*Grenze International Journal of Engineering and Technology, June Issue*

# Harvesting Innovation: Revolutionizing Crop Yields using Machine Learning

Dolly Kashyap[1], Piyush Gangwar[2], Mudit Goel[3] and Kanchan Dixit[4]
[1-4]ABES Engineering College , Ghaziabad, India
Email: dollykashap2003@gmail.com, gangwarpiyush11@gmail.com, muditgoel248@gmail.com, Kanchan.dixit@abes.ac.in

***Abstract*— India's agricultural sector has been grappling with the adverse effects of climate change over the last two decades, resulting in the diminished performance of various crops. Predicting crop yields well in advance of harvest holds immense significance, aiding policymakers and farmers in making informed decisions regarding marketing and storage. This project endeavors to provide farmers with a tool that offers predictive insights into crop yields before they embark on the cultivation process. The proposed solution involves thedevelopment of a credit card firm's need to be on the lookout for a fraudulent prototype for an interactive prediction system, featuring a user-friendly web-based graphical interface and employing machine learning algorithms. The prediction results are readily accessible to farmers, enabling them to plan and strategize effectively. In the face of challenges posed by climatic variables such as weather, temperature, humidity, rainfall, and moisture, this project harnesses various data analytics techniques and algorithms to enhance crop yield predictions. The Random Forest algorithm, a robust and widely-used supervised machine learning method capable of both classification and regression tasks, is leveraged for this purpose**.

***Index Terms*— Agriculture, Machine Learning, crop-prediction,Supervised Algorithms, Crop yield, Data Mining, Decision Tree.**

## I. INTRODUCTION

Crop prediction analysis through machine learning is a cutting-edge research area poised to revolutionize agricultural practices. This field involves the development and application of advanced algorithms to forecast crop yields with unprecedented accuracy. By integrating historical data, real-time environmental factors, and cultivation practices, machine learning models provide invaluable insights for farmers and stakeholders. This research paper explores the significance of machine learning in predicting crop outcomes, addressing the complexities of agricultural systems, and contributing to the sustainability and efficiency of modern farming.

*1. Rising Importance of Rising Agriculture:* Precision agriculture is gaining prominence as a pivotal approach in addressing contemporary challenges in realm of global food production. This research paper underscores the increasing importance of precision agriculture as a transformative strategy. It explores how precision agriculture, leveraging advanced technologies and data analytics, enhances the efficiency and sustainability of farming practices. The rising significance of precision agriculture is examined in the context of optimizing resource utilization, minimizing environmental impact, and fostering a more resilient and responsive agricultural system. Through this lens, the researchaims to contribute valuable insights into the evolving landscape of precision

*Grenze ID: 01.GIJET.10.2.234*

agriculture and its implications for the future of food production.
*2. Data-Driven Agriculture:* The research paper delves into the paradigm shift toward data-driven agriculture, emphasizing the growing reliance on data-driven decision-making in modern farming practices. This exploration highlights how agricultural systems are evolving to harness vast datasets, advanced sensors, and analytical tools. The paper aims to elucidate the transformative impact of data-driven approaches on optimizing crop management, resource allocation and overall farm productivity. By examining the integration of data analytics in agriculture.
*3. Addressing Modern Agricultural Challenges*: This research paper addresses the multifaceted challenges faced by modern agriculture and explores how innovative solutions, particularly machine learning, are being employed to navigate these complexities. By examining issues such as climate change, resource scarcity, and increasing demand, the paper aims to shed light on how technology can contribute to adaptive and resilient farming practices. The research explores the potential of machine learning in mitigating risks, optimizing resource utilization, and enhancing overall productivity in the face of contemporary agricultural challenges. The objective is to provide valuable insights into the transformative role of technology in shaping the future of agriculture.
*4. The Innovation Imperative*: This research paper underscores the innovation imperative in crop prediction analysis, emphasizing the critical need for advanced methodologies, technologies, and approaches to meet the evolving challenges in agriculture. By examining the current state of crop prediction practices and the limitations of traditional methods, the paper aims to highlight the urgency of embracing innovative solutions. It explores how cutting-edge technologies, particularly machine learning, can revolutionize crop prediction, offering more accurate, timely, and actionable insights for farmers.

### *A. Motivation*

*Optimal Resource Utilization*: To allocate resources like water, fertilizer, and pesticides as efficiently as possible, machine learning algorithms employ massive volumes of data pertaining to weather patterns, crop health, and soil quality analysis. Efficiency is raised while expenses and environmental effect are decreased as a result.
*Climate Change Resilience*: Unpredictable weather patterns, the development of pests and diseases, and other issues are only a few of the major difficulties that climate change presents to agriculture. Farmers may better adjust to these changes by using machine learning, which offers insights about how to modify crop varieties, planting schedules, and management techniques to reduce risks and preserve yield.

### *B. Scope*

The scope of this research paper encompasses an in-depth exploration of the application of machine learning techniques in enhancing crop yield across various agricultural contexts. It involves investigating the utilization of machine learning algorithms for analyzing agricultural data, predicting crop performance, optimizing resource management, and facilitating precision farming practices. The research will delve into the development and evaluation of machine learning models tailored to specific crops, environmental conditions, and farming practices. Additionally, the paper will examine the potential challenges, limitations, and ethical considerations associated with implementing machine learning solutions in agriculture. Overall, the scope aims to provide comprehensive insights into how machine learning can revolutionize crop yield enhancement strategies.

### *C. Use (Background)*

Crop detection projects integrate, machine learning, to identify crop correctly. Utilizing tools like preprocess dataset, and extract features for classification. Algorithms such as Decision Tree, and K-Nearest Neighbour (KNN) and ensemble methods distinguish between healthy to identify suitable based on factor. IoT devices collect real-time data for risk assessment, while mobile apps enable farmers to capture images and receive instant diagnoses. GIS aids spatial insights, and cloud computing facilitates scalable model training and hosting. This amalgamation of techniques and tools enhances agricultural practices by providing timely disease detection and insights to ensure healthier crops.

### *D. Problem Statement*

In order to maximize crop selection and yield estimation based on meteorological patterns and monsoon cycles, agricultural operations in the past mostly depended on traditional knowledge and prediction methodologies. Nevertheless, as agriculture has become more mechanized and automated, there has been a noticeable decrease in the dissemination of this important knowledge over time.
Although automated systems are efficient and convenient, they frequently lack the sophisticated knowledge of crop management techniques and environmental elements that characterized conventional agricultural approaches.

Consequently, farmers can encounter difficulties in adjusting to shifting weather patterns and maximizing agricultural output.

*E. Methodology*

The methodology for this research papers on crop prediction systems encompass various machine learning methods and their applications in agriculture. One study suggests a system for predicting the best crop for a specific plot of land based on soil composition and weather conditions. Another research paper aims to discover the best model for crop prediction, which can help farmers decide the type of crop to grow based on climatic conditions and soil nutrients. This study compares popular algorithms such as K-Nearest Neighbour (KNN), Decision Tree, and Random Forest Classifier, with results revealing that Random Forest gives the highest accuracy among the three Predict.

Here's how we enhance crop yield step by step:

*1. Data Collection*: We are using the "CROP-Dataset" for our project, which is a collection of crop and agricultural-related metrics and data. The data collection is then downloaded into a Jupyter Notebook.

*2. Cleansing the data*: Finding missing, erroneous, or superfluous portions of data is one step in the process. The dirty data is then replaced, eliminated, or altered.

*3. Feature Extraction* In machine learning-based crop yield prediction, feature extraction entails choosing and modifying pertinent input variables (features) from raw data to reflect significant attributes that are helpful in crop yield prediction. A variety of agricultural characteristics, such as soil characteristics (pH, nutrient levels), meteorological factors (temperature, precipitation), crop management techniques (fertilization, irrigation), and past production information, might be included in these attributes.

*4. Data Preprocessing*: Following the acquisition of datasets from different sources. Preprocessing the dataset is necessary before training the model. Preprocessing the data can be carried out by phases, starting with the examination of the gathered dataset. Data cleansing is the next step in the process. When cleaning data, the Certain attributes in datasets are duplicates; these attributes are not taken into account when predicting crops. Thus, we must stop undesirable characteristics and datasets with some missing values that we must either replace or eliminate these missing variables undesirable nan values to improve accuracy.

