A Smart Prediction Platform for Agricultural Crops Using Machine Learning

Ammar Rafiq[1], Muhammad Usman Younus[2,3], Syed Aqeel Haider[4], Kalsoom Safdar[5,6], Najaf Ali[7], Nusrat Husain[8], Muhammad Asgher Nadeem[9], Zeeshan Ahmad Arfeen[10], Faisal Mumtaz[5]

[1]Department of Computer Sciences, NFC-Institute of Engineering and Fertilizers Research Faisalabad, Pakistan.

[2]Department of Computer Science and IT, Baba Guru Nanak University, Nankana Sahib, Pakistan.

[3]Ecole Math´ematiques, Informatique, T´el´ecommunications de Toulouse, 31000 Universit´e de Toulouse, France.

[4]Department of Computer & Information Systems Engineering, Faculty of Computer & Electrical Engineering, N.E.D. University of Engineering and Technology, Karachi 75270, Pakistan.

[5]Department of Computer Science and IT, University of Jhang, Jhang, Pakistan.

[6]Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis (UniMAP), 02600 Perlis, Malaysia

[7]School of Computer Science and Technology. University of Science and Technology of China.

[8]Department of Electronics & Power Engineering, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan.

[9]Department of Computer Science and IT, Thal University Bhakhar, Punjab – Pakistan.

[10]Department of Electrical Engineering, The Islamia University of Bahawalpur (IUB), Bahawalpur - 63100, Pakistan.

*** Correspondence:** Zeeshan Ahmad Arfeen; zeeshan.arfeen@iub.edu.pk

Citation | Rafiq. A, Younus. M. U, Haider. S. A, Safdar. K, Ali. N, Husain. N, Nadeem. M. A, Arfeen. Z. A, Mumtaz. F, "A Smart Prediction Platform for Agricultural Crops Using Machine Learning", IJIST, Vol. 07 Issue. 01 pp 98-110, Jan 2025

Received | Dec 19, 2024 **Revised |** Jan 13, 2025 **Accepted |** Jan 15, 2025 **Published |** Jan 17, 2025.

Abstract.

It is very critical to have the economic development of emerging countries, like Pakistan. Pakistan, while being one of the world's main suppliers of a wide range of commodities, continues to employ traditional techniques. Pakistani farmers have challenges not just in coping with changing climatic circumstances, but also in meeting increased demands for higher food output of excellent quality. Farmers must be mindful of shifting meteorological circumstances to produce quality crops. Operations are greatly affected by a variety of factors, including the availability of water, the type of soil, the climate, and fertilizer. Farmers in conventional farming must decide on all of these aspects. What to grow, how to use the irrigation schedule, and the kinds of fertilizer are all covered in this event. Decisions made by farmers are primarily dependent on their experience, which can lead to the waste of expensive resources like water, fertilizers, time, effort, etc. Additionally, cultivating crops that are not the best fit for a given soil type and climate by using standard farming methods might cause problems, which can reduce production and profit. The application of machine learning in crop prediction is very widespread. The most popular method is irrigation. The major goal of this paper is to efficiently develop an E-business online platform to enhance farmers' productivity and circulation cycle. In this paper, we develop a platform for smart crop predictions. The platform will help farmers by assisting them in obtaining suggestions based on several metrics like humidity, temperature, pH, moisture, and rainfall. Additionally, the user of our platform will be able to get precise advice about what crop to plant depending on variables like humidity, pH, and other characteristics. The user will also be able to get connected with the buyers of their crops and efficiently meet their requirements.

Abbreviations.

Precision Prediction (PP)
Internet of Things (IoT)
Genetic Techniques (GA)
Convolutional Neural Networks (CNNs)
Recurrent Neural Networks (RNNs)

Keywords: Irrigation, fertilizer, economic, climate, kaggle, and soil.

Introduction:

Agriculture is a cornerstone of global food security and economic stability, particularly in developing regions where efficient resource management is critical. Despite the advancements in technology, farmers and agricultural stakeholders often face challenges in predicting crop yields, assessing risk factors, and making timely decisions due to limited access to accurate and actionable data. Existing solutions frequently lack scalability, adaptability, and the capability to deliver precise predictions under diverse environmental conditions. This study introduces a Smart Prediction Platform for Agricultural Crops, leveraging cutting-edge machine-learning techniques to address these challenges. The platform is distinguished by its integration of ensemble learning models and a live training mechanism, which together ensure robust, real-time adaptability to dynamic agricultural scenarios. Ensemble learning combines the strengths of multiple algorithms to enhance prediction accuracy and reliability, while the live training component allows the platform to continuously learn from incoming data, ensuring its relevance and effectiveness in real-world applications.

To ensure that crops and soil receive the exact nutrients, they require for optimum health and productivity, Precision Prediction (PP) for crops is used [1]. PP is a method of farm management, using machine learning techniques [2]. PP aims to guarantee profitability, sustainability, and environmental preservation. According to the World Bank, if the current rate of global position rise is maintained, food consumption will jump by 50% by 2050. Crop yield has decreased by more than a quart because of dramatic weather changes. To produce high-quality crops in large quantities, there needs to be a focus on the application of machine learning techniques in the field of crop prediction [3].

