International Journal of Research in Agronomy 2024; 7(11): 108-115

International Journal of Research in Agronomy

Combining predictive models: Hybrid approaches in crop recommendation systems

Prity Kumari, GY Chandan, Divya Agarwal, Shivam Joshi and Prashant Vekariya

DOI: https://DOI.org/10.33545/2618060X.2024.v7.i11b.1951

E-ISSN: 2618-0618
P-ISSN: 2618-060X

www.agronomyjournals.com
2024; 7(11): 108-115
Received: 07-09-2024
Accepted: 15-10-2024

Prity Kumari
College of Horticulture, Anand Agricultural University, Anand, Gujarat, India

GY Chandan
B. A. College of Agriculture, Anand Agricultural University, Anand, Gujarat, India

Divya Agarwal
College of Agriculture, Junagadh Agricultural University, Junagadh, Gujarat, India

Shivam Joshi
B. A. College of Agriculture, Anand Agricultural University, Anand, Gujarat, India

Prashant Vekariya
B. A. College of Agriculture, Anand Agricultural University, Anand, Gujarat, India

Corresponding Author:
Prity Kumari
College of Horticulture, Anand Agricultural University, Anand, Gujarat, India

Abstract
This study aims to develop a robust crop recommendation system by using various machine learning (ML) algorithms and hybrid models. The dataset, sourced from Kaggle, having essential agricultural parameters, including nitrogen, phosphorus, potassium, soil pH, humidity, temperature and rainfall, across 22 crops. Using algorithms such as Decision Trees, Random Forest, Naïve Bayes, Support Vector Machines, Logistic Regression and XGBoost, this research evaluates both standalone and hybrid models to determine the optimal approach for accurate crop recommendations. Results indicate that individual models like Random Forest and Gaussian Naïve Bayes perform exceptionally well, achieving accuracies close to 99%, while hybrid models, particularly the SVM + GNB combination, also demonstrate equally strong predictive performance. These findings suggest that a carefully chosen individual model may be equally effective, though hybrid models offer further refinement in certain scenarios. The proposed system provides farmers with data-driven insights for crop selection based on soil and environmental conditions, supporting precision agriculture and sustainable farming practices.

Keywords: Crop recommendation system, machine learning algorithms, hybrid models, precision agriculture, random forest, gaussian naïve bayes

1. Introduction

The agricultural sector is increasingly turning to advanced technologies to enhance productivity and sustainability. Among these technologies, machine learning (ML) has emerged as a powerful tool for crop recommendation systems. These systems control vast amounts of data, including soil characteristics, climatic conditions and historical crop performance, to provide tailored recommendations for optimal crop selection. The application of machine learning (ML) techniques in crop recommendation systems has gained significant attention in recent years, primarily due to the increasing need for precision agriculture. These systems use various algorithms to analyse multiple factors, including soil nutrients, weather conditions and historical crop data, to provide recommendations for optimal crop selection.

Recent research underscores the effectiveness of ensemble learning techniques in these systems. Ensemble learning, which combines the outputs of multiple models to improve accuracy, has shown significant promise across various agricultural applications. For instance, Benos *et al.* highlighted that ensemble methods can compensate for the errors of individual models, leading to superior overall performance [1]. Ugale *et al.*, which highlights the effectiveness of neural networks in achieving a prediction accuracy of 91%, outperforming other algorithms in crop recommendation systems [2]. Rodríguez *et al.* emphasized the importance of ensemble size and composition in achieving reliable recommendations, suggesting that a minimum ensemble size can enhance the quality of predictions [3].

