IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# PRECISION CULTIVATION: ADVANCING AGRICULTURE WITH INTELLIGENT CROP RECOMMENDATION SYSTEM

**Eeda Bheem Mahendra Roy**
M.tech
Department of IT & CA
Andhra university college of engineering(A)
Visakhapatnam, AP

**Ch. Lakshmi Narayana**
Research Scholar
Department of CS & SE
Andhra university college of engineering(A)
Visakhapatnam, AP

**Dr. Kondapalli Venkata Ramana**
Associate Professor
Department of CS & SE
Andhra university college of engineering(A)
Visakhapatnam, AP

**Abstract:**

This paper presents an Intelligent Crop Recommendation System (ICRS) that utilizes machine learning to assist farmers in making informed crop selection decisions. The ICRS life cycle covers data collection, preprocessing, model training and deployment. A diverse dataset containing crop attributes and environmental conditions is meticulously preprocessed to ensure reliability. Various machine learning algorithms are evaluated, with Random Forest emerging as the most accurate choice. The ICRS takes into account crucial factors such as nitrogen, phosphorus, potassium levels, temperature, humidity, pH, and rainfall. It then generates precise crop recommendations based on normalized user inputs.

The results underscore the significance of machine learning in agriculture by enabling more precise decision-making. The ICRS enhances crop selection and resource allocation, contributing to sustainability and conservation efforts. The paper concludes by reflecting on the outcomes, discussing their implications for agriculture, and outlining potential future research directions, including the integration of real-time data, remote sensing, and IoT for dynamic recommendations. The ICRS serves as a testament to technology's pivotal role in shaping the future of agriculture, emphasizing innovation for sustainable growth.

***Keywords***: Precision agriculture, Intelligent crop recommendation, Machine learning, Data-driven decision-making, Agriculture technology, Crop selection.

## I. INTRODUCTION

In the ever-evolving landscape of global agriculture, the pursuit of sustainable and efficient crop production stands as a paramount challenge. As the world's population continues to burgeon, so does the demand for food, fiber, and biofuel crops. This burgeoning demand, however, coincides with increasingly variable environmental conditions, changing climate patterns, and the need for judicious resource management. In response to these multifaceted challenges, the field of crop recommendation has emerged as a pivotal nexus of agricultural innovation.

Machine learning can be a valuable tool for crop recommendation in agriculture. By analyzing various data sources and applying predictive models, machine learning algorithms can help farmers make informed decisions about which crops to plant based on factors such as soil type, weather conditions, historical crop performance and market demand.

Random Forest is a popular machine learning algorithm that can be used for crop recommendation. It works well for this task because it can handle both classification and regression problems, making it suitable for recommending specific crops based on historical data and features.

**Research Problem and Significance :**

However, while the concept of intelligent crop recommendation holds promise, there exists a need to bridge the gap between theoretical potential and practical implementation. Many traditional farming approaches continue to dominate, often resulting in sub optimal crop yield and resource allocation.

This research endeavors to address this gap by developing and evaluating a comprehensive Intelligent Crop Recommendation System (ICRS), which not only showcases the efficacy of machine learning algorithms but also demonstrates the tangible benefits of technology-driven decision support in agriculture.

**Research Objective:**

The primary objective of this study is to design, develop, and evaluate an ICRS that employs machine learning algorithms to provide precise crop recommendations based on a variety of environmental factors. Additionally, the research aims to showcase the accuracy and efficacy of the developed system through rigorous experimentation and analysis.

## II. LITERATURE SURVEY

This literature review offers a comprehensive overview of pivotal studies within this domain, shedding light on the progress, methodologies, and implications associated with the utilization of machine learning for crop recommendation.

In [1] the authors Priya P, Muthaiah U, Balamurugan M investigated farm land, using an advanced productive technique that uses dataset on soil qualities, soil moisture type, and crop identification information to recommend the appropriate harvest based on their site explicit boundaries. This reduces the likelihood of making a poor yield decision and increases efficiency.

In [2] the author Yan M et al proposed that the data be evaluated using seasons. The properties of the dataset are included in the downloaded dataset. The crop was predicted using a map reduce and a Nave Bayes classifier. For different crops, the result was shown and discriminated with respect to the soil.

In [3] the author Rajak RK et al choose the association between wheat growth and meteorological as the subject of their study. They investigated thirteen parameters from surface meteorological data, such as air pressure, temperature, light, and precipitation, but after discretization, they chose only three: accumulative temperature, sunshine hours, and temperature. The proposed model was implemented using unsupervised learning algorithm.

