International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

A machine learning-based model for crop recommendation using Agro-climatic and soil nutrient parameters (Agrismart)

Rebina Ferdous [1, *], Soyaib Rahman [2], Salim Mia [2] and Hafizur Rahman [3]

[1] Genome Research Centre, Bangladesh Jute Research Institute, Dhaka-1207, Bangladesh.
[2] Department of Computer Science and Engineering, Jagannath University, Dhaka-1100, Bangladesh.
[3] ICT cell, Jagannath University, Dhaka-1100, Bangladesh.

International Journal of Science and Research Archive, 2025, 15(03), 1080-1089

Publication history: Received on 25 April 2025; revised on 01 June 2025; accepted on 04 June 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.15.3.1713

Abstract

Agrismart is an innovative decision support system for agriculture, utilizing machine learning to enhance crop and fertilizer recommendations. It integrates advanced data analysis techniques to process a wide range of agricultural data, including soil information and weather data. Through the analysis of soil parameters such as moisture levels, nitrogen, phosphorus, potassium content, and pH, combined with real-time weather data including temperature, humidity, and rainfall, Agrismart generates recommendations for crop selection and optimal fertilizer application. The system also provides real-time information through external APIs like weather forecast accessible via a user-friendly interface. Agrismart's goal is to improve productivity, resource efficiency, and sustainability in agriculture. By providing farmers with data-driven recommendations, Agrismart empowers them to make informed decisions, ultimately leading to increased crop yields, reduced input costs, and a more sustainable farming practice. Moreover, Agrismart is committed to ongoing research and development, continuously refining its models and scaling its capabilities for real-world deployment. With its innovative approach, Agrismart is poised to revolutionize farming practices globally, making agriculture more efficient, sustainable, and profitable.

Keywords: Agriculture Machine Learning; Crop Recommendation; Fertilizer Recommendation; Soil Weather Decision Support User Interface

1. Introduction

The agricultural sector plays a pivotal role in sustaining human life and ensuring food security worldwide. With the ever-growing global population and changing environmental conditions, there is an increasing pressure on farmers to enhance productivity while minimizing resource utilization and environmental impact. In this context, the integration of advanced technologies such as machine learning and data analytics presents a promising avenue for optimizing agricultural practices and improving crop yield. In agriculture, the challenge of crop selection persists due to diverse soil conditions, variable weather patterns, and market dynamics. Traditional methods often lack precision, leading to suboptimal yields and resource inefficiency. Climate change further complicates these issues, necessitating adaptive crop choices. Farmers need personalized, data-driven tools to navigate these complexities effectively. Current solutions often fail to consider individual farm circumstances, highlighting the need for innovative approaches. Agrismart addresses this need by offering a comprehensive crop recommendation system powered by machine learning and integrated with external data sources. By analyzing soil and weather data, Agrismart provides personalized recommendations to optimize crop selection and management practices, enhancing productivity and sustainability in agriculture. This work fills the gap in current decision support tools, providing farmers with a user-friendly platform to make informed decisions and adapt to modern agricultural challenges.

