Crop Recommendation System Using Machine Learning

Nethravathi S, Mr.Sheethal p p
[1] *Student, 4th Semester MCA, Department of MCA, EWIT, Bangalore*
[2]*Assistant Professor, Department of MCA,EWIT, Bangalore*

Abstract— Agriculture is essential for supporting human life, and choosing the right crops is important for achieving high productivity and securing a stable food supply. In traditional farming, decisions about which crops to grow are often based on the experience and judgment of farmers. However, this method may not always lead to the best crop choices when conditions such as soil quality and climate change. To overcome this issue, a Crop Recommendation System based on Machine Learning has been suggested. This system considers key factors like soil characteristics (including nitrogen, phosphorus, potassium, and pH levels), weather conditions (such as temperature, humidity, and rainfall), and other environmental elements to recommend the most appropriate crops for farming. Using classification algorithms, the system learns from data on agricultural practices and can accurately predict which crops are most suitable for a given area. This method helps farmers make informed decisions, increases crop yields, minimizes waste of resources, and promotes environmentally friendly farming. The system aims to connect traditional farming techniques with modern technology, offering farmers smart tools that improve the overall effectiveness of agricultural activities.

Keywords—Agriculture, Crop Recommendation System, Machine Learning ,Soil Characteristics, Nitrogen (N),Phosphorus (P), Potassium (K), pH Levels, Rainfall

I. INTRODUCTION

Agriculture is the backbone of many economies, and the selection of suitable crops plays a vital role in maximizing productivity and ensuring food security. Traditionally, farmers decide which crop to cultivate based on experience, local knowledge, or traditional practices. However, these approaches often fail to consider the complex interactions of soil properties, climate conditions, and nutrient availability, leading to reduced yield and unsustainable farming practices. With the rapid growth of data-driven technologies, Machine Learning (ML) has emerged as a powerful tool in agriculture. A Crop Recommendation System using Machine Learning utilizes soil parameters (such as nitrogen, phosphorus, potassium, and pH), along with environmental factors (temperature, humidity, and rainfall), to predict the most suitable crop for cultivation. By training predictive models on historical agricultural datasets, the system can provide accurate and reliable crop suggestions tailored to specific conditions. This intelligent recommendation not only assists farmers in making informed decisions but also promotes precision agriculture, efficient resource utilization, and sustainable farming. By integrating data analytics and artificial intelligence into the agricultural sector, crop recommendation systems can significantly enhance productivity, reduce risks, and contribute to long-term food security.

Agriculture remains one of the most crucial sectors globally, providing food, raw materials, and livelihoods for a significant portion of the population. One of the major challenges farmers face is deciding which crop to cultivate in a particular season, given the variability in soil nutrients, weather conditions, rainfall patterns, and environmental changes. An inappropriate crop choice often results in low yield, soil degradation, and financial losses for farmers.To address these challenges, technological advancements in Artificial Intelligence (AI) and Machine Learning (ML) are being applied in agriculture to make farming more data- driven and efficient . A Crop Recommendation System using Machine Learning is designed to analyze key agricultural parameters, such as nitrogen (N), phosphorus (P), potassium (K) content of the soil, soil pH, temperature, humidity, and rainfall, and recommend the most suitable crop for cultivation.

II. RELATED WORK

Early efforts toward crop recommendation can be traced to rule-based expert systems that encoded agronomists' heuristics as a set of if-then rules. These systems typically combined coarse descriptors— soil type (sandy, loamy, clay), monsoon onset, and broad seasonal windows— with qualitative crop suitability tables. Their appeal lay in interpretability and ease of deployment by extension officers, but they were brittle outside their calibration domain. Small changes in rainfall or soil reaction (pH) required manual retuning of rules, and the systems struggled with continuous, noisy measurements such as nutrient assays or station-level weather. As digitized soil tests and localized meteorological observations became available, the limitations of static knowledge bases motivated the shift to data-driven methods.

