Intelligent Systems for Crop Recommendation using Machine Learning

SREYA CHUNDURI, HARRY RAJ, NARENDRA V. G.
Computer Science and Engineering, Manipal Institute of Technology,
Manipal Academy of Higher Education,
Manipal, Udupi, Karnataka, 576104,
INDIA

Abstract: - Given the soil and climate, information is of utmost importance in predicting which crop is best suited. Crops can now be grown with higher precision by analyzing data regarding temperature, humidity, soil conditions, and the chemical makeup of the soil, all of which impact crop growth. This is one facet of Precision Agriculture. Precision agriculture is a contemporary farming approach that uses scientific findings on the types, properties, and yields of soil. It guides farmers in selecting the most suitable crops tailored to their specific site conditions, reducing the chance of making unsuitable crop selections and ultimately helping raise overall productivity. The proposed work offers a web application that assists in classifying 22 crops based on various soil and environmental factors using two algorithms: SVM and Decision Trees. It analyzes the classifiers' accuracy using two performance metrics: the confusion matrix and the accuracy score. Farmers are better able to decide on the farming strategy they wish to use after utilizing the application.

Key-Words: - Intelligent system**,** Crop Recommendation, Support Vector Machine, Decision Tree, Machine Learning, Precision Agriculture.

Received: March 17, 2024. Revised: November 7, 2024. Accepted: December 4, 2024.Published: January 10, 2024.

1 Introduction

Agriculture has been a pivotal sector in India's economy for many years, involving more than half of the population. Given the diverse geographical features and varying climate conditions across different regions of India, it's vital to identify the most suitable crops for cultivation according to specific environmental conditions.

In 2021, climate change led to the loss of 5 million hectares of crops in India. To tackle this issue, we are employing various Machine Learning techniques to determine the most appropriate crops for cultivation. By developing an accurate and scalable algorithm, we aim to significantly aid numerous people across the country by improving crop yield and quality.

1.1 Problem Formulation

Our approach focuses on supervised Machine Learning methods, as the availability of labeled training data is crucial for achieving accurate results. In our study, we plan to offer recommendations for the best crops to cultivate, considering major factors such as soil nitrogen, pH level, phosphorus, potassium, temperature, humidity, and rainfall. Our objective is to use the latest advancements in Machine Learning to support the most disadvantaged segments of the agricultural sector.

The dataset we used was picked up from the Indian Chamber of Food and Agriculture, [1]. The first 5 rows of the dataset are followed (Figure 1).

	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

Fig. 1: Dataset attributes and values

2 Related Work

The paper [2] introduced an Efficient Crop Yield Recommendation System for Digital Farming, focused on aiding Indian farm owners in cultivating the right crop based on soil necessities. The project utilized a dataset incorporating various parameters such as N, P, K content, and pH level, along with the location data. Employing the SVM algorithm for classification, the system predicts suitable crops for the given soil conditions and provides information on nutrient deficiencies. The outcome is a user-friendly system that assists farmers, particularly novices, in making informed decisions, ultimately enhancing crop yield and farmer income. The paper [3] explores the application of various machine learning algorithms—DT, KNN, and RFs—to predict the most suitable crop. The dataset includes soil and environmental conditions, encompassing Nitrogen content, Potassium content, Phosphorus content, pH level, rainfall, temperature, and humidity. The project unfolds in three stages: pre-processing the data, applying Feature Selection, and finally, employing ML Algorithms on the modified dataset. Additionally, performance metrics like precision, accuracy score, recall, and F1 score are utilized for algorithm comparison. Consequently, the NBC emerges as the most accurate algorithm.

In their research, the authors of [4] explore a variety of techniques, including NB (Multinomial), SVM, and unsupervised machine learning algorithms like K-Means Clustering, to classify crops based on factors like temperature, humidity, pH level, sunlight, and moisture. The system achieves an impressive overall accuracy of over 92%, which can be further improved to 95% with prolonged usage. Meanwhile, the authors of [5] emphasize the significance of "Precision Agriculture" and propose an ensemble model recommendation system that uses a majority voting technique. This system recommends crops based on various parameters, including Depth, Soil Texture, acid/pH levels, Soil Colour, Permeability, Drainage, Water holding, and Erosion, across different sites. 10 crops were considered in this research. This model performed with an accuracy of 88%.

