

CROP YIELD PREDICTION USING HYBRID MACHINE LEARNING MODEL**K. Johny Elma**

Assistant Professor
Department of Information Technology
Easwari Engineering College
Chennai,India

Sree Kayathri S

Department of Information Technology
Easwari Engineering College
Chennai,India

Yukttha Sree R S

Department of Information Technology
Easwari Engineering College
Chennai,India

Santhosh Kumar R

Department of Information Technology
Chennai,India
Easwari Engineering College

Abstract- Crop yield prediction is a crucial task for farmers to optimize their resources and maximize productivity. This paper proposes a hybrid machine learning and Internet of Things (IoT) approach for crop yield prediction that utilizes Ph, rainfall, temperature, humidity, and NPK sensors. The system combines data from these sensors to provide accurate predictions for crop yield. The IoT-based data collection and processing platform ensures real-time and accurate data collection. The proposed system utilizes machine learning models such as Random Forest, Decision Tree, and Artificial Neural Network to predict the crop yield based on the collected sensor data. The accuracy of the proposed approach is evaluated through experiments, which show that it can achieve high accuracy in crop yield prediction. The results demonstrate the effectiveness of the proposed approach in enhancing the agricultural industry's productivity. The proposed system can help farmers make informed decisions about irrigation, fertilization, and other farming practices. By using the Ph, rainfall, temperature, humidity, and NPK sensors, the system can assist in reducing the use of resources and minimizing waste. The study concludes that the hybrid machine learning and IoT approach with these sensors can provide an efficient and effective solution for crop yield prediction, helping farmers to optimize their resources and maximize their productivity. Overall, the proposed system has the potential to revolutionize the agriculture industry, leading to sustainable and efficient farming practices. The use of the Ph, rainfall, temperature, humidity, and NPK sensors can provide more accurate and relevant data for crop yield prediction, leading to better decision-making and higher yields for farmers.

Keywords—Crop yield, Random Forest, Machine learning

I.INTRODUCTION

Farming is an important part of the economy, and crop yield prediction plays a crucial role in helping farmers optimize their resources and maximize productivity. Historically, crop yields have varied mainly in response to changes in precipitation, humidity, temperature, soil composition, and the effects of global warming on farming. The current crop yield prediction algorithms often have low

accuracy and insufficient knowledge about soil nutrient content, leading to suboptimal decision-making by farmers. To address this issue, this project proposes a solution that combines the power of machine learning techniques and soil sensors such as Ph, rainfall, temperature, humidity, and NPK sensors to accurately forecast crop yields. The system will have a simple user interface (UI) to assist farmers in forecasting their crop yields based on the collected sensor data. The system will utilize machine learning models such as Random Forest, Decision Tree, and Artificial Neural Network to predict the crop yield based on the collected sensor data. The main objective of this project is to provide accurate crop yield predictions for farmers, enabling them to make informed decisions about their planting and harvesting practices. The proposed system will reduce financial losses faced by farmers who plant the wrong crops at the wrong time and evaluate a suitable crop for the area, which will boost the economy. The accuracy of the machine learning algorithm will depend on the parameters used and the correctness of the dataset. Overall, this project aims to revolutionize the agriculture industry by providing an efficient and effective solution for crop yield prediction using machine learning techniques and soil sensors. By utilizing the power of technology, we can enhance agricultural productivity, reduce waste, and promote sustainable farming practices.

The purpose of this paper is to determine the most accurate crop prediction program that can help farmers choose what kind of crop to produce depending on meteorological factors and soil nutrients. As a result, it is viewed as a potential enabling technology for the phenomena of the next generation of application-driven networks, which is growing and giving rise to a wide range of intriguing applications and new commercial opportunities. The accuracy of a system gaining knowledge of set of rules can also additionally depend upon the parameters used and to the quantity of correctness of the dataset (Dr. P.M. Durai Raj Vincent 2020).. This system tends to reduce the financial losses faced by the farmers as they plant the wrong crops at wrong time and also evaluate a suitable crop for the place which itself becomes a boost for the economy.

Reasons for Choosing Machine Learning and IOT Approach

There are several reasons why the machine learning and IoT approach was chosen for crop yield prediction. Firstly, machine learning models have proven to be effective in predicting crop yields based on historical data, and can help farmers make informed decisions about their planting and harvesting practices. Machine learning algorithms can analyse large datasets, identify patterns, and make predictions with high accuracy, making them an ideal choice for crop yield prediction.

Secondly, IoT sensors such as Ph, rainfall, temperature, humidity, and NPK sensors can provide real-time and accurate data about the environmental conditions of crops. This data can be used to improve crop yield predictions, optimize resource use, and reduce waste. The use of IoT sensors can also help farmers monitor the health of their crops, detect pests and diseases early, and take necessary actions to prevent losses.

Furthermore, the combination of machine learning and IoT can create a powerful system for crop yield prediction. Machine learning algorithms can analyse the large amounts of data collected by IoT sensors, identify patterns, and make accurate predictions. This can provide farmers with a complete and detailed picture of the environmental conditions and soil nutrient content of their crops, allowing them to make better decisions and maximize their productivity.

Finally, the machine learning and IoT approach can provide an efficient and effective solution for crop yield prediction. By using technology to analyse and process data, farmers can reduce their reliance on guesswork and intuition, and make informed decisions based on accurate and relevant data. This can lead to improved crop yields, reduced waste, and more sustainable farming practices.

In summary, the machine learning and IoT approach was chosen for crop yield prediction due to its ability to analyse large amounts of data, provide real-time and accurate information, and create a powerful system for predicting crop yields. This approach has the potential to revolutionize the agriculture industry by enabling farmers to make more informed decisions, leading to increased productivity and sustainability.

II. LITERATURE SURVEY

(Aruvansh Nigam 2019), The paper's main objective was to predict agricultural yield using several machine learning algorithms. The output of this technique is compared with mean absolute error. The algorithms like RNN, LSTM are used to get the accurate result this helps the farmer to know which crops to grow to get benefited.

In this study (Aruvansh Nigam 2019), the authors propose using a deep reinforcement learning model to accurately estimate crop yields. The crop feature that aids in prediction is extracted using deep-learning based methods. The RL

algorithm, a collaborative agent learning by contact with the user, makes use of the Q learning technique.

The authors (Potnuru Sai Nishant 2020) of this publication, present a discussion on the subject of predicting crop yields using machine learning. Almost any type of Indian crop may be predicted using this app. In order to complete its goals, this research employs cutting-edge algorithms including Kernel Ridge, Rasso, and Enet. Regression stacking is used to boost prediction accuracy. With this project, root-mean-squared performance is being tracked. To get a prediction result, the user can enter the value of metrics.

In this paper, the authors (Dhivya Elavarasan 2019) have made research over the prediction of crop yield by using the attributes rainfall and temperature and applied ARMA, SARIMA and ARMAX models for accuracy and then fuzzy logic is implemented once the prediction has made. This model works by taking an input of previous year crop yield to predict the upcoming year crop yield.

Using the SVM (support vector machine) algorithm, writers (Niketa Gandhi Et Al 2016). suggest a system to estimate rice crop yield in India. The Sequential Minimum Optimization (SMO) classifier was applied in the WEKA software to the provided datasets. The analytical results were factored into the final conclusions since they are better at estimating the rice crop production.

In this study, the authors offer a method for crop prediction that takes into account soil and weather conditions in order to determine which crops would thrive in a given environment. Here, an ANN (artificial neural network) is employed to get the desired effect. The network learns from samples of inputs and outputs it intends to compute using a back propagation technique, which is trained via supervised learning.

In this paper, the authors has proposed the research on crop yielding prediction technique by data mining process which is widely applied to agricultural problems (Miss.Snehal S. Dahikar 2019) It is applied to the task of classifying and finding patterns in massive datasets. The MLR method was used to develop this model for predicting agricultural yields. The authors of this research suggest a CNN-RNN framework for predicting crop yield. The model, which incorporates CNNs, fully connected layers, and RNNs, is a hybrid one. Using an FC, the W-CNN and S-CNN aspects of the weather and soil conditions were merged (completely connected layer). The RNN model is applied to the multi-year time series data of agricultural yield.

In this paper, the researchers (B.Manjula Josephine, Et Al 2018). have proposed the system for crop Yield Prediction Using Machine Learning. crop yield dataset is inputted and data pre-processing techniques is involved to transformed the raw data into understandable format. The random forest approach is then utilized, which requires data validation and forecasting in order to produce appropriate results. Dimensional reduction is performed to lower the number of random variables.

The authors of this article present a model for estimating future harvests. Predictions of crop yield were made using data mining to ensure maximum crop productivity. After cleaning and clustering the raw data with the k-means technique, we were able to utilize it to make educated predictions about crop productivity. generated rules. The district and crop is input to this prediction mode.

This study suggests combining filter and wrapper strategies for feature selection. (Dr. P.M. Durai Raj Vincent 2020) Combining soil, crop, and meteorological data produces ideal features for a crop recommendation algorithm. By comparing dataset features and recommended approach features, we may evaluate model performance. The suggested feature selection technique is tested using MSE, RMSE, MAE, and R 2. Using the given features, artificial neural networks and decision trees are formed.

The study recommends 22 crops utilizing correlation, distribution, assembly, and majority vote. Three-tiered architecture for crop recommendation. (R K Ray 2022) It's preprocessed, classified, and evaluated. Correlation plots and density distribution classify features. Majority vote determines performance. This article uses ensembling with base learners, such as decision trees, random forests, and Naive Bayes. Majority vote determines performance metrics. Visualize correlation, density-histogram, confusion, and performance charts. Naive Bayes post-implementation accuracy is 99.54%. It's 98.52%. Thus, Nave Bayes classifier is better. This article discusses issues and future research.

For crop yield prediction, advanced technologies such as data mining techniques, machine learning techniques, remote sensing and image processing, etc., can be utilized (R K Ray 2022). Using various parameters such as temperature, precipitation, soil parameters, area sown, etc., In predicting agricultural yield, machine learning techniques are crucial. This study investigated different supervised and unsupervised machine learning approaches to forecast agricultural yield. When predicting multiple crop yields with various parameters, classification techniques have been discovered to be more reliable than regression and unsupervised learning techniques.

Random Forest, classification and regression activities can be carried out using the most well-liked and effective supervised machine learning technique. They are used in different crops to minimise crop production output losses despite potentially distracting environmental conditions (Ranjani 2021). The sustainability of agriculture has been seriously threatened by changes in the weather, climate, as well as other environmental issues. Machine learning (ML) is crucial because it offers Crop Yield Prediction (CYP) a decision-support tool that can help with decisions like which crops to farm and what to do during the crop's growth season. Crop yield estimation uses a number of well-recognized methods to boost agricultural crop production, which is its main goal.

Currently, software can supplement traditional knowledge. Accurate record-keeping is crucial to agricultural risk

management. We suggest creating a crop yield production forecast model with the aid of machine learning techniques. We (Neha Rale 2019) assess the effectiveness of several both linear and nonlinear recursive models using 5-fold cross validation. With basically default parameters, the random forest regression model delivered the best results., followed by nearest-neighbor regression, L 2 linear regression with polynomial features, and support-vector regression with a Radial Basis Function (RBF) kernel.

As a result, machine learning techniques have taken over the task of prediction in recent years, and in this study, a number of these techniques were used to predict crop yield. To ensure that a particular machine learning (ML) model performs with a high degree of accuracy, to convert unprocessed data into a dataset that is suitable for machine learning, it is crucial to employ efficient feature selection strategies. Consequently, efficient feature selection is required to guarantee that the model includes only the most pertinent features. Our model will be unnecessarily complicated if ever y feature extracted from raw data is aggregated without consideration of their role in the model-building process.