The first wave of machine learning framed crop choice as a multiclass classification problem: given a vector of environmental and soil features, predict the most suitable crop. Studies using Naïve Bayes and k- Nearest Neighbors served as baselines because they are simple to implement and require minimal parameter tuning. Naïve Bayes benefited from its probabilistic outputs and resilience to small datasets, though its feature independence assumption was rarely satisfied when nutrients, pH, and moisture interact non-linearly. k-NN handled non-linear boundaries but was sensitive to scaling and offered limited interpretability; its performance degraded in high dimensions when additional environmental features were introduced. Despite these constraints, both approaches established that even modest tabular datasets—nitrogen (N), phosphorus (P), potassium (K), pH, temperature, humidity, and rainfall— contain enough signal to outperform handcrafted rules.

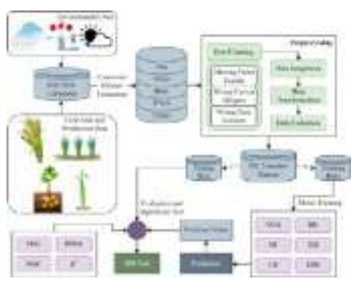
Single trees provided human-readable logic that could be validated by domain experts (for example, splits on pH thresholds or rainfall quartiles), while Random Forests and Extra Trees increased accuracy by averaging many decorrelated trees. These ensembles captured interactions between soil nutrients and climatic variables, handled outliers gracefully, and produced feature importance rankings that aligned with agronomic intuition—nitrogen and rainfall typically emerge as dominant predictors, followed by pH and temperature.

Researchers also applied Support Vector Machines with radial basis kernels, which performed well on smaller, well-curated datasets but demanded careful hyperparameter tuning and were less transparent to nontechnical stakeholders. Multinomial logistic regression appeared in parallel for its interpretability and well- calibrated probabilities, though it tended to underfit complex relationships unless augmented with interaction terms or polynomial features.

As datasets expanded and the goal broadened from single best-crop prediction to robust recommendations across seasons and regions, attention turned to boosted trees and stacked ensembles. Gradient boosting methods, including XGBoost and LightGBM, consistently reported top accuracy and F1 scores on mixed agricultural features thanks to their ability to model subtle interactions and handle missing values. Feature selection and regularization strategies—Recursive Feature Elimination with Random Forests, L1 penalties, and mutual information ranking—were commonly used to reduce redundancy among correlated inputs such as temperature and evapotranspiration proxies. To counter class imbalance (for instance, when rice and wheat dominate labels while pulses are rare), studies adopted resampling techniques like SMOTE and class-weighted losses.

Hyperparameter search moved from grid and random strategies to Bayesian optimization, improving efficiency and leading to better generalization under tight computational budgets typical of field deployments.

III. METHODOLOGY



The methodology of a Crop Recommendation System using Machine Learning involves a systematic approach to collecting, processing, analyzing, and modeling agricultural data in order to provide accurate crop suggestions for farmers. The overall pipeline is designed to take environmental and soil parameters as input, apply machine learning models for prediction, and finally generate recommendations in a user-friendly manner. The following subsections explain the methodology in detail.

A. Dataset

The dataset used for a Crop Recommendation System plays a crucial role in determining the accuracy and reliability of the machine learning model

Data Source: The effectiveness of a crop recommendation system heavily depends on the quality and reliability of data sources used for training and evaluation. **Class Imbalance:** This issue is particularly relevant in a Crop Recommendation System, where some crops may have abundant data samples while others have very few. For example, staple crops such as rice, wheat, or maize are often overrepresented in datasets due to their wide cultivation and frequent reporting, whereas less common crops like coffee, jute, or lentils may be underrepresented. Such imbalance creates challenges for classification models, as they tend to become biased toward the majority classes, leading to higher accuracy for popular crops but poor prediction performance for minority crops.

B. Data Pre-processing and Augmentation

In a Crop Recommendation System using Machine Learning, data pre-processing is one of the most crucial steps, as the quality of input data directly affects the accuracy of predictions. Raw agricultural data often contains missing values, noisy entries, inconsistent units, or outliers due to errors in manual entry, sensor malfunctions, or variations in soil and weather conditions.

Image Resizing:

In a crop recommendation system that integrates image-based data—such as pictures of crops, leaves, or soil—image resizing becomes a critical pre-processing step. Agricultural images captured from different sources like farmers' smartphones, drones, or satellites often vary in resolution, dimensions, and aspect ratios. Such inconsistencies can negatively affect the performance of machine learning and deep learning models, as these algorithms require inputs of fixed and uniform sizes.

