IoT Based Crop Recommendation System Using Machine Learning for Smart Agriculture

S. Siva Priyanka[1](✉), M. Raju[2], G. Smitha[3], J. Lahari[3], G. Akash Reddy[3], and P. Mani Vinay[3]

[1] Department of ECE, CBIT, Gandipet, Hyderabad, India
sivapriyankas_ece@cbit.ac.in
[2] Department of ECE, Kakatiya Institute of Technology and Science, Warangal, India
[3] Department of ECE (Student), Kakatiya Institute of Technology and Science, Warangal, India

Abstract. In India, agriculture is one of the most significant sources of income. For the survival of the human race, agriculture is essential. These days, the climatic conditions are unpredictable and irregular, which in turn impacts the agriculture industry a lot more than any other industry. A change in the climatic condition affects the nutrients in the soil, which, therefore, affects the type of crop to be sown for the best result. This paper help farmers for recommending suitable crops to yield based on the input parameters using Machine Learning algorithm. Temperature and humidity are collected through the DHT11 sensor using NodeMCU, and NPK, pH, (are directly fed from soil analysis report) and rainfall values. To make it a farmer-friendly application a mobile application is built using Kodular Creator and it communicates with the Firebase cloud platform. To measure the accuracy for crop recommendation, different performance metrics are evaluated: Precision Score, Recall Score, and F1 Score. The proposed method shows better performance compared to the various other existing methods.

Keywords: Crop Recommendation · Machine Learning · Firebase Cloud · Kodular Creator

1 Introduction

India is an agriculturally-based country where most people depend on this industry for survival. The economy of the country is greatly impacted by agriculture. Farming is a key development in the world to produce essential items for the whole world. The majority of farming worldwide is conventional or traditional. It involves methods recommended by seasoned farmers [1]. These methods require hard work and take a lot of time because they are not exact. In agriculture, IoT-based smart farming systems are frequently utilized to monitor fields using sensors like the DHT11 (temperature and humidity sensor). Farmers may check on the state of their fields from anywhere. Farmers are unaware of crop cultivation due to climate change. The recommendation of crops is main problem in the present agriculture sector. The farmers have to know the correct way to resolve the issues to meet their exceptions. Several Machine Learning [ML] techniques have been used to

© The Author(s) 2023
B. Raj et al. (Eds.): ICETE 2023, AER 223, pp. 893–904, 2023.
https://doi.org/10.2991/978-94-6463-252-1_90

recommend the suitable crop more accurately. Machine Learning algorithms can be used when the data is large. It is a subset of artificial intelligence [AI] that predicts the outcomes of the model based on the given labeled data. Using different agriculture parameters, crop recommendation systems can be easily implemented. The system comes with a model to predict accurate crop recommendation with a mobile application. Soil data should be collected to know the pH, N, K, and P values. An IoT-based agriculture system with exact crop production forecasts and crop recommendation models is developed as part of the proposed paper. This work includes pre-processing, feature selection, and classification. After the data has undergone pre-processing, the features are chosen via a feature selection technique. Then, an ML model for intelligent farming based on IoT was developed. The proposed method is used to get good crop recommendations for the climatic changes in a farmer-friendly way. A mobile application has been developed. The crop is predicted based on the weather conditions and soil parameters [2]. This would relatively increase the production of the crop, which the country needs. This application will communicate with farmers using a display. The input parameters should be given first, and then the in-built ML algorithms collect the data and process it; the results are displayed on the screen for the given parameters.

2 Related Work and Background

The author introduces a Machine Learning and Internet of Things-based model for yield prediction in this work [1]. By dividing the work into three stages: first, pre-processing the data gathered from various sources, next feature selection (FS) process is used to eliminate the irrelevant features. The last stage was classification, where an adaptive K-nearest neighbor classifier (AKNCN) was used to classify the soil type based on the input parameters (NPK values), and in the next phase, an ELM (Extreme Learning Machine) was used. For predicting the accuracy of the model, the author used different performance metrics like the coefficient of determination (R^2), root mean square deviation (RMSD), mean absolute percentage error (MAPE).

