iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** IV **Month of publication:** April 2020

DOI:

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: 7.429
Volume 8 Issue IV Apr 2020- Available at www.ijraset.com

Machine Learning Approach for Prediction of Crops

A. V. Patil[1], Shubham Khadki[2], Niranjan Dhumal[3], Tejas Kalbhor[4], Aniruddha Raje[5]
[1, 2, 3, 4, 5]*Information Technology Department, Zeal College of Engineering and Research Pune*

***Abstract:** In India, a major portion of its population lives in rural areas and has agriculture as its prime source of income. Since the traditional way of farming has its own disadvantage as it is a static way of handling a business. India has geographically rich land masses providing the farmers' opportunity to cultivate a range of different crops. Not just traditional agriculture but horticulture can also be seen as an option. In order to give farmers range of options this project proposes using machine learning algorithms Support Vector Machine (SVM) and Logistic regression. By considering different data points including dependent and independent variables giving accurate predictions as possible. The datasets in processing are obtained from data sources like Government of India, Government of Maharashtra and local bodies. The introduction of the open data policy of the Indian government helped in order to gain datasets. As the requirement of data was significant some data was generated explicitly. This paper gives a brief description about smart farming choices with help of algorithm and machine learning.*
***Keywords:** Prediction, machine learning, logistic regression, support vector machine, data cleansing*

I. INTRODUCTION

In 2019, the GDP of Indian is about 5% according to government which gives the overview of the industries in India. Many industries are facing some serious downfall, agriculture industry is also affected. Though this industry has the potential to withstand recession as it is the fundamental necessity of people, the industry is out performing. As most of the farmers don't see agriculture as feasible business option. This leads to depression in the production patterns. Market has experienced many staggering up-downs in the demand and supply of commodities, this project is an attempt to regulate the pattern. The agriculture yield prediction is a difficult task for agricultural departments over the world. But by processing datasets like fertility of soil, availability of water, previous weather records as well as current weather conditions, geographical location, etc. in order to generate results suggesting users about the options of crops in order to get maximum profit along with the factors that were taken in consideration to find the respective results. This will help the farmers with gaining more and also for the next generation farming after the green revolution of 1960 this is attempt for digital revolution in farming providing digital solution for their needs. Machine learning was developed from the concept that computers can self-learn to execute specific task without being explicitly programmed and pattern recognition. The iterative behaviour of machine learning is necessary for a reason as the models are exposed to new data, so they become able to capture knowledge accordingly. They continuously self-learn from their past computations to obtain more reliable results. With the humongous expansion of availability of data in almost every complex, large-scale, and networked systems. For example streams like security, surveillance, Internet, and finance, it becomes necessary to upgrade the basic as well as crucial understanding of modern analytic techniques and analysis from raw data to support more accurate predictions.

II. LITERATURE SURVEY

This paper is published by in 2016 IEEE conference paper. In this paper, the authors proposed a system that recommends a crop for farmers according to their region with the help of precision agriculture modern technique. Precision agriculture is the new trend of "site-specific" agriculture farming and it delivers more improvements. Major problems regarding crops in agriculture are solved though proposing a system using various techniques like CHAID, Random Tree, k-nearest neighbour, and Naive Bayes. Also paper contains a valid explanation and introduction to various techniques. They addressed the solution for getting more productivity and minimized the errors in making the decisions. This is the base paper because it gives useful information regarding choices in agriculture. The authors also give the introduction to new topic, ensemble technique for prediction of crops. It is a data mining model, which combines the power and strength of one or more models to obtain more accurate prediction, efficiency. The also insures that their model will give 88% accurate prediction [1] This paper is published by researcher D Ramesh, B Vishnu Vardhan in the year 2015, the journal named IJRET. This paper reveals new techniques. This paper introduced the technique namely density-based clustering techniques and multiple-linear regression (MLR) that help for the prediction of the crops in the specified region by the author in Andhra Pradesh. In this paper various elements or attributes are taken into consideration for prediction, this gives a rough idea about the challenges faced by farmers while farming. This paper focuses on the primary attribute weather conditions according to regions and also considered eight input valid variables of the data. It gives a brief introduction about data mining

