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.2024.62830

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

SmartCrop: A Data-Driven Crop Recommendation System for Precision Agriculture

Huma Verma[1], Akanksha[2], Meeta Chaudhry[3]
[1, 2]*Department of CSIT Kiet Group of Institutions, Delhi NCR, India*
Associate Professor, Department of CSIT, Kiet Group of institutions Delhi NCR, India

Abstract: ***The Indian economy contributes to agriculture. Most farmers in India rely on their instincts when deciding to plant crops during the specific period of the year. They don't know that crop yield is environmentally sensitive and depends on soil and weather. A farmer's ignorant decisions can have consequences for the health of the soil, as well as psychological and financial harm to the actual farmer. Using a machine learning model will be beneficial solve this problem. The data employed in the article was produced by combining records for fertilizer, climate and rainfall for India. Models for machine learning will be applied to the data to obtain the most accurate models that will recommend the right crops for the agricultural field. Indian farmers would be able to choose their crops with more knowledge thanks to these recommendations. Factors such as where the farm is located, planting season, soil characteristics, and climate will all be considered in the suggestions.***
Keywords: Crop suggestion, Nitrogen- Phosphorus- Potassium (NPK), Humidity, Rainfall, pH, Machine Learning (ML), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB).

I. INTRODUCTION

The Indian economy's cornerstone is agriculture, which generates 20% of GDP and employs more than 41.49% of the labor force. It covers more than 60% of the country's territory and feeds part of the population, especially in rural areas, where it is an important source of income for women. Although it has an important place in agriculture, it emerges due to inadequate understanding of important concepts such as traditional agriculture, climate, soil dynamics and good fertilizer use. This leads to lower expectations per worker, affecting economic growth. The lack of technological development further exacerbates these challenges, leading to a cycle of inadequate knowledge and development. Rural farmers therefore face crop failure due to poor understanding of climate change, soil degradation due to overuse of pesticides, and short sightedness. When it comes directly to income, economy and climate change. The resulting economic depression is when the 3.4% increase in agricultural gross value added (GVA) becomes admirable, the importance of all decisions regarding agriculture for the regional economy becomes clear. The wrong choice can lead to negative financial consequences, indicating the need for solutions that go beyond the farmer's knowledge. A solution has emerged at this important time for farmers and business: the sowing recommendation. Guided by parameters such as temperature, weather, humidity and soil parameters such as nitrogen, phosphorus, potassium and pH, the system is designed to guide farmers in decision making, reduce risk and improve crop selection. This article highlights the complexity of the challenges facing Indian agriculture, highlighting the need for advanced systems that enable farmers to make informed decisions and the complex web of changes affecting crop production and economic sustainability. A Crop suggestion system helps farmers choose the best crop, ensuring that the farmer makes an informed decision before planting. The technology can determine the temperature and humidity levels based on the user's location. With escalating food scarcityand population growth, it is critical to find answers to farmer concerns. Agriculture has remained and will continue to be an important industry. As a result, the system can be evolved with the necessary data.

The focus of machine learning is on algorithms, such as reinforcement learning, supervised learning, and unsupervised learning, each of which has pros and cons. The algorithm, which is taught under supervision, builds a mathematical model from a set of data that includes the inputs and the intended outputs. An method that uses unsupervised learning builds a mathematical model created from a set of data that just includes inputs and no specified output labels. Algorithms for semi-supervised learning build mathematical models from partial training data, in which some sample input is label- free.

There is no cost for taking images of the soil or performing difficult operations to obtain crop details, so farmers can concentrate on selecting the best crop. Farmers may readily understand the forms because the application merelyrequires them to provide values. As a result, no expensive equipment or technical knowledge is required to determine which crop is suitable for their farm. This solution can be combined with other agricultural problem solutions to increase farm productivity.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

The progress of agriculture has become evident in recent academic debates that have led to a review of past literature to understand the limitations and limits of agriculture for future development. Jain et al.[1].A crop selection strategy according to meteorological factors was proposed. Their method of predicting seasonal weather patterns and matching suitable crops accordingly is unique to the state of Telangana. The method uses RNN for weather prediction and a random forest distribution to determine the crop and also record the optimal planting time.

