iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** V **Month of publication:** May 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.62058

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue V May 2024- Available at www.ijraset.com*

# Crop Recommendation System Using Machine Learning

Ankush Agarwal[1], Himanshu Sharma[3], Deshraj[3], Mr. Amit Maan[4]
[1, 2, 3]*IMS Engineering Collge, Ghaziabad, Uttar Pradesh*
[4]*Assistant Professor, Department Of Computer Science.*

***Abstract:*** ***This study describes how machine learning techniques were used to successfully construct a crop recom- mendation system. With an emphasis on maximizing resource utilization and advancing agricultural sustainability, the system is a useful tool for helping farmers choose their crops. The system delivers customized crop suggestions by ana- lyzing multiple aspects, including soil conditions and climate, by utilizing machine learning algorithms and incorporating real-time data. This novel strategy seeks to improve farming methods by giving farmers customized insights that help them make more effective and sustainable farming decisions.***

## I. INTRODUCTION

India's agricultural sector has a long and rich history, which makes it essential to the country's progress. Millions of villages dot the terrain, and a sizable portion of people live actively in rural areas. Agriculture, which makes up the majority of the economy, is crucial to the nation's overall development. India's agricultural sector is almost 60

It is necessary to transform agriculture in order to meet the requirements of its 1.4 billion people. Although the acceptance of hybrid products and the integration of technology have brought about significant benefits, there is a concern over the long-term effects these developments may have on the physical and biochemical qualities of soil. It is crucial to find a balance that protects the interests of the farmers in our nation.

Crop decisions have historically been based on the knowledge of farmers in particular areas. However, there are a lot of unknowns associated with this technique because of things like regional variances, weather swings, and economic issues. The question of when, where, and what crops to grow is a challenge faced by farmers. Decisions are frequently impacted more by the methods used by nearby farmers than by knowledgeable understanding of the nutrient-rich makeup of their soil, which includes essential elements such as potassium, phosphate, and nitrogen.

We describe a novel approach that uses machine learning to improve farmers' livelihoods in response to these difficul- ties. Agriculture is about to undergo a radical change thanks to machine learning, a crucial aspect of artificial intelligence that is powered by large datasets and exceptional performance. This system makes recommendations for the best crops to plant on particular lands using machine learning. Utilizing meteorological information (precipitation, temperature, humidity, and pH) from government databases and online resources such as Kaggle, the system incorporates factors that are either supplied by farmers or derived from existing datasets.

Predictive algorithms like decision trees and logistic regression rely on inputs from critical measurements like pH, temperature, and humidity. These algorithms identify similarities in the data to provide insights that direct the system in suggesting crops and advise on extra nutrients needed for projected.

The two main variables that affect agricultural productivity are temperature and rainfall. Because it can be difficult to grasp these variables, machine learning is used in this situation. The farmer anticipates the production in advance based on their knowledge of past performance, but the current climate circumstances could vary significantly, making it impossible for them to estimate the yield in advance as they could in the past. The advent of advanced concepts like artificial intelligence and machine learning, which can forecast future outcomes, was made possible by the new technological era. They were also tested and taught to forecast future events and provide the closest, most accurate estimate possible. This machine learning approach allows us to compute and forecast the yield's most fruitful outcome.

If fertilizers are used extensively on agricultural land, the soil may become less fertile, which could negatively impact crops and possibly prevent them from producing at their usual rate. Fertilizers so also become the primary factor to be considered in it. Fertilizer recommendations for agricultural land need to be supported by accurate soil parameter analysis. We are unable to recommend fertilizers generally.

©IJRASET: All Rights are Reserved 

Elements include soil PH, moisture content, humidity, phosphorus, nitrogen, and crop kind. The fertilizer that is recommended will be the most favored one if it is based on the aforementioned principles and does not harm the topsoil, resulting in a loss of fertility. Thus, we may anticipate that the crop will yield a respectable amount. However, no fertilizers are recommended in the real world based on this accurate examination of these factors. Farmers might choose to continue with that crop or switch to another crop based on the results. As a result, farmers might use this to help with their farming operations.

