SMART CROP ADVISOR

Submitted by

221FA04180	**Sri Rama Sai Venkatesh**
221FA04206	**Lavanya**
221FA04250	**Siva Dinesh**
221FA04607	**Krishna Sri**

Under the guidance of

Mrs.B.Suvarna

Assistant professor, Department of CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH

(Deemed to be UNIVERSITY)

Vadlamudi, Guntur.

ANDHRA PRADESH, INDIA, PIN-522213.

VIGNAN'S
FOUNDATION FOR SCIENCE, TECHNOLOGY & RESEARCH
(Deemed to be University) - Estd. u/s 3 of UGC Act 1956

CERTIFICATE

This is to certify that the Field Project entitled **"Smart Crop Advisor"** that is being submitted by 221FA04180 (Venkatesh), 221FA04206 (Lavanya), 221FA04250 (Dinesh), 201FA04607 (Krishna Sri) for partial fulfilment of Field Project is a bonafide work carried out under the supervision of **Mrs.B.Suvarna, Department of CSE**.

Mrs.B.Suvarna Guide name& Signature Assistant/Associate/Professor, CSE	Dr. S. V. Phani Kumar HOD,CSE	Dr.K.V. Krishna Kishore Dean, SoCI

DECLARATION

We hereby declare that the Field Project entitled **"Smart Crop Advisor"** is being submitted by 221FA04180 (Venkatesh), 221FA04206 (Lavanya), 221FA04250 (Dinesh), 201FA04607 (Krishna Sri), in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of **Mrs.B.Suvarna, M.Tech., Assistant Professor, Department of CSE**.

By

221FA04180 (Venkatesh)

221FA04206 (Lavanya)

221FA04250 (Dinesh)

221FA04607 (Krishna Sri)

Date: 07-11-2024

ABSTRACT

Agricultural productivity is significantly influenced by the choice of crops suitable for the local soil and environmental conditions. The ”Smart Crop Advisor” uses machine learning to assist farmers in selecting the best crop for cultivation based on real-time environmental and soil parameters. This research uses a dataset containing various attributes related to the soil and weather conditions of the farmer’s field. The dataset is pre-processed and scaled using Standard Scalar normalization, and multiple classification algorithms, are trained and evaluated. The Random Forest emerged as the most accurate model for predicting crop suitability with 99.7% followed by Catboost and Naive Bayes with 99.2% and 99% respectively, all other models also performed well with predicting accuracy around 90%-98%. The proposed solution accepts user input for environ- mental parameters, scales the data appropriately, and predicts the most suitable crop using the trained model. This system provides a practical and efficient tool for farmers to make informed decisions about crop cultivation, thereby enhancing yield, optimizing resource usage, and contributing to sustainable agriculture practices. The results demonstrate the effectiveness of machine learning in agricultural decision-making, underscoring its potential to revolutionize crop management and planning.

Index Terms—Smart Crop Advisor, Machine Learning, Data Scaling, Random Forest Classifier, Predictive Modeling.

TABLE OF CONTENTS

1. Introduction
1.1 What is a Smart Crop Advisor System and What Causes the Need for It?
1.2 The Consequences of Ineffective Crop Recommendations
1.3 Economic and Environmental Impacts of Inefficient Crop Recommendation Systems
1.4 Current Methodologies in Crop Prediction
1.5 Applications of Machine Learning in Addressing Agricultural Challenges

2. Literature Survey
2.1 Literature Review on Crop Recommendation Models
2.2 Motivation for Developing a Smart Crop Advisor System

3. Proposed System
3.1 Overview of the Smart Crop Advisor System
3.2 Input Dataset for Crop Recommendation
3.2.1 Detailed Features of the Crop Dataset
3.3 Data Pre-processing for Crop Prediction
3.3.1 Normalization of Input Data
3.3.2 Data Augmentation and Enhancement Techniques
3.4 Model Building for Crop Prediction
3.4.1 Machine Learning Algorithms Used (e.g., Random Forest, CatBoost)
3.5 System Methodology and Workflow
3.6 Model Evaluation Metrics for Crop Prediction Accuracy
3.7 Constraints and Challenges in Building the System
3.8 Cost and Sustainability Impact of Smart Crop Advisory Systems

4. Implementation
4.1 Environment Setup for the Smart Crop Advisor
4.2 Sample Code for Preprocessing and Model Prediction

5. Experimentation and Result Analysis

6. Conclusion

7. References

LIST OF FIGURES

Figure 1. Architecture of the Smart Crop Advisor System
Figure 2. Key Features of the Crop Dataset
Figure 3. Pre-processing Steps Applied to the Dataset
Figure 4. Loaded Machine Learning Models in the System
Figure 5. Flow of Prediction Process
Figure 6. Comparison of Model Performance
Figure 7. Example of a Prediction Result

Introduction

Agriculture is a crucial part of economies around the world, playing a key role in ensuring food security, supporting rural communities, and promoting sustainable development. However, farming today faces many challenges, including unpredictable weather, soil degradation, water shortages, and pest infestations. These problems are made worse by climate change, making traditional farming methods less reliable and effective. Farmers need to make important decisions about which crops to grow to get the best yield and profit while min- imizing environmental impact. Traditionally, these decisions have been based on experience, intuition, or general guidelines that may not consider the specific conditions of a particular location, often leading to less-than-ideal results.

Advanced technologies and data science offer new solutions to these challenges, and machine learning (ML) has become a powerful tool for improving decision-making in agriculture. A ”Smart Crop Advisor” system uses machine learning algo- rithms to analyze various types of data, such as soil properties, weather forecasts, past crop performance, and market trends, to suggest the best crops for a specific location and time. These systems can adjust to local conditions, providing recommen- dations that are more accurate than traditional methods.

Fig. 1. Smart Crop Recommendation System

Machine learning models have shown great potential in this area because they can learn patterns from large amounts of data. Techniques such as Gradient Boosting Machines (GBM), CatBoost, Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA), and Ensemble Learning have been successfully used in agriculture for tasks like predicting yields, detecting crop diseases, and modeling climate.

