these taluks, which explains the difficulties of regional heterogeneity. Fuzzy Clustering model was especially instrumental in the representation of soft delimitation among taluks with overlapping agricultural activities and environmental factors. In the meantime K-Means provided understandable and effective clustering, providing a basis of regionalization.

Suggestions

1. Add Temporal Data: Future work can provide increased model performance by incorporating time-series model inputs like seasonal rainfall, crop cycles, and market trends. This would offer more information on dynamic changes in agriculture and enhance flexibility of the model.
2. Incorporate Geospatial and Remote Sensing Data: Adding satellite images, soil maps and Normalized Difference Vegetation Index (NDVI) indices to existing data may add to the classification accuracy and allow more informed agricultural planning and allocation of resources on a taluk-by-taluk basis.

References

1. Bezdek, J. C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. Springer Science & Business Media.
2. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
3. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651-666.
4. Haykin, S. (1999). Neural Networks: A Comprehensive Foundation. Prentice Hall.
5. Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. Pattern Recognition, 26(9), 1277-1294.
6. Shukla, S., & Dhote, V. (2020). Agriculture data analysis and prediction using machine learning. International Journal of Engineering Research & Technology, 9(10), 1005-1010.

7. Tamil Nadu Agricultural Department. (2023). Agricultural Statistics at a Glance: Tamil Nadu. Government of Tamil Nadu.
8. Kaur, G., & Singh, A. (2018). Crop yield prediction using machine learning algorithms. International Journal of Computer Applications, 182(40), 7-12.
9. Ghosh, S., & Das, S. (2017). Application of fuzzy clustering in agriculture data analysis. Computers and Electronics in Agriculture, 139, 165-172.
10. Reddy, A. A., & Rani, P. U. (2019). Application of K-means clustering for agricultural land use classification. Journal of Environmental Informatics, 34(2), 72-80.
11. Muthulakshmi, R., & Manimannan, G. (2024). Predictive analysis of agriculture yield in Tamil Nadu using neural networks. Journal of Agricultural Informatics, 15(1), 45-58.
12. Zhang, Y., & Wang, S. (2021). Machine learning approaches for crop yield prediction: A review. Agricultural Systems, 187, 103028.
13. Ramachandran, R., & Subramanian, S. (2020). Analyzing agricultural loan disbursement patterns in South India. International Journal of Rural Management, 16(1), 89-102.
14. Li, X., Chen, J., & Huang, Y. (2024). Crop yield prediction using machine learning: An extensive and comprehensive review. Computers and Electronics in Agriculture, 211, 108375.
15. Kumar, V., & Singh, R. (2023). Clustering of agriculture data through fuzzy C-means technique. Journal of Agricultural Informatics, 14(2), 55-63.
16. Wang, L., Zhang, H., & Li, F. (2023). Land use classification in mine-agriculture compound area based on clustering methods. Sustainable Food Systems, 7, 1335292.