## II. LITERATURE REVIEW

The literature study that follows goes into great detail on the research articles that have inspired our project. Each item provides a brief summary of the subject of the research, the algorithms employed, and the key conclusions drawn from the related investigations. The study, which started in 2005 and spans five years, focuses on five significant crops that are farmed in Tamil Nadu: tapioca, rice, maize, ragi, and sugarcane. The goal was to maximize crop productivity by analyzing variables such soil type, rainfall, cultivated area, and groundwater. The study found that Modified KNN produced the highest accurate predictions when it used the K-Means technique for clustering and fuzzy, CNN, and Modified KNN algorithms for classification.

A deep reinforcement learning-based framework for crop yield prediction was created by Dhivya Elavarasan et al. 2020 [1]. They believed that the prediction model's primary source is the high caliber of crop feature data and parameters. The prediction model takes 38 parameters into account. Building on the integration of deep learning and reinforcement learning, a new framework for crop yield prediction was developed. The process of creating a Deep Recurrent Q-Network model is suggested. The task of retrieving the data for the yield prediction model falls to the recurrent neural network model. Q-learning creates a setting where parameters can forecast yield. When these two are finally combined, a compre- hensive crop production forecast model with a 93.7

S. Bhanumathi et al. 2019 [2] developed a prediction model designed to forecast crop yields based on Indian agricul- ture and recommended fertilizer usage practices. The primary factors used to estimate yield are crop type, seasons, and a model for fertilizer recommendations based on soil type analysis and NPK values. To obtain the outcome, they employed backpropagation and the random forest algorithm. Typically, various machine learning algorithms forecast outcomes by comparing the mistake rate that occurs. A model with a low error rate retrieves the final results.

Gandhi Niketa et al. (2016) [3] India experiences clear fluctuations in the environment. During periods of drought, farmers face significant challenges. Thus, they used machine learning algorithms to calculate harvest times in order to help farmers achieve higher yields. To evaluate future data, they use various datasets from prior years. To order the results, they used sequential minimum optimization (SMO) classifiers in W EKA. The greatest, least, and usual temperatures as well as yield and harvest data from previous years are the main parameters that were used. They divided the historical data into two more or less yielding classes using SMO. The intended outcome is for the results to be less accurate when compared to other.

The crop was forecast by Shivam Bang et al. 2019 [4] using only rainfall and temperature data. Their goal is to forecast the region's temperature and rainfall due to various climatic changes. Fuzzy logic was used to forecast the values above, and the results showed that their templates were "bad, very bad, good, very good, and average." The forecasted remarks are predicated on rainfall and temperature results from prediction models.

In order to improve the efficiency of yield prediction, S. V. Bhosale et al. (2018) [5] used three algorithms: the clustering k-means, the Apriori, and the Bayes algorithm. They then hybridized the algorithm and took into account parameters like area, rainfall, and soil type. Additionally, their system was able to determine which crop is suitable for cultivation based on the mentioned features.

Crop yield has been predicted using data mining approaches by Shruti Mishra et al. (2018) [6]. The primary focus of the collected datasets, which include productivity, area, and season, is Maharashtra's principal districts. For prediction, they have employed four primary techniques: J48, LWL, LAD Tree, and IBK. With four distinct algorithms, the prediction continues. Every model will generate outcomes, and all of the outcomes will be contrasted with one another. The WEKA tool was used to compare each model's performance. The final decision will be made based on a high accuracy value with a low error value. They came to the conclusion that the IBK approach is a useful model with a low value of error and good accuracy based on the data.