Precision farming, often known as AI-based crop prediction, is a practical and efficient way to guarantee a sustainable food supply considering the world's constantly expanding population [4]. Machine learning may undoubtedly assist in reducing costs and aid in expanding production. AI-based crop prediction aims to increase worker productivity, increase yield, and improve the quality of crop prediction output [5]. Today's built-smart crop prediction employs a variety of innovative technology on farms, including sensors, GPS, big data analysis, the IoT, machine learning, and robots [6]. The main goals of smart crop prediction are to reduce human effort, increase food self-sufficiency, and make optimal use of available resources [7]. In order to achieve high quality and mass production of crops, we need to focus on implementing intelligent technology in the crop prediction field [7], [8]. Machine learning can certainly help reduce costs and scale production [6]. There are certain factors that play an important role in plant production. Almost 51% of crop yields depend on the influence of these factors. These factors include nitrogen, phosphorus, potassium, temperature, humidity, and pH concentration [9], [10]. Machine learning is popular in various fields such as for instance culture and natural Resources. In horticulture and IoT systems, many machine-learning models have been developed [11]. Right now, there are several works in smart crop prediction based on many machine-learning models that have been developed [12]. There are several works in smart crops prediction based on machine learning as marked in Figure 1.

A smart irrigation system uses machine learning decision trees to predict crop water demand [13]. Humidity and temperature are the most important parameters for determining the amount of water required for crop prediction. There are two types of decisions: yes and no [14].

a) If the decision tree predicts a yes result, a warning is sent to the farmer.

b) If the technique predicts the result as no, a warning is sent to the farmer.

This is a similar study in which authors improved the performance of the DT model in relation to irrigation management, a hybrid approach involving the integration of DT and Genetic Techniques (GA) was proposed to provide the optimal decision tree model for predicting irrigation schedules was Implemented [14], [15]. The irrigation schedule event takes the form of a binary classification problem leading to a decision to irrigate or not [16]. A KNN model that will calculate the characteristics and reveal the crop which is perfect for the system drawing. The bitmap that a definite region instantly [17], [18]. Environmental parameters like soil type, rain, humidity, etc. are collected and crop prediction is completed around with all the accuracy for your crops done utilizing the KNN [19].

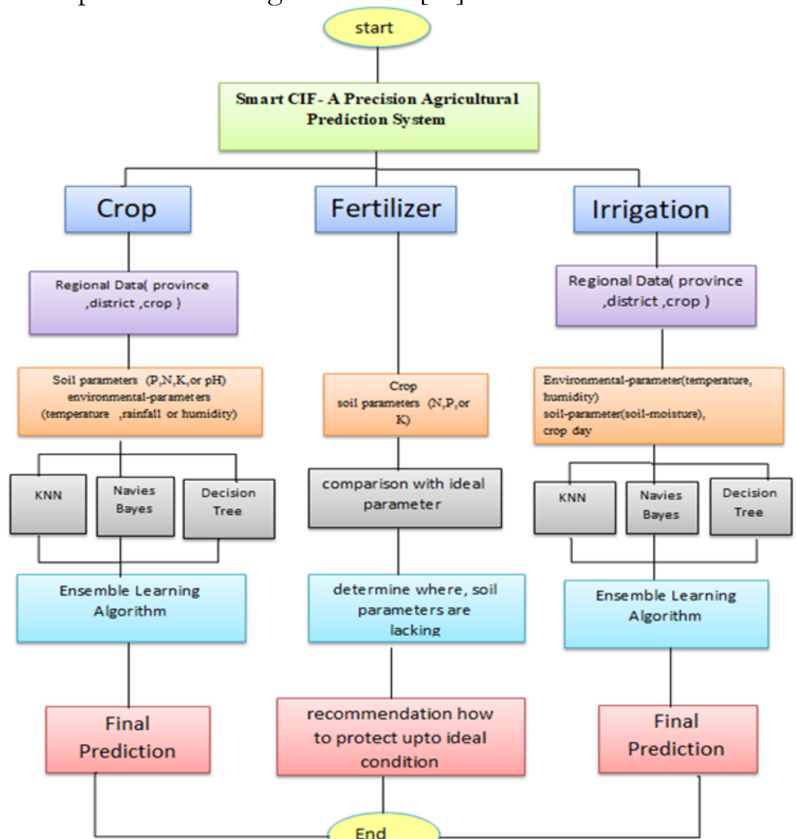

Figure 1. Working flow diagram of smart irrigation system.

Novelty:

The novelty of this platform lies in its ability to offer highly precise predictions tailored to regional and temporal agricultural needs, setting it apart from conventional static models. By focusing on ensemble learning and live training, this research provides a scalable and adaptive solution for stakeholders, empowering them to optimize agricultural practices and mitigate risks effectively. The focus is on proposing IoT-based smart farming systems that help farmers get recommendations based on various factors such as humidity, temperature, pH, humidity, and rainfall. The system also focuses on suggesting fertilizers to farmers based on factors such as nitrogen, phosphorus, and potassium levels in the soil. Various machine learning techniques such as Decision Trees, Naive Bayes, Support Vector Machines, Logistic Regression, Random Forest, and XGBoost were applied to the training data set and compared based on model accuracy. XGBoost was used for the prediction model as it showed the highest accuracy [20].

Objectives:

The objective of this study is to design an accessible and scalable prediction platform for crops, irrigation, and fertilizer recommendations. This platform aims to improve agricultural productivity and resource utilization through real-time, accurate predictions tailored to individual user needs. The novelty of this study lies in its integration of ensemble learning techniques with a live training feature, offering a practical and adaptable solution for farmers in diverse environments.

Literature Review:

The researcher used an ensemble learning model to predict the watering of crops based on a crop prediction IoT system. About four models, including linear SVR, Ad boost DT, and RF, which support linear regression, were trained to evaluate the performance of the intelligent irrigation system. An IoT framework was implemented along with a platform and mobile application to enable the model to be used for real-time irrigation planning [21]. Another framework provides an ensemble learning irrigation model based on crop prediction IoT systems. In the context of Pakistan's agriculture, the use of machine learning and IoT systems has shown significant potential in addressing challenges such as low crop yields, inefficient resource utilization, and climate unpredictability. For instance, Ahmed et al. highlighted the role of IoT-based precision farming in improving irrigation practices in Pakistan, emphasizing the integration of sensor data to optimize water usage [24]. Similarly, Rafiq and Younus demonstrated the effectiveness of an automated irrigation system using Fi-WSN in enhancing water efficiency and crop quality in arid regions [25].

