iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.47649

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 II Feb 2023- Available at www.ijraset.com

Survey on Novel Approach for Crop Yield Prediction Using Machine Learning

Aditya Kamble[1], Kunal Rathod[2], Poonam Hake[3]
[1, 2, 3]*Smt. Kashibai Navale College of engineering, Pune*

Abstract: ***Agriculture relies heavily on the ability to predict crop yields. When it comes to crop yield, there are a number of factors at play. This research is focused on developing cost effective methods for predicting crop yields using available parameters like irrigation, fertilizer, and temperature. Sequential forward FS, sequential backward elimination FS, correlation-based FS, random forest variable importance, and the variance inflation factor algorithm are among the five Feature Selection (FS) algorithms discussed in this study. In general, machine learning techniques are well-suited to a specific region, so they greatly assist farmers in predicting crop yields. Crop prediction can be improved by using a new FS technique called modified recursive feature elimination (MRFE). With the help of a ranking algorithm, the MRFE technique identifies and prioritizes the most important features in a dataset.***

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

I. INTRODUCTION

While there have been recent advancements in agricultural crop forecasting, using a variety of technology resources, techniques, and procedures is still a difficult task to accomplish. Research in agricultural management is focused on developing cost-effective algorithms for predicting crop production based on available data, such as irrigation information, fertiliser details, and temperature. A large crop prediction data set necessitates the selection of critical traits that help identify crops that are suitable for specific land areas. The process makes use of feature selection approaches.

Historical data on crop production and weather data can be used to predict crop yield. Year, area, production, and yield are all examples of variables that could be included in a crop production dataset. Minimum temperature, maximum temperature, average temperature, precipitation, evapotranspiration, and reference crop evapotranspiration are all examples of weather variables that can be included in a dataset. These are the most important parameters in a weather dataset for predicting crop yields, but there may be others in a weather dataset.

A comprehensive picture of paddy crop output can be obtained using feature selection algorithms that identify relevant paddy field circumstances (features). Predictive models in the form of expert systems are being developed by data analysts in order to improve agricultural yield while taking into account environmental factors such as soil quality, irrigation, and land use. Crop yields are estimated using machine learning (ML) in the majority of existing systems, but little has been done to predict crop yields based on soil and environmental factors.

The features collected are fed into k-nearest neighbour (kNN), Naive Bayes (NB), decision trees (DT), support vector machines (SVM), random forests (RF), and bagging classifiers in order to predict an appropriate crop and evaluate the FS process' effectiveness. More than one method of crop prediction has been used in this study (such as the SVM and the kNN), but the most common method is the use of a single prediction model (such as the SVM). Prediction features are unique to each algorithm. For crop prediction, a suitable classifier must be found to work with the FS approach. According to soil and environmental conditions, a permutation crop data set can be used to select the most appropriate key features for feature elimination (MRFE). Because the data set does not need to be updated with each iteration, the algorithm takes less time to run. The Work of MRFE It's more effective than other FS methods. The Bagging Classifier and other Benchmark Data Sets: To ensure that the proposed MRFE approach was applicable to data sets other than crop-related data sets, the UCI Repository was searched for non-crop data sets.

II. LITERATURE SURVEY

Sr no	Author Name	Year of publications	Features and Techniques	Advantages
1	M Gopal P S and B. R	2019	Sequential forward feature Selection, Sequential backward	.

©IJRASET: All Rights are Reserved

			feature removal, correlation based feature selection, random forest Variable Importance, and Variance Inflation Factor. RMSE, MAE, R, and RRMSE metrics.	• To use feature selection algorithms to identify significant paddy field conditions (features) in order to provide a comprehensive view of paddy crop yield. • Forward feature selection is considered to be better than backward elimination algorithm in terms of time complexity. It has an accuracy rate of 84% when all of the features are included in the model. • A better prediction can be made by using the forward feature selection approach.
2	Dipika H. Zala, M.B. Chaudhari	2018	Data Mining, Bootstrap Aggregating Technique, Bagging technique.	In statistical classification and regression, bootstrap aggregation (bagging) is a machine learning ensemble meta-algorithm that aims to increase the stability and accuracy of machine learning algorithms used. BAGGING is a more accurate method than others..
3	A. Bahl et al	2019	Machine learning techniques, PCA – Principle component analysis, PC – Principle component, RF – Random forest, RFE – Reverse feature elimination, kNN – k-nearest neighbors, STIS – Short-term inhalation study, LOAEC – Lowest observable adverse effect concentration, MDA – Mean decrease in accuracy	To reduce the dimensionality of the input feature space, PCA uses linear combinations of the original input variables that show the most variation in a dataset. • Machine learning techniques are well-suited to the selection and ranking of parameters in a data-driven environment.
4	P. S. Maya Gopal and R. Bhargavi	2018	Boruta algorithm,MLR	When predicting crop yield, the Bourta algorithm selects the most important features.
5	K. Ranjini, A. Suruliandi, and S. P. Raja	2020	Machine Learning(ML),Assisted reproductive technology (ART)	When compared to other ensemble models, the proposed model has fewer errors and can

				be used to predict the outcome of ART. • The proposed ensemble outperforms all alternatives on a level of average performance.
6	J.-Y. Hsieh, W. Huang, H.-T. Yang, C.-C. Lin, Y.-C. Fan, and H. Chen	2019	Machine Learning, Neural Network	• Using neural networks, the authors develop a model that predicts the effect of climate on anthracnose severity. • Authors use weather data and previous crops to analyse climatic factors.
7	J. Camargo and A. Young	2019	sequential forward feature selection algorithm,	• In order to get reasonable results and short computation times, this algorithm uses a sequential forward feature selection approach that evaluates each feature in turn. • It was possible to achieve high levels of accuracy (>95 percent) for 33 different movement types using a combination of feature selection.
8	R.Rajasheker Pullanagari, G.Kereszturi, and I. Yule	2018	RF-RFE Method	• . A large-scale study of the proposed approach is required in areas with a wide range of soil types. The final accurate pasture quality spatial maps allow farmers to make the best agronomic decisions. • We found that RF–RFE was an efficient feature selection method, far better than traditional methods, for analysing high dimensional data and selecting important spectral and environmental features sensitive to pasture quality.
9	F. Balducci, D. Impedovo, and G. Pirlo	2018	National Research Council (CNR) scientific dataset, Istat statistical dataset, and the industrial Internet of Things (IoT) Sensors dataset	• . IoT sensors datasets, as well as other data sources, were analysed using machine learning and traditional statistical methods.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

10	M. Lango and J. Stefanowski	2018	Class imbalance-Roughly balanced bagging · Types of minority examples-Feature selection, Multiple imbalanced class.	• The authors distinguish approaches that use re-sampling techniques in ensembles that are not typical adaptive. • The SMOTE method, or varying the oversampling ratios in each bootstrap, is referred to by the authors as a way to increase ensemble diversity.

III. THE EXISTING SYSTEM

1) In existing System, the RFE method is a wrapper-type FS method that searches for a subset of features, starting with all features in the training data set and successfully deleting features until only a small number remain.
2) The RFE technique ranks acceptable traits in order of relevance, rejecting the ones that aren't as vital.
3) This method needs an iterative process for data set updating in the feature elimination process.
4) The most challenging aspect of the RFE is updating the data set, and the greatest time is spent removing weak features.

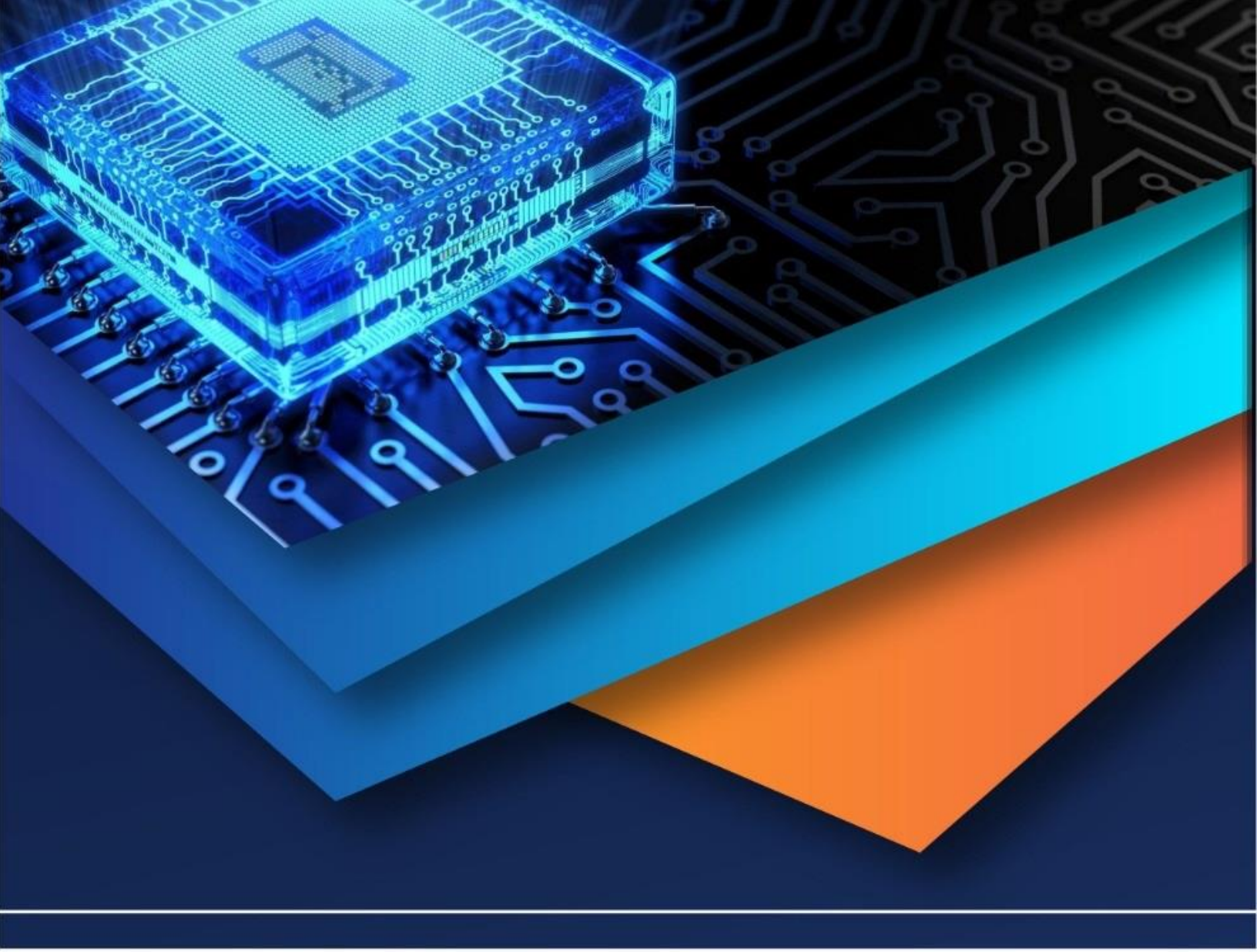
IV. CONCLUSION

In this paper, the suggested MRFE approach is applicable to both crop and non-crop data sets. When compared to previous strategies, the MRFE uses permutation and ranking to choose the most suited characteristics with improved prediction ACC in the shortest amount of time. MRFE technique in predicting the most suitable crops for cultivation using classification algorithms.

REFERENCES

[1] M Gopal P S and B. R, “Selection of important features for optimizing crop yield prediction,” Int. J. Agricult. Environ. Inf. Syst., vol. 10, no. 3, pp. 54–71, Jul. 2019.
[2] D. H. Zala and M. B. Chaudhri, “Review on use of BAGGING technique in agriculture crop yield prediction,” Int. J. Sci. Res. Develop., vol. 6, no. 8, pp. 675–677, 2018.
[3] A. Bahl et al., “Recursive feature elimination in random forest classification supports nanomaterial grouping,” NanoImpact, vol. 15, Mar. 2019, Art. no. 100179.
[4] 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.
[5] K. Ranjini, A. Suruliandi, and S. P. Raja, “An ensemble of heterogeneous incremental classifiers for assisted reproductive technology outcome prediction,” IEEE Trans. Comput. Social Syst.early access, Nov. 3, 2020, doi: 10.1109/TCSS.2020.3032640
[6] 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.
[7] J. Camargo and A. Young, “Feature selection and non-linear classifiers: Effects on simultaneous motion recognition in upper limb,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 27, no. 4, pp. 743–750, Apr. 2019.
[8] R. Rajasheker Pullanagari, 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.
[9] F. Balducci, D. Impedovo, and G. Pirlo, “Machine learning applications on agricultural datasets for smart farm enhancement,” Machine, vol. 6, no. 3, pp. 38–59, 2018.
[10] M. Lango and J. Stefanowski, “Multi-class and feature selection extensions of roughly balanced bagging for imbalanced data,” J. Intell. Inf. Syst., vol. 50, no. 1, pp. 97–127, 2018

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)