iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** IX **Month of publication:** September 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.55592

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 11 Issue IX Sep 2023- Available at www.ijraset.com*

# Crop Prediction Based on Characteristics of the Agricultural Environment

Bodduna Priyanka[1], Dr. V. Uma Rani[2]
[1]*MTech (Computer Science), Student, Department of Information Technology, JNTUHUCESTH, Hyderabad, Telangana – 500085*
[2]*Professor of CSE, Department of Information Technology, JNTUHUCESTH, Hyderabad, Telangana – 500085*

***Abstract:*** ***An ever-increasing number of individuals are keen on concentrating on horticulture. Crop prediction is significant in farming since soil factors like temperature, moistness, and downpour make a major difference. Ranchers used to have the option to pick which food to put, choose if it would develop, and choose when to accumulate it. All things considered, creators can't envision making this stride as a result of how rapidly the climate is evolving. Along these lines, machine learning (ML) recipes are presently used to make forecasts. A few of these ways were utilized in this review to sort out cultivating creation. For a machine learning (ML) model to be precise, successful element choice strategies should be utilized to transform unstructured information into a dataset that can be utilized for AI. To work on the nature of the model and cut down on duplication, just information qualities that hugely affect its result ought to be incorporated. Due to the ideal selection of parts, the model has just the most fundamental subtleties. On the off chance that each element from the first information is added to our model without its worth, it will be too difficult to even think about understanding. Likewise, on the off chance that factors that don't make any difference much were added, the model's outcomes would be less precise. The outcomes show that a gathering strategy is greater at making expectations than the standard framework for portraying individuals.***

***Keywords: Crop Prediction, Machine learning, Ensemble Techniques, Decision Tree, Random Forest, Support Vector Machine.***

## I. INTRODUCTION

Predicting crops is a hard undertaking in horticulture, and many models have been made and tried to help. Since food cultivating is impacted by both living and nonliving things, this issue should be tackled utilizing various datasets. Biotic variables are portions of the climate that are brought about by the immediate or backhanded impacts of living creatures on other living life forms, like microbes, plants, creatures, parasites, hunters, and irritations. In this classification, human variables (like preparation, plant care, the water framework, air contamination, water contamination, and grounds, and so on) are additionally considered. These elements could cause inner imperfections, actual irregularities, and changes in the compound design of food development. Both nonliving and living things significantly affect the temperature, as well as on how things occur and what plants are like. Abiotic components can be isolated into three gatherings: physical, substance, and other. Environment conditions (barometrical tension, temperature, dampness, air development, and daylight), mechanical vibrations (vibration, commotion), radiation, (for example, ionizing, electromagnetic, bright, and infrared), soil type, geology, soil roughness, air, and water science, particularly saltiness, are actual variables. A portion of the most exceedingly terrible things for the climate are sulfur dioxide and its results, polycyclic aromatic hydrocarbons (PAHs), nitrogen oxides and their side effects, fluorine and the mixtures it makes, lead and the mixtures it makes, cadmium and the mixtures it makes, nitrogen composts, pesticides, and carbon monoxide. Different ones are mercury, arsenic, dioxins and furans, asbestos, and aflatoxins. Notwithstanding bedrock, height, temperature, and water conditions, its characteristics are additionally impacted by non-living things. There are numerous manners by which soil-shaping variables influence how soil is made and the way that valuable it is for cultivating.

## II. LITERATURE REVIEW

Applying naive Bayes classification technique for classification of improved agricultural land soils:

Due to the development of portable PCs and information putting away gadgets, a ton of information has been made. To assist with finishing the data partition, new devices and strategies have been made, like information extraction. The difficulty has been sorting out some way to utilize these "crude information" to learn something. The objective of this study was to check whether there were any huge connections between these notable information mining techniques and a dirt science educating measure. At the S.V. Developing School, the Division of Soil Sciences and Rural Science has made various soil data records accessible. Evaluations of soil profiles from a couple of spots in the Chittoor Region, Chandragiri Mandal, were essential for the data that was gathered.