* Corresponding author: Rebina Ferdous

2. Related works

Crop and fertilizer recommendation systems have garnered significant attention in agricultural research and practice due to their potential to improve decision-making processes and enhance productivity. In this section, we review existing literature on crop recommendation systems, machine learning techniques in agriculture, and the integration of external data sources and APIs. Many works that may be utilized to forecast harvests for the user were provided by earlier literature. Nevertheless, different soil characteristics are not the main focus of the study. This means that in order to better match the soil properties and climate conditions, crop prediction and recommendation systems must be made more effective. The appropriate levels of soil nutrients, such as nitrogen, potassium, and phosphorus, were determined using a model, which also provided crop recommendations for the future. A neural network was used to predict the crops in [1] with an accuracy of 89.88%. While appropriate crops are predicted in this article, crop rotation has not been fully examined. A method that takes into account all the factors, such as soil type, sowing season, and geographic location, has been proposed by the study to assist farmers in selecting crops. When recommending a suitable crop to the user, the suggested approach takes into account meteorological aspects like rainfall, temperature, and geographic position within the state, in addition to soil attributes like pH value, nutrient concentration, and soil type. Many machine learning methods were employed, however the outcomes are unimpressive. Similarly, a Naive Bayes algorithm was proposed for crop recommendation [2], which included the temperature, humidity, and moisture content of the soil as important variables. Using machine learning (ML), one of the most advanced technologies for crop prediction, this study provides novice farmers with a strategy that tells them where to plant quality crops. Moreover, the Naive Bayes supervised learning technique offers instructions on how to accomplish it. These models need to forecast things more accurately. Another machine learning method was suggested [3] in order to evaluate the various characteristics of the soil and suggest the crop for planting. They employed k-nearest neighbor (KNN) algorithms, but the sole basis for prediction was the characteristics of the soil. In the agricultural sector, crop yields were predicted using the random forest (RF) method [4]. By taking into account as few models as feasible, the RF technique yields the best crop production model. The findings suggest that agricultural sectors can benefit from crop production prediction. By assessing the properties of the soil using online soil spectroscopy and a prototype sensor, a winter wheat prediction model was suggested [5]. The model combined XY-fused networks, artificial neural networks based on counter-propagation, and supervised Kohonen networks with a self-organizing map. To optimize crop productivity, understanding soil parameters alone will not be suffi- cient, even if the method produces useful data. Feature selection is essential since crop prediction is dependent on a number of variables. A comparative analysis of various feature selection strategies was carried out to forecast crops using various classifiers utilizing soil properties and environmental data, such as rainfall, season, texture, and temperature [6]. A technique for crop classification based on particular data was created by Suresh et al. A support vector machine (SVM) was used to increase productivity and precision (SVM). The two datasets that this inquiry focused on were the crop data sample dataset and the location data sample dataset. According to the availability of the particular nutrients, namely N, P, K, and pH, particular crops, such as rice, black gram, maize, carrot, and radish, were recommended with this suggested strategy [7]. A method for precisely recommending the optimal crop based on the kind and characteristics of the soil, including the surface temperature and typical rainfall, was proposed in [8]. Naive Bayes, RF, and linear SVM were the machine learning techniques employed by this proposed system. The recommended crop varieties for Kharif and Rabi were identified from the input soil dataset by this crop recommendation system. The accuracy rate was 99.91% when the recommended method was used. With an overall improvement over various other ways of 3.6%, the study in [9] accurately compares numerous machine learning algorithms to calculate the crop's recommended yield. The subsequent research aids agronomists in choosing the right crops for farming. The output of the crops will also rise dramatically. Consequently, this method will raise India's income. Numerous aspects were reviewed [10] to advise the crops, including N, P, K, pH, temperature, humidity, and rainfall. There are eight features and 2200 cases in the dataset. Using ML algorithms in the Waikato environment for knowledge analysis (WEKA) yields the best model. Rule-based classifiers, multilayer perceptrons, and decision tree classifiers are the machine learning algorithms selected for classification. The feature importance in this study has not been assessed. Artificial neural networks (ANNs), which are deep learning algorithms, are employed in [11] to generate exact harvests at the right times. To forecast crops, a deep neural network and graphical user interface are utilized to provide inputs like moisture, temperature, pH, and humidity via a sensor network and the Internet of Things. Crop ideas can assist farmers in selecting which crops to plant. The Internet of Things (IoT) and machine learning (ML) system, which uses sensors to enable soil testing, were proposed in [12]. It is predicated on observ- ing and quantifying soil characteristics. This technique promotes crop vitality and lessens the possibility of soil deterioration. Numerous sensors are used by this system to track the soil's nutrients—nitrogen, phosphorus, and potassium (NPK)—temperature, humidity, and soil wetness. These sensors measure pH, soil moisture, temperature, and other variables. While taking into account every feature, they did not examine the significance of each feature for this study. Furthermore, the input parameters are not subjected to hyperparameter optimization. In [13], an ensemble model with majority voting methods was proposed as a recommendation system that uses random trees, chi-squared automatic interaction detection (CHAID), KNN, and Naive Bayes as learners to recommend a good crop with high specific accuracy and efficacy based on soil data. Nevertheless, this study did not include an analysis or comparison of the results, nor a feature importance assessment. Various soli and environmental factors were demonstrated in [14], and crops were predicted using these features. There, machine learning techniques such as Decision Tree, KNN, Random Forest, XGBoost, Support Vector Machine, and others were applied. The best crop prediction model was found in [15], which can help farmers choose the optimal crop to grow based on soil nutrient levels and local climate parameters. In this article, two popular metrics—Gini and Entropy—for algorithms like KNN, decision trees, and RF classifiers are compared. Results indicate that RF has the highest accuracy. In order to identify the best attributes for crop recommendations, more analysis of features is necessary. Different factors like N, P, K, pH, temperature, humidity, and rainfall to advise the crops were discussed [15]. The dataset consists of 2200 instances and eight features. The best model is created by utilizing ML algorithms in Waikato environment for knowledge analysis (WEKA). The ML algorithms chosen for classification are decision tree classifiers, multilayer perceptron, and rule-based classifiers. They have not evaluated the feature importance in this study.

3. Methodology

The concepts and materials utilized for this experiment are described to make the proposed methodology more easily readable and clear. We have two components in the suggested system. These are the sections of the dataset and the website. The pre-processed dataset is included in the dataset sec- tion.The pertinent features were selected, and then these were fed into the machine learning models. Additionally, we divided the data into training and testing sets. We employed several ML classification methods, including Decision Tree, Naive Bayes, SVM, Logistic Regression, RF, and some other algorithms, to construct our ML model. Python is the language that is used for the machine learning models. We test, run, and quantify the correctness of the model. Ultimately, we select the most effective model to incorporate into our online application. The frontend and backend are the two subparts of the web application portion. We used PHP as our backend language for the backend and HTML, CSS, Bootstrap, Javascript, and a few additional dependencies for the frontend. Our web application displayed the outcome at the conclusion, and we stored our input and output to the dataset for further use. In order to improve the aesthetics of our website, we also employed various APIs, such as NewsAPI and WeatherAPI.

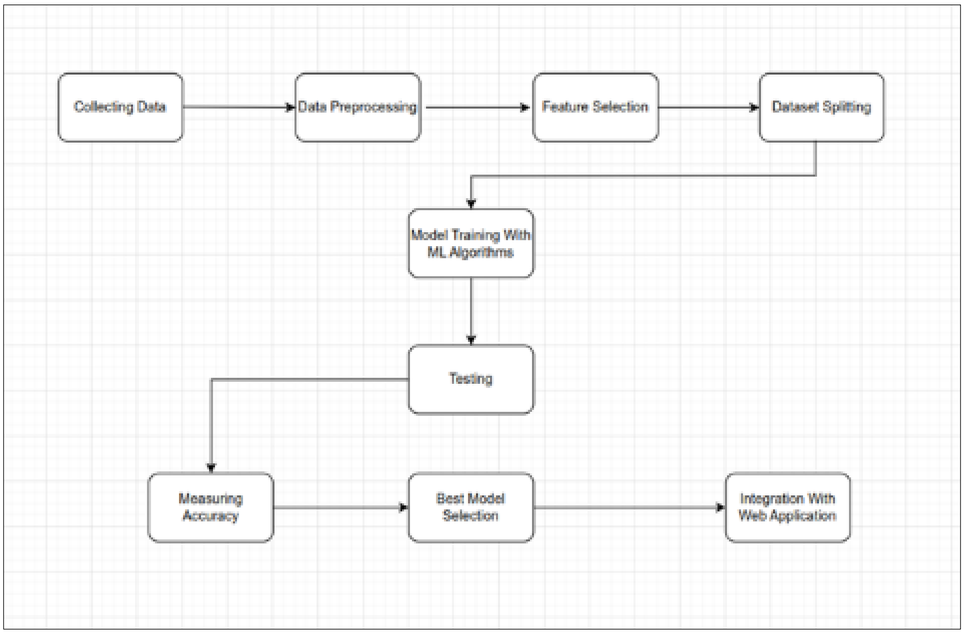

Figure 1 Methodology

3.1. Dataset collection

3.1.1. Crop recommendation dataset

Two thousand records and seven factors (N, P, K, Temperature, Humidity, pH, and Rainfall) were used in a crop recommen dation dataset. For each crop, the necessary soil content was ascertained in order to better comprehend the nature of the data. The dataset was divided into training and validation sets before being subjected to cross-validation. The information was taken from (https://www.kaggle.com/datasets/atharvaingle/ crop-recommendation-dataset)

Kaggle. The overview of the dataset used in this investigation is provided in Table 1. This dataset's characteristics such as humidity, temperature, rainfall, pH, nitrogen, phosphorous, and potassium requirement ratio were considered essential for crop suggestion when training the model. Specific factors for each crop include temperature, humidity, rainfall, and levels of nitrogen, potassium, and phosphorus. There are no blank fields among the attributes in the crop suggestion dataset.

Table 1 Crop Recommendation Dataset Description

Parameters	**Unit**
Nitrogen	Kg/Hect
Phosphorus	Kg/Hect
Potassium	Kg/Hect
Temperature	C
Humidity	Percentage
PH	ph-scale
Rainfall	mm

The data type of the attributes (int64) is established and labels are given once it has been verified that there are no missing values. The information needs to be perceptually ready prior to using machine learning algorithms to examine the experimental research. ML models are not capable of training and testing the uneven dispersal of practical farming information gathered by sensors. As stated by the features such as soil play's N, P, and K values an important function from a biological perspective since these are the main micronutrients involved in crops. The main contributions of these micronutrients can hold significance.

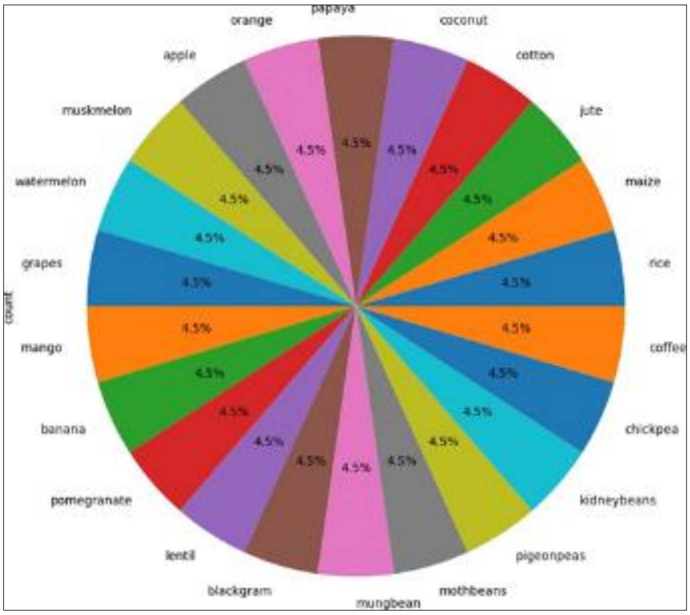

Figure 2 Crop visualization

3.1.2. Fertilizer recommendation dataset

Similar to this, nine characteristics (temperature, humidity, moisture, soil type, crop type, nitrogen, potassium, phosphorus, and fertilizer name) are utilized in a fertilizer recommendation dataset. kaggle (https://www.kaggle.com/datasets/gdabhishek/fertilizer-prediction) is where the dataset was found.

Table 2 Fertilizer Recommendation Dataset Description

Parameters	Unit
Nitrogen	Kg/Hect
Phosphorus	Kg/Hect
Potassium	Kg/Hect
Temperature	C
Humidity	Percentage
Soil Moisture	Percentage
Soil Type	-
Crop Name	-

For each crop, the necessary soil content was ascertained in order to better comprehend the nature of the data. The primary factor influencing the growth of plant leaves is nitrogen, which is present in this soil.

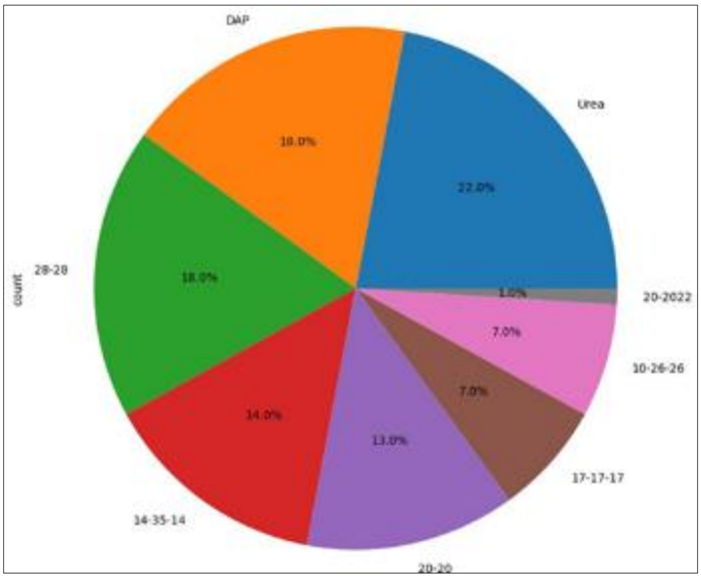

Figure 3 Fertilizer visualization

3.2. Data Preprocessing

Data preprocessing, which involves methods to clean, transform, and arrange raw data into a format appropriate for study, is an essential stage in data analysis and machine learning. This procedure usually entails managing missing values, identify- ing outliers, normalizing, scaling features, and encoding categorical variables, among other things, to guarantee that the data is correct, comprehensive, and prepared for additional examination or modeling.

Most of the dataset we used is good. We still needed to make a few changes. We completed two main duties. If a value is missing, we made an effort to locate the most pertinent information for that row in order to fill the gap. We eliminated the data from the dataset if we could not locate a significant amount of pertinent information.

3.3. Data cleaning

The process of finding and fixing mistakes, inconsistencies, and inaccuracies in a dataset to enhance its quality and dependability for modeling or analysis is known as data cleaning. The data for our model has been cleansed using a number of methods, including the elimination of null values and duplicates.

3.4. Feature selection

Large dimensionality problems in many machine learning applications are solved by feature selection, an important preprocessing step. Choosing a subset of features from the available data is the first step in applying a learning algorithm. By choosing the most important characteristics from the original full feature set and eliminating the redundant, unnecessary, and noisy ones using an evaluation criterion, feature selection reduces the feature set to only the most important or relevant features for the machine learning model. To identify and recommend the optimum crop, a variety of characteristics can be chosen, including pH, temperature, humidity, rainfall, N, P, and K. In this experiment, the names of the different crops serve as the dependent variable. Our suggested strategy takes rainfall-independent elements like pH, temperature, humidity, and N, P, and K into account when recommending crops. Our suggested approach takes into account the following factors when recommending fertilizer: temperature, humidity, moisture, soil type, crop kind, nitrogen, potassium, phosphorus, and fertilizer name.

3.5. Splitting datasets

In machine learning, dividing datasets into subsets for training and testing is essential for assessing how well models function on unobserved data. The model's ability to generalize well beyond the training set of data is ensured by this procedure. Creating two subsets from the available data training set and a testing set is the most popular method for splitting datasets. The model is trained that is, it learns the underlying patterns in the data using the training set. The purpose of the testing set, also known as the validation set, is to analyze the performance of the model and determine whether it can be generalized to new data. Generally, the testing set holds the remaining fraction of the data (e.g., 20–30 percent), while the training set comprises a bigger amount (e.g., 70–80 percent). We choose to use 80% of the data for training in our approach, and 20

3.6. Used Algorithms

We employed five methods for Crop Recommendation:

- Decision Tree: The structure of a decision tree is similar to a flowchart, with each internal node denoting a test on a feature, each branch representing the test's result, and each leaf node representing a class label. It is applied to tasks involving regression and classification.
- Naive Bayes: Naive Bayes is a probabilistic classifier that relies on the "naive" assumption of feature independence and is based on Bayes' theorem. Even though it is straightforward, it works well for jobs like text categorization, particularly when working with high-dimensional data.
- SVM: A supervised learning algorithm used for regression and classification applications is SVM. In order to maximize the margin between the classes, it locates the hyper plane in the feature space that best divides the classes. Using the kernel method, SVM can handle non-linear data as well.
- Logistic Regression: A statistical technique called logistic regression is used to analyze datasets where an outcome is determined by one or more independent variables. Its main application is in binary classification, where it fore- casts the likelihood of an occurrence based on one or more predictor factors.
- Random Forest: During training, the Random Forest ensemble learning method builds a large number of decision trees, which it then uses to pro- duce the mean prediction for regression tasks or the mode of the classes for classification tasks. In comparison to individual decision trees, it decreases overfitting and increases accuracy.

4. Results and discussion

4.1. Crop Recommendation

We employed Decision Tree, Naive Bayes, SVM, Logistic Regression, and Random Forest for crop recommendation. With the Random Forest model, we obtain the maximum accuracy. Therefore, we integrated this model into our website to give visitors the most accurate results possible. Python is the language used to create the models since it offers the necessary libraries and functions, making it the ideal choice for machine learning algorithms. This is the comparison table and bar chart for all models and accuracy:

Table 3 Accuracy for Crop Recommendation

Models	**Accuracy**
Decision Tree	0.90
Naive Bayes	0.99
SVM	0.91
Logistic Regression	0.95
Random Forest	0.990909

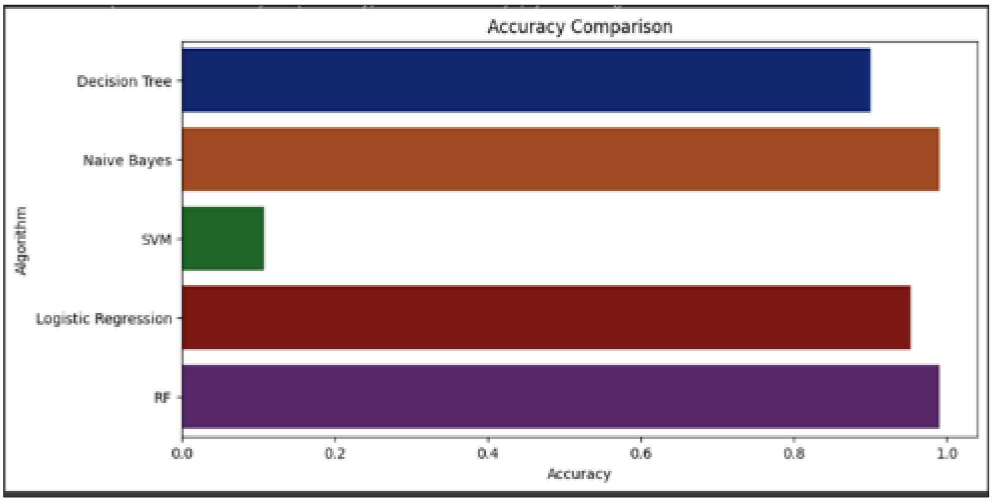

Figure 4 Crop recommendation accuracy comparison

The crop recommendation models' performance metrics are shown below. It displays each model's f1-score, recall, and precision.

Table 4 Performance Matrics for Crop Recommendation

Models	**Precision**	**Recall**	**F1-score**
Decision Tree	0.86	0.90	0.87
Naive Bayes	0.99	0.99	0.99
SVM	0.66	0.11	0.13
Logistic Regression	0.95	0.95	0.95
Random Forest	0.99	0.99	0.99

Here is the Confusion matrix for Random Forest in the crop recommendation model.

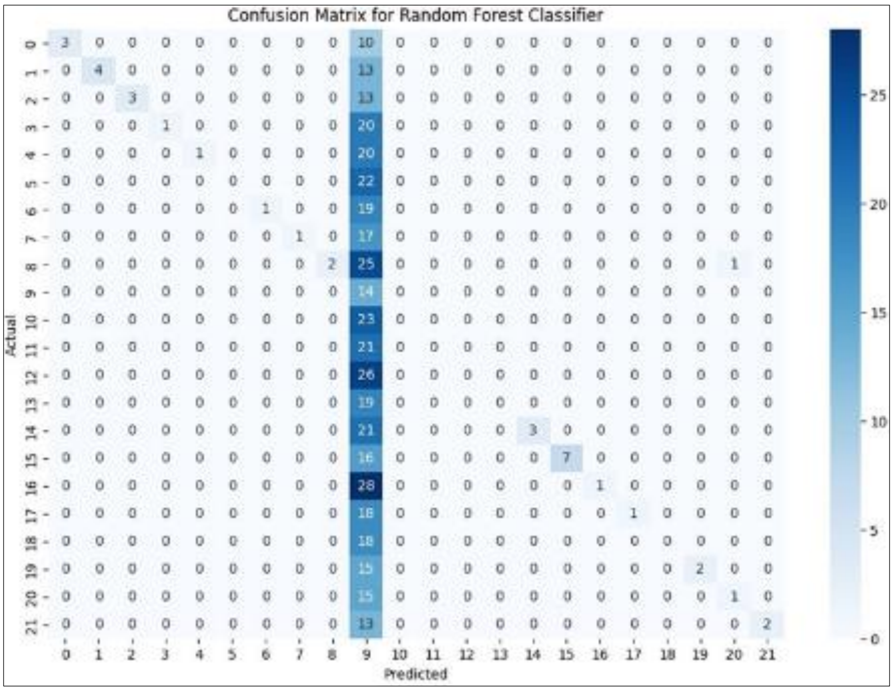

Figure 5 Confusion matrix for RF in crop recommendation

4.2. Fertilizer recommendation

We employ MLP Regressor, Decision Tree, XGBoost, SVM, Logistic Regression, Random Forest, and logistic regression to recommend fertilizer.

Here is the table for all model's accuracy:

Table 5 Accuracy for Fertilizer Recommendation

Models	Accuracy
Decision Tree	0.80
XGBoost	0.95
SVM	0.106
Logistic Regression	0.95
Random Forest	0.90
MLPRegressor	.9167

The confusion matrix for the Decision tree is here:

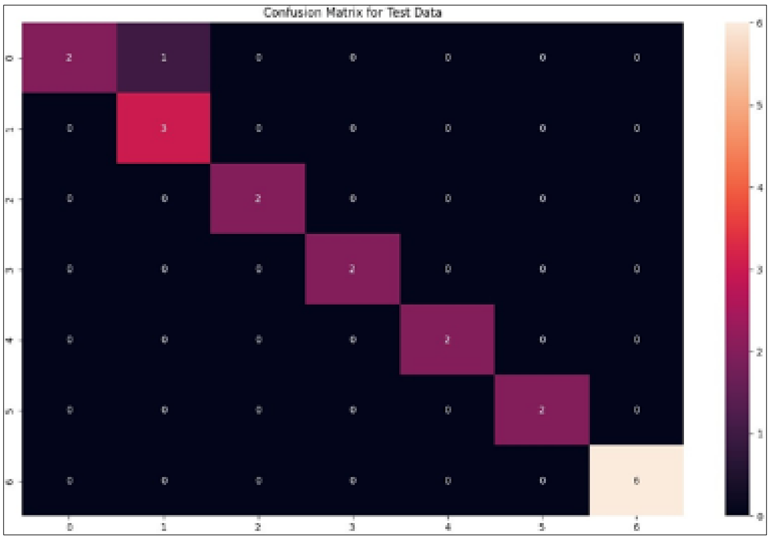

Figure 6 Confusion matrix for XGBoost in fertilizer recommendation

5. Conclusion

A major advancement in agricultural technology is represented by the Agrismart project, which provides a comprehensive machine learning-powered solution for crop and fertilizer advice that is integrated with external APIs. The development and use of Agrismart has enabled a number of noteworthy developments in the field of agriculture. Advanced Decision Support: Agrismart provides farmers with personalized recommendations for crop selection and fertilizer application through the use of machine learning algorithms and soil/weather data analysis. This state-of-the-art decision support system can help farmers maximize resource utilization, improve decision- making, and raise overall productivity. Integration of External APIs: By integrating external APIs like ChatGPT, weather, and news APIs, Agrismart offers real-time access to crucial data and in- sights, including weather forecasts, market trends, and agronomic best practices. This addition enhances the User-Friendly Interface: Farmers of all technical backgrounds can input and comprehend data more easily using Agrismart's user-friendly interface. Users may navigate and understand recommendations more easily with the help of interactive visualizations and intuitive design features, which guarantee usability and accessibility.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Durai SK, Shamili MD. Smart farming using machine learning and deep learning techniques. Decision Analytics Journal. 2022.

[2] Priyadharshini A, Chakraborty S, Kumar A, Pooniwala OR. Intelligent crop recommendation system using machine learning. In 5th International Conference on Computing Methodologies and Communication 2021 (pp. 843-8). IEEE.

[3] Kalimuthu M, Vaishnavi P, Kishore M. Crop prediction using machine learn- ing. In third international conference on smart systems and inventive technology 2020 (pp. 926-32). IEEE.

[4] Mariappan AK, Madhumitha C, Nishitha P, Nivedhitha S. Crop recom- mendation system through soil analysis using classification in machine learning. International Journal of Advanced Science and Technology. 2020; 29(3):12738-47.

[5] Kumar YJ, Spandana V, Vaishnavi VS, Neha K, Devi VG. Supervised machine learning approach for crop yield prediction in agriculture sector. In international conference on communication and electronics systems 2020 (pp. 736-41). IEEE.

[6] Pantazi XE, Moshou D, Alexandridis T, Whetton RL, Mouazen AM. Wheat yield prediction using machine learning and advanced sensing techniques. Computers and Electronics in Agriculture. 2016; 121:57-65.

[7] Anguraj K, Thiyaneswaran B, Megashree G, Shri JP, Navya S, Jayanthi J. Crop recommendation on analyzing soil using machine learning. Turkish Journal of Computer and Mathematics Education. 2021; 12(6):1784-91.

[8] Suresh G, Kumar AS, Lekashri S, Manikandan R, Head CO. Efficient crop yield recommendation system using machine learning for digital farming. International Journal of Modern Agriculture. 2021; 10(1):906-14.

[9] Kulkarni NH, Srinivasan GN, Sagar BM, Cauvery NK. Improving crop pro- ductivity through a crop recommendation system using ensembling technique. In 3rd international conference on computational systems and information technology for sustainable solutions 2018 (pp. 114-9). IEEE.

[10] Garanayak M, Sahu G, Mohanty SN, Jagadev AK. Agricultural recommendation system for crops using different machine learning regression methods. International Journal of Agricultural and Environmental Information Systems. 2021; 12(1):1 20.

[11] Bakthavatchalam K, Karthik B, Thiruvengadam V, Muthal S, Jose D, Kotecha K, et al. IoT framework for measurement and precision agriculture: predicting the crop using machine Technologies. 2022; 10(1). learning algorithms.

[12] Priya PK, Yuvaraj N. An IoT based gradient descent approach for precision crop suggestion using MLP. In journal of physics: conference series 2019 (p. 012038). IOP Publishing.

[13] Gosai D, Raval C, Nayak R, Jayswal H, Patel A. Crop recommendation system using machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2021: 554-69.

[14] Biplob Dey, Jannatul Ferdous , Romel Ahmed Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables

[15] Reddy DA, Dadore B, Watekar A. Crop recommendation system to maximize crop yield in ramtek region using machine learning. International Journal of Scientific Research in Science and Technology. 2019; 6(1):485-9.

[16] Rao MS, Singh A, Reddy NS, Acharya DU. Crop prediction using machine learning. In Journal of Physics: Conference Series 2022 (p. 012033). IOP Publishing.