Several machine learning models, including Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT) and Naïve Bayes (NB), have proven effective for crop recommendation. A study by Reddy *et al.* (2019) implemented a crop recommendation system designed to improve yield in certain regions by using machine learning models, which demonstrated the value of region-specific crop guidance [4]. This aligns with the findings of Sundaresan *et al.*, who discussed the integration of IoT with ML to analyze soil parameters like nitrogen, phosphorus and potassium, thereby improving crop yield recommendations [5]. In [6] the researcher further elaborated on the diverse ML algorithms employed, such as Random Forest, Gradient Boosting and Support Vector Machines, emphasizing their role in enhancing agricultural output while minimizing soil degradation and chemical usage. Rajak *et al*. (2017) explored several machine learning models, including Random Forest and Naïve Bayes, to create a system aimed at optimizing crop yield based on specific soil and environmental parameters. Their research validated Random Forest as a particularly effective model for agricultural decision-making due to its reliability and robustness [7]. Similarly, Doshi *et al*. (2018) introduced AgroConsn intelligent crop recommendation tool that leverages ML to help farmers select suitable crops based on factors like soil type and seasonal conditions [8].

Further advancements have been made by researchers like, Kolhe *et al*. (2023) who developed a machine learning-based recommendation system to aid farmers in selecting suitable crops by analysing historical crop data and factoring in climate, irrigation and location-specific variables. This approach enhances crop yield predictions, mitigates losses from suboptimal crop choices, and provides farmers with an accessible interface to personalize recommendations. Their system illustrates how machine learning can optimize crop selection in the face of climate variability and resource constraints [9]. Similarly, Musanase *et al*. (2023) presented a dual model that combines crop recommendations with a rule-based fertilization system, showcasing the potential for improved agricultural productivity through tailored suggestions [10]. Furthent, in crop recommendation systems have been achieved through integrating real-time data and deploying sophisticated machine learning models. Pudumalar *et al*. (2017) developed a crop recommendation system for precision agriculture that uses data on soil types and environmental parameters, emphasizing the role of such systems in enhancing agricultural decision-making at a granular level [11]. Garanayak *et al*. introduced a system that employs regression-based machine learning techniques to analyze a comprehensive set of parameters, including soil nutrients, weather data, and atmospheric pressure, achieving high accuracy in crop recommendations [12].

This study aims to create a crop recommendation system using machine learning algorithms like Decision Trees, Random Forest, Naïve Bayes, Support Vector Machines, Logistic Regression, and XGBoost. A key aspect of this research is the exploration of hybrid models to improve predictive accuracy, offering recommendations based on soil nutrients, pH, humidity, temperature, and rainfall.

2. Materials and Methods

2.1 Material

The data for this study will be sourced from Kaggle, a popular platform for datasets across various domains, including agriculture. This specific dataset includes essential agricultural parameters like nitrogen (N), phosphorus (P), potassium (K), soil pH, humidity, temperature, and rainfall, which are crucial for analysing crop suitability. Comprising 2,200 records derived from historical agricultural data, this dataset covers 22 different crops, such as rice, maize, chickpea, mung bean, black gram and a selection of fruits including mango, grapes and pomegranate, along with fiber crops like cotton and jute. The dataset can be accessed on Kaggle at https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset.

2.2 Methods

In this study, various machine learning techniques for crop recommendation, including Decision Tree, Naive Bayes, Support Vector Machine (SVM), Logistic Regression and Random Forest. These models were chosen for their diverse approaches to data classification and predictive analysis, each offering unique strengths in handling agricultural data characterized by variability and non-linear patterns.

2.2.1 Decision Tree

A decision tree is a widely utilized machine learning model characterized by its hierarchical structure, which recursively splits data based on feature values to predict outcomes. This model operates through a series of decision nodes and leaf nodes, where each decision node represents a test on an attribute, and each leaf node signifies a class label or outcome. Decision Trees were implemented using the Gini impurity criterion, where each node split minimizes impurity, ultimately forming a tree that predicts variables. The simplicity and interpretability of decision trees make them particularly appealing for various applications, including medical predictions and classification tasks [13].

2.2.2 Naive Bayes: The Naive Bayes [14] classifier is a probabilistic machine learning model based on Bayes' theorem, which is particularly effective for classification tasks. It operates under the assumption of conditional independence, meaning that the presence of a particular feature in a class is independent of the presence of any other feature. This method is particularly useful for datasets with a mixture of continuous and categorical variables. In our study, Naive Bayes provided baseline predictions for yield based on probabilistic reasoning, which served as a comparative method against more complex models [15].