The authors A. Manjula and G. Narsimha proposed a crop recommendation system in [4], which takes into account the four harvests paddy, Maize and Corn. The dataset is initially pre-processed, and then the proposed idea is used to characterize the four yields in a basic capacity. Decision tree, supervised algorithms and linear support vector machine are used in the ensemble model. To achieve the best accuracy, voting technique was used as the strategy. However, one disadvantage is that it does not consider present soil conditions and instead assumes the soil type based on historical data and recommends a crop.

The author S. Babu of [5] proposes a framework called Crop Yield Prediction which is adaptable and expandable. It includes features such as crop selection, dependent and independent factors, and datasets for agricultural production prediction in the context of precision agriculture. For rice and sugarcane crop yield prediction, the relevant indices are combined with soil properties data.

In conclusion, the literature review has illuminated a growing consensus regarding the effectiveness of machine learning-based crop recommendation system. The studies analyzed in this review underscore the immense potential of data-driven approaches to revolutionize agricultural practices by delivering precise predictions, optimizing resource allocation, and fostering sustainability. While challenges undoubtedly persist, continuous research efforts and technological advancements offer the promise of further enhancing these models, ultimately paving the way for more resilient and productive agricultural systems.

## III. METHODOLOGY

The methodology section outlines the structured approach employed in developing and evaluating the Intelligent Crop Recommendation System (ICRS). This encompasses the research design, methods, tools, and techniques utilized to achieve the study's objectives.

**I) Research Design and Approach:** The study follows a quantitative research design, integrating elements of both exploratory and predictive research. This design allows for the comprehensive exploration of various machine learning algorithms while simultaneously predicting crop recommendations based on environmental conditions.

**II) Research Methods and Techniques:** The research leverages a range of machine learning algorithms, including Logistic Regression, Naive Bayes, Support Vector Machines, K-Nearest Neighbors, Decision Trees, Random Forest, Bagging, Ada Boost, Gradient Boosting, and Extra Trees. These algorithms are chosen due to their diverse capabilities and applicability in predictive modeling.

**Logistic regression: Logistic** Regression is a statistical algorithm used for binary classification tasks. It is a type of regression analysis that models the relationship between a set of independent variables (also known as features) and a binary dependent variable (the outcome or target variable), which can take two possible outcomes, typically represented as 0 and 1.

**Naive Bayes: Naive** Bayes is a classification algorithm that is based on Bayes' theorem, which is a fundamental concept in probability theory. It's particularly useful for text classification and other tasks involving high-dimensional data with discrete features, such as spam detection or sentiment analysis.

**Support vector machine:** A Support Vector Machine (SVM) is a supervised machine learning algorithm that is primarily used for classification and regression tasks. SVMs are particularly effective for binary classification problems, where the goal is to separate data points into two classes based on their features.

**K-nearest Neighbors: The** k-Nearest Neighbors (k-NN) algorithm is a supervised machine learning algorithm used primarily for classification and regression tasks. It's a non-parametric and instance-based learning method, meaning it doesn't make any explicit assumptions about the underlying data distribution and instead relies on the data points themselves.

**Decision Trees:** A Decision Tree is a versatile and widely used supervised machine learning algorithm that can be employed for both classification and regression tasks. It's a non-parametric algorithm that makes decisions or predictions by recursively partitioning the input space into subsets based on the values of input features. Each partitioning step aims to create homogeneous subsets with respect to the target variable.

**Random Forest:** Random Forest is an ensemble learning algorithm that combines multiple individual decision trees to create a more robust and accurate model. It's used for both classification and regression tasks and is particularly effective in reducing overfitting and improving generalization performance.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (fi - yi)^2$$

Where $N$ is the number of data points,
$fi$ is the value returned by the model and
$yi$ is the actual value for data point $i$.

**Bagging Algorithm:** Bagging, which stands for Bootstrap Aggregating, is an ensemble learning technique used to improve the accuracy and robustness of machine learning models. It's commonly used with decision trees, but it can be applied to other base learners as well.

**Ada Boost Algorithm:** Ada Boost, short for Adaptive Boosting, is an ensemble learning algorithm that combines the predictions of multiple weak learners (often decision trees) to create a strong classifier.