methods which the author used in his proposed interface. It also described how this method is going to help in analysing and prediction rice. The main aim of authors is to build an interface for the Indian farmers which is user-friendly and will predict the crop. [2] The author's published papers in IEEE (2009). This paper introduces an important topic that is learning and getting a piece of information from imbalanced data. As agriculture is a dynamic field, the available data is in huge quantity and it changes day by day we can term this as an imbalanced data. For processing and learning on this dynamic data this paper is very useful and reduces work by giving a solution. It provides all the review of the improvement of studies (research) in getting useful information from data which is imbalanced and state_of_the_art solution required for the imbalanced data It gives the study about sampling methods like undersampling and oversampling which is consist of adding or modifying the data set of imbalanced raw data to provide a linear and balanced distribution. This paper also helps in the data cleaning process by suggesting various techniques like Tomek links which are used to minimize the overlapping that is required for sampling data. The author also states the nature of the problem and reveals new challenges, opportunities in the research of imbalanced data [3]

This paper is published by authors in Sep 2017. This paper describes the importance of data visualization in every field. Processing of various and large is made easy by machine learning and internet, that processed data, the output produced from that data has to be represented in easily understandable diagrams or charts so that the farmer with minimum technology knowledge will also get the information through our system. For this problem, we referred to paper because the authors focus on more detailed views of these data representations to extract the information from our outputs. One of the main drawbacks of the traditional system is overcome in this paper. The author's aim is to establish a system that will analyse as well as represent the data of the humidity (temperature), rainfall and the water levels in their region which will give them the availability of the growth of the crop in the selected area. [4]

This paper is published in 2013 by author Satish babu. In this paper, the author has given an introduction to precision agriculture. The technique is developed for giving the crop prediction and variety of soil present in the agriculture of developing states. The parameters considered in this paper are very useful to our system. The parameters are soil-crop databases, fields, calendars of crop, temperature, and rainfall using sensors. It also gives a short idea about precision farming previously done in India. The main goal of the author is to deliver a service of advising farmers about the smallest plot of crops using available technologies to produce more yields. [5] This paper is published in 2015 by the author's narishma and Manjula. They proposed a framework of agricultural prediction for the crop which is extensible and flexible. For this, they used a technique called precision agriculture which helps in prediction as well as analysing the data. Their related work gives a brief idea about their framework and the efficiency of their system. This paper also addresses the pre-processing of agricultural data. The methodologies are explained neatly with the help of appropriate figures. [6]

III.METHODOLOGIES

A. Data Collection

The data collection process involves the selection of quality data for analysis. Here we used dataset with four features namely soil moisture, ph, temperature, humidity and rainfall data. The job of a data analyst is to find ways and sources of collecting relevant and comprehensive data, interpreting it, and analysing results with the help of statistical techniques.

B. Data Visualisation

A large amount of information represented in graphic form is easier to understand and analyze. Some companies specify that a data analyst must know how to create slides, diagrams, charts, and templates. In our approach, the data histogram and scatter matrix are shown as data visualization part.

C. Data Pre-processing

The purpose of pre-processing is to convert raw data into a form that fits machine learning. Structured and clean data allows a data scientist to get more precise results from an applied machine learning model. The technique includes data formatting, cleaning, and sampling.

D. Data Splitting

A dataset used for machine learning should be partitioned into three subsets — training, test, and validation sets. Training set. A data scientist uses a training set to train a model and define its optimal parameters it has to learn from data. Test set. A test set is needed for an evaluation of the trained model and its capability for generalization. The latter means a model's ability to identify patterns in new unseen data after having been trained over a training data. It's crucial to use different subsets for training and testing to avoid model overfitting, which is the incapacity for generalization we mentioned above.

E. Model Training

After a data scientist has pre-processed the collected data and split it into train and test can proceed with a model training. This process entails "feeding" the algorithm with training data. An algorithm will process data and output a model that is able to find a target value (attribute) in new data an answer you want to get with predictive analysis. The purpose of model training is to develop a model.

F. Model Evaluation and Testing

The goal of this step is to develop the simplest model able to formulate a target value fast and well enough. A data scientist can achieve this goal through model tuning. That's the optimization of model parameters to achieve an algorithm's best performance.

IV. APPROACH

A. Dataset Collection

The dataset considered has attributes temperature, humidity, soil moisture, rain condition, ph value. The dataset is pre-processed to get the crop suggestion based on the attribute conditions. The below tables shows the dataset considered without pre-processing.

B. Support Vector Machine

Support Vector Machine SVM is a set of related supervised learning method used in medical diagnosis for classification and regression. SVM simultaneously minimize the empirical classification error and maximize the geometric margin. SVM is called Maximum Margin Classifiers and it can be efficiently perform non-linear classification using kernel trick. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a large margin gap that is as wide as possible. Given labelled training data as data points of the form. The SVM classifier first maps the input vectors into a decision value, and then performs the classification using an appropriate threshold value.

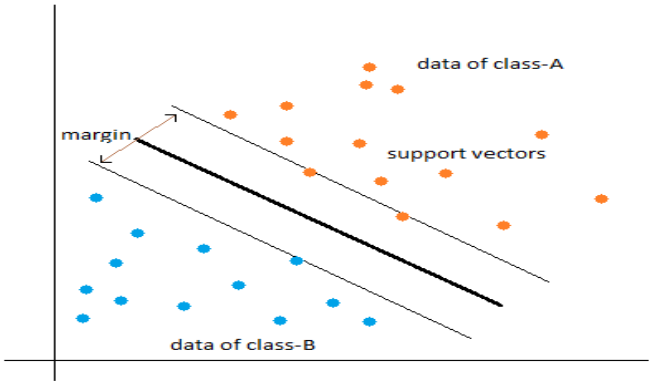


Fig. Support Vector Machine

C. Logistic Regression

Logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

V. SYSTEM IMPLEMENTATION

Implemented system is planned to consider every credible data point in order to generate as accurate result as possible. Attributes like climate, soil, crop history etc. can help the system to generate more accurate results.

The data analysis part of the system signifies the use of machine learning algorithms which does the task of data extraction, cleaning and processing.

The proposed work is implemented in Python 3.6.4 with libraries scikit-learn, pandas, matplotlib and other mandatory libraries. The crop dataset is applied machine learning algorithm such as SVM and Logistic regression. We used these machine learning algorithm and identified best crop. The result shows that crop prediction is efficient using Logistic regression algorithm. Logistic regression achieves 75.6% accuracy, whereas SVM achieves 59.4% accuracy.