Recent studies also indicate the growing adoption of machine learning in crop prediction. Abbas et al. developed a crop yield estimation model tailored to the climatic conditions of Punjab, Pakistan, which accounted for variables such as rainfall, temperature, and soil composition [24]. These region-specific insights underscore the importance of localized machine-learning models that address the unique challenges faced by Pakistani farmers. Despite these advancements, there remains a lack of scalable platforms that integrate real-time user input with adaptive machine-learning models. This research builds on existing studies by providing a platform designed to meet the specific needs of Pakistan's agricultural sector while leveraging cutting-edge ensemble learning techniques and live training capabilities. They mixed regression and classification techniques and applied them to stack and boost different procedures. The proposed model achieves an accuracy of 94.27 [22]. Table 1 portrays a comparison of the most relevant papers reviewed in this study.

Table 1. Comparison of previous studies on crop prediction platforms**.**

Study	Methods Used	Strengths	Limitations
[2]	Decision Trees, Genetic Algorithm	Improved accuracy in irrigation scheduling	Limited to binary classification
Research	Linear SVR, AdaBoost, Random Forest	High performance in real-time irrigation planning	High computational cost
[14]	Optimized Machine Learning Techniques	Enhanced crop quality through parameter tuning	Limited scalability for larger datasets
[17]	Deep Learning	Captures complex patterns in agricultural data	Requires large labeled datasets
[23]	IoT-Enabled Machine Learning	Integrates IoT and ML for crop recommendations	Dependency on IoT infrastructure

In the section above, it has been discussed in the study that they mixed regression and classification techniques and then reservations to different techniques for better results in irrigation prediction. In our research, we've decided to experiment with ensemble learning techniques on classification procedures & techniques [23]. This section explains the problem statement, literature review, and recent advancements. The authors are advised to remain focused on the problem and should avoid discussing misleading stuff [1]. Despite numerous advancements in agricultural machine learning models, existing platforms often lack adaptability, real-time capabilities, and user accessibility, particularly in resource-constrained environments. This research addresses these gaps by developing a platform that leverages ensemble learning and incorporates a live training mechanism. Unlike previous works, which often rely on static models, this platform enables continuous updates and personalized recommendations based on user-provided datasets.

Material and Methods:

The main idea of this system is to develop an E-business online platform that helps users get irrigation predictions and crop recommendations on the basis of their own provided data sets. This allows you to add datasets of various environmental variables such as temperature, humidity, soil moisture, and soil pH. These values are transferred to the platform database and later returned for further processing as explained in Figure 2.

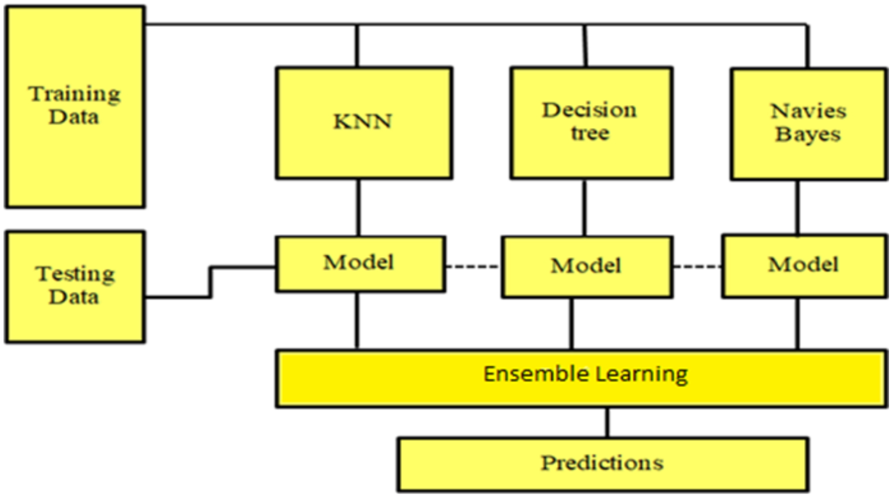

Figure 2. Explain the ensemble learning techniques diagram.

The data is then pushed to a Python server for training records. To connect users to the web, the application is developed using frameworks. We trained the dataset for crop or irrigation prediction and generated three sub-models based on KNN, Naive Bayes, or Decision Trees. The system receives input from the user and makes recommendations from each model. The system applies ensemble learning techniques to these results to make final recommendations. For fertilizer recommendations, we used datasets that provide the most ideal soil parameter values needed to grow specific crops. Depending on the results of various machine learning techniques on the backend, the web platform creates a suggestion template that tells the farmer what to do next.

Assemble Learning Technique:

- Multiple odd numbers of procedures will be used to make our final model.
- First, each model makes its own separate decision.
- Then our final ensemble model will take their decision and consider them as a vote.
- The final prediction is given in favor of the decision with which the majority of the 0 votes lie.
- If the voting system fails, then the decision made by the model with the highest accuracy for that dataset is chosen as the final decision.

Our goal is to add value to regional smart crop prediction in Pakistan. For this purpose, we have extended the system with a live training feature where users can add their own dataset. The system trains that dataset and generates 3 sub-models (KNN, Naive Bayes, and Decision Trees). After getting input values from the user, the system passes those values to each model to get prediction-based recommendations. It then applies ensemble learning to determine the final prediction and presents it to the user.