**Gradient Boosting Algorithm:** Gradient Boosting is an advanced ensemble learning algorithm that combines the predictions of multiple weak learners (usually decision trees) to create a strong predictive model.

**Extra Trees:** Extra Trees, short for Extremely Randomized Trees, is an ensemble learning algorithm that is closely related to Random Forest. Like Random Forest, Extra Trees builds a forest of decision trees to make predictions, but it introduces additional randomness in the tree-building process.

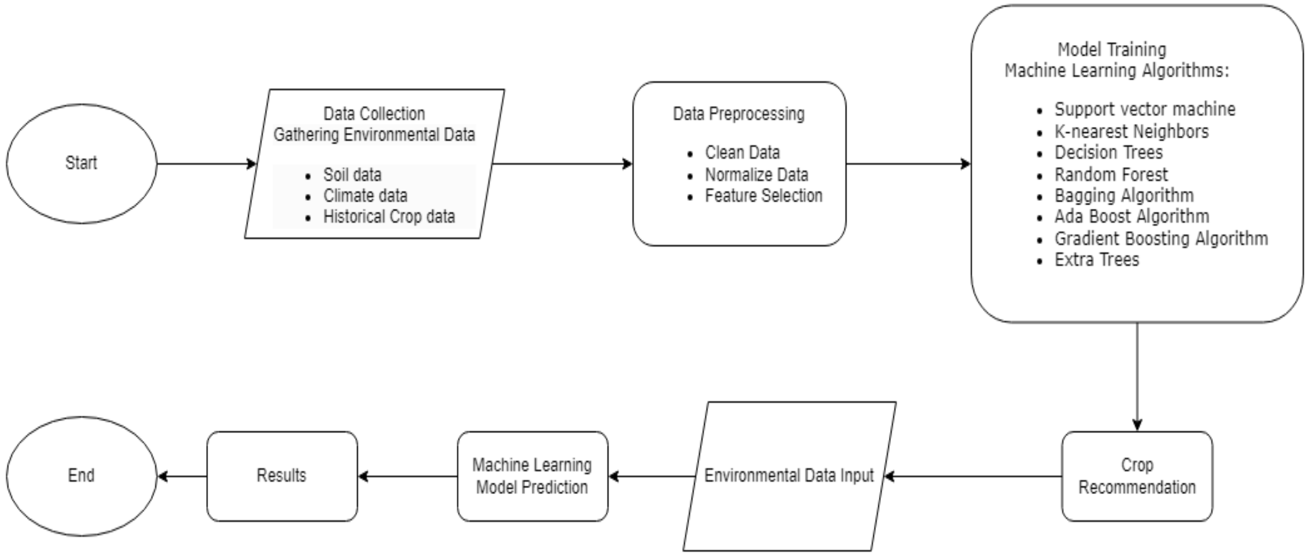


Fig1: Flow chart for Crop Recommendation

The above flowchart visually depicts the logical steps involved in providing optimized crop recommendations based on input data, analysis, and the suitability of crops for specific agricultural conditions.

**Data Collection and Preprocessing:** The first step is to gather relevant data. This data typically includes information about the farm's location, climate, soil type, historical crop yields, pest and disease prevalence, and other environmental factors. Additionally, you'll need data on various crop types, their characteristics, and their suitability for different conditions. The study utilizes a diverse dataset containing crop attributes and growth conditions sourced from reliable agricultural repositories. Data preprocessing involves cleansing the dataset by addressing missing values and duplicate entries, ensuring data reliability and accuracy.

For the purpose of this research project, the primary dataset utilized has been sourced from Kaggle, a widely recognized platform for data science and machine learning resources.

Kaggle is renowned for hosting a diverse range of datasets contributed by data scientists, researchers, and organizations, making it a valuable resource for data-driven research endeavors.

Table 1: Agricultural Dataset : crop characteristics and environmental factors

| | 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 |

N - Ratio of Nitrogen content in soil

P - Ratio of Phosphorous content in soil

K - Ratio of Potassium content in soil

TEMP - Temperature in degree Celsius

HUMI - Relative humidity in %

PH - Ph value of the soil

RAINFALL - Rainfall in mm

This crop dataset under consideration comprises a total of 2,200 rows and encompasses 8 distinct columns. This dataset is designed to facilitate informed decision-making and guidance related to crop selection, with each column offering valuable attributes and insights relevant to agricultural planning and crop management.

