CSE-458: Machine Learning Laboratory

Project Proposal Phase: **02**

Project Name: Optimizing Crop Recommendations through Machine Learning.

Date of Submission: 11.02.2024

Supervised By

Sarnali Basak

Associate Professor

Department of Computer Science and Engineering

Jahangirnagar University

&

Masum Bhuiyan

Lecturer

Department of Computer Science and Engineering

Jahangirnagar University

Submitted By

Group No- 07

Neural Nexas

Class Roll	Exam Roll	Name
352	191334	Sayma Akter Rupa
353	191335	Khandoker Nosiba arifin
356	191337	Mangsura Kabir Oni
371	191350	Shifati Rabbi
377	191356	Md. Ashraful Morsalin

Abstract- The agricultural sector plays an essential role in global food production, and the efficiency of crop cultivation heavily depends on a number of variables such as soil type, climate conditions, and crop-specific requirements. The growing global demand for food production requires the optimization of farming methods. This paper presents a Crop Recommendation System that uses Machine Learning techniques to optimize crop selection for farmers. The system combines a wide range of data sources that helps the farmers to increase the crop yield by recommending a suitable crop for their land with the help of geographic and climatic parameters.

Index Terms- Crop Recommendation, Machine Learning, Soil, Random Forest etc.

I. Introduction

Bangladesh's economy is deeply rooted in agriculture, making it heavily reliant on farming as the cornerstone of its income generation. With a dense population and variable climatic conditions, ensuring food security remains a paramount concern. Traditional agricultural practices often face challenges related to efficiency, cost-effectiveness, and resource utilization. To improve crop yield, factors such as soil type, fertility, and rainfall play crucial roles. Precision agriculture, with its focus on enhancing crop quality through information technology methods, has gained traction in Bangladesh. The country has witnessed significant advancements in agriculture, with precision farming emerging as a key strategy for area-specific cultivation. This study aims to recommend suitable crops based on input parameters like soil pH, nutrients, rainfall, humidity, and temperature, tailored to Bangladesh's agricultural landscape. By employing various algorithms and techniques, the research seeks to build efficient models for predicting crop output accurately, thereby addressing the challenge of increasing crop productivity in Bangladesh's agricultural sector.

II. Literature Review

Bangladesh is largely dependent on agriculture from its economic point of view. The farmers in our country face issues in choosing the right crop for production. As a result, they face considerable reduction in crop production. Precision agriculture has been used to help farmers overcome their problems. Precision agriculture (Crop Recommendation System) is a contemporary agricultural approach which researches data on soil characteristics, the amount of different elements like (Nitrogen, Phosphorus, Potassium) in soil, rainfall, temperature, humidity in that particular area to advise farmers on the optimum crop for their specific region. Our system can also inform farmers about the best practices for growing specific crops in the particular location.

In the previous years, a slight increase has been seen in research in the field of crop recommendation. For example, Alok Kumar Jagadev A et al. (2021) present "Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods. Linear Regression Prediction, Polynomial Regression, Random Forest Regression, Support Vector Regression had been used in their paper with the accuracy of 94.78%. Shilpa Mangesh Pande; Prem Kumar Ramesh; Anmol Anmol; B. R Aishwarya et al. (2021) in their paper presented “Crop Recommender System Using Machine Learning”. In this paper, Support Vector Machine, Artificial Neural Network, Random Forest, Multivariate Linear Regression had been utilized. In their research, Random forest gave the best results with the accuracy of 95%. [1] Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate et al. (2018) "Survey of Crop Recommendation Systems”. This proposed system developed a crop recommendation system for smart farming. This research paper reviewed various machine learning algorithms like CHAID, KNN, K-means, Decision Tree, Neural Network, Naïve Bayes, C4.5, LAD, IBK and SVM algorithms. For this research used the Hadoop framework for the intensive calculations and also helped to get better accuracy for the system.” Rohit Kumar Rajak, Ankit Pawar, MitaleePendke et al. (2017) in their research paper showcase the work “Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique”. They used Support Vector Machine, Naive Bayes, Multi-layer Perceptron , Random Forest in their research. Pudumalar et al., in their paper (most cited on IEEE Xplore) present a similar approach using machine learning on data collected from a district in Tamil Nadu, India; however, the paper does not talk about models’ accuracy or have not described data used. Another famous review on "Machine Learning in Agriculture" by Konstantinos G. Liakos et al. covers most of the applications of machine learning in agriculture, however, it does not mention anything on crop recommendation. There is some other agricultural-related literature that is indirectly related to crop recommendation, for example; Ayaz Muhammad, et al. in their work mainly talks about the Internet of Things and sensors for collecting agricultural data.

The agricultural yield recommendation is determined in this study by comparing multiple machine learning ML regressions and calculating the total percentage improvement over numerous years. Many machine learning techniques, such as prediction, classification, regression, and clustering, are used to anticipate agricultural output. This research examines the requirements and planning required to construct a precision farming software model. We are using machine learning models which include Naïve Bayes, Decision Trees and KNN. These models give us the best results having accuracy more than previous experiences.

III. Methodology

In this study, we used three different machine learning techniques to build models using agricultural data, which included 2201 records with 22 different crop labels. Farmers can receive recommendations from these models regarding the best crops to plant. Our methodology for crop data analysis in **Figure 1** adheres to the standard stages of data analysis. A significant improvement is the inclusion of multiple classifiers, which are fine-tuned and evaluated to identify the most suitable ones for the input data.

