



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Prediction of Crop yield and Suitable Crop

Snehal Guruphale, Manisha Bidve

P.G. Student, Department of Computer Engineering, M. S. Bidve Engineering College, Latur, Maharashtra, India

Assistant Professor, Department of Computer Engineering, M.S. Bidve Engineering College, Latur, Maharashtra, India

ABSTRACT: This research efforts to enhance the precision of crop prediction by leveraging the distinct characteristics of the agricultural environment. Through a comprehensive analysis of diverse environmental factors, such as climate, soil attributes, and weather conditions, this study aims to identify key features crucial for accurate crop yield estimation. Multiple feature selection techniques will be employed to discern the most influential factors. The predictive models will then be constructed using various classifiers, allowing for a comparative evaluation of their performance. The findings of this research contribute valuable insights into optimizing crop prediction models and advancing sustainable agricultural practices.

KEYWORDS: Feature Selection methods, Machine Learning, Performance Metrics, Crop prediction

I. INTRODUCTION

Data analysis is the process of cleansing and modeling data to obtain meaningful information and conclusions. It involves analyzing, extracting, and predicting relevant facts from massive amounts of data to discover patterns. Companies employ this approach to extract relevant information from raw client data. This technique can also be applied in agriculture. Most farmers expect increased profits in the upcoming harvest period based on their long-standing experience in the field with certain crops. However, crop production is occasionally lower than planned due to improper irrigation or crop choices.

Farmers are adamant about the need for an effective framework for forecasting and enhancing crop growth. The majority of agricultural research focuses on biological mechanisms for detecting and improving crop growth. Crop output is generally determined by factors such as crop variety, seed type, and environmental factors such as sunlight (temperature), soil pH, water pH, a sufficient number of plants, and moisture. By assessing the land and atmosphere in a certain place, the optimum crops for higher harvest rates and net crop production can be projected. This forecast will be beneficial to farmers.

Choose crops for your farm based on soil type, temperature, water level, space depth, pH soil, season, fertiliser, and months.

II. LITERATURE SURVEY

K. Rajini et al.: In this paper, by merging the IB1 learner and the AIDE updatable learner, a novel ensemble of heterogeneous incremental classifiers has been presented. The methodology was primarily created to accurately predict ART outcomes. For the ART data set used in the study, the suggested model accomplishes this goal with an ROC area of 94.1. Several ART data sets and various test options are used to evaluate the model's efficacy. For the same data set, the proposed ensemble outperforms the others. To determine if the suggested model is generalizable to other domains, it is compared with additional benchmark data sets and shown to perform better.[1]

J.-Y. Hsieh et al.: In this paper, we attempted to solve the prediction problem by using machine learning and neural network methods. The classification model was built based on climate and historical data. This classifier's projected results have a 72 percent accuracy rate. According to this study, increasing the number of microclimate parameters can improve prediction accuracy. [2]

J. Camargo et al.: This work highlights the importance of feature selection in increasing accuracy levels to above 95%. It was concluded that, on average, accuracy scores were 3.8% lower without feature removal. Feature selection also offers the added benefit of enabling faster real-time computations with fewer features. Chow-Liu trees were

identified as a feature selection method that allows for the rapid selection of a robust set of features with a minimal number of iterations and accuracy comparable to that of a forward selection strategy.[3]

R. Rajashekar Pullanagari et al.: In this paper, the viability of employing RF-RFE to integrate hyperspectral, topographical, and soil data to obtain pasture quality parameters of heterogeneous pasture was investigated. Because many environmental and management factors influence pasture quality, our findings revealed that combining hyperspectral data with commonly accessible environmental characteristics (elevation, slope angle, and soil type) enhanced prediction accuracy compared to using hyperspectral data alone. This finding also revealed that by selecting the most sensitive factors across the spectrum and environmental data, RF-RFE significantly improved estimations of pasture quality, with RPD values ranging from 2.11 to 2.35. Elevation, slope, and soil type were identified as key determinants in predicting crude protein (CP), while the same variables, except elevation, were significant in predicting metabolizable energy (ME).[4]

F. Balducci et al. present research that introduces practical, low-cost, and simple-to-implement tasks to help an agricultural company increase productivity while also deepening the study of the smart farm model. Technological progress in a field that requires control and optimization can significantly contribute to conserving natural resources, adhering to business and international laws, meeting consumer needs, and pursuing economic profits. Machine learning and more traditional statistical techniques were employed to analyze three separate data sources, with a special focus on the IoT sensors dataset. The first task demonstrates that a neural network model can forecast total apple and pear crops from the Istat dataset with a success rate of nearly 90%. Meanwhile, the second task shows that polynomial predictive and regression models are better suited for the CNR scientific data due to the nature of the dataset.[5]