The author [2] focuses primarily on the Naive Bayes classifier, which was used to forecast crop yield under ideal climatic conditions. To begin, data were gathered from various sources, such as Kaggle, Google weather forecasting, and the data government. The dataset was then pre-processed, feature using a correlation matrix. Later, the data is split into testing and training dataset. A Gaussian Naive Bayes classifier is trained using a training dataset, and the model accuracy is determined by providing the input parameters using the DHT11 sensor and soil moisture sensor, and a microcontroller Arduino Uno with an ATMega processor.

The author [3], focuses on a model that helps in predicting the yield of the six specific crops (rice, maize, cassava, seed cotton, yams, and bananas) at the country level in the area of West Africa using a decision tree, multivariate logistic regression, and k-nearest neighbor algorithms to build the system. This system contains four steps: ETL, feature engineering, model training, and evaluation. Finally, the machine is trained using training data and evaluated the model to determine the system's performance for the decision tree model, which has a coefficient of determination (R^2) of 95.3%, while the K-Nearest Neighbor model and logistic regression have $R^2 = 93.15\%$ and $R^2 = 89.78\%$, respectively.

The author [4] discusses the four different models: crop recommendation, weed identification, pesticide recommendation, and crop cost estimation. For crop recommendation, parameters like temperature, humidity, soil pH, rainfall, nitrogen, phosphorus, and potassium requirement values are considered. The dataset is split 50:50 for training and testing and trained the model using various classification algorithms and finally evaluates a model with different performance metrics. The ML algorithms that are used in this work are the Bagging classifier, KNN, Nave Bayes, Random Forest classifier, XGBoost classifier, AdaBoost classifier, and Decision Tree. However, the accuracy for the random forest classifier is only 94%.

In this study [5], the author developed a web application. The data collected from the hardware device will be tested using various Machine Learning algorithms. This application has a maximum focus on data visualization and provides suggestions for every user. The paper gives an explanation of how the web application gets implemented in a unique way in different ways to track and analyze real-time data by using different ML algorithms with the help of different IoT devices like the 1. User application layer (farmer), 2. Data processing and intelligence layer (the cloud), and 3. IoT data collection with the help of a web application, the farmer will be able to know about the crop by using some parameters like soil pH, rainfall, humidity, and temperature.

In this study [6], author mainly proposed a system based on the internet of things (IoT). The major part of this is using the hard ware device like sensors to test the humidity and temperature. The hardware devices collect data and measure the different attributes like soil moisture, humidity, temperature and soil pH values. The proposed model uses four steps for building a model they are 1. The cloud, 2. Sensor interfacing to an IOT device, 3. User interface and 4. Analyzing of real time data. This data will help in showing the quality of soil.

In this work [7], author proposed Ensemble Machine Learning Techniques. The main objective of this research is to improve forecasting of the best traits for overcoming hunger-related issues. Based on the wild blueberry dataset, stacking regression (SR) and cascading regression (CR) with a new combination of MLA are utilized. According to the root mean square error (RMSE) and coefficient of determination (R^2), model is evaluated and received R^2 of 0.984 and RMSE of 179.898, the suggested SR performed better than another investigation, which obtained R^2 of 0.938 and RMSE of 343.026 on the seven characteristics chosen by XFI. On all features and the features chosen by the SBEFS, the SR had the highest R^2 (0.985).

This study [8] developed a mechanism to aid farmers in crop selection. Additionally, it focuses on site-specific crop management. The entire system is composed of three modules: Profit analysis, Crop recommender, Crop Sustainability predictor. In terms of accuracy, the model is evaluated using different algorithms like, K Nearest Neighbor, K Nearest Neighbor with cross validation, Decision Tree, Naive Bayes, and Support Vector Machine. The suggested approach assists farmers in selecting the proper crop by delivering insights that regular farmers do not keep track of, reducing crop failure and improving output.

3 Proposed IoT Based Smart Agriculture Model

IoT refers to the larger system of interconnected devices in addition to the technology that allows them to communicate with one another and with the cloud. AI focuses on imitating human behavior and intelligence to make machines smarter and more intelligent. Instead of natural intelligence, it is the intelligence displayed by machines. Therefore, IoT is best for cloud-based wireless technology.

To build a control system for a specific application, a variety of control devices are available. The ESP8266 was chosen because it provides wireless technology by using the WiFi protocol, which will send data in a serial manner. Arduino is a microcontroller device, and it is bolstered by the ATmega328P. The open-source NodeMCU microcontroller was created primarily for Internet of Things (IoT) applications. The ESP8266 is a common SOC (System on Chip) and Wi-Fi module-equipped Internet of Things (IoT) device. The IP/TCP protocol, which can connect with other devices over Wi-Fi, is supported by this module. Additionally, it may be controlled by a number of IDEs, including the Python IDE and the Arduino IDE.

DHT11 Sensor offers a resistive type humidity measurement tool and a negative temperature coefficient type temperature measurement tool that produce accurate results. This sensor is an integrated, four-pin, single-row, 8-bit microprocessor. Pin description: DHT11 sensor has 4 pins they are Vcc, Data, NC, and Gnd. A power supply is applied to the Vcc pin, it should be in the range of 3.3 to 3 V dc. The data pin will provide digital output, NC represents no connection, and at last, GND represents ground. The circuit connection of the DHT11 sensor is shown in Fig. 1. For this, different hardware and software components are required, like NodeMCU, jumper wires, a breadboard, a micro USB cable, and Arduino IDE software. The data pin is connected to any digital pin in the NodeMCU, Vcc to the power supply, and GND to the ground.

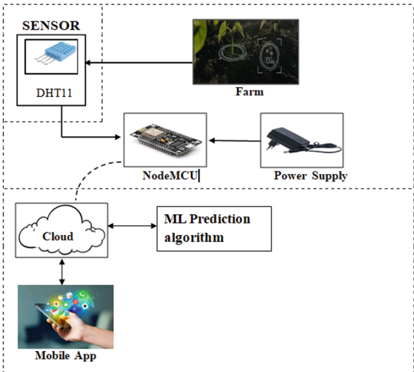

Fig. 1. Proposed IoT based crop recommendation model for smart agriculture

4 Crop Recommendation System

Nutrients in the soil like phosphorus, nitrogen, potassium, and other parameters like temperature, humidity, and rainfall play an important role in cultivation; a lack of these parameters might be one of the reasons for low crop yield. ML algorithms are used to recommend the best crop based on soil and atmospheric parameters. The proposed crop recommendation model is shown in Fig. 2, where trained model is to predict the outcome for new data which is provided through the mobile app and at the same time using DHT11 sensor temperature and humidity are read and passed to cloud which is used for predicting the outcome.

4.1 ML Based Crop Recommendation System

A branch of artificial intelligence known as "Machine Learning" tries to enable computer systems to learn from their experiences and grow without being explicitly programmed. Reinforcement learning, unsupervised learning, and supervised learning are the three categories into which Machine Learning algorithms often fall. Using a labeled dataset, the algorithm is trained to make predictions or classify new, unseen data in supervised learning. Given an unlabeled dataset, unsupervised learning requires the algorithm to identify patterns or correlations on its own. Reinforcement learning is teaching an agent how to behave in a way that maximizes a reward signal.

In the proposed work, pre-processing, FS, and model building are implemented. The proposed flow chart for ML based crop recommendation system is shown in Fig. 3.

Step 1: Importing required libraries- Import libraries like Scikit-Learn, Numpy, Pandas, Matplotlib, Seaborn, and mlxtend which are used to implement the Machine Learning algorithms. Fig. 4, presents a brief overview of the crop recommendation dataset. Temperature, humidity, rainfall, nitrogen, potassium, and phosphorus values are specific to each crop, and Fig. 5 shows the information of the dataset, which displays the number of rows, column names, the data type of each column, and memory usage.

Step 2: Data pre-processing- Kaggle data is collected and pre-processed. It is the initial step in Machine Learning. It is a data mining technique that is used to collect efficient and useful data from raw data. Since ML can't handle noisy data, which will decrease the model's performance. Data should be cleaned before being used in a classification algorithm. In this work, the isna() approach is used for identifying nan values.

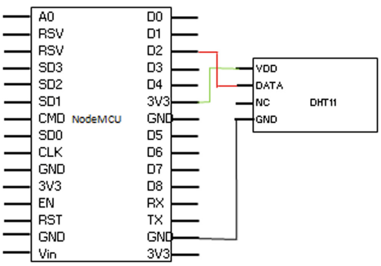

Fig. 2. Schematic of crop recommendation prorotype

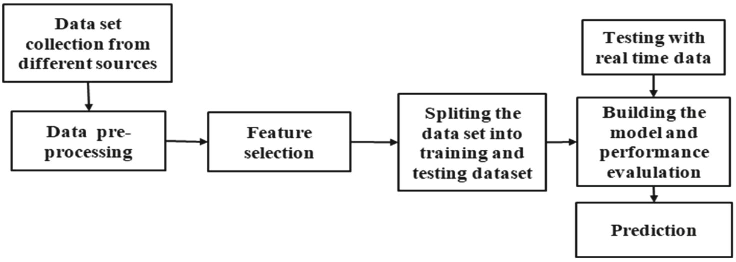

Fig. 3. ML based crop recommendation system

```
****************INFO OF CROP************
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   N            2200 non-null   int64
 1   P            2200 non-null   int64
 2   K            2200 non-null   int64
 3   temperature  2200 non-null   float6
 4   humidity     2200 non-null   float6
 5   ph           2200 non-null   float6
 6   rainfall     2200 non-null   float6
 7   label        2200 non-null   object
dtypes: float64(4), int64(3), object(1)
memory usage: 137.6+ KB
```

Fig. 4. Crop recommendation dataset

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
5	69	37	42	23.058049	83.370118	7.073454	251.055000	rice
6	69	55	38	22.708838	82.639414	5.700806	271.324860	rice
7	94	53	40	20.277744	82.894086	5.718627	241.974195	rice
8	89	54	38	24.515881	83.535216	6.685346	230.446236	rice
9	68	58	38	23.223974	83.033227	6.336254	221.209196	rice

Fig. 5. Information of dataset

Step 3: Feature selection- FS model is applied to identify necessary features and ignore irrelevant features. It is used in fast model training and also increases model accuracy. When necessary features are selected, then it will reduce over fitting. There are various FS models, like filter, embedded, and wrapper methods. Figure 6 shows the output of the FS method. In this work, the wrapper method is used for FS, in which features are combined to form different subsets, and these subsets of features are used to train a model. Evaluation simply means finding the performance metrics, i.e., the regression R^2 (coefficient of determination), and p-value. For classification accuracy, precision score, recall score, F1 score, etc. Finally, it selects a combination of features

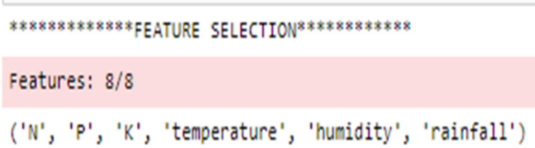

Fig. 6. Output of feature selection method

that provide better accuracy. This method takes more time than other FS methods, but the exhaustive feature selection method is one of the best feature selection methods in FS.

Step 4: Model planning and building- A supervised Machine Learning algorithm has classification and regression algorithms; regression is used for predicting a continuous output value, and the classification algorithm predicts categorical data. Since crop recommendation has categorical-based output, the classification algorithm comes under the supervised learning technique, which helps in classifying new observations based on the model's previous experience. In this case, model will first learn from the dataset and categorize new observations into various classes or groups. The model is evaluated with different classification algorithms and among those accuracy is more for boosting algorithms and bagging algorithms. Since these algorithm comes under ensemble learning [9], will help in enhancing our model performance by combines weak learners to form a stronger one. It mainly concentrates on prediction.

Step 5: Model Evaluation- This stage is crucial for determining the model's accuracy because it involves evaluating the model using several performance criteria. Accuracy, precision, and the confusion matrix are the three measures that are most frequently used to gauge classification performance. Precision is the proportion of expected positives that really turn out to be positive.

Confusion matrix for random forest classification algorithm is shown in Fig. 7. With the help of confusion matrix accuracy which measurement of how effectively it can predict the right output given an input. Formula for accuracy measurement is shown in Eq. 1, precision score formula is represented in Eq. 2, recall score's formula is shown in Eq. 3 and F1 score formula is mention in Eq. 4.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{1}$$

$$Precision = TP/(TP + FP) \tag{2}$$

$$Recall = TP/(TP + FN) \tag{3}$$

$$F1 = [(2 * Precision * Recall)/(Precision + Recall)] \tag{4}$$

where TP- True Positive

TN- True Negative
FP- False Positive
FN- False Negative

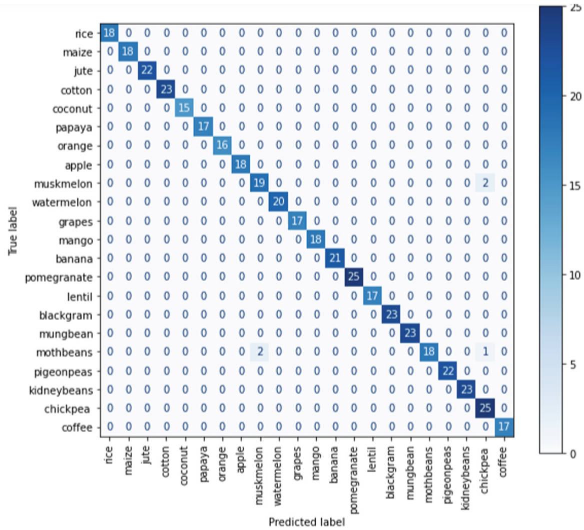

Fig. 7: Confusion Matrix

4.2 Firebase Cloud for Crop Recommendation Data Analysis

In this paper, input parameters like N, K, P, and rainfall are supplied via a mobile application to Firebase Cloud and the trained model is linked to the cloud using the Firebase [10] admin library. The data will be obtained from the cloud, and the Machine Learning model will then categorize the new data point and update the anticipated value to the cloud, which will be shown in the mobile app. The dataset are trained using the machine learning model as mentioned in Table 1.

4.3 Mobile App Creation Using Kodular Creator

Kodular Creator is housed on Google Cloud Platform. A free online toolbox is available for developing mobile applications [11]. It essentially provides an online drag-and-drop tool for building Android apps that anybody can use to create any kind of app without writing a single line of code. Design to provide a straightforward, and user-friendly. For designing the application drag the components needed and drop it in mobile screen and change the properties of each component according to the need like arranging labels in center, font size, providing text as illustrated in designing and for writing the logic it is done in blocks part. Design of crop recommendation model is shown in Fig. 8. After arranging all required components and changing the properties. For building a communication link between firebase cloud and mobile application there is an additional component firebase_database which will help in communicating with cloud by providing firebase credentials like database URL and secret key token. Algorithm for proposed crop recommendation system using kodular creator is shown in Fig. 9.

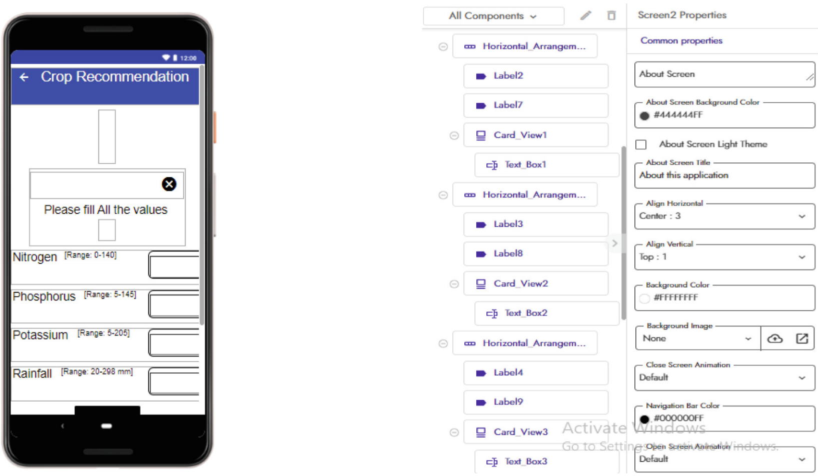

Fig. 8. Proposed crop recommendation app using kodular

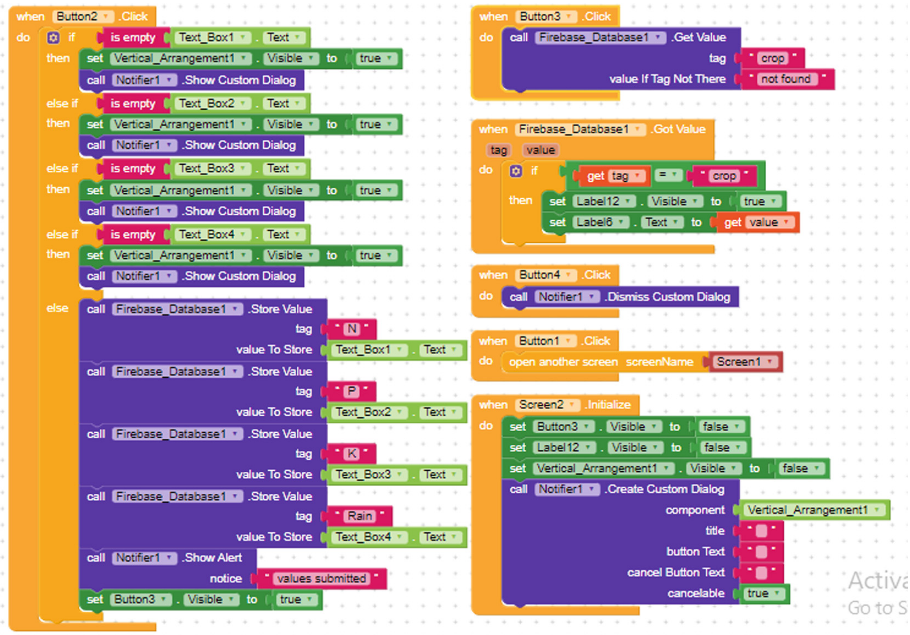

Fig. 9. Proposed crop recommendation system algorithm using kodular creator

5 Results and Discussion

A performance study of the model is provided in this section. The effectiveness of the created technique is assessed using metrics such as Confusion Matrix, Accuracy Score, Precision Score, Recall Score, and F1 Score. Finally, a smart phone app that suggests crops is developed using the Kodular Creator platform. Additionally, climate factors are also taken into consideration. A pop-up notification will be displayed as illustrated in Fig. 10 when any input parameter is not provided by the user. Table 1 displays the accuracy of the model's evaluation using standard Machine Learning classifiers like K-Neighbors, Decision tree, GaussianNB, Support vector, Gradient Boosting, Random forest, Extra tree, LGBM and XGB classifier. From these classifiers accuracy for crop recommendation scores greater than 95%. Figure 11 shows the recommend crop.