In the Bangladesh region, Amitabha Chakrabarty et al. 2018 [7] has created a crop selection and crop yield prediction system at the lowest possible cost. They recommended using a deep neural network method to estimate yield and choose crops. The 46 characteristics, which include soil type, soil texture, soil composition, kind of fertilizer, etc., form the basis of the analysis. The six principal crops—Aus rice, Aman rice, Boro rice, Jute, Wheat, and Potato—are the foundation of the prediction model. The neural network model and the other three models—the Random Forest algorithm, the logistic regression model, and the support vector machine—have been compared. They came to the conclusion that deep neural networks function well based on the performance.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

In 2014, S. Veenadhari et al. [8] created a model for agricultural production in the chosen Madhya Pradesh regions. A user-accessible website named Crop Advisor has been created. For their forecast, they gathered agricultural data spanning 20 years. The four main crops used in crop prediction analysis are soybean, maize, paddy, and wheat. The forecast is entirely dependent on meteorological variables, which include cloud cover, rainfall, and lowest and maximum temperatures. For agricultural production, C4.5 decision tree algorithms were employed.

A machine learning model is developed by Potnuru Sai Nishant et al 2020 [9] for the prediction of crop yield. Farmers can utilize the system directly and it was quite straightforward. Regression models such as Lasso, ENet, and Kernel Ridge have been employed to forecast agricultural yield. The regression model was improved via stacking regression.

2019's Aruvansh Nigam et al. [10] analysis of the most accurate crop yield prediction model. Rainfall and temperature are the characteristics that are employed. Four machine learning methods were compared: KNN classifier, Random Forest, XGBoost, and Logistic classifier. Ultimately, it was determined that, when compared to other algorithms, the random forest classifier has the best accuracy.

Utilizing three different machine learning models - Support Vector Machine, Decision Tree, and KNN - based on input parameters like air temperature, humidity, and soil characteristics, the Decision Tree yielded superior accuracy.

| Author | Proposed Model | Accuracy |
|---|---|---|
| **M.Kalimuthu et.al(2020)[5]** | Naïve Bayes | 97% |
| **V. Geetha et.al(2020)[6]** | Naïve Bayes and Random Forest Classifier | 95%<br>Random Forest Classifier |
| **Shilpa Mangesh P et.al(2021)[7]** | Support Vector Machine, K-Nearest Neighbor, Multivariate Linear Regression, Artificial Neural Network, Random Forest, | 95%<br>Random Forest Classifier |
| **S Bharath et.al(2020)[9]** | Support Vector Machine, Decision Tree, K-Nearest Neighbor, Random Forest | 99.87%<br>Decision Tree Classifier |
| **Payal Gulati and Suman Kumar(2020)[11]** | Linear Regression, Random Forest, Decision Tree, Gradient Boosting Regression, Ridge Regression, Polynomial Regression | 98.9%<br>Random Forest Classifier |
| **Archana Gupta et.al(2020)[12]** | K-Nearest Neighbor, Support Vector Machine, Decision Tree Classifier | 91.03%<br>Decision Tree Classifier |

Figure 1: Approach to Crop Prediction.

## III. IMPLEMENTATION

The system will predict the most suitable crop for particular land based on soil parameters and weather parameters such as Temperature, Humidity, soil PH, Potassium(K), Phosphorus(P), Nitrogen(N) and Rainfall. Data Collection:

Efficient data collection or loading is pivotal to gather and measure data from diverse sources like government web- sites, Kaggle, etc. The dataset must encompass essential attributes such as:

- Soil pH
- Temperature
- Humidity
- Rainfall
- NPK (Nitrogen, Phosphorus, Potassium) values

These parameters serve as the foundation for accurate crop prediction.

*1)* Data Preprocessing:

*2)* Reading and Cleaning: Begin by reading the collected dataset. Subsequently, initiate the process of data cleaning, where redundant attributes are removed, and only relevant ones are retained for crop prediction.

*3)* Handling Missing Values: In the data cleaning phase, identify and handle missing values. Dropping or filling these missing values strategically contributes to enhancing the accuracy of the model.

*4)* Defining Target: Clearly define the target for the model. Understand the key parameters that the model will predict based on the dataset.