M. Lango et al.: Two sorts of methodological contributions are presented in our work. To begin, we propose incorporating a random set of attributes into this ensemble. This approach enhances G-mean and sensitivity measurements for greater dimensional complex datasets, according to experiments. We've also noticed that, unlike the original RB-Bag, it increases the diversity of component classifiers, and employing a larger number of components improves classification. Second, for several imbalanced classes, we have presented an extension of Roughly Balanced Bagging that uses the multinomial distribution to estimate the cardinalities of class examples in bootstrap samples. The under-sampling version of our proposed Multiclass RB Bag enhances G-mean and is better than the oversampling variation and simpler multi-class classifiers, according to testing with synthetic and real datasets. [6]

M. Gopal et al.: In this paper, five feature selection techniques are employed to select the features: sequential forward feature selection, sequential backward feature removal, correlation-based feature selection, Random Forest Variable Importance, and Variance Inflation Factor. Important features are chosen by feature selection algorithms based on their selection criteria. The performance of these algorithms is evaluated using the RMSE, MAE, R, and RRMSE metrics. MLR, ANN, and MSP present the results of the feature selection. The forward feature selection approach proves effective for better prediction, identifying the best features for the paddy dataset in the specified research region as Area, Open Wells, Tanks, Maximum Temperature, and Canal Length. [7]

Dipika H. Zala et al. present an effective Bootstrap aggregating technique for agricultural yield prediction. The accuracy of crop yield prediction will be improved by using a voting cast of the bagging technique with various parameters. The application of crop yield prediction is an important concern in data mining. It can be easily decided which climate or weather is suitable for a crop. [8]

Bahl et al. examined the suitability of an unsupervised strategy, based on PCA and KNN, as well as a supervised approach, based on RFs with and without prior feature selection, for predicting NM toxicity. The reduced RF model, following backward RFE, demonstrated the best performance in terms of the balanced accuracy of the prediction model. In every example, variable selection based on the MDA produced equal or better outcomes than those based on Gini importance. In both the full and reduced models following RFE, the three most critical parameters—zeta potential, redox potential, and dissolution rate—were among the highest-ranking variables in both unsupervised and supervised analyses.[9]

P. S. Maya Gopal et al. discuss in their paper how the Boruta algorithm was utilized to identify the most essential features for predicting crop yields. The critical factors determined include crop area, canal length, number of open wells, number of tube wells, number of tanks, maximum temperature, average temperature, nitrogen, phosphorus, and

potassium fertilizers, sun radiation, and seed rate. These characteristics were incorporated into the MLR model to evaluate its accuracy. With these features, the model achieved an accuracy of 84%. [10]

III. PROPOSED SYSTEM

The proposed approach is especially beneficial in agriculture. Accurate prediction of agricultural production can assist governments, authorities, and farmers in making strategic policy decisions. The emphasis on increasing production without addressing the environmental implications of input resources has led to environmental degradation. This technique identifies the most appropriate crop properties for prediction.

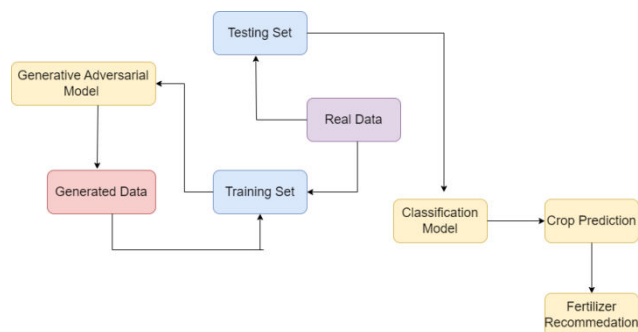


Figure. System Architecture

Algorithm

Random Forest:

Random forests are a machine learning algorithm. The basic premise of the algorithm is that building a small decision tree with few features is a computational process. The algorithm works as follows: for each tree in the forest, we select a bootstrap sample from S where $S(i)$ denotes the i th bootstrap. We then learn a decision tree using a modified decision-tree learning algorithm. The algorithm is modified as follows: at each node of the tree, instead of examining all possible feature-splits, we randomly select some subset of the features $f \subset F$:

Where F is the set of features. The node then splits on

The best feature is f rather than F . In practice f is much, Much smaller than F . Deciding on which feature to split is oftentimes the most computationally expensive aspect of decision tree learning. By narrowing the set of features, we drastically speed up the learning of the tree.

IV. RESULT AND DISCUSSION

For the experimental result evaluation, we have the notation as follows:

TP: True positive (correctly predicted number of instances)

FP: False positive (incorrectly predicted number of instances),

TN: True negative (correctly predicted the number of instances as not required)

FN false negative (incorrectly predicted the number of instances as not required),

Based on this parameter, we can calculate four measurements

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$F1\text{-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

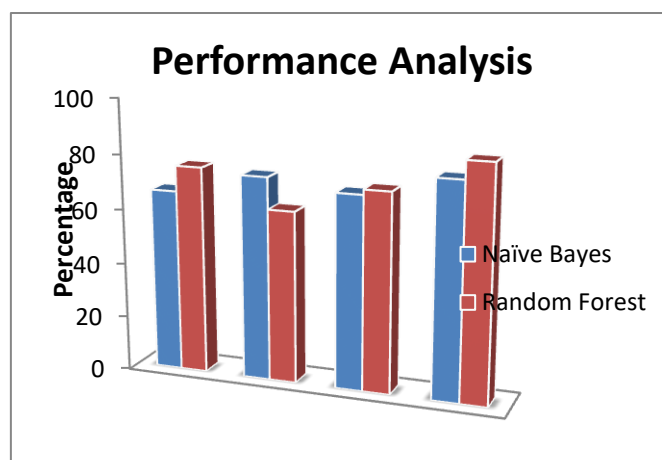


Figure 2. Performance analysis

V. CONCLUSION

In amalgamating these diverse perspectives, our study aims to contribute not merely as an addition to the existing body of knowledge but as a catalyst for advancement. As we delve into the dynamic interplay of feature selection techniques and classifiers, we embark on a journey to refine predictive models that resonate with the intricacies of precision agriculture. Through iterative analysis, we aspire to offer practical solutions, empowering farmers with actionable insights for sustainable and efficient crop management. The echoes of existing works reverberate in our methodologies, and it is with a commitment to building upon these foundations that we step into the realm of precision agriculture, where the future of crop prediction unfolds.

REFERENCES

1. K. Ranjini, A. Suruliandi, and S. P. Raja, "An ensemble of heterogeneous incremental classifiers for assisted reproductive technology outcome prediction," *IEEE Trans. Comp. Social Syst.* early access, Nov. 3, 2020, doi: 10.1109/TCSS.2020.3032640. [1]
2. J.-Y. Hsieh, W. Huang, H.-T. Yang, C.-C. Lin, Y.-C. Fan, and H. Chen, "Building the Rice Blast Disease Prediction Model based on Machine Learning and Neural Networks," *Easy Chair World Sci.*, vol. 1197, pp. 1–8, Dec. 2019.[2]
3. J. Camargo and A. Young, "Feature selection and non-linear classifiers: Effects on simultaneous motion recognition in the upper limb," *IEEE Trans. Neural Syst. Rehab il. Eng.*, vol. 27, no. 4, pp. 743–750, Apr. 2019.[3]
4. R. RajashekerPullanagari, G. Kereszturi, and I. Yule, "Integrating airborne hyperspectral, topographic, and soil data for estimating pasture quality using recursive feature elimination with random forest regression," *Remote Sens.*, vol. 10, no. 7, pp. 1117–1130, 2018.[4]
5. F. Balducci, D. Impedov o, and G. Pirlo, "Machine learning applications on agricultural datasets for smart farm enhancement," *Machine*, vol. 6, no. 3, pp. 38–59, 2018.[5]
6. M. Lango and J. Stefanowski, "Multi-class and feature selection extensions of roughly balanced bagging for imbalanced data," *J. Intel. Inf. Syst.*, vol. 50, no. 1, pp. 97–127, 2018[6]
7. M Gopal P S and B. R, "Selection of important features for optimizing crop yield prediction," *Int. J. Agri. Environ. Inf. Syst.*, vol. 10, no. 3, pp. 54–71, Jul. 2019.[7]
8. D. H. Zala and M. B. Chaudhri, "Review on the use of BAGGING technique in agriculture crop yield prediction," *Int. J. Sci. Res. Develop.*, vol. 6, no. 8, pp. 675–677, 2018.[8]
9. Bahl et al., "Recursive feature elimination in random forest classification supports nanomaterial grouping," *NanoImpact*, vol. 15, Mar. 2019, Art. no. 100179.[9]
10. P. S. Maya Gopal and R. Bhargavi, "Feature selection for yield prediction in Boruta algorithm," *Int. J. Pure Appl. Math.*, vol. 118, no. 22, pp. 139–144, 2018.[10]



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details