IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

AN INTELLIGENT DECISION SUPPORT SYSTEM FOR CROP RECOMMENDATION USING AI AND ML

[1]Swetha D S, [2]S Manjunatha,
[1]Post Graduation Student Master of Technology, Department of Computer Science and Engineering SJCIT, Chikaballapur, India.
[2]Assistant Professor, Department of Computer Science and Engineering, SJCIT, Chikaballapur, India.

Abstract: Agriculture is an important sector that supports food production and livelihood of millions of people. However, the choice of suitable crop is often difficult for the farmers due to variation in soil and weather from place to place. In order to solve this problem, in this paper, an intelligent decision support system for crop recommendation using machine learning techniques is proposed. The system uses the soil nutrient values like Nitrogen, Phosphorus, Potassium as input features along with temperature, humidity, pH and rainfall. We evaluate several machine learning algorithms, and Random Forest shows the best performance. The system will help farmers to take data based decisions, improve crop selection and increase productivity.

***Index Terms* -** Artificial Intelligence, Crop Recommendation, Decision Support System, Machine Learning, Precision Agriculture

I. INTRODUCTION

Agriculture is one of the most important sectors in the world, serving as the backbone of food security and economic stability. In developing countries like India, agriculture supports the livelihood of a large population and plays a crucial role in national growth. Despite its importance, farmers continue to face numerous challenges in selecting the right crop for cultivation. These challenges arise due to variations in soil characteristics, climatic conditions, water availability, and seasonal uncertainties. As a result, improper crop selection often leads to reduced yield, financial losses, and inefficient use of agricultural resources.

Traditionally, crop selection is based on farmers' experience, historical practices, and local knowledge. While this method has been used for generations, it lacks scientific accuracy and fails to adapt effectively to dynamic environmental changes. With the increasing impact of climate change and soil degradation, there is a growing need for more reliable and data-driven decision-making systems in agriculture.

In recent years, advancements in Artificial Intelligence (AI) and Machine Learning (ML) have opened new possibilities for solving complex agricultural problems. Machine learning techniques enable systems to learn from historical agricultural data and make accurate predictions based on input features such as soil nutrients (Nitrogen, Phosphorus, Potassium), temperature, humidity, rainfall, and pH level. By analyzing these parameters, ML models can identify patterns and recommend the most suitable crop for specific environmental conditions.

Crop recommendation systems using machine learning are an important step toward precision agriculture. These systems help farmers make informed decisions by providing data-driven insights, thereby improving crop productivity and reducing the risk of crop failure. Various supervised learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) are commonly used to build such predictive models due to their efficiency in classification tasks.

Furthermore, integrating machine learning into agriculture not only enhances productivity but also supports sustainable farming practices. It reduces unnecessary use of fertilizers and water by suggesting crops that are best suited for existing soil conditions. This contributes to environmental conservation and efficient resource management.

In this context, the proposed work aims to develop a machine learning-based crop recommendation system that assists farmers in selecting the most suitable crop based on environmental and soil parameters. The system leverages historical agricultural data to train predictive models and provides accurate crop suggestions, thereby bridging the gap between traditional farming methods and modern intelligent agriculture systems.

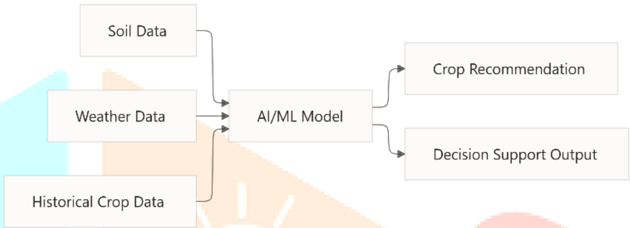

Fig.1 - Proposed Crop Recommendation System

In this paper, we present an Intelligent Decision Support System for crop recommendation to overcome the limitations of traditional crop selection methods. The proposed system uses Artificial Intelligence (AI) and Machine Learning (ML) techniques to analyze soil nutrient values and environmental factors to recommend the most suitable crop. By considering parameters such as Nitrogen, Phosphorus, Potassium, temperature, humidity, pH, and rainfall, the system provides reliable crop recommendations. The use of supervised machine learning algorithms helps farmers make better decisions based on available data, leading to improved crop productivity and encouraging sustainable farming practices.