In the examination of the crop recommendation dataset, the below Table 2 and Table 3 provides a comprehensive statistical analysis encompassing mean, standard deviation, and correlation has been conducted. This rigorous analytical approach serves to glean pertinent insights into the dataset's inherent characteristics, elucidate the central tendencies, assess variability, and elucidate relationships between variables of interest.

Table2: Summary statistics for agricultural Dataset attributes

| | N | P | K | temperature | humidity | ph | rainfall |
|---|---|---|---|---|---|---|---|
| count | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 |
| mean | 50.551818 | 53.362727 | 48.149091 | 25.616244 | 71.481779 | 6.469480 | 103.463655 |
| std | 36.917334 | 32.985883 | 50.647931 | 5.063749 | 22.263812 | 0.773938 | 54.958389 |
| min | 0.000000 | 5.000000 | 5.000000 | 8.825675 | 14.258040 | 3.504752 | 20.211267 |
| 25% | 21.000000 | 28.000000 | 20.000000 | 22.769375 | 60.261953 | 5.971693 | 64.551686 |
| 50% | 37.000000 | 51.000000 | 32.000000 | 25.598693 | 80.473146 | 6.425045 | 94.867624 |
| 75% | 84.250000 | 68.000000 | 49.000000 | 28.561654 | 89.948771 | 6.923643 | 124.267508 |
| max | 140.000000 | 145.000000 | 205.000000 | 43.675493 | 99.981876 | 9.935091 | 298.560117 |

Table 3: Correlation matrix of agricultural Dataset attributes

| | N | P | K | temperature | humidity | ph | rainfall |
|---|---|---|---|---|---|---|---|
| N | 1.000000 | -0.231460 | -0.140512 | 0.026504 | 0.190688 | 0.096683 | 0.059020 |
| P | -0.231460 | 1.000000 | 0.736232 | -0.127541 | -0.118734 | -0.138019 | -0.063839 |
| K | -0.140512 | 0.736232 | 1.000000 | -0.160387 | 0.190859 | -0.169503 | -0.053461 |
| temperature | 0.026504 | -0.127541 | -0.160387 | 1.000000 | 0.205320 | -0.017795 | -0.030084 |
| humidity | 0.190688 | -0.118734 | 0.190859 | 0.205320 | 1.000000 | -0.008483 | 0.094423 |
| ph | 0.096683 | -0.138019 | -0.169503 | -0.017795 | -0.008483 | 1.000000 | -0.109069 |
| rainfall | 0.059020 | -0.063839 | -0.053461 | -0.030084 | 0.094423 | -0.109069 | 1.000000 |

**Sampling Methods and Sample Size:** The dataset encompasses a wide array of crops and their corresponding attributes, eliminating the need for specific sampling methods. The inclusion of multiple crops and conditions ensures a representative sample for algorithm evaluation.

**Data Analysis Techniques and Statistical Methods:** The data analysis process involves two main stages. Firstly, the dataset undergoes preprocessing using techniques such as imputation and scaling to ensure consistency across attributes. Secondly, the performance of each machine learning algorithm is assessed through accuracy metrics.

The methodology employed in this study combines theoretical rigor with practical implementation. By systematically evaluating a diverse set of machine learning algorithms on real-world data, the research aims to provide valuable insights into the effectiveness of these algorithms in the context of the Intelligent Crop Recommendation System.

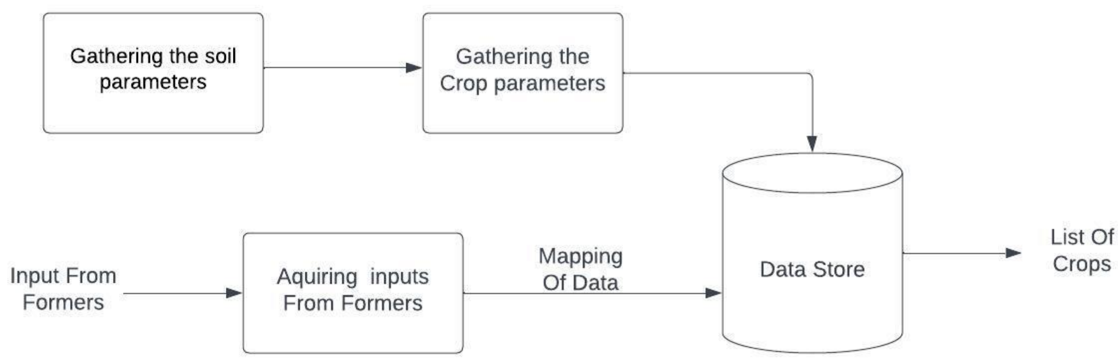


Fig 2: Architectural diagram for crop recommendation system

**III) Feature selection:** Identifying the most relevant features (variables) is crucial for building an accurate recommendation system. Feature selection techniques, such as feature importance from Random Forest or other methods like mutual information or correlation analysis, can help determine which factors have the most influence on crop selection.