2.2.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm primarily used for classification and regression tasks. The core principle of SVM is to find the optimal hyperplane that separates data points of different classes in a high-dimensional space. This hyperplane is chosen to maximize the margin between the closest points of the classes, known as support vectors, thus enhancing the model's generalization capabilities [16]. SVM is particularly effective in scenarios where the number of dimensions exceeds the number of samples, making it suitable for applications in fields such as bioinformatics, text classification, and image recognition. The algorithm can also handle non-linear relationships by employing kernel functions, which transform the input data into a higher-dimensional space where a linear separator can be found [17].

2.2.4 Logistic Regression

Logistic Regression is a widely used statistical method for binary classification problems, where the outcome variable is categorical and typically takes on two possible outcomes (e.g., success/failure, yes/no). The model estimates the probability that a given input point belongs to a particular category by applying the logistic function, which maps any real-valued number into the (0, 1) interval. This model uses the logistic function to map predictions to probabilities, making it interpretable and effective for predicting binary outcomes, such as crop viability. In agricultural studies, Logistic Regression is valued for its straightforward approach to binary classification, which is useful for identifying potential crops based on specific conditions [18].

2.2.5 Random Forest
Random Forest is an ensemble learning method that combines multiple decision trees to improve classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the mode of their predictions (For classification) or the mean prediction (For regression). This technique is particularly effective because it mitigates the risk of over-fitting, which is a common problem with individual decision trees. By averaging the results of many trees, Random Forest achieves greater accuracy and robustness [19].

3. Results & Discussion
This section is divided into two parts: Data Pre-processing and Model Output, each detailing findings in the study.

3.1 Data pre-processing
The box plot in Fig. 1 reveals that the input parameters exhibit varying degrees of dispersion and outliers, which is crucial for interpreting crop suitability. Parameters such as nitrogen and rainfall show high variability and multiple outliers, indicating that they may significantly influence the recommendation system due to their variability across different geographical areas or climatic conditions. Specifically, the presence of high outliers in rainfall suggests regions with potential for excessive water availability, which may be suitable for water-intensive crops or require drainage measures. In contrast, stable parameters like temperature and pH imply a consistent baseline that may favour crops with less stringent environmental requirements, making these parameters easier to model in a predictive system for crop selection.

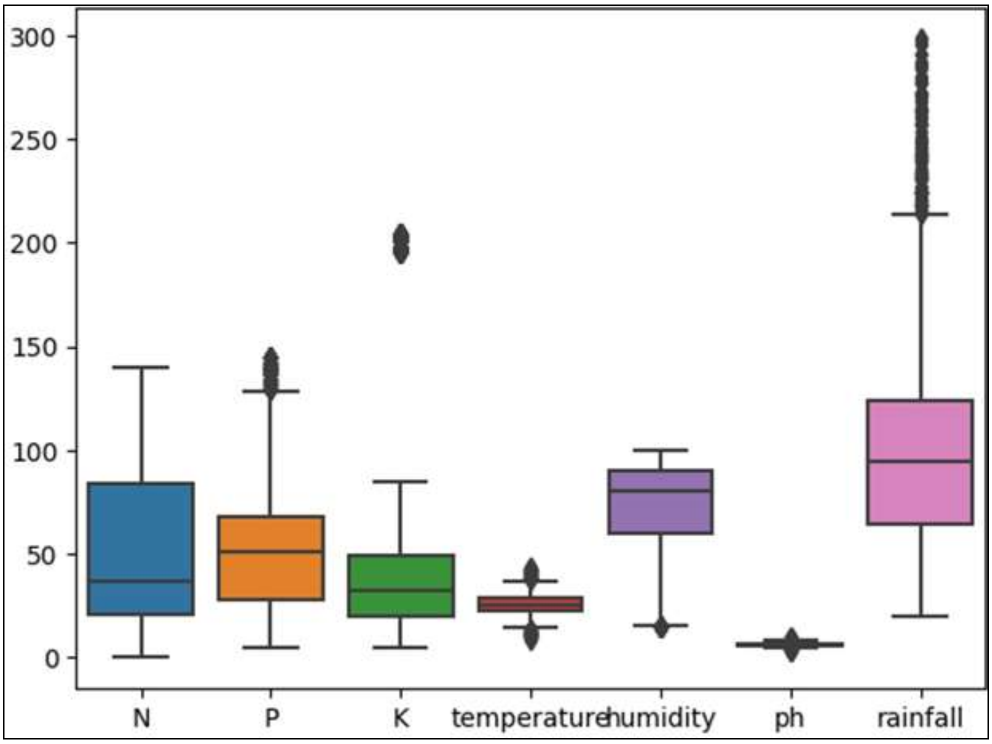

Fig 1: Box plot of input parameters

The heat map in Fig. 2 illustrates the correlation coefficients among various agricultural parameters, including nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. The most notable relationship is between phosphorus (P) and potassium (K), with a correlation coefficient of 0.74, suggesting a significant positive correlation. This implies that higher levels of phosphorus often coincide with higher potassium levels, which may be due to soil management practices or nutrient content in certain regions.

Nitrogen (N) has weak negative correlations with phosphorus (-0.23) and potassium (-0.14), indicating that in some cases, as nitrogen increases, phosphorus and potassium may slightly decrease, although this relationship is not strong. Other parameters, such as temperature, humidity and pH, show very low or near-zero correlations with each other and with nutrient levels, indicating minimal interdependence. For example, temperature and pH have almost no correlation (-0.02), which implies that soil pH is relatively stable across different temperature conditions.

Rainfall exhibits low correlation values with all other parameters, such as a weak positive correlation with humidity (0.09) and a slight negative correlation with pH (-0.11), suggesting that rainfall does not directly influence other parameters in the dataset. This independence of rainfall may reflect its variability in relation to regional climatic conditions rather than soil composition.

The parallel coordinates plot in Fig. 3 provides a visual representation of how different crops align with key agricultural parameters-nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. Each line represents a specific crop type, color-coded as shown in the legend. The normalized values for each parameter are scaled between 0 and 1, allowing for easy comparison across crops.

In the plot, distinct patterns for each crop based on parameter values were observed:

Fig 2: Heat map of correlation among parameters

1. **Nitrogen (N), Phosphorus (P) and Potassium (K):** These three nutrient parameters show varying levels across different crops. For example, some crops like maize (in cyan) require relatively higher levels of nitrogen, which is indicated by their lines trending towards the upper part of the scale on the nitrogen axis. Phosphorus and potassium values also vary, with crops like chickpea (In pink) showing moderate levels.
2. **Temperature:** Most crop lines are clustered around the middle to high range, reflecting an optimal temperature requirement that is similar across many crop types. This suggests that temperature is an important, yet relatively consistent, factor for crop growth.
3. **Humidity:** There is more variability in humidity requirements, with some crops like rice (In dark blue) showing a need for higher humidity, while others, such as jute (In brown), require lower levels. This highlights the adaptability of different crops to various moisture conditions.
4. **pH:** The pH axis shows that most crops have a neutral to slightly acidic requirement, with lines converging in the middle range. This indicates a common preference for moderate pH levels among many crops.
5. **Rainfall:** Similar to humidity, rainfall requirements vary widely. Crops like rice and watermelon have lines near the upper part of the scale, indicating high rainfall requirements, whereas others like cotton and chickpea require comparatively less rainfall, as their lines fall in the lower part of the rainfall axis.

The visualizations by box plot identifies outliers and variability for normalization, the heat map highlights correlations to optimize feature selection and the parallel coordinates plot reveals unique crop profiles, confirming the dataset's suitability for classification. These insights enhance model precision and efficiency by identifying patterns in crop requirements, supporting the crop recommendation system by showing the specific environmental and nutrient conditions suitable for each crop.

