A Web Based Crop Recommendation System Using Various Machine Learning Algorithms

Monisha H M
Dept. of Information Science
B.M.S. College of Engineering
Engineering Bangalore, India
monishahm.ise@bmsce.ac.in

Dhanush Jogi K
Dept. of Information Science
B.M.S. College of Engineering
Bangalore, India
dhanushjogi.is20@bmsce.ac.in

Adarsha Dalavai S M
Dept. of Information Science
B.M.S. College of Engineering
Bangalore, India
adarshadalavai.si20@bmsce.ac.in

Manvith Dalli
Dept. of Information Science
B.M.S. College of
Bangalore, India
manvithreddy.is20@bmsce.ac.in

***Abstract*—** Crop recommendation is a crucial aspect of agriculture, aiding farmers in making informed decisions for optimal yield and profitability. The comparative study of five machine learning algorithms—Logistic Regression, One-vs-All, Decision Trees, Bagging Classifier, and Random Forest—for crop recommendation. The study highlights the significance of crop recommendation in addressing the challenges faced by farmers, such as uncertain weather conditions, soil variability, and market demands. By leveraging machine learning techniques, accurate and efficient crop predictions can be made, aiding farmers in selecting the most suitable crops for their specific conditions. These findings offer insights into the suitability of different machine learning algorithms for crop recommendation. Agricultural stakeholders can leverage this knowledge to make informed decisions regarding the adoption of suitable algorithms for developing efficient crop recommendation systems using the website.

***Index Terms*—** Agriculture, Crop recommendation, Machine learning(ML), Predictive modeling.

I. INTRODUCTION

Crop recommendation is a critical aspect of modern agriculture, aiding farmers in making informed decisions about the selection of suitable crops based on specific conditions. Machine learning algorithms have emerged as powerful tools in crop recommendation systems, offering data-driven approaches to optimize agricultural productivity and resource allocation. This report provides a comprehensive overview of the suitability of different machine learning algorithms, including logistic regression, one-vs-all, decision trees, bagging classifier, and random forest, for crop recommendation.

This model aims to recommend the most suitable crop based on input parameters like Nitrogen(N), Phosphorous(P), Potassium(K), pH, Humidity, Rainfall and temperature. This paper predicts the accurate crop for input parameters for higher yield and profit. It has 20 different crops such as Kidney Beans, Pigeonpeas, Mothbeans, Mungbean, Blackgram, Lentil, Pommegranate, Banana, Mango, Grapes, Watermelon, Muskmelon, Apple, Orange, Papaya, Coconut, Cotton, Jute, Maize and Rice. This proposed system applied different types of Machine Learning Algorithms such as Logistic Regression, Decision trees and Random Forest.

Logistic regression is a widely used algorithm for binary classification tasks. In the context of crop recommendation, logistic regression can be employed to predict the suitability of a specific crop for a given set of input parameters. By learning the relationships between various agricultural factors and crop outcomes from historical data, logistic regression can provide probabilistic predictions and assist farmers in selecting appropriate crops. Its simplicity and interpretability make it a practical choice, particularly when dealing with binary crop recommendation scenarios.

Decision trees are versatile algorithms that can handle both classification and regression tasks. They are particularly well-suited for crop recommendation due to their ability to capture complex relationships between various agricultural parameters. Decision trees recursively split the data based on different features, creating a hierarchical structure that represents decision rules. By traversing the tree, the algorithm can determine the recommended crop based on the input parameters. Decision trees offer interpretability, enabling farmers to understand the reasoning behind the recommended crop choice.

Random forest is an extension of the bagging classifier algorithm that introduces additional randomness by selecting a random subset of features at each node of the decision trees. By further diversifying the individual trees, random forest improves the predictive power and reduces the correlation between trees. In crop recommendation, random forest can effectively handle high-dimensional datasets with numerous features, providing accurate and reliable recommendations based on the collective knowledge of multiple decision trees.

