International Journal of Advanced Engineering: Artificial Intelligence, Robotics and Automation
ISSN XXXX-XXXX
Vol.:1 Issue: 1
Published online on: Dec 2024
PP 77-84

The Role of Machine Learning in Enhancing Wheat Crop Yield Predictions

Sandeep K Tiwari[1], Premanand Singh Chauhan[2,] Anand kumar Singh[3], Anil Kewat[4]
[1 ,2,3] Vikrant University, Gwalior, India
[4]Bundelkhand University, Jhansi , India

Abstract: Agriculture is the key concern and promising field to explore in every country. In India, the population is growing rapidly; with the increase in population, the need for food is also rising. In this field, farmers and agriculture is have to take several decisions day by day , and sophisticated complexities involve the factors affecting them. An important concern for agricultural planning is the precise crop yield prediction and evaluation for a variety of types of crops involved in the planning. Machine learning techniques are the essential approach for achieving the competent solutions for these problems. Therefore, agriculture has been an obvious goal for big data. Environmental conditions, unpredictability in soil types, soil nutrients, rainfall, fertilizers, have made it all the more relevant for farmers to use the information and this information used to make censorious farming decisions. In the proposed work various machine learning algorithms like multiple linear regressions, ridge regression, lasso regression and, random forests are used. The lasso regression technique provides the better prediction results in comparison of others.

Keywords: MLP, LASSO regression, Ridge regression, Clustering, Association rule mining.

I.Introduction:

"Agriculture is the method of growing plants and other crops and for food, other human needs, or economic gain." India is an agricultural country. In India, the majority of the population is dependent on agriculture. Agriculture provides the highest contribution to national income. There are many severe problems like- erosion, diseases, pests, drought, and rainfall that people face trying to grow food today.

Machine learning techniques can be beneficial for better estimation of crop production rate based on various parameters. Classification and prediction techniques are applied to metrological-related data and crop-related data. Various predictions can be made based on predicted results which can help in increasing crop production rate. Various machine learning algorithms can be used to predict crop production rates based on all the available parameters.
In this research work we found that various machine learning and data mining techniques can be used for estimating the crop yield production rate and also how to find the optimal parameters or features for better prediction.

The main objective of this research paper is to study the state of the art in the agriculture field. In this research work various machine learning algorithms like random forest, lasso regression, MLP, ridge regression are used for better crop yield prediction based on area harvested, rainfall, and soil type. In this paper Section II describes the literature survey of research papers these papers are taken from IEEE, Science direct and Springer journal websites. Section III describes the proposed methodology for prediction of wheat crop production rate. Section IV descries the experimental results and finally section V describes the conclusion and future work.

II. Literature Review:

Vinita CN et al. proposed a model for predicting the rice production rate. This research paper researcher described the various data mining techniques and algorithms like support vector machine algorithm, Bayesian classifier, selective attribute network algorithm, K-Nearest neighbors' algorithm, and clustering techniques K-means clustering algorithm. In this paper author also proposed an SVM system flow diagram for providing the alert to farmers.

Copyright © IJAE 2024
Corresponding Author's E-mail ID: soet_sandeep@vikrantuniversity.ac.in

Moreover, the researcher applied the SVM algorithm to make the decisions for farmers and applied the selective attribute network algorithm and K-means clustering algorithm for clustering the data. The researcher predicted the crop yield based on various climatic conditions. The researcher described that for improving results and better crop yield prediction, artificial neural networks and neural networks could be used [1].

Jharna Majumdar et al. focus on analyzing agriculture data and finding optimal features to maximize crop production using data mining techniques. The authors elaborated on data mining techniques and described that data mining techniques play a significant role in extensive agriculture data analysis. Data mining is the method of extracting optimal patterns from a large amount dataset. In this proposed work, the PAM, DBSCAN, and CLARA clustering algorithms are used. The researcher used the dataset of Karnataka state. In this proposed work, PAM, CLARA, and DBSCAN clustering algorithms are applied to the dataset for categorizing the dataset of Karnataka, which has a similar crop yield production rate. The clusters are made according to the temperature, i.e., two clusters are made, and based on rainfall range, i.e., six clusters are formed and according to the soil type, i.e., five clusters formed. For better results, the researcher proposed the modified DBSCAN algorithm. The method is used to cluster the data based on districts with similar temperatures, rainfall, and soil type and find out the optimal features like temperature, soil type, and rainfall for better crop yield production [2].