Waikar et al. [2] suggested a technique that predicts soil type and suggests crops that fit it using soil and rainfall data. Utilizing classification algorithms including SVM, ANN, Ada Boost, Naive Bayes, and Bagged Trees, the system determines which crops are appropriate for which types of ground.

Attaluri et al. [3] Establish a framework that prioritizes recommendations according to results. They used algorithms like random Logistic regression, artificial neural networks, and forests to evaluate performance standards from various indicators, taking into account factors such as planting and harvesting costs, seed and fertilizer costs, rainfall, crop demand and yield.

Kumar et al. [4] proposed recommendations that include product selection and pest analysis to provide pest management strategies. Their research assessed logistic regression, decision tree, and SVM algorithms and showed that SVM is a reliable method for crop predicting.

Tamsekar et al[5].studied GISbased crop selection prediction using tillage maps of Hadgaon and Belgaon for training and validation. They came to the conclusion that SVM with PCA was the best accurate model based on the CSPM model with and without PCA.

Doshi et al. [6] developed a smart farming network and applied multi- label algorithms for crop recommendation. The system combines the state- level precipitation forecast subsystem with the map view. A critical analysis of these studies clearly shows that most studies focus on a single parameter for crop cultivation (soil, weather or crop yield). Some are limited to certain crops or soil types. Given the heavy Dependence of India's workforce on agriculture complex combination of soil properties, climate and fertilizer use is required.

A holistic data set that includes these factors is required for good and diverse crop recommendations. Moreover, given the many situations of agricultural decision- making, rigorous testing of different learning models is also important to determine the most accurate model for product recommendations.

Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. [24]"Machine learning system based Crop recommendation system to optimize crop productivity in Ramtek region". The three parameters that this proposed system was built on were soil types, characteristics, and crop yield data gathering. The system then suggested to the farmer which crop would be best to cultivate considering on these aspects.

The machine learning methods used by this suggested system included random forest, CHAID, K- Nearest Neighbor, and Naïve Bayes. We can forecast a certain crop under particular condition meteorological conditions, as well as state and district values, by putting this suggested system into practice. Our proposed effort will help farmers pant the right seed based on soil requirements, which will increase national output.

Kulkarni, Nidhi H., G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery. [21] "Increasing Crop yield by using a Crop Recommendation System using Ensemblingtechnique". With great accuracy, this suggested method suggests the appropriate crop depending on the kind of soil and its properties, such as typical rainfall and surface temperature. The suggested system operated on a number of machine learning techniques, including Linear SVM, Random Forest, and Naive Bayes. The input soil dataset was categorized by this crop recommendation system into the recommended crop types, Kharif and Rabi. Applying the suggested system produced an accuracy result of 99.91%.

III. CROP RECOMMENDATION ANALYSIS

As shown in Figure 1, the envisioned system initiates a systematic process that will facilitate the cultivation of crops. He initially started by gathering important information to make recommendations about crops. This includes comprehensive assessment of soil parameters, nutrient content, climate nuances and temperature profiles. In this process, quality decisions were expanded to cover 22 different crops. Legumes, fruits, cash crops, and food crops are just a few of the agricultural productsthat can be included in the produced sample