Methodology:

Data Preprocessing:

The datasets were sourced from Kaggle, containing parameters such as temperature, humidity, soil pH, nitrogen, phosphorus, and potassium levels. Prior to model training, the data underwent extensive preprocessing to ensure reliability and consistency. Missing values in numerical variables were imputed using the mean or median, depending on the data distribution, while categorical variables were filled using the mode. Continuous features, including temperature, humidity, and soil pH, were normalized using Min-Max scaling to standardize the range and prevent dominance by variables with larger scales. Outliers were detected using interquartile range analysis, and corrective actions such as capping or removal were applied based on their impact on model performance. Additionally, categorical variables were encoded into numerical formats using one-hot encoding to ensure compatibility with machine-learning models. Table 2 shows a sample dataset.

Table 2. Sample dataset of the prediction model.

N	P	K	ph	rainfall	humidity	temperatu	label
90	42	43	6.502985	202.9355	82.00274	20.87974	rice
85	58	41	7.038096	226.6555	80.31964	21.77046	rice
60	55	44	7.840207	263.9642	82.32076	23.00446	rice
74	35	40	6.980401	242.864	80.15836	26.4911	rice
78	42	42	7.628473	262.7173	81.60487	20.13017	rice
69	37	42	7.073454	251.055	83.37012	23.05805	rice
69	55	38	5.700806	271.3249	82.63941	22.70884	rice
94	53	40	5.718627	241.9742	82.89409	20.27774	rice
89	54	38	6.685346	230.4462	83.53522	24.51588	rice
68	58	38	6.336254	221.2092	83.03323	23.22397	rice
91	53	40	5.386168	264.6149	81.41754	26.52724	rice
90	46	42	7.502834	250.0832	81.45062	23.97898	rice
78	58	44	5.108682	284.4365	80.88685	26.8008	rice
93	56	36	6.984354	185.2773	82.05687	24.01498	rice
94	50	37	6.94802	209.587	80.66385	25.66585	rice
60	48	39	7.042299	231.0863	80.30026	24.28209	rice
85	38	41	6.249051	276.6552	82.78837	21.58712	rice

The choice of KNN, Naive Bayes, and Decision Tree models was driven by their simplicity, interpretability, and efficiency, which align with the objectives of this study. KNN was selected for its ability to handle multi-class classification problems and its effectiveness with smaller datasets, as it relies on instance-based learning. This is particularly useful for real-time predictions where user-provided datasets may be limited in size. Naive Bayes was chosen for its computational efficiency and robustness in handling datasets with categorical features, which are common in agricultural data (e.g., soil type and crop categories). Its assumption of feature independence, while a simplification, works well in scenarios where relationships among features are weak or not dominant. Decision Trees were included for their intuitive structure, which provides clear, rule-based outputs that are easily interpretable by non-technical users, such as farmers. Additionally, Decision Trees perform well with mixed data types and do not require feature scaling, making them ideal for diverse agricultural datasets. The different scenario outputs shown in Table 3 containing values of Decision Tree, KNN, and Naive Bayes are loaded into the Ensemble-learning technique containing different scenarios related to different crop fields. The final prediction is made on behalf of the option that has obtained a majority of votes. If the voting system fails, the final decision is determined using a model with the highest accuracy. (i.e.) the shaded lines in Table 1 illustrate that each technique produces different outcomes, and the ensemble-learning technique selects the crop of the KNN technique because it has the maximum accuracy as compared to other techniques.

Table 3. Different techniques result.

Sr.	Decision Tree	KNN	Naïve Bayes	Ensemble Learning
1	Kidney beans	Pigeon peas	Pigeon peas	*Pigeon peas*
2	Kidney beans	Pigeon peas	Moth beans	*Moth beans*
3	Chickpea	Moth beans	Moth beans	*Moth beans*
4	Banana	Chickpea	Banana	*Banana*
5	Apple	Pigeon peas	Pigeon peas	*Pigeon peas*
6	Watermelon	Maize	Muskmelon	*Maize*
7	Pomegranate	Jute	Coconut	*Jute*
8	Grapes	Pigeon peas	Lentil	*Pigeon peas*
9	Banana	Banana	Papaya	*Banana*
10	Coffee	Pigeon peas	Jute	*Pigeon peas*

Decision Tree, KNN, and Naive Bayes values are fed into the Ensemble-learning procedure which contains different scenarios related to irrigation recommendation as shown in Table 3. The selection with the most votes is the one for which the final prediction is made. If the voting mechanism fails, the final decision is made based on the most accurate model accessible.

Table 4. Results from different procedures.

Sr.	Decision Tree	KNN	Naïve Bayes	Ensemble Learning
1	0	0	0	0
2	0	1	0	0
3	1	1	1	1
4	0	1	1	1
5	0	1	1	1
6	1	1	1	1
7	0	1	0	0
8	0	0	0	0
9	1	1	1	1
10	1	1	0	1

Table 4 shows a different scenario output in binary form 0 means water is not required and Sr. 1 means water is required for soil. Decision Tree, KNN, and Naive Bayes values are fed into the Ensemble-learning procedure which contains different scenarios related to irrigation recommendation. The selection with the most votes is the one for which the final prediction is made. If the voting mechanism fails, the final decision is made based on the most accurate model accessible. Ensemble learning in this study involves the integration of three models: KNN, Naive Bayes, and Decision Tree. Each model provides an independent prediction based on the input data, and the final decision is derived using a majority voting mechanism.