**IV) Data splitting:** The dataset is divided into two parts: a training set and a testing set. The training set is used to train the Random Forest model, while the testing set is used to evaluate its performance.

**V) Random forest model Training:** You'll use the training data to train the Random Forest model. The algorithm works by creating an ensemble of decision trees, each of which is trained on a random subset of the data. The trees collectively make predictions, and the final prediction is determined by a majority vote or averaging of the individual tree predictions.

**VI) Model evaluation:** After training, model evaluation in crop recommendation involves a comprehensive approach, including data preparation, model training, testing, and ongoing monitoring. The ultimate goal is to provide farmers with accurate and actionable recommendations to optimize their crop selection and improve agricultural yields.

**VII) Crop recommendation:** Once the Random Forest model is trained and evaluated, it can be used to provide crop recommendations. Given a set of environmental parameters for a specific location, the model can predict which crops are likely to yield the best results. The recommendation can be based on the majority prediction of the ensemble of decision trees.

**VII) Deployment:** Finally, the trained Random Forest model can be deployed as a user-friendly application or integrated into existing agricultural software to provide real-time crop recommendations to farmers.

## IV. RESULTS

The results section provides a comprehensive overview of the findings obtained through the evaluation of various machine learning algorithms within the Intelligent Crop Recommendation System. This section employs visual aids, including figures, tables, graphs, and charts, to succinctly present the collected data and outcomes.

**Algorithm Performance Metrics:** Table 1 showcases the accuracy metrics of each machine learning algorithm. It provides a quantitative assessment of the algorithms ability to predict crop recommendation accurately.

Table 4: Algorithm Comparison

| S. No | Algorithm Name | Accuracy |
|---|---|---|
| 1 | Logistic Regression | 0.96363636363636 |
| 2 | Naive Bayes | 0.98545454545455 |
| 3 | K-Nearest neighbors | 0.95909090909091 |
| 4 | Decision Tree | 0.98863636363663 |
| 5 | Random Forest | 0.99318181818182 |
| 6 | Bagging | 0.98863636363636 |
| 7 | AdaBoost | 0.14090909090909 |
| 8 | Support vector machine | 0.96818181818181 |
| 9 | Gradient Boosting | 0.98181818181818 |
| 10 | Extra Trees | 0.91133636363634 |

