Prediction of Crop Based on Characteristics of Agricultural Environment Using Machine Learning Techniques

Madhavarapu Prathima Rao[1(✉)], R. Jegadeesan[1], P. Pranitha[1], D. Praveen Kuamar[2], and J. Krishna Chaitanya[2]

[1] Department of CSE, Jyothishmathi Institute of Technology and Science, Karimnagar, Telangana, India
prathimaraom@gmail.com, polsani.pranitha@jits.ac.in
[2] Department of Electronics and Communication Engineering, Vardhaman College of Engineering, Shamshabad, Hyderabad, India

Abstract. Horticulture related research is creating to anticipate crops, farming, specifically, essentially depend on soil and natural components including temperature, stickiness, and precipitation. Ranchers used to be accountable for picking the yield to be developed, watching out for its development, and choosing when to gather it. But since of the climate's fast changes these days, it is hard for the cultivating local area to proceed. Accordingly, AI methods are continuously dislodging ordinary forecast strategies. A few of these procedures have been utilized in this review to gauge rural yield. To guarantee that a specific AI (ML) model capabilities with an elevated degree of accuracy/properness, it is critical to use effective component determination strategies to change over the crude information into a dataset that is AI well disposed. To lessen copy information and increment model exactness, just information qualities that are exceptionally applicable to characterizing the last result of the model ought to be incorporated. It is basic to utilize ideal element determination to ensure that main the most pivotal highlights are remembered for the model. Assuming we coordinate all attributes from the crude information without first considering part in the model-building procetures, This model will turn out to be excessively perplexing. The incorporation of new boundaries that have negligible bearing on the model's presentation will likewise raise the time and spatial intricacy of the ML model. The outcomes exhibit that a group procedure gives higher expectation exactness when contrasted with the ongoing characterization techniques.

Keywords: Crops · Zigbee · Monitoring · Soil · Security · Data models

1 Introduction

Various models have been made and tried because of the many-sided process engaged with foreseeing crops in horticulture. The use of a few datasets is expected since crop development is reliant upon both biotic and abiotic factors [1, 2]. Crop estimating is

X. Cheng (Ed): BROADNETS 2024, LNICST 601, pp. 202–210, 2025.
https://doi.org/10.1007/978-3-031-81168-5_18

neither direct nor easy to-do. As per Myers et al. furthermore, Muriithi, the system for estimating the region under development is an assortment of measurable and numerical devices supportive in an iterative and further developing streamlining process [5, 6].

ZigBee has many applications in accuracy horticulture, where the Web of Things is utilized for Brilliant field the board by unequivocally checking components affecting the cultivated yields to help improved and upgraded farming result. To boost the creation, a framework like this cautiously screens an assortment of development related factors, including temperature, soil quality, pH, saltiness, stickiness, and so on. Like ZigBee, Z-Wave can be used in farming for different observing and control frameworks. Z-Wave innovation's interoperability can be utilized to construct arranged agrarian frameworks with productive correspondence between them.

To screen soil dampness levels and upgrade water system, LoRa innovation has been utilized. Every water system valve in a grape plantation has soil dampness sensors connected. The sensor consistently finds out the dirt dampness content and sends the information to the LoRa door inside its reach. Contingent upon the situation, the water system valves can be controlled. This permits ranchers utilizing LoRa to monitor up to half more water. LoRa can likewise be utilized to follow meteorological factors like temperature and stickiness. Much appreciated by and large to LoRa, which has likewise assisted the improvement of Savvy ranches, farming and IoT with having met up? Also, finding leaks can be utilized. Horticulture utilizes ZigBee, Wi-Fi, Bluetooth, and LoRa to assemble ongoing information for expectation processes. The expectation cycle in this paper utilizes an ongoing static farming dataset from the earlier year. Thus, in this work, ML procedures are used. In the field of study, there are different troubles. Despite the fact that they could perform better, crop expectation models at present produce results that are acceptable. The two principal highlight determination and characterization methodologies were anticipated. To adjust an uneven dataset, inspecting strategies are utilized before highlight determination draws near [11, 12].

The Engineered Minority Over-Examining Procedure (Destroyed), Larger part Weighted Minority Over-Testing Method (MWMOTE), and Irregular Over-Testing Models (ROSE) are utilized in this work to adjust the provided dataset. It is possible to use classification techniques that perform better and can help identify the target class by identifying the most crucial elements in a dataset. In this review, the Boruta, Recursive Element Disposal (RFE), and Adjusted Recursive Component End (MRFE) covering highlight determination techniques are utilized to recognize the critical elements of the dataset. A few directed grouping procedures, including as Innocent Bayes (NB), Choice Tree (DT), k Closest Neighbor (KNN), Backing Vector Machine (SVM), Packing, and Irregular Timberland (RF), are prepared with the chose elements to foresee an OK result from the dataset.

2 Related Work

"Applying naive bayes classification technique for classification of improved agricultural land soils" [1]. India is mostly an agricultural nation. North Indian Soil and South Indian Soil are the two distinct types of soil found in India. North Indian Soil and South Indian Soil are the two sets we have divided into. Deep alluvial soil is typically used

to produce the plains of northern India. The top soil's surface varies from sand to clay, with the majority being light loam, which is naturally fertile and has a leaky texture. The southern India peninsular earth plane, or south Indian soil, is prepared up of hills and river valleys. Naturally, hilly areas are unsuitable for farming, and some moorland is quite warm [Prakash, 2015]. The goal of this research is to categorise soil types according to their traits and fertility potential using classification techniques like naive bayes, and to compare this technique to others like zero and stacking. The objective of this study is to evaluate novel data mining approaches used to classify soil in soil databases, as well as to compare the naive bayes classifier to other classifiers.

By B. B. Sawicka and B. Krochmal-Marczak, "Biotic components influencing the yield and quality of potato tubers" [2]. Over the past ten years, Canterbury's potato yields have been steady at about 60 t/ha. Models of potato growth suggest that yields of up to 19t/ha are feasible, however some commercial producers have already reached this level. Industry and research trends looked into the factors restricting agricultural productivity over a two-year period. In year 1, 11 processing crops were closely observed for evaluations of soil quality, plant health, and final output. The main causes of falling yields were found to be soil-borne diseases (such as Spongospora root infection and Rhizoctonia stem canker), subsurface soil compaction, and inadequate irrigation management. Eight weeks following the emergency, farms with recent histories of potato crop production began exhibiting stem canker symptoms earlier than those with periods of grass growth and no preceding potato crops. In year 2, using a commercial crop that was known to have high levels of soil-borne pathogens, researchers attempted to isolate and quantify the impact of soil-borne diseases on yield in a controlled field experiment.

Response Surface Methodology: A Retrospective and Literature Survey, by R. H. Myers, D. C. Montgomery, G. G. Vining, C. M. Borror, and S. M. Kowiski [3]. Process and product designs are optimised using Response Surface Methodology (RSM), a collection of statistical design and numerical optimisation approaches. Since the initial study in this field was undertaken in the 1950s, it has seen a lot of use, especially in the chemical and process industries. For the past 15 years, RSM has been widely used, and during that time, numerous significant advancements have been made. We specifically focus on RSM efforts from 1989 in this review analysis. We talk about the areas of research that are currently being done and suggest some new ones.

3 Preliminaries

Python is an interpreter, high-level, interactive, and object-oriented scripting language. It is quite easy to read Python code. Compared to other languages, it has fewer syntactical characteristics and frequently uses English terminology in place of punctuation. Learning Python is Fundamental for the two understudies and working experts who wish to become extraordinary programmers, particularly on the off chance that they work in the web improvement area. The principal advantages of learning Python incorporate the accompanying.

Python is interpreted, which means that an interpreter processes it as it is being used. Before running your programme, there is no requirement that it be compiled. These are PERL and PHP-like in nature. Python is intelligent; you can utilize a Python brief

to sit at the console and speak with the mediator as you make your projects. Python upholds the article arranged programming style or approach, which typifies code into objects. Python is dynamite First Programming Language Python is a tremendous first programming language and works with the improvement of a great many applications, from basic text handling to internet browsers to games.

Man-made brainpower (artificial intelligence) and software engineering's part of AI centers around utilizing information and calculations to reenact human growing experiences and consistently further develop exactness. Normally, an assortment of AI calculations is utilized to give characterizations or forecasts. In view of a few unlabeled or named input information, the calculation will give a gauge about an example in the info information. A mistake capability will be utilized to assess the model's forecast. On the off chance that realized models exist, a blunder capability can contrast them with decide if the model is precise.

4 Proposed Work

The notable highlights of the dataset are recognized in this work utilizing covering highlight choice strategies as the Boruta, Recursive Component Disposal (RFE), and Altered Recursive Element End (MRFE). To foresee a proper result from the dataset, various managed order techniques, including Gullible Bayes (NB), Choice Tree (DT), k Closest Neighbour (KNN), Backing Vector Machine (SVM), Sacking, and Arbitrary Backwoods (RF), are prepared with the picked qualities.

Feature Selection Techniques:

Boruta is a random forest-based algorithm; hence it is applicable to classification models like Logistic Regression and SVM as well as tree models like Random Forest and XGBoost. Boruta iteratively eliminates features (artificial noise variables produced by the Boruta method) that are statistically less significant than a random probe. Variables that are rejected in one iteration are not taken into account in the following iteration. It usually yields an excellent overall feature selection optimisation.

The RFE procedure, a covering highlight choice strategy, begins with the whole dataset. RFE is well known in light of the fact that it is clear to set up and utilize and compelling at choosing the highlights (sections) in a preparation dataset that are more or generally important for anticipating the objective variable.

Classification Techniques:

Decision Tree Classifiers - The applications of selection tree classifiers are diverse and useful. Their main strength is their ability to distinguish unambiguous dynamic data from given information. An election tree can be made using preparation kits. A set of objects (S), each with a place in a different class (C1, C2… Ck), is the basis of the method used to follow such an age.

K-Nearest Neighbors (KNN) - KNN is a straightforward yet extremely effective classification method. Based on a similarity metric, it classifies. The algorithm is non-parametric and lazy. Before the test example is supplied, it does not "learn". We use the training data to identify the K-nearest neighbours of any fresh data that needs to be classified.

Logistic Regression Classifiers - The relationship between a bunch of free (illustrative) factors and an all out subordinate variable is examined utilizing strategic relapse investigation. At the point when the reliant variable just has two qualities, for example, 0 and 1 or Yes and negative, the term strategic relapse is utilized. In circumstances where the reliant variable has at least three particular qualities, like Wedded, Single, Separated, or Bereft, the term multinomial strategic relapse is commonly held. Notwithstanding involving an alternate arrangement of information for the reliant variable than various relapse, the technique has a comparable down to earth use. On autonomous factors that are both mathematical and all out, this program processes parallel calculated relapse and multinomial strategic relapse.

Naive Bayes - The oversimplified assumption that the presence or lack of one characteristic in a class has no bearing on the presence or absence of any other feature is the foundation of the naive bayes approach, a supervised learning technique. However, it still appears solid and powerful. It is equally as successful as other supervised learning techniques. Numerous explanations have been put forth in the literature. The naive bayes classifier is a linear classifier, along with logistic regression, linear SVM, and linear discriminate analysis. The method used to calculate the classifier's parameters explains the discrepancy.

Random Forest - Irregular woodlands make a ton of choice trees during the preparation stage as a gathering learning procedure for order, relapse, and different issues. The after effect of the irregular timberland for order issues is the class that most of the trees pick. The mean or normal gauge of every individual tree is returned for relapse undertakings. Choice trees frequently over fit their preparation set, however arbitrary decision woodlands right this issue. Angle upgraded trees much of the time beat irregular woodlands with regards to accuracy. Nonetheless, the idea of the information could influence how helpful they are.

SVM-SVM is a discriminate technique because, unlike evolutionary algorithms or perceptrons, both of which are frequently employed in machine learning for classification, it analytically solves the convex optimisation issue and consistently returns the same optimal hyper plane value. Solutions for perceptions depend heavily on the initialization and termination standards. The perception and GA classifier models are different each time preparing is initialised, though preparing gives particularly characterized SVM model boundaries for a given preparation set for a specific portion that interprets the information from the info space to the component space. Since GAs and perceptrons just objective is to diminish preparing related mistake, numerous hyper planes will fulfil this prerequisite.

5 System Architecture

To find missing qualities, eliminate excess data, normalize the dataset, and change target ascribes into factor credits, the info information is first pre handled. The pre handled information is utilized to separate key credits utilizing covering highlight choice strategies. Preceding applying order strategies to the improved properties, the dataset is parted into preparing and testing stages. The order calculation is prepared utilizing obscure examples from the preparation dataset to recognize the yield that is generally proper for

filling in a specific plot of land. The prepared classifier is applied to the testing dataset to figure (Fig. 1).

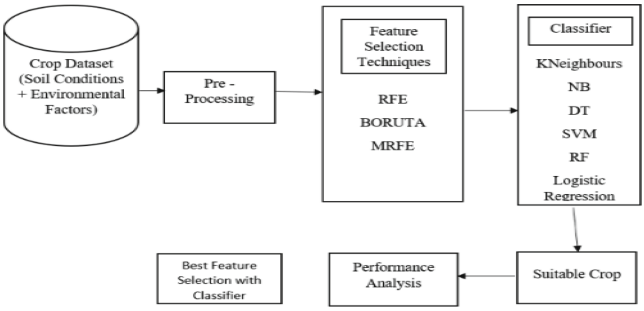

Fig. 1. System Architecture

The crop that will be grown. A suitable crop is finally harvested, and the outcomes are assessed using several performance criteria. The research identifies the most effective feature selection strategy and the most suitable classification approach.

6 Results and Analysis

In the below image represents the input and output of prediction of crop. Based on feature selection techniques and classifiers it predicts whether the crop is suitable or not for that Soil condition and Environment conditions and Recommended zone, and crop name, parameters considered for soil and environment conditions were N,P,K, ph(soil), Environment(temperature, humidity, rainfall) (Figs. 2 and 3).

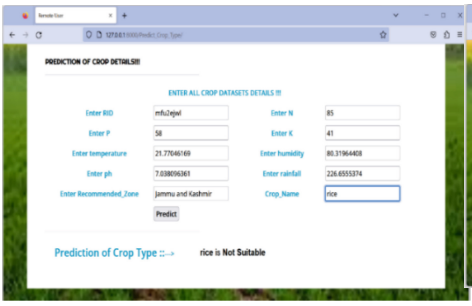

Fig. 2. Result output -1

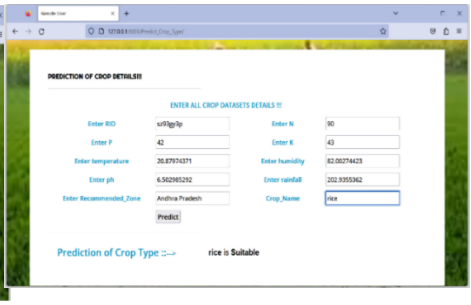

Fig. 3. Result output-2

ANALYSIS:
In this work, the measurements of exactness (ACC), explicitness (S), review (R), accuracy (P), and F1 score were utilized to assess the viability of the element determination and arrangement calculations for expectation.

Analysis of algorithms:
Table: Analysis of algorithms (Accuracy analysis)

Graphs:
The following charts are the graphical representations of our work regarding crop prediction based on soil and environment conditions.

Bar Graph: The visual device of a reference diagram utilizes bars to look at information between classes. It is otherwise called a bar graph or bar outline. A structured presentation might be situated either upward or evenly. It is important to recognise that a bar's value rises with length (Fig. 4).

Model Type	Accuracy (Out of 60)	Precision	Recall	F1-Score	Support
Naïve Bayes	48.40909090909091	0.49	0.42	0.45	222
Support Vector machine	46.81818181818182	0.47	0.45	0.46	222
Logistic Regression	46.81818181818182	0.47	0.46	0.47	222
Decision Tree	46.36363636363636	0.47	0.43	0.45	222
KNeighbours Classifier	53.86363636363637	0.55	0.49	0.52	222
Random Forest	50.0	0.51	0.38	0.43	222

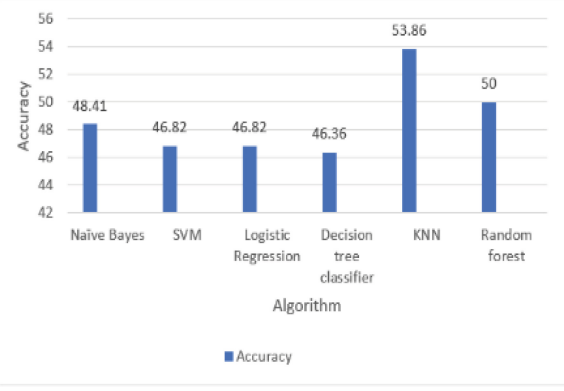

Fig. 4. Bar Graph (Accuracy Analysis)

7 Conclusions and Future Work

Our ranchers might choose the erroneous harvest to plant, which will decrease their pay, since they are at present not really taking advantage of innovation and examination. We have fostered a rancher cordial framework with a graphical UI (GUI) to gauge which harvest would be the most reasonable for a specific land parcel and to give insights regarding the supplements that should be added, the seeds that should be utilized for development, the expected yield, and the market cost. To develop the agrarian business through inventive reasoning, this rouses ranchers to choose the ideal harvest for creation.

By giving GPS directions to a land parcel and accessing the public authority's downpour determining framework, we can expect crops without get-together any extra information. We can likewise make a model to forestall food deficiencies and excesses. Our task just lets about know crop appropriate. In ongoing we need to foster that whether any illnesses might happen, assuming they develop that harvest, and what manure is reasonable for yield.

References

1. Jahan, R.: Applying naïve Bayes classification technique for classification of improved agricultural land soils. Int. J. Res. Appl. Sci. Eng. Technol. **6**(5), 189–193 (2018)
2. Sawicka, B.B., Krochmal-Marczak, B.: Biotic components influencing the yield and quality of potato tubers. Herbalism **1**(3), 125–136 (2017)
3. Sawicka, B., Noaema, A.H., Gáowacka, A.: The predicting the size of the potato acreage as a raw material for bioethanol production. In: Zdunek, B., Olszáwka, M. (eds.) Alternative Energy Sources, pp. 158–172. WydawnictwoNaukowe TYGIEL, Lublin (2016)
4. Sawicka, B., Noaema, A.H., Hameed, T.S., Krochmal-Marczak, B.: Biotic and abiotic factors influencing on the environment and growth of plants. In: Proceedings of the BioróżnorodnośćŚrodowiskaZnaczenie, Problemy, Wyzwania. MateriałyKonferencyjne, Puławy (2017). https://bookcrossing.pl/ksiazka/321192. (in Polish)
5. Myers, R.H., Montgomery, D.C., Vining, G.G., Borror, C.M., Kowalski, S.M.: Response surface methodology: a retrospective and literature survey'. J. Qual. Technol. **36**(1), 53–77 (2004)
6. Muriithi, D.K.: Application of response surface methodology for optimization of potato tuber yield. Amer. J. Theor. Appl. Statist. **4**(4), 300–304 (2015). https://doi.org/10.11648/j.ajtas.20150404.20
7. Marenych, M., Verevska, O., Kalinichenko, A., Dacko, M.: 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. **16**(2), 183–188 (2014)
8. Olędzki, J.R.: Thereporton the state of remotesensing in Poland in 2011–2014. Remote Sens. Environ. **53**(2), 113–174 (2015). (in Polish)
9. Grabowska, K., Dymerska, A., Poáarska, K., Grabowski, J.: Predicting f blue lupine yields based on the selected climate change scenarios. Acta Agroph. **23**(3), 363–380 (2016)
10. Li, D., et al.: Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning. Remote Sens. **13**(16), 3322 (2021). https://doi.org/10.3390/rs13163322
11. Chanamarn, N., Tamee, K., Sittidech, P.: Stacking technique for academic achievement prediction. In: Proceedings of the International Workshop Smart Info-Media Systems, pp. 14–17 (2016)

12. Paja, W., Pancerz, K., Grochowalski, P.: Generational feature elimination and some other ranking feature selection methods. In: Advances in Feature Selection for Data and Pattern Recognition, vol. 138, pp. 97–112. Springer, Cham (2018)
13. Jegadeesan, R., et al.: Stable Route Selection for Adaptive Packet Transmission in 5G-Based Mobile Communications. Wirel. Commun. Mob. Comput. **2022**, Article ID 8009105 (2022). https://doi.org/10.1155/2022/8009105
14. Akshitha, M., Jegadeesan, R., Akshaya, G., Akhilac, P., Pavan Kalyan, M., Sindhusha, G.: Covid-19 future forecasting using supervised machine learning models. Zeichen J. **7**(6), 257–269 (2021). https://doi.org/10.1109/ACCESS.2020.2997311. ISSN 0932-4747
15. Priyavarshini, P., Jegadeesan, R., Vaishnavi, T., Sahithi, K., Shivani, B., Balakishan, P.: Cyber money laundering detection using machine learning. Zeichen J. **7**(6), 231–238 (2021) (2021). 15.10089.ZJ.2021.V7I6.285311.2422. ISSN 0932-4747
16. Jegadeesan, R., Srinivas, D., Umapathi, N., Karthick, G, Venkateswaran, N.: Healthcare chatbot for medical suggestions using artificial intelligence and machine learning. Eur. Chem. Bull. **12**(S3), 6004–6012 (2023). https://doi.org/10.31838/ecb/2023.12.s3.670
17. Jegadeesan, R., Srinivas, D., Umapathi, N., Karthick, G.: Utilizing ensemble learners help prevent unauthorized access into IoT networks. Eur. Chem. Bull. 12 (S3), 5994–6003 (2023). https://doi.org/10.31838/ecb/2023.12.s3.669
18. Jegadeesan, R., Vasania, I., Rehaan, B.U., Goyal, A.: Implications of Machine Learning for autonomic network operation and management. J. Xidian Univ. (Sci. Technol. Edn.), **15**(8), 307–315 (2021). https://doi.org/10.37896/jxu15.8/031. ISSN1001-2400