Kirtan Jha et al. reviewed the problems occurring in the agriculture field and described that machine learning and data mining techniques could increase the production rate. First, the researcher reviewed the previous work in the agriculture field and emphasized the problems in the agriculture field. Then, the researcher proposed a model in which an artificial neural network is used to predict the crop yield; after that, automation and wireless network system are used in the agriculture field. The researcher concluded that machine learning, deep learning, fuzzy logic, and IOT, artificial intelligence can be used for automation in this field[3].

A.T.M Shakil Ahamed et al. tested few data mining techniques to predict the annual yield of major crops in Bangladesh. In this paper, the clustering technique is used to predict the results. There are parameters such as temperature, humidity, minimum temperature, maximum temperature, average sunshine, Soil PH, and salinity used to predict the Annual crop production. K- means the author uses clustering to recommend plant crops in Bangladesh's districts [6].

Belabed Image et al. proposed an approach for extracting information using data mining approaches in three domains, bioinformatics, medicine, and the agriculture industry. Initially, the variables are clustered to increase the functionality, and the association rules are generate among the target variables and the previously identified set of variables [7].

Shreya et al. reviewed and identified the problems and challenges that are faced by farmers in India. They have also collected and analyzed the dataset available online by using data visualization techniques of data mining for better understanding the data; then, they applied an unsupervised learning approach, i.e., K-Means clustering algorithm for finding out appropriate and valid clusters of the dataset. They have also used the Apriori algorithm of association rule mining for counting the frequently occurring features. Then naïve Bayes algorithm is implemented for the estimation of crop yield prediction[4].

Mercelin Francis et al. proposed a model for disease diagnosis along with classification in the agriculture field. He made a classification based on the spots found on the leaves so that agricultural productivity can be improved. He applied deep learning techniques to identify and classify apple leaf images. In their framework, authors first applied their methodology on the multi-space image reconstruction inputs and generated a set of gradient images. Next, in their work, they extracted the high-level semantic features from the original and reconstructed images using the depth-wise convolutional layers. The classification was achieved using the SoftMax classifier[5].

T.R.Lekhaa et al. studied various machine learning and data mining techniques. He suggested that WEKA(Waikato setting for information Analysis) be used to implement machine learning algorithms. He described that the WEKA tool could be used for data preprocessing, data visualization, and predicting the crop production rate[18].

Michael L. Mann et al. reviewed and estimate that precisely agricultural impact assessments for droughts are critical for designing appropriate interventions. In the proposed methodology, he proposed a data fusion method in which they use collective remotely sensed data with agricultural analysis data that may overcome challenges. They used the technique in Ethiopia, which is hit by droughts. They trained machine learning models to predict losses and discover the most significant features. This work examined the benefits of using machine learning models and remotely sensed data to predict agricultural losses due to drought in Ethiopia [17].

Copyright © IJAE 2024
Corresponding Author's E-mail ID: soet_sandeep@vikrantuniversity.ac.in

III. Research methodology:

Dataset description- Data used in this experimental is verified from- data.gov.in and indiawateerportal.org. There were wheat crop data of Madhya Pradesh. The attributes are in this data set are- district, year, area (in hectare), production (in metric ton), rainfall (in mm) and soil type and PH value etc.

1	District	year	area	production	rainfall	soiltype
2	Allirajpur	2007	14048	25310	922.2	1
3	Anuppur	2007	11964	8230	948.2	1
4	Ashoknagar	2007	104899	119871	810	1
5	Balaghat	2007	14672	13166	1257.1	1
6	Barwani	2007	34398	51908	733.6	1
7	Betul	2007	99067	153407	1301	2
8	Bhind	2007	73949	102508	478.6	3
9	Bhopal	2007	68225	120087	952.8	1
10	Bhurhanpur	2007	10380	19911	1060	1
11	Chhatarpur	2007	57707	60129	575.2	4
12	Chhindwara	2007	110732	223408	927.5	2
13	Damoh	2007	63063	99767	888.9	1
14	Datia	2007	88482	151820	568.4	1
15	Dewas	2007	118746	215031	1013.3	1
16	Dhar	2007	216267	500035	1161	1
17	Dindori	2007	29692	14958	975.2	4
18	Guna	2007	79286	113969	765.8	3
19	Gwalior	2007	67870	111365	540.5	3
20	Harda	2007	125786	412619	1140.6	1
21	Hoshangabad	2007	217612	698551	1092.8	1