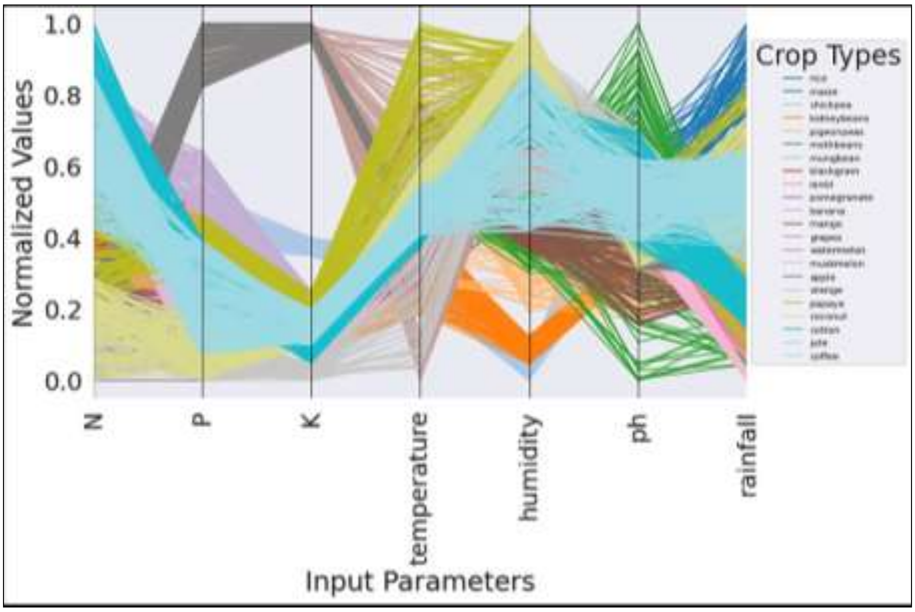

Fig 3: Parallel Coordinates Plot of Input Parameters for Crop

3.2 Model output

3.2.1 Decision Tree

Table 1: Accuracy matrix of Decision Trees

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.90	0.84	0.88	0.85
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.86	0.86	0.86

The performance metrics presented in the Table 1 are based on test data, obtained by splitting the dataset into an 80:20 ratio for training and testing. The test dataset consisted of 440 samples, representing 20% of the total data. The Decision Tree model achieved an overall accuracy of 90%, with high precision and recall scores for most crops, including apple, banana, and mango, all achieving perfect scores across metrics. The macro-average metrics, which provide an un-weighted average across all crop classes, reveal a precision of 0.84, recall of 0.88, and F1-score of 0.85. These values indicate the model's generally good performance across diverse crop categories, although not perfectly balanced for each. The weighted average metrics, precision, recall and F1-scor, each score at 0.86, reflecting the model's balanced accuracy across classes while accounting for class representation.

3.2.2 Gaussian Naive Bayes (GNB)

Table 2: Accuracy matrix of Gaussian Naive Bayes

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.99	0.99	0.99	0.99
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.99	0.99	0.99

Table 2 provides an accuracy matrix summarizing the performance metrics of the Gaussian Naive Bayes model on test data. The test dataset consisted of 440 samples, representing 20% of the total data. The model achieved a high overall accuracy of 99%, reflecting its effectiveness in correctly classifying the majority of crop types. The macro-average scores; precision, recall, and F1-score-all reached 0.99, indicating consistent performance across various crop classes without favouring any particular category. Additionally, the weighted average scores for precision, recall, and F1-score also measured at 0.99, which suggests the model maintained high accuracy while accounting for the varying representation of each crop class in the dataset. These metrics collectively highlight the robustness and reliability of the Gaussian Naive Bayes model for crop classification tasks.

3.2.3 Support Vector Machine (SVM)

Table 3: Accuracy matrix of Support Vector Machine

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.11	0.66	0.13	0.14
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.66	0.13	0.14

Table 3 provides an accuracy matrix summarizing the performance metrics of the Support Vector Machine model on test data. The test dataset consisted of 440 samples, representing 20% of the total data. Support Vector Machine (SVM) model shows an overall accuracy of 11%, indicating significant limitations in correctly classifying most samples. Similarly, the macro-average metrics; precision (0.66), recall (0.13), and F1-score (0.14), suggest that performance varies significantly across different crop classes, with low recall and F1-scores indicating difficulties in capturing certain classes consistently. The weighted average metrics, which consider class distribution, also show low recall (0.13) and F1-score (0.14), alongside a precision of 0.66. These results highlight substantial challenges for the SVM model in accurately identifying crop categories, especially with class imbalances and implying areas for potential model improvement.