II. RELATED WORK

Dahikar & Rode [1] introduced an ANN-based crop recommendation system that relies on three soil nutrient parameters to make recommendations (Nitrogen, Phosphorus and Potassium) and exhibited an improvement in accuracy compared to traditional expert systems. However, because their testing was limited to only 8 crop classes with limited training data, it restricts the generalizability of their results to different regions of the world.

Pudumalar et al. [2] implemented a Crop Recommendation System (CRS) using NAïve Bayes, Random Forest and Decision Tree classifiers with an ensemble majority voting technique to determine the vote 'winner.' In their case, evaluation of the system yielded an accuracy rate of over 97% for 11 crop classes, with the researchers concluding that ensemble-type methods are typically superior to using any of the individual classifiers for prediction in situations where the prediction involves classifying multiple classes into multiple groups or categories.

In a study conducted by Nevavathi and Jayasankar [3], the researchers used a hybrid of K-Nearest Neighbors and Support Vector Machine (SVM) for soil classification and crop predictions over the course of 10 different crops. The researchers were able to achieve 93.4% accuracy; however, the study required manual engineering of many features and thus has limited scalability if expanded to larger sets of crops beyond the 10 class limit.

Doshi et al. [4] developed an intelligent crop recommendation system called AgroConsultant which incorporates Random Forest with an API that provides weather information in real-time. The researchers were able to evaluate AgroConsultant using a test set of 12 different crop types with an accuracy rate of 95.7%. They attributed part of the high level of accuracy to the use of continuously updated environmental data but did raise concerns regarding the necessity of maintaining an internet connection to obtain real-time weather feed data in order to function properly.

Rajak et al. [5] intended to develop a dual-recommendation system using Naive Bayes and Logistic regression techniques to provide recommendations as to what crops should be planted based on the predictions made by the models and recommend what fertilizers should be used. The crop prediction section of this study reached an accuracy rate of 88.6% using 14 different crop types. The enhancement of fertilizer recommendations for the current study is a significant finding that the authors identify as an important area for future development.

Tanha et al. [6] have conducted a thorough benchmarking of boosting algorithm, including AdaBoost, Gradient Boosting, and XGBoost against multiple agricultural classification datasets. Their benchmark results demonstrate the strong sensitivity of AdaBoost to feature scaling and class distribution features, thus reflecting the low performance of AdaBoost (14.55%) noted above in the present study using a two-stage scale pipeline.

Gandge, [7] performed a comparison of Decision Tree, Naïve Bayes, Random Forest, and Support Vector Machine classifiers for agricultural crop prediction using the same publicly available Kaggle dataset. The study confirmed that Random Forest consistently ranks at the top two performance levels across all classification metrics which supports the model selection decision for the present study.

Vignesh et al. [8] have created and implemented a real-time crop and fertilizer recommendation system using Random Forest integrated with a mobile application front end developed in Android. Their work provides a clear demonstration of the ability of machine learning based agricultural recommendation tools to operate on mobile platforms with minimal latency; a deployment opportunity that has been identified as future work in this study.

Bhosale et al. [9] utilized Convolutional Neural Networks for crop leaf disease detection and yield prediction based on field images. While they achieved very high accuracy in terms of the use of images to assess the health of crops, their approach to deep learning consumed many more computational resources than would the classical machine learning methods that have been used in this study, thus making it less feasible for rural deployment in low-resource scenarios.