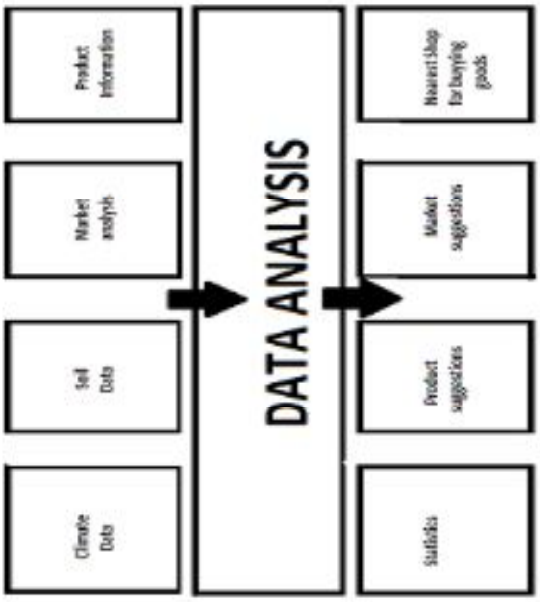


Fig. System architecture for proposed model

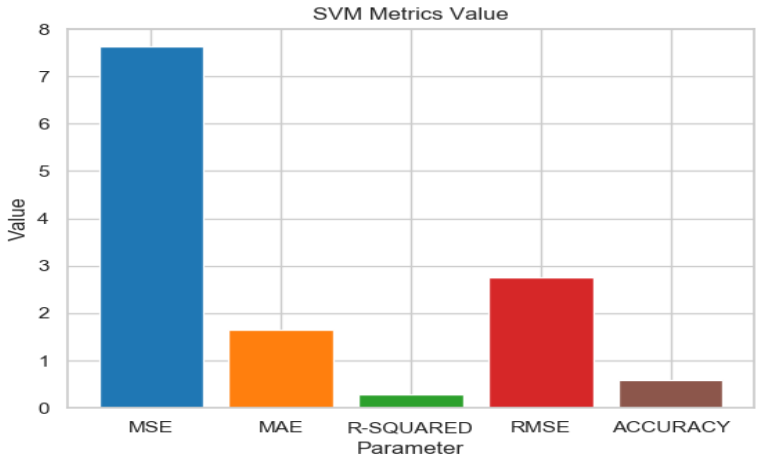


Fig. Evaluation of SVM algorithm

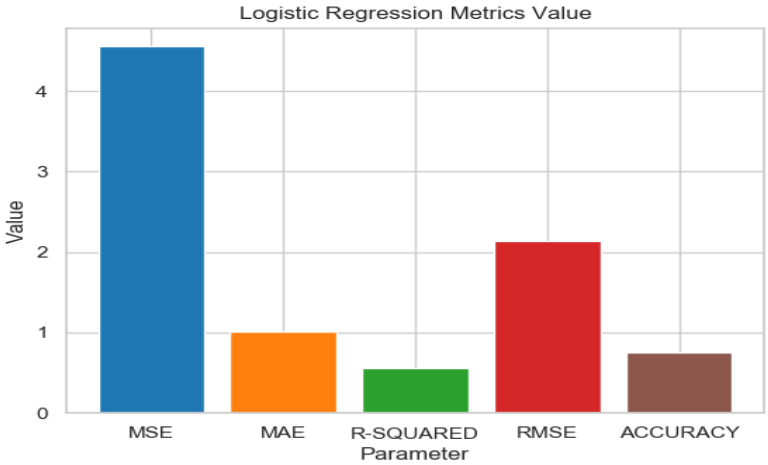


Fig. Evaluation of Logistic Regression

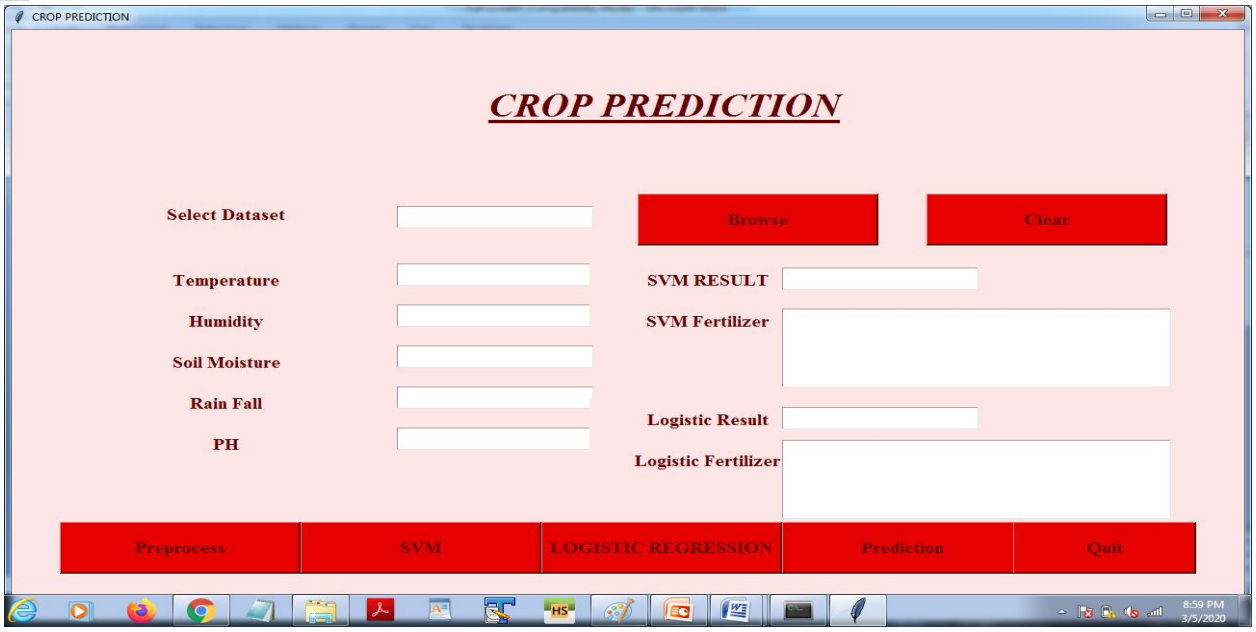


Fig. Home page of the system

VI. CONCLUSION

The method was created with the aim to overcome three obstacles. The first one is to find the effective amount of rainfall the second being to use the effective rainfall to find the irrigation water required and third to suggest suitable crop that should be implemented by the farmers to increase crop productivity. The method is successful in all three aspects and in the future it is expected to bring more areas under inspection and also bring more crops into the picture.

In future, we are interested to work on deep learning techniques such as DNN and CNN and compare the results.

VII. ACKNOWLEDGMENT

We express our sincerely thanks to all those who have provided us the valuable guidance towards the successful completion of this system as a part of syllabus for the bachelor's course. We express gratitude towards our co-operative department for providing us with all the valuable assistance.

We hereby take this opportunity to sincerely thank Asst. Prof. A.Yadav for her valuable guidance, inspiration, whole hearted involvement during every stage of this project and her experience, perception through professional and constant supervision as well as for providing necessary information regarding the project.

We are also thankful to Dr. Y. B. Gurav - Head of Department – Information Technology for his constant enlightenment, support and motivation which has been highly instrumental in successful completion of our project.

We will fail in our duty if we won't acknowledge a great sense of gratitude to the Principal Dr.A.M.Kate and the entire staff member in Information Technology department for their co-operation.

REFERENCES

[1] S.Pudumalar*, E.Ramanujam*, R.Harine Rajashreen, C.Kavyan, T.Kiruthikan, J.Nishan(2016),Crop Recommendation System for Precision Agriculture, IEEE Eighth International Conference on Advanced Computing (ICoAC).