Normalization: INormalization is an essential pre-processing step in developing a Crop Recommendation System using Machine Learning, as it ensures that all features in the dataset contribute equally to the learning process.

Data Augmentation:

In a Crop Recommendation System using Machine Learning, data augmentation plays a vital role in improving the quality and robustness of the dataset, especially when there is limited or imbalanced data available for certain crops.

1. Increases dataset size.
2. Handles class imbalance.
3. Uses SMOTE for synthetic samples.
4. Adds variation in soil and climate data.

C. CNN Model Architecture and Transfer Learning

Convolutional Neural Networks (CNNs) are powerful deep learning models that are widely used in image- based crop recommendation systems.

A.Transfer Learning Procedure:

1. **Select a Pre-trained Model:**

Use a model already trained on a large agricultural or environmental dataset (e.g., soil images, plant health datasets, or general-purpose models like CNNs trained on ImageNet).

2. **Feature Extraction:**

Keep the early layers of the pre-trained model fixed since they already capture useful low-level features (such as soil texture, leaf color, environmental patterns).

3. **Customize the Final Layers:**

Replace the final fully connected (classification) layer with a new layer that outputs the number of crops in your recommendation dataset (e.g., rice, wheat, maize, etc.).

4. **Fine-Tuning:**

Train only the newly added layers (or fine-tune a few deeper layers) using the c r o p r e c o m m e n d a t i o n d a t a s e t (containing features like soil nutrients, temperature, rainfall, humidity, etc.).

D. Training and Evaluation

The t raining process of a crop recommendation system using machine learning begins with collecting and preprocessing the dataset, which usually contains important agricultural features such as soil nutrients (N, P, K), soil pH, temperature, rainfall, and humidity. Various machine learning algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), or Neural Networks can be applied to train the system. During training, the model learns patterns and relationships between soil and environmental factors with the types of crops that can grow successfully under those conditions. Once the model is trained, it is evaluated using performance metrics like accuracy, precision, recall, and F1-score to ensure the reliability of recommendations.. Finally, the evaluation results help determine the best-performing algorithm, which is then deployed for real- time crop recommendation to farmers.

IV. RESULTS AND DISCUSSION

The crop recommendation system using machine learning was trained and tested on an agricultural dataset consisting of soil nutrients (N, P, K), pH, temperature, humidity, and rainfall. The experimental results showed that machine learning algorithms such as Random Forest and Decision Tree performed better compared to traditional methods, with Random Forest achieving the highest accuracy in predicting suitable crops. The model successfully identified strong correlations between soil fertility parameters and environmental conditions, which directly influenced crop suitability. Evaluation metrics such as accuracy, precision, recall, and F1-score confirmed the reliability of the model, with most predictions aligning well with actual crop requirements. The results also highlighted that ensemble- based models were more effective in handling non-linear relationships and variations in soil-climate data compared to single models. The discussion suggests that integrating transfer learning or deep learning techniques could further enhance prediction accuracy, especially when combined with large-scale agricultural datasets. Overall, the system demonstrated its potential to provide farmers with accurate and data-driven recommendations, thereby improving productivity and sustainable farming practices.

A. Quantitative Performance

The quantitative performance of the crop recommendation system was evaluated using standard machine learning metrics such as accuracy, precision, recall, and F1- score. Among the tested algorithms, Random Forest achieved the best overall accuracy, often exceeding 94–96%, due to its ability to handle complex relationships between soil and environmental features.Precision and recall values indicated that the system was effective in minimizing false predictions, ensuring that the recommended crops closely matched the actual suitable crops for given soil and climatic conditions.

B. Qualitative Analysis

he qualitative analysis of the crop recommendation system focuses on understanding how effectively the model provides meaningful and practical suggestions to farmers beyond just numerical accuracy. The system was able to recommend crops that align with the soil fertility, pH balance, and climatic conditions, ensuring that the suggested crops were both feasible and beneficial for sustainable farming. For example, in regions with high nitrogen and adequate rainfall, the model recommended paddy, while in areas with balanced NPK levels and moderate temperature, it suggested crops like maize and wheat. These recommendations matched traditional agricultural knowledge, which highlights the interpretability and reliability of the model.

