International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 10, ISSN No. 2455-2143, Pages 117-123
Published Online February 2023 in IJEAST (http://www.ijeast.com)

CROP RECOMMENDER SYSTEM

Shivanoori Sai Samhith, Dr. T. V. Rajinikanth, Burma Kavya, Alley Yashwanth Sai Krishna
Department of CSE
Sreenidhi Institute of Science and Technology,
Hyderabad, Telangana, India

Abstract— **The development of farming techniques improved food supply to the nation's GDP and made it simpler for farmers to grow the precise and suitable crop without any risk or loss in productivity. Agriculture is important to human existence and livelihood because it has become a necessary and crucial part of our daily lives. To maintain the sustainability in rate and quality of production, we need to involve and introduce the best portable and noncomplex strategies such as machine learning tools to carry on the needed operations and procedures to get desired and expected crop by coordinating and initiating data exchange between the scientific and practical platforms which is the trending existing system. The idea behind this project, called the crop recommendation system, is to define and state that the appropriate crop should be grown based on a number of relative parameters, including soil features like nitrogen, phosphorus, and potassium that are extracted from the soil through filtration, and weather conditions that are embedded in a dataset in the form of structured data. This dataset is taken over by machine learning algorithms that will perform some operations like classification and will be finding the accuracy where in detail to be explained we will be splitting the given dataset into training and testing data and compare the results of those algorithms based on accuracy that each model gives and that will be our preferred algorithm. Along with a few dimensionality reduction techniques including PCA, LDA, and cross validation, we applied machine learning techniques like Decision Tree, Random Forest, and KNN.**

Keywords— **Crop prediction, Decision Tree, Random Forest, K-Nearest Neighbour, Cross Validation.**

I. INTRODUCTION

Today's globe, where population is growing, depends heavily on agriculture. Farmers must consider many factors when growing crops, such as which crop will produce a higher yield, the future weather, and many more. Predicting which crop would give better yield involves different factors like soil ph value, rainfall, area, temperature and many more. This helps the farmer to get a good crop with good revenue. By this farmers will also be able to lead a better life. Knowing which crop to grow is a very important aspect before we cultivate any crop.

Here in this paper we would be using few factors and predicting which crop to grow so that it gives a better yield. The factors which we are going to use in this are from a dataset which has N, P, K values, temperature in that area, humidity in that area, ph values and rainfall. Here N, P, K values involves the values of Nitrogen, Phosphorus and Potassium levels in the soil. These values play a vital role in knowing how the soil is and will help in knowing which crop would grow better that soil. Rainfall is also an important aspect to know the weather conditions and how many centimetres rainfall might occur. We would be using those factors and predicting which crop to grow like Rice, Paddy, Coffee, Maize, etc.

We are going to split the data and then apply few machine learning based techniques to predict which crop to grow and then predicting the accuracy percentage to know which algorithm is going to work better to predict the crop. The Decision Tree Classifier, Random Forest, and K-Nearest Neighbour algorithms are examples of machine learning algorithms. After predicting which crop to grow we would calculate some accuracy score so that we would be able to know which algorithm is giving best prediction among them. By getting the accuracy score we can help farmer by predicting which crop to grow in their farm by using the best algorithm among them which we applied. We also applied few dimensionality reduction techniques like PCA and LDA to get better accuracy and also cross validated the accuracy predicted.

By predicting the accuracy percentage to know which algorithm works better we also represented that graphically to get it easily understandable.

