Proceedings of the Sixth International Conference on Trends in Electronics and Informatics (ICOEI 2022)
IEEE Xplore Part Number: CFP22J32-ART; ISBN: 978-1-6654-8328-5

A Machine Learning-based Approach for Crop Yield Prediction and Fertilizer Recommendation

Jeevaganesh R
Department of Computer Science and Engineering
Sri Sairam Engineering College, Chennai
jeevaganesh.2812@gmail.com

Harish D
Department of Computer Science and Engineering
Sri Sairam Engineering College, Chennai
harishdevan0603@gmail.com

Priya B
Department of Computer Science and Engineering
Sri Sairam Engineering College, Chennai
bpriya.cse@sairam.edu.in

2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) | 978-1-6654-8328-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICOEI53556.2022.9777230

***Abstract*—Agriculture plays a critical role to Indian global economy and contributes a major part to GDP. With the expansion of the human population, it is necessary to maintain food security, it is achieved and controlled by the agricultural yield produced by the nation. The yield of a crop is mainly determined by the climatic conditions like temperature, rainfall, soil conditions, and fertilizers. Due to these variable factors, the production gets affected and remains a huge problem for the agricultural sector to strengthen the need for exactness for proper analyzing the crop production in variable climatic conditions. Recently, the machine learning algorithms are used by the researchers to predict the yield of a crop before its actual cultivation. This research study has proposed a machine learning algorithm: AdaBoost to predict the yield of crops based on the parameters like state, district, area, seasons, rainfall, temperature, and area. To enhance the yield, this research study also suggests a fertilizer based on the soil conditions like NPK values, soil type, soil PH, humidity, and moisture. Fertilizer recommendation is primarily done by using the Random Forest [RF] algorithm.**

Keywords—Crop yield prediction, Fertilizer recommendation, Machine learning- AdaBoost, Random Forest.

I. INTRODUCTION

India is known for its widespread agricultural production. Agribusiness shares almost a high edge in the nation's Gross Domestic Product GDP and incorporates tremendous specialists. Crop production fluctuates in light of the various variables, climatic and weather conditions. At some point without appropriate investigation on these variables or boundaries, the yield of a harvest gets impacted and results in huge loss. With the goal that many long stretches of difficult work put to the development of harvest by the farmers are getting extinguished.

The ascent of innovation had made tremendous transformations in all fields all over the world. Be that as it may, India has only experienced a marginally less transformation as most of the farmers have not procured to present-day agricultural practices. However, as a developing country, only few present-day methods of farming are contrasted with different fields. If it is made to follow some new agricultural practices, more yield can be acquired and considerably a more percentage can be contributed to the country's GDP. Let us think how it could be, like knowing how much we get yield before the cultivation itself. That's what machine learning does, it enables you to predict the future based on the past experience. It is all about train and testing the datasets. How much we train the machine learning model that is more datasets we provide; we can expect more accuracy.

From an early period agribusiness has turned into the foundation of our country. In recent days, the climatic conditions differ frequently. Thus, it is difficult to develop crops under these climate conditions. We want to utilize some innovation to find or comprehend the yield subtleties and guide the farmers to develop crops appropriately and additionally, fertilizers are an essential one and also one of the main considerations to cultivate crops.

Temperature and rainfall are the major factors that influence crop production, understanding these factors is quite complex, this is where machine learning comes into existence. Beforehand yield is anticipated on the farmer's knowledge on the before experience however presently climate conditions might change definitely so they can't predict the yield as they would do earlier. The new age of technology paved the way for the rise of advanced concepts like artificial intelligence, machine learning, these can predict future results. Also, they were trained and tested to predict and give a most accurate and close assumption for future events to be held. With the help of this machine learning algorithm, we can calculate and predict the most productive result of the yield.