Various information mining strategies are utilized in the review to see whether sands are arranged. Likewise, a connection was made between the clarification of the Naive Bayes technique and the survey of the best methodology. The survey's outcomes could be utilized in various tasks associated with regular security, land the executives, and cultivating.

*A. Biotic Components Influencing the Yield and Quality of Potato Tubers*

Throughout recent years, how much potatoes filled in Canterbury has remained around 60 t/ha. Potato development expectations, then again, call for harvests of up to 90 t/ha, which a few modern cultivators have proactively reached. Business and school individuals cooperated for quite a long time to concentrate on the cutoff points on cultivating results. In anticipation of the fundamental establishing season, 11 yields were chopped down in an exuberant manner. on-going checks of result, plant wellbeing, and soil quality) Low yields have been accused on soil-borne sicknesses like Rhizoctonia stem annoy and Spongospora root contamination, soil compaction, and insufficient water the executives. The impacts of Rhizoctonia stem bruises appeared quicker (by deteriorating) where potatoes had never been developed and where plants were simply beginning to develop. In the subsequent year, a controlled field study was finished on a market crop that had a great deal of soil-borne microorganisms to figure out what soil-borne sicknesses meant for crop yield. As insect spray controls, we utilized flusulphamide (400 ml for every ha), azoxystrobin (1.5 l per ha), and chloropicrin (90, 112, and 146 kg for each ha). After treatment (showering fumigant on the plots), soil-borne microbial DNA estimations showed a little drop in the DNA levels of Rhizoctonia solani and Spongospora subterranea, however the outcomes were as yet unique. The strategy gave a mean worth of 58 t/ha, which was the main new finding at the biggest size. Contrasted with different drugs, the azoxystrobin treatment reliably decreased the harm that R. solani did to profound roots generally through the season.

*B. Response Surface methodology: A Retrospective and Literature Survey*

Response surface methodology (RSM) is a technique that utilizes both genuine preparation and mathematical standardization to further develop cycle and thing plans considerably more. During the 1950s, the primary examinations were finished around here. From that point forward, they have been utilized constantly, particularly in the medication and bicycle business. Throughout the course of recent years, RSM has changed a ton and has been utilized a great deal. The principal focal point of this rundown is on the RSM rehearses that started in 1989. We take a gander at the things that are being concentrated on now and recommend more things to study.

*C. Application of Response Surface Methodology for Optimization of Potato Tuber Yield*

In this review, the creator takes a gander at the parts that need to work to work on the development of potato establishes in Kenya. Potato ranchers are not permitted to give out any more data about this. Utilizing the response surface technique and the Factorial Plans 2 and 3, the most common way of making potatoes was made great. Utilizing a technique called "response surface," the impacts of water, nitrogen, and phosphorous material enhancements were contemplated and displayed to cooperate. The best potato root blend has 70.04 percent water structure water, 124.75 kilograms per hectare of urea nitrogen, and 191.04 kilograms per hectare of triple super phosphate phosphorus. In the best circumstances, a plot that is 1.80 meters by 2.25 meters can deliver 19.36 kilograms of potato roots. Limited scope ranchers in Kenya who develop potatoes might have better everyday environments and lower costs on the off chance that they develop more potatoes. I likewise feel that the strategy utilized in this potato study will be utilized in different examinations, which will assist us with grasping cultivating yield in a more complete manner.

*D. Improving Potato Yield Prediction by Combining Cultivar Information and UAV Remote sensing data using Machine Learning*

In accuracy farming, definite huge standard yield ideas are utilized to find worldwide yield anomaly patterns, make sense of significant variables that influence yield changeability, and give data for site-explicit administration. Changes in the sorts of guaging potato (Solanum tuberosum L.) could influence tuber yield assuming remote following devices are utilized. The objective of this study was to increment potato yield by consolidating crop data with machine learning (ML) checks and utilizing Unmanned Aerial Vehicles (UAVs) for distant identification. In little plot preliminaries in 2018 and 2019, various harvests and nitrogen (N) levels were utilized. During the development season, UAV photographs were taken consistently. The strategies for machine learning (ML) Random Forest Regression (RFR) and Support Vector Regression (SVR) were utilized to incorporate assortment information and explicit advances. Unmanned Aerial Vehicles (UAV) information from the tuber start stage in the early developing season (late June) was found to have a preferred association with potato taste over information from later in the season. Each kind of potato has its own arrangement of ideal indications of development and designated spots for advancement.