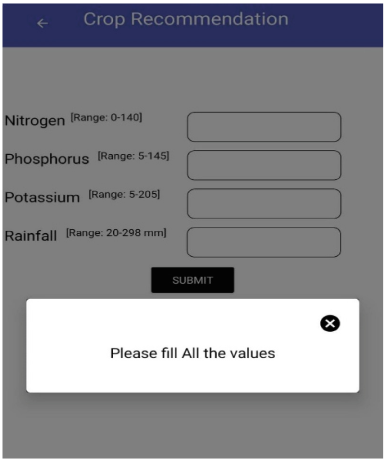

Fig. 10. Pop-up notification for missing fields

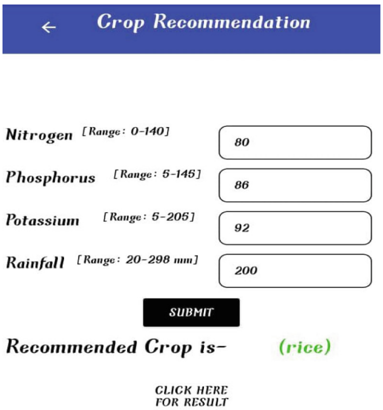

Fig. 11. Recommends the suitable crop for provided inputs

Table 1. Model performance details for crop recommendation

ML Algorithm	Accuracy
K-Neighbors Classifier	0.9772727
Decision Tree Classifier	0.9977272
SVC (Support Vector Classifier)	0.9818181
GaussianNB	0.9909090
Random Forest Classifier	0.9977272
Gradient Boosting Classifier	0.9886363
Extra Trees Classifier	0.99318181
LGBM Classifier	0.99545454
XGB Classifier	0.99318181

6 Conclusion

This proposed system suggests a solution to the rising number of farmer suicides and to aid in their financial development. The crop recommender system assist farmers in making decisions regarding which crop to cultivate Machine Learning tools were used to gather, analyses, and train relevant datasets. This system benefits the agricultural industry. The system's recommends user to choose the suitable crop, additionally, the system offers, and a user-friendly application where the input parameters are provided by users through mobile application which is processed by backend and show results but there are limited number of features in kodular creator platform which makes the application inflexible. In the future by providing GPS coordinates for the field and gaining access to the government's rain forecasting system, we can gather all the necessary data and anticipate crops simply by providing the coordinates.

References

1. Gupta, Akanksha, and Priyank Nahar. "Classification and yield prediction in smart agriculture system using IoT." Journal of Ambient Intelligence and Humanized Computing (2022): 1–10.
2. D. J. Reddy and M. R. Kumar, "Crop Yield Prediction using Machine Learning Algorithm," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1466–1470, doi: https://doi.org/10.1109/ICICCS51141.2021.9432236.
3. Lontsi Saadio Cedric, Wilfried Yves Hamilton Adoni, Rubby Aworka, Jérémie Thouakesseh Zoueu, Franck Kalala Mutombo, Moez Krichen, Charles Lebon Mberi Kimpolo. "Crops yield prediction based on machine learning models: Case of West African countries" Elsevier (2021).
4. D. Naga Swetha, Savadam Balaji. "Agriculture Cloud System based Emphatic Data Analysis and Crop Yield Prediction Using Hybrid Artificial Intelligence." International Conference on Physics and Energy (2021).
5. Amine Dahane, Rabaie Benameur, Kechar Bouabdellah, Abou El Hassan Benyamina. "An IoT Based Smart Farming System Using Machine Learning." International Symposium on Networks, Computers and Communications (2020).

6. Sonal Agarwal and Sandhya Tarar, “Am hybrid approach for crop yield prediction using machine learning and deep learning algorithm.” Institute of Physics Publishing (2020).
7. Yanping Wang, Zongtao Chi. “System of Wireless Temperature and Humidity Monitoring Based on Arduino Uno platform” (2016).
8. Priyadharshini A, Aayush Kumar, Swapneel Chakraborty, and Omen Rajendra Pooniwala. “Intelligent Crop Recommendation System using Machine Learning”, IEEE Xplore (2021).
9. Scikit-learn: Machine Learning in Python (cite for ML algorithms).
10. https://firebase.google.com (Site link for Firebase Cloud).
11. https://creator.kodular.io (link for Kodular Creator).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.