For crop recommendations, these models can analyze factors like soil pH, nutrient levels (such as nitrogen, phosphorus, and potas- sium), temperature, rainfall, and humidity to predict which crops will perform best in certain conditions. Research shown that machine learning models can achieve high levels of accuracy in recommending crops. For example, methods that combine multiple algorithms, known as ensemble learning, have been found to reach accuracies over 95%. These advance- ments highlight the potential of machine learning to improve crop selection by providing reliable, data-driven details related to local environments. Smart crop recommendation systems can help farmers maximize yields, reduce input costs, and promote sustainable farming practices by making better use of resources like water, fertilizer, and labor.

This research aims to develop a strong crop recommendation model using the latest machine-learning techniques. The study will use a detailed dataset that includes different environmental and farming variables from various regions to build a model that is both flexible and applicable in different contexts. The research will focus on selecting the most important variables, fine-tuning the model to improve performance, and validating it with independent datasets to ensure it works well in real- world situations.

By examining different machine learning methods and how they can be applied to crop recommendations, this paper contributes to the growing field of precision agriculture. The ultimate goal is to develop a smart crop recommendation system that helps farmers make informed decisions, reduce un- certainty in crop planning, and support sustainable agricultural practices that can adapt to changing climate and economic conditions.

Literature Survey

The research [1] explores the use of machine learning to improve agricultural decision-making in India, focusing on crop selection and yield prediction to help farmers make informed choices. The study developed two modules: one for predicting suitable crops and another for estimating yields. For crop prediction, various algorithms were tested. The Naive Bayes classifier achieved the highest accuracy with a Cohen's Kappa score of 95.13%. For yield prediction, models including Multiple Linear Regression, Random Forest Regression, Sup- port Vector Regression, and KNN Regression were evaluated. The Random Forest Regressor performed best, with R-squared values of 82.85% and 91.79% across two datasets, indicating strong predictive capability. The researchers also developed a web application using Flask, enabling farmers to input specific conditions to receive tailored crop recommendations and yield predictions.

Paper [2] investigates the use of ML for crop prediction, which is vital for agricultural planning. It compares three algorithms—K-Nearest Neighbor (KNN), Decision Tree, and Random Forest Classifier—using the crop recommendation dataset. The dataset was divided into 80% training and 20% testing data, and the models were assessed based on various metrics. Results showed that the Random Forest Classifier performed best, achieving a 99.32% accuracy, followed by the Decision Tree Classifier with 98.86% (Gini criterion) and 97.95% (Entropy criterion). KNN had the lowest accuracy at 97.04%. The study highlights the potential of machine learning in aiding farmers with crop selection based on environmental factors. Future research could involve using more advanced techniques like Artificial Neural Networks (ANN) or Convo- lutional Neural Networks (CNN) to expand crop classification.

The study [3] presents a Machine Learning-based Crop Recommendation System that addresses challenges like un- predictable climate, diverse soil types, and limited modern farming access. The researchers evaluated nine machine learn- ing algorithms, including Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Decision Tree, and Random Forest, using historical data on climate, soil properties, and crop yields. Data preprocessing and fine-tuning techniques like cross-validation were applied to optimize the models. The Random Forest algorithm achieved the highest accuracy at 99.31%. The study highlights the system'spotential to improve crop productivity, manage risks from climate and market changes, and promote sustainable practices. While recognizing limitations in capturing land quality and climate variations, the authors suggest future research to address these gaps, contributing to global food security and precision agriculture.

The paper [4] introduces an AI-enabled crop recommenda- tion system that uses ML to suggest optimal crops based on soil composition and weather patterns, aiming to help farmers maximize yields and profitability. Using a dataset from Kaggle containing soil parameters and climate factors, the researchers tested seven machine learning models, including Decision Tree, Naive Bayes, Random Forest, k-nearest Neighbors (k- NN), Support Vector Machine (SVM), XGBoost, and Logistic Regression. Naive Bayes and XGBoost achieved the highest accuracy of 99.55%, while Random Forest and SVM followed with 99.31%. The system's high accuracy suggests it could ef- fectively support precision agriculture by optimizing resource use, reducing crop failures, and boosting productivity. The study also discusses potential enhancements, like integrating IoT and remote sensing technologies, and suggests future research involving diverse datasets and deep learning methods.

Research [5] explores the use of ML for crop recommenda- tion in smart farming the authors tested seven algorithms, in- cluding Logistic Regression, Decision Trees, Random Forests, K-Nearest Neighbors, Naive Bayes, Support Vector Machines, and Neural Networks, using a Kaggle dataset with various soil and weather features. The Random Forest algorithm achieved the highest accuracy at 99.5%, while the Neural Network reached 97.73% with 1000 epochs. The study highlights the potential of these models to improve crop yields, sustainabil- ity, and profitability, and discusses challenges like climate change, water scarcity, and data quality. Future directions include assessing the economic impact through farmer surveys, developing mobile apps, and incorporating real-time sensor data.

The research paper [6] introduces a new approach to crop recommendation using ML through a hybrid model called the Wrapper-PART-Grid method. This model combines a wrapper feature selection method to identify key soil features, grid search (GS) for hyperparameter optimization, and the PART (partial C4.5 decision tree) classifier for crop prediction. Using a dataset of 2,200 instances with various soil and weather pa- rameters, the model achieved the highest accuracy of 99.31%, outperforming other ML models such as multilayer percep- tron, instance-based learning, and decision trees. The study demonstrates that this method improves prediction accuracy and reliability.

The paper [7] presents a system that combines machine learning (ML) and deep learning (DL) models to improve crop recommendations for farmers. Utilizing a crop recommenda- tion dataset available in Kaggle, the study evaluated models including Decision Trees, Random Forests, XGBoost, SVM, KNN, Naive Bayes, ANN, DNN, and Temporal Convolutional Networks (TCN). The TCN model achieved the highest accu- racy of 99.9% by effectively capturing temporal dependencies, while the Random Forest model also performed well with

99.2% accuracy. The study emphasized parameter optimization for XGBoost and TCN, highlighting their precision. A web- based interface was developed to allow farmers to input soil and environmental data and receive immediate crop recom- mendations.

The study [8] develops a crop recommendation system to help farmers choose suitable crops based on various en- vironmental factors. Using ML algorithms like K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logis- tic Regression (LR), Decision Tree (DT), and Na¨ıve Bayes (NB)—on a dataset of 2,200 records covering these factors and 22 crop types, the study found Na¨ıve Bayes to be the most accurate with 99% precision. Logistic Regression followed with 98%, KNN with 97%, Decision Tree with 94%, and SVM with 82.27%. A Google Assistant-based platform was also created to offer real-time crop recommendations by inputting environmental data, aiming to boost agricultural productivity and assist especially young farmers in making informed crop choices.

The research [9] examines the use of ML algo- rithms—PART, JRip (RIPPER), and Decision Table—for crop recommendation in precision agriculture. Using a Kaggle dataset with 2,200 instances featuring attributes like soil composition, weather conditions, and crop details, the data was divided into 60% for training and 40% for testing. The PART algorithm performed best, achieving a prediction accuracy of 98.33% and a model-building time of 0.05 seconds, along with high precision, recall, and F-measure scores. JRip and Decision Table also showed good performance with accuracies of 95.90% and 87.42%, respectively. The study highlights the potential of these models in offering precise crop recom- mendations, helping farmers make better decisions, optimize resources, and increase yields, particularly in regions impacted by climate change and resource limitations.

The study [10] introduces a web-based recommendation system for farmers using ML to improve agricultural decisions. The researchers employed random forest and decision tree algorithms to develop highly accurate predictive models for temperature and rainfall, achieving near-perfect R-squared val- ues (0.9999) and low error rates (MAE of 0.003223 for temper- ature and 0.002021 for rainfall). The system, accessible via a user-friendly web platform, provides real-time insights to help farmers optimize crop management and resource allocation. Additionally, a Power BI dashboard visualizes agricultural trends, aiding in data interpretation. The paper suggests future enhancements, such as incorporating pest and soil moisture predictions, to further support sustainable farming practices.

The research paper [11] presents an ML-based system for crop and fertilizer recommendations, it uses algorithms like linear regression, random forests, k-nearest neighbors, and SVM to provide personalized suggestions based on soil composition, pH, moisture, temperature, and precipitation. The study utilized a dataset from Kaggle and achieved high accuracy, with SVM showing particularly strong performance. The system is accessible via a user-friendly website where farmers can input soil data to receive recommendations, potentially increasing crop yields and promoting sustainable farming. The authors propose future enhancements, including deep learning for plant disease prediction and IoT integration for real-time analysis.

This research paper [12] provides a review of machine learn- ing techniques used in crop recommendation systems based on soil characteristics and crop yield prediction. It focuses on supervised algorithms like K-Nearest Neighbors (KNN), Random Forest, Decision Trees, Naive Bayes, Support Vector Machines (SVM), and Artificial Neural Networks (ANN), analyzing their effectiveness and limitations. The study dis- cusses various research efforts, noting that some models, like SVM, achieved up to 96% accuracy, while ensemble models combining multiple algorithms reached up to 99% accuracy. The paper highlights the importance of considering factors like soil pH, NPK ratios, temperature, and rainfall, and suggests that future research could benefit from incorporating more data sources, using hybrid models, and applying deep learning techniques to enhance prediction accuracy and provide better crop recommendations to farmers

Motivations

A common challenge faced by many Indian farmers is the difficulty in selecting the right crop based on the specific requirements of their soil. Precision agriculture has emerged as a solution to this problem. This method involves gathering a comprehensive crop database from farms, which is then analyzed by agricultural experts. The data is fed into a recommendation system that uses machine learning models to suggest the most suitable crop for a particular location, based on site-specific parameters. This approach ensures high accuracy and improved agricultural performance.

PROPOSED SYSTEM

A. *Data Collection*

The foundational step in developing the ”Smart Crop Ad- visor” is the collection of relevant data that encompasses both environmental and soil parameters crucial for crop growth. The ”Crop recommendation” dataset available in Kaggle is used which includes various features such as Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and rainfall, which are essential for understanding the con- ditions that influence crop suitability. Additionally, the dataset contains crop labels that indicate which crops are most suitable under specific conditions. This dataset provides a compre- hensive basis for training and evaluating machine learning models, enabling the system to learn patterns and relationships between environmental factors and crop types. The quality and representativeness of the data are vital as they directly impact the accuracy and reliability of the predictions made by the ”Smart Crop Advisor.”

B. *Data Preprocessing*

Data preprocessing is a critical step in preparing the raw dataset for machine learning analysis. Initially, categorical crop labels in the dataset are converted into numerical values using a mapping dictionary. This transformation is necessary because ML algorithms generally require numerical inputs to perform computations. After encoding the crop labels, the dataset is divided into features and target variables. Features include the environmental and soil attributes, while the target variable is the crop label. Following this, the data is split into training and testing sets, with 80% of the data allocated for training the models and the remaining 20% for testing. This separation ensures that the model can be trained on a subset of the data and evaluated on an independent subset, providing a realistic assessment of its performance.

C. *Feature Scaling*

Feature scaling is an essential preprocessing step that nor- malizes the range of feature values, ensuring that all features contribute equally to the model’s learning process. In the ”Smart Crop Advisor” system, Standard Scalar Normalization is employed to rescale the features such that they have a mean of 0 and a standard deviation of 1. This normalization helps to address issues related to features with different units or scales, which can otherwise lead to biased model perfor- mance. For instance, features with larger numerical ranges can dominate the learning process, overshadowing features with smaller ranges. By applying Standard Scalar Normalization, each feature is standardized to have a consistent distribution around zero, thus enhancing the stability and effectiveness of the machine learning algorithms. This preprocessing step is crucial for achieving accurate and reliable predictions from the models, as it ensures that all input features are treated uniformly during the training and evaluation phases.

D. *Model Training and Evaluation*

Once the data has been preprocessed and scaled, the next step is to train and evaluate various ML models to determine which offers the best crop recommendations based on envi- ronmental conditions. In this research, nine models are used: Gradient Boosting Machines (GBM), CatBoost, Linear Dis- criminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Nearest Centroid (NC), Decision Tree, Support Vector Machine (SVM), Naive Bayes and Random Forest. Each model contributes uniquely to the recommendation process by leveraging different approaches. GBM is an ensemble technique that builds decision trees sequentially, where each tree corrects the errors of the previous ones, leading to high predictive accuracy and robustness. CatBoost is another powerful gradient-boosting algorithm that efficiently handles categorical features, providing strong performance by reducing overfitting and improving accuracy. Both GBM and CatBoost excel in modeling complex, nonlinear relationships between environmental features like soil type, rainfall, and temperature. LDA is used to find linear combinations of features that best separate different crop classes, reducing dimensionality and projecting data into a lower-dimensional space. QDA extends LDA by assuming that each class follows a differ- ent covariance structure, allowing for more flexible decision boundaries and better classification performance when crop classes exhibit nonlinear separability. Nearest Centroid (NC) offers a simple yet effective approach by classifying crops based on proximity to the average environmental conditions for each crop class, making it computationally efficient and useful for large-scale or real-time applications. The Decision Tree model creates a series of splits based on environmental conditions, offering easy-to-interpret decisions and performing
well with low-variance data.

The Support Vector Machine (SVM) finds an optimal hy-perplane that separates different crop classes, excelling in high-dimensional spaces and ensuring robust performance with nonlinear relationships. Lastly, Random Forest, an ensemble of decision trees, enhances prediction stability and accuracy by averaging the results from multiple trees, reducing overfit- ting. Together, these models offer diverse methodologies that account for the variability in environmental factors affecting crop growth, leading to accurate, reliable, and efficient crop recommendations.

E. *Model Selection*

After training and evaluating all the machine learning models, the selection of the most suitable model is based on performance metrics such as accuracy and confusion matrix. Accuracy measures the proportion of correct predictions made by the model, while the confusion matrix provides detailed in- formation on the types of errors—true positives, true negatives, false positives, and false negatives. Among the models tested, the one with the highest accuracy and the most favorable confusion matrix results is selected. This chosen model is considered the best fit for making accurate crop predictions based on the given environmental conditions. The selection process ensures that the final model is reliable and performs well in providing crop recommendations, which is crucial for the effectiveness of the ”Smart Crop Advisor.”

F. Recommendation System

The Recommendation System utilizes the selected ML model to provide crop suggestions based on user inputs from a web interface. The system takes various environmental parameters—Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and rainfall—from the user. It normalizes these inputs using the same scaling method applied during preprocessing. The input data from the web interface is normalized and then fed into the trained machine-learning model by using Flask to generate a prediction. This pre- diction, represented by a numerical value, is mapped to the corresponding crop label using a predefined dictionary. The system’s output is a specific crop recommendation tailored to the user’s environmental conditions which is displayed on the web interface. This approach ensures that the ”Smart Crop Advisor” delivers accurate and relevant crop suggestions, aiding users in making informed agricultural decisions.

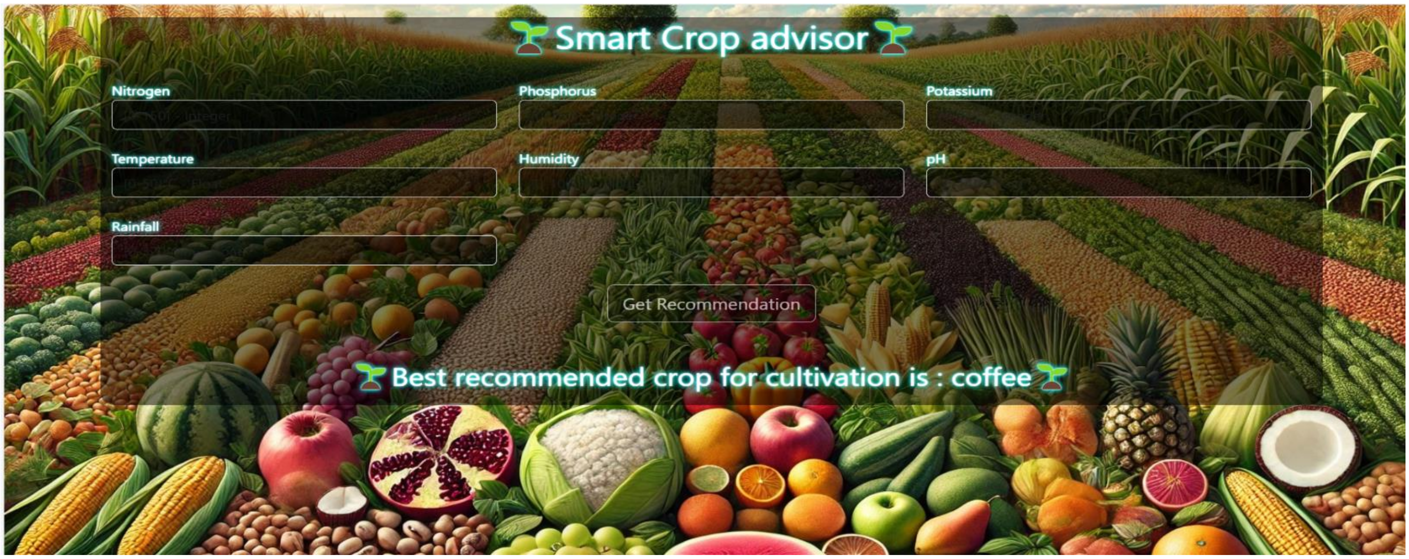

Fig. 2. Recommendation 1

Fig. 3. Recommendation 2

IMPLEMENTATION

Model Evaluation using Performance Metrics

In this study, various performance metrics such as recall, F1-score, precision, and accuracy are used to compare the methods under investigation. These metrics are calculated using the following components: true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

The accuracy of the model is calculated as shown in Equation (1), which is the ratio of correctly predicted instances (both positive and negative) to the total number of predictions. Recall (Equation 2) measures the proportion of actual positive cases that are correctly identified by the model. Precision (Equation 3) evaluates how many of the predicted positive cases are true positives. The F1-score (Equation 4) is the harmonic mean of precision and recall, offering a balanced measure when there is an uneven class distribution.

The formulas are expressed as follows:

1. Accuracy = (TP + TN) / (TP + FP + FN + TN)

2. Recall = TP / (TP + FN)

3. Precision = TP / (TP + FP)

4. F1 Score = 2 * (Recall * Precision) / (Recall + Precision)

Cost and sustainability Impact

There are several cost implications as well as sustainability considerations associated with the Smart Crop Advisor project that would need to be carefully evaluated to ensure that their implementation is a success.

i. **Cost Implications:**The major cost factors for the Smart Crop Advisor project include the following:

a. **Hardware Costs:**Hence, the Smart Crop Advisor will need an initial huge investment in hardware, such as sensors, storage for data, and even IoT for real-time monitoring of agriculture conditions. These technologies can be very expensive in the initial stages, especially when implemented using high-precision equipment that has to collect accurate data for any useful good analysis.

b. Software and Development Cost: The development of the Smart Crop Advisor incurring software development-related expenses including licensing proprietary tools and frameworks. While open-source solutions might save cost, the requirement of the specialized algorithms and machine learning models for crop prediction and analysis is quite likely to increase development costs.

c. **Data Acquisition and Storage:**Machine learning acquires expensive good-quality training datasets. Though not all of this may be at the point of purchase, farmers or organizations may have to pay money for acquiring commercial data sets especially huge volumes of agricultural data requiring efficient and safe storage solutions.

d. **Operating Cost:**Operating the Smart Crop Advisor would attract constant cost of operation: electricity for hardware, maintaining sensors, and other subscription costs related to cloud services meant for data processing and analytics.

ii. Sustainability Impact:The impact of sustainability for the Smart Crop Advisor can be monitored through various perspectives as follows:

a. Resource Efficiency: The Smart Crop Advisor would optimize resource usage in agriculture, including water, fertilizers, and pesticides. Precise recommendations following data analysis lead to a major wastage reduction, thus sustainable agriculture.

b. Energy Consumption: Though the technology may consume more energy in processing data and training models, there is a need for efforts that may reduce the carbon footprint. That is through sourcing power from renewable sources, as well as the use of energy-efficient technologies.

c. **Environmental Impact:**Environmental impacts have to be taken into consideration when deploying the Smart Crop Advisor, especially concerning the production and disposal of hardware parts. Use of materials that have been responsibly sourced and promoting recycling can help mitigate negative environmental impacts.

d. **Social Impact:**But a more critical issue is the matter of data privacy ethics and equitable sharing of benefit of technological advancement for smallholder farmers, as well The Smart Crop Advisor can be used to empower farmers through data-driven insights, which enhance productivity and decision-making.

RESULTS

I. TABLE
PERFORMANCE METRICS OF VARIOUS MODELS ON CROP RECOMMENDATION DATASET

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	99.7%	0.997	0.997	0.997
Naive Bayes	99.2%	0.992	0.992	0.992
CatBoost	99.2%	0.991	0.992	0.991
GBM	98.9%	0.990	0.989	0.989
QDA	98.9%	0.989	0.989	0.989
SVM	97.7%	0.982	0.977	0.974
Decision Tree	97.2%	0.974	0.972	0.970
LDA	97.2%	0.972	0.972	0.972
NC	89.8%	0.911	0.898	0.898

Table 1 presents the performance metrics of various ma- chine learning models trained on the Crop Recommendation Dataset available on Kaggle. These models were evaluated to recommend crops efficiently based on soil and weather conditions, with metrics including accuracy, precision, recall, and F1 score. Among the models, Random Forest achieved the highest accuracy of 99.7%, with nearly perfect precision, recall, and F1 score, indicating its robustness in predicting the most suitable crops. Naive Bayes and CatBoost also performed exceptionally well, both achieving high validation accuracies of 99.2%. On the other hand, the Nearest Centroid (NC) model had the lowest accuracy at 89.8%, though it still demonstrated a reasonable performance. Models like GBM and QDA also showed strong results with accuracies of 98.9%. These results highlight the varying capabilities of different models in predicting crop recommendations based on the dataset, providing valuable insights for future research and applications in agriculture.

Confusion Matrices of various trained models are shown in Fig4-Fig12. They show the classification capacity and efficiency of each model in learning the dataset and predicting

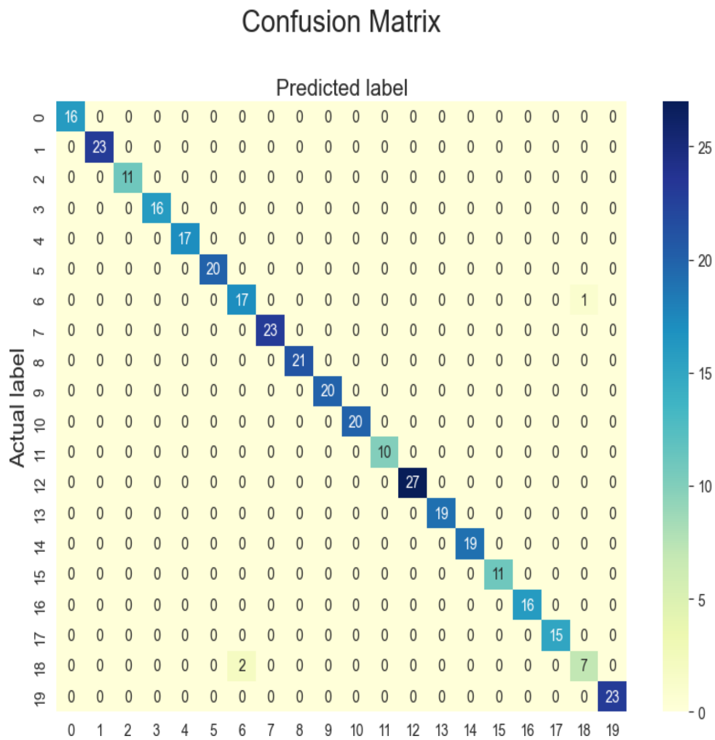

Fig. 4. GBM

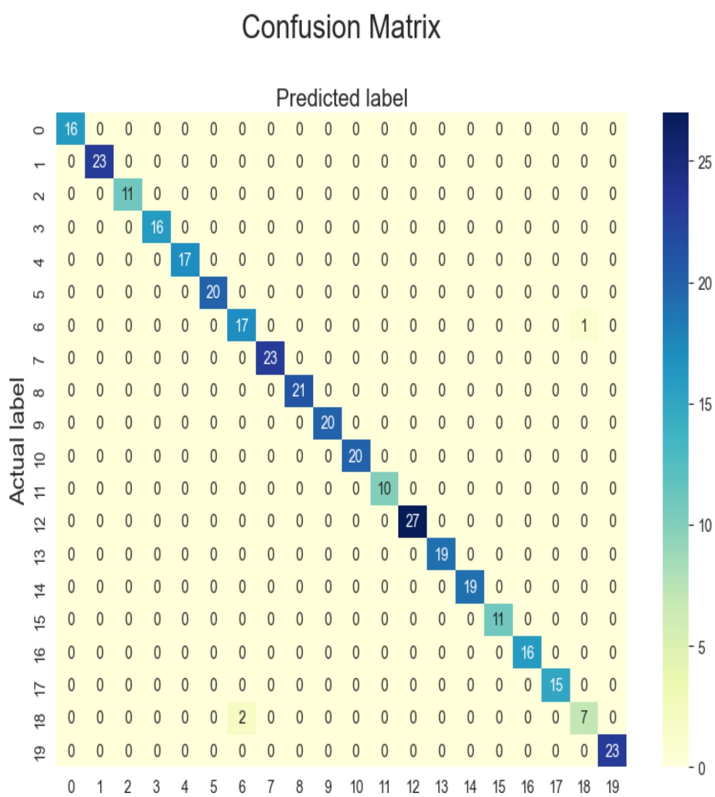

Fig. 5. C

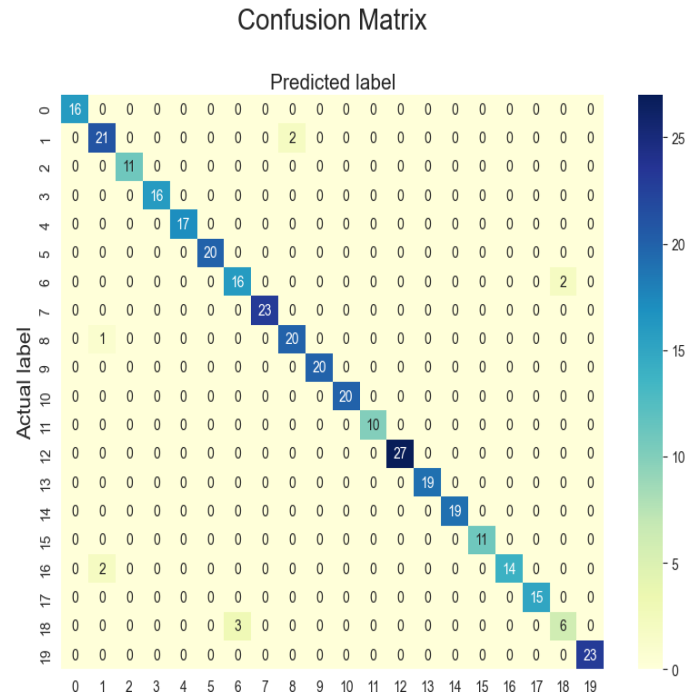

Fig. 6. LDA

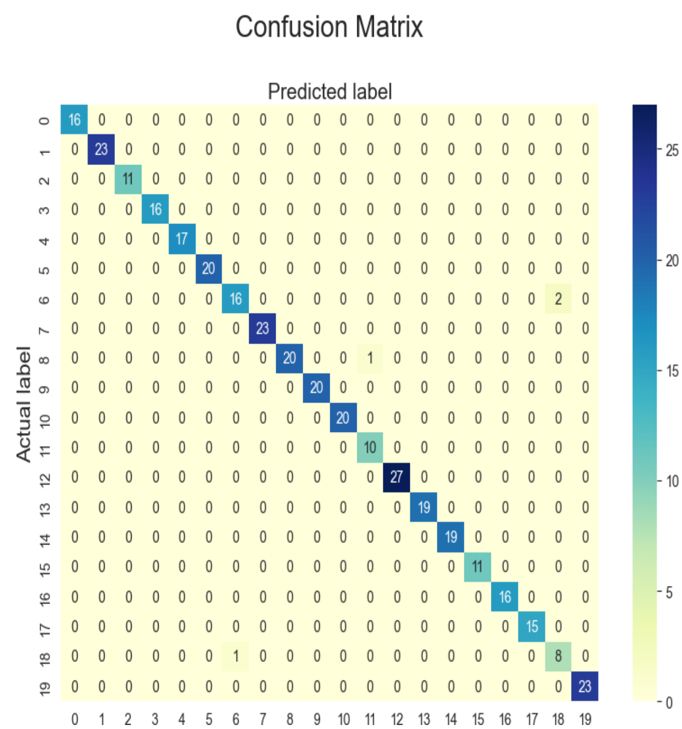

Fig. 7. QDA

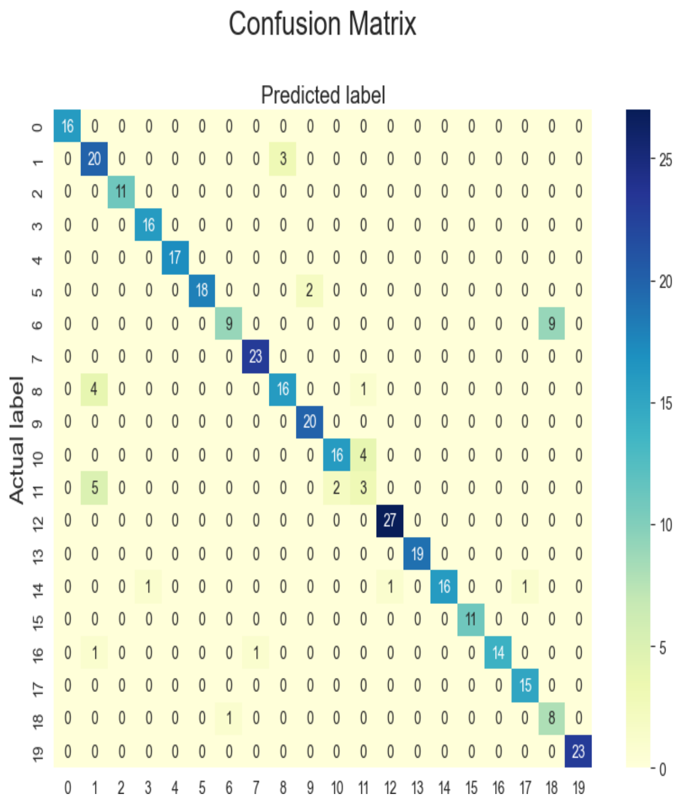

Fig. 8. Nearest Centroid

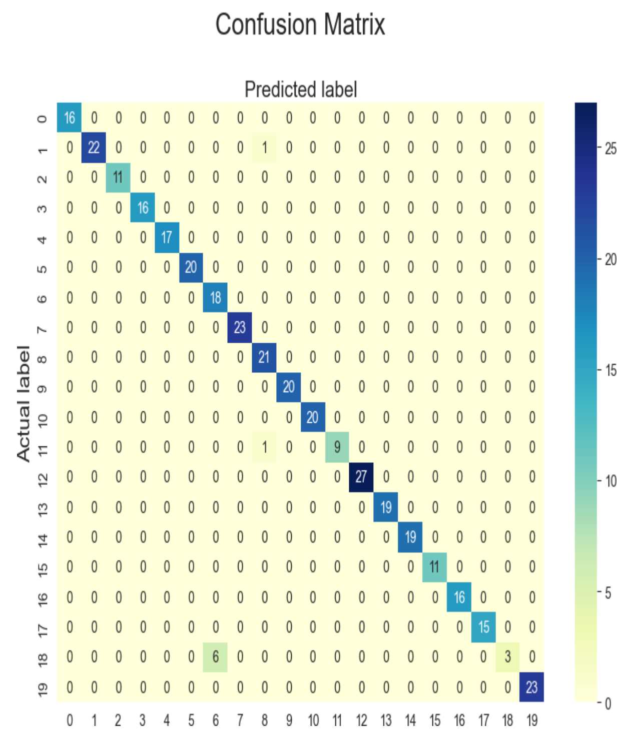

Fig. 10. Support Vector Machine

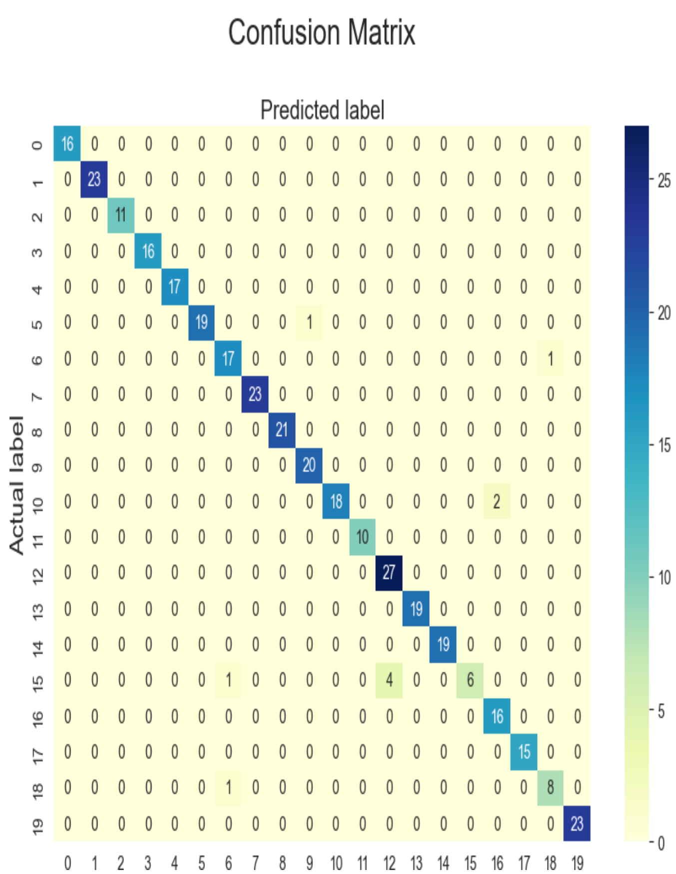

Fig. 9. Decision Tree

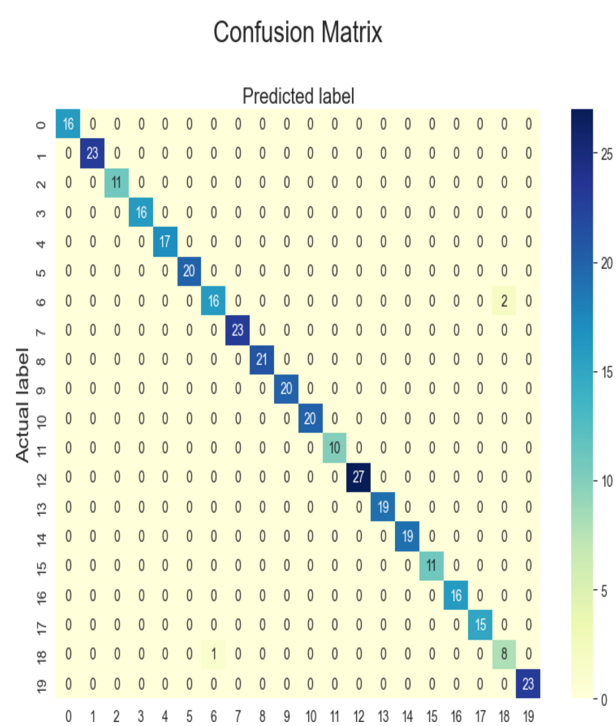

Fig. 11. Naive Bayes

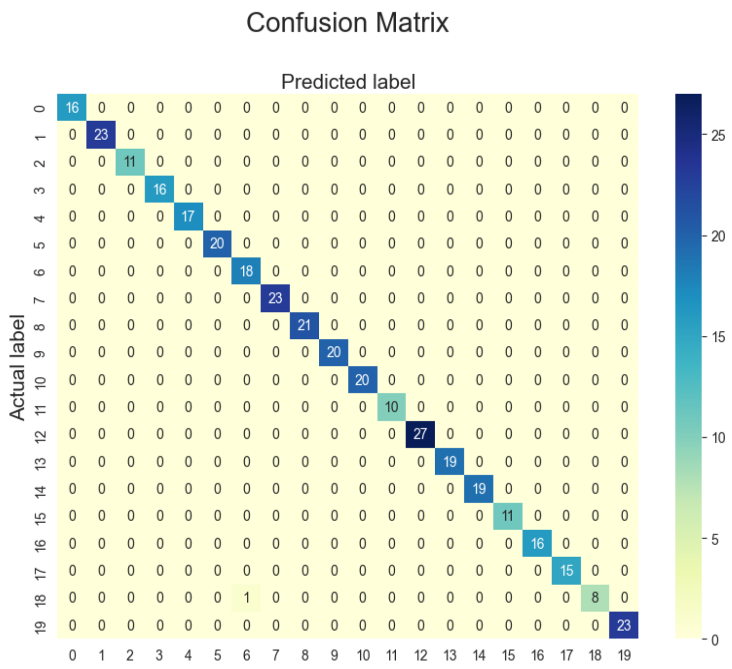

Fig. 12. Random Forest

CONCLUSION

Smart Crop Advisor is a novel innovation that can potentially change agriculture through the application of data-driven insights and advanced algorithms using machine learning. This is optimized both in the use of water, fertilizers, and pesticides by informing farmers' decisions that promote crop yield with minimal environmental impact. On the implementation base, there are many cost considerations starting from hardware investment during initial stages up to operational costs over time. Besides, sustainability is paramount, in that the success of the project directly depends on minimizing energy use and lessening deterioration of the environment to be sustained over a long period of time.

Summary In brief, the crop advisor is an innovation technology that has much in store for the modern farmer, particularly in terms of improving productivity and sustainability. Appropriate costing and resource efficiency and social impacts are a step that should be taken seriously to ensure the right implementation of the Smart Crop Advisor. If done right in terms of ethics and responsibility, this technology will play an important part in what is coming for future farms-that is, sustainable, efficient, and resilient agricultural systems.

REFERENCES

[1] Pawar, Shubham, et al. ”Smart farming using machine learning.” Smart Comput (2021): 73-81.

[2] Rao, Madhuri Shripathi, et al. ”Crop prediction using machine learning.” Journal of Physics: Conference Series. Vol. 2161. No. 1. IOP Publishing, 2022.

[3] Prity, F.S., Hasan, M.M., Saif, S.H. et al. Enhancing Agricultural Productivity: A Machine Learning Approach to Crop Recommendations. Hum-Cent Intell Syst (2024).

[4] Sharma, Priyanka, Pankaj Dadheech, and AV Senthil Kumar Senthil. ”AI-Enabled Crop Recommendation System Based on Soil and Weather Patterns.” Artificial Intelligence Tools and Technologies for Smart Farm- ing and Agriculture Practices. IGI Global, 2023. 184-199.

[5] Dahiphale, Devendra, et al. ”Smart farming: Crop recommendation using machine learning with challenges and future ideas.” Authorea Preprints (2023).

[6] Garg, Disha, and Mansaf Alam. ”An effective crop recommendation method using machine learning techniques.” International Journal of Advanced Technology and Engineering Exploration 10.102 (2023): 498.

[7] Ghosh, Asit, et al. ”A Comprehensive Crop Recommendation System Integrating Machine Learning and Deep Learning Models.” 2024 1st In- ternational Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU). IEEE, 2024.

[8] ReenaSri, S., U. Dhanushree, and K. Sivadhanu. ”Crop Recommenda- tion Systems using Machine Learning Algorithms.” 2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET). IEEE, 2023.

[9] Parameswari, Ponnusamy, N. Rajathi, and K. J. Harshanaa. ”Machine learning approaches for crop recommendation.” 2021 international con- ference on advancements in electrical, electronics, communication, com- puting and automation (ICAECA). IEEE, 2021.

[10] Mewar, Amritya, et al. ”Design of Web Based Recommendation System for Farmers using Machine Learning.” 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET). IEEE, 2024.

[11] Iniyan, S., et al. ”Crop and Fertilizer Recommendation System Apply- ing Machine Learning Classifiers.” 2023 International Conference on Emerging Research in Computational Science (ICERCS). IEEE, 2023.

[12] Shedthi, Shabari, et al. ”Machine Learning Techniques in Crop Recom- mendation based on Soil and Crop Yield Prediction System–Review.” 2022 International Conference on Artificial Intelligence and Data Engi- neering (AIDE). IEEE, 2022.