The RFR and SVR models didn't do well when they were given just information about how far separated things were (R2 = 0.48-0.51 for idea), yet they improved when they were given harvest information (R2 = 0.75-0.79 for help). At the point when ML calculations are utilized to join high-spatial-objective UAV pictures with cultivar information, the forecast of potato yield is obviously superior to with strategies that don't utilize cultivar information. More exploration will very likely be finished to further develop potato yield guaging by consolidating more precise reap information, soil and scene subtleties, guideline data, and solid ML gauges.

## III. METHODOLOGY

The principal test is to sort out the agroclimatic factors that influence the development of winter plants in the low-temperature zone, particularly grains. Winter yield is impacted by both the quantity of days with temperatures over 5 degrees Celsius and the quantity of days with temps somewhere in the range of 0 and 5 degrees Celsius. A significant number of these can be investigated utilizing public information and proportions of how frequently they occur over the long run. Utilizing notable models, the requirement for a state contribution plan in the food market has been checked out. For definite plans to be made, meteorological line expectations is required. Since these parts are not steady, a specific issue could occur. Specialists have attempted to take care of this issue in various ways, with changing levels of accomplishment.

### *A. Disadvantages*

*1)* Climate and soil factors, like temperatur humidity, and rainfall, hugely affect gather expectations.
*2)* Farmers haven't had the option to develop in light of the abrupt changes in the environmental factors.

There are a couple of issues with this area of examination. Despite the fact that crop expectation computations are great now, they could be better. This study gives a more exact method for assessing yields, which tends to these concerns. Succession and feature selection (FS) techniques are the fundamental pieces of the conjecture strategy. Before FS techniques are utilized to adjust an assortment, the cycle is checked out.

### *B. Advantages*

*1)* Just information scores that have a high measure of significance in concluding the model's end result ought to be utilized. This will eliminate apparent duplication and make ML models more exact.
*2)* An ensemble technique is superior to a past approach to grouping as far as how accurate it is.

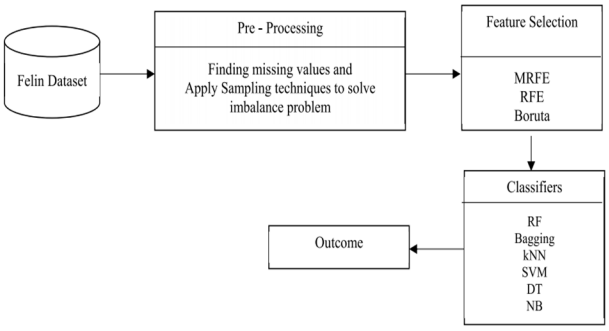



Fig 1: System Architecture

### *C. Procedure*

*1)* Import Shipment: All bundles will be stacked with this device.
*2)* This device will be utilized to send in the datasets with altered information from information revelation.
*3)* Cleaning and treatment of data: This device will give us information to deal with.
*4)* Perception with seaborn and matplotlib: This program shows data and information to the rest of the world.
*5)* With this device for handling, the information will be parted into sets called "train" and "test".
*6)* Voting Classification, Decision Tree, Support Vector Machine, and Gradient Boosting are a portion of the discretionary capabilities (SMOTE, ROSE, RFE, MRFE, BORUTA, and MEMOTE).
*7)* Training as a model: Utilizing the techniques you learned in this example, make the Voting Classifier model for handling and anticipating. It is more precise than different models you might have seen.

*8)* Sqlite and the Carafe System are utilized for join and sign-in: With this element, clients can join and sign in by bringing in programs.

*9)* The client adds numbers for things like: When this module is utilized, the client gives contribution to expectation, and the info is utilized to make forecasts. Assumptions depend on the arrangement that was educated: The point of interaction is utilized to show the outcomes that this module will make conceivable.

Note: The Gradient Boosting Voting Classifier is utilized as an expansion for highlight esteems that are utilized in estimates. This is on the grounds that it is around 100 percent more accurate than different models.

*D. Algorithms*

A variety of Methods from machine learning can beused on cleaned-up data, with a focus on methods that provide clear and transparent decision-making processes. Some understandable techniques include:

*1)* *Random Forest:* A popular ensemble option tree technique for evaluating each characteristic.

*2)* *Decision Tree:* Decision trees are accessible and can be depicted visually, making choosing an avenue to take easier.

*3)* *SVM (Support Vector Machine):* SVM is useful in information categorization, and its support vectors can be used to study the processes of decision-making.

*a)* *KNN:* "K-Nearest Neighbour" maybe inscribed as "KNN." It is a forethought that is to say conditional information created by society. With this approach, you can handle both relapse and plenty questions immediately. The letter "K" talks about how many determinants accompanying close neighbours maybe anticipated or sink groups.

*b)* *Naive Bayes:* Naive Bayes' chance categorization is used to interpret entity. It maybe proved by way of hints and models of probabilities that maybe altered. Few nation's plans of exemption examine accompanying how belongings honestly are. Because of this, society anticipate they're stupid.

*c)* *Random Forest:* The Random Forest (RF) Random Forest is an main machine learning form for learning in a regulated habit. ML issues like preparation and duplication maybe answered cognizant. It is established the plan of group realising, that leads together various orders to answer a hard question and boost the habit the model looks. As allure name plans, the Random Forest classifier "takes the average of diversified decision trees on various subsets of the likely dataset" and "everything on the widespread veracity of that dataset." Instead of depending just individual preferred shrub, the spread forest resolve what will take place by adjoining up the estimates each sapling and utilizing ultimate likely consequence.

*d)* *Decision Trees:* Decision trees use any of determinants to elect if to split a center into not completely two substitute-centers. Because of how they look, the subnodes are more alike. To set it another habit, the centre gets more exact as it gets tighter to the aim changing.

*e)* *Support Vector Machine or SVM:* Support Vector Machine (SVM), a familiar Unsupervised Learning design, is secondhand by both Backslide and Depiction. But most of allure ML uses are established order questions. The aim of the SVM approach for n-flaky room design search out find highest in rank line or judgement limit for fear that new news maybe surely increased the right arrangement. A hyperplane is the borderline that create ultimate sense.

## IV. EXPERIMENTAL RESULTS

The experiments in this work were done using a PC with 4GB RAM, an Intel Core i5 5th generation CPU, and a Jupyter Notebook with 4GB storage.

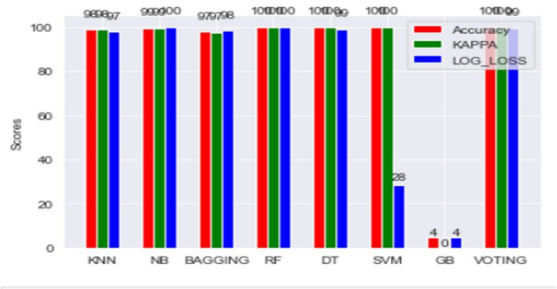



Fig 2: Accuracy of all the Algorithms

The above figure Shows the accuracy of classification techniques and ensemble techniques, in which voting classifier has the highest accuracy of 97.7272 comparing to the other algorithms.

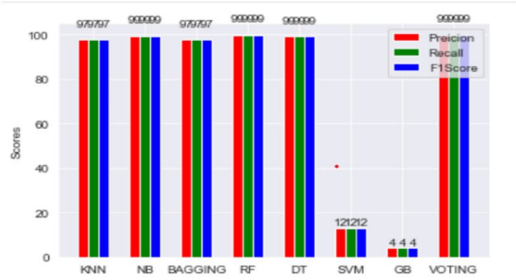



Fig 3: Precision, Recall and F1score of the Algorithms

Fig 3 shows that the precision, recall and F1score of the Classification and ensemble techniques.
Based upon the above results of Accuracy, Precision, Recall and F1score Voting classifier which ensembled Decision Tree, Random forest and Support Vector Machine has the highest prediction rate comparing with the other classification Techniques.

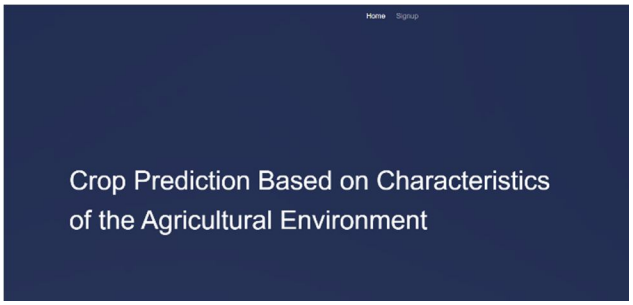



Fig 4: Home screen

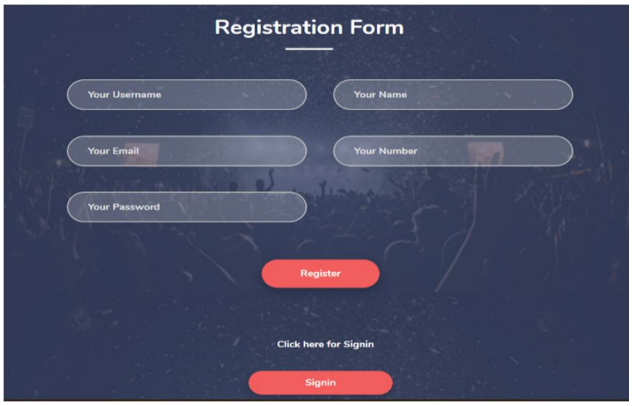



Fig 5: User signup

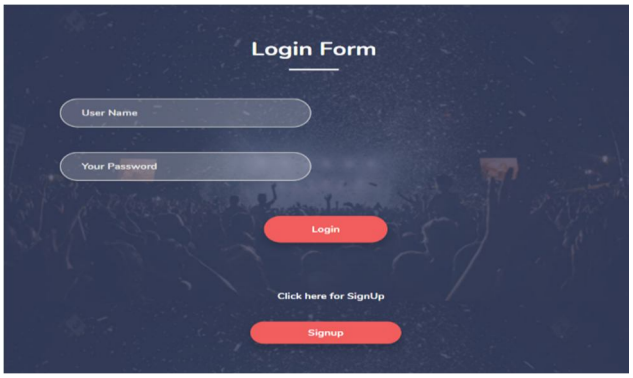



Fig 6: User Login

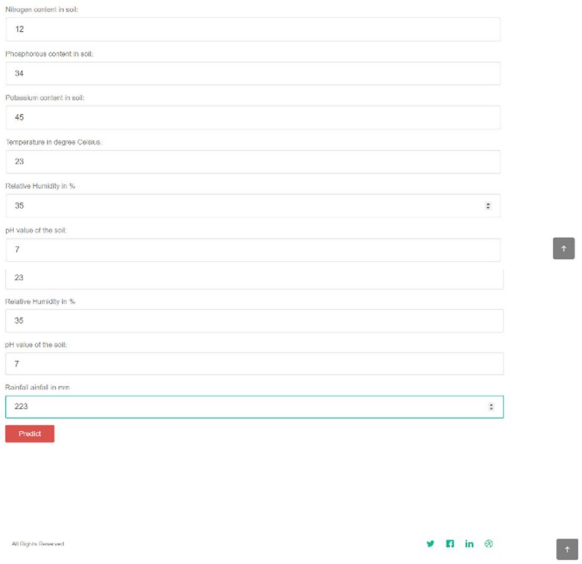



Fig 7: User Input

Fig 7 shows the input parameters which are collected to predict the crop those input parameters are Nitrogen, Phosphorous, Potassium, Temperature, Rainfall, Humidity and Ph value.
Based upon all these input parameters the output will be generated after analysing the data. i.e. Summer crops & winter crops.

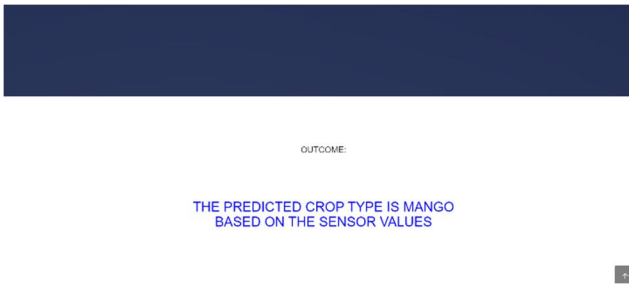



Fig 8: Prediction result

Fig 8 ssshows the predicted crop based upon the given input parameters.

## V. CONCLUSION

A difficult issue in agriculture is predicting the crops that will be grown. This study will predict the size of plant cultivation yields using a variety of feature selection and classification techniques. The findings will show that compared to the current classification technique, an ensemble technique which will help to deliver greater prediction accuracy. Planning the organization of their planting, both on the farm and national scales, can be done by forecasting the area of cereals, potatoes, and other energy crops. Utilizing contemporary forecasting methods can result in quantifiable financial advantages.

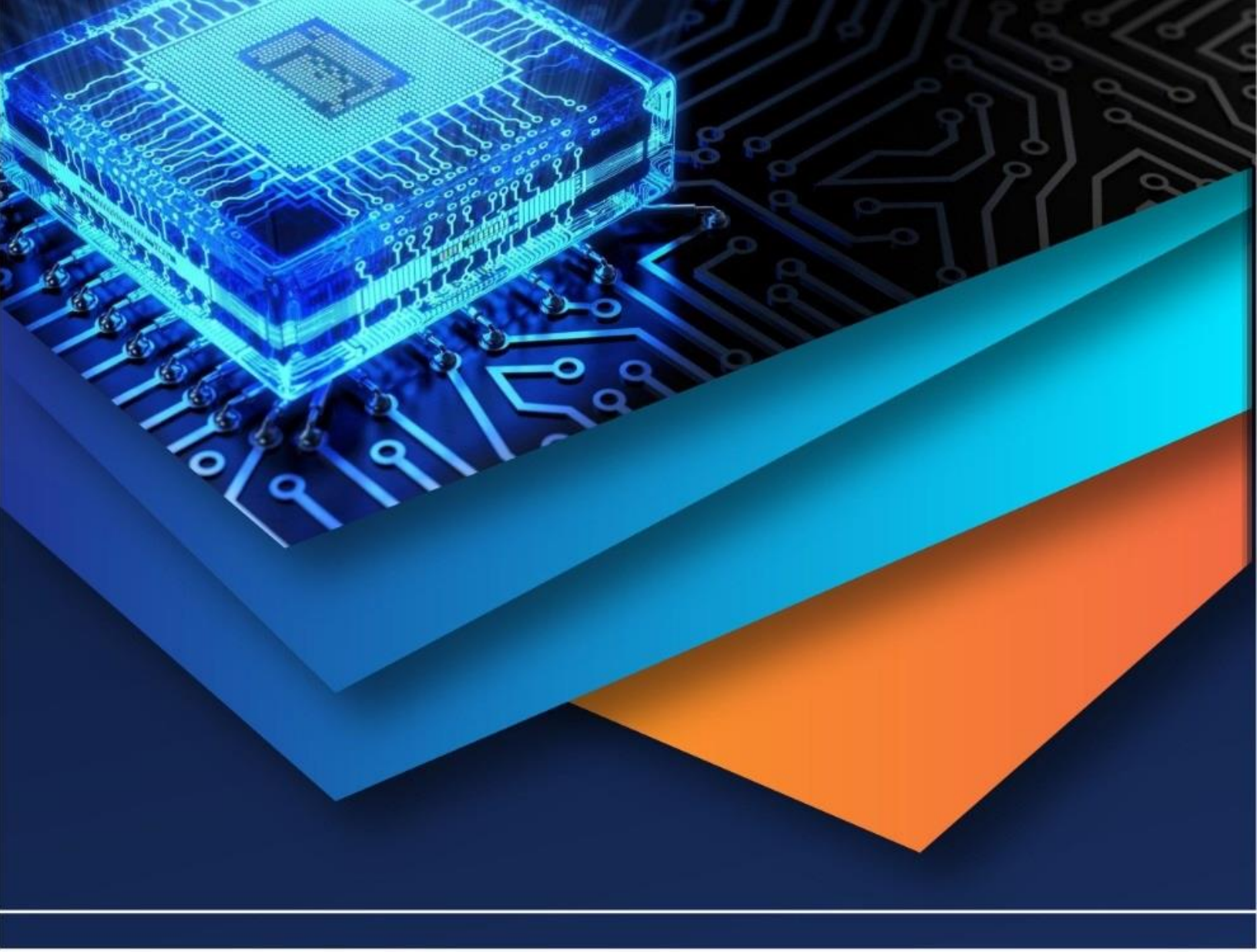
## VI. FUTZRE ENHANCEMENT

Future research will be focused on growing the dataset's data and adding more classes in order to enhance precision, recall, and F1 score and by using sensors.

## REFERENCES

[1] R. Jahan, ‘‘Applying naive Bayes classification technique for classification of improved agricultural land soils,’’ Int. J. Res. Appl. Sci. Eng. Technol., vol. 6, no. 5, pp. 189–193, May 2018.
[2] B. B. Sawicka and B. Krochmal-Marczak, ‘‘Biotic components influencing the yield and quality of potato tubers,’’ Herbalism, vol. 1, no. 3, pp. 125–136, 2017.
[3] B. Sawicka, A. H. Noaema, and A. Gáowacka, ‘‘The predicting the size of the potato acreage as a raw material for bioethanol production,’’ in Alternative Energy Sources, B. Zdunek, M. Olszáwka, Eds. Lublin, Poland: Wydawnictwo Naukowe TYGIEL, 2016, pp. 158–172.

[4] B. Sawicka, A. H. Noaema, T. S. Hameed, and B. Krochmal-Marczak, ‘‘Biotic and abiotic factors influencing on the environment and growth of plants,’’ (in Polish), in Proc. Bioróżnorodność Środowiska Znaczenie, Problemy, Wyzwania. Materiały Konferencyjne, Puławy, May 2017. [Online]. Available: https://bookcrossing.pl/ksiazka/321 192
[5] R. H. Myers, D. C. Montgomery, G. G. Vining, C. M. Borror, and S. M. Kowalski, ‘‘Response surface methodology: A retrospective and literature survey,’’ J. Qual. Technol., vol. 36, no. 1, pp. 53–77, Jan. 2004.
[6] D. K. Muriithi, ‘‘Application of response surface methodology for optimization of potato tuber yield,’’ Amer. J. Theor. Appl. Statist., vol. 4, no. 4, pp. 300–304, 2015, doi: 10.11648/j.ajtas.20150404.20.
[7] M. Marenych, O. Verevska, A. Kalinichenko, and M. Dacko, ‘‘Assessment of the impact of weather conditions on the yield of winter wheat in Ukraine in terms of regional,’’ Assoc. Agricult. Agribusiness Econ. Ann. Sci., vol. 16, no. 2, pp. 183–188, 2014.
[8] J. R. Olędzki, ‘‘The report on the state of remotesensing in Poland in 2011–2014,’’ (in Polish), Remote Sens. Environ., vol. 53, no. 2, pp. 113–174, 2015.
[9] K. Grabowska, A. Dymerska, K. Poáarska, and J. Grabowski, ‘‘Predicting of blue lupine yields based on the selected climate change scenarios,’’ Acta Agroph., vol. 23, no. 3, pp. 363–380, 2016.
[10] D. Li, Y. Miao, S. K. Gupta, C. J. Rosen, F. Yuan, C. Wang, L. Wang, and Y. Huang, ‘‘Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning,’’ Remote Sens., vol. 13, no. 16, p. 3322, Aug. 2021, doi: 10.3390/rs13163322.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY