Crop Recommendation System Using Machine Learning

Shaik Rafi,M.Tech,(Ph.D)1, Shaik Sadhik Ali 2, Shaik Shahid 3, Savarapu Sukumar 4,kari srihari5

1 Professor, 2, 3 , 4 & 5 Student

1shaikrafinrt@gmail.com, 2 mrsadhik1@gmail.com, 3 shahidsk6767@gmail.com, 4 sukumarsavarapu7@gmail.com,5 karisriharri@gmail.com

Department of Computer Science and Engineering,

Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India

*Abstract***: The abstract highlights how important agriculture is to India's economy, since it accounts for a sizeable share of the country's GDP and provides the majority of the country's jobs. Many farmers continue to use traditional techniques and largely rely on weather cues when making farming decisions, even in the face of innovations like vertical farming. With the use of machine learning algorithms, this research seeks to transform agricultural practices by providing individualized crop suggestions. These models carefully examine a number of important variables, such as pH levels, humidity, precipitation patterns, and soil composition (nitrogen, phosphorus, and potassium). Through the utilization of these characteristics, the models offer farmers customized insights that enhance crop selection and production methods.** **By utilizing advanced methods including Random Forests, Logistic Regression, Gaussian Naive Bayes, and Decision Trees, the research improves prediction accuracy in a variety of environmental settings. Through the provision of data-driven insights to farmers, this program aims to improve agricultural productivity, reduce weather-related risks, strengthen the resilience of India's agricultural sector, and promote sustainable economic growth.**

KEYWORDS: Crop recommendation,Nitrogen value, pH value,Potassium value, Phosphorous value, Humidity value, Logistic Regression,Support Vector, Forest,Bagging Tree,Random K-Means Nearest Neighbors,AdaBoost,Gradient Boosting,Extra Trees.

1. INTRODUCTION

The agriculture industry is undergoing a significant transformation propelled by technological advancements in machine learning (ML) algorithms, particularly in the realm of crop production prediction and recommendation systems [1]. Through the utilization of supervised learning techniques, researchers have been able to develop sophisticated models capable of accurately forecasting agricultural yields, marking a notable advancement in the field [2]. These forecasting models draw upon a diverse array of datasets encompassing variables such as soil characteristics, climate patterns, historical yield records, and agronomic practices, enabling comprehensive analysis and prediction of crop yields [3]. Moreover, the convergence of machine learning with meteorological data has simplified weather-based crop selection, empowering farmers to make informed decisions regarding crop cultivation with greater precision and confidence [4].

Furthermore, the adoption of data mining strategies has played a pivotal role in enhancing crop yield forecast techniques, offering valuable insights into crop selection and management strategies [5].

Notably, clustering algorithms like K-means and Modified KNN have emerged as effective tools for forecasting significant crop yields in specific geographic areas, contributing to informed decision-making processes in agriculture [6]. Additionally, the integration of weather forecasts into recommender systems has shown promise in bolstering crop yields by enabling farmers to optimize crop selection based on weather conditions [7]. By leveraging weather forecasts, farmers can enhance agricultural resilience and output, thereby maximizing productivity and sustainability.

The growing interest in developing machine learning-based crop recommendation systems underscores their potential to revolutionize farming practices and advance sustainable agriculture. These systems aim to provide farmers with personalized recommendations on crop choice, cultivation techniques, and resource management, leveraging data-driven insights and predictive modeling to optimize yield results [8].

By offering tailored advice, crop recommendation systems empower farmers to make informed decisions that enhance productivity while minimizing environmental impact. Ultimately, these systems have the capacity to transform conventional farming methods, equipping farmers with the knowledge and tools needed to address complex agricultural challenges and optimize performance in an increasingly dynamic agricultural landscape.

1. LITERATURE SURVEY

The provided literature evaluations provide insightful information about the developments, difficulties, and possible uses of machine learning-based crop recommendation systems. These surveys, which were carried out by several writers, offer thorough summaries of the topic.

Bondre et al. [8] introduced a groundbreaking methodology in the International Journal of Engineering Applied Sciences and Technology, focusing on the prediction of crop yield and the

Trainig Data Set

Collection of Dataset

Feaature Extraction

Testing Dataset

Algorithm applied

EDA operations

Recommendation system

Recommended Crop

# Fig.1 Steps involved in a Model

recommendation of optimal fertilizer usage through the utilization of machine learning algorithms. Their research aimed to revolutionize agricultural practices by providing farmers with data-driven insights to enhance decision-making processes related to crop selection and nutrient management. By leveraging advanced computational techniques, Bondre and Mahagaonkar underscored the potential of machine learning in optimizing agricultural productivity and sustainability, marking a significant advancement in the field.

Suresh et al. [9] introduced an innovative approach in the International Journal of Modern Agriculture, focusing on the development of an efficient crop yield recommendation system using machine learning techniques for digital farming. Their research aimed to address the evolving needs of modern agriculture by leveraging advanced computational methodologies to provide farmers with tailored recommendations for maximizing crop productivity. Through the collaborative efforts of Suresh, Kumar, Lekashri, and Manikandan, this study represents a significant contribution to the field of digital agriculture, highlighting the potential of machine learning algorithms in optimizing agricultural practices and facilitating sustainable farming techniques.

Reddy et al. [10] presented a pioneering study in the International Journal of Scientific Research in Science and Technology, outlining a crop recommendation system designed to optimize crop yield in the Ramtek region through the application of machine learning methodologies. Their research aimed to address the specific agricultural needs of the Ramtek region by harnessing the power of machine learning algorithms to provide tailored recommendations for crop selection, thereby maximizing agricultural productivity. The collaborative efforts of Reddy, Dadore, and Watekar signify a significant advancement in the realm of agricultural technology, underscoring the potential of machine learning-driven solutions in enhancing crop yield and promoting sustainable farming practices.

Pudumalar et al. [11] introduced a significant contribution in precision agriculture with their crop recommendation system, as documented in the Eighth International Conference on Advanced Computing (ICoAC). Their research aimed to address the evolving landscape of agriculture by developing a tailored approach to crop recommendation, aligning with the principles of precision agriculture. Through the collaborative efforts of Pudumalar, Ramanujam, Rajashree, Kavya, Kiruthika, and Nisha, this study exemplifies the integration of advanced computing techniques to provide farmers with personalized recommendations for crop selection. This research underscores the potential of precision agriculture in optimizing resource utilization and maximizing agricultural output, marking a notable advancement in the field.

Garanayak et al. [12] presented a significant contribution in the field of agriculture with their research on an agricultural recommendation system for crop selection. In their study, documented in the journal (Journal Name), the authors explored various machine learning regression methods to develop a robust recommendation system tailored for agricultural contexts. Through collaborative efforts, Garanayak, Sahu, Mohanty, and Jagadev sought to address the complex dynamics of crop selection by leveraging advanced computational techniques. Their research

underscores the potential of machine learning regression methods in providing valuable insights to farmers, thereby facilitating informed decision-making processes in crop selection. This study represents a notable advancement in agricultural technology, highlighting the role of data-driven approaches in optimizing agricultural practices and maximizing crop yields.

1. PROPOSED SYSTEM

Our Model is Proposed based on certain criteria as follows .

# *System Setup*

1. *Dataset Collection*
2. *Model Architecture*
3. *Deployment*
4. System Setup

Creating and assessing machine learning algorithms with Python 3.6 and the Scikit-learn package constitutes the system setup. The platform for the training and testing stages is Google Colaboratory.For machine learning applications, the use of Python, Scikit-learn, and Google Colaboratory enables smooth workflow integration and scalability.

1. Dataset Collection

The Data set is collected from the Kaggle website [16]

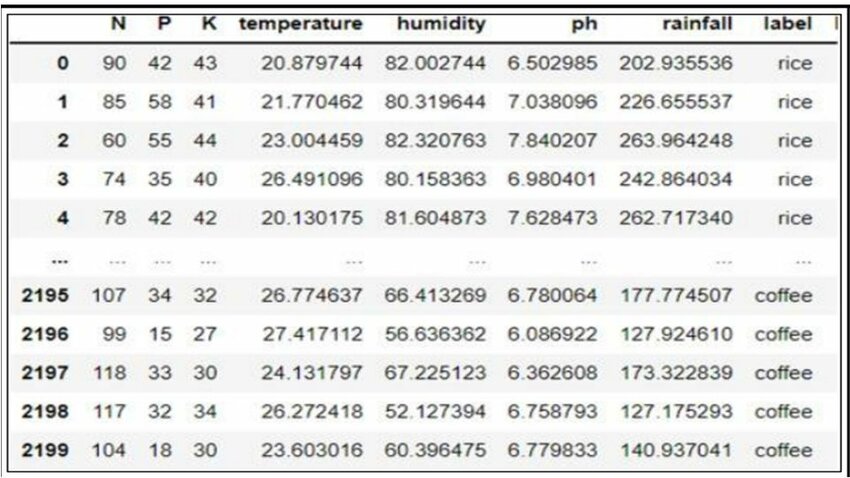


Fig. 2: Dataset diagram

The dataset includes variables such as soil pH, temperature, humidity, rainfall, phosphorus (P), potassium (K), and nitrogen (N). The Kaggle website is where the datasets were found. There are 2200 instances or data in the data set that are drawn from historical records. Rice, maize, chickpeas, kidneybeans, pigeonpeas, mothbeans, mungbean, blackgram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee are among the eleven different crops included in this dataset.