Figure 1: Dataset

Proposed Methodology: In the proposed methodology for prediction of wheat crop yield prediction on the basis of area harvested, rainfall, soil type, and soil nutrients by using various machine learning techniques like multiple linear regression, random forest, LASSO regression, and ridge regression. Figure 2 shows the flowchart of the proposed methodology. The descriptions of phases are as follows:

Copyright © IJAE 2024
Corresponding Author's E-mail ID: soet_sandeep@vikrantuniversity.ac.in

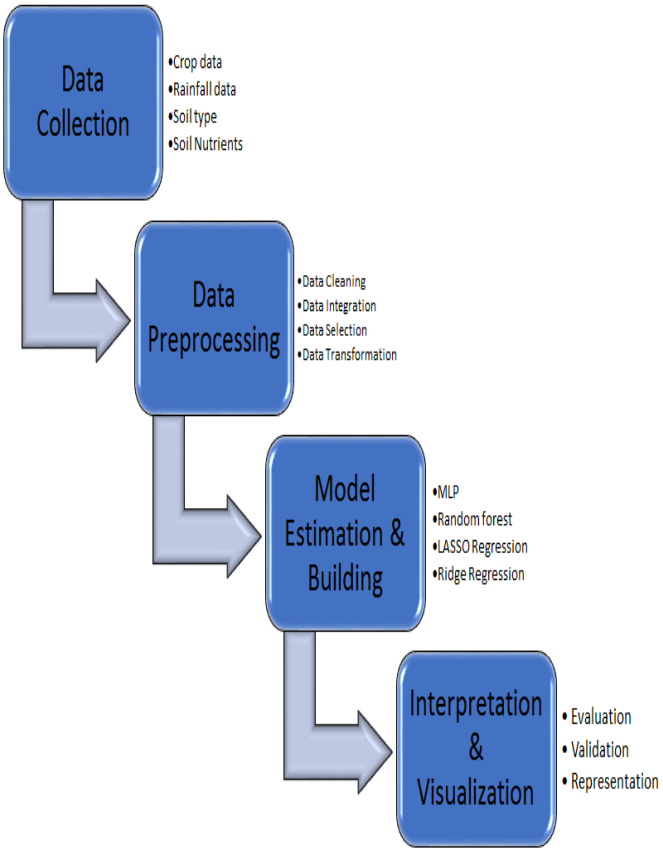

Figure 2: Proposed Methodology

Ø *Data Collection*:

A lots of data generated every year in the field of agriculture. Certain features associated with a crop make the data quite large.

Ø *Data Preprocessing:*

This step is used for making the data ready to process. This step includes removing noisy values, removal of redundant values, selection of optimal attributes, etc.

Ø *Model estimation & Building:*

In this step, a model is estimated and builds for the prediction of crop yield production rate on the basis of area harvested, rainfall, soil type.

For training of the model definite amount of data which is the part of main data is necessary. Likewise, for testing and validation, the rest amount of data is required. Data Splitting is the method that separates training dataset and testing dataset. In this step, machine learning algorithms like the random forest, multiple linear regression, ridge regression, and lasso regression algorithms are applied to the dataset. And lasso regression technique provides better results in comparison to others.

Ø *Interpretation & Visualization:*

In this step, predicted crop production is interpreted and visualize by using different charts and graphs.

- Multiple Linear regression- MLP is a kind of supervised learning algorithm with labeled dataset of machine learning. It describes the relationship between two or more independent variables and a single dependent variable by fitting an algorithm to data.
- LASSO Regression: LASSO(Least absolute shrinkage and selection operator) is a regression technique that performs both variable selection and regularization and also enhances the prediction accuracy.
- Ridge Regression: A ridge regression is a kind of regularized version of linear regression; that is, to the original cost function of linear regression, add an extra term that forces the algorithm to fit the data and manage the weights as lower as possible. The regularized term has the parameter which controls the regularization of the model.

Copyright © IJAE 2024
Corresponding Author's E-mail ID: soet_sandeep@vikrantuniversity.ac.in

- Random Forest: It is an ensemble learning method for regression, classification, and other tasks. For regression tasks, the mean or average prediction of the individual trees is returned.

IV. Experimental results:

From figure 3 to 6 shows the actual production vs. predicted production of wheat crop using multiple linear regression, random forest, lasso regression and ridge regression.

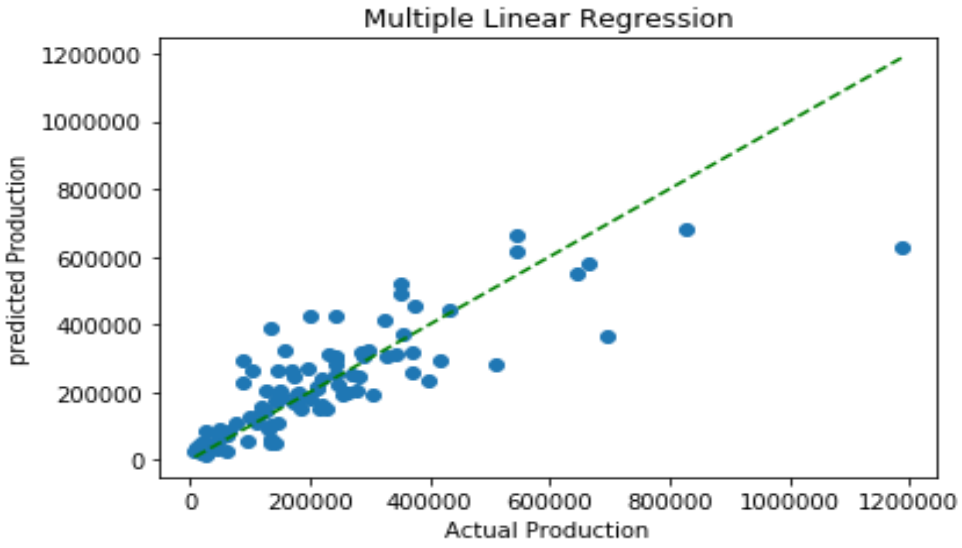

Figure 3: Actual production vs. predicted production using MLP

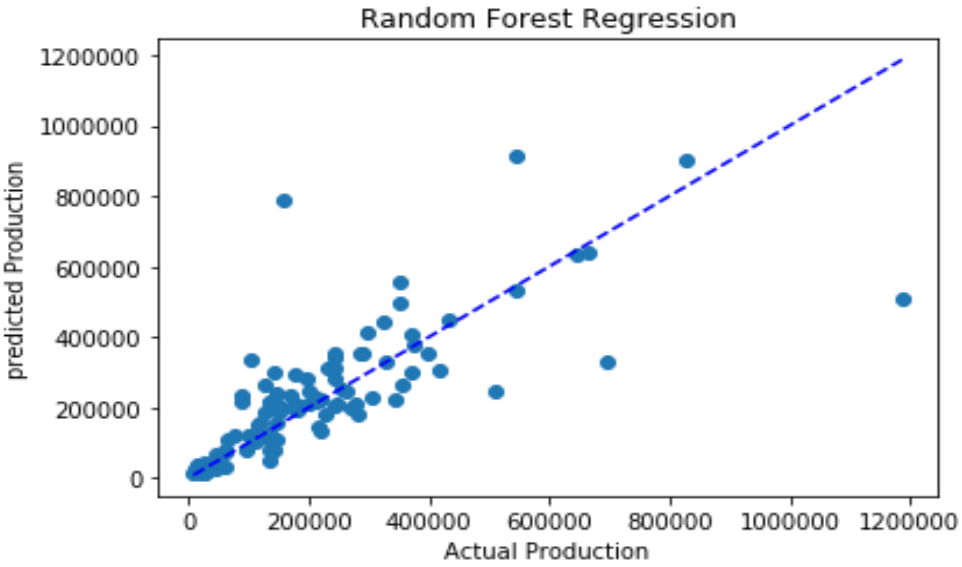

Figure 4: Actual production vs. predicted production using Random Forest

Copyright © IJAE 2024
Corresponding Author's E-mail ID: soet_sandeep@vikrantuniversity.ac.in

