IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, April 2023

Survey on Novel Approach for Crop Yield Prediction using Machine Learning

Aditya Kamble[1], Patil Dhairyasheel[2], Kunal Rathod[3], Poonam Hake[4], Prof. Dhanashri Nevase[5]
Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India[1,2,3,4,5]

Abstract: *Predicting crop yields is crucial to agriculture. Crop production is affected by a number of factors. The goal of this study is to provide low-cost techniques for forecasting agricultural yields utilising existing variables like irrigation, fertiliser, and temperature. The five Feature Selection (FS) algorithms described in this article are sequential forward FS, sequential backward elimination FS, correlation-based FS, random forest variable significance, and the variance inflation factor algorithm. Machine learning techniques are typically well adapted to a particular area, therefore they substantially help farmers forecast agricultural output. With a novel FS method termed modified recursive feature removal, crop prediction can be improved (MRFE). The MRFE approach locates and ranks the most crucial characteristics in a dataset with the use of a ranking algorithm.*

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

I. INTRODUCTION

Although recent improvements in crop forecasting for agriculture, employing a range of technical resources, strategies, and procedures is still a challenging undertaking to complete Agriculture's objective. The goal of management research is to develop algorithms that can precisely predict agricultural yield using data on irrigation, fertiliser, and temperature. The identification of crucial qualities that aid in identifying crops that are suited for certain land locations is required in order to make use of a big crop prediction data collection. Approaches for feature selection are used in the process.

Crop output may be predicted using meteorological information and historical crop production data.A agricultural production dataset could contain variables such as year, area, production, and yield, for example. Weather variables that can be included in a dataset include minimum temperature, maximum temperature, average temperature, precipitation, evapotranspiration, and reference crop evapotranspiration. There may be more variables in a meteorological record, but these are the most crucial ones for predicting agricultural yields.

Using feature selection algorithms that recognise pertinent paddy field situations, a thorough image of paddy crop production may be generated (features). Data analysts are creating predictive models in the form of expert systems to increase agricultural productivity while accounting for environmental elements like irrigation, land usage, and soil quality. In the bulk of current systems, crop yields are predicted using machine learning (ML), but little has been done to forecast crop yields based on soil and environmental parameters.

In order to forecast a suitable crop and assess the efficiency of the FS process, the characteristics gathered are input into k-nearest neighbour (kNN), Naive Bayes (NB), decision trees (DT), support vector machines (SVM), random forests (RF), and bagging classifiers. The SVM and the kNNare two crop prediction techniques that were utilised in this study, however the usage of a single prediction model is the most popular technique (such as the SVM). Each algorithm has its own prediction characteristics. To use the FS technique for crop prediction, an appropriate classifier must be found. Based on the soil and environmental conditions, a permutation crop data collection may be utilised to select the most appropriate key features for feature removal (MRFE).The method runs more quickly since the data set does not need to be changed after each iteration. Based on the soil and environmental conditions, a permutation crop data collection may be utilised to select the most appropriate key features for feature removal (MRFE). Other benchmark data sets and the bagging classifier: The UCI Repository was searched for non-crop data sets to make sure the suggested MRFE technique was applicable to data types other than crop-related data sets

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 6, April 2023

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 feature removal, correlation based feature selection, random forest Variable Importance, and Variance Inflation Factor. RMSE, MAE, R, and RRMSE metrics.	• to offer a thorough understanding of rice crop production by using feature selection algorithms to pinpoint important paddy field characteristics (features). • In terms of temporal complexity, forward feature selection is thought to be superior to backward elimination algorithms. • When all of the characteristics are incorporated into the model, accuracy is 84%. • Using the forward feature selection method, a better forecast may be achieved.
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	• PCA employs linear combinations of the original input variables that exhibit the greatest variance in a dataset to reduce the dimensionality of the input feature space. • In a data-driven context, machine learning approaches are ideally suited for parameter selection and ranking.
4	P. S. Maya Gopal and R. Bhargavi	2018	Borutaalgorithm, 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	● The authors create a model that forecasts how the climate will affect the severity of anthracnose using neural networks. ● Writers analyse climatic parameters using meteorological information and prior crops.
7	J. Camargo and A. Young	2019	sequential forward feature selection algorithm,	● This technique employs a sequential forward feature selection strategy, evaluating each feature in turn, to get satisfactory results and fast calculation times. ● Using a mixture of feature selection, it was feasible to obtain high levels of accuracy (>95%) for 33 different movement types.
8	R.RajashekerPullanagari, G.Kereszturi, and I. Yule	2018	RF-RFE Method	● .In places with a broad variety of soil types, a comprehensive assessment of the suggested technique is necessary. The final, precise geographic maps of pasture quality enable farmers to choose the optimal agronomic options. ● For the analysis of high dimensional data and the identification of significant spectral and environmental features sensitive to pasture quality, we discovered that RF-RFE was a highly effective feature selection approach, significantly superior than conventional methods.
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.
10	M. Lango and J. Stefanowski	2018	Class imbalance-Roughly balanced bagging · Types of minority examples-Feature selection, Multiple imbalanced class.	● The authors make a distinction between methods that employ resampling methods in ensembles that are not typically adaptive. ● The authors suggest using the SMOTE approach, or altering the oversampling ratios with each bootstrap, to boost ensemble diversity.

2.1 The Existing System

- In the current system, the RFE technique is a wrapper-type FS method that begins with all features in the training data set and successfully eliminates some of them features until only a small number remain.
- The RFE method evaluates acceptable features in order of importance and disqualifies the less significant ones. This method needs an iterative process for data set updating in the feature elimination process.
- Updating the data set is the most difficult part of the RFE, and eliminating weak features takes up the most time.

III. CONCLUSION

The MRFE method presented in this paper may be used to data sets from sources that are not related to crops. In comparison to prior techniques, the MRFE uses permutation and ranking to choose the attributes with the best predictive ACC in the shortest period of time. Using classification algorithms to find the best crops for farming is done using the MRFE method.

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. 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.

[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