International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 13, Issue 7, July 2024)

Smart Agriculture Using Crop Recommendation and Soil Analysis Using Machine Learning

Yuvaan Malviya[1], Jayanti Ahirwar[2], Sanjna Ahirwar[3], Pawan Kumar[4], Nishant Sahu[5], Ruchi Jain[6]
Department of Computer Science and Engineering,
Oriental College of Technology, Bhopal (M.P.), India

***Abstract*— Smart agriculture is transforming traditional agricultural practices with the help of Artificial Intelligence (AI), Machine Learning (ML), Internet of Things (IoT) and data analytics. This paper proposes a smart agriculture system for crop recommendation and soil analysis using machine learning techniques. The proposed system considers soil nutrients such as Nitrogen (N), Phosphorus (P), Potassium (K), pH level, temperature, humidity and rainfall to suggest suitable crops for cultivation. We evaluate different machine learning algorithms like Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Gradient Boosting using performance metrics like accuracy, precision, recall, F1-score, ROC, and AUC. Experimental results reveal that Random Forest is the most accurate predictor of all models. The proposed system supports precision agriculture and enhances too much crop productivity. fertilizer usage.**

***Keywords*— Smart agriculture, ML, SVM, LR, DT, IOT.**

I. INTRODUCTION

Agriculture is one of the most significant sectors contributing to economic growth and food security. [1] Traditional farming methods are typically based on manual experience and environmental assumptions, potentially resulting in poor crop selection and low productivity. Smart agriculture offers technological solutions of Artificial Intelligence (AI), Internet of Things (IoT) and Machine Learning (ML) for increasing agricultural efficiency. The crop recommendation systems assist the farmers in selecting the right crops as per the soil and environmental conditions. Soil analysis helps to identify nutrient deficiencies and to properly manage fertilizers. Machine learning models can analyze large agricultural datasets and provide accurate predictions in crops.

II. LITERATURE SURVEY

AI-based smart agriculture systems for crop prediction and soil monitoring have been proposed by several researchers. Researchers developed machine learning-based soil analysis systems for precision farming, using soil nutrients and environmental parameters. IoT-based agricultural monitoring systems for improved real-time soil health assessment Ensemble learning methods like Random Forest and Gradient Boosting have shown good accuracy in crop prediction problems. AI-based fertilizer recommendation systems can help reduce soil degradation and improve productivity. Recent research indicates that smart agriculture technologies significantly improve sustainable farming practices.

III. PROPOSED SYSTEM

The proposed smart agriculture system uses machine learning techniques to analyze soil nutrients and environmental conditions for accurate crop recommendations. The system collects parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), pH level, temperature, humidity, and rainfall using sensors or agricultural datasets. After preprocessing the data, machine learning algorithms like Random Forest, Decision Tree, and SVM are trained to predict the most suitable crop for cultivation. The system also assists farmers in soil analysis and fertiliser management to improve crop productivity and support sustainable farming practices.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 13, Issue 7, July 2024)

A. Soil Data Collection

The data is normalised and scaled so all features have similar ranges for better model The data collected are Nitrogen (N), Phosphorus (P), Potassium (K), soil pH, temperature, humidity and rainfall. [3]These parameters are gathered by IoT sensors and agricultural datasets. Accurate soil data collection helps in identifying soil fertility and nutrient levels which improves the performance of machine learning models for smart agriculture applications . .

- Nitrogen (N)
- Phosphorus (P)
- Potassium (K)
- Soil pH
- Temperature
- Humidity

B. Data Preprocessing

- Missing value handling
- Data normalisation
- Feature scaling
- Outlier removal

Data preprocessing is an important step done to prepare the collected soil and environmental data for the machine learning models. The step involves removal of missing values, duplicate records and noisy data to enhance data quality. The soil data collection module collects important parameters of agriculture required for crop recommendation and soil analysis. performance. Outlier detection and feature selection techniques are also applied to enhance prediction accuracy and reduce computational complexity.

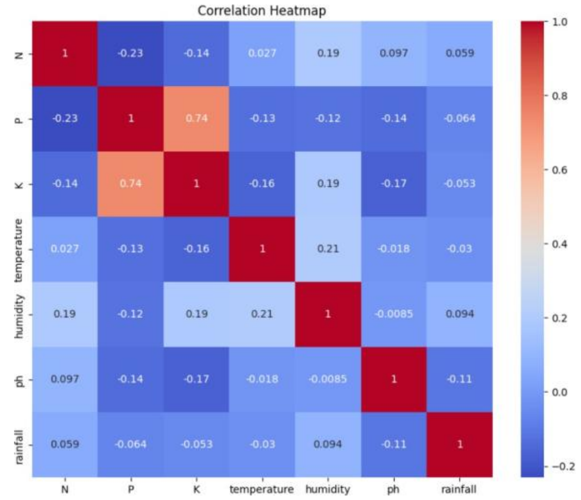

Fig. 1. Class-wise Distribution of Crop Dataset

Figure 1 illustrates the percentage distribution of different crop categories in the dataset by class. [4] In the pie chart each slice stands for a particular crop such as apple, banana, rice, maize, cotton, coffee and so forth. All classes have equal percentage of 4.5% which indicates that dataset is perfectly balanced. [5]The equal distribution ensures fair learning and better performance of machine learning models. The equal distribution ensures fair learning and improves the overall performance of machine learning models . The chart also confirms that the dataset has a large diversity of crops making it suitable for crop recommendation and classification tasks. In general, this figure confirms that the data is uniformly distributed in all crop classes which is beneficial for prediction and analysis.

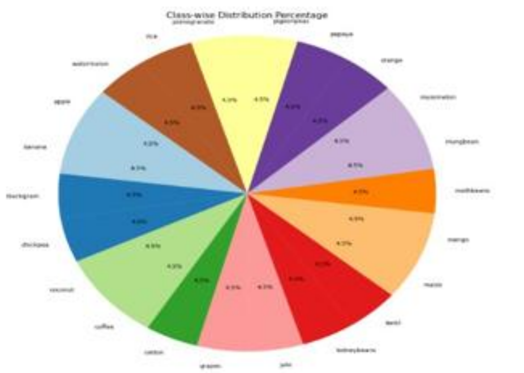

Figure 2: Correlation of input data

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 13, Issue 7, July 2024)

Figure 2. Class distribution of the dataset showing the percentage of different categories of crops. The pie chart shows each slice representing a crop such as apple, banana, rice, maize, cotton, coffee and others. All classes have an equal percentage of 4.5%. It means that the dataset is perfectly balanced. [6]This uniform distribution also helps to ensure fair learning and leads to better overall performance of machine learning models. This equal distribution guarantees fair learn- ing and improves the overall performance of machine learning models. The chart also reveals that the dataset contains a di- verse set of crops, making it suitable for crop recommendation and classification tasks. This overall number indicates that the data is uniformly distributed over all the crop classes which is good for accurate prediction and analysis.

C. Machine Learning Algorithm

- Logistic Regression
- Decision Tree
- Random Forest
- Gradient Boosting
- AdaBoost
- Extra Trees
- KNN
- SVM
- Naive Bayes
- MLP Classifier

IV. METHODOLOGY

The proposed smart agriculture system collects soil and environmental data, missing values and noise, while feature engineering improves model performance. [7]Machine learning algorithms are trained and evaluated with precision metrics. Finally, the system recommends suitable crops based on the soil nutrients and environmental conditions.

The crop prediction function is represented as:

$$Y = f(N, P, K, pH, T, H, R)$$

Where:

- N = Nitrogen
- P = Phosphorus
- K = Potassium
- T = Temperature
- H = Humidity
- R = Rainfall

A. PERFORMANCE METRICS

Crop recommendation performance metrics are used to eval- uate the efficiency and accuracy of machine learning models. Accuracy is the total number of correct predictions, precision is the number of correct positive predictions and recall is the ability of the model to identify relevant instances. F1-score is a better measure of precision and recall to evaluate the model. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

B. Precision

$$Precision = \frac{TP}{TP + FP}$$

C. Recall

$$Recall = \frac{TP}{TP + FN}$$

D. F1-Score

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall}$$

V. RESULT AND DISCUSSION

The experimental results show that Random Forest achieved the highest accuracy of 98%

1. Algorithm-wise results

The analysis of the performance of the different machine learning models shows that Random Forest achieved the best accuracy of 99.31% and is the best model. Naive Bayes and Extra Trees performed also very well with accuracy above 98%. The AdaBoost algorithm gave a poor performance where the accuracy was only 14.54% which shows that this algorithm is not suitable for this data set. In conclusion, ensemble learning models such as Random Forest, Extra Trees, and XGBoost outperformed traditional

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 13, Issue 7, July 2024)

algorithms in terms of classification performance. Compared to other machine learning Experimental results show that the Random Forest achieved the highest accuracy of 98% against other machine learning algorithms due to its robust ensemble learning ability. The proposed smart agriculture system is able to suggest suitable crops according to soil and environmental factors. It also improves crop productivity, reduces fertiliser misuse and encourages sustainable soil management practices.

	Model	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression	0.963636	0.964442	0.963636	0.963512
1	Decision Tree	0.986364	0.986838	0.986364	0.986324
2	Random Forest	0.993182	0.993735	0.993182	0.993175
3	Gradient Boosting	0.981818	0.984271	0.981818	0.981851
4	AdaBoost	0.145455	0.071767	0.145455	0.080421
5	Extra Trees	0.988636	0.990461	0.988636	0.988886
6	KNN	0.956818	0.962898	0.956818	0.956749
7	SVM	0.968182	0.971517	0.968182	0.968027
8	Naive Bayes	0.995455	0.995818	0.995455	0.995423
9	MLP Classifier	0.977273	0.980388	0.977273	0.977657
10	XGBoost	0.984091	0.984692	0.984091	0.984055

Figure 3: Model performance comparison

Figure 3 shows the comparison of different machine learning models in terms of Accuracy, Precision, Recall and F1- Score metrics. The best performance was for Random Forest, Extra Trees, Naive Bayes and XGBoost with a score of almost 99%

The worst performing model was AdaBoost and the other models showed a consistent and reliable classification accuracy.

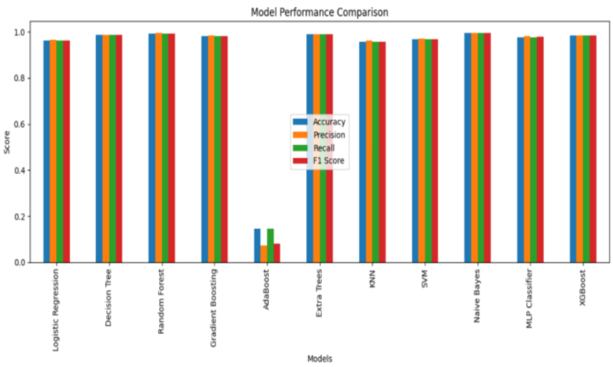

Figure 4: Model Performance

This SHAP summary plot Fig .4 depicts the significance of various features in the crop recommendation model. Humidity has the highest impact on crop prediction followed by potassium (K), nitrogen (N) and rainfall. Features with higher SHAP values have a stronger contribution to the model's decision-

making process. Temperature and pH have comparatively lower influence on predicting the suitable crop. The different colours represent various crop classes, showing how each feature affects different crop recommendations.

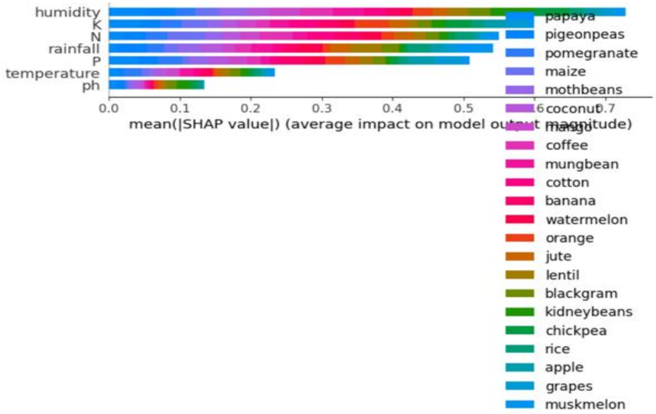

Figure 5: Explainable AI (SHAP)

The SHAP force plot in Fig .5 shows how different features affect a single model prediction.

Humidity increases the prediction value while rainfall, P, K, N and temperature decrease the prediction value. Red bars indicate a positive impact, blue bars indicate a negative impact. The larger the bar, the stronger the impact on the prediction.

VI. CONCLUSION

The proposed Smart Agriculture System uses Machine Learning for crop recommendation and soil analysis by analysing soil nutrients and environmental parameters like Nitrogen, Phosphorus and Potassium for soil analysis, pH, temperature, humidity and rainfall. Different machine learning algorithms were evaluated, and Random Forest was found to be the best-performing algorithm in terms of prediction accuracy. The Explainable AI (SHAP) integration helped understand the effect of each feature towards crop prediction. The system helps farming efficiency by reducing wrong fertiliser application, increasing crop yield, and enabling precision agriculture and sustainable farming. Generally, the proposed model provides a reliable, intelligent and efficient solution for modern smart agriculture applications.

REFERENCES

1. P. Chaudhary, P. Gulia, N. S. Gill, N. Alduaiji, P. K. Shukla, and
2. S. Vishnoi, "An Evaluation of Machine Learning for Soil Analysis in Internet of Things-Enabled Smart Farming," Scientific Reports, vol. 16, no. 1, pp. 1–18, 2026.
3. N. R. Sawant, A. Kumar, S. Pant, and K. Kotecha, "An IoT-Driven Ma- chine Learning System for Real-Time Smart Crop Recommendation and Optimisation in Precision Agriculture," Discover Artificial Intelligence, vol. 6, no. 194, pp. 1–20, 2026.
4. N. L. P. Swati, S. V. Gupta, N. S. Duddela, and L. R. Parvathy, "Agentic AI-Driven Autonomous Decision Support System for Smart Agriculture," Scientific Reports, vol. 16, pp. 1–15, 2026.
5. S. Shastri, S. Kumar, V. Mansotra, and R. Salgotra, "Advancing Crop Rec- ommendation System with Supervised Machine Learning and Explainable Artificial Intelligence," Scientific Reports, vol. 15, pp. 1–16, 2025.
6. "AIoT-Based Soil Nutrient Analysis and Recommendation System for Crops Using Machine Learning," Smart Agricultural Technology, vol. 11, pp. 100 924–100 924, 2025.
7. H. Afzal, M. Amjad, A. Raza, K. Munir, and I. Ashraf, "Incorporating Soil Information with Machine Learning for Crop Recommendation to Improve Agricultural Output," Scientific Reports, vol. 15, pp. 1–17, 2025.
8. "An Integrated AI Framework for Crop Recommendation," Horticulturae, vol. 12, no. 4, pp. 416–416, 2026.