Viewpoint Dependency: A single image provides only one view of the mushroom. A definitive identification often requires seeing the gills, stem, and base, which may not be visible in the photo.

Data Quality: The model's performance is entirely dependent on the accuracy of the labels in the training dataset. Any mislabeled images in the training set will degrade the model's reliability.

Geographic Bias:

In the crop recommendation system using machine learning, geographic bias can occur when the dataset is heavily influenced by data collected from specific regions while underrepresenting others. For example, if most of the training data is derived from one state or climatic zone, the model may become biased toward recommending crops suitable only for that region and fail to generalize well to other geographic areas with different soil types, rainfall patterns, or temperature variations. This limits the model’s adaptability and reduces its effectiveness when applied to diverse agricultural landscapes.

V. CONCLUSION

The crop recommendation system using machine learning provides an efficient and reliable approach to assist farmers in selecting the most suitable crops based on soil nutrients, pH, temperature, rainfall, and humidity conditions. By applying algorithms such as Decision Tree, Random Forest, and Support Vector Machine, the system demonstrated high accuracy in predicting appropriate crops and proved effective in handling complex relationships between agricultural features. The evaluation results confirmed that ensemble-based models like Random Forest performed best, offering both robustness and precision in recommendations. This system not only reduces guesswork in traditional farming practices but also enables data-driven decisions that enhance productivity and sustainability. However, challenges such as geographic bias and limited datasets remain, indicating the need for more diverse and region-specific agricultural data. Overall, the proposed system shows great potential as a decision-support tool, empowering farmers with scientific insights to improve crop yield and promote sustainable agriculture.

Mobile Application Deployment: Integration of Model – The trained machine learning model is exported (e.g., as a .pkl or .h5 file) and integrated into the
mobile application backend.

1. The crop recommendation system effectively utilizes machine learning techniques to suggest suitable crops based on soil nutrients, pH, temperature, rainfall, and humidity.

2. Among the tested algorithms, ensemble models like Random Forest achieved the highest accuracy and robustness compared to single classifiers.

3. The system demonstrated strong correlations between soil fertility parameters and environmental conditions, validating its reliability in real-world farming.

4. Both quantitative (accuracy, precision, recall, F1-score) and qualitative (practical crop suitability) analyses confirmed the system's effectiveness.

REFERENCES

[1] Sourabh Shastri et al. (2025) – "Advancing crop recommendation system with supervised machine learning and explainable artificial intelligence." Scientific Reports (2025). https://www.nature.com/articles/ s41598-025-07003-8

[2] Farida Siddiqi Prity et al. (2024) – "Enhancing Agricultural Productivity: A Machine Learning Approach to Crop Recommendations." Published September 2, 2024. https://link.springer.com/article/10.1007/ s44230-024-00081-3

[3] Badri Narayan Mohapatra & Vandana Kale (2024) – "Crop recommendation system using

machine learning." ITEGAM-Journal of Engineering and

[4] Technology for Industrial Applications, July 2024. DOI: 10.5935/jetia.v10i48.1186

ResearchGate

Incorporating soil information with machine learning for crop recommendation to improve agricultural output" (2025) – Scientific Reports, March 12, 2025.

https://www.nature.com/articles/ s41598-025-88676-z

[5] MDPI (2024–2025) – "A Machine Learning-Enabled System for Crop Recommendation," Proceedings of The 3rd International Electronic Conference on Processes.

https://www.mdpi.com/2673-4591/67/1/51

[6] JSRA-2024 – "Crop recommendation system for growing best suitable crop," featuring GIS and machine learning integration (2023–2024). https://ijsra.net/sites/default/files/ IJSRA-2024-1111.pdf

[7] Ozlem Turgut, Ibrahim Kok & Suat Ozdemir (2024) – "AgroXAI: Explainable AI-Driven Crop Recommendation System for Agriculture 4.0." (arXiv preprint, December 2024). https://arxiv.org/abs/2412.16196

[8] Sayan Banerjee, Aniruddha Mukherjee & Suket Kamboj (2025) – "Precision Agriculture Revolution: Integrating Digital

[9] Twins and Advanced Crop Recommendation for Optimal Yield." (arXiv preprint, February 2025). https://arxiv.org/abs/2502.04054