RANDOM FOREST APPLICATION FOR CROP YIELD PREDICTION

Abbas Maazallahi[1], Sreehari Thota[1], Naga Prasad Kondaboina[1], Vineetha Muktineni[1], Deepthi Annem[1], Abhi Stephen Rokkam[1], Mohammad Hossein Amini[2], Mohammad Amir Salari[1], Payam Norouzzadeh[1], Eli Snir[2], Bahareh Rahmani[1]

[1] Saint Louis University, Computer Science, Saint Louis University, Saint Louis, USA
[2] Washington University in Saint Louis, Business School, Saint Louis, USA

ABSTRACT

This study analyzes crop yield prediction in India from 1997 to 2020, focusing on various crops and key environmental factors including crop typesand years, cropping seasons, specific details for each state, areas of cultivation, production quantities, annual rainfall, and the usage of fertilizers and pesticides. We applied advanced machine learning techniques like Logistic Regression, Decision Tree, KNN, Naïve Bayes, K-Mean Clustering, and Random Forest to predict agricultural yields. The main goal of this study is offering the best model to predict crop yields. Based on our study, Random Forest demonstratesalmost high accuracy. Naïve Bayes shows high precision indicating the high quality of a positive prediction made by this model. In this study, we are discovering the best machine learning models to predict the crop yield. If people know that their yield will be decreased next year, they find a way increase the crop yields.

1. INTRODUCTION

Machine learning has significantly influenced agricultural practices, particularly in crop yield prediction. In this study, various machine learning techniques have been employed to enhance the accuracy and efficiency of forecasts. These techniques offer valuable insights into the complex nature of agricultural data and the prediction of crop yields.

LogisticRegression explores show the probabilistic relationships between variables. The simplicity and interpretability of Logistic Regression make it a popular choice in many fields, including agriculture[1].The Decision Tree emerges as a robust method in classification and regression toolkit. It aids clear decision-making by splitting data into branches based on variable values[2].

K-Nearest Neighbors (KNN) as a non-parametric method, identifies the similarities between new and existing data points, making it suitable for classification and regression problems.Random Forest (RF) is a popular ensemble machine learning algorithm to combine the output of several decision trees to classify and predict the future outcomes.As the field continues to evolve, exploring and implementing these techniques remain critical in addressing the complexities of crop yield prediction.

2. LITERATURE REVIEW

A study featured in Nature's Scientific Reports presents an "Interaction Regression Model for Crop Yield Prediction." This research selects robust features and interactions to predict crop yields, utilizing an elastic net regularization model. This model is instrumental in identifying high-quality features across various environmental and management categories, due to reducing the risk of overfitting and increasing the robustness of predictions across different geographic locations and timeframes [2].

Ziegel's seminal work, "The Elements of Statistical Learning" (2003), remains a foundational text in the field of machine learning and statistical modeling. Published in Technimetrics, this book provides an extensive overview of various statistical learning techniques, including regression, classification, and ensemble methods. Ziegel’s work is particularly relevant to the domain of crop yield prediction, as it lays the theoretical groundwork for many of the advanced ML algorithms employed in agricultural research today. The comprehensive nature of this text makes it an essential reference for understanding the underlying principles and applications of statistical learning in diverse fields, including agriculture [3].

Another research paper, "Using Machine Learning for Crop Yield Prediction in the Past or the Future," published in Frontiers, offers a unique perspective by simulating sunflower and wheat yields over a twenty-year period from 2000 to 2020. This research emphasizes the significance of continuous nutrient and water balance in the simulation process and explores the impact of changes in cultivars and planting densities on crop yields. The detailed simulation models provide valuable insights into long-term yield prediction and resource management, marking a significant advancement in the field[4].

The study "Analysis of Crop Yield Prediction using Machine Learning Algorithms" in IEEE Xplore reports the uncertainties of weather and its impact on farming.The paper evaluates the efficacy of machine learning algorithms—K-Nearest Neighbors (KNN), Random Forest, and Linear Regression—using parameters like state, crop, temperature, and rainfall to predict crop yields. The results showcase a remarkable 97% accuracy for KNN, outshining the Random Forest's 75% and Linear Regression's 54%, highlighting the promise of KNN inpredictive agriculture and offering a data-driven example for enhancing agricultural productivity[5]. A Machine Learning Approach to Predict Crop Yield and Success Rate" from IEEE Xplore details an innovative study within India's agricultural sector. Focusing on improving farmers' decision-making by predicting crop yields, this research employs neural network regression modeling with an extensive dataset drawn from government sources. The researchers reported a 45% accuracy using RMSprop optimizer, which was substantially improved to 90% by refining the network architecture and shifting to the Adam optimizer. The model applies a 3-Layer Neural Network with the Rectified Linear Activation Unit (ReLU) function and leverages both backward and forward propagation techniques to establish a robust model for crop yield prediction[6].

Moreover, "Utilizing Naïve Bayes Algorithm for Crop Yield Prediction" explores the application of Naïve Bayes algorithm in predicting crop yields based on various agricultural parameterscontaining weather information, soil characteristics, and crop management practices. This study shows that Naïve Bayes in accurately predicting crop yields across different regions and crop varieties, highlighting its potential as a valuable tool for agricultural decision-making[7].

