iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47793

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 10 Issue XII Dec 2022- Available at www.ijraset.com

Crop, Fertilizer, & Irrigation Recommendation using Machine Learning Techniques

Avi Ajmera[1], Mudit Bhandari[2], Harshit Kumar Jain[3], Supriya Agarwal[4]
[1, 2, 3]*Department of Computer EngineeringMukesh Patel School of Technology Management & Engineering Mumbai, India*
[4]*Assistant Professor, Department of Computer Engineering, Mukesh Patel School of Technology Management & Engineering, Mumbai, India*

Abstract: ***Agriculture is the majority source of income for many people not just in the Indian subcontinent but around the world and hence forms the backbone of the economy. Present-day difficulties like unpredictability in weather conditions, water scarcity, and volatility due to demand- supply fluctuations create the need for the farmer to be equipped with modern day techniques. More specifically, topics like less yield of crops due to unpredictable climate, faulty irrigation resources, and soil fertility level depletions need to be communicated. Hence there is a requirement to modify the abundant agriculture data into modern day technologies and make them conveniently accessible to farmers. A technique that can be implemented in crop yield prediction is Machine learning. Numerous machine learning techniques like regression, clustering, classification and prediction can be employed in crop yield forecasting. Algorithms like Naïve Bayes, support vector machines, decision trees, linear and logistic regression, and artificial neural networks can be employed in the prediction. The wide array of available algorithms poses a selection dilemma with reference to the selected crop. The purpose of this study is to investigate how different machine learning algorithms may be used to forecast agricultural production and present an approach in the context of big data computing for crop yield prediction and fertilizer recommendation using machine learning techniques.***
Keywords: Random Forest Classifier, SVM, XG-Boost, Geographical Attributes, Crop Management, Supervised Learning

I. INTRODUCTION

Agriculture, which incorporates agricultural production, is indeed a vital role in the Indian economy. Crops may be both food and commercial crops. Crop production is substantially impacted by several circumstances such as soil type, nutrient content, fertilizer type, and irrigation supply. As a result, accurate yield prediction is a critical issue that must be addressed. Early production prediction would allow farmers to take preventative measures to increase output. Early forecasting is feasible by collecting prior experience of farmers, weather conditions, and other influencing elements and storing it in a vast database.

The typical input variables are as follows Temperature, air humidity, pressure, wind speed, wind gust, wind direction By the application of machine learning algorithms like logistic regression, K-nearest neighbour, decision tree, naïve bayes, random forest, XG-Boost early predictions about the yield can be made. Forecasting agricultural output prior to harvest is a critical topic in agriculture since crop production fluctuations affect international trade, food availability, and worldwide market pricing. Crop production forecasting at the within-field level has become more common in recent years. Soil and irrigation conditions have the greatest impact on crop output. If those forecasts are more exact, farmers may be warned well in advance, allowing the significant loss to be avoided and contributing to economic growth. In the event of a severe emergency, the forecast will also help farmers make decisions such as choosing alternate crops or discarding a crop at an early stage. Furthermore, crop output prediction can help farmers have a better understanding of seasonal crop production and scheduling. Thus, crop output must be modelled and predicted prior to cultivation for effective crop management and desired results. Because crop yield and crop-influencing factors have a non-linear connection, machine learning approaches may be effective for yield prediction.

II. LITERATURE REVIEW

A technique called the Crop Selection Method (CSM) was suggested to address the crop selection problem, optimize the net yield rate of crop over the season, and achieve maximum economic growth of the country. Plantation days and anticipated yield rate are used as inputs to discover a sequence of crops that will produce the most per day over the course of a whole season, according to the program [1]. To estimate crop output based on input circumstances, the proposed system uses machine learning and prediction algorithms such as Multiple Linear Regression. The suggested method is put to the test using data gathered from 1997 to 2014. Data is gathered from all 640 districts and 5924 sub districts across India's states. The overall accuracy achieved is 78% using polynomial regression [2].

©IJRASET: All Rights are Reserved

The method proposed in the paper is a unique way to determine the best nutrients for increasing yield output while retaining soil fertility using time-series soil nutrients enhanced the population initialization technique. The suggested technique aids in the reduction of search space as well as the elimination of missing local optimization parameters. The model employs a genetic algorithm to analyze the data and make recommendations for remoteenvironment optimization. This study suggested animproved genetic algorithm (IGA) for recommending the best nutrient settings for various crops [3].

This paper recommended the suitable classifier to classify African's countries and their crops in the harvested area using algorithms such as Naive Bayes classification, Decision Table classification, PART classification, J48 classification and Lazy classifier IBK. The dataset used,contains 163 crops which are produced in 59 African countries. The outcome demonstrated that the J48 algorithm has performed the best [4].

The spatial fluctuation of individual soil nutrients inKerala's Wayanad district was simulated using geostatisticalapproaches in this research. The soil data included pH, Organic Carbon (OC), Phosphorus (P), Potassium (K), Zinc (Zn), Iron (Fe), Copper (Cu), Manganese (Mn), Boron (B), and Sulphur (S) values, as well as the WGS84 coordinatesof sample sites. There was a total of 19275 soil sample sites.With nugget values greater than 0, pH, Boron, and Copper show a smooth slope. With a range of more than 10 kilometers, pH and Boron show modest autocorrelation. Copper and Sulphur have a strong autocorrelation with a range of less than 1 kilometer [5].

The proposed effort was based on historical agriculturaloutput, and crops can be recommended to farmers using a recommendation system. Farmers will be provided recommendations based on the crop output season. A totalof 1,20,000 entries from the Tamil Nadu Agriculture Datasetwere collected. The user was provided recommendationsbased on crop productivity and the season in which thecrops were grown [6].

The researchers created a fuzzy logic-based software that provides an appropriate quantity of fertilizers to soildepending on nutrient levels. This research uses the triangle membership function. The fuzzy system calculates the appropriate amount of fertilizer based on the NPK and season values. The outcome demonstrates that the fuzzylogic system was effectively designed and simulated to provide appropriate fertilizer recommendations [7].

This model incorporates Artificial Neural Networks becausethey can provide extremely precise results and can evaluate and handle a huge quantity of data on soil moisture, pH, andtemperature at various locations quickly, which is criticalfor any systems that need to produce immediate answers [8].

The methodologies used in the research were time-series analysis, ANN, Multi-Linear Regression, Moving Average. The application is used to control the irrigation system. Inputs are supplied dependent on the crop and its stage. If there was an intruder or if the temperature rose, the farmer would receive a notification. Romyan's technique is used to calculate the amount of water required for each crop [9]. This study describes an advanced technology-based smart system that uses K-NN algorithm to anticipate a field's irrigation needs by monitoring ground parameters such as soil moisture, temperature-humidity, and water level. Engineering of the framework to effectively deal with the water system process, has been proposed to acquire,communicate, and process the physical parameters of the agricultural area. The proposed method (K-NN) is meant to address the overfitting problem, consuming less memory, and making it compatible for all types of low memory devices and has given an accuracy of 93% [10].

Based on the area and crop type Naive Bayesian, Linear Support, Vector Classification, and K Nearest Neighbor algorithms were used to determine the type of fertilizer to use. The numerous types of soil, their healthiness and features, and the link between various types of pesticides and manures and the types of soil they are to be identified with are all needed for this undertaking. The results revealedthat KNN had the highest accuracy of 0.8145 in terms of thetype of fertilizer to be used, while Naive Bayesian and Linear SVC had 0.759 and 0.777 accuracy, respectively [11].

III. DATA COLLECTION

A dataset consists of a wide assortment of different attributes corresponding to a single variable. To implement the system, we decided to focus on selecting the database for major crops cultivated in the Indian geographic region. The first step involved collection of data from several websites like data.gov.in, Kaggle, GitHub etc. The data gathered from data.gov.in contains district-wideinformation about major production area and yield fromevery state since 2000. The insights gained from the dataset was about the cultivation of major crops and their productionquantity in the respective states of India.

Another crop dataset was obtained from a GitHub repositorythat provided with the detailed environmental requirements for the crop being cultivated. As the temperature, rainfalland other factors directly affect the crops, it is necessary to monitor the factors for the yield to be maximized.

A dataset was extracted from Kaggle to obtain the fertilizer requirements for every crop since every crop requires a balanced diet of essential nutrients during their entire growth cycle. Many of the required nutrients are obtained directly from the soil, but often in case of insufficient quantities, fertilizers are required to sustain high crop yields.
The last dataset was obtained from a GitHub repository to accurately predict the soil moisture to help the farmers prepare their irrigation schedules more effectively.

IV. DATA DESCRIPTION

The crop dataset obtained from GitHub contains 5 attributes – temperature, humidity, pH, rainfall, and the croplabel. The dataset contains information for 31 variety ofcrops which is spread over 3100 entries.
Temperature is mainly spread over the range of 18ºC to 37ºC. Humidity levels are spread across quite uniformly. pHscale is skewed rightwards towards the basic medium. The entries for rainfall are left-skewed indicating less heavy rainfall.

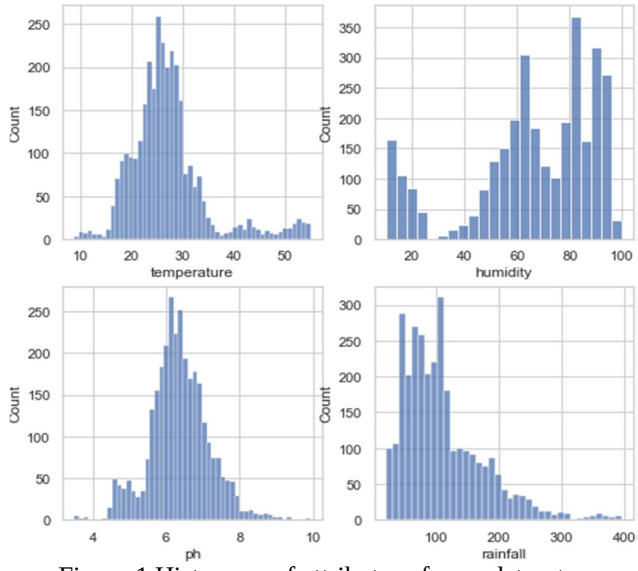


Figure 1 Histogram of attributes of crop dataset

The dataset for fertilizer requirements contains 5 attributes –crop label, Nitrogen levels, Phosphorus levels, Potassium levels, and pH scale. The dataset contains information for 97variety of crops which is spread over 1843 entries.
The dataset for irrigation contains 8 attributes – Air temperature (C), Air humidity (%), Pressure (KPa), Wind speed (Km/h), Wind gust (Km/h), Wind direction (Deg),Soil humidity, Irrigation. The dataset is spread over 23,995 entries giving detailed information about irrigation scheduling.

V. DATA PREPROCESSING

A. Crop and Fertilizer Dataset

1) Both the datasets were checked for null values.
2) Next, the crop labels were changed to lowercase letters.
3) Some changes were made in the names of croplabels of fertilizer dataset to ease the labelextraction process.
4) Labels were extracted from Fertilizer dataset.
5) Then, the extracted labels were used to create the processed datasets.
6) Finally, the datasets were combined.

B. Irrigation Dataset

1) The dataset was checked for missing values.
2) Missing values were treated by removing the rows containing the same.
3) Outliers were detected using histogram plots.
4) Outliers were treated by removing the rowscontaining the same due to abundance of data availability.

VI. METHODOLOGY

Machine learning is an application of Artificial Intelligence which allows user to feed computer algorithman immense amount of data and with the help of data machine learns itself and gives the prediction based only on the input data. The goal is to use historical data and predictive algorithm to find out the likelihood of future event and going beyond what has already happened. In our project we have used supervised machine learning techniques and artificial neural networks to recommend crop and fertilizers and the need of irrigation. Learning algorithmthat we have used in our research are:

A. Logistic Regression

Logistic regression algorithm is a classification algorithm used for classifying binary labels. The dataset which we have used contains multiple crops labels and so we had usedone VS rest for multiclass classification. One vs rest is the method that is used to classify multiclass classification usingbinary classifiers. It splits the multiclass into several binary classes and the binary classifier which in this case is logistic regression is trained on each binary classification problem and prediction is the class that gets the highest accuracy. The downside with this approach is that the model is applied to each class and would take a significantly large amount of time and memory in case of too many class labels.

B. Support Vector Machine (SVM-SVC)

The Support vector machine or SVM is supervised machine learning algorithm which is primarily used for classification problem but can also be used for regression problems. For a classification problem SVM separate different classes through a best fit line/plane called hyperplane which is formed with the help of support vectors which are the extreme points/vectors, so that the new data point can be easily put into correct category in future. In case of large or noise data SVM fails because the target class overlaps.

C. K-Nearest Neighbor:

Also known as Lazy learner algorithm is one of the simplest supervised machine learning algorithms. It assumes the similarity among the given dataset and the new data point and put the data point in the category which is most like it. KNN does not make any assumption on the underlying dataset hence it is a non-parametric algorithm.

D. Decision tree classifier

They are again a non-parametric supervised machinelearning algorithm which can be applied to both classification and regression problem. It is called decision tree because it starts with a root node which future expands its branches and form a tree like structure. It consists of internal nodes, branches and leaf nodes which represent feature of dataset, decision rules and output of those decision respectively. The goal of the decision tree is to compare the data point to the root node and based on the comparison follow the respective branch and jumps to next node till it reaches to the leaf node. Decision tree are one of the best ML algorithms as its mimic human decision makingability while making decision.

E. Naive Bayes Classifier

Naïve Bayes Classifier follows the principle of bayes theorem where it assumes that each features makes an independent and equal contribution to the outcome hence it called naive. It predicts the output based on the probability so it's a probabilistic classifier. The naïve bayes greatlyoutperform in the case of large and categorical dataset. Its assumption of independent predictors is a limitation becauseit is nearly impossible in real world that we get a set of predictors which are completely independent.

F. Random Forest Classifier

It is a customized version of decision tree classifier. It is based on the concept of ensemble learning where instead of using a single classifier a group of classifiers are used to solve a complex classification problem and improve theperform of the model. In the case of random forest multiple decision trees are trained on the subset of the dataset and takes the average of the output to improve the accuracy of the dataset. The more the number of trees in the forest the better the accuracy of the model and reduces the problem of overfitting.

G. Extreme Gradient Boosting

It is an ensemble machine learning algorithm that can be usein predictive analysis of classification and regressionproblem. It uses gradient boosting algorithm where decision trees are sequentially ensemble and fit to reduce the error from the prior model.

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

VII. RESULT

After the models are being trained on the train dataset, they are tested on test dataset so that the model with the best accuracy on the train as well as on the test dataset could be used for recommendations.

A. Accuracy Comparison

Here on both modules i.e., the crop recommendation and the irrigation recommendation the random forest classifier outperformed with an accuracy of 99.31% and 98.37% for crop and irrigation recommendationrespectively.

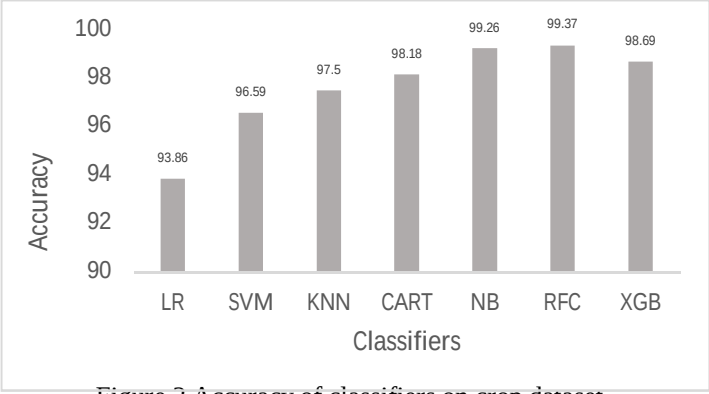


Figure 2 Accuracy of classifiers on crop dataset

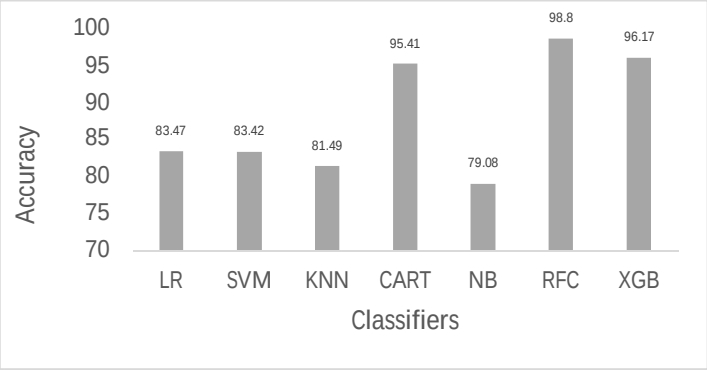


Figure 3 Accuracy of classifiers on irrigation dataset

B. Evaluation Matrix

1) Confusion Matrix

The confusion matrix is an N*N matrix which gives a summary of prediction results on a classification problem where N is the classes of the dependent variable.

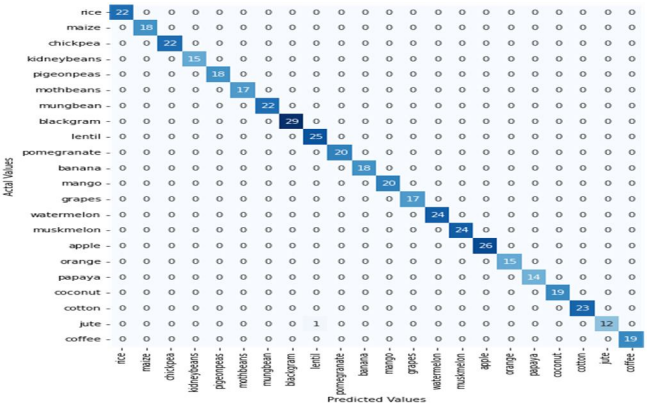


Figure 4 Confusion Matrix for Crop Prediction

This is the confusion matrix of dimension 22*22 for croprecommendation model where N is the number of crops. This matrix shows the comparison of actual crops to the crops predicted by the model

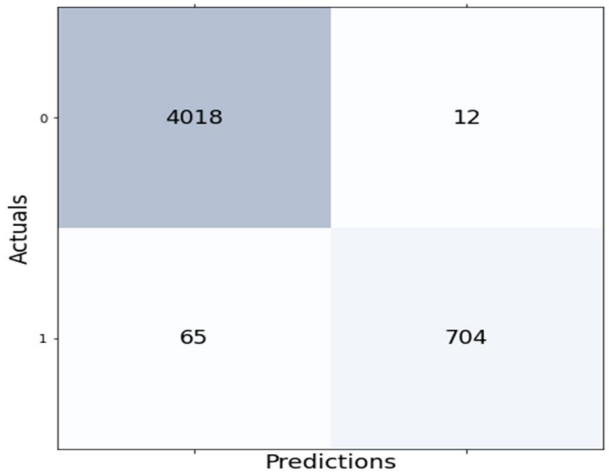


Figure 5 Confusion Matrix for Irrigation Recommendation

This the confusion matrix of dimension 2*2 for irrigation recommendation model which shows a comparison of actual need of irrigation to the predicted need of irrigation.

2) Classification Report

The classification report measures the qualityif prediction made by the classification model in the form of precision, recall, & f1-score.

	precision	recall	f1-score	support
rice	1.00	1.00	1.00	22
maize	1.00	1.00	1.00	18
chickpea	1.00	1.00	1.00	22
kidneybeans	1.00	1.00	1.00	15
pigeonpeas	1.00	1.00	1.00	18
mothbeans	1.00	1.00	1.00	17
mungbean	1.00	1.00	1.00	22
blackgram	1.00	1.00	1.00	29
lentil	0.96	1.00	0.98	25
pomegranate	1.00	1.00	1.00	20
banana	1.00	1.00	1.00	18
mango	1.00	1.00	1.00	20
grapes	1.00	1.00	1.00	17
watermelon	1.00	1.00	1.00	24
muskmelon	1.00	1.00	1.00	24
apple	1.00	1.00	1.00	26
orange	1.00	1.00	1.00	15
papaya	1.00	1.00	1.00	14
coconut	1.00	1.00	1.00	19
cotton	1.00	1.00	1.00	23
jute	1.00	0.92	0.96	13
coffee	1.00	1.00	1.00	19
accuracy			1.00	440
macro avg	1.00	1.00	1.00	440
weighted avg	1.00	1.00	1.00	440

Figure 6 Classification Report for Crop Prediction

	precision	recall	f1-score	support
0	0.98	1.00	0.99	4030
1	0.98	0.92	0.95	769
accuracy			0.98	4799
macro avg	0.98	0.96	0.97	4799
weighted avg	0.98	0.98	0.98	4799

Figure 7 Classification Report for IrrigationRecommendation

VIII. CONCLUSION

The recommendation of crops and fertilizer inputs based on geospatial data and various other parameters has proved that a high yield rate can be achieved. The recommendation of irrigation at appropriate durations can also result in high yield output.

Hence, we arrive at a conclusion that crop recommendation works best with Random Forest at 99.37% accuracy. For irrigation recommendation, Random Forest gives best resultswith an accuracy of 98.80%.

We believe that work in the agriculture industry with high technology integration can be extended further by creating aneasy-to-use mobile application for farmers where they can also detect the crop diseases by uploading images.

Agri-tech holds a broad scope for humankind and with significant efforts noteworthy developments can be achievedin this domain.

REFERENCES

[1] Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015), “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).

[2] Hardik Joshi, Monika Gawade, Manasvi Ganu, Prof. Priya Porwal, “Crop Yield Prediction Using Supervised Machine Learning Algorithm,” 2020 IOSR Journal of Engineering (IOSR JEN), ISSN (p): 2278-8719.

[3] Sudhansu Bisoyi, Archana Panda, “Designing of an Improved Agricultural Classification System,” IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE), Volume 12, Issue 6 Ver. III (Nov-Dec. 2017), PP 83-86.

[4] S. Vaishnavi., M. Shobana., R. Sabitha. and S. Karthik., "Agricultural Crop Recommendations based on Productivity and Season," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 883-886.

[5] R. T. Iyer, M. T. Krishnan, S. K. K. Pillai and L. P. P. Lalitha, "Geostatistical Modeling of Soil nutrients and development of a mobile application for site-specific fertilizer recommendation," 2019 IEEE Recent Advances in Geoscience and Remote Sensing : Technologies, Standards and Applications (TENGARSS).

[6] S. Vaishnavi., M. Shobana., R. Sabitha. and S. Karthik., "Agricultural Crop Recommendations based on Productivity and Season," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).

[7] J. J. I. Haban, J. C. V. Puno, A. A. Bandala, R. Kerwin Billones, E. P.Dadios and E. Sybingco, "Soil Fertilizer Recommendation System using Fuzzy Logic," 2020 IEEE REGION 10 CONFERENCE (TENCON).

[8] R. Kondaveti, A. Reddy and S. Palabtla, "Smart Irrigation System Using Machine Learning and IOT," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking(ViTECoN).

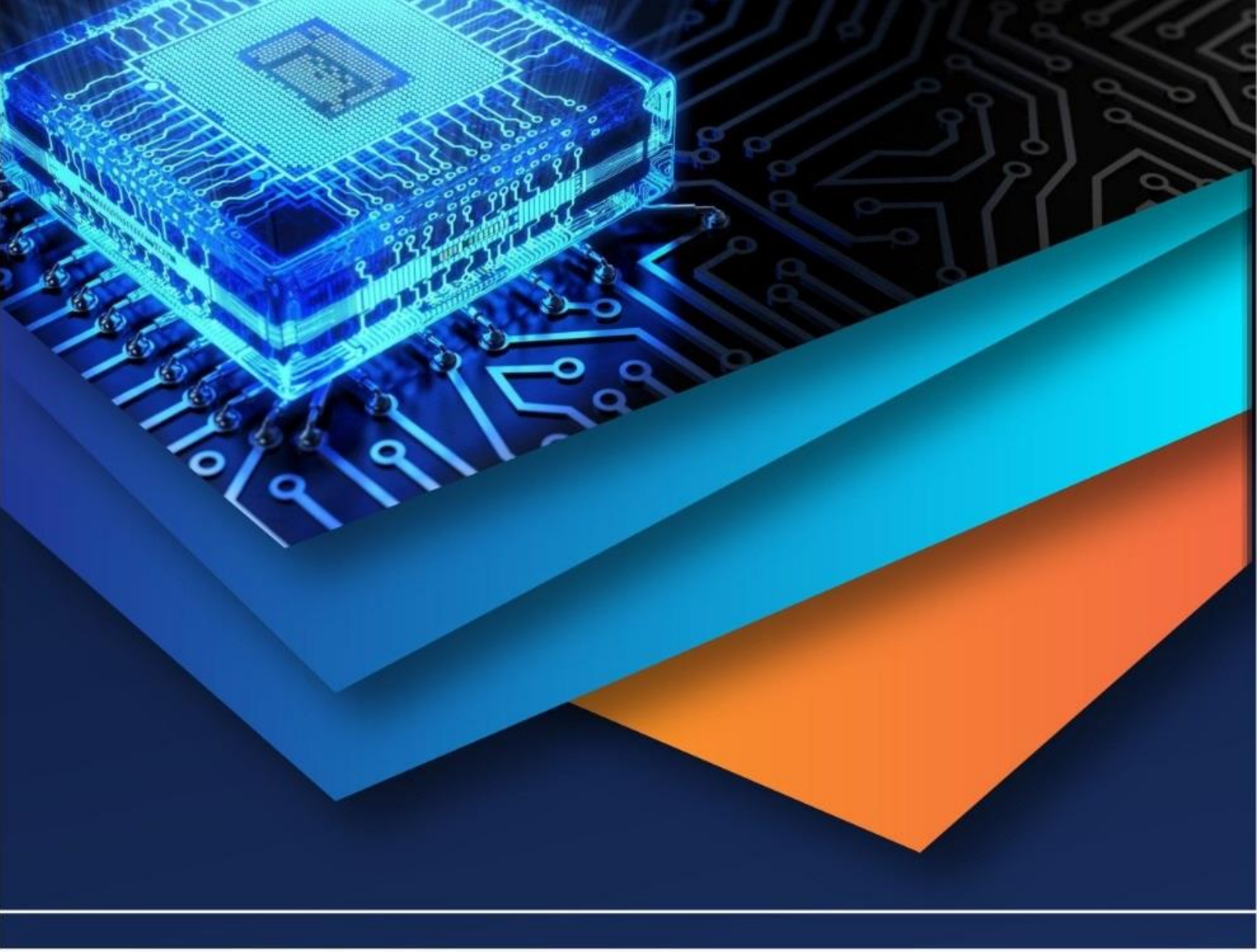
[9] J. Cardoso, A. Glória and P. Sebastião, "Improve Irrigation Timing Decision for Agriculture using Real Time Data and Machine Learning," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI).

[10] S. Akshay and T. K. Ramesh, "Efficient Machine Learning Algorithm for Smart Irrigation," 2020 International Conference on Communication and Signal Processing (ICCSP).

[11] L. Kanuru, A. K. Tyagi, A. S. U, T. F. Fernandez, N. Sreenath and S. Mishra, "Prediction of Pesticides and Fertilizers using Machine Learning and Internet of Things," 2021 International Conference on Computer Communication and Informatics (ICCCI).

[12] S. Aggarwal and A. Kumar, "A Smart Irrigation System to Automate Irrigation Process Using IOT and Artificial Neural Network," 2019 2nd International Conference on Signal Processing andCommunication (ICSPC)

©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)