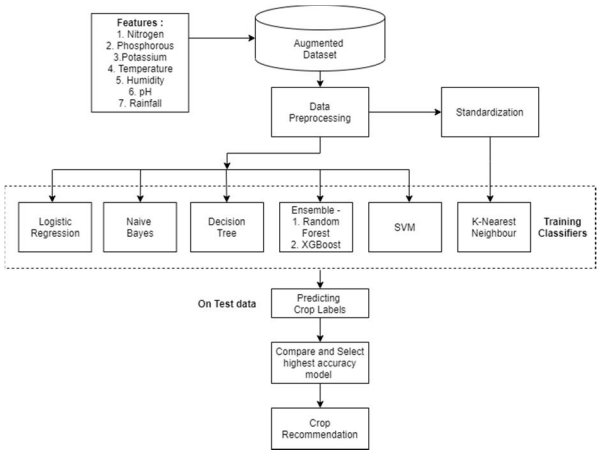


Figure 1. The system architecture for the suggested Crop- Recommendation system

thanks to this wise section. The utilization of machine learning classification models to determine the optimal recommendations is the crux of this approach. The system performs a segmentation process using carefully selected data based on relevant agricultural practices. The department allows training and rigorous testing of models to ensure their validity and reliability in predicting suitable crops based on specific farming conditions. Through the integration of big data and machine learning technology, the proposed system attempts to close the distance between traditional agricultural practices, usability and today's technological advances. It is hoped that by using information through understanding, farmers will be helped to know crop selection and make better decisions, leading to agricultural productivity and economic sustainability.

IV. METHODOLOGY

The following procedure can be used to construct the facility approval form [7, 8]:

1) *Advice:* Create a CSV file with the optimized data in it. It then needs to be pre-treated.
2) *Preprocessing:* It is done after checking the data set parameters. These were done by changing the mean and type of numerical and categorical data respectively. Standardize data to increase some of the models that our system uses for training more accurate.
3) *Data set Segmentation:* System first divides the data into training set and test set by determining the 80:20 ratio. Additionally, the distribution is done randomly using a function that generates the same random sequence of integers each time data is entered.
4) *Classifiers:* Insert training data into any machine learning classification model for training.
5) *Test all Classifiers:* Then test the training model using the test data. Contract estimates are obtained and then compared to actual harvest contracts. Each model's accuracy was assessed by calculating the test data yields.
6) *Most Precise Model:* Examine the precision of each model and choose the most accurate model that will provide the best contribution to the crop.

Attribute	Specifics
Nitrogen (N)	Ratio of nitrogen content
Rainfall (R)	Rainfall in mm
Phosphorus (P)	Ratio of Phosphorus content in soil
pH	pH value in soil
Humidity (H)	Relative humidity in %
Potassium (H)	Ratio of Potassium content in soil
Temperature (T)	Temperature in degree celsius

Table 1. Schema for datasets

Consider the use of data involving various agricultural activities to overcome previous limitations. The data used here was created by combining rainfall, climate and fertilizer datasets for India. These data include 2200 observations and 7 variables are listed in Table 1 [16]. Contains 22 crops, including various legumes, fruits, cash crops and food crops. Crops include wheat, corn, chickpeas, kidney beans, pigeon peas, broad beans, mung beans, black beans, lentils, pomegranates, bananas, mangoes, grapes, watermelons, melons, apples, oranges, papaya, coconut, cotton, jute and coffee.

V. CROP RECOMMENDATION THROUGH CLASSIFICATION

Precision agriculture uses a lot of machine learning algorithms to give farmers reliable results. There exist countless things to consider when planting plants. Farmers will lose some crops when they want to cultivate their l and. This could result in crop failure and losses to farmers. Therefore, Reliable machine learning models are employed by the perfect system. As just one of the 22 lists is provided, more than one list would be ideal. There is support for many things in this category, as well as testing of many machine learning algorithms such as logistic regression, naive Bayes, KNearest Neighbors (KNN), decision trees, random forests, and hybrid models like XGBoost. Label classification Vector machines are used under this framework [10, 12].

A. Logistic Regression

Only binary distribution issues can be solved with logistic regression. An expansion of logistic regression that can handle multivariate distribution issues is called polynomial logistic regression. To find the probability of each target, use the soft max function rather than the sigmoid function. These outcomes must be transformed to thermal coding and the cross entropy method is used during training and calculation of visual weight. Weights are use d to predict the list. [17] Here, the "lbfgs" solver uses polynomial logistic regression to find the losses.