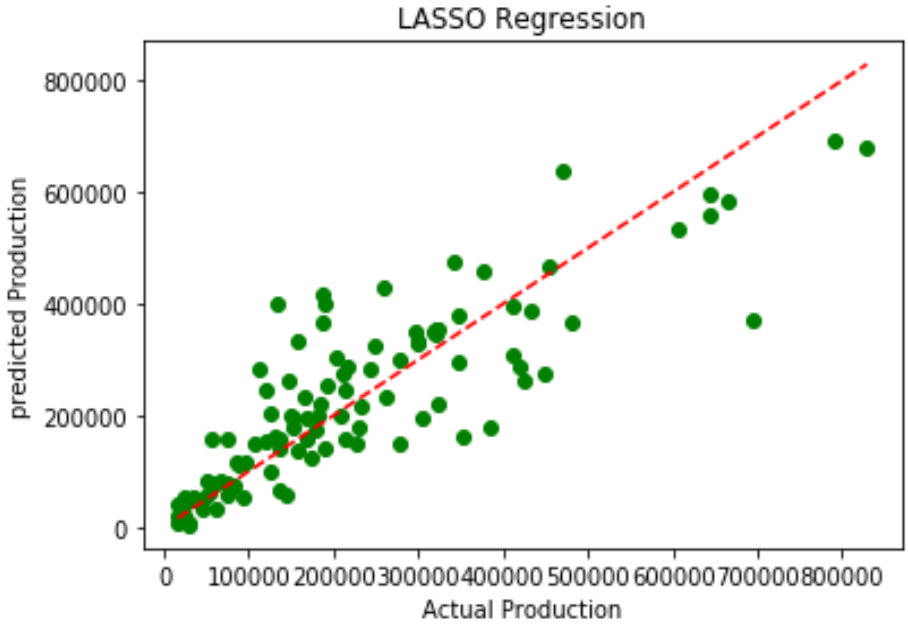

Figure 5: Actual Production vs. Predicted production using LASSO Regression

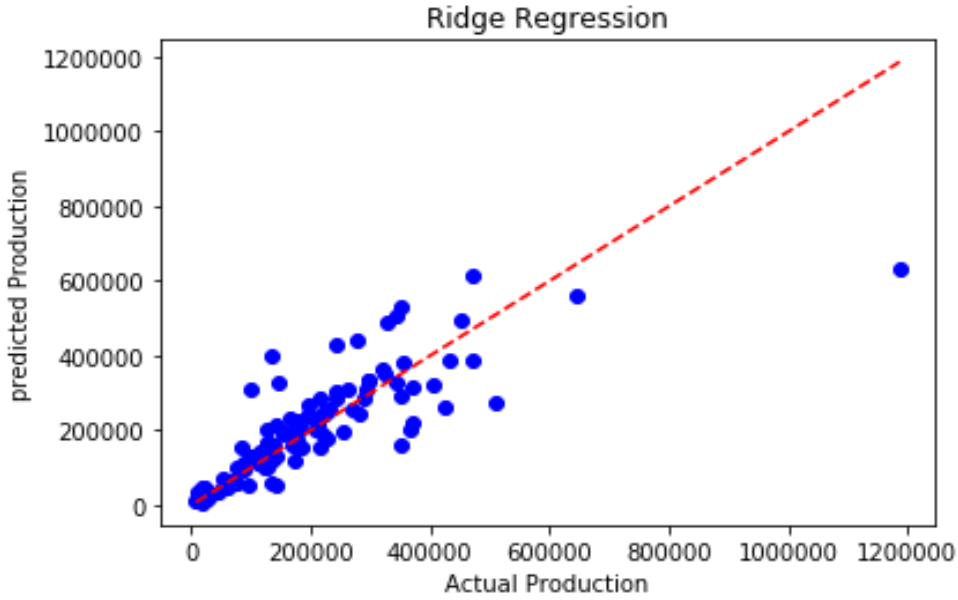

Figure 6: Actual Production vs. Predicted production using Ridge Regression

Figure 7 shows the comparison among mean squared error and mean absolute error of machine learning models.

Copyright © IJAE 2024
Corresponding Author's E-mail ID: soet_sandeep@vikrantuniversity.ac.in

Figure 7: Result Analysis of various algorithm

Table 1 shows the MSE, MAE and R^2 score values.

Algorithm	Mean Squared Error(MSE)	Mean Absolute Error(MAE)	R^2 Score
Multiple Linear Regression	10149338578	63544	0.71
Random Forest	15702206633	67476	0.55
LASSO Regression	89072.85	62267	0.85
Ridge Regression	93962.78	56496	0.66

Table 1: Mean squared error and mean absolute error and R^2 Score of algorithms

The results show that LASSO regression provides the better prediction results of wheat crop production rate.