Agriculture is one of India's most prevalent and lowest-paying industries. (A Nigam 2019) Machine learning can bring about a boom in the agriculture industry by altering the income scenario through the cultivation of the optimal crop. This paper focuses on predicting crop yield using a variety of machine learning techniques. The results of these methods are compared based on their mean absolute error. The predictions made by machine learning algorithms will assist farmers in deciding which crop to cultivate for maximum yield, taking into account factors such as temperature, precipitation, area, etc.

This study forecasts the yield of almost all crops grown in India. (Shreya 2018) By utilizing simple parameters such as State, district, season, and area, this script enables the user to predict the crop yield for any given year. The paper utilizes advanced regression techniques such as Kernel Ridge, Lasso, and ENet algorithms to predict the yield and employs the concept of Stacking Regression to improve the algorithms and provide a more accurate forecast.

III. EXISTING SYSTEM

As more people become aware of the issue in Indian agriculture, more researchers are investing time and effort to finding solutions. Regularized Greedy Forest is used by many different works to select the optimal crop sequence for a given time stamp. A different approach is to use the past's worth of weather records as a training set for the model. And this is used to predict the yield of the crop.

To identify the characteristics that prevent apples from flourishing, a model is educated. The monthly weather trends are then used to make an accurate forecast of apple production. Crops are selected using various methods, such as Artificial Neural Networks, K Nearest Neighbors, and Regularized Greedy Forests, to maximize the expected yield under varying environmental and agronomic

conditions. It also includes pesticide forecasting and online trade based on agricultural commodities.

All these significant studies had the same issue, in our opinion, which was that the authors relied too heavily on a single factor (weather or soil) to make predictions about the optimal conditions for crop growth. But we think that both of these factors should be considered together for the best and most accurate prediction. The reason behind this is that while a specific soil type may be good for growing a certain crop, the yield will decline if the local climatic circumstances are unfavorable to that crop type.

IV. PROPOSED SOLUTION

The agricultural industry is in need of a solution that provides accurate and real-time insights into crop yield predictions, and IoT sensors are at the forefront of this technological revolution. Our proposed system combines the latest technology, including Ph sensors, rainfall sensors, temperature sensors, humidity sensors, and NPK sensors, to collect data about crop environments. Machine learning algorithms such as random forest, decision tree, and artificial neural network are then used to analyse this data and predict crop yields.

The system's easy-to-understand user interface provides farmers with crop yield predictions, a reporting system, and an alert system that notifies farmers of any issues. This flexible and scalable system can adjust to different crops and environmental conditions, ultimately revolutionizing the agricultural industry by providing real-time insights to farmers.

The use of IoT sensors can optimize resources, increase productivity, and minimize waste, contributing to sustainable and efficient farming practices. By improving the agricultural industry's productivity, this solution can help ensure food security for the growing global population. The proposed system has the potential to revolutionize the agriculture industry by providing accurate, real-time insights to farmers, enabling them to make informed decisions and optimize their resources for maximum productivity.

V. ARCHITECTURE DIAGRAM

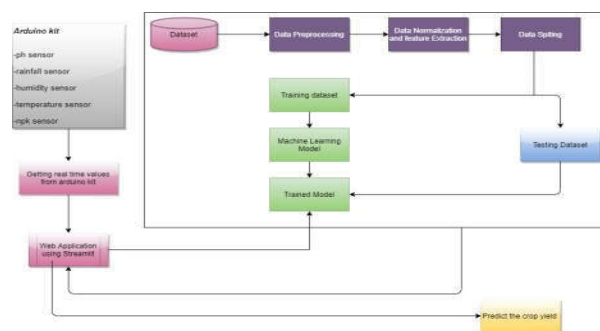


Fig 1.1 Architecture Diagram

The IoT sensors installed on the farm include Ph sensors, rainfall sensors, temperature sensors, humidity sensors, and Npk sensors, which collect data in real-time. The data collected by the sensors is pre-processed to remove noise and redundant information. The architecture diagram for the machine learning project is mentioned in Fig 1.1. The preprocessed data then is fed to the machine learning algorithms, such as random forest, decision tree, and artificial neural network, to predict crop yield. As a result, when the user gives the input parameters, the model predicts the suitable crop for the area.

V. Pseudo code

Phase of learning: Create a training instance data set

Classification section:

For each unidentified instance, x_n

Identify x_1, x_2, \dots, x_n obtained from the data set using ML algorithms.

Set class label until it equals the class that occurs most frequently.

Return type;

End

VI. METHODOLOGY

In the proposed system for precision agriculture, the machine learning model is trained to predict crop yields and make recommendations about fertilizers. One of the major challenges in precision agriculture is predicting crop yields, which requires the use of real-time data from various sources such as climate, weather, soil, temperature, and seed type.

The proposed system in our task includes 3 modules. The first module is data collection and preprocessing, where data from various sources such as Ph sensors, rainfall sensors, temperature sensors, humidity sensors, and NPK sensors are collected and preprocessed. This module also includes data cleaning, feature selection, and data transformation to prepare the data for model training.

The second module is model training, where machine learning models such as random forest, decision tree, and artificial neural network are trained on the preprocessed data to predict crop yields and recommend fertilizers. The models are optimized for accuracy using techniques such as cross-validation and hyperparameter tuning.

The third module is the web application, where farmers can access the predictions and fertilizer recommendations through an easy-to-use interface. The web application also includes features such as visualization of historical data, comparison of different crop yields, and alerts for adverse weather conditions.

Overall, the proposed system has the potential to revolutionize precision agriculture by providing accurate and real-time predictions of crop yields and fertilizer recommendations. By leveraging the power of machine learning and real-time data, this system can help farmers optimize their resources, increase productivity, and minimize waste, contributing to sustainable and efficient farming practices.

A. Data Collection & Preprocessing

For this project, IoT sensors are deployed on the farmland, comprising of various types of sensors, namely Ph, rainfall, temperature, humidity, and Npk sensors, that constantly gather real-time data. The collected data undergoes preprocessing to eliminate irrelevant or duplicated data, before feeding it into machine learning models like random forest, decision tree, and artificial neural network, for the prediction of crop yield.

It was important to choose the parameter with which we were supposed to work with on the project. Those parameters are humidity levels, average temperature conditions, as well as sodium, and potassium found in the soil. The next major step is data splitting. The preprocessing work is shown in the Fig 1.2.

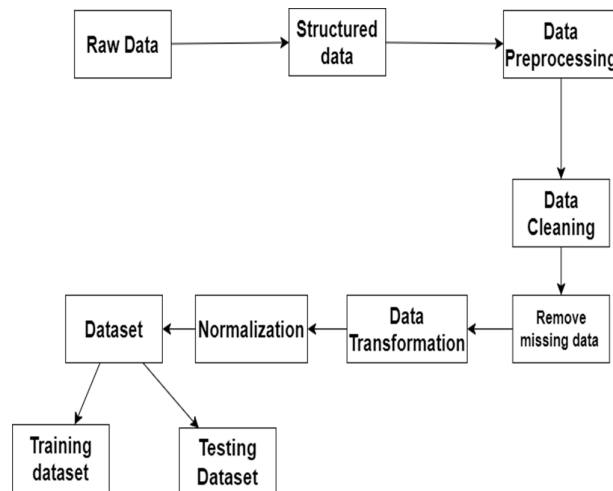


Fig 1.2 Dataset Collection

We will just split our data into 2 segments - training and testing data in the ratio of 80:20 ratio. The machine learning model is trained using the training dataset according to the machine learning algorithms. Once the model is trained using the 20% of the test data to check if model is trained well. It is crucial for us to train our model on the selected algorithm. We are using logistic regression and random forest. We do this step and then hyper tune our models so that the best models can be chosen to give maximum efficiency.



Fig 1.3 NPK sensor

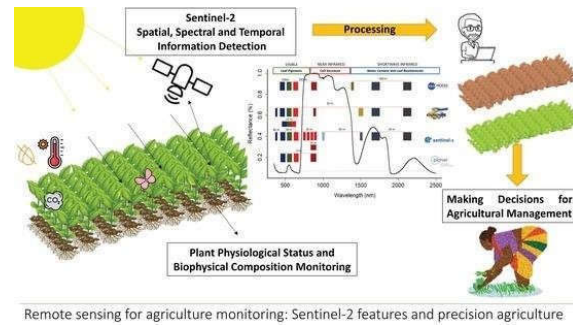


Fig 1.4 Precision agriculture data preprocessing

B. Model Training

In this module, the following steps are performed:

C. Splitting the Dataset

Training Set: The data set we feed into our model for learning potential underlying models and relationships. The training set is the set of data we use to create our model. It's that set of data that our model uses to learn the underlying patterns or relationships that will allow predictions to be made later.

Test Set: The set of data we use for estimating the unbiased accuracy of our model in nature. Once we have used the validation set to determine the algorithm and the selection of parameters we want to use in production, the test set is used to estimate the real performance of models in the wild.

The Working process of the random forest algorithm applied on training dataset.

- Step-1: From the training set, choose K data points at random.
- Step-2: Create the decision trees linked to the chosen data points (Subsets).
- Step-3: For any decision trees you intend to construct, select N.
- Step-4: Repeat Step 1 & 2.
- Step-5: Locate each decision tree's forecasts for new data items, then place them in the category that receives the most votes.

C. Arduino Kit

The ph sensor, rainfall sensor, temperature sensor, humidity sensor and npk sensor are the following sensors which are connected to the Arduino board. After inserting the sensors to the soil the respective values are displayed in the Arduino LED display.

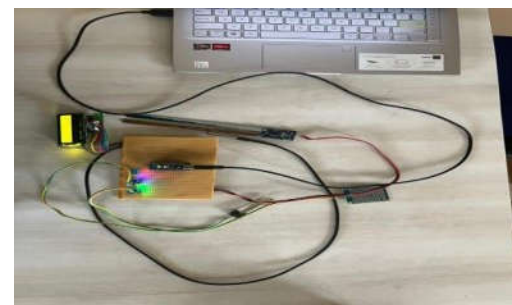


Fig 1.5 Arduino kit

D. Web Application

In this project a webapp is built using streamlit that will take multiple parameters as input and will suggest crops based on the climate and all. A crucial aspect of developing a machine learning model is sharing the model with others. No matter how many models we create, if they remain offline, very few individuals will be able to observe our accomplishments. Therefore, we must deploy our models so that anyone can interact with them via a pleasant User Interface (UI). This system's user interface is a single-page web application built with Flask. It will take input and predict whether or not the user has a chance of developing chronic heart disease within the next decade. A tiny web framework based on Python is called Flask. As it doesn't require any specific tools or libraries, it is referred to as a microframework. It doesn't have a form validation layer, database abstraction layer, or any other feature that is already covered by third-party libraries.

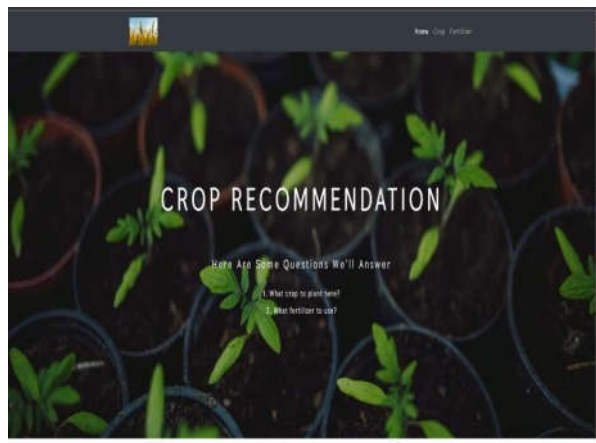


Fig 1.6 Home page

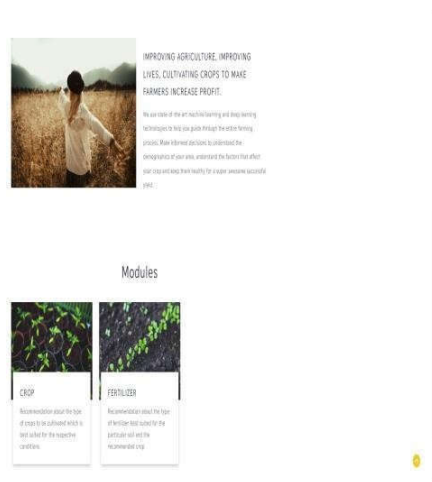


Fig 1.7 Home page



Fig 1.8 Arduino connected to the soil



Fig 1.9 Arduino output

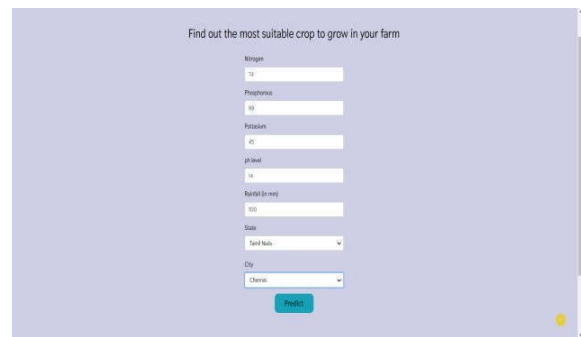


Fig 1.10 Crop prediction input page

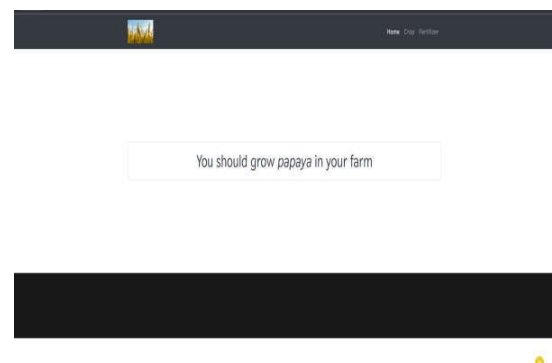


Fig 1.11 Crop prediction result

Fig 1.12 Fertilizer recommendation input page

Fig 1.13 Fertilizer recommendation output

VII.

RESULTS

The suggested method has a great deal of potential to assist farmers in forecasting agricultural yields, which have a significant impact on the nation's economy. Future call for the ranching equipment, such as the tractor, to be connected to the internet, which will allow ranchers to receive information gradually regarding yield gathering and potentially contaminated harvests, assisting them in taking the appropriate action. Additionally, the most fruitful harvest may be determined by taking the financial and expansion ratio into account. Users are urged to submit information in this application, such as the temperature, so that the prediction process may begin when the location of the user is determined automatically. Also, the future endeavors include the suggestion of the crops that has to be grown in the particular type of field and the most suitable fertilizer that will further increase the yield.

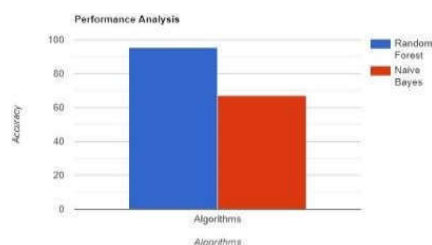


Fig.1.14 Random forest and Naïve Bayes comparison

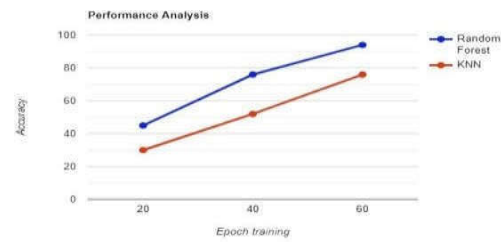


Fig.1.15 Random forest and KNN comparison

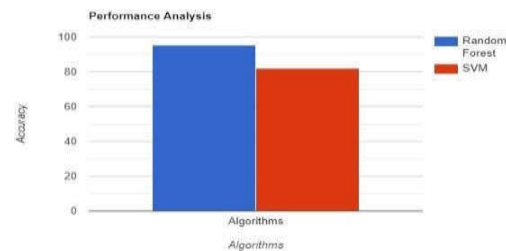


Fig.1.16 Random forest and SVM comparison

VII. CONCLUSION

This system helps the farmer to choose the right crop by providing insights that ordinary farmers don't keep track of thereby decreasing the chances of crop failure and increasing productivity. It also prevents them from incurring losses. The system can be extended to the web and can be accessed by millions of farmers across the country.

REFERENCES

- [1] Shivam Bang, Akshaya Kumar Dixit Rajat Bishnoi, Indu Chawla, Ankit Singh Chauhan, "Predicting Crop Yield using Fuzzy Logic using ARMA, SARIMA, and ARMAX Model Predicted Temperature and Rainfall Parameters, IEEE-2019
- [2] Crop yield prediction in Indian agriculture using machine learning. Aruvansh Nigam, Saksham Garg, Parul Agarwal, Archit Agarwal.
- [3] Crop Yield Prediction based on Indian Agriculture using Machine Learning" by Potnuru Sai Nishant, Pinapa Sai Venkat, and Bollu Lakshmi Avinash B.
- [4] Crop yield prediction using a deep reinforcement learning model for environmentally responsible agricultural applications. Dhivya Elavarasan and P. M. Durairaj Vincent. IEEE.
- [5] Involving Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong, and Amiya Kumar Tripathy Reference: IEEE, "Rice Crop Yield Prediction in India Using Support Vector Machines,"

[6] Miss.Snehal S. Dahikar and Dr.Sandeep V.Rode's article "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach" published in the international journal of creative research

[7] D. Ramesh and B. Vishnu Vardhan, Analysis of crop yield prediction utilising data mining techniques, International Journal of Research in Engineering and Technology

[8] Crop yield prediction using machine learning; B.Manjula Josephine; K.Ruth Ramya; K.V.S.N. Rama Rao; Swarna Kuchibhotla; P.Venkata Bala Kishore; S. Rahamathulla; international journal of scientific & technology research

[9] M. J. and I. M., "Hybrid filter and wrapper approaches-based feature selection for crop recommendation." 2020 International Conference on Electronic Systems and Intelligent Computing (ICESIC),

[10] R. K. Ray, S. K. Das, and S. Chakravarty, 12th International Conference on Cloud Computing, Data Science, and Engineering (Confluence), 2022. Smart Crop Recommender System-A Machine Learning Approach

[11] S. G L, N. V and S. U, "A Review on Prediction of Crop Yield using Machine Learning Techniques," 2022 IEEE Region 10 Symposium (TENSYP), Mumbai, India, 2022, pp. 1-5, doi: 10.1109/TENSYP54529.2022.9864482.

[12] N. Rale, R. Solanki, D. Bein, J. Andro-Vasko and W. Bein, "Prediction of Crop Cultivation," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019, pp. 0227-0232, doi: 10.1109/CCWC.2019.8666445.

[13] S. P. Raja, B. Sawicka, Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers," in IEEE Access, vol. 10, pp. 23625-23641, 2022, doi: 10.1109/ACCESS.2022.3154350.

[14] A. Nigam, S. Garg, A. Agrawal and P. Agrawal, "Crop Yield Prediction Using Machine Learning Algorithms," 2019 Fifth International Conference on Image Information Processing (ICIIP), Shimla, India, 2019, pp. 125-130, doi: 10.1109/ICIIP47207.2019.8985951.

[15] S. V. Bhosale, R. A. Thombare, P. G. Dhemey and A. N. Chaudhari, "Crop Yield Prediction Using Data Analytics and Hybrid Approach," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697806.