3.2.4 Logistic Regression

Table 4: Accuracy matrix of Logistic Regression

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.95	0.95	0.95	0.95
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.95	0.95	0.95

Table 4 provides an accuracy matrix summarizing the performance metrics of the Logistic Regression model on test data. The test dataset consisted of 440 samples, representing 20% of the total data. Logistic Regression model achieved an overall accuracy of 95%. This high accuracy indicates that the model performs well in correctly classifying the majority of crop types in the test data. The macro-average scores; precision, recall, and F1-score, all reached 0.95, suggesting that the model maintains consistent and balanced performance across different crop classes, without particular bias toward any specific category. The weighted average metrics for precision, recall, and F1-score are also each 0.95, reflecting the model's robustness and reliability while accounting for the distribution of each class in the dataset. These metrics underscore that the Logistic Regression model is much better than SVM and also effective for crop classification, with high accuracy.

3.2.5 Random Forest (RF)

Table 5: Accuracy matrix of Random Forest

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.99	0.99	0.99	0.99
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.99	0.99	0.99

Table 5 provides an accuracy matrix summarizing the performance metrics of the Random Forest model on test data. The test dataset consisted of 440 samples, representing 20% of the total data. The accuracy matrix for the Random Forest model, which achieved an impressive overall accuracy of 99% is mentioned in Table 1. This high accuracy indicates that the model is highly effective at correctly classifying most crop types. The macro-average metrics; precision, recall, and F1-score, all scored 0.99, demonstrating the model's balanced and strong performance across various crop classes, with no significant bias toward specific categories. Similarly, the weighted average metrics for precision, recall and F1-score also reached 0.99, reflecting the model's robustness and ability to maintain high accuracy across the dataset, accounting for the different class sizes.

3.2.6 Hybrid Model 1 (SVM + Decision Tree)

Table 6 Accuracy matrix of Hybrid Model 1

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.98	0.98	0.98	0.98
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.98	0.98	0.98

Table 6 shows the accuracy matrix for Hybrid Model 1, a combination of Support Vector Machine (SVM) and Decision Tree algorithms. This hybrid approach achieved a high overall accuracy of 98%, indicating that it is effective at accurately classifying crop types. The macro-average scores, precision, recall, and F1-score, all reached 0.98, demonstrating that Hybrid Model 1 maintains consistently high performance across different crop classes, providing balanced predictions without favoring any specific class. Additionally, the weighted average metrics for precision, recall, and F1-score are each at 0.98, reflecting the model's reliability in maintaining high classification accuracy even when accounting for class imbalances. Here again, test dataset consisted of 440 samples, representing 20% of the total data split.

3.2.7 Hybrid Model 2 (SVM + Naive Bayes)

Table 7: Accuracy matrix of Hybrid Model 2

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.99	0.99	0.99	0.99
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.99	0.99	0.99

Table 7 provides the accuracy matrix for Hybrid Model 2, which combines Support Vector Machine (SVM) and Naive Bayes algorithms. This hybrid model achieved an overall accuracy of 99%, indicating its high effectiveness in classifying crop types accurately. The macro-average metrics; precision, recall, and F1-score, all scored 0.99, highlighting the model's ability to maintain excellent and consistent performance across different crop categories without showing bias towards any particular class. The weighted average metrics for precision, recall, and F1-score also reached 0.99, demonstrating the model's robustness and stability, even when accounting for class distribution within the dataset. These metrics suggest that Hybrid Model 2 successfully integrates the strengths of SVM and Naive Bayes, resulting in a reliable and highly accurate predictive model for crop classification tasks.

3.2.8 Hybrid Model 3 (RF + SVM)

Table 8: Accuracy matrix of Hybrid Model 3

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.90	0.87	0.90	0.88
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.87	0.90	0.88

Table 8 presents the accuracy matrix for Hybrid Model 3, a combination of Random Forest (RF) and Support Vector Machine (SVM). This model achieved an overall accuracy of 90%, indicating reasonable effectiveness in crop classification but lower performance compared to Hybrid Model 2. The macro-average metrics; precision (0.87), recall (0.90), and F1-score (0.88), show that while Hybrid Model 3 provides balanced predictions, it does not match the consistency or accuracy of Hybrid Model 2, which scored 0.99 across all metrics.

The weighted average scores for precision (0.87), recall (0.90) and F1-score (0.88) similarly reflect this reduction in performance, especially in precision and F1-score, when accounting for the class distribution. These results suggest that, despite combining three algorithms, Hybrid Model 3 does not outperform the simpler combination in Hybrid Model 2.

3.2.9 Hybrid Model 4 (RF + SVM + Naive Bayes)

Table 9: Accuracy matrix of Hybrid Model 4

Metric	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
Score	0.15	0.14	0.18	0.14
Metric		Weighted Avg Precision	Weighted Avg Recall	Weighted Avg F1-Score
Score		0.11	0.15	0.11

Table 9 displays the accuracy matrix for Hybrid Model 4, which combines Random Forest (RF), Support Vector Machine (SVM), and Naive Bayes algorithms. This model achieved an overall accuracy of only 15%, making it the least effective among all hybrid models tested. The macro-average metrics, precision (0.14), recall (0.18), and F1-score (0.14), indicate very low performance, with the model struggling to provide consistent and accurate predictions across crop categories.

The weighted average metrics are similarly low, with precision at 0.11, recall at 0.15, and F1-score at 0.11, reflecting poor performance even when accounting for the distribution of different crop classes. These results demonstrate that Hybrid Model 4 underperformed significantly compared to other models, including simpler hybrid models like Hybrid Model 2, suggesting that the combination of RF, SVM, and Naive Bayes in this configuration did not yield an effective or reliable predictive model for crop classification

In Fig. 4, the accuracy comparison of various machine learning algorithms and their hybrid combinations for crop recommendation is illustrated. The analysis reveals that among individual models, Random Forest (RF) achieves the highest accuracy, followed by Naïve Bayes (GNB) and Logistic Regression, both performing well. Decision Tree demonstrates a moderate level of accuracy, while Support Vector Machine (SVM) exhibits the lowest accuracy among all individual models.

For hybrid models, the SVM + GNB (Gaussian Naïve Bayes) combination stands out by achieving the highest accuracy across all models, outperforming other hybrid configurations. Other combinations, such as RF + SVM and SVM + Decision Tree, also show enhanced performance compared to individual models. However, the RF + SVM + GNB combination ranks the lowest among the hybrid approaches, indicating that not all model integrations contribute equally to accuracy improvement.

The results indicate that Random Forest (RF) and Gaussian Naïve Bayes (GNB), as individual models, achieved very high accuracy, close to 99%, which suggests that these algorithms are inherently well-suited for the dataset and effectively capture the underlying patterns necessary for crop recommendation. The SVM + GNB hybrid also performed exceptionally well with 99% accuracy, similar to the individual RF and GNB models.

The high performance of both individual and hybrid models suggests that while hybrid models can be beneficial, a well-chosen single model like RF or GNB might be equally effective in this context, offering simpler yet powerful solutions for the crop recommendation system.

4. Conclusion

This study developed a crop recommendation system utilizing various machine learning algorithms to classify and suggest optimal crops based on parameters such as soil nutrients, pH, humidity, temperature, and rainfall. The study concludes that both individual and hybrid models offer high accuracy for crop recommendation, with Random Forest (RF) and Gaussian Naïve Bayes (GNB) standing out as particularly effective individual models. The SVM + GNB hybrid model achieved the highest accuracy among hybrid combinations, underscoring the potential benefits of combining machine learning algorithms to enhance model robustness. However, the performance of RF and GNB as individual models suggests that in this context, simpler models can be equally effective. The findings highlight the potential of machine learning in precision agriculture, providing a tool that can help farmers make informed decisions based on critical parameters like soil nutrients, pH, and climate conditions. Future work may explore the integration of real-time data and scalability for broader regional applications.

5. References

1. Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D. Machine learning in agriculture: A comprehensive updated review. Sensors. 2021;21:3758. DOI:10.3390/s21113758.
2. Ugale S, *et al*. Review of machine learning techniques for crop recommendation system. Int J Res Appl Sci Eng Technol. 2022;10:40559. DOI:10.22214/ijraset.2022.40559.
3. Rodríguez DA, *et al*. Implications of crop model ensemble size and composition for estimates of adaptation effects and agreement of recommendations. Agric For Meteorol. 2019;265:98-109. DOI:10.1016/j.agrformet.2018.09.018.
4. Reddy V, Varshini V, Gireesh N, Naresh M. A comprehensive study on crop recommendation system for precision agriculture using machine learning algorithms. EAE. 2023;2:30-36. DOI:10.46632/eae/2/1/5.
5. Sundaresan S, *et al*. Machine learning and IoT-based smart farming for enhancing crop yield. J Phys Conf Ser. 2023;2466:012028. DOI:10.1088/1742-6596/2466/1/012028.
6. Reddy D, Anantha, Dadore B, Watekar A. Crop recommendation system to maximize crop yield in Ramtek region using machine learning. Int. J Sci Res Sci Technol. 2019;6:485-489. DOI:10.32628/IJSRST196172.
7. Rajak RK, Pawar A, Pendke M, Shinde P, Rathod S, Devare A. Crop recommendation system to maximize crop yield using machine learning technique. Int. Res J Eng Technol. 2017;4:950-953.
8. Doshi Z, Nadkarni S, Agrawal R, Shah N. AgroConsultant: Intelligent crop recommendation system using machine learning algorithms. In: 2018 Fourth International Conference on Computing; c2018. p. 1-5.
9. Kolhe V, Sable P, Dongare A, Mohite A, Kulkarni M. Machine learning techniques for crop yield prediction and recommendation. Int Res J Mod Eng Technol Sci. 2023;5:40641. DOI:10.56726/IRJMETS40641.
10. Musanase C, Vodacek A, Hanyurwimfura D, Uwitonze A, Kabandana I. Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. Agriculture. 2023;13:2141. DOI:10.3390/agriculture13112141.
11. Pudumalar S, Ramanujam E, Rajashree H, Kavya C, Kiruthika T, Nisha J. Crop recommendation system for precision agriculture. In: 2016 Eighth International Conference on Advanced Computing (ICoAC); c2016. p. 32-36. IEEE.
12. Garanayak M, Sahu G, Mohanty SN, Jagadev AK. Agricultural recommendation system for crops using different machine learning regression methods. Int J Agric

Environ Inf Syst. 2021;12:1-20.
13. Wakiya Y, *et al*. Prediction of massive bleeding in pancreatic surgery based on preoperative patient characteristics using a decision tree. PLoS One. 2021, 16. DOI:10.1371/journal.pone.0259682.
14. Wibowo A, *et al*. Hybrid top-K feature selection to improve high-dimensional data classification using naïve Bayes algorithm. Sci. J Inform. 2023;10:42818. DOI:10.15294/sji.v10i2.42818.
15. Zhang Y, Gao Y. Performance analysis and improvement of naïve Bayes in text classification application. In: Proceedings of the 2013 IEEE International Conference on Data Mining; c2013. DOI:10.1109/anthology. 2013.6784818.
16. Shalev-Shwartz S, Ben-David S. Pegasos: Primal estimated sub-gradient solver for SVM. Math Program. 2010;127:3-30. DOI:10.1007/s10107-010-0420-4.
17. Schölkopf B, Smola AJ. Learning with Kernels. MIT Press; 2001. DOI:10.7551/mitpress/4175.001.0001.
18. Peng CYJ, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. J Educ. Res. 2002;96:3-14. DOI:10.1080/00220670209598786.