B. Naive Bayes

For binary and multivariate classification class problems, naïve Bayes classification approaches are employed. It makes the assumption that the data points used as input for its catalogs are not mutually exclusive. Greater size of the base is advantageous. It predicts classroom material using recent and past occurrences and is according to the Bayes theorem [13,15]. The Gaussian Naive Bayes model is employed by the system sustains continuous data following a normal Gaussian distribution. Targets are determined using Z-scores.

C. K-Nearest Neighbors

The K-Nearest Neighbors. The (KNN) method utilizes a learning process that requires historical data patterns into account when predicting the new list. The "k" value is selected when it is determined that the accuracy of the model is maximum and the error is lowest. A fresh concept is categorized as soon as it is created. According to its proximity through distance measurements Euclidean distance, for example, Minkowski distance or the Manhattan distance. [6, 9] Because the algorithm is dependent on distance, the size of the space has an impact on it. For this reason, information is generated prior to the model is trained along with the k worth of 5 is chosen for the training prototype [8, 9].

D. Logistic Regression

Only binary distribution issues can be solved with logistic regression. An expansion of logistic regression that can handle multivariate distribution issues is called polynomial logistic regression. To find the probability of each target, use the soft max function rather than the sigmoid function. These outcomes must be transformed to thermal coding and the cross entropy method is used during training and calculation of visual weight. Weights are use d to predict the list. [17] Here, the "lbfgs" solver uses polynomial logistic regression to find the losses.

E. Decision Tree

Tree structures are employed in decision trees to express lists and attributes. [4] The last predictions are contained in this tree's leaves. The root node's attributes are contrasted with the properties of each new data additionally, a fresh choice is made. This procedure is repeated until a leaf containing the list of predicted classes is reached. To keep track of this process, use the "entropy" process, which has a maximum of 5. The maximum likelihood Index helps prevent over fitting to the tree and can identify when the accuracy rate is higher than the training accuracy for a given depth.

F. Support Vector Machines

Support Vector Machines (SVM) classify data and make a difference by detecting and finding a large plane of data points shown in a space of N dimensions. Using kernel functions, one can transform low- dimensional information into high- dimensional data. Data points are classified using this technique using the Gaussian radial basis function.

G. Gradient Boosting

With gradient descent, each new model is trained to minimize the loss function, such as mean squared error or cross-entropy of the preceding model. Gradient boosting is a potent boosting procedure that turns multiple weak learners into strong learners. The approach calculates the gradient of the loss function in relation to the current ensemble's predictions for each iteration, and then trains a new weak model to minimize this gradient. Next, the new model's predictions are included in the ensemble, and the procedure is continued until a stopping requirement is satisfied.

H. AdaBoost Regression

Adaptive Boosting, or AdaBoost, is a machine learning algorithm for ensembles that may be applied to a range of classification and regression applications. This supervised learning approach creates a strong learner by merging several weak or base learners (like decision trees) to classify data. In order for AdaBoost to function, the training dataset's instances are weighted according to how well prior classifications worked.

I. Random Forest

Rather than depending just on one decision tree, the random forest takes the prediction from each tree and predicts the final output based on the majority votes of predictions. Random forest is a classifier that contains multiple decision trees on different subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

J. Bagging

Bagging, alternatively referred to as Bootstrap aggregating, is an ensemble learning method that enhances machine learning algorithms' accuracy and performance. It lowers a prediction model's variance and handles bias-variance trade-offs. Bagging is used in decision tree methods, specifically for regression and classification models, to prevent overfitting of the data.

K. Extra Trees Classifier

Extra Trees Classifier is an ensemble learning method fundamentally based on decision trees. Extra Trees Classifier, like Random Forest, randomizes certain decisions and subsets of data to minimize over- learning from the data and overfitting.

VI. RESULT ANALYSIS

Providing training on various distributed machine learning methods of data. This training model is then evaluated. Consider the accuracy index, accuracy, memory, and F1 score. The most accurate next, a model is chosen to make recommendations for later crops.

Logitsic regression with accuracy	0.9636
Naive Bayes with accuracy	0.9954
Support Vector Machine with accuracy	0.9681
K-Nearest Neighbors with accuracy	0.9590
Decision Tree with accuracy	0.9818
Random forest with accuracy	0.9931
Bagging with accuracy	0.9886
AdaBoost Regression with accuracy	0.1409
Gradient Boosting with accuracy	0.9818
Extra Trees with accuracy	0.8977

Figure 2. A comparison of the accuracy of the models

After analyzing the accuracy obtained in Figure 2, From Figure 2, obtained separately for each example, it can be seen t hat the Of the following algorithms: Random Forest, XGBoost, Decision Tree, Naive Bayes, KNN, and SVM; the most precise is logistic regression. The model of Random forest achieved the test data's maximum accuracy at 99.55, XGBoost comes next at 99.3. Therefore, it can be considered that the design model is superior to the system's other classification models. In contrast to the systems examined in the literature review, the accuracy rate was expanded by 9% for Random Forest, 3% for SVM, additionally 2% for Naive Bayes. Ensemble was also used to test the system with unknown values technology and Table 2 displays a comparison of these outcomes. It is evident that the majority of the yield estimates made by the XGBoost and Random Forest classifiers are comparable.

VII. CONCLUSION

So far the facility approval process has been completed and put into action and is simple enough for Indian farmers to use. The system will assist farmers in producing informed choices regarding which crops to plant. This advice will be contingent upon various soil, environmental and weather conditions. The random forest's high level of accuracy integration method makes this system excellent for all practical needs. The prototype presented within this document able to expand in the future to include elements such as insecticide use recommendations, crop demand recommendations, second choice recommendations if many similar-type crops are planted in the surrounding field. Crop related economic factors Such as price, demand, and supply can be considered as the optimal pattern of the crop. This will provide a health forecast based notjust on land, weather and environmental elements, but furthermore on economics.

REFERENCES

[1] Machine learning convergence for weather-based crop selection, S. Jain and D. Ramesh, 20202020 IEEE SCEECS (International Students' Conference on Electrical, Electronics, and Computer Science)

[2] Mahesh S. Shinde, Priya P. Rajput, Ashlesha A. Ghute, Sheetal Y. Thorat, and Vrushali C. Waikar (2020). "Crop Prediction using Machine Learning with Classifier Ensembling based on Soil Classification." Engineering and Technology International Research Journal (IRJET), Volume: 07Problem: May 5, 2020

[3] "Crop plantation recommendation using feature extraction and machine learning techniques," Attaluri, S.S., Batcha, N.K., and Mafas, R. (2020). Applied Technology and Innovation Journal, 4(4), 1 (e-ISSN:2600-7304).

[4] Sarkar, S., Pradhan, C., and Kumar, A. (2019). "Agricultural Crop Identification and Pest Control Technique Recommendation System." The 2019 International Conference on Signal Processing and Communication (ICCSP).doi:10.1109/iccsp.8698099 (2019).

[5] Tamsekar, P., Bhalchandra, P., Kambarde, K., Deshmukh, N., Kulkarni, G., & Husen, S. (2019). "Evaluative Comparison of Supervised Machine Learning Techniques for GIS-Oriented Crop Selection Forecasting Model." Sustainability of Computing and Networks, 309-314. doi:10.1007/978-981-13-7150-9 33

[6] Doshi, Z., Shah, N. (2018), Nadkarni, S., and Agrawal, R."Agro Consultant: Machine Learning Algorithms-Based Intelligent Crop Recommendation System."2018 saw the Fourth International Conference on Computing, Automation, and Communication Control (ICCUBEA).The doi is the same as iccubea.2018.8697349

[7] Mohidul Islam, S. M., Chandra Mitra, K., and Rahman, S. A. Z. (2018). "Soil Series-Based Crop Suggestion and Machine Learning-Based Classification" 21st International Conference of Information Technology and Computers, 2018 (ICCIT). 2018; doi:10.1109/iccitechn.

[8] Kulkarni, N. H., Cauvery, N. K. (2018), Srinivasan, G. N., and Sagar, B. M. "Using an Ensembling Technique to Improve Crop Productivity through a Crop Recommendation System." 2018 will be the third international conference on information technology and computational systems for sustainable solutions, or CSETS.10.1109/csitss.2018.8768790

[9] Munishwara M.S., Gracias Abigail Angela, C. G. Neha, Preethi, Akarsh S. (2020). "Using Machine Learning for Soil Fertility Analysis and Crop Prediction." Volume 9, Issue 6, April 2020, International Journal of Innovative Technology and Exploring Engineering (IJITEE)

[10] Guo, L., Di, L., Zhang, C., and Lin, L. (2019). "Annual crop planting in the U.S. Corn Belt predicted by machine learning using historical crop planting maps." Agriculture's Use of Computers and Electronics,166,104989.1016/j.compag.2019 is the doi.

[11] Panchamurthi, S. (2019) [11]."Machine learning for soil analysis and crop prediction that is suitable for agriculture."Journal of Res. Appl. Sci. Eng. Technol.

[12] A Naïve Bayes MapReduce Precision Agricultural Model, R. Priya, D. Ramesh, and E. Khosla, 2018 International Conference on Advances in Computing, Communications, and Informatics (ICACCI),2018

[13] Shreedhara, K. S., and Akshatha, K. R. (2018). "Using precision agriculture to implement machine learning algorithms for crop recommendation." Journal of Engineering, Science, and Management Research International (IJRESM), 1(6), 58–60.

[14] Nie Y, De Santis L, Carratù M, O'Nils M, Sommella P, Lundgren J. Deep melanoma classification with k-fold cross- validation for process optimization. In international symposium on medical measurements and applications (MeMeA) 2020 (pp. 1-6). IEEE.

[15] Belete DM, Huchaiah MD. Grid search in hyper parameter optimization of machine learning models for prediction of HIV/AIDS test results. International Journal of Computers and Applications. 2022; 44(9):875-86.

[16] PANDE, S. M., RAMESH, P. K., ANMOL, A., Aishwarya, B. R., ROHILLA, K., & SHAURYA, K. (2021, April). Crop recommender system using machine learning approach. (pp. 1066-1071),IEEE Xplore.

[17] Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].

[18] Eswari, K. E., & Vinitha, L. (2018). Crop yield prediction in Tamil Nadu using Baysian network. International Journal of Intellectual Advancements and Research in Engineering Computations, 6(2), 1571-1576.

[19] Sriram Rakshith.K, Dr.Deepak.G, Rajesh M, Sudharshan K S, Vasanth S, Harish Kumar N, "A Survey on Crop Prediction using Machine Learning Approach", In International Journal for Research in Applied Science & Engineering Technology (IJRASET), April 2019, pp(3231- 3234)

[20] Bandara, P., Weerasooriya, T., Ruchirawya, T., Nanayakkara, W., Dimantha, M., & Pabasara, M. (2020). Crop recommendation system. International Journal of Computer Applications, 975, 8887.

[21] Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].