II. LITERATURE REVIEW

M. Suganya et al [1] Using various machine learning algorithms, including K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Random Forest, and Decision Tree, the paper attempts to forecast the crop's production. By applying all those algorithms they concluded that logistic regression works better among all the algorithms and support vector machine gives worst performance among all of them. Anakha Venugopal etal [2] paper is about predicting which crop to be grown and which crop gives the maximum yield for the farmer. Temperature, rainfall, area, and other variables are used to predict the crop

and production. In order to estimate the crop and determine the accuracy rate, various machine learning techniques, including Naive Bayes, Logistic Regression, and Random Forest, are applied to the dataset containing the aforementioned elements. Among them Random Forest gives the best accuracy. Kalaiarasi Sonai Muthu Anbananthen et al [3] paper is about helping the farmer to know which crop to grow so that it gives good yield. Here different hybrid machine learning techniques are applied like Gradient Boosted Tree Regression, Random Forest Regressor, and Stacked Generalization so that we can analyse which algorithm gives better accuracy for recommending the crop to the farmer. From the above three hybrid machine learning algorithms Stacked Generalization gives the best accuracy of yield and this algorithm can be used to know which crop can be grown. Aksheya Suresh et al [4] paper is about predicting the crop and also yield of the crop based on the factors of location such as state name, district name and also based on the season. Here, the crop was predicted using the machine learning algorithm Decision Tree, and the yield was projected using the method Linear Regression. The model has a good prediction accuracy rate. Ganesh Sastry Kakaraparthi et al [5] paper is about predicting the price of the crop based on the factors rainfall and wholesale price index. Here they used machine learning technique named Decision Tree for estimating the price of the crop for next twelve months. This prediction of price would help the farmer in getting a good revenue. S. Nagini et al [6] paper is about predicting the yield of a crop in a particular region based on the factors like soil characterictics, soil moisture, surface temperature, nitrogen, water, rain water. Here they used various regression based machine learning techniques like Multiple Linear regression, Linear regression, non- linear models. The accuracy percentage of predicted yield among different algorithms is compared to know which algorithm works better. Rohini Jadhav et al. [7] had stated about working with some advanced farming technologies like GPS technology app that captures the user details and location are given to it and it will provide yield prediction and crop through various machine learning algorithms applied on the data. A. Suruliandi et al. [8] has conveyed that for healthy production we need to prefer latest procedures like comparing and selecting the feature selection methods with various classification or machine learning models based on the accuracy to get the optimal result. Mahendra Choudhary et al. [9] has expressed about providing the importance of latest farming techniques and machine learning algorithms to farmers that prevents imprecision farming and also identifies crop as well as classifies plant disease through a GUI app. Prasad D Kadam et al. [10] has portrayed the content which explains about consistency in yield will be regularized by a framework called crop recommendation model that includes deep learning algorithm and predefined data is linked with AgroSYS neural network architecture to give valid accuracy. K Anguraj et al. [11] has mentioned that the procedure for a crop to be grown is of capturing soil parameters with help of IOT sensor device that collects the data and that is given to the GUI which performs operation and displays the required crop based on given attributes. Vaishnavi.S et al. [12] has explained about planting crops with help of its content related to productivity and season are giving to machine learning models that allows to choose the right crop for cultivation that reduces the climatic risk factors and maintains healthy agricultural environment. Jyothi Gupta et al. [13] has stated about an android application that operates on machine learning platform which depicts the crop yield and type as soil and weather parameters are concatenated as well as given to train the models and will provide suitable crop according to given attributes gained from soil testing lab and agriculture experts. Kodimalar Palanivel et al [14] stated that, How different machine algorithms are helpful in yield prediction given that a number of factors, such as water scarcity and soil fertility, have an impact on crop production and accuracy. They used algorithms from the big data computing paradigm to estimate agricultural yield. José R.Romero et al [15] The main goal is to forecast the durum wheat production using computer algorithms, then compare the results to determine which algorithm performs best. SuYing-xue et al [16] This project will create a model to predict the yield and stage of rice development, which it will then utilise to connect into the SBOCM system. Within specific scale ranges, the SBOCM system is employed for perpetual modelling and one-year rice projections. Avat Shekoofa et al [17] selected the most pertinent elements for the possibility of accurately boosting maize grain output after analysing a huge number of physiological and agronomic traits using screening, clustering, and decision tree models. The most common tool is a decision tree. The outcomes demonstrated that crop physiologists can benefit from using model techniques as tools. Yvette Everingham et al [18] stated that, In order to create a sugarcane prediction model, a data mining technique called random forest is used. The algorithm is flawless, and they have noticed a considerable change in the sugarcane production. Farmers can enhance their nitrogen management to match the needs of the new crops thanks to better crop predictions. Mayank Champaneri et al [19] stated that, The ability to anticipate crop yield in advance of harvest would be helpful to farmers and decision-makers in determining the best course of action for selling and storage. The most potent method for both classification and regression problems is random forest. It reliably forecasts the crop yield. Takalani Orifha Mufamadi et al [20] had seen the idea of selecting the best crop based on the application and execution of machine learning algorithms and its related operations such as support vector machine and random forest on the soil features of nitrogen, potassium, and phosphorus as well as ph value, which in turn results in the right crop.

