JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

OPTIMIZED CROP PREDICTION USING MACHINE LEARNING ALGORITHMS

A.Krishnaveni*[1], K.Vanitha *[2], K.M.Buvaneswari*[3]
*[1&*2]Assistant Professors, Department of Computer Science
*[3]Assistant Professor, Department of Home Science
*[1&*2*3]Thassim Beevi Abdul Kader College for Women, Kilakarai, Ramanathapuram, Tamilnadu, India
Email id: *[1]drakrishnaveni@gmail.com, *[2]rkvanithamca@gmail.com,*[3]srithaguna@gmail.com

Abstract: Prompt and optimized classification of soil type with the suitable crop is a crucial problem-solving in the agriculture field. All the crofters needed beneficiary outcomes through the cultivation of crops done in peculiar soil. Choosing the appropriate soil type and the apt crop is successful only when the crofter has the awareness and prefers the suitable methods. Soil consists of nutrients, which are used by the plants to grow. Variety of soil and various properties of those soils gave an opportunity to the crofter to choose one particular soil with different types of crop cultivation. Particular land type is the primary concern for satisfactory outputs from crop cultivation. To increase crop cultivation every farmer should be aware to make decision-making about soil and crop. This can be done by first analyzing the soil then classifying it into different soil groups. Based on these soil groups, one can decide which crop is best suited and is beneficial. Based on the crofter skill and knowledge, the land was chosen for soil classification. This work proposes a novel approach to select appropriate features from a data set for crop prediction. The experimental results show that the Naïve Bayes technique helps accurately predict a suitable crop. The performance of the proposed technique is evaluated by various metrics such as accuracy (ACC), precision, recall, specificity, and F1 score, mean absolute error, and time is taken. From the performance analysis, it is justified that the proposed technique performs better than other methods.

***Index Terms* – Optimization, Soil, Crop, Classification, Accuracy, Naïve Bayes, F1 measure, Recall, Precision**

I. INTRODUCTION

For a country, one of the main parts of its development rotates around its capability to create food. Machine learning (ML) approaches are utilized in many fields, going from stores to assess the conduct of clients to the expectation of clients' telephone use. AI is likewise being utilized in horticulture for a very long time. The crop yield forecast is one of the difficult issues in accuracy horticulture, and many models have been proposed and approved up until now. This issue requires the utilization of a few datasets since crop yield relies upon a wide range of variables like environment, climate, soil, utilization of manure, and seed assortment. This demonstrates that harvest yield forecast is anything but an insignificant errand; all things considered, it comprises of a few convoluted advances. These days, crop yield expectation models can assess the real yield sensibly, yet a superior execution in yield forecast is as yet alluring.

Machine learning, which is a part of Artificial Intelligence (AI) zeroing in on learning, is a commonsense methodology that can give better yield expectation dependent on a few highlights. AI (ML) can decide examples and connections and find information from datasets. The models should be prepared to utilize datasets, where the results are addressed dependent on previous experience. The prescient model is assembled utilizing a few elements, and thusly, boundaries of the not really settled utilizing verifiable information during the preparation stage. For the testing stage, part of the chronicled information that has not been utilized for preparing is utilized for the presentation assessment reason.

1.1 REVIEW OF LITERATURE

Many of the researchers examined machine learning algorithms that are very useful to detect the soil types and choose the crop based on the types of the land. Unmistakably, a rancher is the best leader in the determination and development of harvests. Today, notwithstanding, cultivar expectation is done physically in research centers, and ranchers need the assistance of specialists to decide the most appropriate yield/s for a particular land parcel. The specialists gather soil tests from a specific piece of land and test them in the research facility, following which they submit ideas on the best yield/s to be raised. Expectation sets aside time, and choosing the most reasonable yield/s is a perplexing errand in agribusiness. Manual expectation has generally fizzled, inferable

from climatic changes and natural factors that influence crop development. Exact expectations of reasonable yields for development further develop creation levels. Crop expectation ascribes are characterized by different factors like genotype, environment, and the collaborations between the two. An exact yield forecast needs a principal comprehension of the useful connection among development and intuitive variables like the genotype and environment. Further, it requires both definite datasets and proficient calculations to analyze these connections. Advocated by these realities, AI procedures are utilized in this review to anticipate the most reasonable yield for a particular stretch of land, and this method is great for considering factors like dirt and ecological conditions. Various related examinations are talked about in this survey.

Sanmay Das [1] talked about the advantages and disadvantages of the channel and covering strategies, and executed another crossbreed highlight determination approach utilizing the helping method. The investigations were done utilizing genuine world datasets from the University of California, Irvine (UCI) vault. The outcomes demonstrated that the proposed technique is a lot quicker than the covering strategy. Huan Liu and Lei Yu [2] assessed the current component choice calculation for grouping and bunching procedures. Along these lines, a middle-of-the-road step on a bringing-together stage was proposed in their work.

Al Maruf et al. [3] showed the prevalence of the gapped k-mer creation and converse supplement elements of the k-mer structure over different organizations. The Support Vector Machine (SVM) with the Radial Basis Function (RBF) part was utilized as an order calculation. Contrasted with different methodologies, the iRSpot-SF performs extensively better compared to the Matthews, with a relationship coefficient and affectability of 69.41% and 84.57% and it has 84.58% exactness. Jana Novovicova et al. [6] proposed an element choice strategy with no hunt system and one most appropriate for multimodal information. Isabelle Guyon and Andre Elisseeff [10], likewise momentarily talked about a component choice technique dependent on the channel and covering draws near and, furthermore, characterized include positioning and multivariate element choice. Jia-You Hsieh et al. [12], in their review, talked about Rice Blast Disease (RBD). The Recursive Feature Elimination (RFE) calculation with the Auto-Sklearn was utilized to choose key elements affecting RBD. The point of their work was to fabricate a model as a notice component for RBD.

Ron Kohavi and George H. John [13] analyzed the covering and enlistment strategies without highlighting subset choice and continued from there on to contrast them with Relief, a channel strategy with include subset determination. The qualities and shortcomings of the covering approach were talked about, and a progression of further developed plans was shown. Isabelle Guyon et al. [14], carried out a Support Vector Machine (SVM) strategy dependent on the Recursive Feature Elimination (RFE) for quality choice. Of the various strategies used to choose highlights, the RFE is a recently evolved strategy that chooses highlights for little example order issues.

II. METHODOLOGY

The proposed methodology involves the following steps to accomplish a particular aim: • Dataset collection • Pre-processing • Classification • Prediction • Results

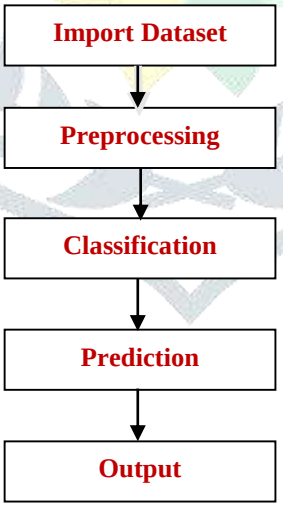

Figure 1: System architecture for classification

A. Dataset assortment: In this segment, the WEKA Datamining instrument examines the proposed strategies for crop yield forecast. For the exploratory reason, distinctive soil boundaries like temperature, Humidity, precipitation, pHare considered. The dataset with these boundaries and its qualities are utilized in the proposed strategy to foresee the legitimate harvest yield. The dataset contains 3100 occasions and 5 ascribes. The informational collection is isolated into preparing and testing the dataset. 70% of information has been utilized for preparing the model and 30% of the information utilized for testing the exhibition of the model. The upsides of these are taken as contributions for the calculation [7].

B. Pre-processing: For successful completion of a model, a huge set of data is required. The data that is collected from real-world might be in raw format. It may contain some missing values, inconsistent and noisy values. In this step, such redundant values should be filtered. The data is made normalized. Discretization is applied for numeric fields. Feature selection was applied to select only necessary attributes for classification.

C. Classification: For a fruitful culmination of a model, a colossal arrangement of information is required. The information that is gathered from the genuine world may be in a crude arrangement. It might contain some missing qualities, conflicting and loud qualities. In this progression, such repetitive qualities ought to be separated. The information is made standardized. Discretization is applied for numeric fields. Highlight determination was applied to choose just vital qualities for arrangement.

D. Prediction: The introduction of grouping calculation related dependent on exactness and execution examination and will give an idea of yield to the ranchers to develop in a specific soil type.

E. Results: The end result gives the suggestion of crops.

III. MODELS AND ANALYSIS

CLASSIFICATION ALGORITHMS: Some of the classification algorithms used in this research work are Naïve Bayes, J48, and Random Forest.

A.Naive Bayes: Naive Bayes is a sort of classifier, which utilizes the Bayes Theorem. It predicts participation probabilities for each class, for example, the likelihood that a given record or information point has a place with a specific class. The class with the most elevated likelihood is considered the most probable class.

B. J48: This algorithm depends on a hierarchical system, a recursive gap, and vanquish technique. You select which property to part on at the root hub, afterward you make a branch for every conceivable trait esteem, and that divides the examples into subsets, one for each branch that stretches out from the root hub.

C.Random Forest: A Random Forest is an AI method that is utilized to take care of relapse and order issues. It uses troupe realizing, which is a method that consolidates numerous classifiers to give answers for complex issues. An irregular woods calculation comprises numerous choice trees.

3.1Proposed model

In this section, the WEKA Datamining tool analyzes the proposed methods for crop yield prediction is discussed. For the experimental purpose, different soil parameters such as temperature, humidity, rainfall, pHare considered. The dataset with these parameters and their values are used in the proposed method to predict the proper crop yield. The dataset contains 3100 instances and 5 attributes. The data set is divided into training and testing datasets. 70% of data has been used for training the model and 30% of the data used for testing the performance of the model.

Table 1: Attributes and its Data types

Attribute	Data type
Temperature	numeric
Humidity	numeric
pH	numeric
Rainfall	numeric
Crop	Nominal

Replacing NULL values:

Information is seldom spotless and regularly you can have bad or missing qualities. Distinguish, imprint, and handle missing information when creating AI models to get the absolute best exhibition.

Feature Selection using wrapper: The algorithm of feature selections is used to select a subset of the features that are considered to be important so that the redundant data can be removed. The aim of this research work is achieved through unwanted features identification and removal of unrelated features. Retrieval of land and types of soil is easily stored without any irrelevant details and occupied the allocated storage place. A feature selection algorithm selects a subset of vital features and removes superfluous, unrelated, and raucous features for simpler and more precise data illustration. The classifier subset evaluator method was used for attribute evaluator. Best First search method was applied.

Discretization: Information discretization is a strategy for changing over credits upsides of ceaseless information into a limited arrangement of spans with the least information misfortune. An **f-measure** that consolidates accuracy and review is the consonant mean of **accuracy** and review. Discretization further develops **f-measure**.

Result shown that the proposed approach is better compared to the run of the mill Naïve Bayes arrangement calculation with characterization exactness, accuracy, review, f measure and time taken to assemble a model. Supplanting NULL qualities, Feature determination and Discretization further develops execution of forecast.

IV. RESULTS AND DISCUSSION

4.1 Results of Descriptive classifiers

The proposed model depends on soil and yield data set. A few AI calculations are utilized to order the dirt sort. For a specific soil type, an appropriate harvest is proposed. From the exploratory outcome, we see that Naive Bayes has gotten the greatest exactness and takes less time contrasted with other algorithms. This paper utilizes a methodology for further developing the grouping exactness, review, accuracy, f-measure, and decreasing the time needed to execute.

Table 2: Comparative study of different classification algorithms

Classifier	Time is taken (Sec)	Accuracy (%)
J48	0.3	91.96
Random Forest	2.58	91.09
Naïve Bayes	0.01	93.45

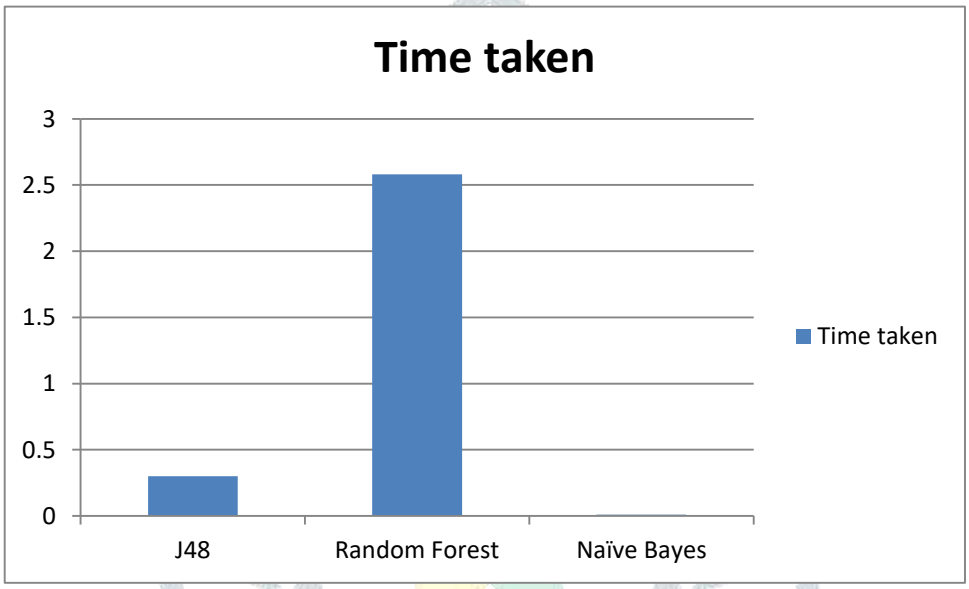

Figure2: Time is taken to build the model

Table 3. Performance measures of Naïve Bayes algorithm

Parameters	Conventional Method	Proposed System
Time is taken (Sec)	0.1	0.03
F-Measure	0.834	0.931
Precision	0.836	0.921
Recall	0.835	0.919

The proposed naïve Bayes model has high accuracy and takes less time to the classification of soil type and crop prediction with high precision and recall value given in Table 3.

4.2 CONCLUSION

Classification and prediction of soil types and crop cultivation is a challenging problem in the agriculture field. Without food, no one can survive in this world. So satisfying the basic need of human being the productivity of crop management might be increased. In this research, the proposed model was helpful in increasing the accuracy of soil classification and crop prediction using machine-learning algorithms. Utilized classifiers are random forest (RI), Naïve Bayes (NB) and J48. The exactness of the

current model is most extreme than the current models. The proposed approach is better compared to the common Naïve Bayes with classification accuracy and time is taken to build a model. Discretization, appropriate Feature Selection, and Replacing NULL values, are very useful to improve the performance of prediction.

In the future, reasonable manures are recommended for the good development of the harvest cultivated. The current model manages with available old data whereas the future model contains real-time information that is straightforwardly gotten from the farming area that is set with sensors. The sensors detect soil fertility and different minerals contained in the soil.

V. Acknowledgment

The authors would be thankful for the publicly available dataset to soil classification and crop prediction.

References

[1] Meghana S, Dharani N V, “Crop and Yield Prediction with Fertilizer Estimation”, International Journal of Emerging Technologies and Innovative research, Vol.8, Issue 7, pp.428-431 (2021)

[2] V. Vanarase, V. Mane, H. Bhute, A. Tate and S. Dhar, "Crop Prediction Using Data Mining and Machine Learning Techniques," Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1764-1771, (2021), doi: 10.1109/ICIRCA51532.2021.9544724.

[3] Bhatti, U. and Hanif. M. 2010. Validity of Capital Assets Pricing Model.Evidence from KSE-Pakistan.European Journal of Economics, Finance and Administrative Science, 3 (20). P. Malik, S. Sengupta and J. S. Jadon, "Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), (2021), pp. 1004-1007, doi: 10.1109/Confluence51648.2021.9377147.

[4] G. Mariammal, A. Suruliandi, S. P. Raja and E. Poongothai, "Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique With Various Classifiers," in IEEE Transactions on Computational Social Systems, vol. 8, no. 5, pp. 1132-1142, Oct. (2021), doi: 10.1109/TCSS.2021.3074534.

[5] C. P. Wickramasinghe, P. L. N. Lakshitha, H. P. H. S. Hemapriya, A. Jayakody and P. G. N. S. Ranasinghe, "Smart Crop and Fertilizer Prediction System," 2019 International Conference on Advancements in Computing (ICAC), (2019), pp. 487-492, doi: 10.1109/ICAC49085.2019.9103422.

[6] S. Bhanumathi, M. Vineeth and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0769-0773, doi: 10.1109/ICCSP.2019.8698087.

[7] S. Afrin et al., "Analysis of Soil Properties and Climatic Data to Predict Crop Yields and Cluster Different Agricultural Regions of Bangladesh," 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), 2018, pp. 80-85, doi: 10.1109/ICIS.2018.8466397.