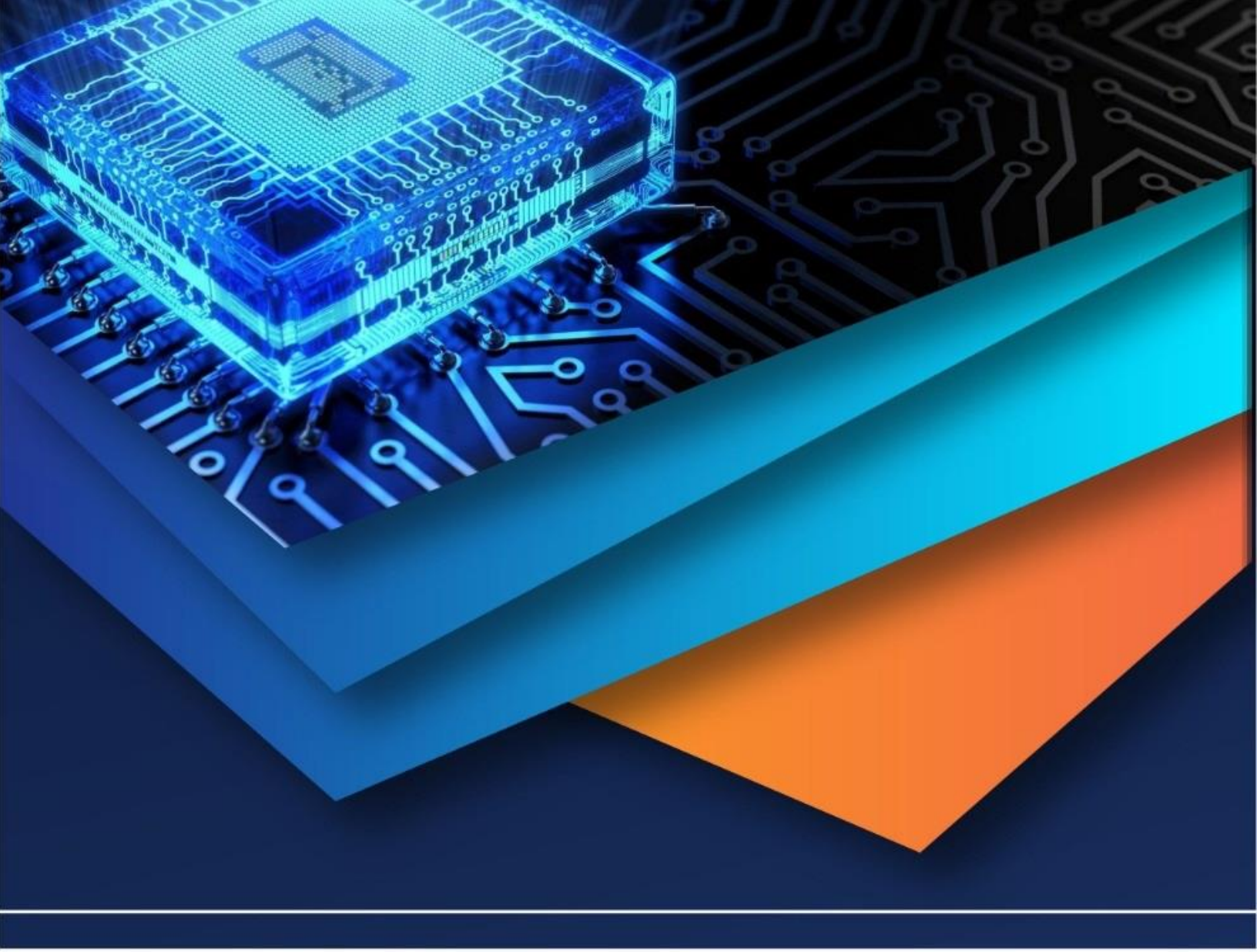
[22] Paja, K. Pancerz, and P. Grochowalski, ''Generational feature elimination and some other ranking feature selection ,'' in Advances in Feature Selection for Data and Pattern Recognition, vol. 138. Cham, Switzerland: Springer, 2018, pp. 97–112.

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

[23] Reddy, D. A., Dadore, B., & Watekar, A. (2019). Crop recommendation system to maximize crop yield in ramtek region using machine learning. International Journal of Scientific Research in Science and Technology, 6(1), 485-489

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

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



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)