Another study "Enhancing Crop Yield Prediction through Random Forest Algorithm" investigates the use of Random Forest algorithm to improve crop yield prediction’s accuracy. By constructing an ensemble of decision trees and aggregating their predictions, Random Forest leverages the strength of multiple models to capture complex nonlinear relationships between predictor variables and crop yields. This research demonstrates the superior performance of Random Forest over traditional regression models, making it an asset for precision agriculture[8].

Zhang et al. (2023) investigate the use of the Random Forest algorithm to improve crop yield prediction accuracy. By constructing an ensemble of decision trees and aggregating their predictions, Random Forest captures complex nonlinear relationships between predictor variables and crop yields. This research illustrates the superior performance of Random Forest over traditional regression models, making it an asset for precision agriculture and informed decision-making in farming practices [9].

Dhaliwal and Williams (2024) provide an insightful exploration into the prediction of sweet corn yield using machine learning models and field-level data. Their study, published in Precision Agriculture, utilizes a combination of ML algorithms to enhance yield prediction accuracy. By integrating extensive field-level data, including soil properties, weather conditions, and crop management practices, the researchers demonstrate a robust framework for predicting sweet corn yields. Their findings emphasize the importance of high-resolution field data in improving the predictive performance of ML models in agriculture [10].

Rashid et al. (2021) offer a comprehensive review of crop yield prediction using machine learning approaches, with a particular emphasis on palm oil yield prediction. Published in IEEE Access, this review synthesizes existing research and methodologies, providing a detailed analysis of various ML techniques applied to crop yield prediction. The authors discuss the challenges and advantages of different ML models, highlighting how advanced algorithms like deep learning and ensemble methods have been successfully employed to predict yields in complex agricultural systems. This review serves as a valuable resource for researchers and practitioners aiming to leverage ML for enhanced agricultural productivity [11].

Hussain, Sarfraz, and Javed (2021) conducted a systematic review on crop-yield prediction through Unmanned Aerial Vehicles (UAVs) presented at the 16th International Conference on Emerging Technologies (ICET 2021). The study highlights the prevalent use of Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Networks (CNN) in crop yield prediction. The review underscores the significance of these algorithms and their adoption in developing countries, reflecting the growing reliance on UAVs for data collection and analysis in agriculture [12].

Saraiya, Chaudhari, and Verma (2022) discussed the challenges of crop yield prediction and crop selection based on climatic sensor data and historical yield data in their book section, "Monitoring Agricultural Essentials," from the "Application of Machine Learning in Agriculture." The authors emphasize the importance of machine learning in addressing major agricultural problems and improving crop yield predictions by leveraging climatic and past data [13].

Van Wart et al. (2015) explored the creation of long-term weather data for crop simulation modeling in their article published in Agricultural and Forest Meteorology. This study highlights the necessity of high-quality daily weather data, such as uncorrected gridded solar radiation, for accurate crop yield simulation and variability prediction. The authors demonstrate how propagating long-term weather data significantly enhances the reliability of crop simulation models [14].

Mahmood (1998) conducted a comparative study on air temperature variations and rice productivity in Bangladesh, published in Ecological Modelling. The study compares the performance of the YIELD and CERES-rice models, finding that rice productivity predictions at Mymensingh are higher using the YIELD model. This research underscores the critical role of accurate temperature data in predicting crop productivity [15].

Venkatesh and Saravanan (2022) investigated the prediction of crop yield using Simple Linear Regression (SLR) and Polynomial Regression (PR) in their study presented at the 3rd International Conference on Smart Electronics and Communication (ICOSEC 2022). Their findings suggest that SLR significantly outperforms PR in predicting crop yields, indicating the effectiveness of simpler models for specific types of agricultural data [16]. In similar works we applied machine learning methods to predict weather patterns [17], and customer churn [18].

These studies underscore the dynamic and evolving nature of crop yield prediction research. They not only highlight the potential of machine learning in agriculture but also set a foundation for future studies. Our research aims to build upon these methodologies, introducing novel approaches to further enhance the precision and applicability of crop yield predictions.

3. DATA DESCRIPTION

The dataset used in this study, available at Kaggle (https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian-states-dataset), includes extensive agricultural data from India from 1997 to 2020. It covers a wide range of crops grown across different Indian states. Data includes crop types and years, cropping seasons, specific details for each state, areas of cultivation, production quantities, annual rainfall, and the usage of fertilizers and pesticides. The data features are:

Crop: This field identifies the crop type. The dataset includes a diverse array of 55 crops, reflecting India's 55 rich agricultural variety including rice, maize, onion, potato, coconut, and banana.

Crop Year: The dataset covers crop years from 1997 to 2020, providing a comprehensive temporal view of agricultural trends over 24 years.
Season: The data categorizes cultivation of 4 distinct seasons, including major seasons Autumn and Spring to analyzethe seasonal impacts on agriculture.

State: Includes data from 30 Indian states which offers a wide geographical perspective, to find the regional agricultural patterns.

Area: Represents the land area under cultivation in hectares.The mean of area is approximately 179,926 hectares, ranging from a minimal 0.5 hectares to a vast 50.8 million hectares toindicate the varied scale of farming practices across regions.

Production: The quantity of crop production, measured in metric tons, shows an average of around 16.4 million tons. It varies greatly, with a maximum recorded production of about 6.3 billion tons.

Annual Rainfall: This feature, measured in millimeters, indicates the climatic conditions affecting crop growth. The average annual rainfall is about 1,438 mm, ranging from 301.3 mm to a significant 6,552.7 mm.

Fertilizer: The total amount of fertilizer used, in kilograms, with an average of around 24.1 million kg. It shows a diverse nutrient management strategy across different crops and regions.

Pesticide: This field presents the total pesticide usage in kilograms. On average, around 48,848 kg of pesticides are used, with the maximum 15.75 million kg.

Yield: This attribute indicates production per unit area with an average of approximately 79.95 and an extremely varied range, peaking at 21,105. This metric is evaluating the efficiency of agricultural practices.

4. STATISTICAL ANALYSIS

These statistical insights provide a more understanding of the dataset, highlighting the complexity and diversity of agricultural practices in India.

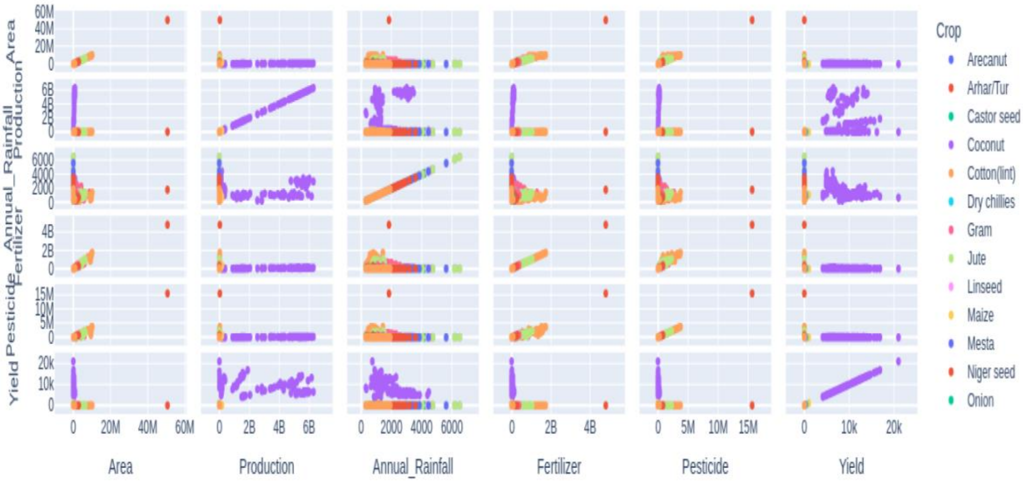

Figure 1.Scatter Plot of Key Features with Crop Categories

The scatter plot shows the relationships between various key features of the agricultural dataset differentiated by color. Each subplot in the matrix compares two different features,area vs production, annual rainfall vs fertilizer, and so on (Figure 1).From the scatter plot, we can observe the following:

Area vs Production: There is a positive correlation between the area of cultivation and the production formost crops, which is expected as larger cultivation areas generally lead to higher production volumes.

Annual Rainfall vs Production: The relationship between annual rainfall and production varies among crops, suggesting that some crops may be more sensitive to rainfall than others.

Fertilizer vs Production: There seems to be a positive correlation for some crops, indicating that increased fertilizer usage may cause higher production. However, this relationship does not hold uniformly across all crop types.

Pesticide vs Production: Pesticide usage does not show a clear correlation with production in this visualization which means the effectiveness or necessity of pesticides may vary greatly depending on the crop.

Yield: The yield scatter plots across different features show varied patterns for different crops, indicating that yield is influenced by a complex interplay of factors, not just a single feature.

Each crop type, represented by a unique color, exhibits its own pattern of distribution and correlation across the different features, which can inform targeted agricultural practices and policies. The data points for crops like coconut are notably distinct due to high-volume output, which skews the distribution.

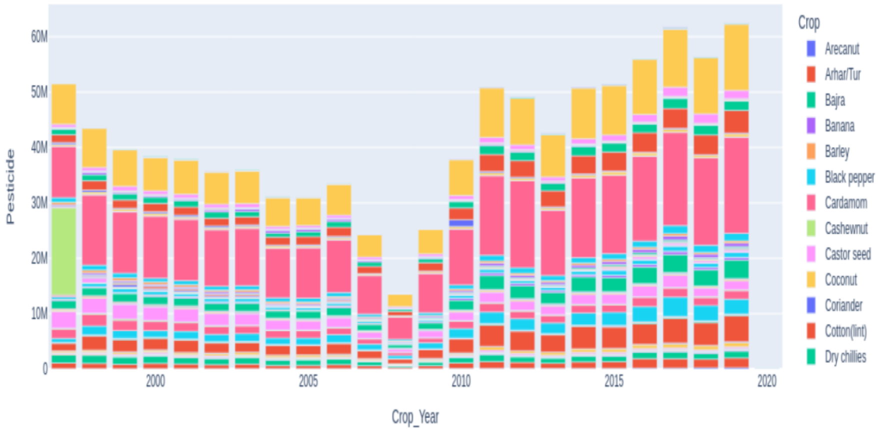

Figure 2. Aggregated Pesticide Usage by Crop and Year

The bar chart would offer a comprehensive view of the trends in pesticide use across different crops over the years (Figure 2). It shows how pesticide usage has varied over time for each crop type. This chart is an instrument to identify patterns and potential correlations between pesticide use and other factors like crop yield, cultivation practices, or environmental changes. It would serve as a critical tool for understanding the dynamics of pesticide management in agriculture, aidingin developing more sustainable and efficient farming practices.

4.1. Features Distribution

The feature distribution plots for area, production, annual rainfall, fertilizer, and pesticide from the agricultural dataset provide a visual summary of the underlying data characteristics and variability (Figure 3).The histograms reveal the frequency distribution of values for each feature. The Area histogram shows a concentration of values in smaller land areas, suggesting that most of the crop cultivation occurs in relatively smaller lands. The Production histogram is rightly skewed with a few instances of very high yields, indicative of a small number of highly productive operations. Annual Rainfall appears more consistently distributed, suggesting a level of predictability in this environmental factor. Fertilizer and Pesticide usage are both right skewed, indicating that lower usage rates are more common across the dataset.

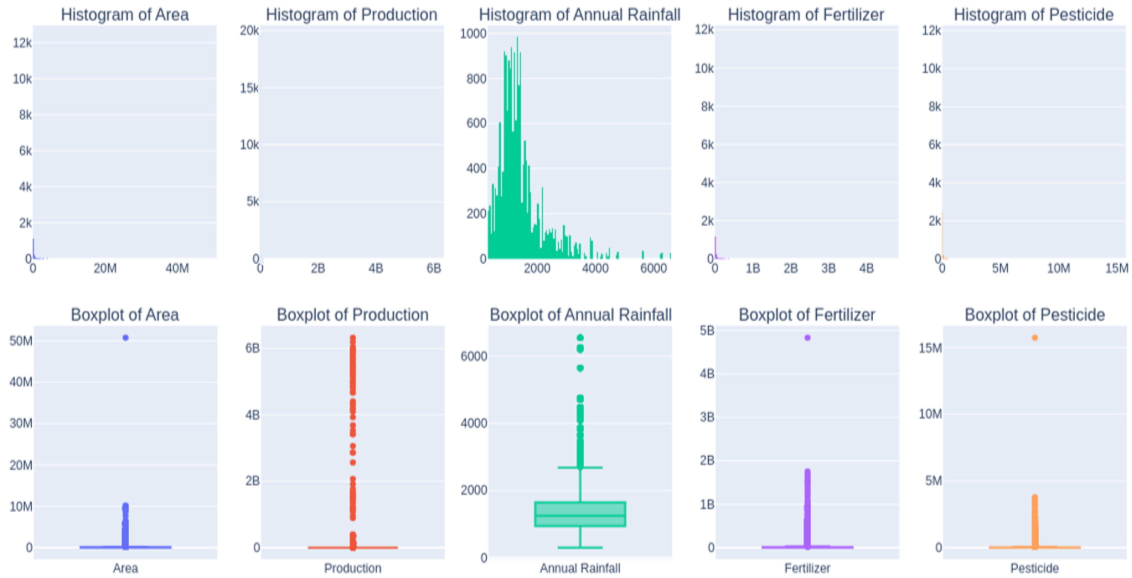

Figure 3. Histogram, Density and Box plot of selected features

Density plots provide a smoothed representation of the data distribution, revealing the probability density of the different values. These plots show the likelihood of specific values occurring within the dataset and highlight the central tendencies and the spread of data more clearly than histograms. For area, production, fertilizer, and pesticide, the peaks of the density plots suggest the most common values and verify the skewness seen in the histograms.

Box plots offer a summary of the data's statistical distribution, including the median, quartiles, and outliers. The box plots for area and annual fertilizer and pesticide do not show significant outliers, which indicates a more homogeneous distribution within the interquartile range whereproduction, fertilizer, and pesticide box plots display several upper-end outliers. These outliers represent values that are exceptionally higher than the typical range of the data and may correspond to instances of intensive farming practices or atypical environmental conditions.

5. Normalization and Labeling of Crop Yield Data (Target Variable)

Normalization per crop is aessential preprocessing step in agricultural data analysis. This process involves scaling the yield data for each crop type within a specified range (commonly 0 to 1) to ensure a uniform scale across various crops. The primary reasons for this normalization include:

Comparability: Different crops may have inherently different yield scales due to varying biological and cultivation factors. Normalization allows for a fair comparison of yields across diverse crop types on a common scale.

Outlier Mitigation: Some crops might have extreme yield values (either high or low) that can skew the overall analysis. Normalization helps in mitigating the impact of such outliers.

Uniformity in Analysis: It ensures that the yield data across all crops are treated uniformly, making the subsequent analysis more robust and less biased towards crops with larger or smaller yield values.

5.1. Purpose of Labeling into Four Classes

Labeling the normalized yield data into four distinct classes is a method of discretization that simplifies complex continuous data into categorical segments. This is beneficial for several reasons:

Simplification of Data: It simplifies the continuous range of yield values into distinct categories, making it easier to analyze and understand patterns within the data.

Facilitates Classification Analysis: By converting yields into classes, the data is prepared for classification algorithms in machine learning, predictive modeling or trend analysis.

Enhanced Interpretability: Labeling yields into categories like 'Low', 'Medium', 'High', and 'Very High' provides a more intuitive understanding of the yield performance for each crop.

Labeling is often done using quartiles, dividing the data into four equal parts based on their distribution. This method ensures that each class has an equal number of data points, providing a balanced categorization of the yield data.

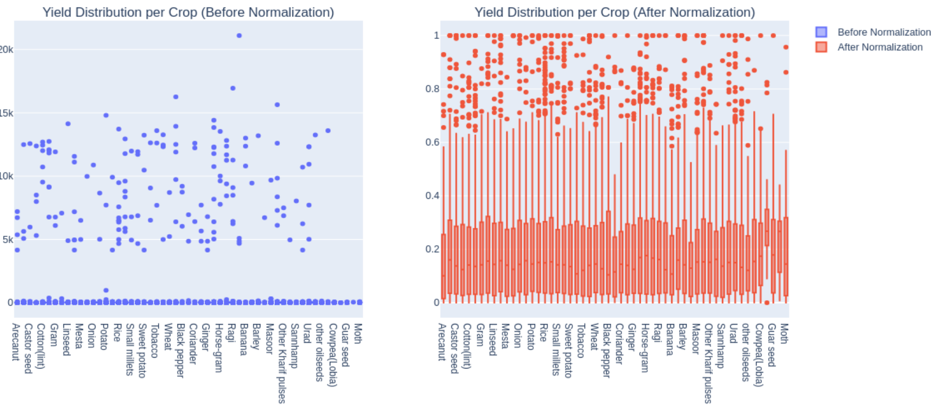

Figure 4. Yield Distribution per Crop, Before and After Normalization

Figure 4 illustrates the impact of normalization on yield data across various crops. On the left, we see the yield distribution before normalization, where each crop's yield values span a wide and disparate range, making it difficult to compare between crops. Outliers and variances are prominent, and the scales are imbalanced, with some crops showing yields reaching 20,000 units. On the right, after normalization, all yields are scaled between 0 and 1. This transformation standardizes the data, bringing all crops onto an even playing field and highlighting the relative distribution within each crop type without being overshadowed by the absolute yield values. This normalized view allows for more straightforward comparisons across different crops and a clearer interpretation of yield performance relative to each crop's potential.

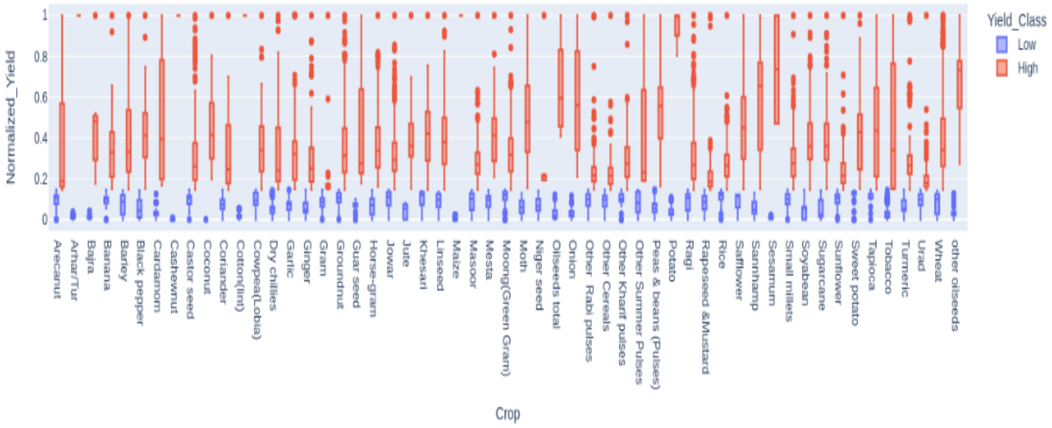

Figure 5. Normalized Yield Distribution per Crop

Figure 5 presents a boxplot illustrating the normalized yield distribution for a variety of crops, with yield data segmented into two distinct classes: Low and High. Each crop type is represented by a series of boxplots along the horizontal axis, with the normalized yield values plotted on the vertical axis ranging from 0 to 1. The Low-yield class is depicted in blue, and the High-yield class in red, allowing for a clear visual distinction between the two categories. For each class, the box plots show the median yield value (the line within the box), the interquartile range (the box itself), and potential outliers (the individual points beyond the whiskers). This graph effectively communicates the variability in yield within each crop type, as well as between the two yield classes, providing insights into the distribution patterns of agricultural productivity across different crops.

6. METHODOLOGY

This section evaluates and compares the performance of various machine learning classifiers on a crop yield. The dataset, preprocessed with feature normalization, includes key agricultural indicators such as area, production, annual rainfall, fertilizer, and pesticide usage, all normalized to ensure uniformity and comparability across different scales. The target variable, 'Yield_Class_Int', represents yield categories encoded as integers, facilitating a multi-class classification approach.

The selected classifiers include a diverse array of algorithms: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Gradient Boosting. These methods cover a spectrum from simple linear models to more complex ensemble methods, each with its strengths in handling different types of data distributions and relationships. The dataset is split into training and testing sets, with 80% of the data used for training and the remaining 20% for testing to ensure a robust evaluation framework. Figure 6 shows one tree of Random Forest, K-Nearest Neighbor (K=3), and Naïve Bayes Classifiers.

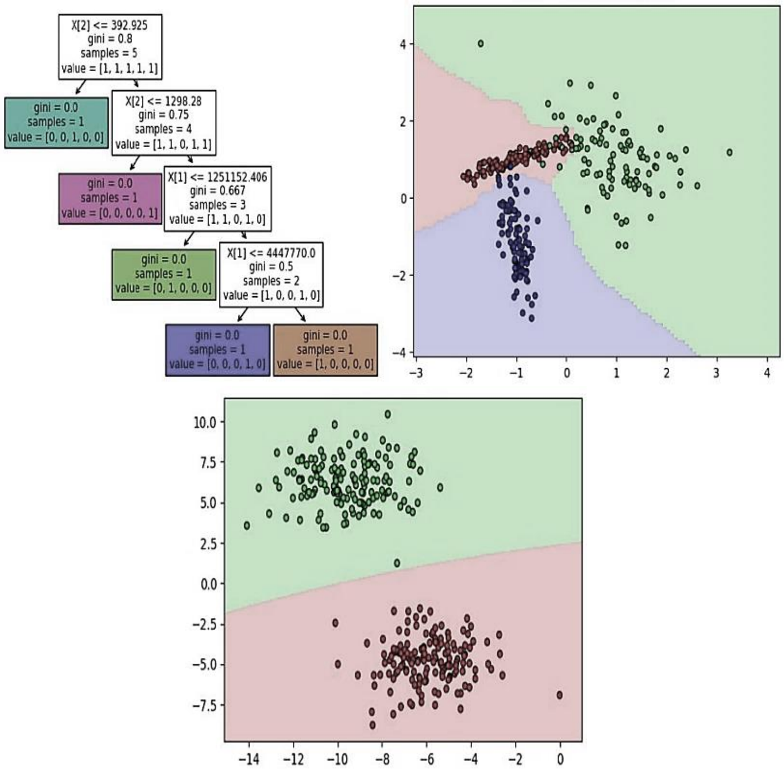

Figure 6. left to right: One tree of Random Forest, K-Nearest Neighbor (K=3), Naïve Bayes Classifiers

Each model is trained on the training set and then evaluated on the test set. Performance metrics such as accuracy, precision, recall, f1-score, and the confusion matrix are computed for each model. These metrics provide a multi-dimensional view of the models' performance, with accuracy indicating the overall correctness, precision and recall offering insights into the models' ability to identify each class correctly, and the f1-score presenting a balance between precision and recall. The confusion matrix further elucidates the specific areas of strength and weakness for each classifier, by showing the distribution of predictions across actual classes. This rigorous assessment allows for a detailed comparison of the models, highlighting their efficacy and fitness for the crop yield classification task.

7. RESULTS

Figure 7 compares various machine learning models used for classification tasks. The metrics includes Accuracy, Precision, Recall, and F1 Score.

Accuracy reflects the overall rate of correctly predicted the class labels. Precision indicates the proportion of true positives among all positive predictions. Precision is a key measure when the cost of a false positive is high. Recall measures the proportion of actual positives that were identified correctly, which is particularly important when missing a positive is costly. The F1 Score is the harmonic means of precision and recall, providing a single metric that balances both the false positives and false negatives.

model with high precision but lower recall might be conservative in its positive predictions but miss out on several actual positives. In contrast, a model with high recall but lower precision might capture most of the positives but at the cost of increased false positives. The F1 Score helps to balance these aspects and is often a crucial metric when choosing the best deployment model.

In confusion matrix shown in figure 8, the horizontal axis represents the predicted classifications, while the vertical axis represents the actual classifications, each divided into 'Positive' and 'Negative' categories. The top left quadrant represents true positives (TP), where the model correctly predicts the positive classes. The bottom right quadrant represents true negatives (TN), where the model correctly predicts the negative class. The top right quadrant shows false negatives (FN). In these instances, the model incorrectly predicts the negative class, and the bottom left quadrant shows false positives (FP), where the model incorrectly predicts the positive class.

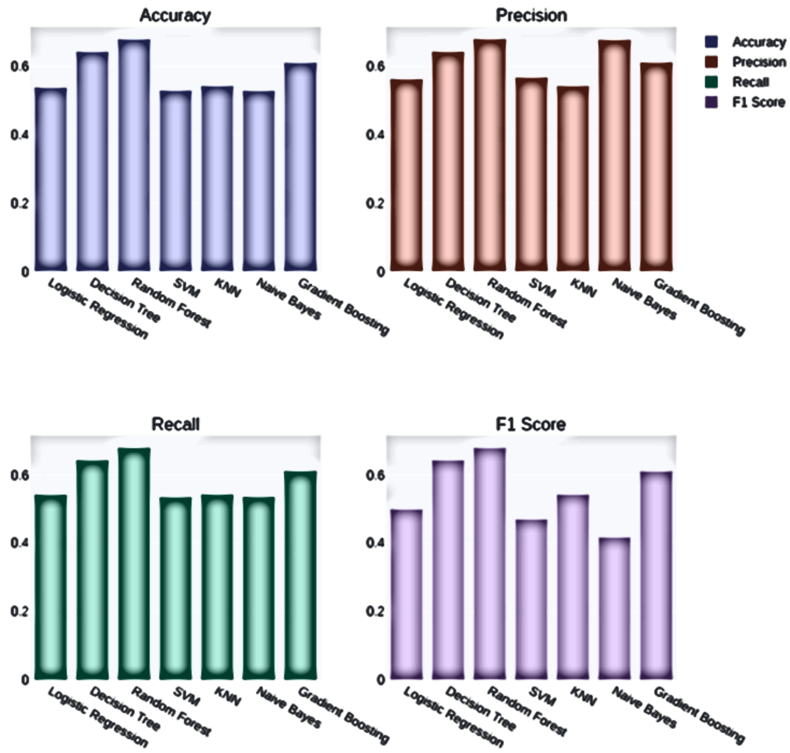

Figure 7.Performance Metrics of Different Methods

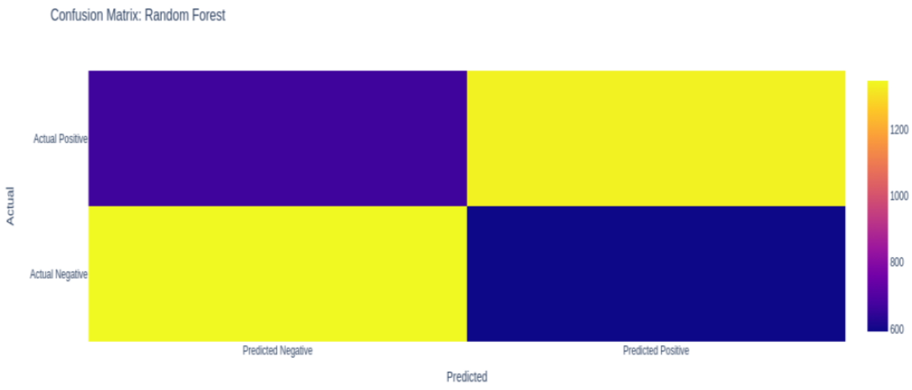

Figure 8. Confusion Matrix

The intensity of the colors corresponds to the number of observations in each category, with darker colors typically representing higher numbers. This visualization helps in quickly assessing the model's performance, particularly in terms of its ability to distinguish between the classes. For example, if the TP and TN quadrants are much darker than the FN and FP quadrants, this indicates a high level of accuracy like what we have in figure 8.

Our explorations into machine learning models for agricultural yield prediction have yielded significant insights. Based on Figure 7 and Table 1 the Random Forest model, tailored to our specific dataset, has demonstrated almost high accuracy, reaching a 73% success rate in predicting yield when considering crucial features such as area and production. This high level of precision underscores the model's capability to handle the discrete nature of our data effectively.

Table 1: Performance Metrics of Different Models Based on Figure 7

	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.52	0.58	0.55	0.5
Decision Tree	0.62	0.62	0.62	0.62
Random Forest	**0.73**	**0.73**	**0.73**	**0.73**
SVM	0.51	0.58	0.51	0.48
KNN	0.52	0.55	0.52	0.52
Naive Bayes	0.51	**0.64**	0.51	0.39
Gradient Boosting	0.61	0.6	0.61	0.6

8. Discussion

Our project's findings indicate that Random Forest models surpass in accuracy for discrete data sets, a characteristic that is particularly relevant to our agricultural domain. With its ensemble approach, the Random Forest model has complemented the probabilistic predictions of Naïve Bayes, which is offering a robust alternative for yield classification.

Throughout this project, we have not only applied various machine learning techniques but also honed our ability to discern the most appropriate methods for our dataset. The process has enhanced our analytical skills, enabling us to create informative visualizations that succinctly convey the efficacy of different machine learning strategies.In the future work we will apply boosting methods to increase the accuracy of the predictor.

Compliance with Ethical Standards

Conflict Interest Statement

There is no conflict of interest declared by authors. All authors have reviewed and agreed with the manuscript. We state that the submission is an original paper and is not under review at any other journal.

Research's human participants and/or animals

There are no humans or animals participating in this project.

Consent to Participate

Authors consent to participate in this project, and we know that: the research may not have direct benefit to us. Our participation is entirely volunteer. There is a right to withdraw from the project at any time without any consequences.

Data Availability

The dataset used for this project is collected from Kaggle. https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian-states-dataset

Funding

No Funding has been applied for this project.

Ethical Approval

All subjects gave their informed consent for inclusion before they participated in the study.

Consent to Publish

We give our consent for the publication of exclusive details, that could be included figures and tables and details within the manuscript to be published in Computational Brain & Behavior.

REFERENCES

[1] V. A. Barbur, D. C. Montgomery, and E. A. Peck, "Introduction to Linear Regression Analysis.," The Statistician, vol. 43, no. 2, 1994, doi: 10.2307/2348362.

[2] J. Ansarifar, L. Wang, and S. V. Archontoulis, "An interaction regression model for crop yield prediction," Sci Rep, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-97221-7.

[3] E. R. Ziegel, "The Elements of Statistical Learning," Technimetrics, vol. 45, no. 3, 2003, doi: 10.1198/tech.2003.s770.

[4] A. Morales and F. J. Villalobos, "Using machine learning for crop yield prediction in the past or the future," Front Plant Sci, vol. 14, 2023, doi: 10.3389/fpls.2023.1128388.

[5] V. Krishna, T. Reddy, S. Harsha, K. Ramar, S. Hariharan, and A. Bhanuprasad, "Analysis of Crop Yield Prediction using Machine Learning algorithms," in Proceedings - 2022 2nd International Conference on Innovative Sustainable Computational Technologies, CISCT 2022, 2022. doi: 10.1109/CISCT55310.2022.10046581.

[6] S. S. Kale and P. S. Patil, "A Machine Learning Approach to Predict Crop Yield and Success Rate," in 2019 IEEE Pune Section International Conference, PuneCon 2019, 2019. doi: 10.1109/PuneCon46936.2019.9105741.

[7] M. Gupta, B. V. Santhosh Krishna, B. Kavyashree, H. R. Narapureddy, N. Surapaneni, and K. Varma, "Various Crop Yield Prediction Techniques Using Machine Learning Algorithms," in Proceedings of the 2nd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2022, 2022. doi: 10.1109/ICAIS53314.2022.9742903.

[8] M. Manafifard, "A new hyperparameter to random forest: application of remote sensing in yield prediction," Earth Sci Inform, vol. 17, no. 1, 2024, doi: 10.1007/s12145-023-01156-8.

[9] Q. Zhang et al., "Maize yield prediction using federated random forest," Compute Electron Agric, vol. 210, 2023, doi: 10.1016/j.compag.2023.107930.

[10] D. S. Dhaliwal and M. M. Williams, "Sweet corn yield prediction using machine learning models and field-level data," Precis Agric, vol. 25, no. 1, 2024, doi: 10.1007/s11119-023-10057-1.

[11] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, “A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction,” 2021. doi: 10.1109/ACCESS.2021.3075159.

[12] N. Hussain, S. Sarfraz, and S. Javed, “A Systematic Review on Crop-Yield Prediction through Unmanned Aerial Vehicles,” in ICET 2021 - 16th International Conference on Emerging Technologies 2021, Proceedings, 2021. doi: 10.1109/ICET54505.2021.9689838.

[13] J. P. Sarvaiya, A. P. Chaudhari, and J. P. Verma, “Monitoring agricultural essentials,” in Application of Machine Learning in Agriculture, 2022. doi: 10.1016/B978-0-323-90550-3.00004-7.

[14] J. Van Wart, P. Grassini, H. Yang, L. Claessens, A. Jarvis, and K. G. Cassman, “Creating long-term weather data from thin air for crop simulation modeling,” Agric forMeteorology, vol. 209–210, 2015, doi: 10.1016/j.agrformet.2015.02.020.

[15] R. Mahmood, “Air temperature variations and rice productivity in Bangladesh: A comparative study of the performance of the YEILD and the CERES-rice models,” Ecole Modell, vol. 106, no. 2–3, 1998, doi: 10.1016/S0304-3800(97)00192-0.

[16] N. G. E. Hackland and R. I. Jones, “Voo spelling van seasonableproduce van ergotiscurvelet,” Proceedings of the Annual Congresses of the Grassland Society of Southern Africa, vol. 15, no. 1, 1980, doi: 10.1080/00725560.1980.9648891.

[17] K Miller, G Yi, E Snir, B Rahmani, Precipitation analysis and forecasting weather of Texas, United States, International Journal of Information Technology, Springer Nature, 15 (2), 549-556.

[18] Y. M. Meda, B. M. S. Bokka, H. Jamallamudi, P. Norouzzadeh, E. Snir, B. Rahmani, Customer Churn Prediction with Machine Learning, A. Maazallahi, the proceedings of International Business Analytics Conference (IBAC).