*5. Model Training*: When a machine learning model is trained, its internal parameters are changed to help the model identify patterns or correlations in the data. The model gains knowledge from labeled data during training, and it modifies its parameters to reduce the discrepancy between the target values and its predictions. This is usually carried out iteratively with an optimization algorithm to get the optimal parameter values that yield precise forecasts. The model can forecast outcomes on fresh, unobserved data once it has been trained.

*6. The Real Test*: Once our computer has learned from the training data, we give it new dataset to evaluate. It checks these profiles and decides if they're real or fake, based on what it learned during training. This is how we identify fake profiles.

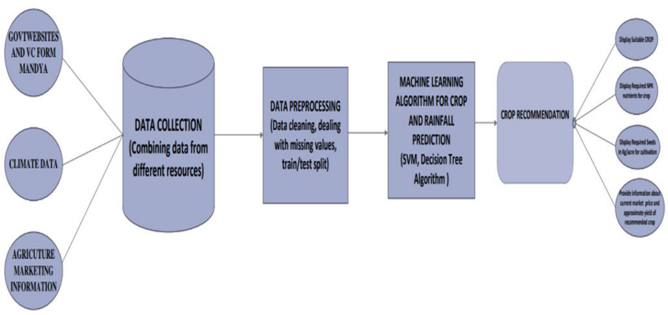



Figure 1. - Flow Diagram to Predict Crop Yield

## II. ALGORITHMS

*A. DECISION TREE*: The decision tree is a technique for choosing the best root nodes; once components of the same class are obtained, the tree is divided further according to attributes. It is a kind of supervised learning technique mostly utilized for classification issues, with adaptable features that aid materialization of both continuous and categorical dependent variables. This method generates groups that are as unlike from one another as is practical by splitting the population into two or more homogenous set.

We can select the best crop for the population by using the decision tree approach, which will give us the best split based on a variety of criteria. The decision tree classifier is suitable for crop prediction because of its feature selection procedure.

*B. Random Forest:* Random Forest is a machine learning technique that uses multiple decision trees to make better predictions. It works by creating multiple trees using different parts of the data and then combining their results. It can also tell you which factors are most important. Each tree in a random forest is created using a subset of the data; This helps prevent overfitting and makes predictions more accurate.

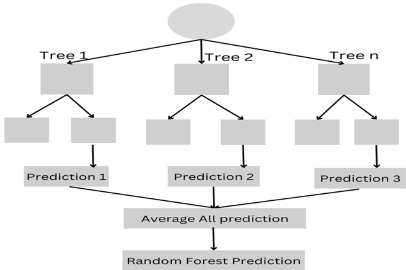



Figure 2: Random Forest Architecture

*C. K MEANS NEAREST ALGORITHM*: When using KNN, an instance-based learning technique, all computations are delayed until the classification is finished and the performance is only roughly estimated at the regional level. For both classification and regression, it will be helpful to provide weights to neighbour contributions such that the closer neighbours add more to the average than the farther neighbours do.

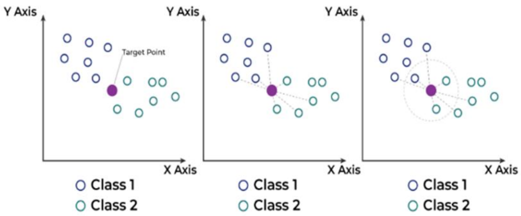



Figure-3 K-Nearest neighbour

## III. LITERATURE SURVEY

2.1 In [1] This paper talks about using a method called Random Forest to predict how much crop will be produced. They used real data from Tamil Nadu to build their models and tested them with samples. They found that the Random Forest method can accurately predict crop yields.

2.2 In [2] This study found that using Random Forest, a type of machine learning, is really good at predicting how much crops will be produced both on a global scale and in specific regions. They found it's more accurate and useful compared to another method called multiple linear regression. So, Random Forest is a great tool for forecasting agricultural production.

2.3 In [3] This study suggests using two ensemble models, AdaNaive and AdaSVM, to predict crop production over a period of time. These models use a technique called AdaBoost to enhance the performance of Naive Bayes and Support Vector Machine (SVM) algorithms. Basically, they're using these models to make better predictions about how much crops will be produced.

2.4 In [4] This research introduces a machine learning method to predict agricultural production based on climate conditions. They created a user-friendly webpage called Crop Advisor, which helps anticipate how weather affects crop yields. They used the C4.5 algorithm to identify which climate factors most affect crop yields in specific districts of Madhya Pradesh. They implemented Decision Trees in their study to make these predictions. Essentially, they're using technology to figure out how weather impacts crop growth in certain areas.

2.5 In [5] In the October 2016 issue of the International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), volume 5, issue 10, there's a discussion about how soil analysis and interpreting test results are currently done using paper-based methods. This outdated approach sometimes leads to mistakes in understanding soil test results. As a result, farmers might receive incorrect recommendations for crops to grow, soil improvements, and fertilizers to use. This can lead to lower crop yields, soil lacking in important nutrients, and using too much or too little fertilizer.

2.6 In [6] At Symbiosis International University, the Symbiosis Institute of Geoinformatics is conducting a research project using precision agriculture techniques on aerial platforms to calculate crop yield for insurance purposes. Precision agriculture involves using remote sensors and geospatial methods to detect differences in fields and respond accordingly. These methods help address factors like crop stress, irrigation methods, and pest or disease occurrences that can affect crop growth. In this research, they are employing Ensemble Learning (EL) techniques to implement their methods. Essentially, they're using advanced technology to gather data from above the fields, analyze it, and then use that information to accurately calculate crop yields. This can be especially useful for insurance purposes, ensuring that farmers receive fair compensation based on their actual crop production.

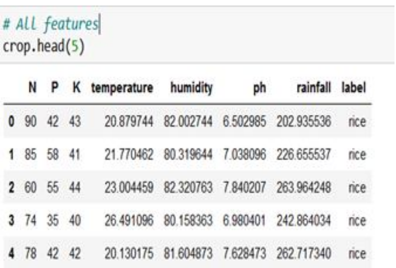
2.7 In [7] In the International Journal of Research in Technology and Engineering, there's a study discussing crop yield prediction using data mining techniques. The main aim of this research is to create a user-friendly tool that helps farmers analyze rice production based on available data. They used various data mining methods to forecast agricultural output, aiming to maximize crop production. One of the techniques they used is the K-Means method, which predicts factors related to atmospheric pollution.

2.8 At the Institute on the Environment at the University of Minnesota, they're using a method called random forests to predict crop yields around the world and in specific regions. They found that random forests are really accurate, making them a great tool for forecasting agricultural production on both small and large scales. In their research, they also used other techniques like K nearest neighbour and support vector regression to help with their predictions. This shows they're using multiple methods to ensure the accuracy of their results.

## IV. FIGURES AND TABLES

### *A. Result*

As a result we get visual of leaf with actual, predicted and confidence label and history parameter which shows loss and accuracy probability .And on an average we found that with the help of CNN algorithm our accuracy reach approximately 98% in detection of disease in plant because of the training we given to our dataset which is as shown below.



```
# All features
crop.head(5)
```

| | N | P | K | temperature | humidity | ph | rainfall | label |
|---|---|---|---|---|---|---|---|---|
| 0 | 90 | 42 | 43 | 20.879744 | 82.002744 | 6.502985 | 202.935536 | rice |
| 1 | 85 | 58 | 41 | 21.770462 | 80.319644 | 7.038096 | 226.655537 | rice |
| 2 | 60 | 55 | 44 | 23.004459 | 82.320763 | 7.840207 | 263.964248 | rice |
| 3 | 74 | 35 | 40 | 26.491096 | 80.158363 | 6.980401 | 242.864034 | rice |
| 4 | 78 | 42 | 42 | 20.130175 | 81.604873 | 7.628473 | 262.717340 | rice |

Figure 4. Features of Dataset

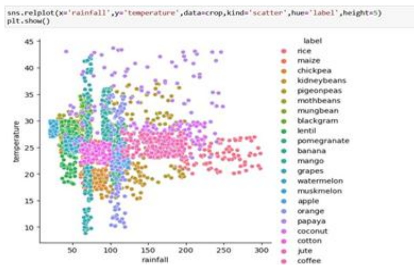



Figure 5. Image Visualization samples

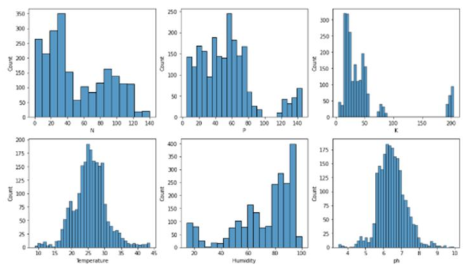


Figure 6. Predicted model visualization

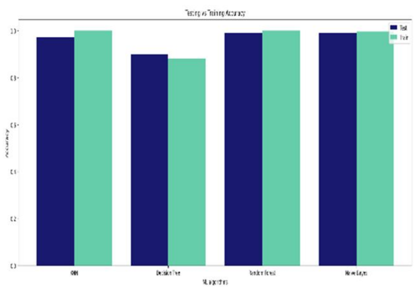


Figure 7.Testing dataset Prediction

### *B. Abbreviations and Acronyms*

*ML: Machine Learning*

Machine learning is a subset of artificial intelligence (AI) that focuses on the development of algorithms and models that allow computers to learn and improve from experience without being explicitly programmed. Used to analyze large amount of data and pattern in plant disease detection.

*KNN: K- Nearest Neighbour*

instance-based learning algorithm used for classification and regression tasks. But it has limitation when applied to high dimensional data such as images we used this for comparison of result with CNN algorithm.

*Decision -Tree*

For both classification and regression tasks, a decision tree is one of the most effective supervised learning techniques. The system constructs a tree structure resembling a flowchart, in which every internal node signifies a test on an attribute, every branch indicates a test outcome, and every leaf node, or terminal node, has a class label.

### *C. Equations*

Mathematical representation to find the distance between two point:
K-Nearest-Neighbour :

$$d=\sqrt{(x2-x1)^2+(y2-y1)^2}$$

### *D. Other Recommendations*

Predictive Modeling for Pest and Disease Management: Create machine learning models to forecast pest and disease outbreaks by taking historical data, crop traits, and environmental factors into account. Farmers can reduce risks and decrease production losses by proactively identifying potential hazards and taking appropriate action.

## V. CONCLUSIONS

Especially in agro-based nations like India, the Crop Yield Enhancement System (CYES) described in this paper is a viable answer to the pressing problem of crop yield improvement. Improving crop output is crucial to meeting the demands of a growing population because it is a key factor in determining economic growth. Novel techniques have been developed as a result of the current agricultural problems, such as crop hybridization, poor seed quality, and the use of chemicals in fertilizers.

The Crop Selection Method (CSM), a complete system offered by CYES, is intended to maximize crop sequencing in order to attain a greater net production rate throughout the season. According to their length growth, crops are categorized by CSM into four types: whole-year, short-term, long-term, and seasonal plantations. It offers suggestions for crop combinations that maximize the cumulative yield rate over the season. CSM's adaptability to the limited land resources enables the reusability of land, further enhancing agricultural productivity.

## ACKNOWLEDGMENT

We owe special debt of gratitude to our Professor Ms. Kanchan Dixit Department of Computer Science & Engineering for guiding us for her constant support and guidance throughout the course of our work. We also take the opportunity to acknowledge the contribution of Professor (Dr.) Divya Mishra, Head, Department of Computer Science & Engineering for her full support and assistance during the development of the project.

## REFERENCES

[1] P. Priya, U. Muthaiah M. Balamurugan. Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.

[2] J. Jeong, J. Resop, N. Mueller and team. Random- forests for global and regional crop yield prediction. PLoS ONE Journal.

[3] Narayanan Balkrishnan and Dr. Govindarajan Muthukumarasamy. Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE).

[4] S. Veenadhari, Dr. Bharat Misra, Dr. CD Singh. Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI).

[5] Shweta K Shahane, Prajakta V Tawale . Prediction On Crop Cultivation. International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016.

[6] Baisali Ghosh. A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Atur Centre, Gokhale Cross Road, Model Colony, Pune– 411016.

[7] D. Ramesh and B. Vardhan, “Analysis of crop yield prediction using data mining techniques”, International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2015.