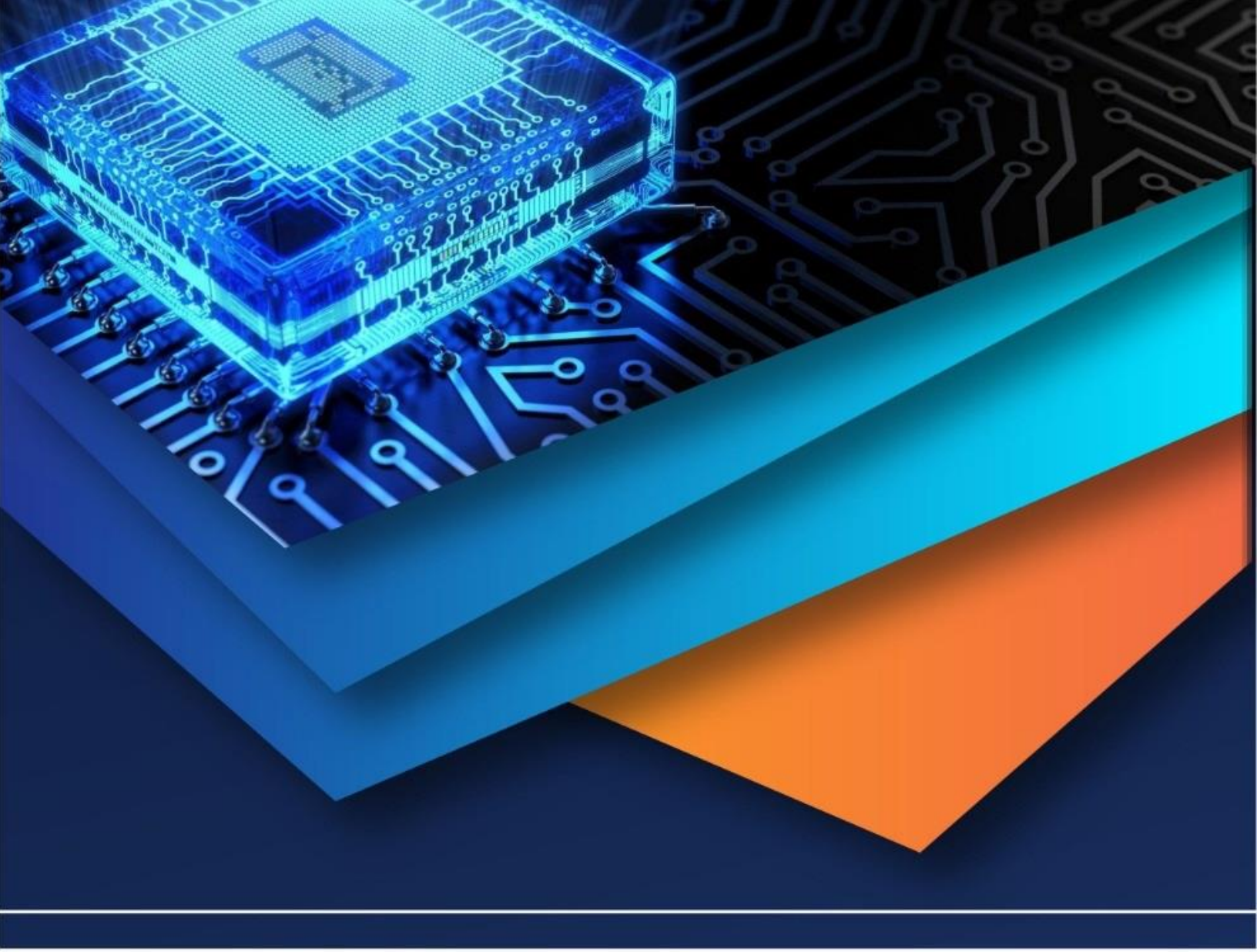
[2]]D Ramesh 1, B Vishnu Vardhan2 1Associate Professor, Department of CSE, JNTUH College of Engineering, Telangana State, India 2Professor, Department of CSE, JNTUH College of Engineering, Telangana State, India (1 Jan 2015),ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES

[3] Haibo He, Member, IEEE, and Edwardo A. Garcia: Learning from Imbalanced Data,(VOL. 21, NO. 9, SEPTEMBER 2009),IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING

[4] Harshitha B P, Amith R, Abhishek S, Rohit Vibhu C (Volume 49 Number 3 July 2017), Agricultural Data Visualization for Prescriptive Crop Planning ,International Journal of Computer Trends and Technology (IJCTT)

[5] Satish Babu (2013), 'A Software Model for Precision Agriculture for Small and Marginal Farmers', at the International Centre for Free and Open Source Software (ICFOSS) Trivandrum, India.

[6] Aakunuri Manjula, Dr.G .Narsimha (2015), 'XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction' , Conference on Intelligent Systems and Control (ISCO)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY