International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Smart Crop Recommendation System Using Machine Learning for Optimized Agricultural Practices

Yashwant Pradhan [1], Momita Kundu [2], Harsh Vardhan Sinha[3], Paritosh Kumar [4] and Raja Pandey [5]

[1,3,4,5] *B. Tech, CSE Dept., R. V. S College of Engg. & Tech., Jamshedpur*
[2]*Professor, MCA Dept., R. V. S College of Engg. & Tech., Jamshedpur*

Abstract

The goal of the Smart Crop Recommendation System is to use machine learning and data specific to each farmer, to inform them of the best crops to grow depending on their environment, as well as past and real-time conditions. The integrated system utilizes both static agronomic data, such as soil type, nutrient levels and soil pH, and dynamic environmental conditions—such as the temperature, humidity and the amount of rainfall—to effectively recommend the best crop. Much traditional farming consulting approaches offer often generalized 'solutions' that ignore the context at which the crops is intended to be grown. By using additional advanced preprocessing and engineering techniques, and machine learning techniques like Random Forest, Decision Trees, and Support Vector Machine, recommendations are provided as adaptive and location specific recommendations. Inputted data are pre-processed and then vectorized utilizing different numeric encoding methods, to allow for complex pattern recognition in real-time, that affects where a crop can be grown. Further contextual static data that captures seasonal trends, used in conjunction with local growing practices will also be included to improve prediction performance accuracy. Preprocessing wraps around and inputs the static and contextual data together in the same pipeline workflow. Model architectures are trained using state of the art hyperparameter optimization and search techniques like Grid Search and cross validation to identify the most optimal model to showcase. At last, hyperparameter-optimized model is integrated into a solid deployment pipeline with an interactive Streamlit interface for farmers to input their soil and environmental data and receive real-time, educated and informed crop recommendations . We expect this tool will be a scalable and data-driven decision support program will ultimately drive sustainable agriculture and improve farm efficiency with crop planning across different agro-climatic zones.

Keywords: Artificial Intelligence (AI), Machine Learning (ML), Precision Agriculture, Crop Recommendation System, Environmental Data Analysis, Soil Analysis, Weather Forecast Integration, Feature Engineering, Data Preprocessing, Random Forest, Decision Trees, Support Vector Machines (SVM), Hyperparameter Tuning, Predictive Modeling, Agricultural Informatics, Sustainable Farming, Streamlit Interface, Geospatial Intelligence, Smart Farming, Decision Support System (DSS).

1. Introduction

Identifying an ideal crop from climatic and soil information is a significant application of artificial intelligence in our understanding of precision agriculture. Practicing farmers often defer decisions based on standard advisory systems or standardized practices. Historical systems of informing decisions in farming always involve some manual and slow processes (the more "people involved" the slower), which also limits the scale of that information to an entire country can't be adapted to local agro-climatic conditions. The weakness of this traditional system may result in under performances due to poor selection and yield losses, but also wastes resources and time. The rationale for choosing the right crop to grow using AI can consider static and dynamic agricultural variables, such as the condition of the soil and live monitoring of the weather on decision-making, to produce effective, timely, personal, and scientific recommendations. In a situation where machine learning assisted by agronomy can overcome the limitations of past methods and support sustainable farming decisions at a regional scale.

In this study, the Smart Crop Recommendation System is presented. It is a machine learning (ML) platform designed to use information acquired from the agricultural and environmental domains to analyze which crops can be considered for real-time cultivation. By using a structured set of agronomic inputs, this system can evaluate important factors like soil type, pH, nutrient content, temperature, moisture/ relative humidity and rainfall to make accurate crop recommendations for a specific area. This platform also employs feature engineering and pre-processing methods to convert raw input features to a meaningful vector format suitable for an ML model. The recommendation logic is defined by an ML algorithm (once trained and tuned), and a new Python-based backend and user interface that was built from `streamlit`, allows users to enter their environmental inputs and immediately receive a crop recommendation based on machine-generated suggestions. Ultimately, it helps foster precision agriculture through the decision-making process and it provides a means to increase crop yields through a data-driven intelligent process.

A primary strength of this system is its ability to provide efficient, real-time classification already knowing the usual language patterns of people who are depressed. The system allows for early identification of possibly depressed individuals and also allows the person to track their emotional history, in a systematic way. The use of the data is supported by secure ways of storing and retrieving the data and users and researchers can view historical data on prior entries, for ongoing assessment and more nuanced analysis.

Safety, privacy, and accessibility are key tenets of the Smart Crop Recommendation System. The platform leverages data protection best practices, storing no personal identifiable information and only processing user provided agricultural inputs. The system is lightweight and can be delivered by cloud service providers; making it scalable, device-independent, and extremely accessible to farmers, researchers, and agricultural personnel. By providing a machine-learning automated crop selection process to the user and incorporating real-time environmental data; the platform is a smart and trustable tool to enhance crop planning; while supporting sustainable agricultural development.

2. Literature Survey

Sustainable farming practices and careful data-driven solutions are becoming the priority in agricultural development, which has intensified the area of smart crop recommendation in agricultural technology. There is now more access to agronomic and environmental data, and the use of computational approaches can further improve the processes of recommending what crops to grow, with data such as soil profile, weather forecasts and historic data of the crops. Therefore, machine learning (ML) will be central for analysing the depth and breadth of the variables available as a dataset. ML enables agronomists or advisory systems to develop accurate and contextual crop recommendations for specific places and situations.

Many studies have investigated various methods for agricultural crop recommendation. For example, conventional approaches, such as manual consultation with agricultural professionals or rule-based advisory systems, often lack the contextual sensitivity needed to accommodate place-specific socio-economic factors and dynamic, real-time variations in the environment. They are static, generic, and further removed from the unpredictable complexity of current-day agricultural environments; thus, they dismiss emergent nuanced patterns in soil and climate emergent behaviours that help to make viable crop selections and bed yields.

To address these hindrances, machine learning (ML)-based models have gained wide usage in contemporary agricultural systems. There are several commonly used ML algorithms, including Support Vector Machines (SVM), Decision Trees, Random Forest, K-Nearest Neighbours (KNN), and Logistic Regression—as well as more reasonably advanced models such as Multi-Layer Perceptron (MLP). The models use a wide range of agronomic and environmental features—soil type, soil pH, soil temperature, soil humidity, rainfall patterns, etc.—to make predictions. Usually, The "raw" inputs are changed by feature extraction and pre-processing techniques—normalization, one-hot encoding for categorical data, feature scaling, and dimension reduction—into formats that the predictive models can robustly train on. These models, as a result of their reliance on data, make the predictions more accurate, scalable, and flexible than conventional predictive approaches for crop based on agronomics and environmental variables.

Typically, datasets used to train these models are sourced from publicly available agricultural repositories, governmental databases, academic publications, and open-source platforms such as Kaggle. Many of these datasets are linked to soil characteristics, weather, crop types, and yield data. For example, many studies have collected datasets from the Indian Ministry of Agriculture and ICAR (Indian Council of Agricultural Research), as well as other organisations like FAO and CGIAR from an international source. In working and preparing the datasets, they were all previously checked and labelled according to each region and the history of the crops. Therefore, the machine learning models—Decision Trees, Random Forest, etc.—could be trained for crop recommendation.

Also, recent papers have presented comprehensive techniques with multiple models (such as Support Vector Machines (SVM), Stochastic Gradient Descent (SGD), Multi-Layer Perceptron (MLP), Naïve Bayes, Random Forest) implemented and tested on various agricultural datasets with preprocessing conducted to carry out feature scaling, different encoding techniques and dimensionality reduction. In this research showed how preprocessing and feature engineering can impact model development. Furthermore, when deep learning models are trained on clean structured datasets and labelled datasets containing environmental variables and crop variables, they present more of an ability to generalize and accuracy.

In addition, several researchers have attempted to use supervised learning strategies for crop recommendation with data derived from regional agricultural surveys or meteorological records. Experimental results have also shown that classifiers such as Random Forest and Logistic Regression perform better than others due to higher accuracy and interpretability, achieving performance levels of over 90% on datasets containing thousands of labelled records with soil type, climate factors and crop yield. Some researchers have also attempted to use deep learning frameworks such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks focused on modeling seasonal cropping patterns which are time-oriented. However, these deep learning models usually require larger well-structured datasets and increased computational resources for training.

While many studies have reported high accuracy results, the challenges of developing trustworthy crop recommendation systems are considerable. These include challenges such as datasets with underrepresented areas or crop types (i.e., they are spatially and structured imbalanced), limited amount high-quality labeled data, and the challenge of interpretability of complex models. Many improvements have been proposed as a response to these challenges, including a combination of hybrid machine learning or ensemble learning used with crop classifiers where hybrid models may incorporate multiple classifiers or combinations of machine learning models and existing knowledge, in the form of rules, has useful properties such as more generalization, robustness and transparency of decision-making, leading to more trust, reliability and acceptance by the end-users (i.e., farmers and agronomists).

Research studies like those conducted by agricultural researchers have indicated that there is a significant increase in classification performance in crop recommendation models using feature selection methods. Cost-sensitive learning methods, and cloud infrastructures to create real-time systems, have been proposed to provide a scalable, accessible, and efficient solution, facilitating accurate and context-aware crop recommendations for farmers in various agro-climatic contexts.

In broad terms, crop recommendation models can be classified in three categories - traditional rule-based, machine-learning based, and hybrid. Rule-based systems are more limited in scope and adaptability, but machine-learning models provide improved accuracy and scalability. Hybrid models offer advantages of domain knowledge and data-driven approach to improve performance and also provide more credible, location-based crop recommendations.

Smart crop recommendation using advanced machine learning techniques not only enables informed decision-making and optimizes resource use, but also marks a step forward when combining data-driven intelligence with modern, sustainable agricultural methods in various farm communities.

3. Methodology

The Smart Crop Recommendation System is made in Python with machine learning models that help to predict what crops are suitable based on the soil and environmental aspects. The approach taken was to start with data collection and end with data preprocessing, feature engineering, selecting the model to apply, training the model, evaluating the model's performance, and deploying the models in an interactive and usable system.

3.1 Data Collection

A labeled dataset of agricultural records is obtained to train and evaluate the crop recommendation strategy. Puplicly available sources are used to guarantee reliability, diversity, and applicability to the crop recommendation strategy. These sources include:

- **ICAR & KVK Datasets:** Official datasets provided by the Indian Council of Agricultural Research (ICAR) and Krishi Vigyan Kendras (KVKs), containing region-wise data on soil characteristics, crop types, and productivity.
- **Kaggle Agricultural Datasets**: Various open-source datasets from Kaggle that include features such as soil pH, nitrogen-phosphorus-potassium (NPK) levels, temperature, humidity, rainfall, and the corresponding recommended crops.
- **Custom Labeled Dataset**: Field-collected data or farmer-submitted records annotated manually by agricultural experts or validated using domain knowledge to ensure high-quality, real-world relevance for model training and evaluation.

3.2 Data Preprocessing

To prepare the data used in the machine learning models, a variety of preprocessing techniques are utilized:

- **Dealing with Missing values:** Records with missing environmental, or soil parameters will either have the missing values statistically imputed, or the records will be removed entirely if they are incomplete.
- **Encoding categorical variables:** Categorical features such as soil type, or crop names are converted to numerical format using various techniques such as one-hot encoding or label encoding.
- Feature scaling: Continuous variables e.g., temperature, pH, humidity, and rainfall are normalized and/or standardized to have consistency across features.
- Outlier identification and deletion: Outliers (very high or low measurements, i.e., extreme rainfall, nutrient levels.) are detected and dealt with to prevent training the model on faulty data.
- **Class imbalance proceeding:** If some crop labels are under-represented, and SMOTE (Synthetic Minority Oversampling Technique) or class weighting can be used to obtain class balance in the dataset.
- **Feature Selection/Engineering:** Selecting feature(s) that are appropriate in order to increase accuracy and reduce the number of dimensions.

3.3 Feature Engineering

Feature engineering assists in relating agricultural and environmental data so that users can utilize their unfortunate situation to better model predictive capability. Important engineering features include:

- **Soil-Based Features:** Values such as pH, nitrogen (N), phosphorus (P), and potassium (K) values that might better represent where crops can be grown.
- **Weather-Based Features:** Objects used to capture the temperature, humidity, and rainfall over time to accommodate for seasonal variability and potential climate impacts.
- **Derived Agronomic Indices:** Created building blocks from the raw data, such as moisture index, ratio of nutrients (N:P:K), or growing degree days (GDD), adding agronomic context - (or measurement).
- **Location-Based Metadata:** Referencing regional or zone-specific identifiers to account for regional-specific farming practices or crop behaviors.

3.4 Model Selection

Multiple supervised machine learning algorithms are taken into account, evaluated as if they were predicted in terms of predictive performance of crop are:

- Logistic Regression: The simple logistic regression is used as a baseline model for crop types' multiclass classification.
- Support vector machines: Effective for high-dimensional feature space and nonlinear decision boundaries.
- Random Forest: An ensemble learning technique that improves accuracy and reduces overfitting by averaging many decision trees.
- **XGBoost / Gradient Boosting**: A widely known high-performance boosting algorithm suitable for complex patterns.
- **Neural networks (optional):** used to model complex non-linear relationships in data, specifically where larger, high-quality datasets are available.

3.5 Model Training

The dataset is divided into training and testing subsets (80/20 split, for example). The chosen model is trained on the training data to learn patterns for optimal crop selection.

3.6 Model Evaluation

The trained model will be evaluated based on other classification metrics to confirm accuracy and robustness:

- **Accuracy:** The degree of correct predictions of all the crops.
- **Precision:** Number of crops predicted as suitable against the number of accurate predicted-suitable crops.
- **Recall (Sensitivity):** Measures all crops a model has correctly identified under the given conditions.
- **F1-Score:** A harmonic mean between precision and recall so to balance false positives and false negatives.
- **ROC-AUC Score:** The ability of a model to discriminate between several classes of crops.

.

3.7 Hyperparameter Tuning

Hyperparameters are optimized using techniques like Grid Search or Randomized Search with Cross-Validation to improve model performance; reduce overfitting and improve generalization ability.

3.8 Model Deployment

The last iteration of the trained model is deployed into a real-time system using some sort of interface commonly such as streamlit so the results for each farmer are user-friendly. The following components are highlighted when bought to deployment:

- **Web interface:** A low-overhead front-end (i.e. Streamlit app) for inputs like soil and weatherization parameters and crop recommendations for outputs.

- **Backend integration:** Backing the trained model is hosted on a backend framework, such as FastAPI and Flask for real-time predictions.

- **User feedback mechanism:** Providing the farmers or agricultural service officer the capability to give feedback on crop suggestions for re-deploying the model using feeding back to re-train.

- **Monitoring and analytics:** Dashboards and logging are built to assess the performance of model predictions and identifies any drifts or changes to the patterns of recommendations.

Security, privacy and ethical implications, especially as they relate to user contributions of agricultural data, are considered during the development phase to produce safe, secure, and ethical ML-based decision support.

Fig1: Interface for the user to enter different parameters

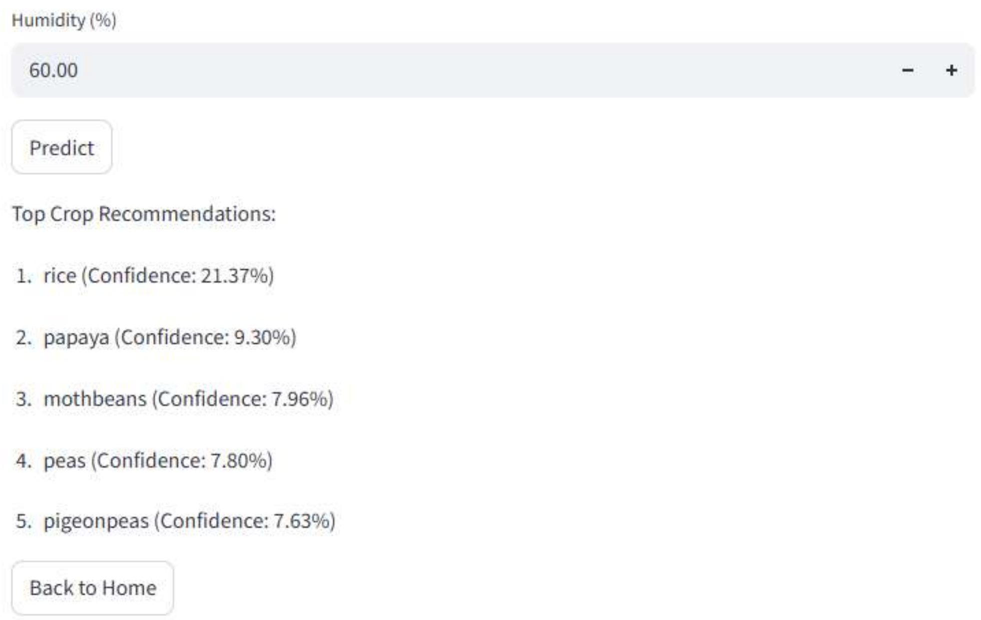

Fig2: Output showing the top 5 recommended crops

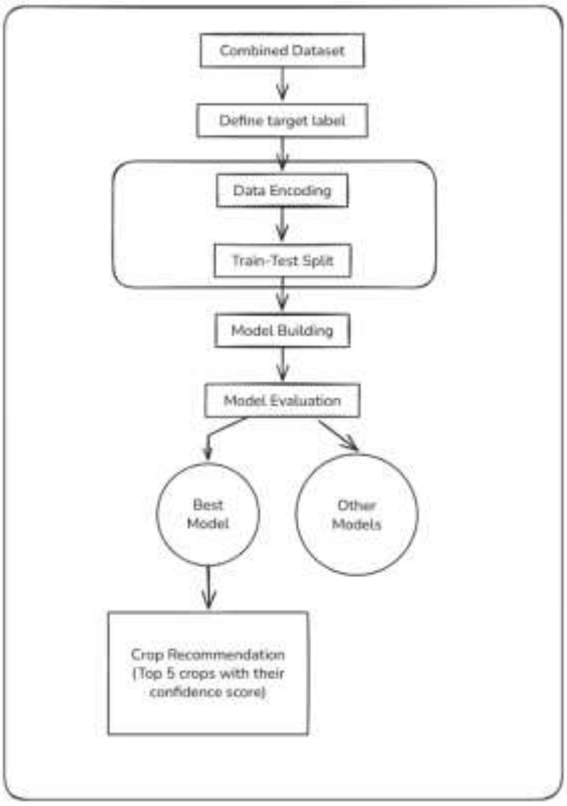

Fig3: Block Diagram of Smart Crop Recommendation System

It is worth mentioning that the Smart Crop Recommendation System offers improved crop selection by considering environmental and soil data, but some challenges persist, including reducing predictive errors, processing incomplete or inconsistent data, and the need to achieve interpretability and trust among end-users, like farmers, that leverage these outputs. Addressing these challenges and enhancing the system's potential for real agricultural applications such as seasonal crop planning, sustainable farming and resource promoton require continued research, model updates and working with agronomists and domain experts.

4. Conclusion and Future Work:

4.1 Conclusion :

A good crop recommendation system is important in contributing to sustainable agriculture and maximizing farm productivity. Nonetheless, the maintenance of a high accuracy in recommending crops according to varied agro-climatic conditions is a challenging problem, encouraging researchers to explore a large set of machine learning algorithms and data-driven methods. The use of these technologies in agriculture offers an adaptable and scalable solution that can be tailored for different regional and environmental settings. Through ongoing model improvement and advance prediction methods, the system is able to keep pace with changing climate trends and soil behaviors—ultimately aiding farmers, agronomists, and policymakers in making education, location-based crop planning choices.

4.2Future Work:

We're looking to enhance the Smart Crop Recommendation System by integrating deep learning models and context-aware data representations. This could involve using transformer-based models or seasonal trend-based temporal models. By improving feature engineering—especially by adding satellite images, IoT sensor data, and geospatial information—we can boost the system's accuracy and adaptability. Plus, optimizing the model pipeline will help reduce computational overhead, enabling real-time deployment on larger, region-specific datasets without compromising responsiveness or scalability across various farming conditions.

5. References:

1. Pantazi, X. E., Moshou, D., & Tamouridou, A. A. (2016). Automated leaf disease detection in different crop species through image features and machine learning. *Computers and Electronics in Agriculture, 132, 141–148*

2. Jeong, H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., & Kim, S.-H. (2016). Random Forests for global and regional crop yield predictions. PLoS ONE, 11(6), e0156571.

3. Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, 621.

4 Pham, H. X., Nguyen, H. T., & Nguyen, D. T. (2021). A machine learning approach for crop recommendation using soil health and environmental conditions. *International Journal of Advanced Computer Science and Applications, 12(2), 409–415*

5. Saini, H. K., Duhan, M., & Khera, A. (2021). Smart agriculture using machine learning: A comprehensive review. *Materials Today: Proceedings, 46(14), 6750–6756*

6. Ghosh, S., & Sanyal, S. (2019). Application of machine learning in crop recommendation: A review. In *Journal of Emerging Technologies and Innovative Research, 6(6), 680–687.*

7. Jain, N., Singh, A., & Goyal, S. (2020). Soil fertility and crop yield prediction using machine learning algorithms. *Procedia Computer Science, 167, 528–536.*

8. Govindaraj, M., & Ramesh, S. (2020). A survey on crop recommendation using data mining and machine learning techniques. *International Journal of Engineering and Advanced Technology, 9*(3), 3862–3866.

9. Rani, S., & Arora, A. (2019). A machine learning approach for crop selection and yield prediction in India. *International Journal of Computer Applications, 178(16), 15–20.*

10. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18(8), 2674.*

11. Tripathi, R., & Maktedar, D. (2021). Soil analysis and crop recommendation system using decision tree. *International Research Journal of Engineering and Technology, 8(2), 404–408.*

12. Muthukannan, M., & Sakthivel, S. (2021). Crop prediction based on soil and weather parameters using ML algorithms. *International Journal of Recent Technology and Engineering, 9(6), 172–176.*

13. Mohanraj, I., Ashokumar, K., & Naren, J. (2016). Field monitoring and automation using IoT in agriculture domain. *Procedia Computer Science, 93, 931–936*

14. Kaur, H., & Kaur, P. (2021). Machine learning in agriculture: A review of current trends and future directions. *Materials Today: Proceedings, 45, 2891–2895*

15. Sharma, R., & Patel, S. (2020). Big data and machine learning in agriculture: A review. *Journal of Statistics and Management Systems, 23(1), 25–36.*