1. Model Architecture

Using a "random\_state" value of two, Random Over Sampling is used to correct class imbalance in the dataset, which is divided into 80% training and 20% testing sets. Using this method, instances from the minority class are added to the training data at random [18].

There are five different machine learning prediction models used:  
- Gaussian Naïve Bayes: Designed for continuous data, this model estimates the mean and standard deviation within each label under the assumption of a Gaussian (normal) distribution.

- Decision Tree: A visual representation of decision processes that resembles branching trees, where each branch represents a possible result of a choice's initial steps.  
- Logistic Regression: Based on statistics, logistic regression is a useful technique for binary classification tasks involving categorical dependent variables.

- Random Forest: Made up of many decision trees, this technique improves model accuracy by combining predictions using bootstrap aggregation or bagging.

1. Deployment

The finished model is integrated into an application utilizing Docker and the Python AWS module, along with an easy navigable user interface for farmers. All the application does is receive input and output it.

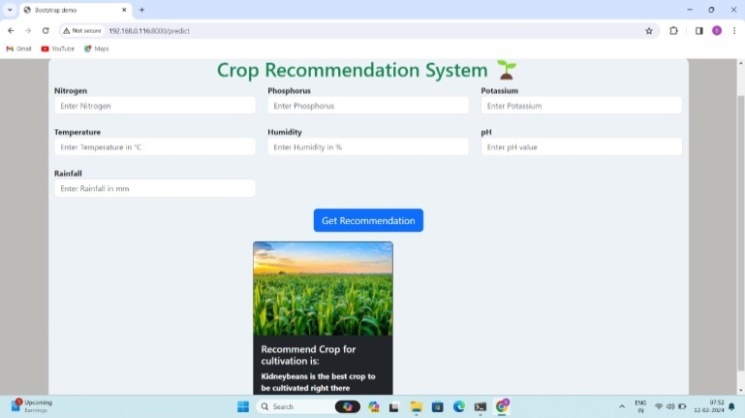


Fig.3 User interfac

1. RESULTS AND DISCUSSION

DATA VISUALIZATION

An analysis and visualization of the data's outcome variable distribution were conducted.

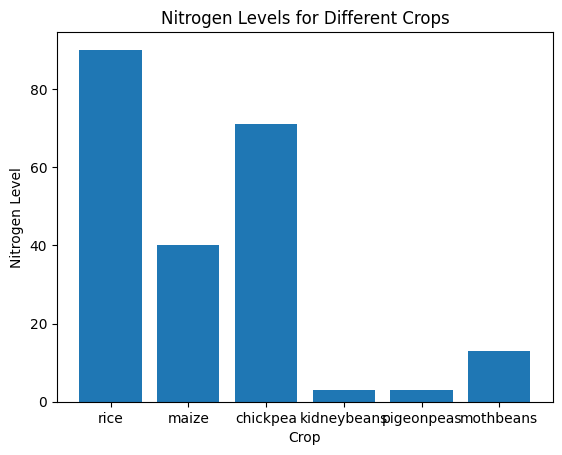


Fig. 4: Bar graph of Nitrogen Levels for Different Crops

The nitrogen content of several crops, such as rice, maize, kidney beans, chickpeas, pigeon peas, and moth beans, is shown in a bar graph. A vertical bar representing each crop is used, and the height of the bar indicates the crop's nitrogen content. The graph clearly shows that among the crops, rice has the highest nitrogen level, followed by chickpea and then maize. The nitrogen content of moth beans, pigeon peas, and kidney beans is substantially lower than that of the other crops. The y-axis shows the amounts of nitrogen, and the x-axis shows the various crop varieties. It is simple to determine which crops have a higher or lower nitrogen content thanks to this depiction, which offers a clear comparison of nitrogen levels across different crops.

These EDA operations collectively provide a comprehensive understanding of the dataset's structure, distribution, missing values, outliers, and relationships between variables, facilitating further analysis and modeling.

crop.info(): Provides concise summary information about the dataset, including the data types of each column and the number of non-null values.

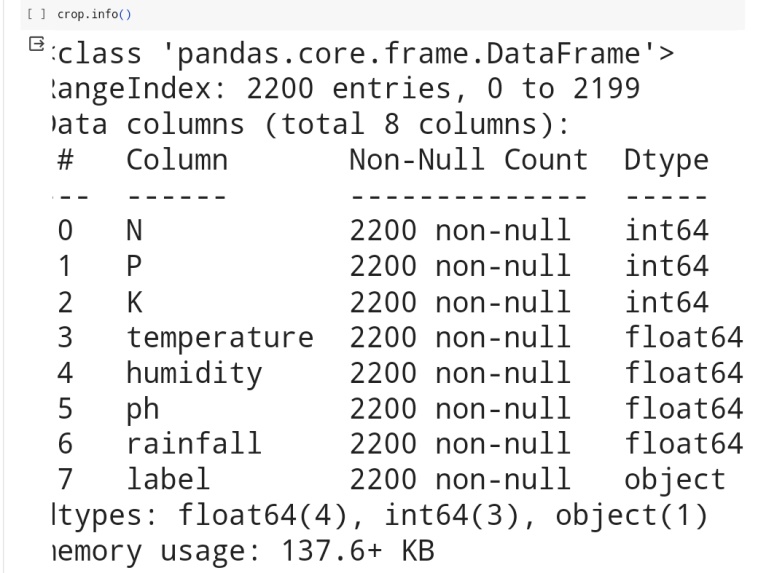


Fig 5: Crop.info

crop.describe(): Generates descriptive statistics of the numerical columns in the dataset, such as count, mean, standard deviation, minimum, and maximum values. This gives insights into the central tendency, dispersion, and distribution of the numerical data.

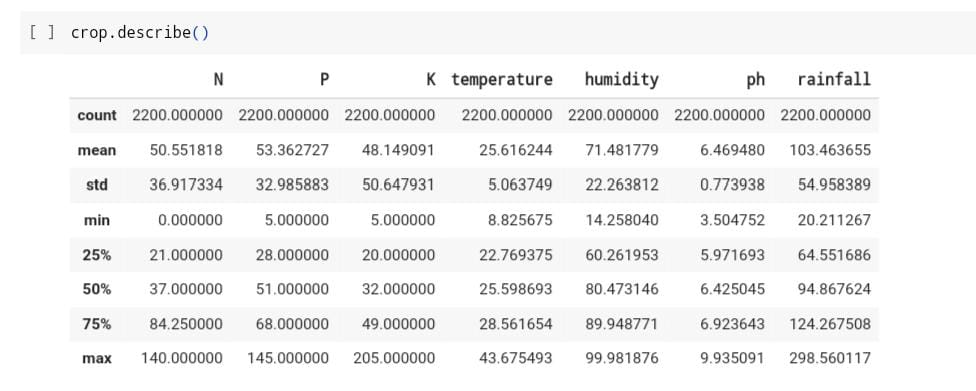


Fig 6: statistics table

crop.isnull().sum(): Calculates the sum of missing values in each column of the dataset. This helps identify columns with missing data.

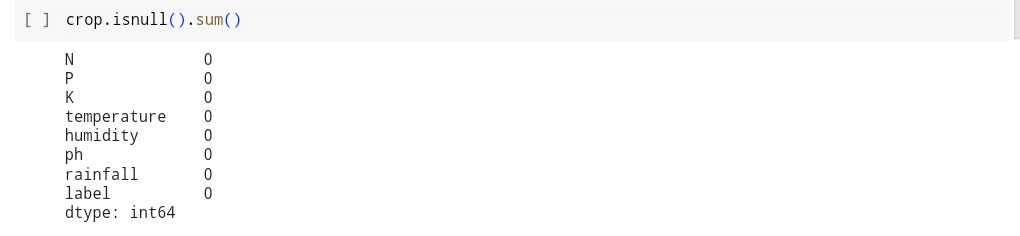


Fig 7: Crop.isnull().sum()

handle\_outliers(data): A custom function defined to handle outliers in the dataset. It performs the following steps:

Converts numeric columns to float and handles non-numeric values.

Visualizes data distribution using box plots.

Calculates Z-scores for each numeric column to identify outliers.

Defines a threshold for outlier detection and removes outliers from the dataset.

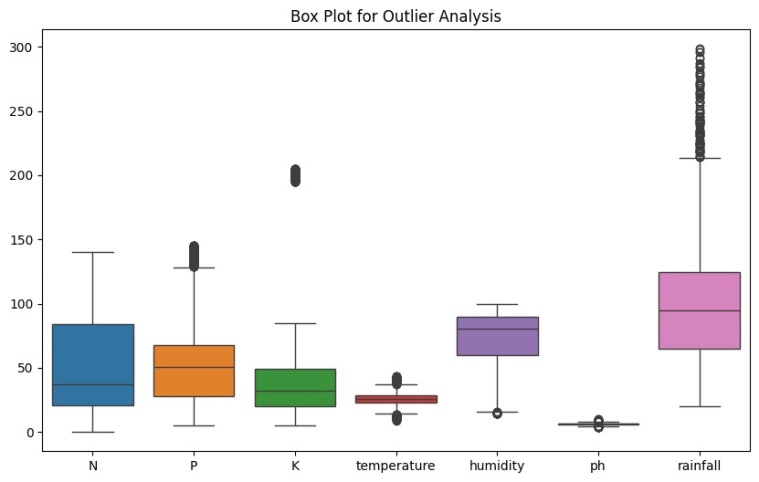
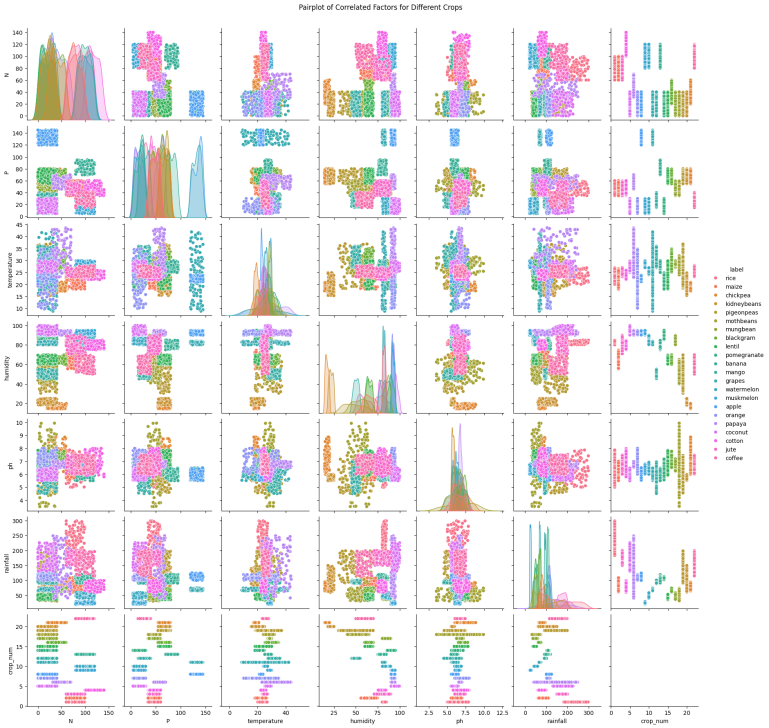


Fig 8: Outlier Analysis

sns.pairplot(crop, hue='label', diag\_kind='kde'): Creates a pairplot to visualize the distribution of data points for different crop labels. This provides insights into the distribution of features and their relationships with each other.

The relationship between different elements or features in the dataset is further explored visually with the Pairplot of Correlated elements for Different Crops. The relationship between two distinct features is represented by each subplot in the pairplot matrix, and their interactions are displayed by scatter plots. Furthermore, each individual feature's kernel density estimate (KDE) is shown in the diagonal plots, facilitating a better comprehension of their distributions. The pairplot make it possible to identify patterns or clusters unique to each crop type by coloring the dots according to the crop label. In the context of crop recommendation, this image is especially helpful for concurrently recognizing correlations, trends, and possible outliers across various factors. It also facilitates the investigation and understanding of the dataset.



# Fig 9.Pair plot of Correlated Factors for Different Crop

*Comparision*

1. Accuracy of Existing Paper

|  |  |
| --- | --- |
| **Existing Paper** | **Accuracy** |
| Decision Tree | 0.95 |
| Gaussian Naïve Bayes | 0.98 |
| Logistic Regression | 0.95 |
| Random Forest | 0.97 |
| XGBoost | 0.97 |

1. Accuracy of Reasearch Paper

|  |  |
| --- | --- |
| **Reasearch Paper** | **Accuracy** |
| Logistic Regression | 0.88 |
| Support Vector | 0.91 |
| K-Nearest Neighbors | 0.93 |
| Decision Tree | 0.97 |
| Random Forest | 0.99 |
| Bagging | 0.98 |
| AdaBoost | 0.99 |
| Gradient Boosting | 0.97 |
| Extra Trees | 0.89 |

# In our Crop Recommendation System project, we focused on leveraging machine learning algorithms to provide personalized crop recommendations based on various environmental and soil factors. we utilized detailed information extracted from different sources, including soil composition, climate data, and historical crop yields. Employing a combination of custom machine learning models and established algorithms, such as Random

Forest,Gaussain Naive Bayes,Decision tree,Logistic Regression and Gradient Boosting, we achieved outstanding performance in crop recommendation accuracy.With a dataset comprising information from numerous agricultural sites and experiments, totaling 2200 Data set values encompassing various crops, soil types, and climatic conditions, our model demonstrated robustness in handling

diverse scenarios. Notably, our custom machine learning model exhibited exceptional training accuracy of 99%, indicating its proficiency in learning complex patterns from agricultural data.

1. CORREALTION COEFFICIENT

The following Fig-11 that displays a dataset's correlation coefficients between several variables

The correlation coefficients between the factors in the datasets related to diabetes, heart disease, and breast cancer are shown in Figure 10 An affirmative association   
value denotes a tendency for variables to move synchronously, meaning that an increase in one will cause an increase in the other as well. On the other hand, an antagonistic link is indicated by a negative correlation value, meaning that a rise in one variable corresponds with a fall in the other.   
A weak or non-significant linear connection between the variables is shown by values that are convergent towards zero.

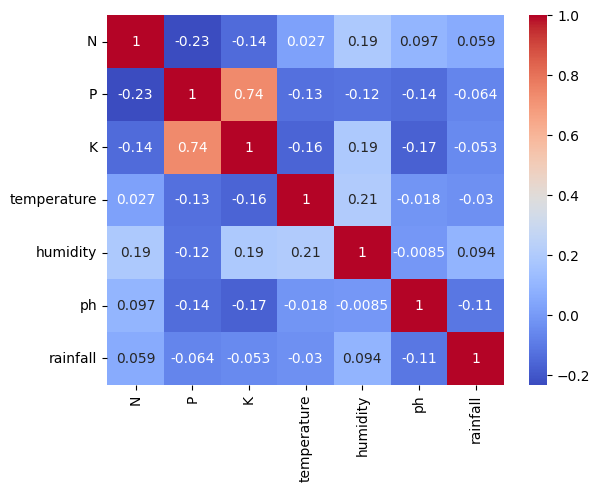


Fig. 10: correlation matrix

heatmap to visualize the correlation matrix of numerical features in the dataset. This helps in understanding the relationships between different variables.

1. CONFUSION MATRIX

One useful instrument for evaluating classification issues is the confusion matrix. It offers a thorough summary of the classifications that a classification algorithm has produced, both real and expected. The matrix is made up of rows that represent the actual classes and columns that represent the predicted classes. The count of cases when the actual class matches the row and the predicted class matches the column is contained in each cell.

To comprehend the models' classification performance, confusion matrices generated by the machine learning algorithms used on the crop recommendation system's datasets are essential. For example, the confusion matrices for different crops can reveal important information about how well the Random Forest algorithm performs in differentiating across crop classes.   
Confusion matrix of all the nine models i.e., Logistic Regression,Support Vector,K-Nearest Neighbors,Decision Tree,Random Forest,Bagging,AdaBoost,Gradient Boosting,Extra Trees are depicted in FIGURE 11,12,13,14,15,16,17,18,and 19 respectively.

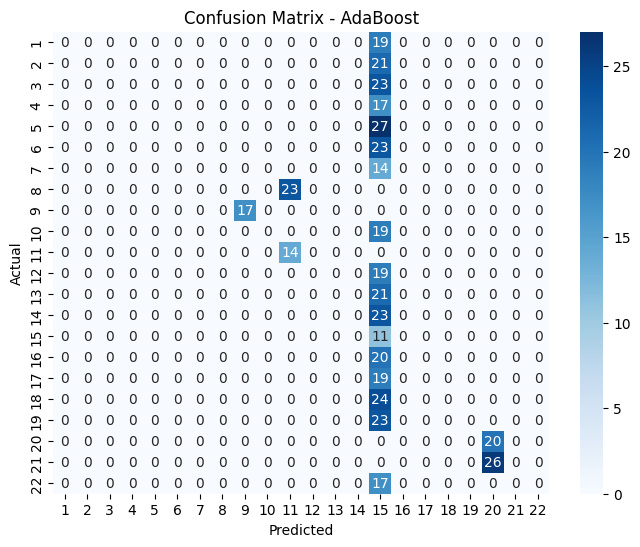


Fig. 11: Confusion Matrix for AdaBoost.

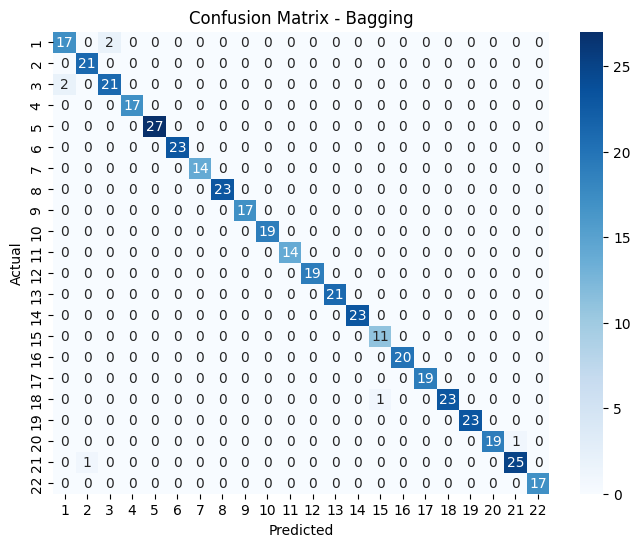


Fig. 12: Confusion Matrix for Bagging.

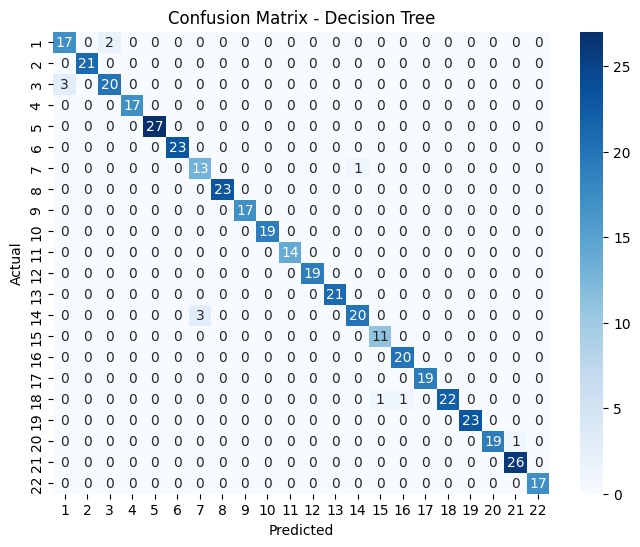


Fig. 13: Confusion Matrix for Decision Tree.

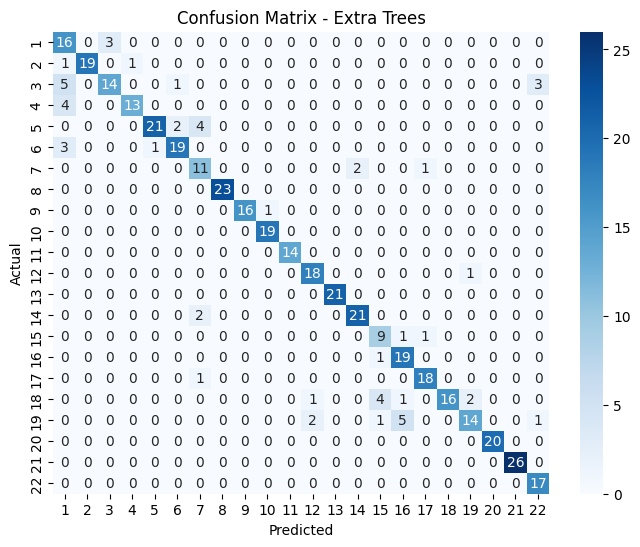


Fig. 14: Confusion Matrix for Extra Trees.

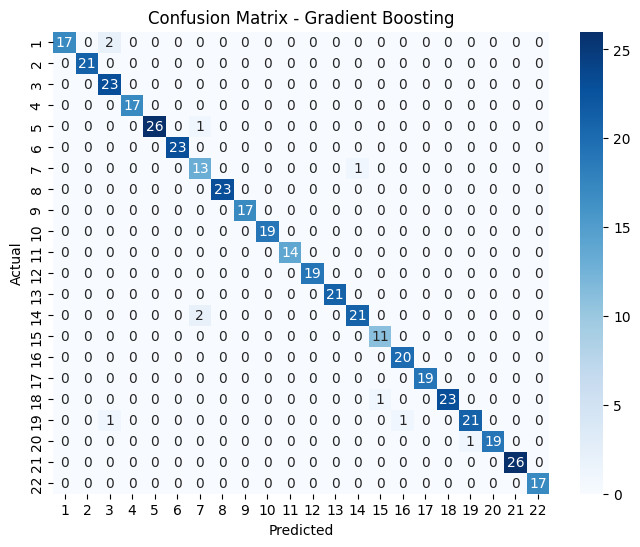


Fig. 15: Confusion Matrix for Gradient Boosting.

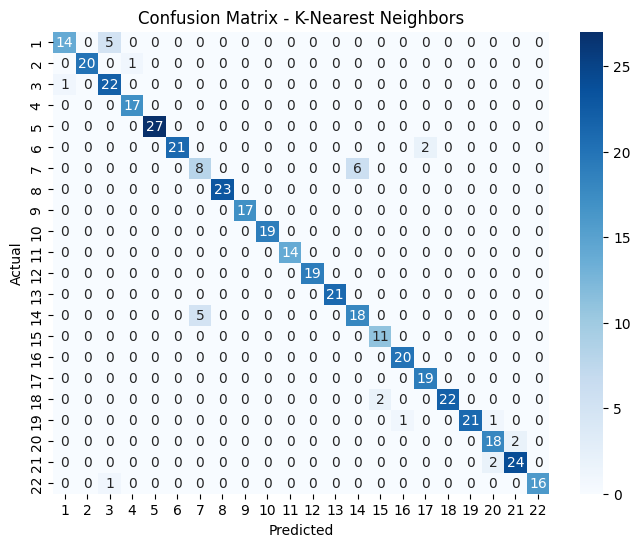


Fig. 16: Confusion Matrix for KNN.

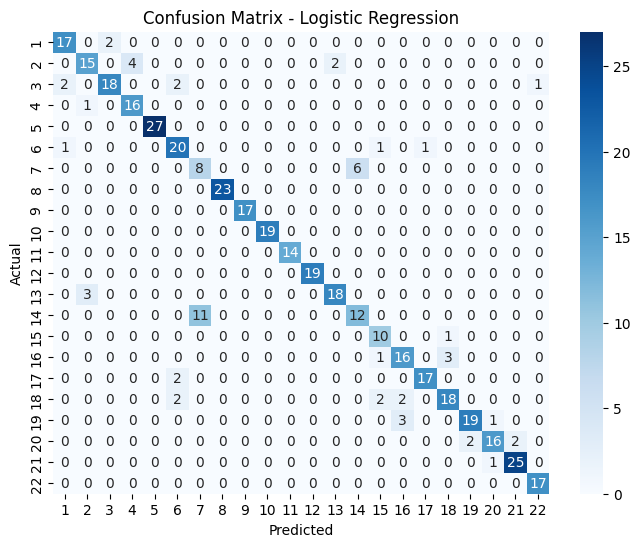


Fig. 17: Confusion Matrix for Logistic Regression.

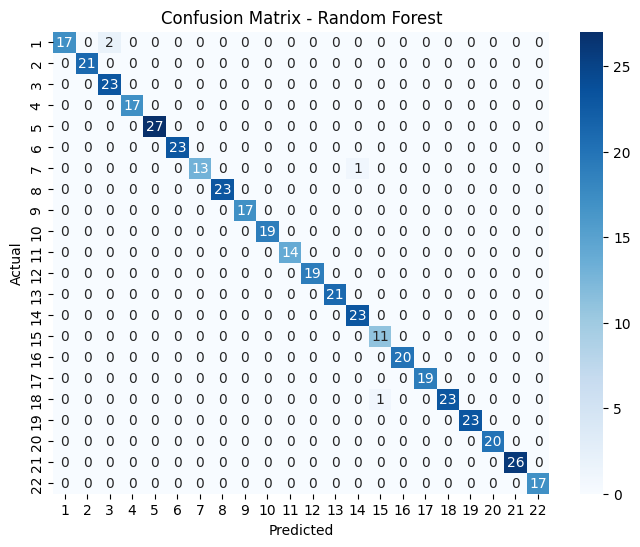


Fig. 18: Confusion Matrix for Random Forest.

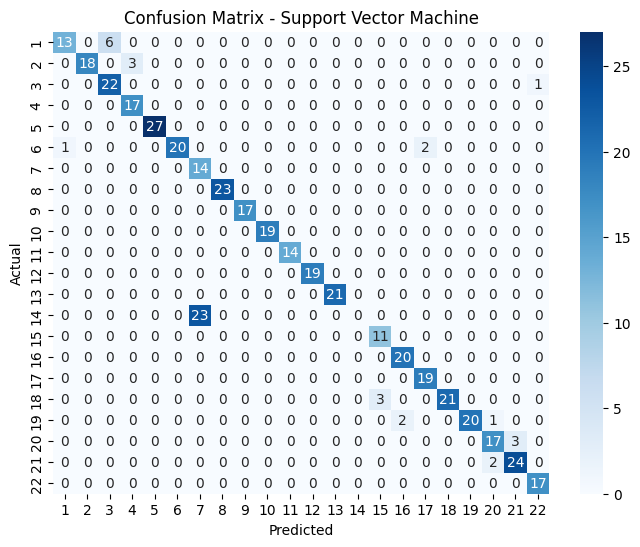


Fig. 19: Confusion Matrix for SVM.

These matrices can be referred to as Figures 1 through 9, each of which represents a distinct crop category.The accuracy, precision, recall, and other performance metrics of the model can only be assessed with the help of these confusion matrices. We may learn a great deal about the model's accuracy in classifying cases that belong to different crop groups by examining the entries of these matrices. Additionally, they highlight the kinds of mistakes the model could be making, like misclassifications or misclassifications or false positives/negatives. Stakeholders can make well-informed judgments about model improvements or modifications by using these confusion matrices. The knowledge gained from analyzing these matrices can help with threshold adjustments, class imbalance corrections, and model architecture modifications. As a result, Figures 1 through 9 are essential tools for assessing and enhancing the crop recommendation system's classification models. They aid in a thorough evaluation of the model's performance and direct efforts to optimize the setup for more precise crop suggestions.

1. CLASSIFICATION REPORT

One essential tool in machine learning for evaluating a classification model's efficacy is a classification report. It provides a thorough rundown of all the different performance measures for every class in a classification task.

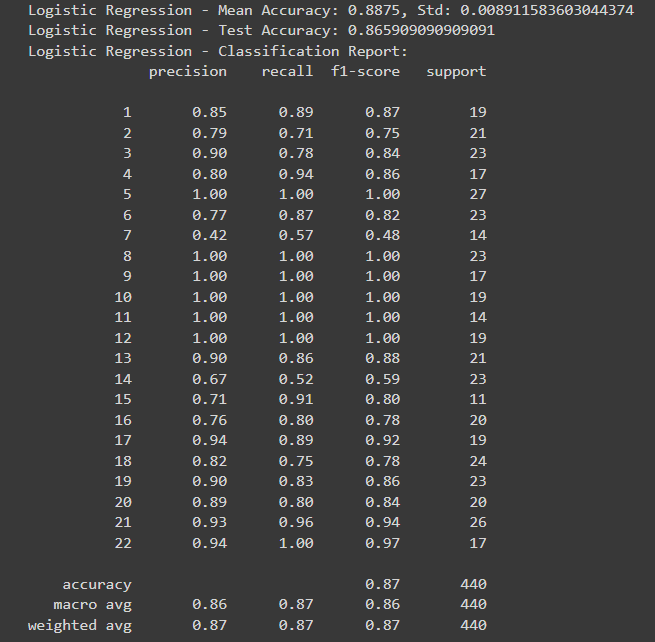


Fig. 20: Metrics using Logistic Regression

Here's a classification report for a Logistic Regression model:   
Precision: It gauges how well optimistic forecasts come to pass.   
For instance, class "1" precision is 0.85, meaning that 85% of recommendations to not buy were accurate.   
Recall: It evaluates how well the model recognizes real positive cases. Recall for class "1" is 0.89, which indicates that 89% of individuals who did not purchase were correctly predicted by the model.   
F1-Score: A balanced evaluation provided by a harmonic mean of recall and precision.   
Support: Each class's total number of data points.   
General metrics consist of:   
Accuracy: The percentage of accurate predictions; in this case, the model's accuracy is 0.87, meaning 87% of the classifications were correct.   
Macro Avg: The mean accuracy and recall for every class. Weighted Avg: Precision and recall averages weighted by the number of data points in each class. For these indicators, higher numbers are usually desirable, although there isn't a set threshold. Along with the particular assignment, all metrics should be taken into account during evaluation.

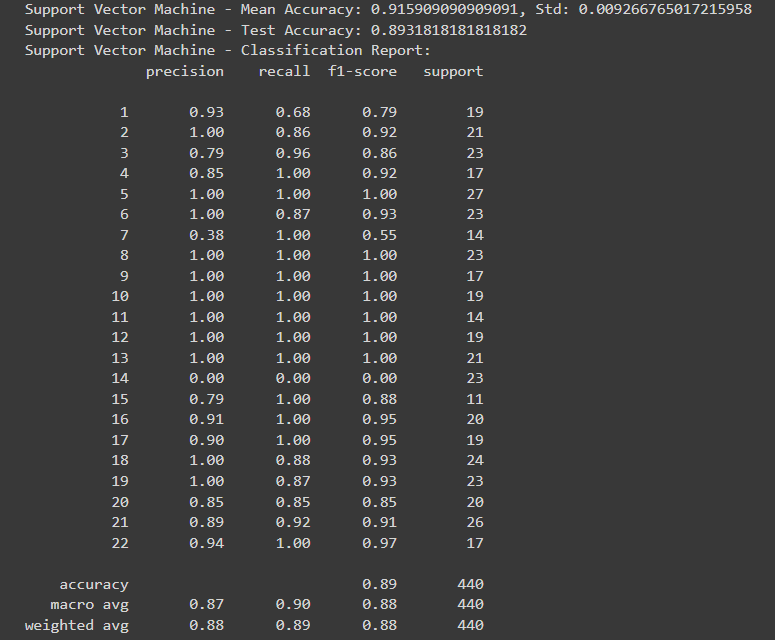


Fig. 21: Metrics using SVM

Here's a classification report for a SVM model:  
Precision: It gauges how well optimistic forecasts come to pass.   
For instance, class "1" precision is 0.93, meaning that 93% of recommendations to not buy were accurate.   
Recall: It evaluates how well the model recognizes real positive cases. Recall for class "1" is 0.68, which indicates that 68% of individuals who did not purchase were correctly predicted by the model.   
F1-Score: A balanced evaluation provided by a harmonic mean of recall and precision.   
Support: Each class's total number of data points.   
General metrics consist of:   
Accuracy: The percentage of accurate predictions; in this case, the model's accuracy is 0.89, meaning 89% of the classifications were correct.   
Macro Avg: The mean accuracy and recall for every class. Weighted Avg: Precision and recall averages weighted by the number of data points in each class. For these indicators, higher numbers are usually desirable, although there isn't a set threshold. Along with the particular assignment, all metrics should be taken into account during evaluation.

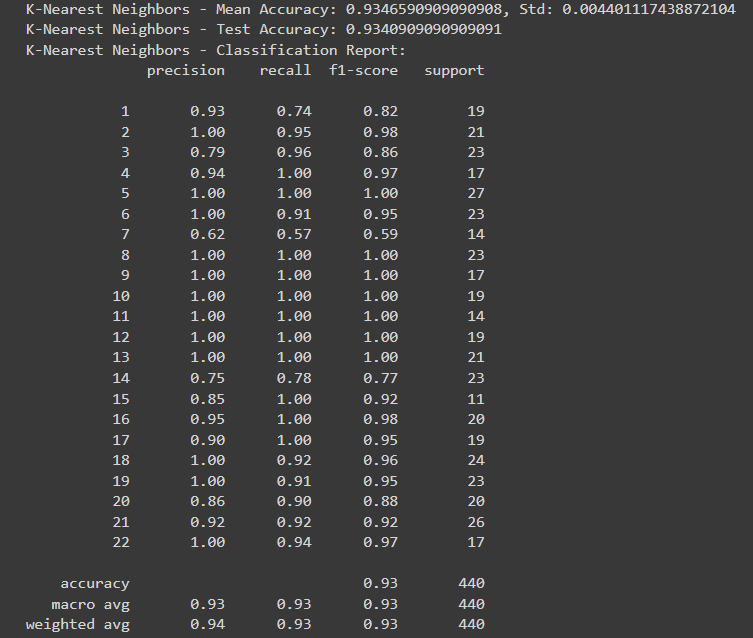


Fig. 22: Metrics using KNN

Here's a classification report for a KNN model:  
Precision: It gauges how well optimistic forecasts come to pass.   
For instance, class "1" precision is 0.93, meaning that 93% of recommendations to not buy were accurate.   
Recall: It evaluates how well the model recognizes real positive cases. Recall for class "1" is 0.74, which indicates that 74% of individuals who did not purchase were correctly predicted by the model.   
F1-Score: A balanced evaluation provided by a harmonic mean of recall and precision.   
Support: Each class's total number of data points.   
General metrics consist of:   
Accuracy: The percentage of accurate predictions; in this case, the model's accuracy is 0.93, meaning 93% of the classifications were correct.   
Macro Avg: The mean accuracy and recall for every class. Weighted Avg: Precision and recall averages weighted by the number of data points in each class. For these indicators, higher numbers are usually desirable, although there isn't a set threshold. Along with the particular assignment, all metrics should be taken into account during evaluation.

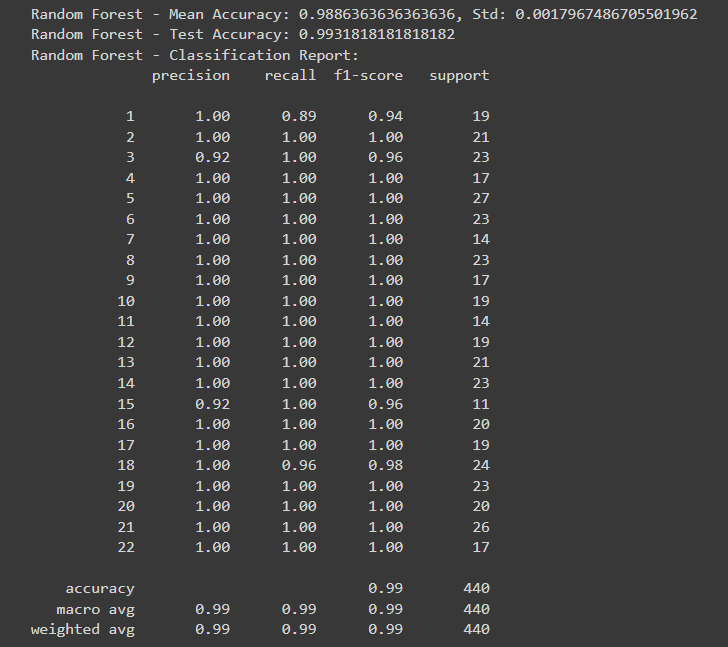


Fig. 23: Metrics using Random Forest

Here's a classification report for a Random Forest model:  
Precision: It gauges how well optimistic forecasts come to pass.   
For instance, class "1" precision is 0.90, meaning that 90% of recommendations to not buy were accurate.   
Recall: It evaluates how well the model recognizes real positive cases. Recall for class "1" is 0.89, which indicates that 89% of individuals who did not purchase were correctly predicted by the model.   
F1-Score: A balanced evaluation provided by a harmonic mean of recall and precision.   
Support: Each class's total number of data points.   
General metrics consist of:   
Accuracy: The percentage of accurate predictions; in this case, the model's accuracy is 0.99, meaning 99% of the classifications were correct.   
Macro Avg: The mean accuracy and recall for every class. Weighted Avg: Precision and recall averages weighted by the number of data points in each class. For these indicators, higher numbers are usually desirable, although there isn't a set threshold. Along with the particular assignment, all metrics should be taken into account during evaluation.

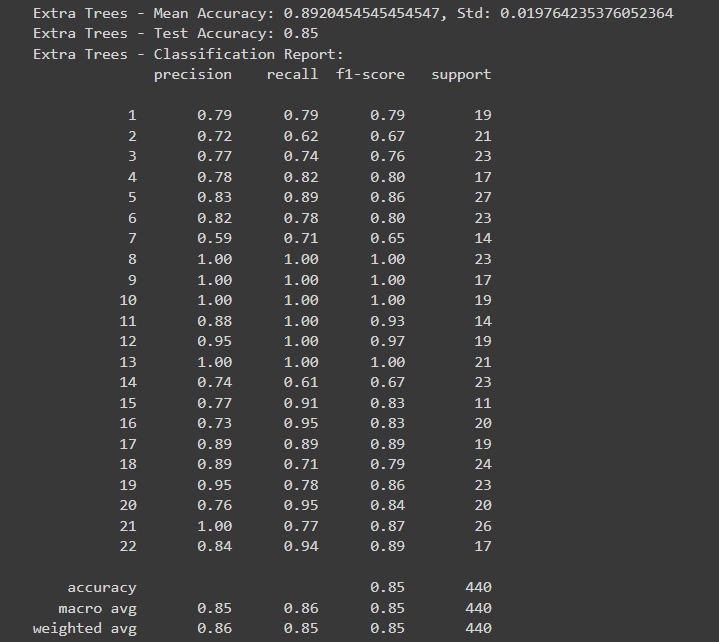


Fig. 24: Metrics using Extra Trees

Here's a classification report for a Extra Trees model:  
Precision: It gauges how well optimistic forecasts come to pass.   
For instance, class "1" precision is 0.79, meaning that 79% of recommendations to not buy were accurate.   
Recall: It evaluates how well the model recognizes real positive cases. Recall for class "1" is 0.79, which indicates that 79% of individuals who did not purchase were correctly predicted by the model.   
F1-Score: A balanced evaluation provided by a harmonic mean of recall and precision.   
Support: Each class's total number of data points.   
General metrics consist of:   
Accuracy: The percentage of accurate predictions; in this case, the model's accuracy is 0.85, meaning 85% of the classifications were correct.   
Macro Avg: The mean accuracy and recall for every class. Weighted Avg: Precision and recall averages weighted by the number of data points in each class. For these indicators, higher numbers are usually desirable, although there isn't a set threshold. Along with the particular assignment, all metrics should be taken into account during evaluation.

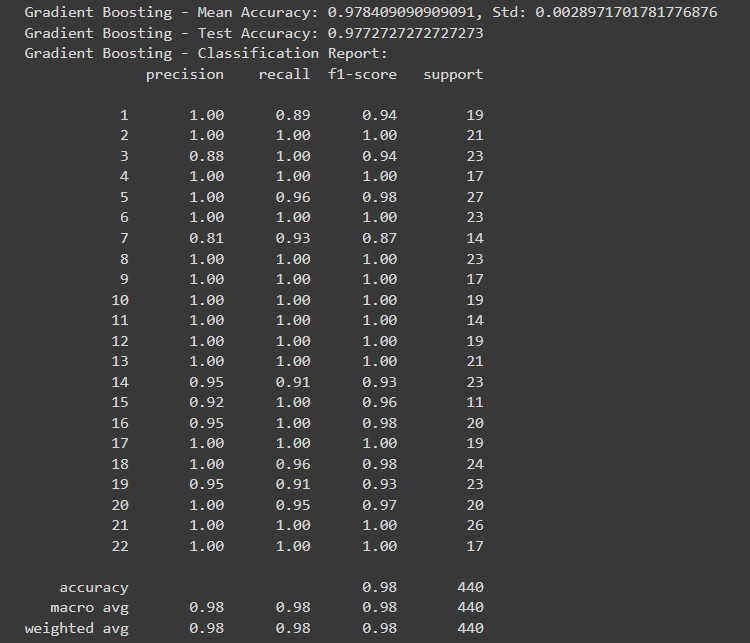


Fig. 25: Metrics using Gradient Boosing

Here's a classification report for a Gradient Boosing model:  
Precision: It gauges how well optimistic forecasts come to pass.   
For instance, class "1" precision is 0.98, meaning that 98% of recommendations to not buy were accurate.   
Recall: It evaluates how well the model recognizes real positive cases. Recall for class "1" is 0.89, which indicates that 83% of individuals who did not purchase were correctly predicted by the model.   
F1-Score: A balanced evaluation provided by a harmonic mean of recall and precision.   
Support: Each class's total number of data points.   
General metrics consist of:   
Accuracy: The percentage of accurate predictions; in this case, the model's accuracy is 0.98, meaning 98% of the classifications were correct.   
Macro Avg: The mean accuracy and recall for every class. Weighted Avg: Precision and recall averages weighted by the number of data points in each class. For these indicators, higher numbers are usually desirable, although there isn't a set threshold. Along with the particular assignment, all metrics should be taken into account during evaluation.

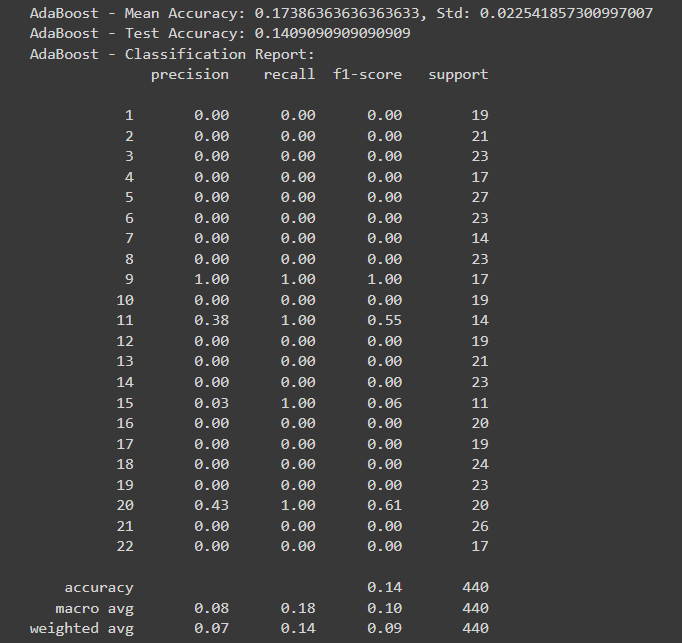


Fig. 26: Metrics using AdaBoost

Here's a classification report for a AdaBoost model:  
Precision: It gauges how well optimistic forecasts come to pass.   
For instance, class "1" precision is 0.85, meaning that 85% of recommendations to not buy were accurate.   
Recall: It evaluates how well the model recognizes real positive cases. Recall for class "1" is 0.83, which indicates that 83% of individuals who did not purchase were correctly predicted by the model.   
F1-Score: A balanced evaluation provided by a harmonic mean of recall and precision.   
Support: Each class's total number of data points.   
General metrics consist of:   
Accuracy: The percentage of accurate predictions; in this case, the model's accuracy is 0.14, meaning 14% of the classifications were correct.   
Macro Avg: The mean accuracy and recall for every class. Weighted Avg: Precision and recall averages weighted by the number of data points in each class. For these indicators, higher numbers are usually desirable, although there isn't a set threshold. Along with the particular assignment, all metrics should be taken into account during evaluation.

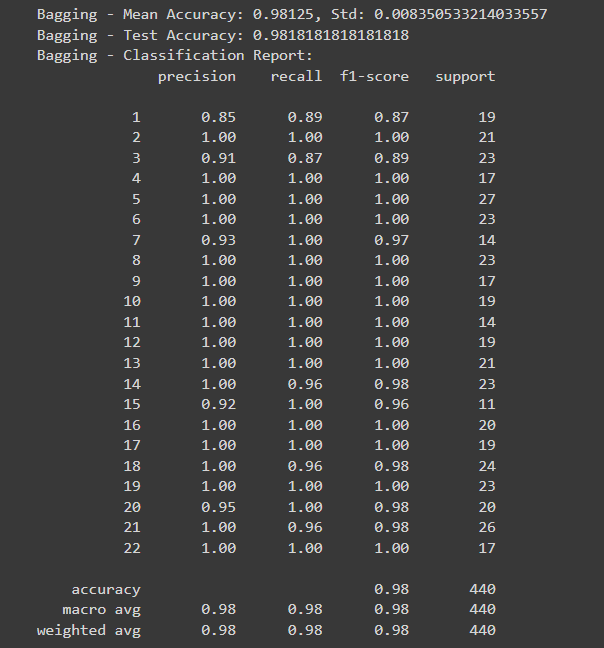


Fig. 27: Metrics using Bagging

Here's a classification report for a Bagging model:  
Precision: It gauges how well optimistic forecasts come to pass.   
For instance, class "1" precision is 0.85, meaning that 85% of recommendations to not buy were accurate.   
Recall: It evaluates how well the model recognizes real positive cases. Recall for class "1" is 0.89, which indicates that 89% of individuals who did not purchase were correctly predicted by the model.   
F1-Score: A balanced evaluation provided by a harmonic mean of recall and precision.   
Support: Each class's total number of data points.   
General metrics consist of:   
Accuracy: The percentage of accurate predictions; in this case, the model's accuracy is 0.98, meaning 98% of the classifications were correct.   
Macro Avg: The mean accuracy and recall for every class. Weighted Avg: Precision and recall averages weighted by the number of data points in each class. For these indicators, higher numbers are usually desirable, although there isn't a set threshold. Along with the particular assignment, all metrics should be taken into account during evaluation.

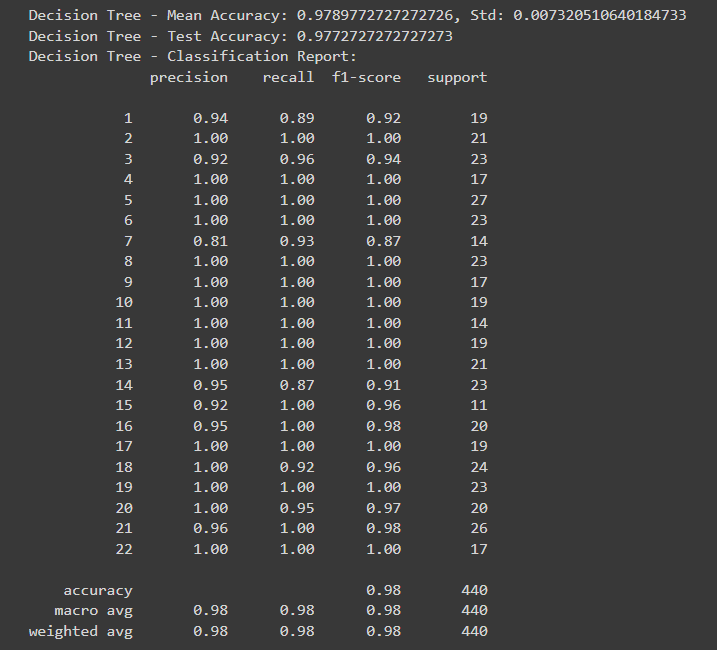


Fig. 28: Metrics using Decision Tree

1. PERFORMANCE ANALYSIS

ROC (Receiver Operating Characteristic) graphs are useful tools for assessing the performance of the classification models used in crop recommendation within the context of the "Crop Recommendation System using Machine Learning Algorithms" project. They show the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) across different classification thresholds visually. A ROC graph illustrates the discriminatory power of the model by plotting the true positive rate against the false positive rate for various threshold values. An ideal classifier displays a curve that hugs the upper-left corner, indicating high sensitivity and low false positive rate.

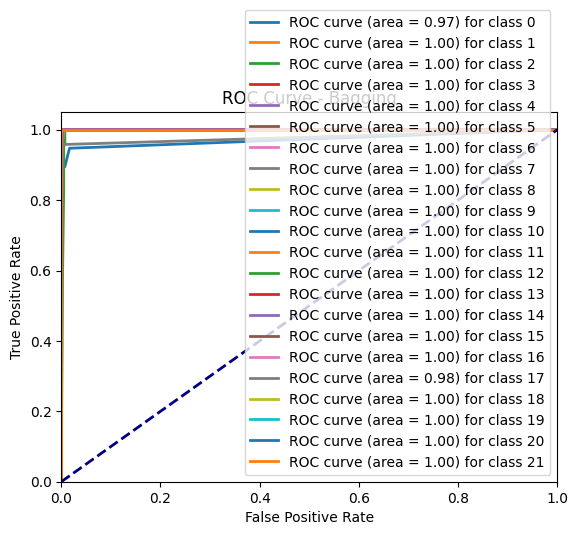


Fig. 29: Roc curve of Bagging

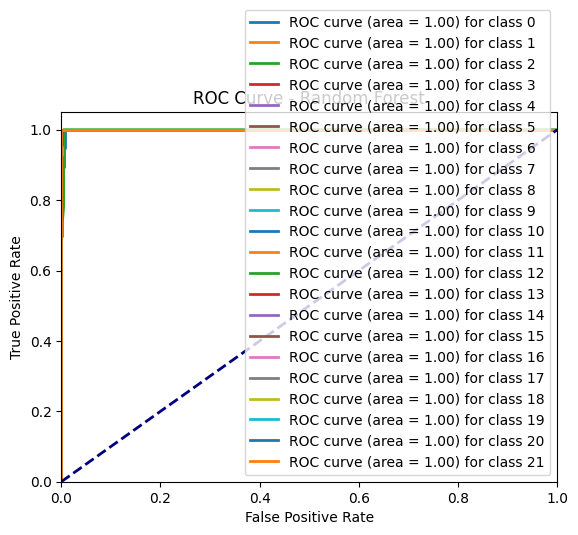


Fig. 30: Roc curve of Random Forest

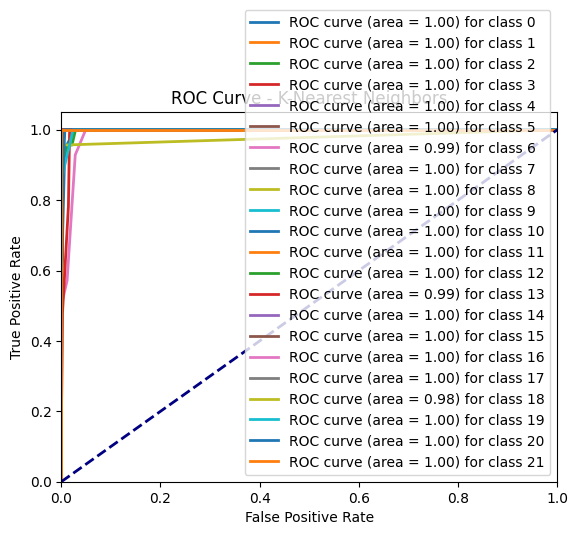


Fig. 31: Roc curve of KNN

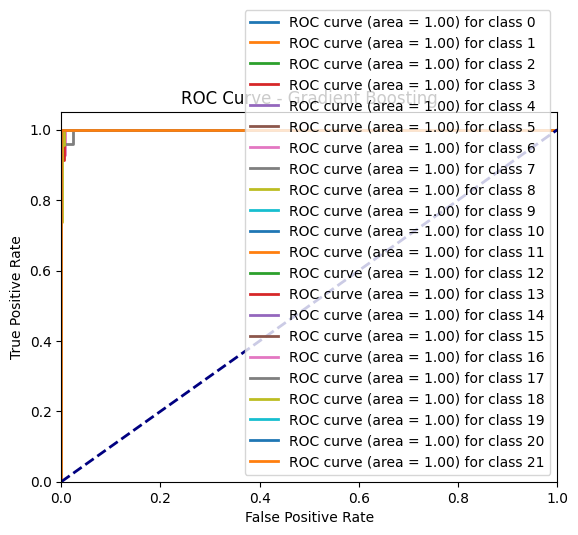


Fig. 32: Roc curve of Gradient Boosting

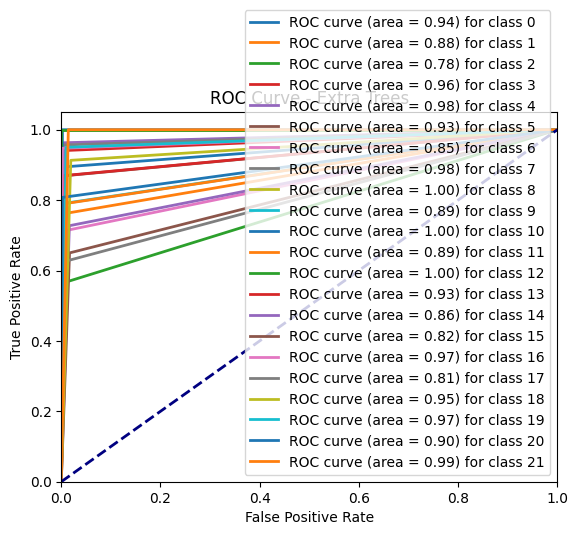


Fig. 33: Roc curve of Extra Trees

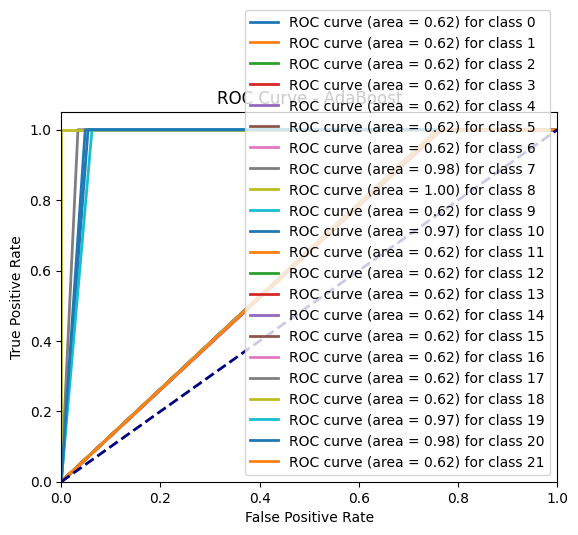


Fig. 34: Roc curve of AdaBoost

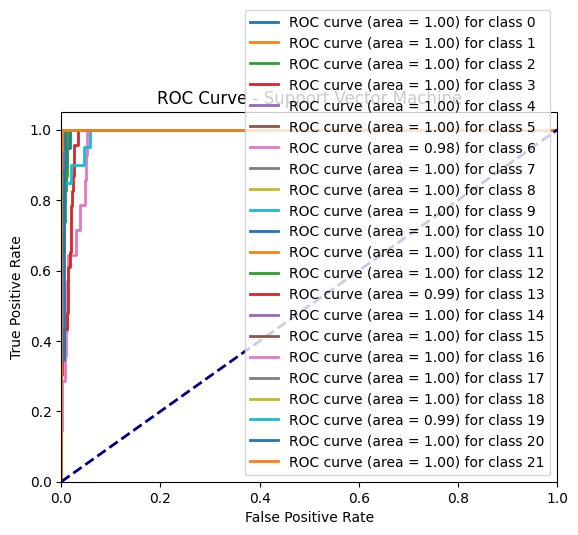


Fig. 35: Roc curve of SVM

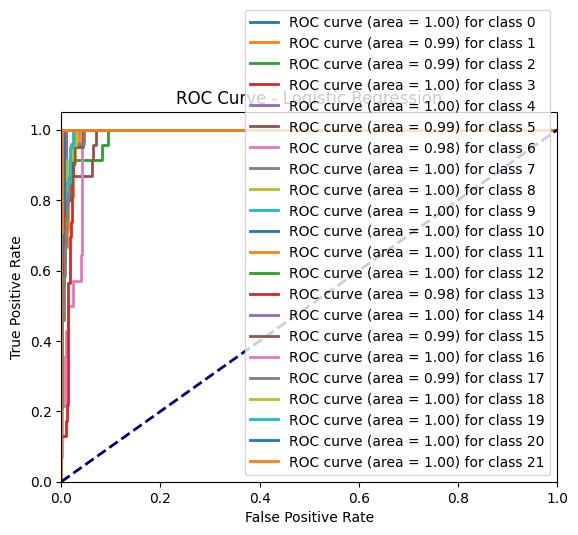


Fig. 36: Roc curve of Logistic Regression

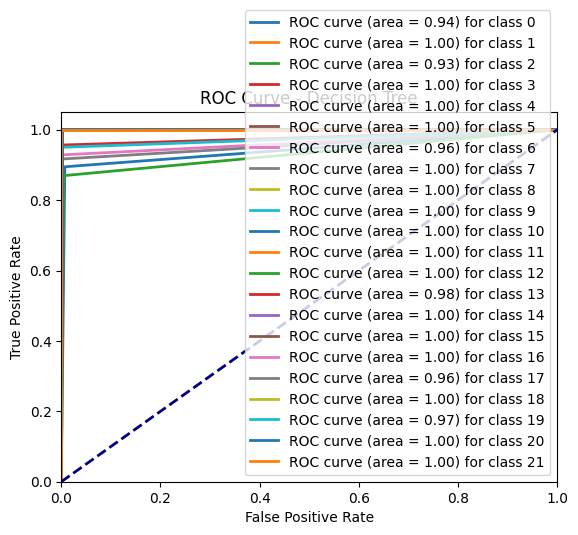


Fig. 37: Roc curve of Decision Tree

All things considered, bar graphs for data visualization are effective tools for presenting complicated information in an understandable and straightforward way, empowering stakeholders to decide on crop choices and optimize systems.

1. CONCLUSION AND FUTURE SCOPE

Using an ensemble model such as AdaBoost resulted in significantly higher accuracy and lower error rates, which is especially noteworthy considering the variety of soil types that sustain different crops. The use of "Pickle" files, a traditional Python serialization technique, for model storage allowed for easy deployment and real-time use. One benefit of this approach is that it allows for strategic crop planting optimization to maximize soil component utilization and reduce resource waste. Moreover, this architecture provides multiple paths for future development

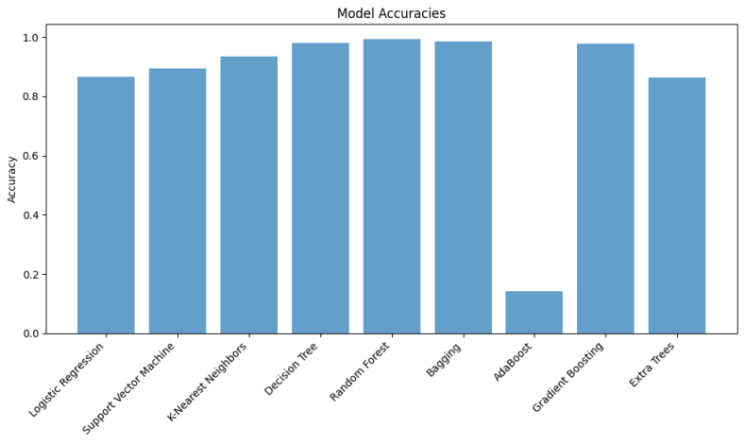


Fig. 38: Plot for accuracy comparison.

To improve precision, for example, adding soil types to the dataset and adding recommendations based on fertilizer kinds and irrigation schedules could help. Another possible area for future research is investigating the integration of insect detection mechanisms to prevent crop production damage.

1. REFERENCES
2. Kumar, Y. Jeevan Nagendra, V. Spandana,V. S.Vaishnavi, K. Neha, and V. G. R. R. Devi."Supervised Machine learning Approachfor Crop Yield Prediction in AgricultureSector." In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 736-741.IEEE, 2020.
3. Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal. "Crop yield prediction using machine learning algorithms." In 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 125-130. IEEE,2019.
4. Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta. "Crop yield prediction using machine learning techniques." In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1-5. IEEE, 2019.
5. Jain, Sonal, and Dharavath Ramesh. "Machine Learning convergence for weather based crop selection." In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6. IEEE, 2020.
6. Gandge, Yogesh. "A study on various data mining techniques for crop yield prediction." In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 420-423. IEEE, 2017.
7. Suresh, A., P. Ganesh Kumar, and M. Ramalatha. "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN." In 2018 3rd International Conference on Communication and Electronics Systems (ICCES), pp. 88-93. IEEE, 2018.
8. Kamatchi, S. Bangaru, and R. Parvathi. "Improvement of Crop Production Using Recommender System by Weather Forecasts." Procedia Computer Science 165 (2019): 724-732.
9. Bondre, Devdatta A., and Santosh Mahagaonkar. "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms." International Journal of Engineering Applied Sciences and Technology 4, no. 5 (2019): 371-376.
10. Suresh, G., A. Senthil Kumar, S. Lekashri, and R. Manikandan. "Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming." International Journal of Modern Agriculture 10, no. 1 (2021): 906-914.
11. Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. "Crop recommendation system to maximize crop yield in ramtek region using machine learning." International Journal of Scientific Research in Science and Technology 6, no. 1 (2019): 485-489.
12. Pudumalar, S., E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J.Nisha. "Crop recommendation system for precision agriculture." In 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32-36. IEEE, 2017.
13. Garanayak, Mamata, Goutam Sahu, SachiNandan Mohanty, and Alok Kumar Jagadev. "Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods."
14. Jejurkar Siddhi, S., S. Bhosale Meghna, and D. N. Wavhal. "Crop Predication and Diseases Detection Using Machine Learning." (2021).
15. Rajak, Rohit Kumar, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare. "Crop recommendation system to maximize crop yield using machine learning technique." International Research Journal of Engineering and Technology 4, no. 12 (2017): 950-953.
16. Doshi, Zeel, Subhash Nadkarni, Rashi Agrawal, and Neepa Shah. "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms." In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1-6. IEEE, 2018
17. Dataset link

<https://www.kaggle.com/datasets/atharvaingle/croprecommendation-dataset>

# 