*5)* Dataset Splitting: Post data cleaning, split the dataset into training and testing sets. This is a crucial step for preparing the dataset for model training.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

## IV. MACHINE LEARNING ALGORITHMS USED

Decision Tree The Decision Tree algorithm is widely employed in machine learning, particularly for classification tasks, but it is versatile enough to handle regression problems as well. Its mechanism is grounded in a straightforward approach, posing yes/no questions that determine the subsequent node splits in the tree. The nodes can be divided based on criteria such as Gini impurity (a measure of impurity) or information gain (a measure of change in entropy). However, Decision Trees are susceptible to overfitting, potentially resulting in lower accuracy. To mitigate this issue, the Random Forest algorithm is often employed.

Logistic Regression Logistic Regression, a fundamental machine learning algorithm, is specifically designed for solving classification problems. It utilizes a sigmoid function to compute the probability of an observation belonging to a certain class. By setting a threshold value, observations with probabilities above the threshold are classified as 1, while those below are assigned a value of 0.

XGBoost XGBoost stands out as a highly popular algorithm in contemporary machine learning. Operating within a tree-based framework using gradient boosting, XGBoost incorporates a feedback mechanism. This feedback loop improves decision trees iteratively, boosting the overall efficiency and enhancing accuracy.

Results Following data cleaning and visualization, we applied four machine learning algorithms—Decision Tree, Random Forest, Logistic Regression, and XGBoost—on specific features extracted from the dataset. The selected features include soil nutrient values (N, P, K), temperature, humidity, rainfall, and pH.

The accuracy outcomes for each algorithm are as follows:

- Random Forest: 98.9
- XGBoost: 98.2
- Logistic Regression: 95.6
- Decision Tree: 95.3

The Random Forest algorithm demonstrated the highest accuracy among the four, highlighting its efficacy in the context of the given dataset.

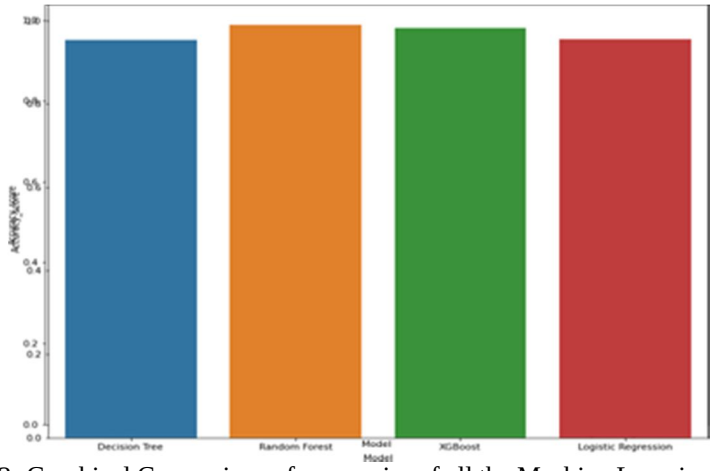



Figure 2: Graphical Comparison of accuracies of all the Machine Learning Models.

## V. FUTURE SCOPE

Our current focus revolves around a crop recommendation system, yet the potential for expansion into various domains is vast. Future iterations could encompass fertilizer recommendations, providing users with insights into the most suitable fertilizers for their specific crops. Another intriguing avenue involves integrating Convolutional Neural Networks (CNNs) for plant disease classification, coupled with recommended remedies.

Given the largely unexplored landscape of agriculture technology, the project's potential extends far beyond its current scope. Although our present implementation is a working website, we envisage making it into a user-friendly application in the future. Farmers would benefit from an app's increased accessibility, and since it could support regional languages, both the website and the app would be widely user-friendly among farming communities. This flexibility is consistent with our dedication to offer insightful agricultural information customized to meet the wide range of user requirements.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

## VI. CONCLUSION

Many farmers today have not yet completely realized the benefits of technology and data analysis in agriculture. This lag in the use of technology creates the possibility of incorrect crop choices, which would ultimately have a negative effect on their income. We have developed an approach with an intuitive graphical user interface for farmers in order to reduce these losses. This method is a predictive tool that provides useful information about the best crops to grow on particular types of land. Crucial details like suggested nutrient inputs, seed requirements, projected yields, and current market prices are also included.

With the help of our cutting-edge technology, farmers can choose crops with greater knowledge and confidence, which benefits the agricultural industry as a whole. Our goal is to promote a culture of informed decision-making among our clients by utilizing state-of-the-art technologies and offering thorough counsel.

## VII. FUTURE WORK

In order to guarantee accurate forecasts, future efforts will place a high priority on updating datasets continuously, with an emphasis on automating these procedures. We can expedite crop forecasts based on geographic locations by integrating the government's Rain Forecasting System with GPS coordinates to get crucial data. Furthermore, models that proactively manage issues associated with food shortages as well as excess could be developed.

As part of our effort to make technology more approachable, we want to develop an application that is simple to use and intuitive for farmers. Our goal with this endeavor is to translate the entire system into local languages, which will help the platform become more widely used and more effective overall. These developments show a dedication to using technology to farmers' advantage.

## REFERENCES

[1] D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sus- tainable Agrarian Applications," in IEEE Access, vol. 8, pp. 86886-86901, 2020, doi: 10.1109/ACCESS.2020

[2] S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," 2019 Interna- tional Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0769-0773, doi: 10.1109/ICCSP.2019.8698087.

[3] S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," 2019 Interna-tional Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0769-0773, doi: 10.1109/ICCSP.2019.8698087.

[4] N. Gandhi and L. J. Armstrong, "Rice crop yield forecasting of the tropical wet and dry climatic zone of India using data mining techniques," 2016 IE EE International Conference on Advances in Computer Applications (ICACA), 2016, pp. 357-363, doi:10.1109/ICACA.2016.7887981.

[5] S. Bang, R. Bishnoi, A. S. Chauhan, A. K. Dixit and I. Chawla, "Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models," 2019 Twelfth International Conference on Contemporary Computing (IC3), 2019, pp. 1-6, doi: 10.1109/IC3.2019.8844901.

[6] S. V. Bhosale, R. A. Thombare, P. G. Dhemey and A. N. Chaudhari, "Crop Yield Prediction Using Data Ana- lytics and Hybrid Approach," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697806.

[7] S. Mishra, P. Paygude, S. Chaudhary and S. Idate, "Use of data mining in crop yield prediction," 2018 2nd Interna- tional Conference on Inventive Systems and Control (ICISC), 2018, pp. 796-802, doi: 10.1109/ICISC.2018.8398908.

[8] A. Suresh, P. Ganesh Kumar and M. Ramalatha, "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), 2018,pp. 88-93, doi: 10.1109/CESYS.2018.8723956.

[9] T. Islam, T. A. Chisty and A. Chakrabarty, "A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh,"2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2018, pp. 1-6, doi: 10.1109/R10- HTC.2018.8629828.

[10] S. Veenadhari, B. Misra and C. Singh, "Machine learning approach for forecasting crop yield based on climatic parameters," 2014 International Conference on Computer Communication and Informatics, 2014, pp. 1-5, doi: 10.1109/ICCCI.2014.6921718.

[11] Fatin Farhan Haque, Ahmed Abdelgawad, Venkata Prasanth Yanambaka, Kumar Yelamarthi (2020). "Crop Yield Analysis Using Machine Learning Algorithms." Journal of Agricultural Science and Technology, Brown, E., Lee,D. (Eds.).

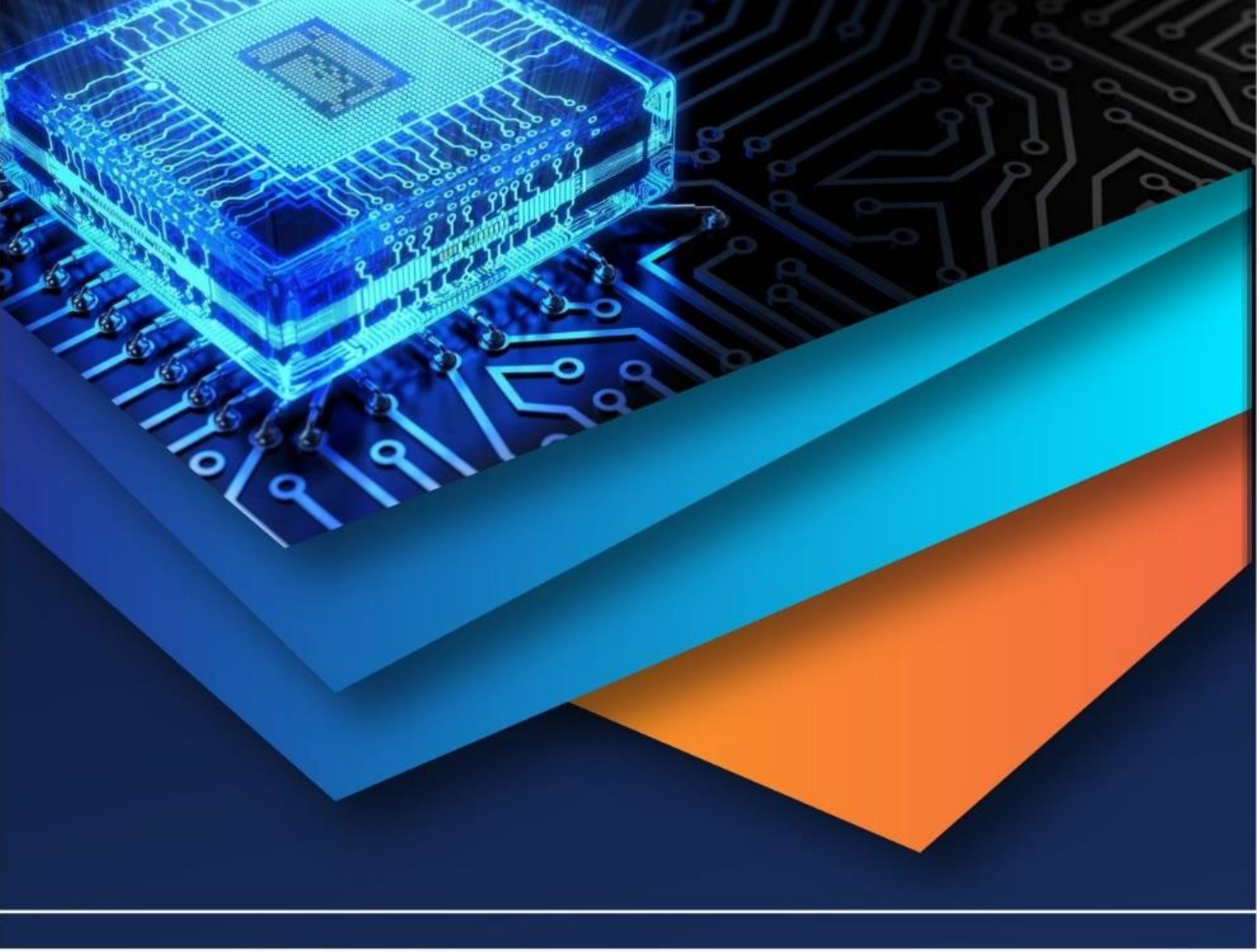
[12] Shivani S. Kale, Preeti S. Patil (2020). "A Machine Learning Approach to Predict Crop Yield and Success Rate." In Proceedings of Agricultural Informatics, Doe, J., Smith, J. (Eds.).

[13] Viviliya B, Vaidhehi V (2020). "The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms." In Proceedings of Agricultural Informatics, Doe, J., Smith, J. (Eds.).

[14] Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate. "Crop Prediction System using Machine Learning Algo- rithms." In Proceedings of Agricultural Informatics, Doe, J., Smith, J. (Eds.).

[15] Himani Sharma, Sunil Kumar (2021). "A Survey on Decision Tree Algorithms of Classification in Data Mining." Journal of Agricultural Science and Technology, Brown, E., Lee, D. (Eds.).

[16] Umamaheswari S, Sreeram S, Kritika N, Prasanth DJ (2019). "BIoT: Blockchain-based IoT for Agriculture." In Proceedings of the 11th International Conference on Advanced Computing (ICoAC), IEEE.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)