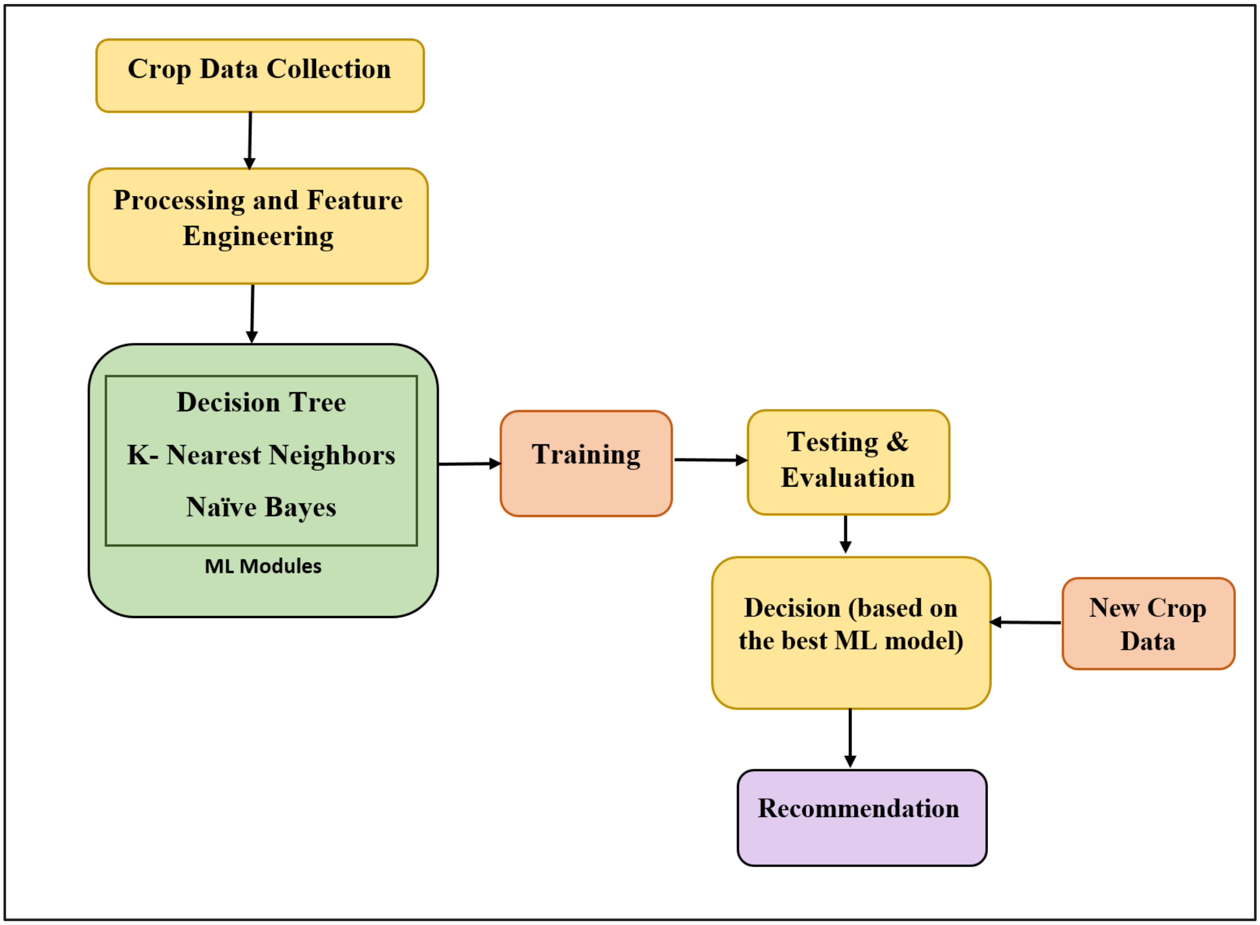


Figure 1: Crop Data Analysis and Recommendation

A. Machine learning algorithm used

Although many machine learning algorithms are commonly used, we are only highlighting the following algorithms in this survey because they are the ones that we used in our study.

(1) Decision Tree: A decision tree is a supervised learning algorithm that uses a tree-like structure to represent the relationship between the input and output data. A decision tree is made up of nodes and branches. The nodes represent decisions, and the branches represent the possible outcomes of those decisions.

(2) KNN: K-Nearest Neighbor (KNN) algorithm is one of the most popular machine learning algorithms which is used for supervised learning techniques such as classification or regression problems. It tries to find out the similarity between the new input data and available training dataset and put the new data into the category that is most similar to the available categories. KNN normally follows some steps to predict the correct category of the input data:

- First of all, select the number of K of the neighbor.
- Then calculate the Euclidean distance of K number of neighbors. $\text{distance}(x, X_i) = \sqrt{\sum_{j=1}^{d}(x_j - X_{i_j})^2}]$
- After that, take the K nearest neighbors as per the calculated Euclidean distance.
- Then count the number of the data points in each category among these k neighbors.
- After counting the number, choose the maximum number of count value
- Then assign the new input data as the maximum valued neighbor's category

There is no exact way to determine the best value for "K", for this we need to try some values to find the best out of them. The most preferred value for K is odd values to avoid ties in classification. If we select a very low value for K such as K=1 or K=2, it can be noisy and lead to the effects of outliers in the model, a higher value of k would be better. It is preferable to select k's value not too low or not too high, considering the modest one.

(3) Naive Bayes (NB) Classifier: The naive Bayes algorithm is a supervised learning algorithm that uses Bayes' theorem. It is a simple and versatile algorithm that can be used for various tasks, such as spam filtering, text classification, medical diagnosis and crop recommendation. The naive Bayes algorithm works by assuming that the presence of a particular feature in a class is unrelated to the presence of any other feature. This assumption is not always true. Therefore, it is called "naive". However, it is a good approximation in many cases, making the algorithm very simple to train and interpret. To classify an object, the Naive Bayes algorithm first calculates the probability of each class. It then calculates the probability of each feature given to each class. The class with the highest probability is the class the object is assigned to.

B. Extraction of dataset

In our project, the dataset plays a crucial role in training and evaluating our model for Crop Recommendation. The dataset comprises information on 2201 instances from kaggle which would allow it to build a predictive model to recommend the most suitable crops to grow in a particular farm based on various parameters. Each instance has seven input feature variables and one output target variable.

We have collected dataset for our project from the following link:

Crop Recommendation (kaggle.com)

There are no missing values and outliers in our dataset .The dataset is already cleaned. So, we don't need to process our dataset again.

The main attributes of our dataset are given below-

Index	Attribute	Attribute Description
1	Nitrogen(N)	It is a nutrient that helps for growth of crop leaves.
2	Phosphorus(P)	It is a nutrient that helps for growth of root and fruits and flower development.
3	Potassium(K)	That nutrient helps perform the overall function of crops.
4	Temperature	Temperature in Degree Celsius.
5	Humidity	Relative humidity in percentage.
6	pH	The pH value of the soil.
7	Rainfall	Rainfall in millimeters.

And our target variable is "label".

The first 5 rows of our dataset are shown here-

index	Nitrogen_Soil	Phosphorus_Soil	potassium_Soil	temperature	humidity	ph	rainfall	label
0	90	42	43	20.87974371	82.00274423	6.502985292	202.9355362	rice
1	85	58	41	21.77046169	80.31964408	7.038096361	226.6555374	rice
2	60	55	44	23.00445915	82.3207629	7.840207144	263.9642476	rice
3	74	35	40	26.49109635	80.15836264	6.980400905	242.8640342	rice
4	78	42	42	20.13017482	81.60487287	7.628472891	262.7173405	rice

Visualizing of all attributes of dataset:

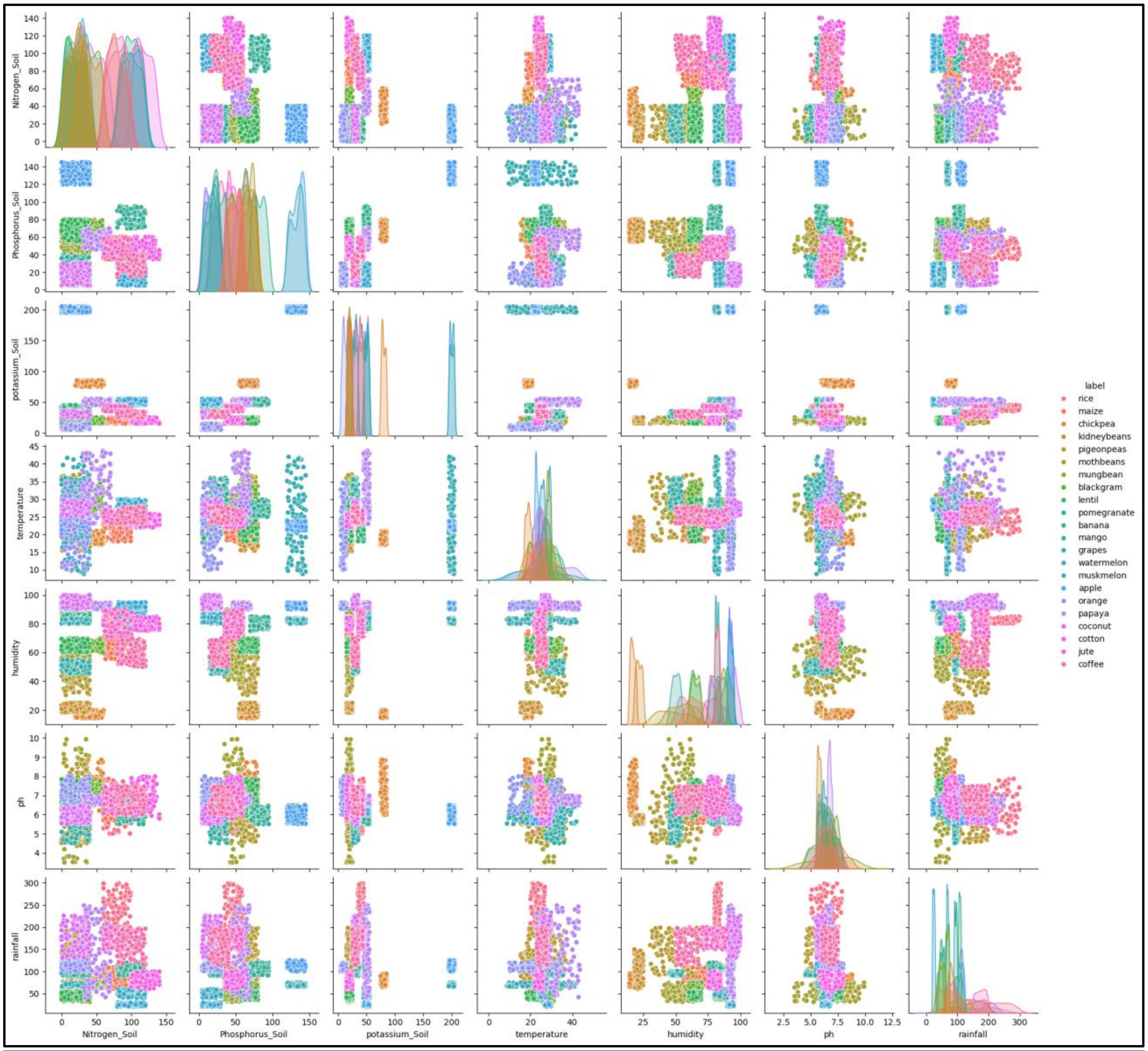


Fig: Pair Plotting of all data

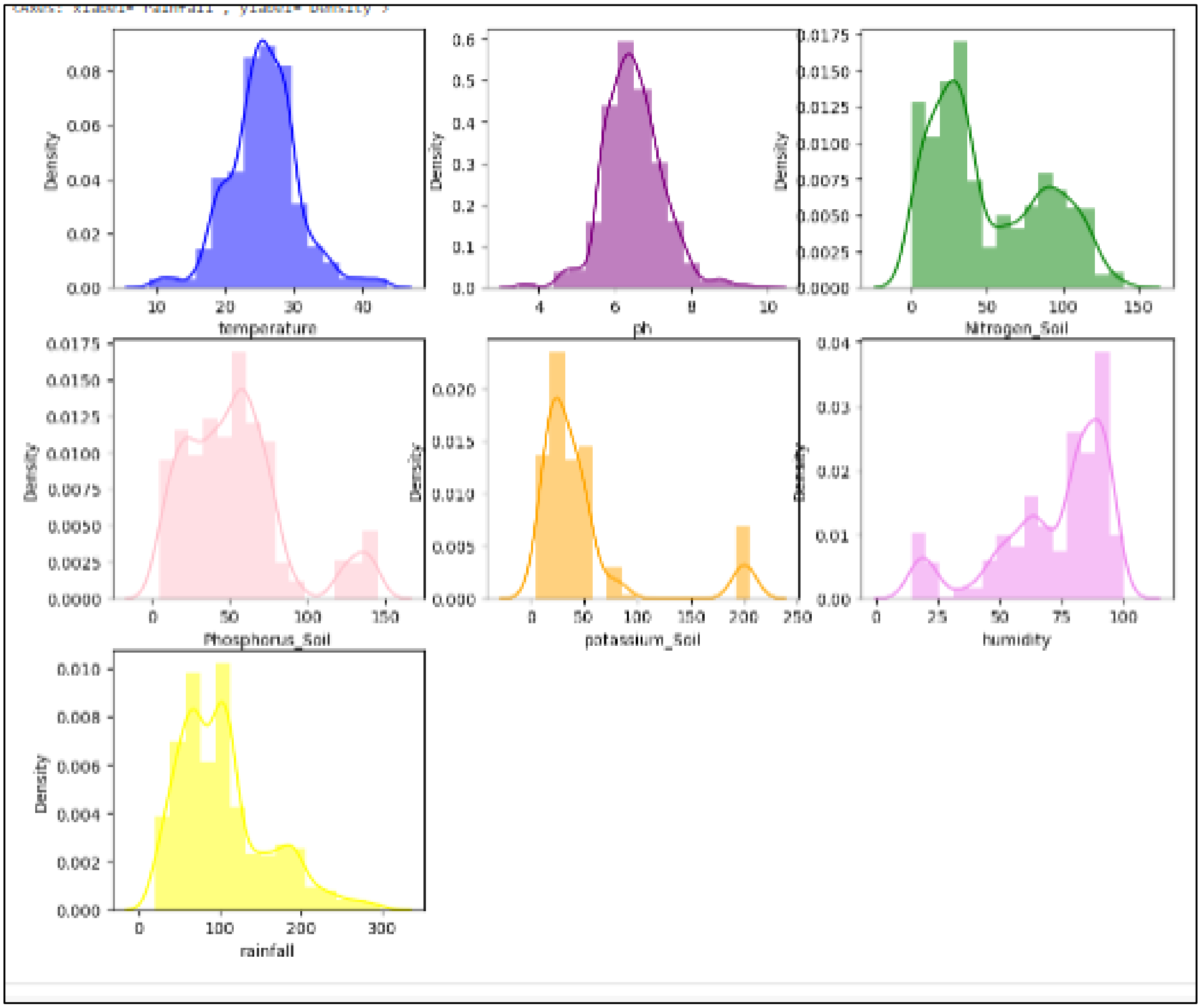


Fig: Visualization of attributes individually

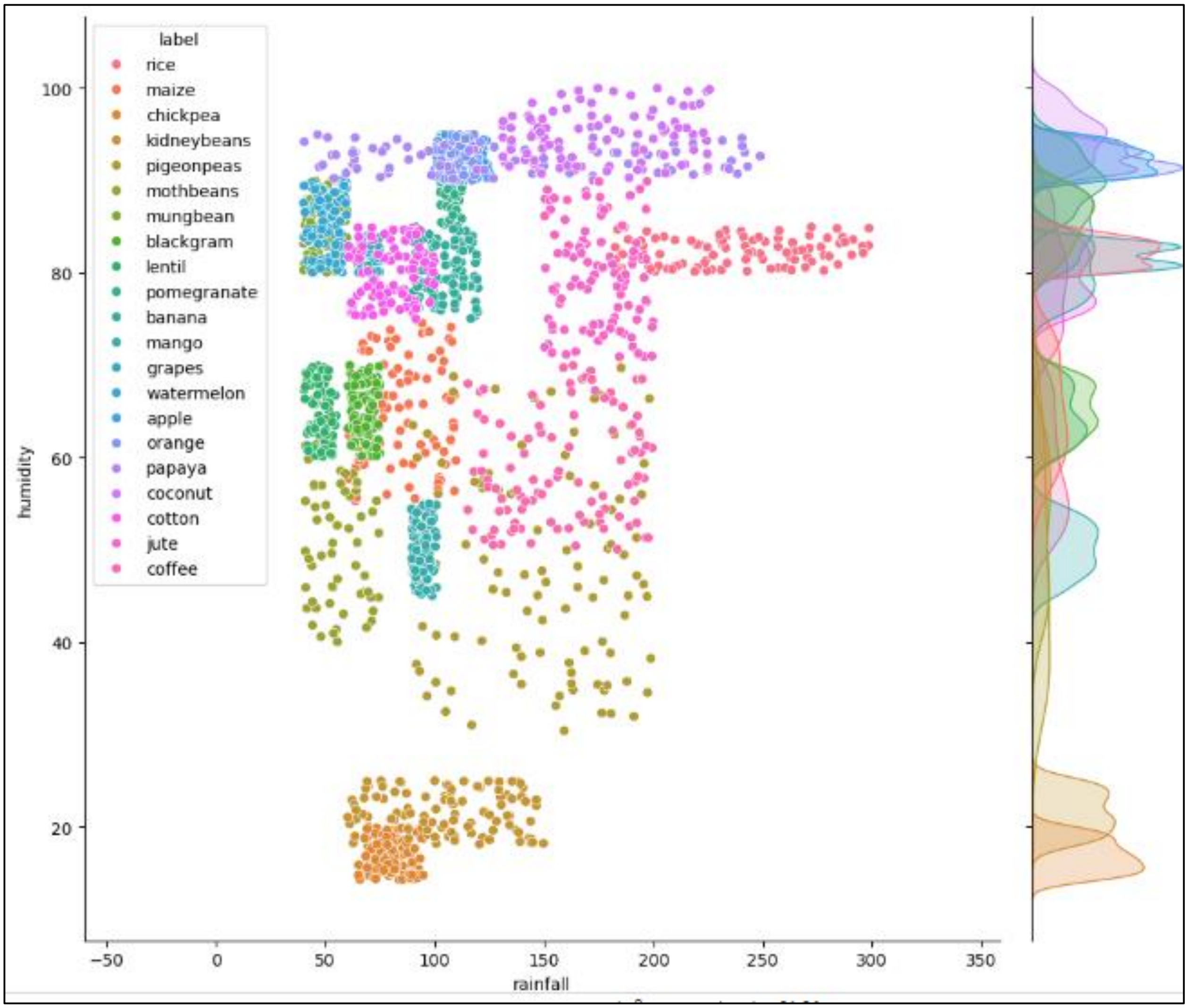


Fig: Join plot of humidity, rainfall

Visualizing all attributes of dataset using weka:

These are the specific colors for a specific crop used in weka. It indicates the color codes of individual crops.

Class colour

rice maize chickpea kidneybeans pigeonpeas mothbeans mungbean blackgram lentil pomegranate banana mango grapes watermelon muskmelon apple orange papaya coconut cotton jute coffee

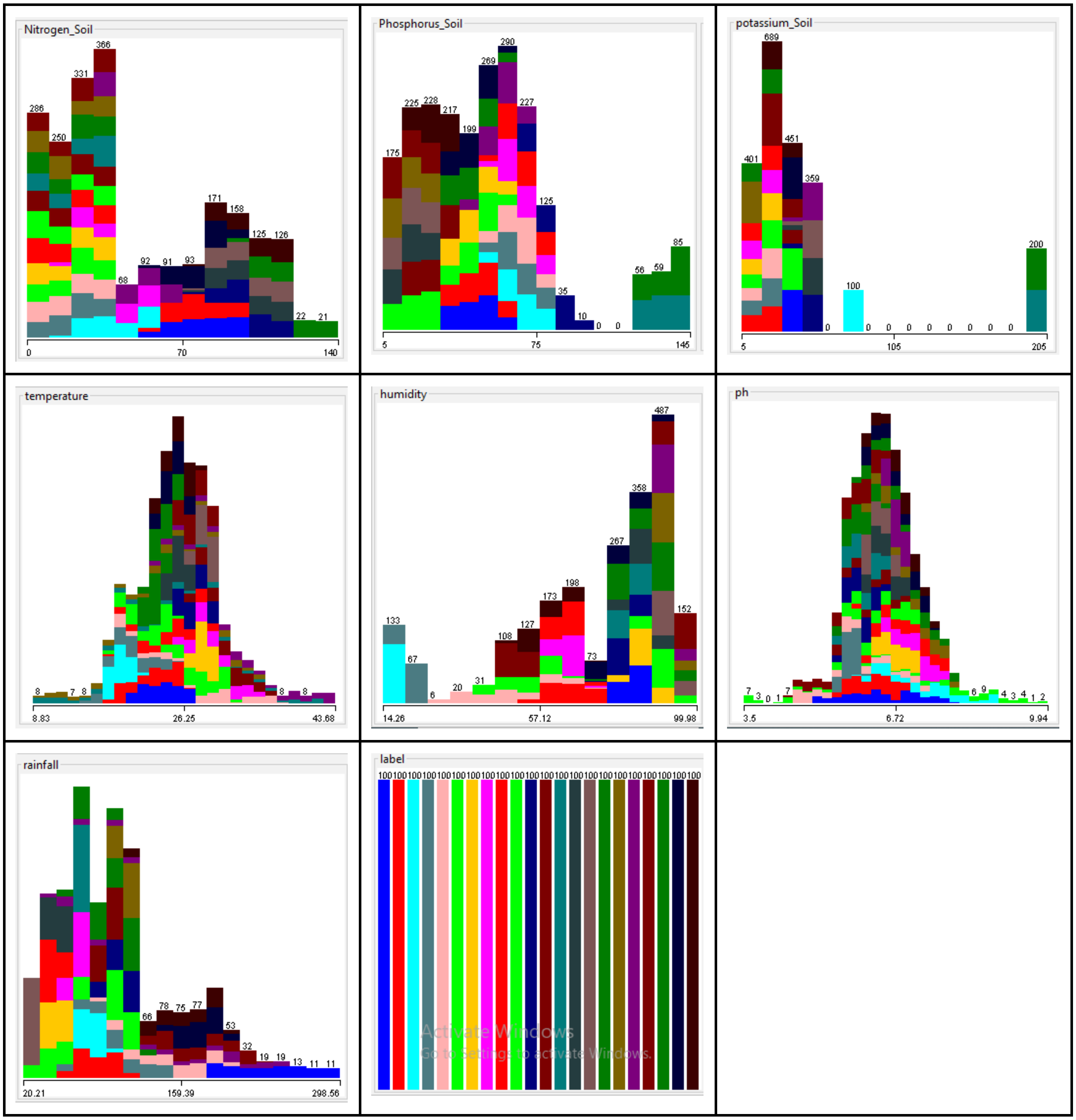


Fig: Visualization of attributes individually using Weka

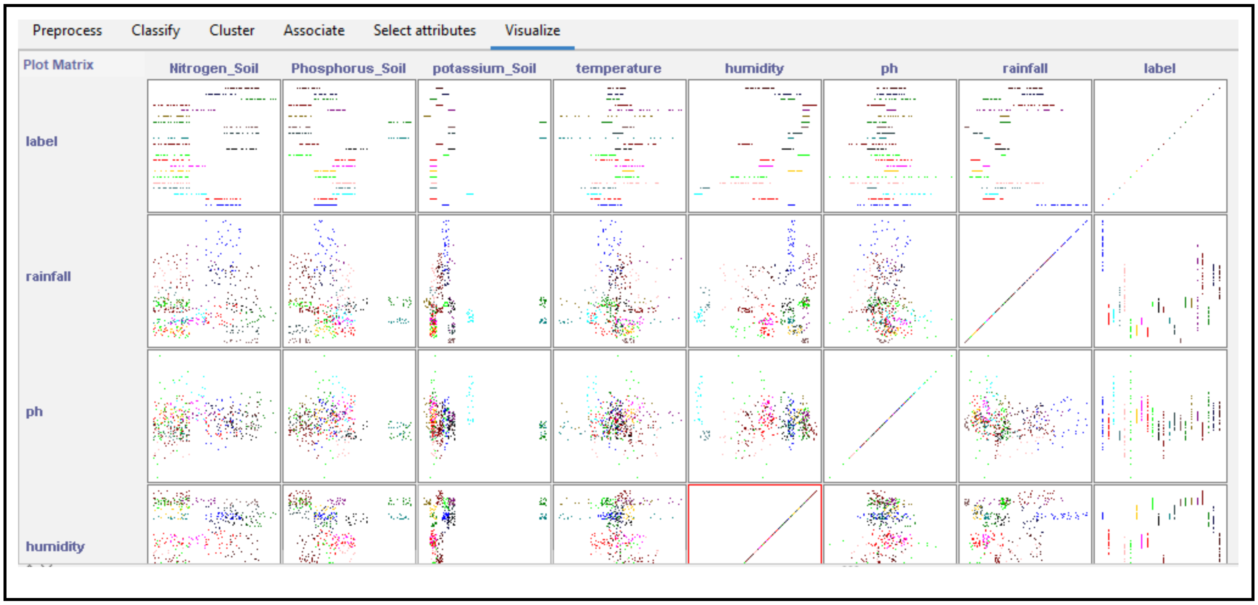


Fig: Visualization of attributes in matrix form

Fig: Join plot of label,pH using Weka

When we select an instance from the above plot we can get the following information about the instance.

Weka: Instance info

```
Plot : Master Plot
Instance: 112
  Nitrogen_Soil : 90.0
Phosphorus_Soil : 49.0
 potassium_Soil : 21.0
    temperature : 24.84016732
       humidity : 68.3584573
             ph : 6.472523287
       rainfall : 74.05474936
          label : maize

Plot : Master Plot
Instance: 133
  Nitrogen_Soil : 84.0
Phosphorus_Soil : 57.0
 potassium_Soil : 25.0
    temperature : 22.53510514
       humidity : 67.99257471
             ph : 6.489040367
       rainfall : 64.40866039
          label : maize

Plot : Master Plot
Instance: 143
  Nitrogen_Soil : 74.0
Phosphorus_Soil : 48.0
 potassium_Soil : 17.0
    temperature : 21.63162756
       humidity : 60.27766379
             ph : 6.430616465
       rainfall : 69.21803098
          label : maize
```

Weka: Instance info

```
Plot : Master Plot
Instance: 1040
  Nitrogen_Soil : 91.0
Phosphorus_Soil : 84.0
 potassium_Soil : 52.0
    temperature : 29.14827211
       humidity : 78.71024836
             ph : 6.390741836
       rainfall : 117.536781
          label : banana

Plot : Master Plot
Instance: 1047
  Nitrogen_Soil : 106.0
Phosphorus_Soil : 70.0
 potassium_Soil : 55.0
    temperature : 25.86824781
       humidity : 78.52399914
             ph : 5.74055541
       rainfall : 116.3019555
          label : banana
```

IV. Result and Analysis

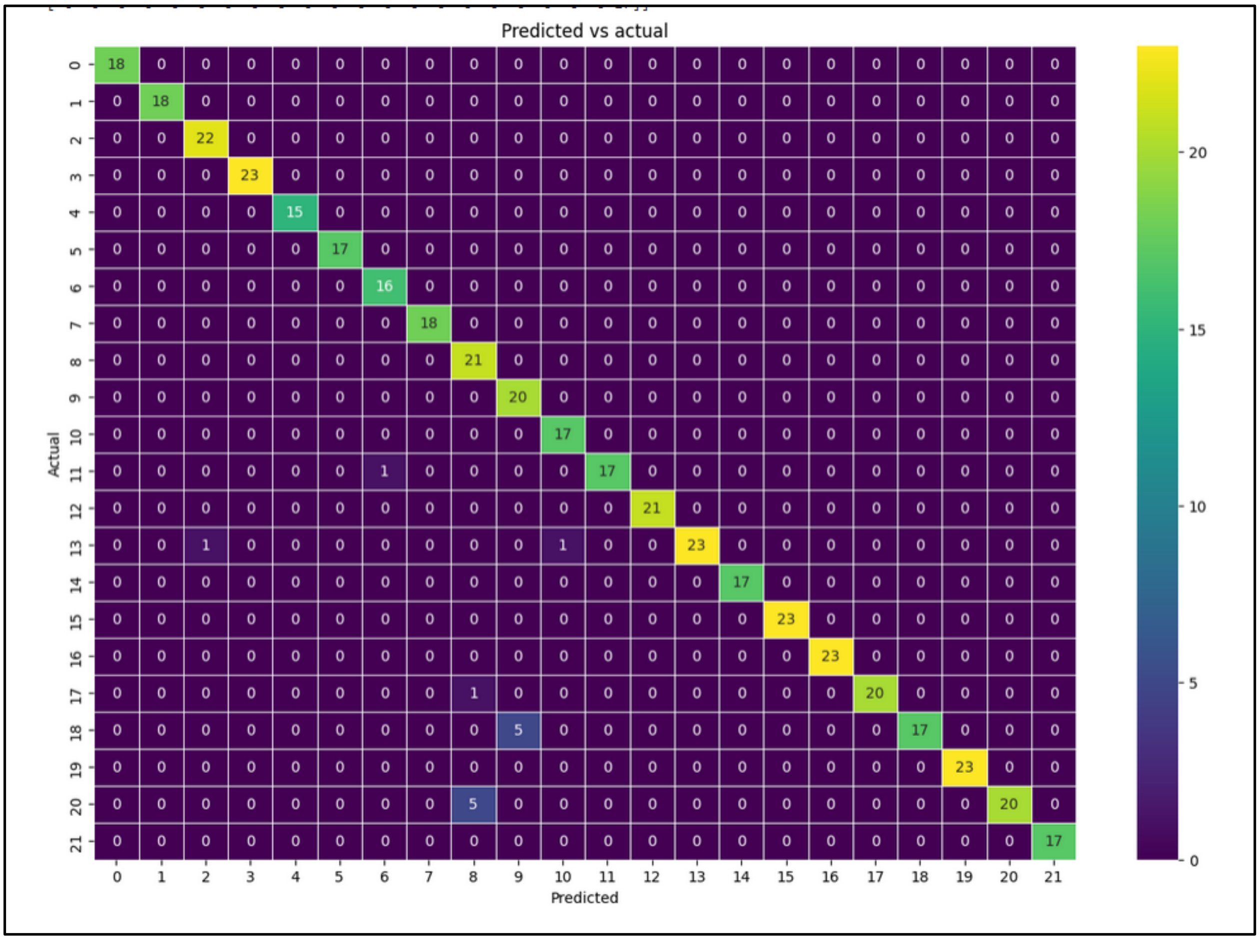
Here we use, **Fold = 5, test_size = .2, random_state = 0**

- **For Decision Tree:**

  Cross-Validation Scores for Decision Tree: [0.99090909 0.98409091 0.99318182 0.98863636 0.98181818]

  Mean Cross-Validation Score for Decision Tree: 0.9877272727272727

  Confusion Matrix for Decision Tree:

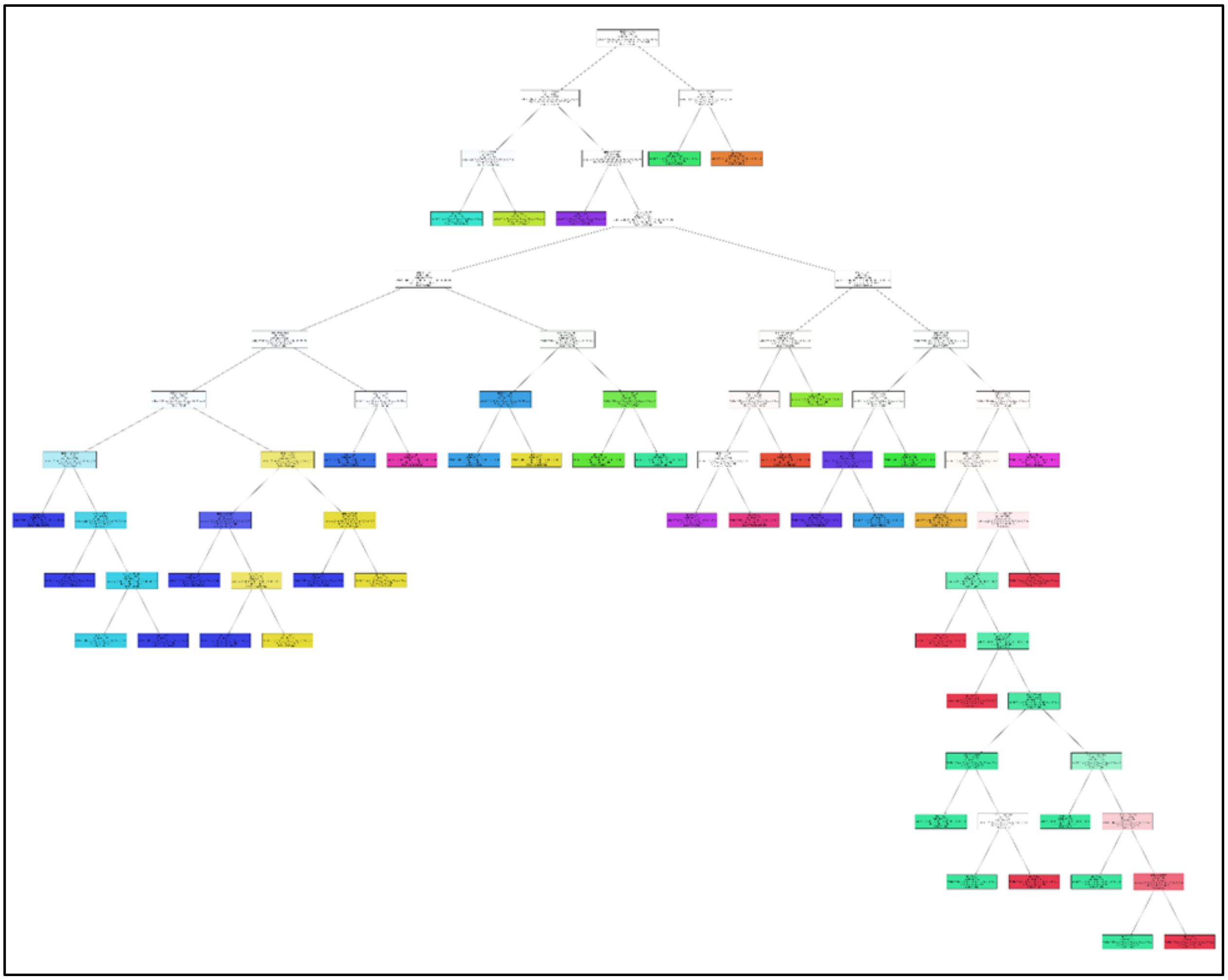


  Accuracy For Decision Tree: 0.988636

  Precision For Decision Tree: 0.987716

  Recall For Decision Tree: 0.989545

  F1 Score For Decision Tree: 0.988603

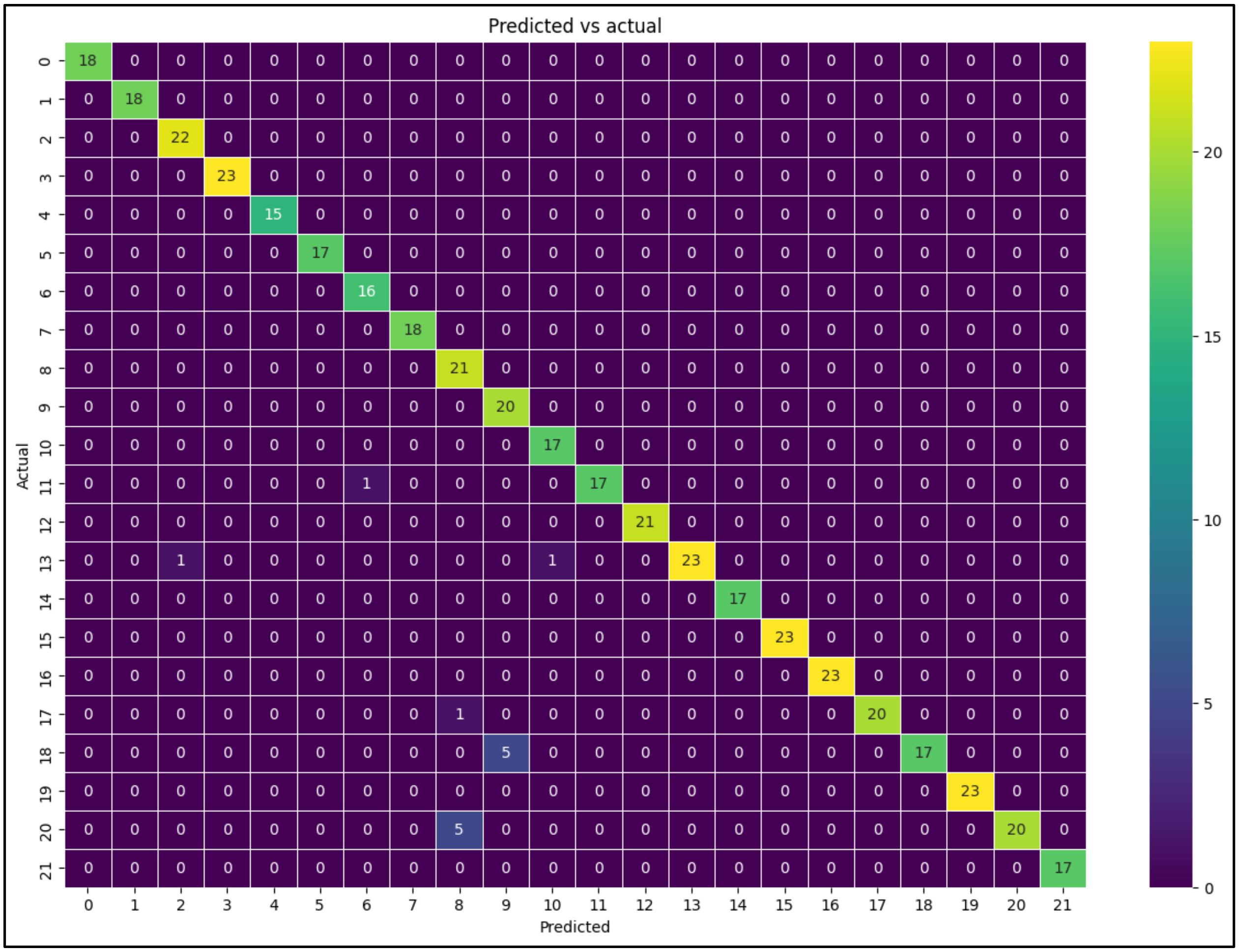


- **<u>For KNN(K-Nearest Neighbor):</u>**

  Cross-Validation Scores for KNN: [0.96590909 0.95909091 0.97954545 0.96363636 0.96818182]

  Mean Cross-Validation Scores for KNN: 0.9672727272727272

  Confusion Matrix for KNN:

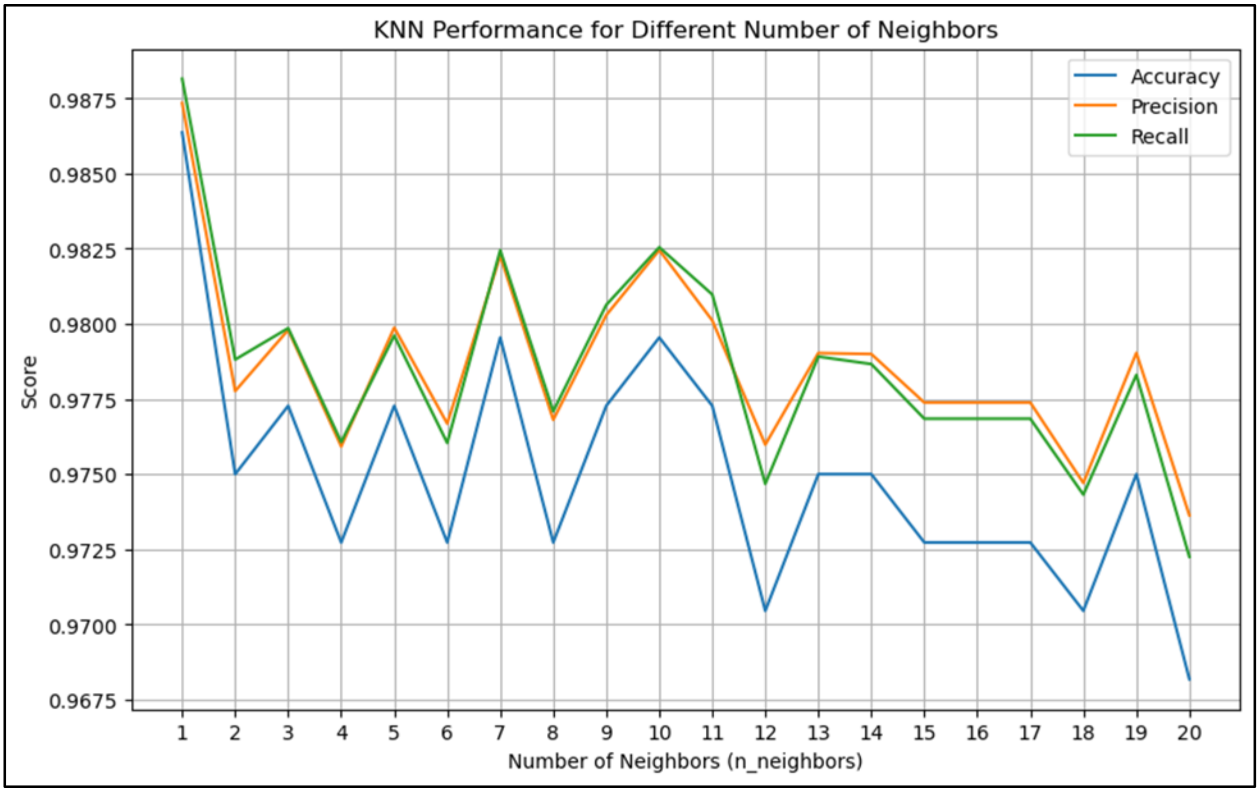


Accuracy For KNN: 0.979545

Precision For KNN: 0.981214

Recall For KNN: 0.979545

F1 Score For KNN: 0.979525

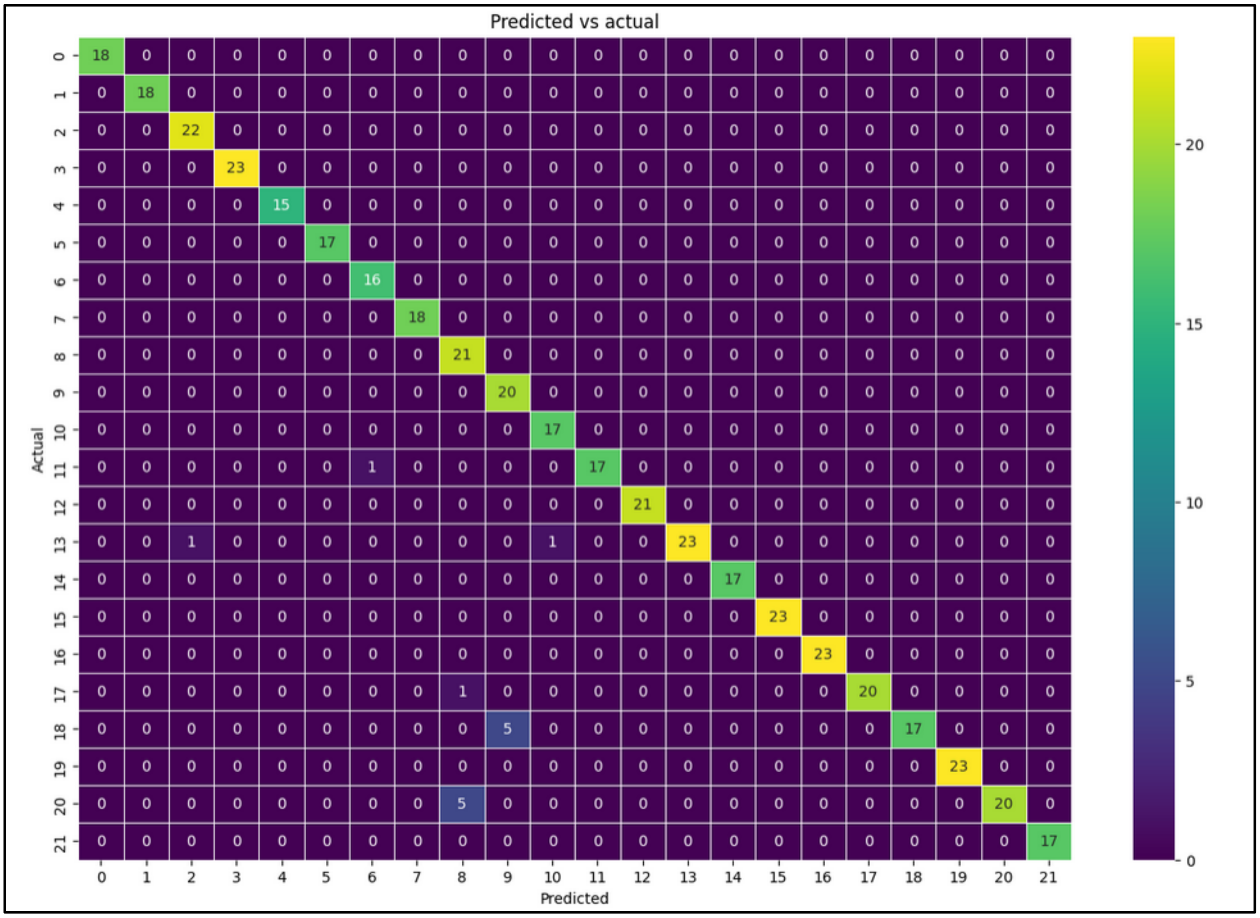


- **For Naive Bayes:**

  Cross-Validation Scores for naive bayes: [0.99772727 0.99545455 0.99545455 0.99545455 0.99090909]

  Mean Cross-Validation Score for naive bayes: 0.9950000000000001

  Confusion Matrix for naive bayes:



Accuracy For Naive Bayes: 0.995000

Precision For Naive Bayes: 0.995451

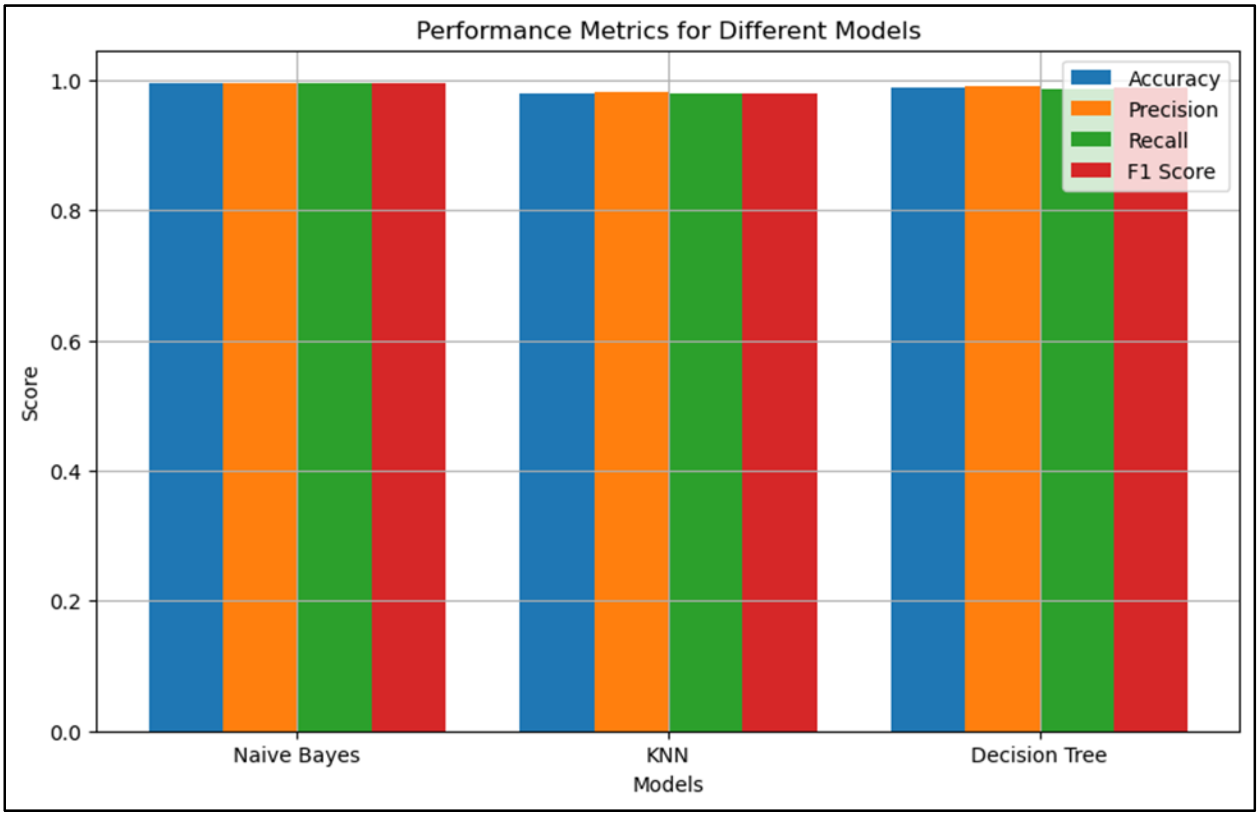
Recall For Naive Bayes: 0.995000

F1 Score For Naive Bayes: 0.994979

<u>Comparison of the performance metrics among different Models:</u>

Model Name	Accuracy	Precision	Recall	F1 Score
Decision Tree	0.988636	0.987716	0.989545	0.988603
KNN	0.979545	0.981214	0.979545	0.979525
Naive Bayes	0.995000	0.995451	0.995000	0.994979

<u>For better visualization, we can consider the below bar chart:</u>



V. Conclusion

In conclusion, this research project has presented optimized crop recommendation models to predict the best crops to grow using multiple machine learning algorithms. The technique is scalable and easily adapted to new data and regions or countries.

The results of this study have several positive implications for the agricultural industry. One key advantage of this system is its ability to enhance productivity and resource utilization. In addition to helping farmers to achieve better financial results, this promotes ecologically friendly and sustainable farming methods. Furthermore, the machine learning-based method incorporates real-time data to continuously learn and adapt, allowing recommendations to be improved over time. Because of its versatility, the system is guaranteed to stay pertinent and responsive to shifting climatic trends and changing farming techniques.

In summary, the introduced technique stands as a valuable asset for the agricultural sector, providing a scalable, accurate, and user-friendly solution that meets with the needs of farmers, governments, and businesses alike.

VI. References

[1] D. Dahiphale, Devendra Dahiphale, Pratik Shinde, Koninika Patil and Vijay Dahiphale, "Smart Farming: Crop Recommendation using Machine Learning with Challenges and Future Ideas," *JOURNAL OF IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE.*

[2] D. Gosai, Chintal Raval, Rikin Nayak, Hardik Jayswal and Axat Patel, "Crop Recommendation System using Machine Learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* vol. 7, no. 3, pp. 554-557, May-June - 2021.

[3] J. Neha, Gawali Sayali, Tikole Snehal, Gandhale Pradnya and Prof. Sayyed J.I, "CROP RECOMMENDATION SYSTEM," *International Journal of Advance Research, Ideas, and Innovations in Technology (IJARIIE),* no. Vol-9 Issue-6, p. 194, 2023.

[4] P. G. Dashmukhe, Ganesh Bharte, Chetan Bhalerao, Shripal Ahire and Bhushandigvijay Thakare, "CROP RECOMMENDATION SYSTEM USING MACHINE LEARNING," *International Research Journal of Modernization in Engineering Technology and Science,* vol. 05, no. 03, March-2023.