By exploring and evaluating the performance of these machine learning algorithms, agricultural stakeholders can gain insights into their suitability for crop recommendation. The comparative analysis of these algorithms will enable farmers, policymakers, and agricultural technology developers to make informed decisions regarding the adoption of specific algorithms based on their requirements and the available data. Ultimately, the application of machine learning in crop

recommendation holds the potential to enhance agricultural practices, optimize resource utilization, and contribute to sustainable and efficient farming systems.

II. RELATED WORK

[1] Dhruvi Gosai, Chintal Raval, Rikin Nayak, Hardik Jayswal, Axat Patel have proposed a system which considers past historical data which includes parameters like Nitrogen (N), Phosphorous (P), Potassium (K), PH value of soil, Humidity, Temperature, and Rainfall is used to recommend suitable crop to the farmer. The dataset contains around five different crops. This crop prediction can be done by Decision Trees, Naïve Bayes (NB), Support Vector Machine (SVM), Logistic Regression, Random Forest (RF), and XGBoost .The Naïve Bayes algorithm gave an accuracy of around 99%. By applying XGBoost algorithm got best accurate value result of around 99.31% accuracy. More accurate results gave more profit to the crop yields.

[2] Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H, W.J.M. Nanayakkara, Dimantha M.A.C, Pabasara M.G.P. have proposed a system which uses a Arduino board to collect data on several parameters like pH, moisture, humidity, temperature, erosion to predict suitable crop. The dataset contains around five different crops. This crop prediction can be done by Native Bayes Classifier, Support Vector Machine (SVM). By applying Native Bayes algorithm aquired higher productivity and accuracy of around 96%.

[3] Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, Avinash Devare have proposed system which uses several parameters like pH, depth, texture, water holding capacity, erosion to predict suitable crop. This proposed system worked on various crops like groundnut, pulses, sorghum, sugarcane, coriander. This crop prediction can be done by Decision Trees, Support Vector Machine (SVM), Naïve Bayes, Random Forest (RF), Artificial Neural Network(ANN). By applying XGBoost algorithm got best accurate value result. This paper would help farmers to increace productivity and thereby increase profit.

[4] Anguraj.K , Thiyaneswaran.B , Megashree.G , Preetha Shri.J.G , Navya.S , Jayanthi. J have proposed a system which uses a IoT device to collect data on several parameters like pH, moisture, humidity, temperature, erosion to predict suitable crop.They developed a Graphical User Interface(GUI). This crop prediction can be done by Naïve Bayes Classifier, Random Forest (RF). By applying Naïve Bayes algorithm got best accurate value result with accuracy of 96.89%.

[5] K.R.Akshatha, K.S. Shreedhara have proposed a system which considers several attributes such as Depth, Texture, pH, Soil Color, Permeability, Drainage, Water holding and Erosion. The dataset contains around nine different crops. The learners used in this model were Random Forest, K-nearest neighbor and Naïve Bayes algorithm. The Naïve Bayes algorithm gave the best accurate results.

[6] Tedy Setiadi, Fiftin Noviyanto, Hendika Hardianto, Ali Tarmuji, Abdul Fadlil, Merlinda Wibowo have proposed a system which considers weather attributes such as sun exposure, rainfall, average temperature, humidity and number of days of rain. Also, considers attributes related to harvest such as productivity,yields in tons. This crop prediction can be done by Naïve Bayes and the best accurate result obtained was around 85.71% with mean absolute error of around 0.1051 .