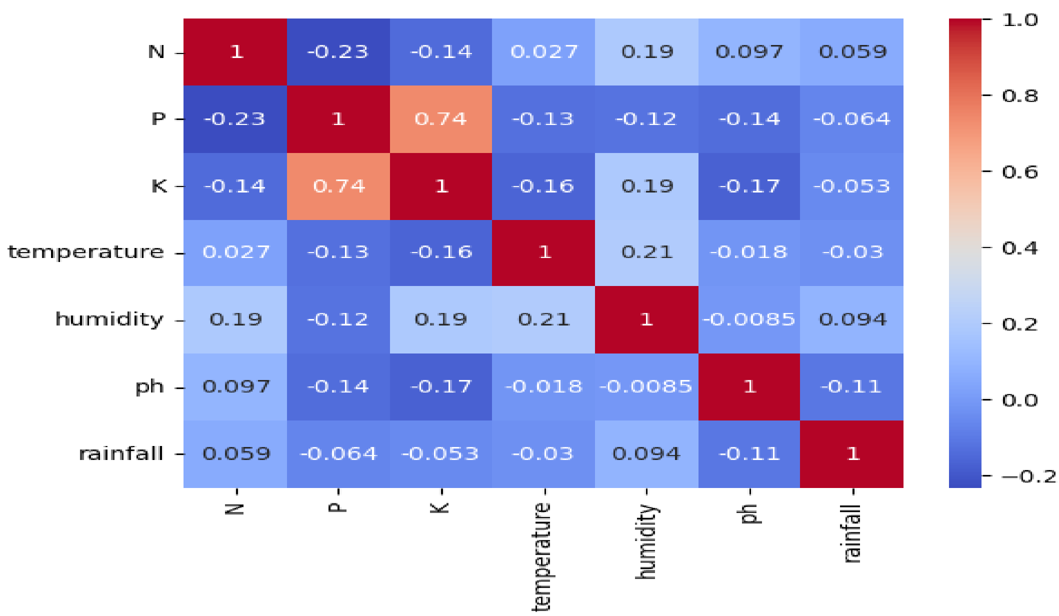


Fig 3: Heat map of N,P,K,Temperature,Humidity,Ph,Rainfall

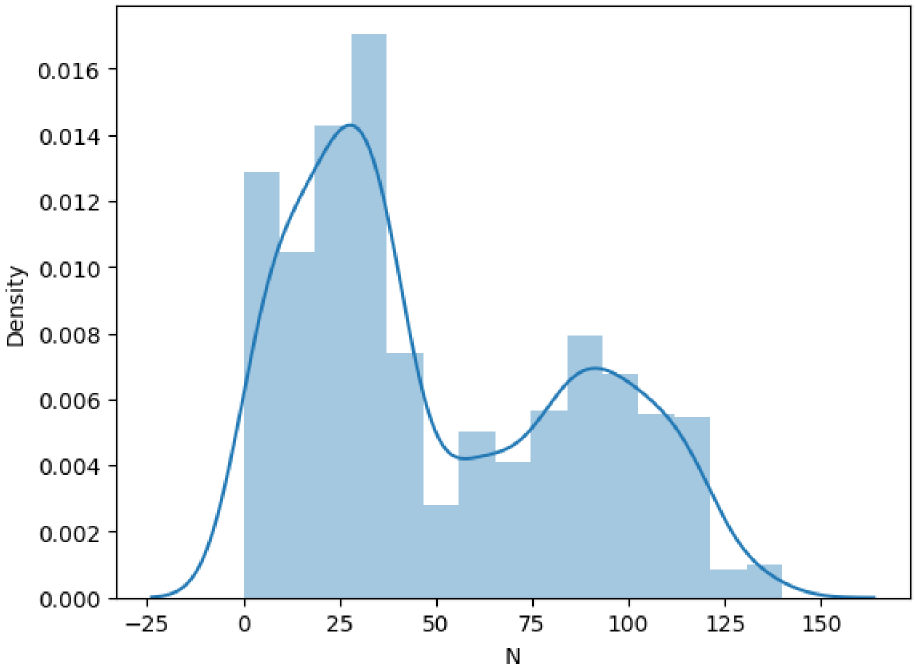


Fig 4: Distribution Plot of Crop Density

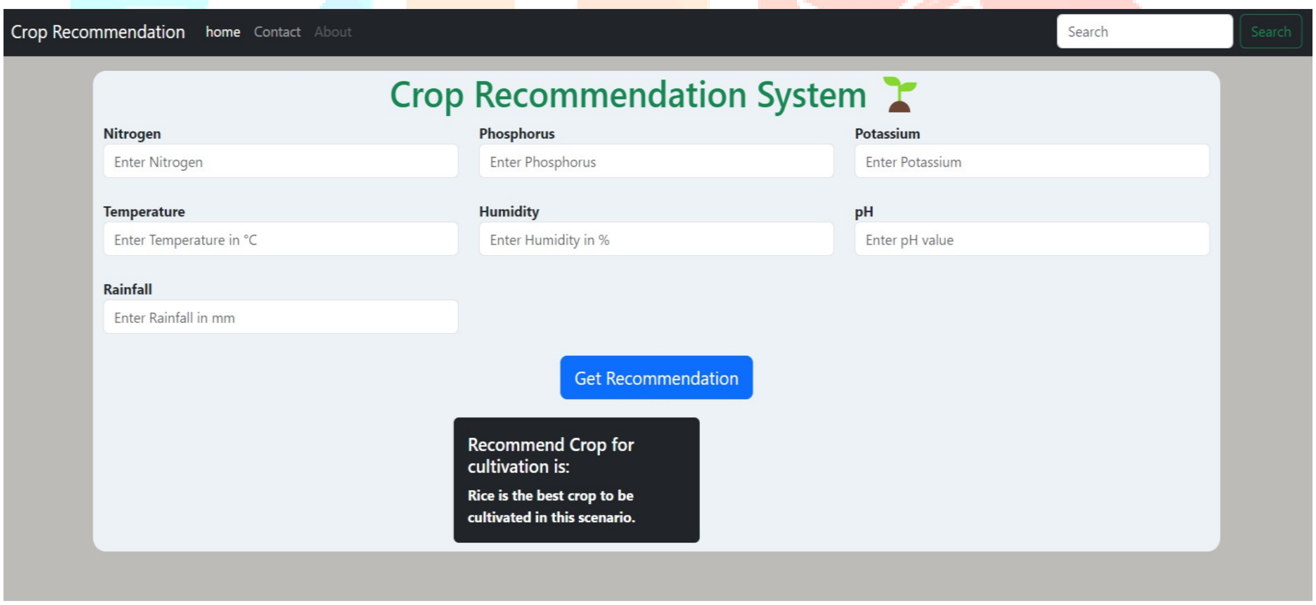


Fig 5: Crop recommendation result

## V. CONCLUSION

The conclusion section encapsulates the key findings derived from the comprehensive evaluation of machine learning algorithms within the Intelligent Crop Recommendation System. This section provides a concise summary of the study's outcomes, their significance, and implications for the field of precision agriculture.

This research embarked on the journey of assessing the efficacy of machine learning algorithms in developing an Intelligent Crop Recommendation System. By leveraging data-driven insights, the study aimed to assist farmers in making informed decisions regarding crop selection. The significance of this study lies in its contribution to bridging the gap between theoretical potential and practical implementation of such systems in real-world agricultural contexts.

**Further Scope:**

While this research significantly advances the understanding of machine learning in crop recommendation, several avenues for further exploration emerge. Future research can delve into integrating real-time weather data, remote sensing technologies, and Internet of Things (IoT) devices to create more dynamic and adaptable recommendation systems. Additionally, an exploration of the system's implementation in different geographical regions could shed light on its universal applicability.

**REFERENCES**

[1] Priya P, Muthaiah U, Balamurugan M (2020) Predicting yield of the crop using machine learning algorithm. Int J Eng Sci Res Technol 7(1):1–7

[2] Yan M et al (2018) Research on precision management of farming season based on big data. EURASIP J Wireless Commun Networking 2018(1):139-143

[3] Rajak RK et al (2017) Crop recommendation system to maximize crop yield using machine learning technique. Int Res J Eng Technol 4(12):950–953