In a recent study [6], researchers aim to help farmers choose the best crops for their specific environmental conditions. They use machine learning techniques like Decision Trees, Support Vector Machines, Logistic Regression, and Gaussian Naive Bayes. By considering factors related to soil and weather, their model achieves a high prediction accuracy of 99.3% for twenty-two different crops. Another study [7] presents the "Crop Selection Method (CSM)," a novel approach to simplifying crop selection. CSM's main goal is to optimize crop yield throughout the season, boosting the nation's economic growth. This method tackles the complexities of crop selection by factoring in predictive yield rates based on weather, soil type, water density, and crop type. It also considers crucial details like crop varieties, ideal sowing times, planting durations, and expected seasonal yields. By strategically planning a sequence of crops, CSM ensures maximum daily production over the season, aligning with the goal of sustainable economic development.

In the paper [8], the main goal is to develop a crop recommendation system that enhances growth rates by analyzing soil characteristics like phosphorus (P), nitrogen (N), magnesium (Mg), and potassium (K). The dataset also includes soil pH, which influences nutrient availability, alongside external factors such as temperature, humidity, and precipitation that significantly affect the accuracy of crop recommendations. The study concludes that Random Forest outperforms Support Vector Machines in soil classification. The paper [9] delves into “Precision Agriculture” with a focus on “crop recommender systems.” It outlines various methodologies, including K-Nearest Neighbors (KNN), similarity-based models, ensemble-based models, and neural networks (NNs). These algorithms take into account external factors such as meteorological data, temperature, soil color, and texture to provide optimal crop recommendations. The paper evaluates the performance of four widely-used models—Random Forest (RF), Logistic Regression, Naive Bayes (NB), and Decision Trees (DT)—in classifying crops into four seasonal categories: Rabi, Kharif, Whole Year, and Summer. The performance of these models is assessed using datasets from Uttar Pradesh (UP) and Karnataka, two Indian states, employing metrics such as Relative Absolute Error (RAE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), accuracy, and Root Relative Squared Error (RRSE).

In a recent study [10], researchers used a Support Vector Machine (SVM) to develop a crop recommendation system for farmers. The SVM was utilized to classify soil parameters and identify the best crop for the given conditions. The metrics considered included N, Phosphorus, K, Moisture level, rainfall, Acidity, and temperature for various soil types. Non-numeric data values were converted to numeric values using Label Encoding. The data was pre-processed and then divided int the ratio of 80:20 into training and testing datasets respectively. The training set contained 1726 rows and 8 columns, while the test set had 440 rows and 7 columns. The Confusion Matrix was used for evaluation, with a total of 333 correct predictions out of 341, resulting in an accuracy of nearly 98%. Another paper [11] discussed multiple algorithms, including SVM, ANN, RF, MultivariateLR, and KNN, for predicting crop yield. The paper also delved into techniques such as Artificial NNs, Fuzzy Networks, and different data mining techniques and their advantages. The accuracy obtained from these methods ranged from 76% to 90%, with an average accuracy of 82%. Among all the algorithms used, Random Forest had the most promising results with an accuracy of 95%.

3 Methodology

Figure 2 illustrates the general outline of the carried research work. As mentioned, the dataset consists of 22 crops grown under different soil and environmental conditions. The parameters they vary on are Nitrogen levels, Potassium levels, Phosphorus levels, and pH levels of the soil along with the Precipitation, Humidity, and Temperature. To classify the Crop based on these values, we employed the SVM and Decision Tree Classifiers. Although different classifiers, we have used a similar approach for both.

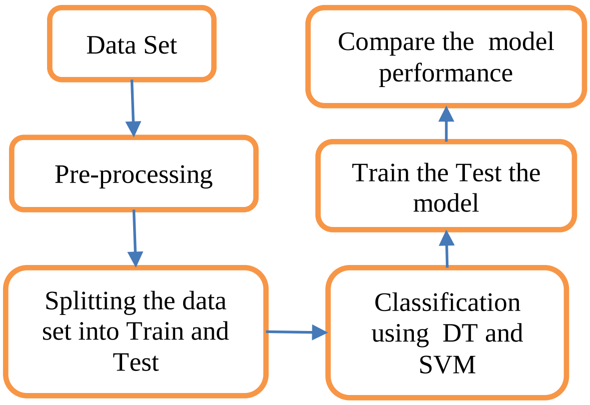


Fig. 2: General Outline of Program

3.1 Decision Tree

The Decision Tree is a versatile supervised learning approach that can be employed in both regression and classification situations. With its distinctive tree-like configuration, this classifier features internal nodes that represent the dataset characteristics, branches that indicate decision rules, and leaf nodes that contain the outcome. Within the Decision Tree, two critical node types are present - Decision Nodes, which play a crucial role in decision-making by offering multiple branches, and Leaf Nodes, which indicate the final result without any further branching.

In this project, the Decision Tree is used as a Classifier to map certain parameters to a particular crop (Classification task). The following steps were used:

1. Load the dataset using pandas
2. Clean the dataset
3. Encode the categorical target variable using Label Encode
4. Splitting the dataset into Train and Test sets using sklearn library's train_test_split
5. Creating an instance of the model using the sklearn library's DecisionTreeClassifier() [Default – Gini Index]
6. Fit the training data to the model
7. Predict the test set
8. Use performance metrics – accuracy_score and confusion_matrix to evaluate the performance of the classifier.

We get a Decision Tree with a maximum depth of 14 (Figure 3). The accuracy_score comes to be 98.64% and confusion_matrix (Figure 4) gets only 6 incorrect classifications.

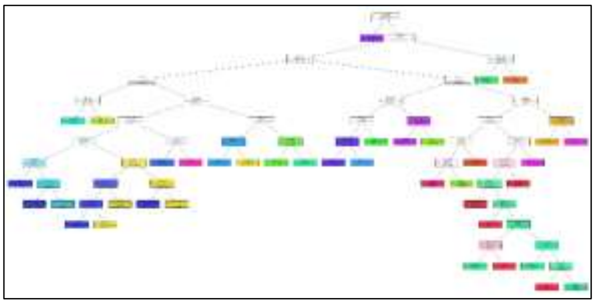
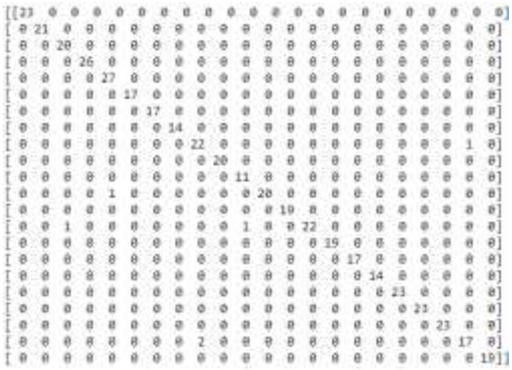


Fig. 3: Decision Tree Plot



```
[[23  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 21  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0 20  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0 26  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0 27  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 17  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 17  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0 14  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 22  0  0  0  0  0  0  0  0  0  0  0  1  0]
 [ 0  0  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0 11  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  1  0  0  0  0  0  0 20  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0 19  0  0  0  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0  1  0  0 22  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0 19  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 17  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 14  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 23  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 21  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 23  0  0]
 [ 0  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0 17  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 19]]
```

Fig. 4: Confusion Matrix

3.2 Support Vector Machines

The Support Vector Machine (SVM) algorithm is a powerful tool for supervised learning that can handle both classification and regression tasks. Its primary focus in classification is to create an optimal decision boundary, or hyperplane, that accurately divides an n-dimensional space into distinct classes, making it easy to categorize new data points. The key to this boundary lies in identifying critical data points, known as support vectors, which are located at the farthest extremes. These vectors play a crucial role in determining the hyperplane, hence the name Support Vector Machine.

In this project, SVM is used as a Classifier to map certain parameters to a particular crop (Classification task). 3 kernels were used – linear,

polynomial, and Radial Basis Function. The following steps were used:
1. Load the dataset using pandas
2. Clean the dataset
3. Encode the categorical target variable using Label Encode
4. Splitting the dataset into Train and Test sets using sklearn library's train_test_split
5. Creating an instance of the model using the sklearn library's SVC() – Here the kernel used can be linear (Figure 4a), polynomial (Figure 4b), and Radial Basis Function (Figure 4c).
6. Fit the training data to the model
7. Predict the test set
8. Use performance metrics – accuracy_score and confusion_matrix (Figure 5) to evaluate the performance of the classifier.

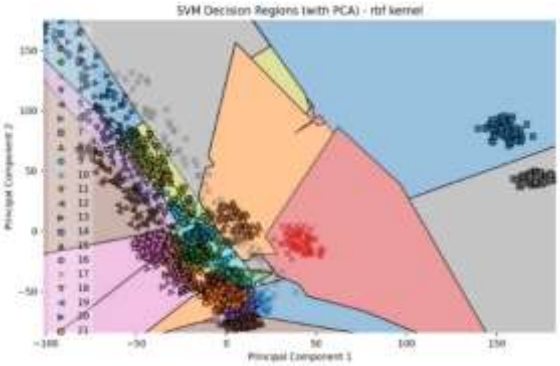


Fig. 4a: SVM – RBF Kernel Plot

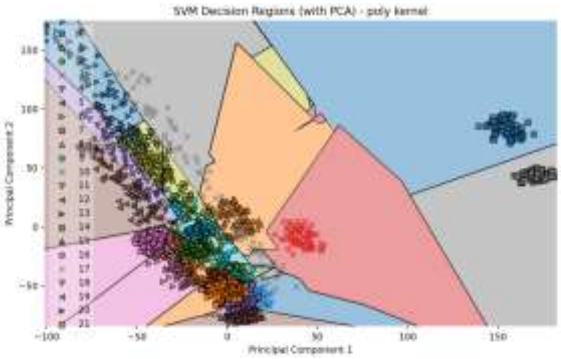


Fig. 4b: SVM – Polynomial Kernel Plot

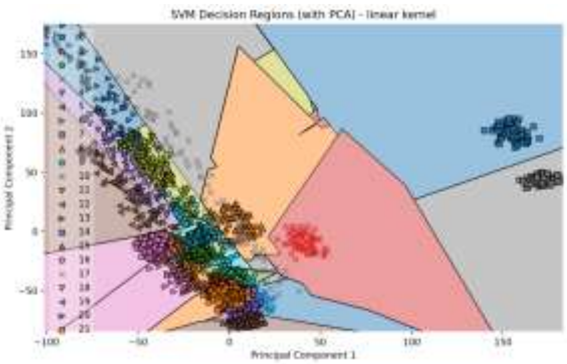
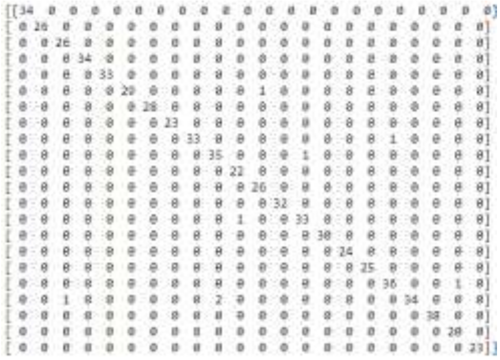


Fig. 4c: SVM – Linear Kernel Plot


```
[[34  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 26  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0 26  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0 34  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0 33  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 20  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 28  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0 23  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 33  0  0  0  0  0  0  0  0  1  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0 35  0  0  0  1  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0 22  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0 26  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0 32  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  1  0  0 33  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 24  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 25  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 36  0  0  1  0]
 [ 0  0  1  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0 34  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 38  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 28  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 23]]
```
Fig. 5: Confusion Matrix for SVM – Polynomial Kernel

4 Results and Discussion

We have developed the web application. The following Figure 6a, Figure 6b and Figure 6c illustrate the user-friendly web application used to take user input and predict the most suitable crop to grow in terms of the soil and environmental conditions.

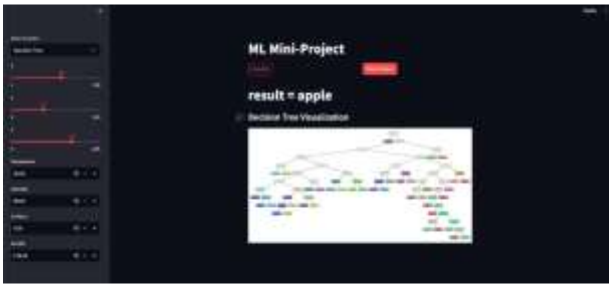


Fig. 6a: Web application

Fig. 6b: Different parameters setting

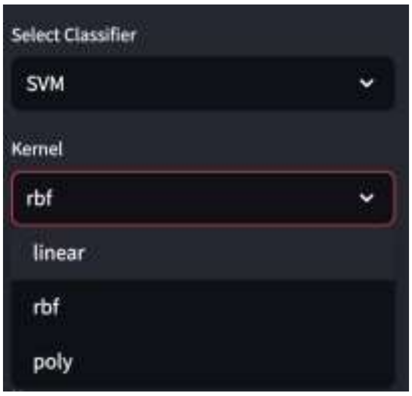


Fig. 6c: Different Classifiers Selection

On comparing the accuracy score and the classification report for Decision Trees and SVM. Table 1 illustrates the obtained accuracy comparison of different classifiers.

Table 1. Decision Tree vs SVM

Algorithm	Accuracy	Precision	Recall	F1-Score
Decision Tree	98.64%	0.99	0.99	0.99
SVM(linear)	97.87%	0.98	0.98	0.98
SVM(poly)	97.57%	0.98	0.98	0.98
SVM(rbf)	96.36%	0.97	0.96	0.96

After careful analysis of the chosen dataset and parameters, it was determined that the Decision Tree method yielded the most favorable results with an accuracy rate of over 98% in the majority of cases. Further examination revealed that rainfall had the lowest Gini index among the parameters. However, a Support Vector Machine with a 'rbf' kernel proved to be the least effective technique. In contrast, SVM utilizing 'linear' and 'poly' kernels exhibited similar outcomes to the Decision Tree method in most cases.

We have also compared with previous work our developed classifiers were given good results. We found that, no web application developed for prediction by considering all the parameters.

5 Conclusion

Precision Agriculture is the future of agriculture. It can be used to improve crop yield and productivity. Although the current dataset comprises only 22 labels, India boasts a plethora of crops that are not yet included. Expanding the dataset to include more crops would benefit a wider audience and have a greater impact across the country. While a web application is in existence to help users classify the best crop based on certain parameters, the creation of a user-friendly mobile application would be a practical and convenient addition for frequent use.

Acknowledgement:
The Department of Computer Science and Engineering Manipal Institute of Technology Manipal Academy of Higher Education Manipal India-576104 is acknowledged by the authors for providing the facilities and resources that were required.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the authors used QuillBot/Grammatically reconstruct the sentences, Grammarly/Grammar check in order to check grammar as well as reconstruct the sentence. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

[1] Indian Council of Food and Agriculture, "Official Website," ICFA, [Online]. https://www.icfa.org.in/ (Accessed Date: July 26, 2024).

[2] Dr.G. Suresh, Dr.A. Senthil Kumar, Dr.S. Lekashri, Dr.R. Manikandan, Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming, International Journal of Modern Agriculture, vol. 10, no. 1, 2021, pp. 906-914, [Online]. https://www.modern-journals.com/index.php/ijma/article/view/688 (Accessed Date: July 26, 2024).

[3] Liza Kansal, Anoushka Pandey, Sanidhya Madhav Shukla, Parneeta Dhaliwal, Review of Machine Learning Techniques for Crop Recommendation, *IC3-2023: Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*, Noida India, August 3-5, 2023, pp.443-449. https://doi.org/10.1145/3607947.3608045.

[4] Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H., W.J.M. Nanayakkara, Dimantha M.A.C, Pabasara M.G.P, Crop Recommendation System, *International Journal of Computer Applications, vol. 175(22), pp. 22-25*, 2020. https://doi.org/10.5120/ijca2020920723.

[5] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, J. Nisha,

Crop Recommendation System for Precision Agriculture, *2016 Eighth International Conference on Advanced Computing (ICoAC)*, Chennai, India pp. 32-36. https://doi.org/10.1109/ICoAC.2017.7951740.

[6] Maaz Patel, Anagha Rane, Crop Recommendation System, *Engrxiv*, Vol.1, 2023, https://doi.org/10.31224/2959.

[7] R. Kumar, M. P. Singh, P. Kumar, J. P. Singh, Crop Selection Method to Maximize Crop Yield Rate Using Machine Learning Technique, *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, Avadi, India, pp. 138-145. https://doi.org/10.1109/ICSTM.2015.7225403

[8] R. Ajoodha, T.O. Mufamadi, Crop Recommendation Using Machine Learning Algorithms and Soil Attributes Data, *Proceedings of 3rd International Conference on Artificial Intelligence: Advances and Applications (ICAIAA) 2022*, Springer. https://doi.org/10.1007/978-981-19-7041-2_3.

[9] R. Katarya, A. Raturi, A. Mehndiratta, A. Thapper, Impact of Machine Learning Techniques in Precision Agriculture, *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, Jaipur, India 2020, pp. 1-6. https://doi.org/10.1109/ICETCE48199.2020.9091741.

[10] D. Modi, A. V. Sutagundar, V. Yalavigi and A. Aravatagimath, "Crop Recommendation Using Machine Learning Algorithm," *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 2021, pp. 1-5, https://doi.org/10.1109/ISCON52037.2021.9702392.

[11] S. M. Pande, P. K. Ramesh, A. Anmol, B. R. Aishwarya, K. Rohilla And K. Shaurya, Crop Recommender System Using Machine Learning Approach, *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2021, pp. 1066-1071, https://doi.org/10.1109/ICCMC51019.2021.9418351.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Each author made an equal contribution to the current each author made an equal contribution to the current study, from the conceptualization of the problem to the final conclusions and resolution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest:

The authors disclose no conflicts of interest.