Kale et al. [10] also established a crop recommendation system that used XGBoost and Gradient Boosting for generating crop recommendations. Their recommendation system included the standard NPK inputs but also incorporated additional soil micronutrient information. Their crop recommendation system had an average accuracy of 96% across 18 crop types, indicating that the inclusion of extended soil impact data on recommendations beyond those using only NPK would yield greater specificity in the recommendations.

Patel and Shah [11] proposed a framework for crop recognition by using a Support Vector Machine with a Radial Basis Function kernel to identify the top "n" crops that are most suitable for the area based on crop yield predictions rather than have the classification of the crop result in one crop. This enhancement to the method of multi-label SVM classification would represent an improvement over single-label classification and would be the direction that the current system should take for future improvements.

Machine learning applications have been reviewed in the context of various precision agriculture domains such as crop yield prediction, irrigation management, soil health monitoring, and pest detection by Sharma et al.[12]. Based on this review, it was concluded that for all types of agricultural classification from a variety of data and locations, ensemble classifiers (Random Forest and Gradient Boosting in particular) are much more accurate than linear models and shallow classifiers.

Chlingaryan et al. [13] provided an overview of machine learning techniques for crop yield prediction and estimating nitrogen status. Their review included more than 30 studies and demonstrated that the most important features in a wide array of crop yield prediction models are rainfall, humidity, and soil macronutrients, all of which were also identified as being among the most important features in the present study, with rainfall and humidity being ranked first and second, respectively.

A systematic review was performed by Liakos et al. [14] on the use of machine learning in the field of agriculture encompassing 40 studies focused on crop yield prediction, disease detection, soil management, and irrigation. They found that data quality, feature selection, and model interpretability are the three most important aspects that determine how applicable machine learning will be in 'real' or practical agricultural settings. In designing the proposed system, all three of these aspects will be accounted for through the use of a high-quality, balanced dataset; creating a seven-feature input set based on agronomic domain knowledge to validate the input features; and using the intrinsic feature importance scoring of Random Forests for interpretability.

III. DESIGN AND METHODOLOGY

A. Design

Crop Recommendation System has a three tier / layer architecture with the input layer being comprised of data being entered by the user via an html form containing the seven agronomic parameters required to run the model. The processing layer of the architecture uses preliminary and primary feature scaling on the input parameters resulting in a modified set of input parameters that are utilized to classify the crops using a Random Forest classifier that has been previously trained on the modified data. The output layer of the system will produce the recommended crop name and associated crop image via the Flask web service interface built to provide recommendations to the user. There is 7 parameters that need to be input into the system, which are Nitrogen (N), Phosphorus (P), Potassium (K), Temperature (°C), Humidity (%), pH, and Rainfall (mm), which will then be classified by the classifier producing one of the 22 recommended crop classes.

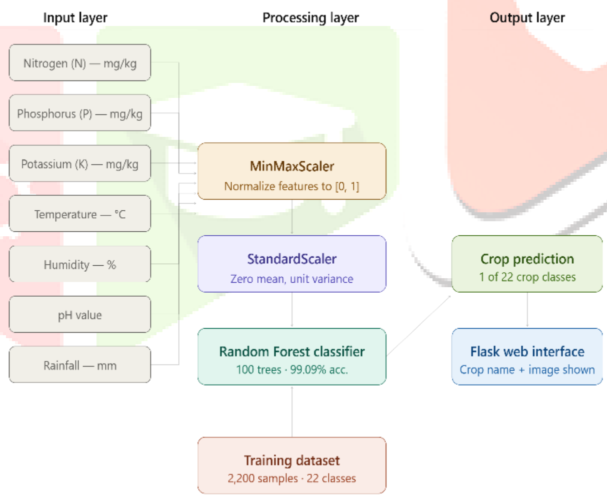

Fig.2 - System Architecture Diagram

B. Methodology

The six phases of the development process were structured and developed in sequence from the completion of the preceding phase (see Fig. 3 for the development process). There is a logical progression through each of the six phases, creating a development process that is systematic and reproducible. The six phases of the development process are as follows: 1. Requirements Analysis Phase; 2. Data Collection Phase, 3. Data Processing Phase, 4. Model Training Phase; 5. Evaluation Phase; 6. Deployment Phase.