III. PROPOSED SYSTEM

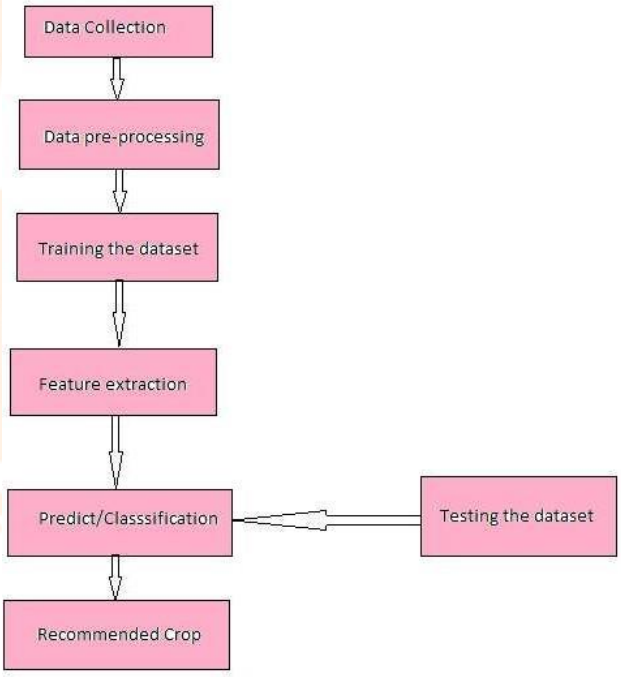

Fig. 1. Block Diagram of Overall Methodology of Proposed System

FLOW OF THE PROPOSED SYSTEM :

As shown in Fig. 1, our proposed system contains the following steps described below :

A. Data Collection :

The dataset contains soil-specific attributes we got from the Kaggle, which has information on the levels of

Nitrogen(N), Phosphorus(P), and Potassium(K) in the soil, and Temperature, Humidity, pH, and Rainfall, which directly impact the growth of crops. The data can be used to make data-driven recommendations for achieving optimal nutrient and environmental conditions to improve crop yield.

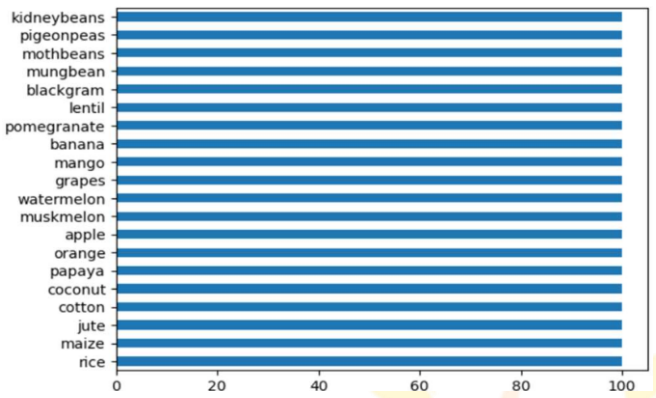

Fig. 2. Various crops present in the dataset

The dataset contains around 20 different crops such as kidneybeans, pigeonpeas, mothbeans and several other crops as shown in Fig. 2.

B. Data Pre-Processing :

Since, the real-world data contains some missing values, noises, or some unusual format that cannot be used directly in machine learning. So, it is a prerequisite for every machine-learning project to pre-process the data.

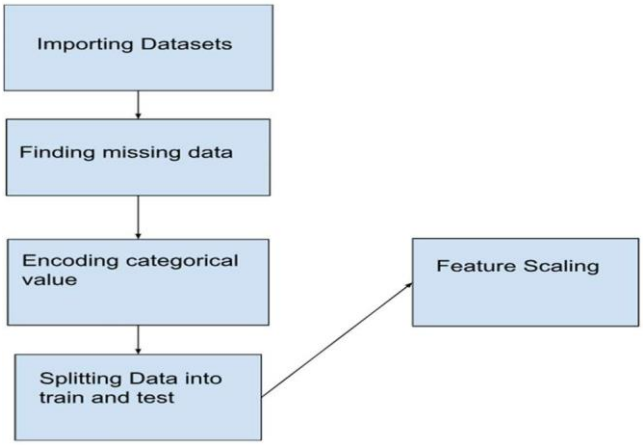

Fig. 3. Flow of Data Pre-Processing

The steps in Fig. 3. are as follows :

(i) The first step is to import the dataset
(ii) Then finding the missing values means deleting that row or calculating the mean. In larger datasets, if we delete the null value there is no problem. But in smaller datasets, it is recommended to calculate the mean
(iii) Machine learning cannot handle categorical data so categorical data has to be converted into numerical data by using Label Encoding technique
(iv) Splitting data into train and test set is a crucial step in machine learning. Suppose we are training our model by a dataset and testing with completely different datasets does not make any sense
(v) Feature scaling is a process of standardizing the variable in a specific range using StandardScaler

C. Prediction :

Prediction in machine learning involves using trained models to forecast outcomes based on input data. Algorithms learn patterns from past data to make accurate predictions on new, unseen data.

D. *Crop recommendation :*

Crop recommendation in agriculture involves using data such as soil quality, climate, and historical crop performance to suggest the most suitable crops for a specific area, maximizing yield and sustainability.

IV. IMPLEMENTATION

In this proposed system, several different Machine Learing Algorithms are used such as Logistic Regression, Decision Trees, Random Forest and XGBoost.

A. LOGISTIC REGRESSION :

Logistic Regression is a base model which is primarily used for classification problems. Logistic regression is used to predict two-class classification problems. To some extent one vs rest classifier will help to multi-class classification. In short, the logistic regression model computes a sum of the input features and calculates the logistic of the result. And the algorithm gave 96.59 % as accuracy.

The steps applied for Logistic Regression in our model are:
(i) Importing LogisticRegression library from sklearn.linear_model class.
(ii) Create classifier object
(iii) fit the object to our data

```
from sklearn.linear_model import
LogisticRegression
softReg = LogisticRegression(multi_class
= 'multinomial', solver = 'lbfgs')
softReg.fit(x_train, y_train)
```

B. DECISION TREE :

Decision Tree is a machine learning algorithm used for classification and regression problems. It is basically used for classification problems in which internal nodes represent the features. Branches represent decision rules, and each leaf represents the output. In order to build a tree it uses the CART algorithm. The algorithm starts from the root node, and its value compared with the records based on comparison follows

the branch and falls into the next node .It continues until it reaches the leaf node. And the algorithm gave 97.80% as accuracy.

The steps applied for Decision Tree in our model are:

(i) Importing DecisionTreeClassifier library from sklearn.linear_model class.
(ii) Create Decision Tree classifier object
(iii) fit the object to our data

```
from sklearn.tree import
DecisionTreeClassifier
clf1 = DecisionTreeClassifier()
```

C. RANDOM FOREST :

Random Forest is a popular ensemble learning algorithm like the decision tree algorithm it can also be used for both classification and regression problems. Rather than using one decision tree, it creates a number of decision trees and takes a prediction from each decision tree, and a prediction from a random forest is based on majority voting of prediction from each decision tree. Highest the number of trees leads to the highest accuracy and avoids overfitting. And the algorithm gave 98.86% as accuracy.

The steps applied for Random Forest in our model are:

(i) Importing RandomForestClassifier library from sklearn.ensemble class.
(ii) Create Random Forest classifier object
(iii) fit the object to our data

```
from sklearn.ensemble import
RandomForestClassifier
classifier= RandomForestClassifier()
classifier.fit(x_train, y_train)
```

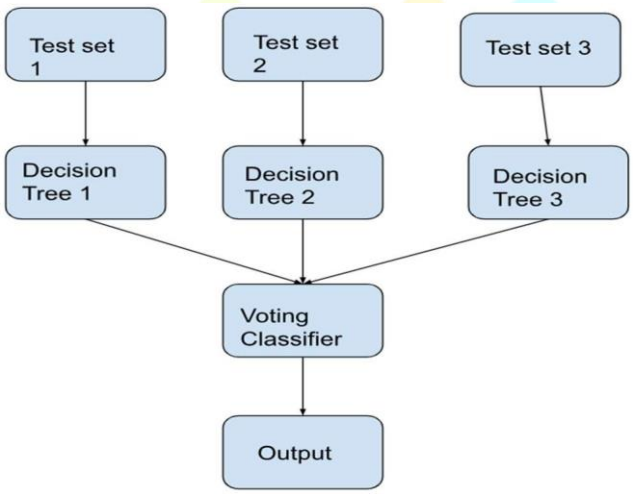

Fig. 4. The dataset is divided into three subsets and is given as input for different Decision Trees. The output is selected from the voting classifier.

D. XGBOOST :

XGBoost (eXtreme Gradient Boosting) is a improved rendition of gradient boosting algorithm .Using advance regularization and second order gradients makes it different from the gradient boosting algorithm. It runs on windows, linux and macOs. It also supports various programming languages. There is a chance of overfitting in XGBoost algorithm. eta parameter in the XGBoost may reduce the chances of overfitting. And the algorithm gave 96.36% as accuracy.

The steps applied for Decision Tree in our model are:

(i) Importing XGBClassifier library from xgboost class.
(ii) Create XGBoost classifier object
(iii) fit the object to our data

```
from xgboost import XGBClassifier
model= XGBClassifier ()
model.fit(x_train , y_train)
```

V. EXPERIMENTAL RESULT

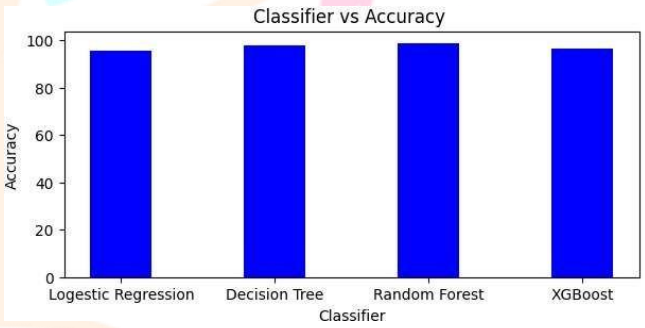

Fig. 5. Accuracy Comparision

Fig. 5. depicts that the Random Forest algorithm gave the highest accuracy among all algorithms. It compares accuracy of different algorithms.

Classifier	Accuracy
Logistic Regression	96.59%
Decision Tree	97.80%
Random Forest	98.86%
XGBoost	96.36%

Table. 1. Classifier and its Accuracy result. It depicts the accuracy to its corresponding classifier. The Random Forest classifier gave 96.86% as accuracy which is the highest accuracy among all classifiers.

A. GRAPHICAL USER INTERFACE

We built a website where users can give inputs related to various environmental factors and soil characteristics to predict crop suitable for higher yeild and profit.

Fig. 6. User interface to take inputs from farmers or users. It shows the website where the user can input the recorded values of different components and predict the suitable crop to grow.

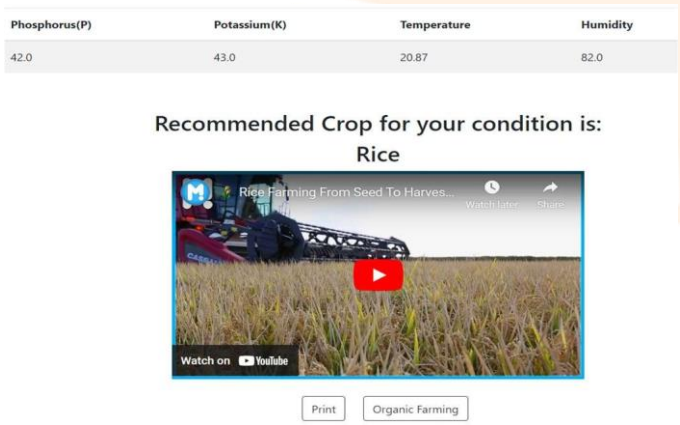

Fig. 7. Crop recommended for given input values. It shows the recommended crop for the user and a video related to crop cultivation for more information about the crop.

VI. CONCLUSION

The recommendation of crops yield based on the machine learning algorithm have shown that higher crop yield can be achieved. . From the obtained results we can conclude that from Random Forest achieved the highest accuracy of about 99%. We recommend to use Random Forest since overfitting will not happen easily. And it gives accurate results.

The system will help farmers in making an informed decision about which crop to grow for better yield and profit considering parameters such as Nitrogen(N), Phosphorous(P), Potassium(K) and environmental factors such as Rainfall, Temperature and humidity.

VII. FUTURE WORK

The work can be developed further to add following functionalities :

(i) To give Fertilizer recommendation for the crop predicted.
(ii) To build mobile application can be built for easy access to farmers from any location.
(iii) By building a model to classify between unhealthy and healthy crop.
(iv) To give organic farming techniques to get high yields for predicted crop.

VII. REFERENCES

[1] Dhruvi Gosai, Chintal Raval, Rikin Nayak, Hardik Jayswal, Axat Patel. “Crop Recommendation System using Machine Learning”.
[2] Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H, W.J.M. Nanayakkara, Dimantha M.A.C, Pabasara M.G.P. “Crop Recommendation System”.
[3] Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, Avinash Devare. “Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique”.
[4] Anguraj.K , Thiyaneswaran.B , Megashree.G , Preetha Shri.J.G , Navya.S , Jayanthi. J. “Crop Recommendation On Analyzing Soil Using Machine Learning Technique”.
[5] K.R.Akshatha, K.S. Shreedhara. “Implementation of Machine Learning Algorithms for Crop Recommendation using Precision Agriculture”.
[6] Tedy Setiadi, Fiftin Noviyanto, Hendika Hardianto, Ali Tarmuji, Abdul Fadlil, Merlinda Wibowo. “Implementation Of Naïve Bayes Method In Food Crops Planting Recommendation”.
[7] Soumya Sri Attaluri , Nowshath K Batcha , Raheem Mafas. “Crop Plantation Recommendation using Feature Extraction and Machine Learning Techniques”.
[8] SML Venkata Narasimhamurthy, AVS Pavan Kumar. “Rice Crop Yield Forecasting Using Random Forest Algorithm”.
[9] D.Anantha Reddy, Bhagyashri Dadore, Aarti Watekar. “Crop Recommendation System to Maximize Crop Yield in Ramtek region using Machine Learning”, 2019.
[10] Deepti Dighe, Harshada Joshi, Aishwarya Katkar, Sneha Patil, Prof. Shrikant Kokate. “Survey of Crop Recommendation Systems”, NOV 2018.
[11] Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. IEEE, 2020.
[12] Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta. "Crop yield prediction using machine learning techniques". IEEE, 2019.
[13] Gandge, Yogesh. "A study on various data mining techniques for crop yield prediction". IEEE, 2017.
[14] Suresh, A., P. Ganesh Kumar, and M. Ramalatha. "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN". IEEE, 2018.
[15] Kamatchi, S. Bangaru, and R. Parvathi. "Improvement of Crop Production Using Recommender System by Weather Forecasts". 2019
[16] Jejurkar Siddhi, S., S. Bhosale Meghna, and D. N. Wavhal. "Crop Predication and Diseases Detection Using Machine Learning" , 2021.

[17] Ghadge, Rushika, Juilee Kulkarni, Pooja More, Sachee Nene, and R. L. Priya. "Prediction of crop yield using machine learning." , 2018.
[18] Kulkarni, Nidhi H., G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery. "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique." IEEE, 2018.
[19] Pudumalar, S., E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha. "Crop recommendation system for precision agriculture." IEEE, 2017.
[20] Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal. "Crop yield prediction using machine learning algorithms." IEEE, 2019.