V. Conclusion

In this research paper, various machine learning techniques like multiple linear regression, random forest, lasso regression, and ridge regression are used for better estimation of wheat crop yield. Machine learning techniques are the necessary approach for achieving efficient solutions for this problem. Therefore, agriculture has been an obvious target for big data. Environmental conditions, variability in soil types, soil nutrients, rainfall, fertilizers, have made it more relevant for farmers to use the information and this information used to make censorious farming decisions. The experimental results show that lasso regression has R^2 score of 0.85, and the mean squared error is 89072.85, mean absolute error is 62267. The result analysis shows that lasso regression provides better results in comparison to other algorithms. In future work need an efficient model in the agriculture field which can predict the crop production rate based on all the parameters, which also provides some recommendations to farmers like which fertilizer is best for this crop and how to maximize crop production rate and also for disease detection in crops. That is why various machine learning and deep learning algorithms like CNN, RNN, LSTM, capsule network can be used for improving prediction results.

Copyright © IJAE 2024
Corresponding Author's E-mail ID: soet_sandeep@vikrantuniversity.ac.in

References:

[1] C. N. Vanitha, N. Archana and R. Sowmiya, "Agriculture Analysis Using Data Mining And Machine Learning Techniques," *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, Coimbatore, India, 2019.

[2] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," Springer, 2017.

[3] Kirtan Jha, Aalap Doshi, Poojan Patel, "A Comprehensive review on automation in agriculture using artificial intelligence," Artificial intelligence in agriculture, Science direct, 2019.

[4] P. Surya, Dr. I. Laurence Aroquiaraj, "Crop Yield Prediction in Agriculture Using Data Mining Predictive Analytic Techniques", IJRAR December 2018.

[5] Mercelin Frances, C.Deisy, "Mathematical and Visual Understanding of the Deep Learning Model Towards m-Agriculture for Disease Diagnosis, " Springer 2020.

[6] A. T. M. S. Ahamed et al., "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh," 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Takamatsu, 2015, pp. 1-6.

[7] B. Imane, B. Abdelmajid, T. A. Mohammed, T. A. Mohammed and T. A. Youssef, "Data mining approach based on clustering and association rules applicable to different fields," 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, 2018, pp. 1-5.

[8] SML Venkata Narasimhamurthy, AVS Pavan Kumar, "Rice Crop Yield Forecasting Using Random Forest Algorithm", IJRASET, October 2017.

[9] Michael L. Mann, James M. Warner, Predicting high-magnitude, low-frequency crop losses using machine learning: an application to cereal crops in Ethiopia, springer 2019.

[10] Abraham Sudharson Ponraj, Vigneswaran T, "Machine Learning Approach for Agricultural IoT", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6, March 2019.

[11] S. Mishra, P. Paygude, S. Chaudhary and S. Idate, "Use of data mining in crop yield prediction," *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, 2018, pp. 796-802.

[12] D.Rajesh, "Application of Spatial Data Mining for Agriculture", International Journal of Computer Applications (0975 – 8887) Volume 15– No.2, February 2011.

[13] M. Paul, S. K. Vishwakarma and A. Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach," *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, 2015, pp. 766-771.

[14] A. Savla, N. Israni, P. Dhawan, A. Mandholia, H. Bhadada and S. Bhardwaj, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2015, pp. 1-7.

[15] Yogesh Gandge, Sandhya, "A Study on Various Data Mining Techniques for Crop Yield Prediction", 2017 International Conference on Electronics, Communication, Computer and Organization Techniques.

[16] K. Pavya, Dr. B. Srinivasan, "Feature Selection Techniques in Data Mining: A Study", IJSDR, June 2017.

[17] Michael L. Mann & James M. Warner & Arun S. Malik, "Predicting high-magnitude, low-frequency crop losses using machine learning: an application to cereal crops in Ethiopia" springer 2020.

[18] T.R.Lekhaa, A.Aruna, M.Malarmathi, "Budding Novel applications in agriculture Victimization data processing", 2018 International Conference on Soft-computing and Network Security (ICSNS), 2018.

[19] https://www.geeksforgeeks.org/Machine _learning.

[20] Akanksha Gahoi, "Data Analytics for Predicting Wheat Crop Production", Artificial & Computational Intelligence.

Copyright © IJAE 2024

Corresponding Author's E-mail ID: soet_sandeep@vikrantuniversity.ac.in