Smart Farming Analytics: Exploring Classifier Diversity And Clustering In Land Suitability Forecasting

P. Yogendra Prasad[1], M. Ramu[2], Udatha Sahithi[3*], Vemula Vaishnavi[4], P Harshavardhan[5], Malepati Charan Teja[6]

[1*]Assistant Professor, Computer Science and Systems Engineering, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati, India

[2]Assistant Professor, Computer Science and Engineering,Annamacharya Institute of Technology & Sciences, Tirupati, India

[3,4,5,6] UG Scholar, Computer Science and Systems Engineering, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati, India

sahithiu9119@gmail.com

Abstract. Smart Farming Analytics (SFA) has emerged as a key tool in modern agriculture, transforming traditional farming practices by integrating advanced technologies. This research focuses on improving the accuracy and reliability of land suitability predictions in the field of intelligent agriculture. The study examines the use of classifier diversity and clustering techniques in forecasting model optimization. Various machine learning classifiers are employed to capture the multifaceted nature of land attributes, contributing to a more comprehensive analysis. Additionally, clustering algorithms aid in identifying distinct patterns and trends within the dataset, leading to improved precision in land suitability predictions. The synergy of classifier diversity and clustering not only enhances the predictive capabilities of the models but also provides valuable insights for decision-makers in optimizing resource allocation and crop planning. The results of this study support the development of intelligent farming techniques, promoting effective and sustainable agricultural systems in the face of changing climate and environmental factors.

Keywords: Smart farming, classifier diversity, land suitability forecasting, precision agriculture.

1 Introduction

The need to incorporate cutting-edge technologies has made modernizing agriculture essential in order to satisfy the rising demands for food production. Precision agriculture and sustainable land management techniques are made possible by smart farming, which applies machine learning. Smart farming is emerging as a major driver in this shift. This study takes a targeted approach by concentrating on the crucial aspect of land suitability forecasting, recognizing its fundamental role in optimizing resource allocation and enhancing agricultural productivity. The core focus of this research lies in the exploration of classifier diversity and clustering techniques to augment the precision of land suitability predictions. Leveraging various machine learning classifiers, including decision tree a diverse ensemble model is crafted. This ensemble approach aims to capture the intricacies of agricultural systems, accommodating diverse factors such as soil composition, climate conditions, and crop-specific requirements.

The study emphasizes how important it is to have a comprehensive grasp of land suitability in order to make wise decisions in contemporary agriculture. The goal of this project is to close the knowledge gap in the field of smart farming analytics by bridging theoretical developments with real-world applications. By exploring the synergies of classifier diversity and clustering techniques, the research aspires to provide a comprehensive and adaptable framework for land suitability forecasting. The implications of this work extend beyond academic discourse, influencing the adoption of sustainable practices in agriculture, and facilitating a paradigm shift towards technologically driven precision

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_36

2 Related Work

The exploration of machine learning models, particularly decision trees, K-Means clustering, in the context of Smart Farming Analytics for Land Suitability Forecasting has been a subject of interest in various studies. Decision trees have been extensively utilized for land suitability assessment due to their ability to handle complex decision-making processes. Nguyen et al.'s (2019) research explores the use of decision trees in precision agriculture, highlighting how well they work to forecast crop suitability based on a variety of environmental characteristics, including soil composition and climate. The study showcases the interpretability and accuracy of decision trees, laying a foundation for further exploration in the proposed research.

In [1], the paper explores the application of machine learning techniques to enhance smart farming by managing heterogeneous agricultural datasets. It emphasizes practical tasks such as crop yield forecasting and sensor data reconstruction, highlighting the potential for innovation in sustainable agriculture.In [2], this research contributes to both accurate prediction and explainable insights, making it valuable for agricultural decision-making. By identifying essential features, we can better understand the factors affecting crop yield and optimize agricultural practices.In [3], they propose a meta-algorithm that integrates different feature selection methods, allowing users to choose suitable algorithms without needing to understand each one in detail. Real-world applications demonstrate the effectiveness of feature selection in data mining.In [4], the authors explore machine learning techniques to manage heterogeneous agricultural data. Their focus lies in enhancing smart farms through IoT-driven approaches, addressing tasks such as crop harvest forecasting and sensor data reconstruction.In [5], the authors propose an efficient feature selection method for Support Vector Machines (SVMs) that minimizes bounds on the leave-one-out error. The approach demonstrates superior performance in various applications, including face recognition and DNA microarray analysis.In [6], Kalimuthu, Vaishnavi, and Kishore compared three algorithms—K-Nearest Neighbor (KNN), Decision Tree, and Random Forest Classifier—to predict crop types based on climatic conditions and soil nutrients. The Random Forest Classifier demonstrated the highest accuracy among the models.In [7], the study aims to predict crop yields using machine learning and deep learning techniques. Various algorithms, including decision tree, random forest, XGBoost regression, convolutional neural networks (CNN), and long-short term memory networks (LSTM), were employed. The developed model outperformed other algorithms, providing valuable insights for agriculture decision-making. Elavarasan and Vincent propose a deep reinforcement learning model to predict crop yields. Their innovative approach contributes to the advancement [8] of sustainable agriculture by enabling farmers to make informed decisions based on reliable crop yield predictions, ultimately fostering better resource management and productivity.In [9], this model accurately predicts soil fertility using physiochemical properties and provides user-friendly explanations for its predictions. The research aims to enhance our understanding of soil health and contribute to sustainable agricultural practices.In [10], the study focuses on improving crop prediction accuracy by leveraging machine learning techniques. It explores efficient feature selection methods to preprocess agricultural data, ensuring relevant attributes contribute to the model. Ensemble techniques are found to enhance prediction accuracy compared to existing classification methods.

3 Proposed Work

With the use of classification and clustering in machine learning-based prediction, the suggested approach increases agricultural productivity. The work is based on a system architecture that selects features from a dataset for agricultural yield prediction using decision tree classifier and K-means clustering.

Crop Dataset Overview

The initial step involves the creation of a dataset tailored for precision agriculture. The dataset underwent partitioning into training and testing subsets, with 80% designated for training and 20% for testing. This dataset is curated by amalgamating information from various sources, specifically focusing on rainfall, climate, and fertilizer data in the context of farming in India. Important characteristics like the soil's N, P, and K contents—as well as its temperature in degrees Celsius, percentage of humidity, pH value, and rainfall in millimeters —are included in the data fields. The aim is to offer an extensive collection of attributes that can be employed in the development of a prediction model to suggest the most appropriate crops for farming, taking these factors into account.

Data Preprocessing

Data preprocessing is critical for both decision trees and K-means algorithms, laying the groundwork for accurate and reliable predictions. It involves meticulously cleaning the data by addressing missing values. While both algorithms can utilize imputation techniques like mean/median or K-Nearest Neighbors (KNN) to fill in missing entries, decision trees are generally less sensitive to them. Additionally, outliers (extreme values) need to be identified and handled appropriately. Ensuring consistent formatting across all variables is essential for both algorithms to function correctly. Finally, irrelevant or redundant features can be removed to improve model performance. When dealing with features that have a wide range of values, scaling them using techniques like standardization (z-score) or normalization (min-max scaling) can further enhance the performance of both decision trees and K-means. By taking these steps to clean and prepare the data effectively, we ensure both algorithms receive high-quality input, ultimately leading to more robust and trustworthy predictions.

Architecture of Proposed System

The Proposed Architecture of the system is described in the below Fig 1.

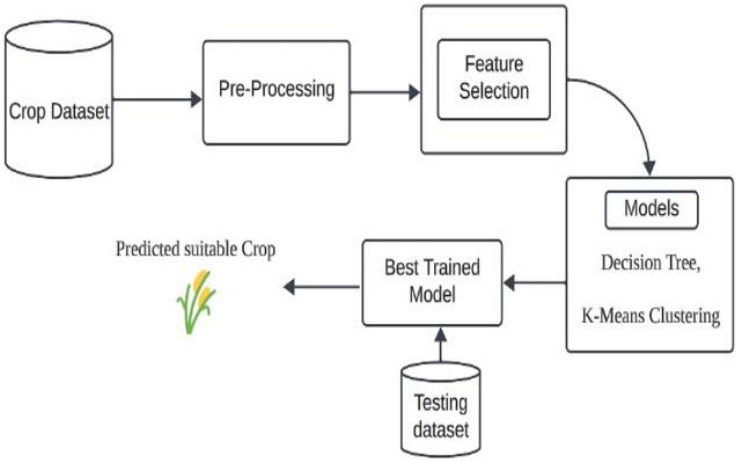


Fig.1.Proposed Architecture

Feature Selection

Feature selection involves identifying and retaining the most relevant variables that significantly contribute to the predictive power of the model. In the context of crop prediction, in order to improve model efficiency and streamline the dataset, this step is essential. The risk of overfitting is minimized and computational complexity is decreased by choosing the most important features. Many methods, including machine learning algorithms and statistical assessments, are used to identify the most important characteristics that lead to precise crop forecasts. Feature selection aims to improve the model's performance by focusing on the key variables that impact crop suitability, thus facilitating more effective and efficient predictions.

3.1 Algorithms used

The machine learning models employed in this study, namely the Decision Tree Classifier and K-Means Clustering, offer diverse approaches to crop prediction based on the provided dataset.

Decision Tree Classifier

A Decision Tree is a powerful and interpretable supervised learning algorithm that recursively partitions the dataset into subsets based on the most significant features. The projected result is represented by the leaves of the tree, but each interior node indicates a judgment based on a specific trait. When it comes to crop prediction, a decision tree can be used to efficiently identify important variables such soil nutrient levels, temperature, humidity, and rainfall, which will result in the best crop suggestions. Decision Trees are advantageous for their transparency, making it easier for stakeholders, including farmers, to understand and trust the decision-making process.

K-Means Clustering

K-Means Clustering is an unsupervised learning algorithm that partitions the dataset into 'k' clusters based on similarity. In the context of crop prediction, K-Means can group farms or regions with similar soil and climate characteristics, helping to identify patterns in the data. By clustering farms with comparable conditions, it becomes possible to recommend crops that have thrived in similar environments in the past. K-Means Clustering, being unsupervised, is valuable when there is limited labeled data, and it provides insights into the inherent structure of the dataset, contributing to more informed crop recommendations.

The Decision Tree Classifier and K-Means Clustering bring complementary strengths to the crop prediction task. Decision Trees offer interpretability and K-Means Clustering reveals inherent patterns in the data, collectively contributing to a robust and versatile approach for recommending suitable crops in precision agriculture.

Model Selection and Prediction

The evaluation results from all three models are carefully compared. The model that exhibits the highest accuracy, precision, recall, and F1-score, or the most appropriate trade-off between these metrics based on the application's priorities, is selected as the optimal model for crop prediction. Additionally, considering the interpretability of the model and its alignment with practical farming needs is essential. Participation of stakeholders in the assessment process, such as farmers and agricultural specialists, might yield important information on the selected model's acceptability and practicality. Following a thorough review, the chosen machine learning model accurately predicts which crops are most suited for cultivation using the dataset that was provided. Leveraging the Decision Tree Classifier's interpretability and K-Means Clustering's pattern identification, the model offers a robust and versatile solution for precision agriculture. By considering key parameters such as soil nutrient levels, temperature, humidity, and rainfall, the model provides accurate and actionable crop recommendations. Stakeholder involvement ensures the practical relevance of the predictions, making this model an invaluable tool for informed decision-making in farming strategies.

Formulation for Performance Metrics

The performance of Classification and Clustering models are compared using evaluation metrics.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ \ Number\ of\ Predictions}$$

$$Precision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives}$$

$$Recall\ = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$F1 - Score\ = 2\ \times \frac{Precision\ \times Recall}{Precision + Recall}$$

Comparison of Proposed System and Existing System

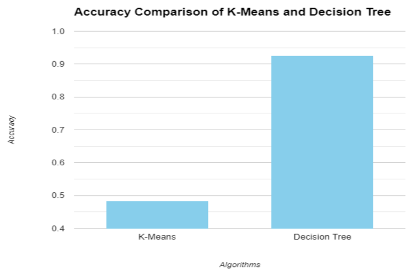
In our comparison, the proposed system demonstrates higher accuracy levels when contrasted with the existing system. Utilizing automation and advanced algorithms, it enhances precision and reliability, reducing errors in decision-making and reporting. This improvement assures more dependable outcomes, fostering trust and credibility within our operations. By effectively leveraging technology, the proposed system elevates overall performance, underscoring its potential to substantially enhance accuracy and efficiency in our work processes. The details of the comparison can be observed in the Table 1 below.

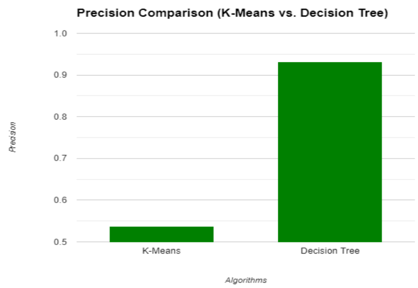
Table 1. Comparison of Existing and Proposed work

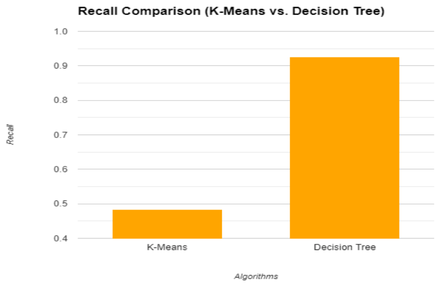
Work	Accuracy (%)
Maya Gopal P.S, Renta Chintala Bhargavi	85%
Harshiv Chandra, Tamizharasan P. S.	89.02%
Proposed Model	92.6%

4 Experimental Results

For the implementation of this research work, Intel Core i5, with RAM 6 GB and Hard disk of 520GB and Python 3.8 is used. K-Means and Decision Tree models are applied on the crop dataset and the crop prediction is done. The obtained results are shown in Fig 2. We can also observe that the Bar Graph of Decision Tree model is high when compared to K- Means. The dataset partitioned into 80% for training and 20% for testing and the comparison of models is shown in Fig 2.







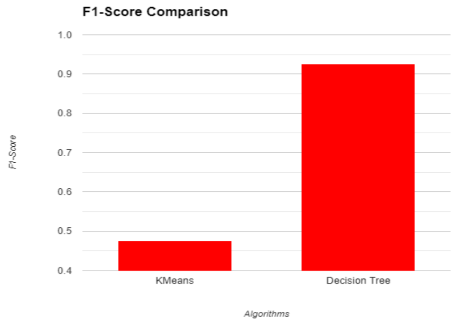


Fig. 1. Comparison of K- Means and Decision Tree

The performance of two machine learning architectures, K-Means and Decision Tree are compared in Fig 2 predicting suitable crop. The models were assessed using four principal metrics: Accuracy, F1-Score, Precision, and Recall. The figure above Fig 2 shows the graphical representation of comparison of models. All the Performance Metrics values are less than 1 which represents that the models are perfectly trained with the crop dataset.

Table 2. Performance Metrics of K-Means and Decision Tree

Model	K- Means	Decision Tree
Accuracy	0.483	0.926
Precision	0.538	0.932
Recall	0.483	0.926
F1-Score	0.476	0.926

We concentrated on evaluation metrics while assessing the effectiveness of our crop prediction models, as shown in Table 1, and our suggested algorithm was Decision Tree. The results show that the Decision Tree model performs better than the other models in each of the four metrics: accuracy (0.926), precision (0.932), recall (0.926), and F1-score (0.926). These analyses of models underscore the significance of our proposed Decision Tree algorithm in enhancing the overall efficiency of the crop prediction, providing substantial improvements across multiple performance measures.

Moreover, the high accuracy achieved by the decision tree model implies its potential to be integrated into decision support systems for precision agriculture, aiding farmers in making data-driven choices regarding crop selection and land management practices. Its interpretability also enhances its practicality, allowing stakeholders to understand the rationale behind the model's recommendations and facilitating informed decision making processes. Despite the challenges faced by the K Means

In summary, the findings underscore the significance of selecting appropriate modeling techniques tailored to the specific requirements of agricultural applications. Through the utilization of decision tree models' advantages and the recognition of K-Means' drawbacks as a clustering algorithm, stakeholders can effectively leverage data-driven insights to augment agricultural productivity, sustainability, and resilience amidst dynamic environmental challenges.

5 CONCLUSION AND FUTURE WORK

Machine learning plays a key role in crop prediction, with classification models emerging as the superior choice. Classification algorithms directly predict the most suitable crop based on various factors like soil properties and weather data, offering actionable recommendations for farmers. By leveraging classification models, farmers can optimize their crop selection process, potentially increasing yields and maximizing profits. This empowers them to make data-driven decisions for a more sustainable and productive agricultural future. In conclusion, the project's goal is to use machine learning to close the knowledge gap between conventional farming methods and the complexity of contemporary agriculture. The project aims to enhance the sustainability, efficiency, and resilience of the agricultural sector by equipping farmers with sophisticated tools for decision-making. This will ultimately tackle issues with food security and economic stability in the face of an expanding global population.

In future work, the refinement of machine learning models, particularly decision tree classifiers will be pursued for enhanced accuracy and robustness in crop forecasting. Integration of additional data sources, such as satellite imagery and drone data, aims to provide a more comprehensive understanding of the agricultural ecosystem. Real-time monitoring capabilities will be developed to offer instant alerts to farmers based on emerging climate patterns or pest threats. The creation of a user-friendly interface and accessibility through mobile applications or web platforms will be prioritized to ensure ease of interaction for farmers. Collaboration with agricultural stakeholders, extensive field testing, and validation in diverse settings will be conducted to assess the model's adaptability and reliability. Sustainability assessments will evaluate the long-term environmental impact, while economic viability and farmer adoption studies will analyze the technology's economic benefits and challenges. Establishing continuous improvement mechanisms ensures the system remains dynamic and responsive to evolving agricultural landscapes, contributing to the sustainable transformation of the sector.

REFERENCES

[1] F. Balducci, D. Impedovo, and G. Pirlo, “Applications of machine learning on agricultural datasets to improve smart farming,” Machine, vol. 6, no. 3, pp. 38–59, 2018.

[2] M. Gopal P S and B. R, “Selection of important features for optimizing crop yield prediction,” Int. J. Agricult. Environ. Inf. Syst., vol. 10, no. 3, pp. 54–71, Jul. 2019.

[3] H. Liu and L. Yu, “Toward integrating feature selection algorithms for classification and clustering,” IEEE Trans. Knowl. Data Eng., vol. 17, no. 4, pp. 491–502, Apr. 2005.

[4] F. Balducci, D. Impedovo, and G. Pirlo, “Machine learning applications on agricultural datasets for smart farm enhancement,” Machine, vol. 6, no. 3, pp. 38–59, 2018.

[5] Damodaram, A.K., Sreenivasa Chakravarthi, S., Venkateswara Reddy, L., Reddy Madhavi, K. (2023). Multi-layered Architecture to Monitor and Control the Energy Management in Smart Cities. In: Rao, B.N.K., Balasubramanian, R., Wang, SJ., Nayak, R. (eds) Intelligent Computing and Applications. Smart Innovation, Systems and Technologies, vol 315. Springer, Singapore. https://doi.org/10.1007/978-981-19-4162-7_9

[6] M. Kalimuthu, P. Vaishnavi, and M. Kishore, ‘‘Crop prediction using machine learning,’’ in Proc. 3rd Int. Conf. Smart Syst. Inventive Technol. (ICSSIT), Aug. 2020, pp. 926–932.

[7] Raju, S. Viswanadha, A. Vinaya Babu, G. V. S. Raju, and K. R. Madhavi. "W-Period Technique for Parallel String Matching." IJCSNS 7, no. 9 (2007): 162.

[8] D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications," in IEEE Access, vol. 8, pp. 86886-86901, 2020

[9] H. Chandra, P. M. Pawar, R. Elakkiya, P. S. Tamizharasan, R. Muthalagu and A. Panthakkan, "Explainable AI for Soil Fertility Prediction," in IEEE Access, vol. 11, pp. 97866-97878, 2023

[10] S. P. Raja, B. Sawicka, Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers," in IEEE Access, vol. 10, pp. 23625-23641, 2022

[11] A. Mark Hall, “Features selection for discrete and numeric class machine learning,” Computer. Sci., Univ. Waikato, pp. 359–366, Dec. 1999.

[12] I. Guyon and A. Elisseeff, “An introduction to variables and features selection,” J. Mach. Learn. Res., vol. 3, pp. 1157–1182, Jan. 2003.

[13] Madhavi, K. Reddy, S. Viswanadha Raju, and J. Avanija. "Data Labeling and Concept Drift Detection using Rough Entropy For Clustering Categorical Attributes." HELIX 7, no. 5 (2017): 2077-2085.

[14] Madhavi, K. Reddy, A. Vinaya Babu, A. Anand Rao, and S. V. N. Raju. "Identification of optimal cluster centroid of multi-variable functions for clustering concept-drift categorical data." In Proceedings of the International Conference on Advances in Computing, Communications and Informatics, pp. 124-128. 2012.

[15] Reddy Madhavi, K., A. Vinaya Babu, and S. Viswanadha Raju. "Clustering of Concept-Drift Categorical Data Implementation in JAVA." In International Conference on Computing and Communication Systems, pp. 639-654. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[16] Z. Karimi, M. Mansour Riahi Kashani, and A. Harounabadi, “Feature ranking in intrusion detection dataset using combination of filtering methods,” Int. J. Comput. Appl., vol. 78, no. 4, pp. 21–27, Sep. 2013.

[17] V. Pihur, S. Datta, and S. Datta, “RankAggreg, an R package for weighted rank aggregation,” BMC Bioinf., vol. 10, no. 1, p. 62, 2009.

[18] S. Nembrini, I. R. König, and M. N. Wright, “The revival of the gini importance?” Bioinformatics, vol. 34, no. 21, pp. 3711–3718, Nov. 2018.