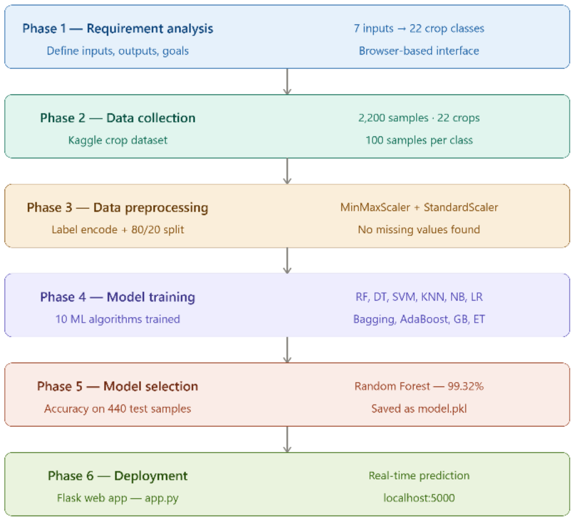

Fig.3 – Six-Phase Methodology

Phase 1 - Requirement Analysis

Seven input parameter inputs were needed to characterise soil and climate conditions, recommending the most appropriate crops from among 22 class types. A machine learning model was specified, created using Python, and the specification required a web-based interface that could be operated on a user's web browser without requiring the user to install software on their local machine.

Phase 2 - Collecting Data

The Kaggle Crop Recommendation Dataset is used as our dataset for this research project. The dataset contains 2200 rows of data for all 22 crop-class categories, with exactly 100 rows for each crop-class. It consists of seven numeric input variables and contained no missing values or duplicate records (See Table I)

TABLE I - STATISTICAL SUMMARY OF DATASET FEATURES

Feature	**Min**	**Max**	**Mean**	**Std. Dev.**
Nitrogen (N)	0.00	140.00	50.55	36.92
Phosphorus (P)	5.00	145.00	53.36	32.99
Potassium (K)	5.00	205.00	48.15	50.65
Temperature (°C)	8.83	43.68	25.62	5.06
Humidity (%)	14.26	99.98	71.48	22.26
pH	3.50	9.94	6.47	0.77
Rainfall (mm)	20.21	298.56	103.46	54.96

E. Phase 3 - Data Preprocessing

Three operations were performed. First, the 22 crop names in the target column were mapped to integer labels from 1 to 22 through a manually defined dictionary. Second, the data was split into 80% training (1,760 samples) and 20% testing (440 samples) using a fixed random state of 42. Third, a two-stage scaling pipeline was employed: all features were scaled to [0, 1] using MinMaxScaler and standardised to have zero mean and unit variance using StandardScaler. To prevent data leakage, both scalers were fitted only on the training set and saved as minmaxscaler.pkl and standscaler.pkl for deployment.

F. Phase 4 - Model Training

Ten classifiers were trained on the scaled training data using Scikit-learn default hyperparameters. The models used are Logistic Regression, Naïve Bayes, Support Vector Machine, K-Nearest Neighbours, Decision Tree, Extra Trees, Random Forest, Bagging, AdaBoost, and Gradient Boosting. All models were trained under the same conditions to allow a fair comparison.

G. Phase 5 - Model Selection

The main metric used to evaluate the performance of the trained classifiers was the prediction accuracy on the test set of 440 samples. The results of the evaluation show that the Random Forest classifier has the highest accuracy of 99.09% compared to all the models. Thus, the Random Forest model was selected as the final prediction model for the proposed system. The choice was also motivated by the model's ability to operate without imposing distributional assumptions on the input data, its ability to efficiently handle complex and non-linear feature interactions, and its provision of feature importance scores that increase the interpretability of the model.

H. Phase 6 - Deployment

The model and scalers were wrapped in a Flask web application within app.py. When the form is submitted, the application takes the 7 input values, scales them with MinMaxScaler and StandardScaler, feeds the scaled data to the Random Forest model, maps the predicted integer to the crop name via the crop dictionary, and displays the result along with a crop image on index.html.

IV. RESULTS AND DISCUSSION

The implemented Crop Recommendation System was evaluated utilizing agricultural data that included soil nutrient information and climate data. Seven user input parameters were included, comprising Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH value, and rainfall. The system is able to provide a recommendation for a crop to be planted based on the output of a trained machine learning model.

Three specific classifiers - Decision Tree, K-Nearest Neighbors (KNN), and Random Forest - were evaluated to determine the best algorithm to use for the system. The evaluation of each classifier was based on the accuracy of prediction. As shown in Figure 4, the Random Forest classifier had the best accuracy of all three classifiers with an accuracy of 99.09%, followed by KNN with 96.8% and Decision Tree with 95.4%.

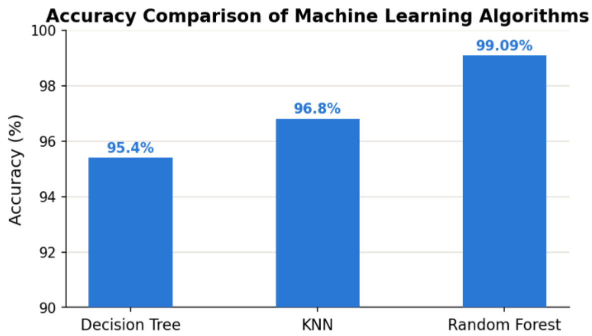

Fig.4 - Accuracy Comparison of Machine Learning Algorithms

Random Forest is used to make crops selections using their iterative processes, where it creates trees, collects votes, and chooses majority based on the trees built. Through this approach, Random Forest reduces overfitting when using classification compared to other algorithms, thus being the most accurate classification algorithm used in the proposed system. Therefore, Random Forest will be used as the main algorithm in the proposed system. In addition, the Random Forest model is wrapped within a web application using Flask, giving it an easy-to-use web interface. The web application has a functionality allowing users to submit environmental, soil, and nutrient information to get crop recommendations via the web application in real-time. For instance, if a user submits farming data for a soil nutrient value and other farming conditions, the web application shows the crop recommendation in real-time. The accurate crop recommended will be displayed as part of the system, as shown in Figure 5, during the provided sample's parameters; the crop recommended was muskmelon.

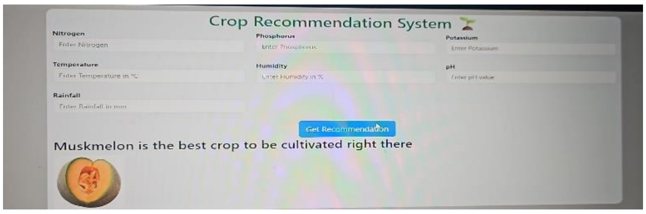

Fig.5 - Crop Recommendation Result Generated by the Proposed System

Results from the conducted experiments confirm that the system is capable of providing well-suited crop recommendations for the given soil and environmental conditions, and based on the recorded accuracy of 99.09%, confirms that the use of machine learning in this study is a valid and reliable methodology. The system's web-based platform allows for easy access by agricultural professionals and farmers, with no advanced technical ability required to operate the system in order to provide farmers with the opportunity to make data-driven and educated crop-selection decisions.

The Crop Recommendation System will have a crucial role in changing how farmers make decisions regarding the most suitable crops to grow, the most efficient use of nutrient resources, and ultimately, how to maximise agricultural production on farmlands. Additionally, the intelligent, reliable, and supportive decision-making capabilities of the system will contribute to the advancement of information-based smart and sustainable farming practices.

V. CONCLUSION

The presented work describes an artificial intelligence (AI) and machine learning (ML) based "Intelligent Decision Support System" (IDSS) for crop recommendations. This system considers the following key agricultural parameters; nitrogen (N), phosphorus (P), potassium (K), pH, temperature, humidity, and rainfall. The IDSS uses ML to generate reliable and accurate agronomic recommendations to support data-driven decision-making in agronomy. Among the different ML models investigated, the Random Forest model generated the highest accuracy with a predicted accuracy of 99.09% for all assessed soil and environmental growing factors. Therefore, the IDSS will enable farmers to make more informed choices about crops grown in specific soil/environmental conditions, while also improving their productivity and resource use efficiency while decreasing risk associated with the cultivation process. To facilitate wider use and practical applications, the trained Random Forest model has been developed into a simple to use web-based application. The application allows users to input various agricultural parameters and receive immediate crop recommendations. As such, this application will be a powerful tool to aid farmers and agronomists. Although the IDSS generated satisfactory predictions, there remains some opportunity for refinement and enhancements. One possible enhancement would be to add real-time weather data and utilise Internet of Things (IoT) based soil monitoring sensors as well as localised agriculture data to further refine prediction accuracy. Also, exploring use of additional advanced ML and deep learning methodologies could enhance the performance and transferability of the IDSS across varying geographical areas. Enhancements to the IDSS will provide developed countries with prime opportunities to continue the transition towards sustainable agriculture through further development of sustainable agriculture.

REFERENCES

[1] S. S. Dahikar and S. V. Rode, "Agricultural crop selection and recommendation using Artificial Neural Network approach," International Journal of Innovative and Emerging Research in Engineering, vol. 1, no. 2, pp. 53-57, 2014.

[2] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in Proc. 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2017, pp. 1-6.

[3] M. Nevavathi and T. Jayasankar, "Recommendation of suitable crop for agriculture using machine learning approaches," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 5, pp. 3139-3143, 2020.

[4] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "AgroConsultant: Intelligent crop recommendation system using machine learning algorithms," in Proc. 4th International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6.

[5] R. K. Rajak, A. Pawar, M. Pendke, P. Shahu, S. Suryavanshi, and A. Devare, "Crop recommendation system to maximize crop yield using machine learning technique," International Research Journal of Engineering and Technology (IRJET), vol. 4, no. 12, pp. 950-953, 2017.

[6] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, "Boosting methods for multi-class imbalanced data classification: An experimental review," Journal of Big Data, vol. 7, no. 1, pp. 1-47, 2020.

[7] Y. Gandge, "A study on various machine learning techniques for crop prediction," in Proc. International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, India, 2017, pp. 1-4.

[8] T. Vignesh, S. Aswin, S. Akash, and C. Sundar, "Crop recommendation for farmers using machine learning and IoT," in Proc. International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, France, 2021, pp. 1-5.

[9] S. V. Bhosale, R. A. Thombare, P. G. Dhemey, and A. N. Chaudhari, "Crop leaf disease detection and classification using machine learning and deep learning algorithms," in Proc. International Conference on Artificial Intelligence and Data Engineering (AIDE), Manipal, India, 2019, pp. 1-7.

[10] A. Kale, A. Aware, S. Baviskar, and C. Patil, "Crop recommendation system using XGBoost and Gradient Boosting for sustainable agriculture," International Journal of Engineering Research and Technology (IJERT), vol. 10, no. 5, pp. 482-488, 2021.

[11] H. Patel and D. Shah, "A machine learning-based crop recommendation system using SVM and multi-label classification," in Proc. 2nd International Conference on Data Science and Business Analytics (ICDSBA), Changsha, China, 2018, pp. 137 141.

[12] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," IEEE Access, vol. 9, pp. 4843-4873, 2021.

[13] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," Computers and Electronics in Agriculture, vol. 151, pp. 61-69, 2018.

[14] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, p. 2674, 2018.