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 10, ISSN No. 2455-2143, Pages 117-123
Published Online February 2023 in IJEAST (http://www.ijeast.com)

III. MOTIVATION

The world's largest industry, agriculture, is finding it harder and harder to estimate earnings as the global population is expanding at an exponential rate. Over 70% of the people in our nation, India, depend on agriculture. Recently most of the farmers are leaving farming and doing other daily wages work for their livelihood because of the losses they are making the agriculture they can't live with loses. If they know the exact yield they can produce before the harvesting then they can escape from their losses, to do that there are different approaches in the technologies to predict the yield before the harvesting. With the perfect decision making algorithm we can predict the accuracy of each crop and make them cultivate so that they can get profits. With the exact values of the soil content and the factors affecting the agriculture we predict which crop should be grown.

IV. METHODOLOGY

In this methodology we would be discussing about the different aspects of the project such as, the dataset attributes, the algorithms applied and the steps involved in getting the crop recommendation.

IV.I Dataset Retrieval

In Agriculture Crop Prediction we consider various factors for prediction and each has a unique importance for their consideration. Before any machine learning algorithms are used on the dataset, the dataset needs be pre- processed. In this project we considered few factors and based on those factors we would be predicting the crop. Temperature, pH, Humidity, Rainfall, and Nutrients such as Nitrogen (N), Potassium (K), and Phosphorous are included in the dataset's characteristics (P). Using these we would be predicting the crop name. The independent attributes of the dataset involves Temperature, pH value, Humidity, Rainfall and Nutrients. The dependent variable will be the Label which is the name of the crop and it is predicted using all the independent variables. We would be applying different algorithms on this dataset like Decision Tree, Random Forest and KNN. Before using any method, we would also use dimension reduction techniques like PCA and LDA to increase the accuracy of crop prediction.

IV.II Algorithm Building

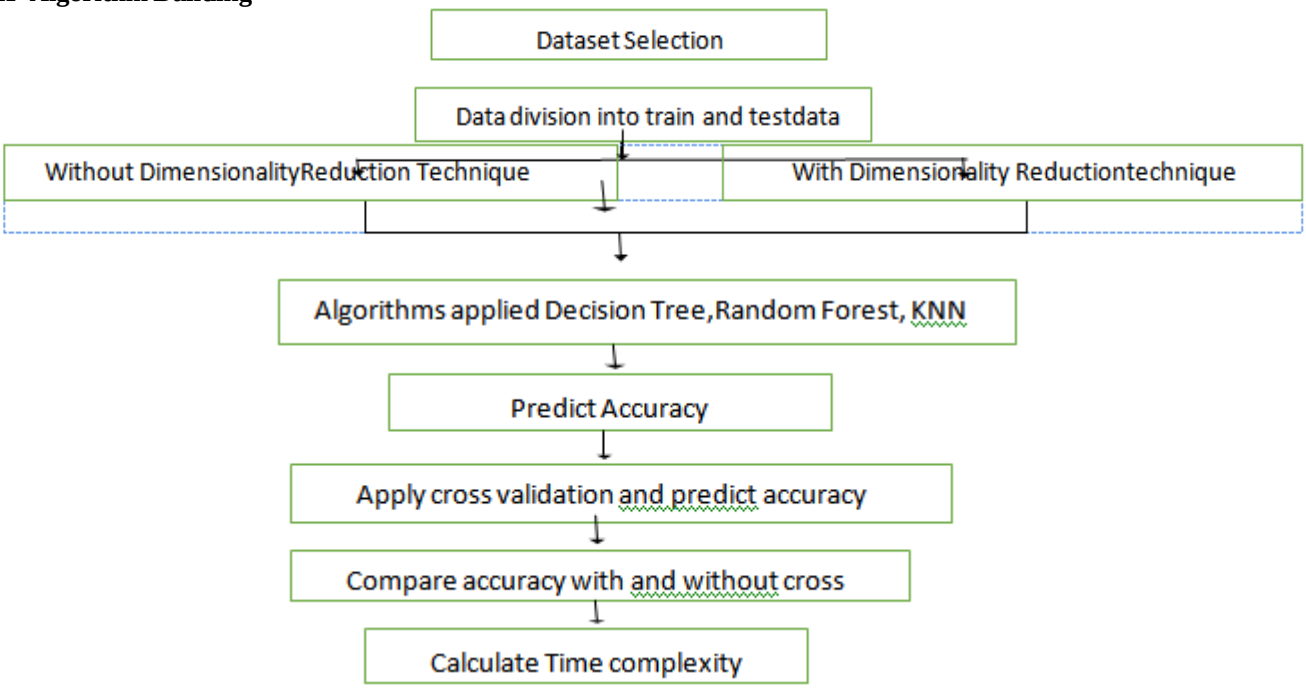

Figure 1: Architecture of Crop Recommendation System.

Three machine-learning algorithms make up the model in this study. The algorithms make predictions depending on the correctness of each individual's work. In some cases, the accuracy supplied by the algorithms may be identical, but we have incorporated extra features to forecast the best method in addition to the accuracy provided by the algorithms. K-Nearest Neighbour, Decision Tree, Random Forest are the algorithms employed. To acquire the projected crop accuracy, we apply several algorithms to the dataset during algorithm development. We compare the predicted crop accuracy with the different algorithms with cross validation and without cross validation.

We have applied Dimensionality reduction methods like Principal Component Analysis (PCA) and Linear

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, No. 8, Pages 1-6
Published Online June – July 2016 in IJEAST (http://www.ijeast.com)

Discriminant Analysis (LDA) to the algorithms and we have compared the predicted crop accuracy with cross validation and without cross validation. In PCA dimensionality reduction, it reduces the number of input variables for dataset and increase the accuracy of the crop prediction. In LDA, used to project the features in higher dimension space to the lower dimension space and increase the accuracy and it is compared with cross validation and without cross validation.

The key benefit of applying dimensionality reduction is to improve the accuracy of the model and the model to run as soon as possible with least amount of time complexity while still providing the best output with the best-predicted values

IV.II Accuracy Prediction and Comparison

IV.II.I Accuracy Prediction without any Dimensionality Reduction Technique

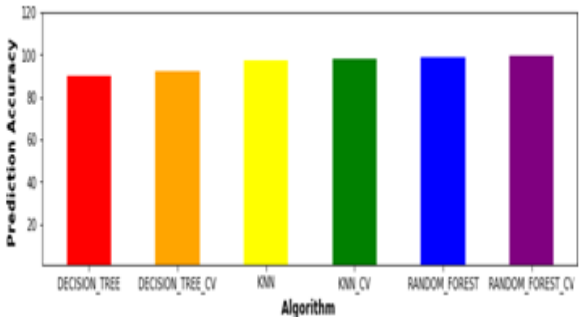

Figure 2: Accuracy using different algorithms

The above graph shows the potentiality of different algorithms with cross validation and without cross validation. As we can see they are two axis on the chart such as x axis represent the algorithm names and y-axis represent the predicted accuracy on the scale of 20 in the range of 0-120.

IV.II.II Accuracy Prediction using PCA Dimensionality Reduction Technique

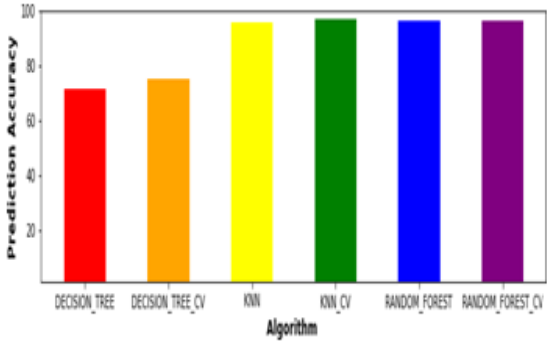

Figure 3: Accuracy using different algorithms with PCA Dimensionality Reduction Technique

This graph shows the accuracy of the different algorithms when PCA dimensionality reduction method is applied with cross validation and without cross validation. In above figure x-axis contains the names of the algorithms, y-axis is the predicted accuracy on the scale of 20 in the range of 0-100.

IV.II.III Accuracy Prediction using LDA Dimensionality Reduction Technique

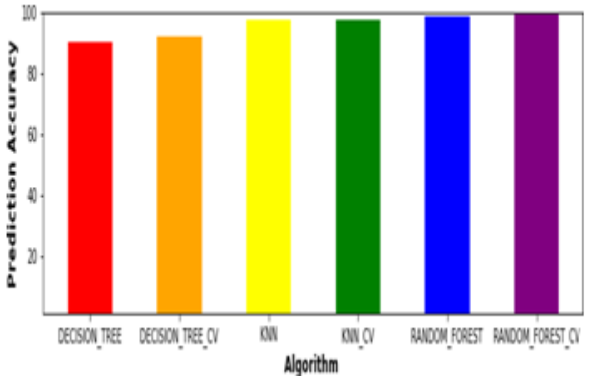

Figure 4: Accuracy using different algorithms with LDA Dimensionality Reduction Technique

This graph shows the accuracy of the different algorithms when LDA dimensionality reduction method is applied with cross validation and without cross validation. Where x-axis represent the names of the algorithm and y-axis gives predicted accuracy on the scale of 20 in the range of 0-100.

V. RESULT

V.I Accuracy Prediction Table

V.I.I Without any Dimensionality Reduction Technique

Algorithms	Without Cross Validation	With Cross Validation
Decision Tree	90.0	92.18
KNN	97.2	97.6
Random Forest	98.8	99.18

Table 1: Accuracy Percentages without any Dimensionality Reduction Technique

These are the accuracy values of different machine learning algorithms obtained before and after applying cross validation on the dataset, the above table shows that random forest has given the highest percentage and responded

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, No. 8, Pages 1-6
Published Online June – July 2016 in IJEAST (http://www.ijeast.com)

positively for both with and without cross validation compared to other algorithms.

V.I.II With PCA Dimensionality Reduction Technique

Algorithms	Without Cross Validation	With Cross Validation
Decision Tree	71.3	75
KNN	95.6	96.7
Random Forest	96.1	96.4

Table 2: Accuracy Percentages with PCA Dimensionality Reduction Technique

The above table represents the accuracy results of three algorithms gained after applying dimensionality reduction technique called Principal component analysis to get much better output, here a difference to be noticed is that by applying cross validation on KNN we got highest value of 96.7 than other techniques.

V.I.III Accuracy Prediction Table with LDA Dimensionality Reduction Technique

Algorithms	Without Cross Validation	With Cross Validation
Decision Tree	90.0	92.18
KNN	97.2	97.6
Random Forest	98.8	99.18

Table 3: Accuracy Percentages with LDA Dimensionality Reduction Technique

The second dimensionality reduction technique that is used to improve the working of algorithms is Linear discriminant analysis applied on the dataset which leads to the compression of dataset gives more efficient and highest results that random forest has got up to 99.18% with cross validation.

V.II Time Complexity for Accuracy Predicted
V.II.I Without any Dimensionality Reduction Technique

Algorithms	Without Cross Validation	With Cross Validation
Decision Tree	0.05secs	0.06 secs
KNN	0.06 secs	0.19 secs
Random Forest	0.3 secs	0.91 secs

Table 4: Time Complexity without any Dimensionality Reduction Technique

The time complexity table shows and states that in how much duration that each algorithm has taken to provide the result and accuracy. From above dataset we can say that the random forest has taken the least and minimum time of 0.3 seconds to provide the result.

V.II.II With PCA Dimensionality Reduction Technique

Algorithms	Without Cross Validation	With Cross Validation
Decision Tree	0.19 secs	0.64 secs
KNN	0.07 secs	0.23 secs
Random Forest	0.2 secs	1.14 secs

Table 5: Time Complexity with PCA Dimensionality Reduction Technique

The above data in table shows the time complexity for Principle component analysis has provided the least time that random forest has taken without cross validation.

V.II.III With LDA Dimensionality Reduction Technique

Algorithms	Without Cross Validation	With Cross Validation
Decision Tree	0.10 secs	0.58 secs
KNN	0.08 secs	0.32 secs
Random Forest	0.24 secs	1.43 secs

Table 6: Time Complexity with LDA Dimensionality Reduction Technique

The above table shows about time taken by each algorithm after applying linear discriminant analysis, we can see that without applying cross validation on KNN has taken 0.08

milliseconds of time to complete the task which is very less compared to other algorithms.

VI. CONCLUSION

This project ensures the reduction of farmer's suicide rate by improving the production per yield, due to which there will be a healthy society. By using machine learning algorithms we are able to estimate which kind of crop to be cultivated based on given conditions, as it performs some operations like classification and regression and applies classifying techniques such as decision tree, random forest and KNN on the given dataset which contains values related to weather and soil. Among all the classifiers Random Forest has provided the best precision of 98.8%.On comparing dataset accuracy percentages of different machine learning algorithms along cross validation we can state that without and with applying dimensionality reduction technique such as principle component analysis both got exact potentiality values and there percentages got increased. We can conclude from above research that the cross validation has increased the accuracy percentage of crop prediction in each and every case whether we applied dimensionality reduction technique or not. But we can also say from above results that the accuracy prediction when LDA dimensionality reduction technique is applied gives very similar accuracy to that when no dimensionality reduction technique is applied. Among all the algorithms applied i.e., Decision Tree, Random Forest and KNN, the standard one was Random Forest has given the best accuracy with or without any dimensionality reduction technique and also Random Forest works the best among three algorithms even when cross validation is applied.

VII. REFERENCES

[1] M. Suganya, Dayana R, Revathi. R, "Crop Yield Prediction Using Supervised Learning Techniques", International Journal of Computer Engineering and Technology, 11(2), 2020, pp. 9-20.

[2] Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Vinu Williams, "Crop Yield Prediction using Machine Learning Algorithms", IJERT ,NCREIS – 2021 (Volume 09 – Issue 13).

[3] Kalaiarasi Sonai Muthu Anbananthen, Sridevi Subbiah, Deisy Chelliah, Prithika Sivakumar, Varsha Somasundaram, Kethaarini Harshana Velshankar, M.K.A.Ahamed Khan, "An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms", 2021 Nov 11. doi:10.12688/f1000research.73009.1.

[4] Aksheya Suresh, K. Monisha, R. Pavithra,B. Marish Hariswamy, "Crop Selection and it's Yield Prediction", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020.

[5] Ganesh Sastry Kakaraparthi, B.V.A.N.S.S.Prabhakar Rao, "Crop Price Prediction Using Machine Learning", e-ISSN: 2582-5208 International Research Journal of Modernization in Engineering Technology and Science Volume:03/Issue:06/June-2021 Impact Factor- 5.354.

[6] S. Nagini, T. V. R. Kanth and B. V. Kiranmayee, "Agriculture yield prediction using predictive analytic techniques," 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), 2016, pp. 783-788, doi: 10.1109/IC3I.2016.7918789.

[7] Rohini Jadhav, Dr. PawanBhaladhare, "A Machine LearningBased Crop Recommendation System: A Survey", Journal of Algebraic Statistics, Volume 13, No. 1, 2022, p.426-430https://publishoa.comISSN:1309- 3452.

[8] A. Suruliandi , G. Mariammal, S.P. Raja, "Crop prediction based on soil and environmental characteristics using feature selection techniques", Mathematical And Computer Modelling Of Dynamical Systems2021, VOL. 27, NO. 1, 117-140.

[9] Mahendra Choudhary, Rohit Sartandel, Anish Arun, Leena Ladge, "Crop Recommendation System and Plant Disease Classification using Machine Learning for Precision Agriculture", 2022. In Saroj Hiranwal & Garima Mathur (eds.), Artificial Intelligence and Communication Technologies, 39–49. Computing & Intelligent Systems, SCRS, India.

[10] Prasad D. Kadam, Riddhi S. Chavan, Aditya M. Kulkarni3, Surekha R. Janrao, "AgroSys – A Crop Recommendation System", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 05 | May 2021.

[11] Anguraj.Ka , Thiyaneswaran.Bb , Megashree.Gc , Preetha Shri.J.Gd , Navya.Se , Jayanthi. J f, "Crop Recommendation on Analyzing Soil Using Machine Learning", Turkish Journal of Computer and Mathematics Education Vol.12 No.6 (2021), 1784-1791.

[12] Vaishnavi., S., Shobana., M., Sabitha., R., & Karthik., S. (2021). Agricultural Crop Recommendations based on Productivity and Season. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). doi:10.1109/icaccs51430.2021.9441736.

[13] Jyoti Gupta, Ayushi Chauhan, Aastha Agarwal, Abhijeet Singh Ranghuvanshi, Richa Saxena, "Machine Learning Approach for Crop Yield Prediction and Crop Variety Recommendation in Android Application", MIT International Journal of Computer Science and Information Technology, Vol. 7, No. 1, January 2018 ISSN 2230-7621.

[14] Kodimalar Palanivel, Chellammal Surianarayanan, "An Approach for Prediction of Crop Yield Using

Machine Learning and Big Data Techniques", International Journal of Computer Engineering and Technology 10(3), pp. 110-118, 2019.

[15] José R.Romero, Pablo F.Roncallo, Pavan C.Akkiraju, IgnacioPonzoni, Viviana C.Echenique, Jessica A.Carballido, "Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires", Computers and Electronics in Agriculture, Volume 96, August 2013.

[16] SuYing-xueXuHuanYanLi-jiao, "Support vector machine-based open crop model (SBOCM): Case of rice production in China", Saudi Journal of Biological Sciences, Volume 24, Issue 3, March 2017.

[17] Avat Shekoofa, Yahya Emam, Navid Shekoufa, Mansour Ebrahimi, Esmaeil Ebrahimie, "Determining the Most Important Physiological and Agronomic Traits Contributing to Maize Grain Yield through Machine Learning Algorithms: A New Avenue in Intelligent Agriculture", Published: May 15, 2014.

[18] Yvette Everingham, Justin Sexton, Danielle Skocaj, Geoff Inman-Bamber, "Accurate prediction of sugarcane yield using a random forest algorithm", Published: 19 April 2016.

[19] Mayank Champaneri, Darpan Chachpara, Chaitanya Chandvidkar, Mansing Rathod, "Crop Yield Prediction Using Machine Learning", April 2020 International Journal of Science and Research (IJSR) 9(4 April 2020):2.

[20] Takalani Orifha Mufamadi, Ritesh Ajoodha, "Crop Recommendation using Machine Learning Algorithms and Soil Attributes Data", 3 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC- SAS).