[4] A. Manjula and G. Narsimha, ”XCYPF”: A flexible and extensible framework for agricultural Crop Yield Prediction,” 2018 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), 2018, pp. 1-5.

[5] S. Babu, ”A software model for precision agriculture for small and marginal farmers,” 2019 IEEE Global Humanitarian Technology Conference:South Asia Satellite (GHTC-SAS), 2019, pp. 352- 355.

[6] U.Gupta, ”Comparison of Self organizing maps and Sammon’s mapping on agricultural datasets for precision agriculture,” 2018 International Conference on Innovations in Information, Embed- ded and Communication Systems (ICIIECS), 2018, pp. 1-5.

[7] A. J. Challinor, J. M. Slingo, T. R. Wheeler, & F. J. Doblas-Reyes, “Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles”, Tellus A: Dynamic Meteorology and Oceanography, vol. 57, no. 3, pp. 498-512, January 2005.

[8] A. Ceglar, and L. Kajfež-Bogataj, "Simulation of maize yield in current and changed climatic conditions: addressing modelling uncertainties and the importance of bias correction in climate model simulations", European Journal of agronomy, vol. 37, no. 1, pp. 83- 95, February 2012.

[9] (Open Source Article) “A complete guide to the random forest algorithm”, Built In, Available: https://builtin.com/datascience/random-forest-algorithm

[10] (Open Source Article) “Understanding Gradient Boosting Machines”, Available: https://towardsdatasecience.com/understanding-gradientboosting-mac hines-9be756fe76ab

[11] (Open Source Article) “An intuitive guide to Gaussian processes”, Medium, Available: https://towardsdatascience.com/an-intuitiveguide-to-gaussian-processes-ec2f0b45c71d

[12] P. Vidyullatha, & D. R. Rao, “Machine learning techniques on multidimensional curve fitting data based on R-square and chi-square methods”, International Journal of Electrical and ComputerEngineering, vol. 6, vo. 3, pp. 974-979, June 2016.