In the voting process, each model casts a vote for its prediction. The ensemble model aggregates these votes, and the option receiving the majority is selected as the final prediction. For instance, if two models predict Crop A and one model predicts Crop B, the ensemble model outputs Crop A as the final decision. To address cases where no majority decision is reached (e.g., each model predicts a different outcome), a conflict resolution strategy is employed. In such scenarios, the decision is based on the accuracy of the individual models for the specific dataset. The model with the highest accuracy, as determined during validation, is given precedence. This ensures that the ensemble model not only reflects the consensus of its components but also leverages the strengths of the most reliable model when necessary.

This approach ensures robustness and adaptability in the ensemble predictions, making the platform more effective in providing reliable recommendations for crop, irrigation, and fertilizer management.

Result and Discussion:

Nitrogen, potassium, phosphorus, temperature, humidity, pH, and precipitation are the parameters used to train the crop prediction model. For irrigation prediction, the training data consists of environmental parameters (temperature and humidity) and soil and crop parameters such as soil moisture, crop days, and crop type. Fertilizer recommendation models consist of parameters such as nitrogen, phosphorus, potassium, pH, and soil moisture. When the user enters values, these values are given to each model individually and the prediction results are saved. The proposed study then applied ensemble learning to the results to generate final predictions based on voting technique or model accuracy. As a farmer, one gets precise recommendations on which crops to grow based on various characteristics such as humidity, pH, and precipitation. Also, users can get recommended fertilizers based on different characteristics such as moisture, nitrogen, phosphorus, and potassium. The three procedures used to generate sub-models and apply ensemble learning are Decision Tree, Naive Bayes, and KNN.

Crop Recommendation:

For crop recommendation, we trained a dataset and generated three sub-models based on KNN, naive Bayes, and decision trees. The system receives input from the user and makes recommendations from each model. The system applies ensemble learning techniques to these results to make final recommendations. Users add nitrogen, phosphorus, potassium, pH, temperature, humidity, or precipitation values to predict what will be appropriate.

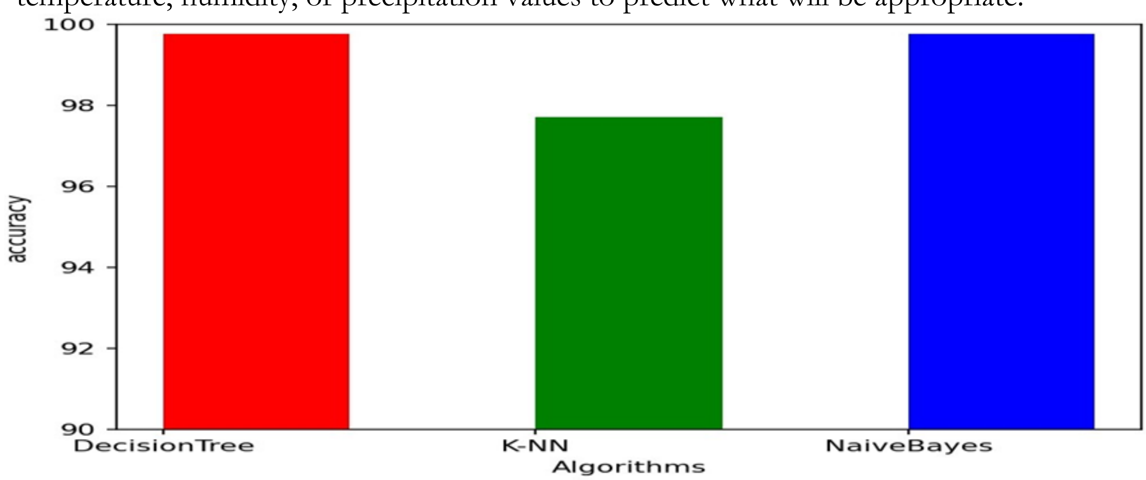

Figure 3. Comparison of crop prediction accuracy using different models.

The purpose of the test was to Figure out how the different procedure works and how it predicts the crop when parameters are given as input. Figure 3 delineates the comparison of different models' crop prediction and the accuracy of a decision tree, KNN, and Naïve Bayes is 99.77%, 97.72%, and 99.77% respectively.

Irrigation Prediction

For irrigation, prediction users are required to add soil and crop-related data values such as soil moisture, crop day, and crop type. The weather-related values will be automatically added when the user chooses his state and district, and the final prediction of irrigation can be achieved.

The purpose of the test was to figure out how the different techniques work and how they predict the irrigation when parameters are given as input. The irrigation-prediction of different techniques accuracy are decision tree 86.0%, KNN 90.0%, and Naive Bayes 90.0% as shown in Figure 4.

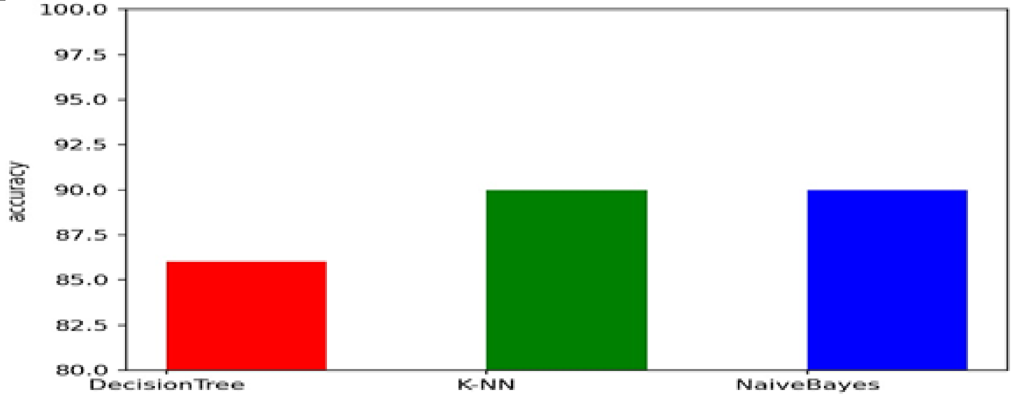

Figure 4. Graph for irrigation prediction accuracy of different models.

To evaluate the performance of individual models (KNN, Naive Bayes, Decision Tree) and the ensemble model, their respective accuracies were compared as shown in Table 5.

Table 5. Accuracy comparison of individual models and Ensemble models.

Model	Crop Prediction Accuracy (%)	Irrigation Prediction Accuracy (%)
KNN	97.72	90.0
Naive Bayes	99.77	90.0
Decision Tree	99.77	86.0
Ensemble Model	99.90	92.5

The ensemble model outperformed the individual models in both crop and irrigation prediction tasks, achieving an overall accuracy of 99.90% and 92.5%, respectively. This improvement can be attributed to the ensemble method's ability to combine predictions from multiple models, leveraging their individual strengths while compensating for their weaknesses. Explanation of Ensemble Model Superiority: The ensemble model aggregates predictions using a majority voting mechanism, ensuring that the final prediction aligns with the consensus of the models. This approach reduces the impact of biases or errors from any single model. For example:

- KNN is effective for small datasets but sensitive to noise, which the ensemble mitigates.
- Naive Bayes assumes feature independence, which may oversimplify relationships, but the ensemble offsets this limitation by incorporating results from more complex models like Decision Trees.
- Decision Trees are prone to overfitting, which the ensemble balances by integrating predictions from more generalizable models like Naive Bayes.

By combining these models, the ensemble achieves improved robustness and higher accuracy, making it more reliable for real-world applications.

In Figure 5 and Figure 6, different models performance is compared. In Figure 5, the crop VS irrigation prediction accuracy of different models is compared. The ensemble model is outperforming in terms of accuracy among KNN, Naïve Bayes, Decision Tree, and Ensemble models. Similarly, in Figure 6 Model performance in terms of Precision, Recall & F1 Score is measured. The ensemble model is giving better performance as compared to other ML techniques.

Fertilizer Recommendation:

For fertilizer recommendations, users are required to add values such as nitrogen, potassium, phosphorous, and crops. Then the user will be able to receive fertilizer advice based on a variety of characteristics, including moisture, nitrogen, phosphorus, and potassium.

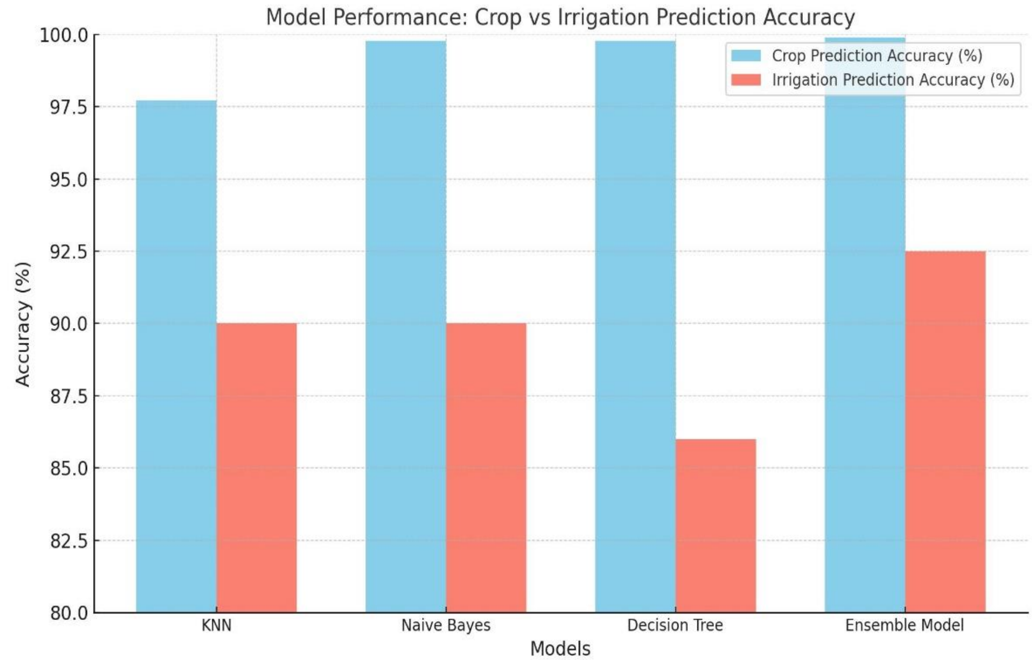

Figure 5. Comparison of crop vs irrigation prediction accuracy.

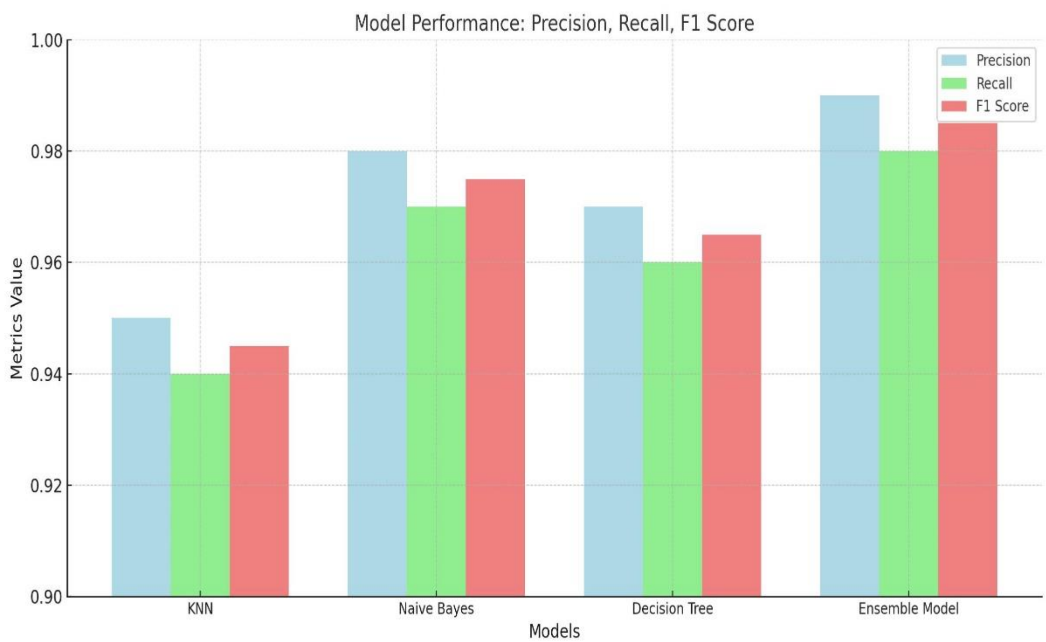

Figure 6. Comparison of model performance in terms of Precision, Recall & F1 score.

Statistical Validation:

The performance of the proposed models (KNN, Naive Bayes, Decision Tree) was evaluated using key metrics such as accuracy, precision, recall, and F1-score. To further validate the reliability of the results, confidence intervals (95%) were calculated for the accuracy of each model. For crop prediction, the accuracy of KNN, Naive Bayes, and Decision Tree models was found to be 97.72%, 99.77%, and 99.77%, respectively. The corresponding 95% confidence intervals for these models were as follows:

- KNN: 96.9% to 98.5%
- Naive Bayes: 99.3% to 100%
- Decision Tree: 99.3% to 100%.

The variance in accuracy for the models was minimal, reflecting the consistency of predictions across different subsets of the dataset. For instance, the variance in accuracy for the Decision Tree model was 0.002%, demonstrating its robustness.

Similarly, for irrigation prediction, the accuracy achieved was 86.0% for Decision Tree, 90.0% for KNN, and 90.0% for Naive Bayes, with 95% confidence intervals:

- Decision Tree: 84.5% to 87.5%
- KNN: 88.5% to 91.5%
- Naive Bayes: 88.5% to 91.5%.

These statistical validations confirm the reliability of the proposed ensemble learning system and highlight its applicability to real-world agricultural datasets.

Conclusion:

This paper discusses the use of ensemble learning techniques for prediction-based recommendations for crops, fertilizers, and irrigation systems in Pakistan using machine learning. The study also added a live training feature to our system. This means that users can get precise predictions by using their own provided datasets to train the models by taking input values from users, the system passes those values to each sub-model to get prediction-based recommendations. It then applies ensemble learning to determine the final prediction and presents it to the user. Moreover, when the ensemble learning procedure fails to make a final decision in the voting system, the system compares the accuracy of the three sub-models (KNN, Naive Bayes, or Decision Tree). While the ensemble model outperformed individual models in most scenarios, some underperforming cases were observed. For datasets with highly imbalanced classes, Naive Bayes tended to favor the majority class due to its assumption of equal class distribution. Similarly, KNN showed sensitivity to noisy data, as outliers in the feature space affected the nearest neighbor calculations. These issues underline the importance of robust data preprocessing, such as balancing class distributions and removing outliers, to enhance overall performance.

Future Work:

Conservative crop prediction communities, to which farmers belong, have a variety of issues, including low crop growth, ineffective fertilizer management, and unfavorable weather circumstances. In addition to giving farmers advice for fertilizer, real-time sensor readings and the use of machine learning procedures are required to enable farmers to make judgments about which crops to cultivate in a certain location. The decision should be determined by a variety of elements, including soil conditions, water requirements, and climatic circumstances. The work presented in this paper was very keen to recognize the future of water scarcity that weighs heavily on crop prediction in Pakistan. Crop prediction is a visionary profession in Pakistan and these fields consume large amounts of water. Irrigation forecasting can be used in real environments with intelligent irrigation systems to improve land productivity.

In the future, the platform can be further enhanced by incorporating deep learning techniques, such as CNNs and RNNs, which can process complex, high-dimensional data more effectively. This would enable the platform to handle unstructured data, such as satellite images or sensor-generated time-series data, providing richer insights for crop and irrigation predictions. Additionally, efforts can be directed toward scaling the platform for broader applicability by integrating distributed computing frameworks like Apache Spark or TensorFlow. These frameworks would allow the platform to process larger datasets and support real-time predictions for multiple regions simultaneously.

Another potential direction is the integration of IoT devices and real-time sensors to automate data collection and enhance the precision of predictions. This would reduce the dependency on manually uploaded datasets and provide dynamic updates for recommendations. Finally, expanding the platform to support mobile and multilingual interfaces can increase its accessibility to farmers in diverse geographic and cultural settings, fostering widespread adoption.

Acknowledgment.
The authors would like to thank all contributing institutions.

Author's Contribution.

Ammar Rafiq (AR) & Muhammad Usman Younus (MUY) proposed the methodology and carried out the experiments, whereas Muhammad Younus (MY) and Najaf Ali designed the study. Arsalan Malik (AM) & Kalsoom Safdar (KS) gathered and examined data. Nusrat Husain (NH), Aqeel Haider & Ahmad Arfeen made contributions to the draft of the work, while Muhammad Asghar Nadeem (MAN), Najaf Ali and Faisal Mumtaz (FM) offered insightful commentary and crucial edits for the finished version. Their combined efforts produced a thorough and significant study result.

Conflict of Interest. There is no conflict of interest in this paper.

Reference:

[1] 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, 2020.

[2] A. Muniasamy, "Machine learning for smart farming: a focus on desert agriculture," in 2020 International Conference on Computing and Information Technology (ICCIT-1441), IEEE, 2020, pp. 1–5.

[3] R. Akhter and S. A. Sofi, "Precision agriculture using IoT data analytics and machine learning," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 8, pp. 5602–5618, 2022.

[4] B. Buakum et al., "Two-Stage Ensemble Deep Learning Model for Precise Leaf Abnormality Detection in Centella asiatica," AgriEngineering, vol. 6, no. 1, pp. 620–644, 2024.

[5] A. Rehman, T. Saba, M. Kashif, S. M. Fati, S. A. Bahaj, and H. Chaudhry, "A revisit of internet of things technologies for monitoring and control strategies in smart agriculture," Agronomy, vol. 12, no. 1, p. 127, 2022.

[6] S. K. S. Durai and M. D. Shamili, "Smart farming using machine learning and deep learning techniques," Decision Analytics Journal, vol. 3, p. 100041, 2022.

[7] A. Rafiq and M. U. Younus, "Implementation of Fi-WSN in precision agriculture for automated irrigation system.," Journal of Agricultural Research (03681157), vol. 60, no. 1, 2022.

[8] N. Ahmed, H. M. S. Asif, G. Saleem, and M. U. Younus, "Image quality assessment for foliar disease identification (AgroPath)," arXiv preprint arXiv:2209.12443, 2022.

[9] L. N. P. Ndjuluwa, J. A. Adebisi, and M. Dayoub, "Internet of Things for Crop Farming: A Review of Technologies and Applications. Commodities 2023, 2, 367–381." 2023.

[10] Z. Abbas, M. R. Anjum, M. U. Younus, and B. S. Chowdhry, "Monitoring of Gas Distribution Pipelines Network Using Wireless Sensor Networks," Wireless Personal Communications, pp. 1–20.

[11] D. Vallejo-Gomez, M. Osorio, and C. A. Hincapie, "Smart irrigation systems in agriculture: A systematic review," Agronomy, vol. 13, no. 2, p. 342, 2023.

[12] N. G. Rezk, E. E.-D. Hemdan, A.-F. Attia, A. El-Sayed, and M. A. El-Rashidy, "An efficient IoT based smart farming system using machine learning algorithms," Multimed Tools Appl, vol. 80, no. 1, pp. 773–797, Jan. 2021, doi: 10.1007/s11042-020-09740-6.

[13] F. K. Shaikh, M. A. Memon, N. A. Mahoto, S. Zeadally, and J. Nebhen, "Artificial intelligence best practices in smart agriculture," IEEE micro, vol. 42, no. 1, pp. 17–24, 2021.

[14] G. Venkatakotireddy, C. L. Reddy, J. Prabhakaran, T. Nivethitha, and R. Mahaveerakannan, “Development of High-Quality Crops using Optimized Machine Learning in Smart Agriculture Environment,” in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), IEEE, 2023, pp. 532–536.
[15] N. Ahmed, H. M. S. Asif, G. Saleem, and M. U. Younus, “Development of Crop Yield Estimation Model using Soil and Environmental Parameters,” Feb. 10, 2021, arXiv: arXiv:2102.05755. doi: 10.48550/arXiv.2102.05755.
[16] E. Elbasi et al., “Crop prediction model using machine learning algorithms,” Applied Sciences, vol. 13, no. 16, p. 9288, 2023.
[17] N. Zhu et al., “Deep learning for smart agriculture: Concepts, tools, applications, and opportunities,” International Journal of Agricultural and Biological Engineering, vol. 11, no. 4, pp. 32–44, 2018.
[18] M. U. Younus, M. K. Khan, and A. R. Bhatti, “Improving the software-defined wireless sensor networks routing performance using reinforcement learning,” IEEE Internet of Things Journal, vol. 9, no. 5, pp. 3495–3508, 2021.
[19] B. Fulkar, S. Mendhe, and P. Patil, “Artificial Intelligence Cultivation: Transforming Agriculture for a Smart and Sustainable Future,” in 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), IEEE, 2024, pp. 960–964.
[20] A. A. S. Aliar, J. Yesudhasan, M. Alagarsamy, K. Anbalagan, J. Sakkarai, and K. Suriyan, “A comprehensive analysis on IoT based smart farming solutions using machine learning algorithms,” Bulletin of Electrical Engineering and Informatics, vol. 11, no. 3, pp. 1550–1557, 2022.
[21] S. Sundaresan, S. D. Johnson, V. M. Bharathy, P. M. P. Kumar, and M. Surendar, “Machine learning and IoT-based smart farming for enhancing the crop yield,” in Journal of Physics: Conference Series, IOP Publishing, 2023, p. 012028.
[22] S. A. Bhat and N.-F. Huang, “Big data and ai revolution in precision agriculture: Survey and challenges,” Ieee Access, vol. 9, pp. 110209–110222, 2021.
[23] G. Davrazos, T. Panagiotakopoulos, S. Kotsiantis, and A. Kameas, “IoT-Enabled Crop Recommendation in Smart Agriculture Using Machine Learning,” in 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA), IEEE, 2023, pp. 1–4.
[24] Ahmed et al., “Crop Recommendation in Smart Agriculture Using Machine Learning,” in 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA), IEEE, 2023, pp. 1–4.
[25] Rafiq and Younus, “Implementation of Fi-WSN in Precision Agriculture for Automated Irrigation System,” Journal of Agricultural Research, vol. 60, no. 1, 2022.