**Crop Prediction Using Machine Learning**

Karan Kumar Chourasiya, Jharkhand Rai university, Ranchi

Manjeet Singh, Jharkhand Rai university, Ranchi

Dr. Ashish Kumar Sinha, Associate Professor, Jharkhand Rai University, Ranchi

Kumar Amrendra, Assistant Professor, Jharkhand Rai University, Ranchi

Anuradha Sharma, Assistant Professor, Jharkhand Rai University, Ranchi

## Abstract

Crop prediction is a web application that use machine learning algorithm that aims farmers and agriculture stakeholders in making informed decisions by accurately forecasting crop yields and identifying suitable crops choices. The dataset will work to give appropriate farmers his choice Crop recommendation.

Crop prediction employs advanced data analytics techniques and machine learning algorithms to analyses their land soil and can help them make informed decisions about crop selection, planting and harvesting. The factors which we have studied include climate conditions, soil quality, precipitation levels, temperature patterns and crop-specific parameters. By combining these datasets, we can see accurate predictions regarding crop performance, growth stages.

We also highlight the benefits of the crop prediction like firstly; it empowers farmers to maximize their crops production while minimizing input costs, resulting in improved profitability. Secondly, we notice that the methods in the proposed system include increasing the yield of crops, selecting efficient parameters, making smarter decisions. Thirdly, we highlight the risk of crop failure and financial losses by providing early warnings for potential threats like weather condition and soil quality.

***Keyword: -*** prediction, recommendation, PH level, testing, Machine Learning.

## 

## Introduction

Machine learning is a valuable decision-making tool for predicting agricultural yields and deciding the type of crops to sow and things to do during the crop growing season. In order to aid crop prediction studies, several machine learning methods have been used.

Machine learning techniques are utilized in various sectors, from evaluating customer behavior in supermarkets to predicting customer phone usage. For some years, agriculture has been using machine learning techniques. Crop prediction is one of agriculture's complex challenges, and several models have been developed and proven so far. Because crop production is affected by many factors such as atmospheric conditions, type of fertilizer, soil, and seed, this challenge necessitates using several datasets. This implies that predicting agricultural productivity is not a simple process; rather, it entails a series of complicated procedures. Crop yield prediction methods can now reasonably approximate the actual yield, although more excellent yield prediction performance is still desired.The project aims to compare various supervised learning algorithms like Naive Bayes, Decision Tree, and Random Forest and SVM on the dataset containing 22 varieties of crops. For the Decision Tree and Random Forest Classifier, the model’s performance is calculated under two criterions- Entropy and Gini Index. The results reveal that the suggested machine learning technique's effectiveness is compared to the best accuracy with precision, recall, and F1 Score. Given the significance of crop prediction, numerous suggestions have been proposed in the past with the goal of improving crop prediction accuracy. After the analysis, Random forest gave the best prediction result of the three algorithms. An application for farmers can be created that will aid in the reduction of many problems in the agriculture sector In this application, farmers perform single/multiple testing by providing input such as crop name, season, and location. As soon as one provides the input, the user can choose a method and mine the outputs. The outputs will show you the crop's yield rate. The findings of the previous year's data are included in the datasets and transformed into a supported format. The machine learning models used are Naïve Bayes and SVM, Random Forest and Decision tree .A comparative study of various machine learning can be applied on a dataset with a view to determine the best performing methodology. The prediction is found by applying the Regression Based Techniques such as Linear, Random Forest, and Decision Tree

## Literature review

Ridge on the dataset containing details about the types of crops, different states, and climatic conditions under different seasons. The parameters used to estimate the efficiency of these techniques were mean absolute error, root mean square error, mean squared error, R-square, and cross validation. Gradient Boosting gave the best accuracy- 87.9% for the target variable ‘Yield’ and Random Forest- 98.9% gave the best accuracy for the target value ‘Production’.

The DHT22 sensor is recommended for monitoring live temperature and humidity [12]. The surrounding air is measured with a thermostat and a capacitive humidity sensor and outputs a digital signal on the data pin to the Arduino Uno port pin. The humidity value ranges from 0-100% RH and - 40 to 80 degrees Celsius to read the temperature. The above two parameters and soil characteristics are considered as input to three different machine learning models: Support Vector Machine, Decision Tree, and KNN. The Decision Tree gave better accuracy results.

## Approach to Crop Prediction

**Author Proposed Model Accuracy**

**M.Kalimuthu et.al(2020)[5]** Naïve Bayes 97%

**V. Geetha et.al(2020)[6]** Naïve Bayes and Random Forest

Classifier

95%

Random Forest Classifier

## Shilpa Mangesh P et.al(2021)[7]

Support Vector Machine, K- Nearest Neighbor, Multivariate Linear Regression, Artificial Neural Network, Random Forest

95%

Random Forest Classifier

**S Bharath et.al(2020)[9]** Support Vector

Machine,

Decision Tree, K- Nearest Neighbor, Random Forest

99.87%

Decision Tree Classifier

## Payal Gulati and Suman Kumar(2020)[11]

Linear Regression, Random Forest, Decision Tree, Gradient Boosting Regression, Ridge Regression, Polynomial Regression

98.9%

Random Forest Classifier

**Archana Gupta et.al(2020)[12]** K-Nearest Neighbor, Support 91.03%

Vector Machine, Decision Tree Decision tree Classifier

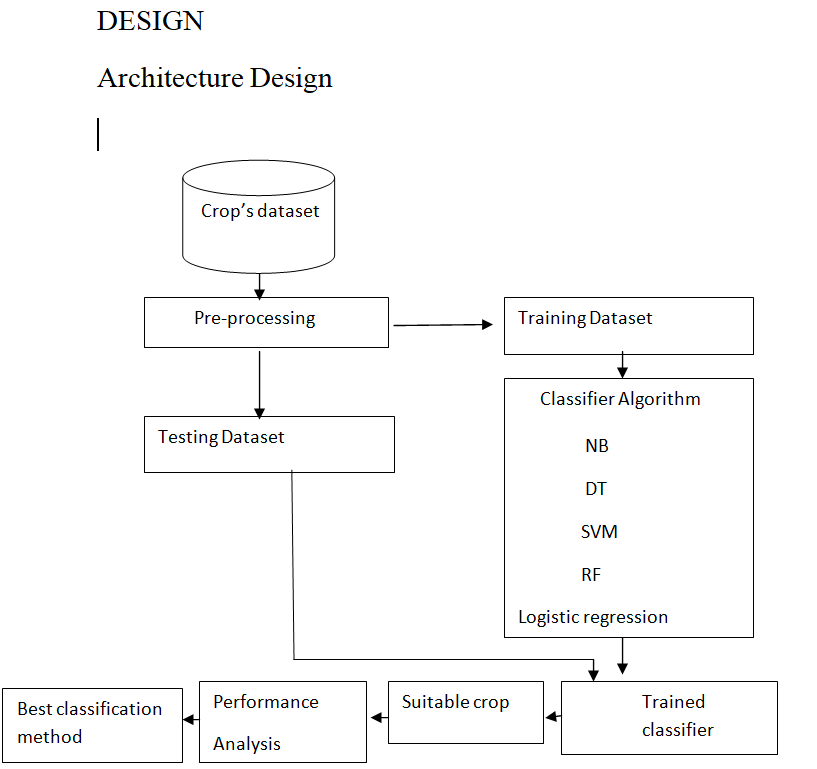
**Process of methodology**

# **Problem Statement**

Crop prediction is a problem of designing an efficient crop prediction system to enhance agriculture productivity and optimize resource allocation. The aim of crop prediction is to provide a way where farmers and agriculture stakeholders in making informed decisions by accurately forecasting crop yields and identifying suitable crops choices. The development of crop prediction should concentrate on many factors like weather patterns, soil quality, crop performance, and market trends to provide accurate predictions of crops to make profitability. By implementing crop prediction system, it would be easy to develop a reliable and accurate system that can forecast the most suitable crops to be cultivated in a given period where the factors like climatic condition, soil quality and crop specific parameters perform a important role. The system will empower farmers to make informed decisions, improve productivity, and reduce the risk associated with improved prediction for their crops which make them leading to sustainable agriculture practices.

# **Proposed System**

In crop prediction we collect relevant data such as climate condition, soil quality, precipitation level and other relevant factors. Crop prediction is a efficient system where it will predict any crops which will be beneficial for farmers. Where user can create a profile on the app browses the available crops, reduce the risk of crop failure, maximize the crops yields which has predicted by crop prediction application. Overall, the system empowers farmers with data-driven decision-making tools, enabling them to make informed choices and achieve greater success in agriculture effort. For Implementation we are using various machine learning algorithms to make a effective crop prediction system. Crop prediction system aims to revolutionize traditional farming practices and make a new era of progressive and effective agriculture.



# Figure: Block Diagram

**In the above architecture**

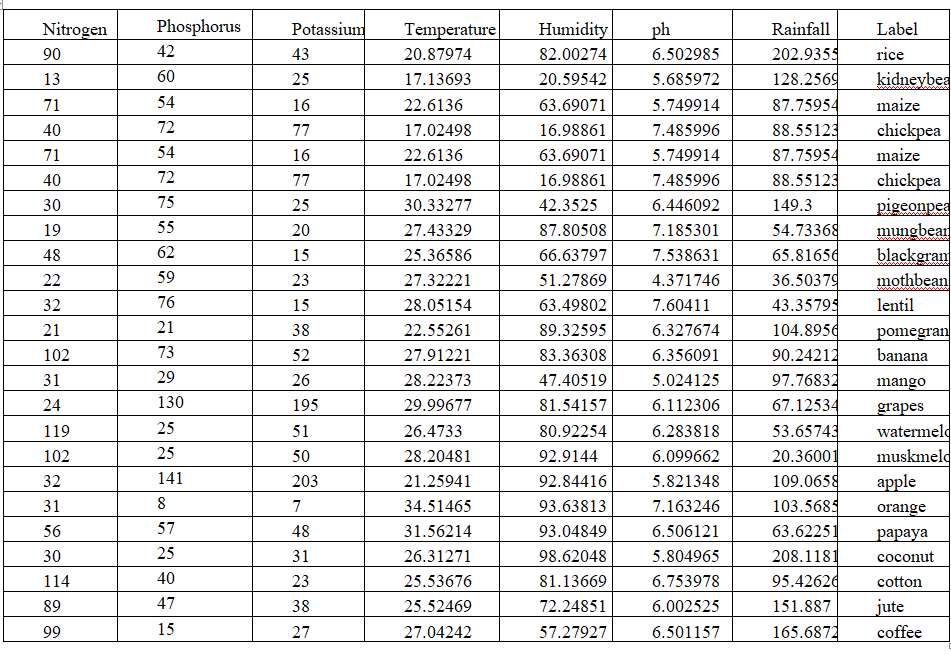
The first step is to collect the dataset for crop prediction then we have import the dataset then we check the unwanted data in the dataset then pre-processing of data takes where we check the outliers and missing value and scale the dataset then we get update the data set and split the data in to Training Dataset, Testing Dataset and apply the classifier algorithm like Random forest, SVM, Logistic Regression, Decision tree , Naïve bayes algorithm for train the model and applied the classified algorithm using test dataset then after we will analysis the performance and choose the algorithm with the best accuracy for predict suitable crop

**Dataset Collection and processing:**

Crop prediction the dataset collected form that is available on the kaggle website (crop recommendation dataset|kaggle) in this dataset have In the crop prediction dataset, we found 2202 row and 8 Column

.Each attributes different parameters and has columns and specific row. In which column A represents Nitrogen Column B represent the

Phosphorus and. Column C represents the Potassium and column D represent Temperature of area next column E represents Humidity of the area and the after that other column PH level, Rainfall and the last column is for the crop



**Dataset Collection and processing:**

Crop prediction the dataset collected form that is available on the kaggle website (crop recommendation dataset|kaggle) in this dataset have In the crop prediction dataset, we found 2202 row and 8 Column

.Each attributes different parameters and has columns and specific row. In which column A represents Nitrogen Column B represent the

Phosphorus and. Column C represents the Potassium and column D represent Temperature of area next column E represents Humidity of the area and the after that other column PH level, Rainfall and the last column is for the crop

**Methodology:**

**Logistic regression:**

p(y=1|x) = 1 / (1 + exp(-z))

where p(y=1|x) is the probability of the dependent variable (y) being equal to 1 given the independent variables (x), exp is the exponential function, and z is the linear combination of the independent variables and their coefficients:

z = b0 + b1x1 + b2x2 + ... + bn\*xn

Where b0 is the intercept and b1...bn are the coefficients for each independent variable (x1...xn). The coefficients are estimated during the model fitting process, where the goal is to find the values that minimize the error between the predicted probabilities and the actual values.

The logistic regression model uses the sigmoid function, which takes the linear combination z as input and maps it to a value between 0 and 1. This allows the model to output a probability that the dependent variable is equal to 1, given the values of the independent variables. If the probability is greater than a chosen threshold, the model predicts that the dependent variable is equal to 1, otherwise it predicts that it is equal to 0

**Random Forest**

Random forest is an ensemble learning algorithm that combines multiple decision trees to make a final prediction. Each decision tree in the random forest is constructed using a random subset of the training data and a random subset of the features.

Here is the formula for random forest:

Prediction = mode(y1, y2, ..., ym)

Where yi is the predicted class for the ith decision tree, and m is the total number of trees in the random forest.

**Decision Tree:** a decision tree can be represented by a set of nested if-else statements that split the data into smaller subsets based on the values of the input features.

There are following step used in Decision Tree algorithm

1. Start with the entire dataset.
2. For each feature in the dataset:
3. Calculate the information gain by splitting the data on that feature.

IG(S, A)= Entropy(S)-∑v€A|SV|/ S. Entropy(S)

1. Choose the feature with the highest information gain as the node to split on.
2. Create a branch for each possible value of the chosen feature and recursively repeat steps 1-3 for each branch, using only the subset of the data that corresponds to that branch.
3. Continue splitting the data until a stopping criterion is met, such as reaching a maximum depth or minimum number of samples per leaf node.
4. Assign a class label to each leaf node based on the majority class of the samples that reach that node.

**Support Vector Machine:**

Support Vector Machine (SVM) algorithm used for classification, regression, and outlier detection.

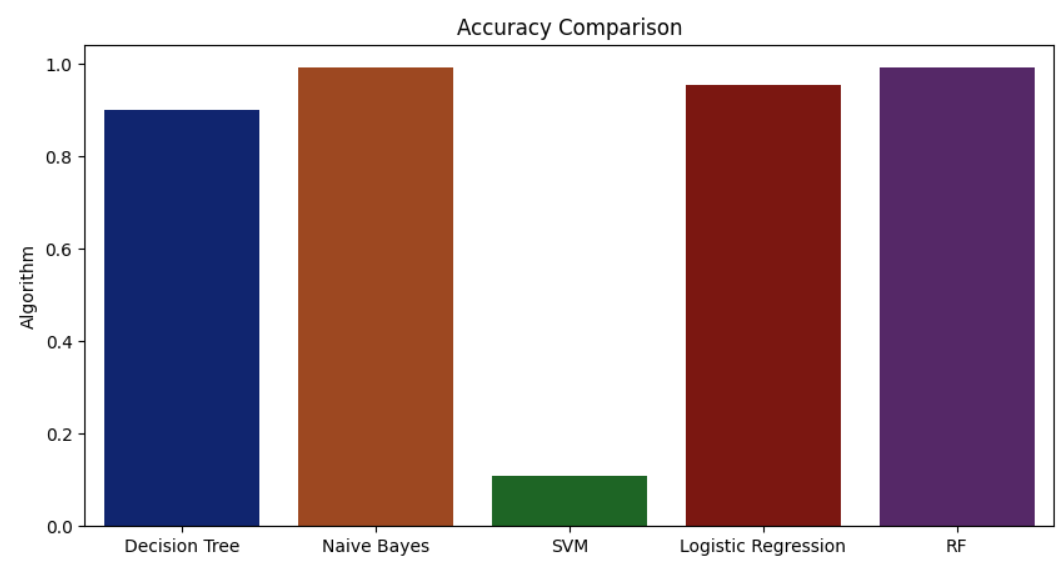
The standard formula of the SVM for binary classification is:

y(x) = w^T x + b

where:

* x is the input vector.
* w is the weight vector.
* b is the bias term.
* y(x) is the predicted output of the SVM for the input vector x.

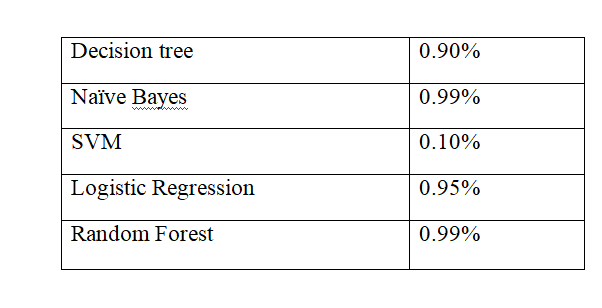
**Accuracy Comparison of all models**

****

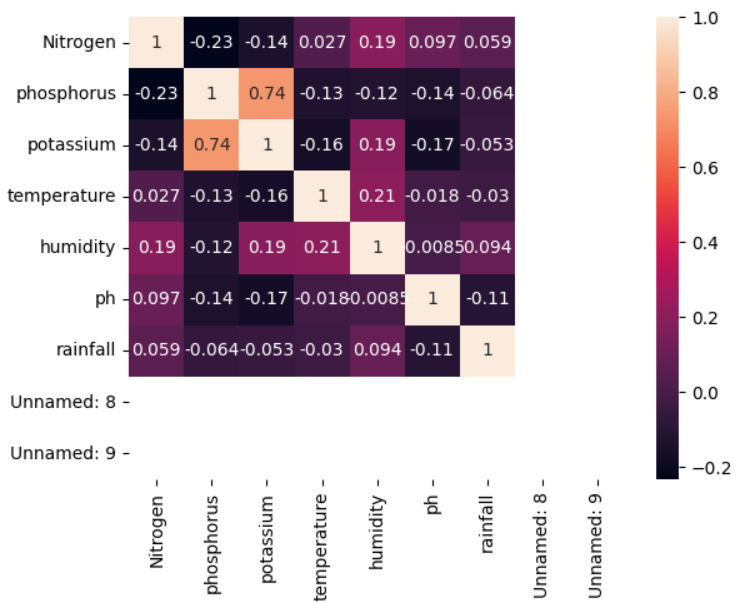
For crop prediction, here we applied following algorithm to compare the accuracy score. After comparison, we get Decision tree 0.90%, Naive Bayes 0.99%, SVM 0.10%, Logistic Regression 0.95%, Random Forest 0.99%

**Result**

The comparative study of three different supervised machine learning models (SVM, Decision Tree, and Random Forest, Logistic Regression) is done to predict the best-suited crop for the particular land that can help farmers to grow crops more efficiently. In completion, we concluded that the crop prediction dataset showed the best accuracy with Random Forest Classifier both in Entropy and Gini Criterion with 0.99%. In contrast, SVM has the lowest accuracy among the three with 0.10%, and the accuracy of Decision Tree Classifier is 0.90% and Random Forest Classifier and Naive Bayes 0.99%. When comparing the accuracy value, Random forest criterion gave a better accuracy of 0.99% compared to Decision Tree Entropy Criterion. In the future, new data from the fields can be collected to get a clear image of the soil and incorporate other machine learning algorithms In the Crop prediction system, used SVM, Decision Tree, random forest, Naive Bayes, Logistic Regression algorithm for as these gave the best accuracy accordingly.



Matrix correlation: Matrix correlation metric used in machine learning is the Pearson correlation coefficient, which measures the strength of the linear relationship between two continuous variables. It ranges from -1 to 1, where -1 indicates a perfect negative correlation.



## 

## Conclusion

In crop prediction system is a valuable tool that can help farmer take necessary steps to prevent or making smarter decisions. By analyzing soil, crop history, and other relevant data, these systems can provide predictions conditions that a person may be at risk of developing. Overall, the use of crop prediction systems can lead to improved and increasing the yield of crops.

## Reference

1. Dahikar S and Rode S V 2014 Agricultural crop yield prediction using artificial neural network approach *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering* **vol 2** Issue 1 pp 683-6.
2. Suresh A, Ganesh P and Ramalatha M 2018 Prediction of major crop yields of Tamilnadu using K-means and Modified KNN *2018 3rd International Conference on Communication and Electronics Systems (ICCES)* pp 88-93 doi: 10.1109/CESYS.2018.8723956.
3. Medar R, Rajpurohit V S and Shweta S 2019 Crop yield prediction using machine learning techniques *IEEE 5th International Conference for Convergence in Technology* (I2CT) pp 1-5 doi: 10.1109/I2CT45611.2019.9033611.
4. Nishant P S, Venkat P S, Avinash B L and Jabber B 2020 Crop yield prediction based on Indian agriculture using machine learning *2020 International Conference for Emerging Technology* (INCET) pp 1-4 doi: 10.1109/INCET49848.2020.9154036.
5. Kalimuthu M, Vaishnavi P and Kishore M 2020 Crop prediction using machine learning *2020 Third International Conference on Smart Systems and Inventive Technology* (ICSSIT) pp 926-32 doi: 10.1109/ICSSIT48917.2020.9214190.
6. Geetha V, Punitha A, Abarna M, Akshaya M, Illakiya S and Janani A P 2020 An effective crop prediction using random forest algorithm 2020 *International Conference on System, Computation, Automation and Networking* (ICSCAN) pp 1-5 doi: 10.1109/ICSCAN49426.2020.9262311.
7. Pande S M, Ramesh P K, Anmol A, Aishwaraya B R, Rohilla K and Shaurya K 2021 Crop recommender system using machine learning approach 2021 *5th International Conference on Computing Methodologies and Communication* (ICCMC) pp 1066-71 doi: 10.1109/ICCMC51019.2021.9418351.
8. Sellam V, and Poovammal E 2016 Prediction of crop yield using regression analysis *I*
9. *Journal of Science and Technology* **vol 9**(38) pp 1-5.
10. Bharath S, Yeshwanth S, Yashas B L and Vidyaranya R Javalagi 2020 Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction *International*