Assuming fertilizers are utilized pretty much in the agricultural land it may result in loss of fertility in soil and crop cultivated may affect badly and even could not meet up the regular yield. In this way, fertilizers additionally turn into the main consideration in it. Fertilizers for agricultural land must be recommended based on the proper analysis of soil parameters. We can't suggest fertilizers on the common basis

of the crop to be cultivated. So that fertilizers recommended are to consider the necessary factors such as Nitrogen, Potassium, Phosphorous, soil PH, humidity, moisture, soil type, and crop type. If the fertilizers are suggested based on the above parameters, then the fertilizer suggested will be the most preferred one and does not cause any damage to the topsoil that is there is no loss of fertility. Hence, we can expect a good amount of yield from the crop. But in the real world, there is no fertilizers are suggested based on this proper analysis of these parameters.

In this way, the present technology can see through the future that is can calculate the yield before the actual cultivation. Based on the results farmers can decide whether to go with that crop or drop it shift to other crops. Hence farmers could use this as assistance for their agricultural activity.

II. LITERATURE SURVEY

Dhivya Elavarasan et al 2020 [1] had developed a crop yield prediction framework that relies on deep reinforcement learning. They considered that the high quality of crop features data and parameters are mainly responsible for the prediction model. 38 Parameters are taken for consideration of the prediction model. Merging of two fields: reinforcement learning and deep learning a new crop yield prediction framework was built. The proposed methodology is to develop a Deep Recurrent Q-Network model. The recurrent neural network model's job is to fetch the data to the yield prediction model. Q-learning develops an environment for parameters to predict yield. Finally merging of these two will produce a complete crop yield prediction model with less error rate and accuracy of 93.7%.

S. Bhanumathi et al 2019 [2] build a prediction model which is made to predict the crop yield based on Indian agriculture and suggested the efficient way of using and recommended fertilizer. Mainly yield prediction is based on the parameters like crop type, seasons, and fertilizer recommendation model based on the analysis of soil type, NPK values. They used a random forest algorithm and backpropagation to get the result. Usually, results are predicted by different machine learning algorithms by comparing the error rate occurred final results are fetched by the model which has a low error rate.

Niketa Gandhi et al 2016 [3] In India, environmental conditions fluctuate unequivocally. In the hour of drought conditions, farmers deal with major issues. So, this thought about they utilized machine learning algorithms calculations to help the farmers to make the harvest for the better yield. They take different datasets from the earlier years to assess future data. They involved sequential minimal optimization (SMO) classifiers in WEKA to arrange the outcomes. The primary parameters taken are least temperature, most temperature, normal temperature, and earlier year's harvest data and yield data. Utilizing SMO, they ordered the past information into two classes that are more and less yield. The results produced are meant to be less accurate compared with others.

Shivam Bang et al. 2019 [4] predicted the crop, based on temperature and rainfall only. Due to various climatic changes, their aim is to predict the temperature and rainfall of the region. Based on the above-predicted values with the use of fuzzy logic, their templates are "bad, very bad, good, very good, and average". The comments predicted are based on outcomes temperature and rainfall prediction model.

S. V. Bhosale et al. 2018 [5] used three algorithms namely clustering k-means, Apriori and Bayes algorithm, then they hybridized the algorithm for better efficiency of yield prediction and they considered parameters like Area, Rainfall, Soil type and also their system was able to tell which crop is suitable for cultivation based on the mentioned features.

Shruti Mishra et al 2018 [6] have used data mining techniques to predict the yield of the crop. The datasets gathered are production, area, and season and are mainly focused on the major districts of Maharashtra. They have used four main methods: J48, LWL, LAD Tree, and IBK for prediction. The prediction goes on with four different algorithms. Each of the models will produce results and all results will be compared to each other. The performance of each model was compared using the WEKA tool. Finally based on the accuracy of high and has less value of error will be finalized. Based on the results they concluded the IBK method is an effective model which has high accuracy and less value of error.

Amitabha Chakrabarty et al 2018 [7] had developed a crop selection and crop yield prediction in the region of Bangladesh at minimum cost. They suggested a deep neural network algorithm for crop selection and yield prediction. Mainly the analysis is based on the 46 parameters like type of soil, soil texture, soil composition, type of fertilizers, etc. The prediction model is mainly based on the six major crops- Aus rice, Aman rice, Boro rice, Jute, Wheat, and Potato. They have compared the neural network model with the other three models like Support vector machine, Logistic Regression, and Random Forest algorithm. Based on the performance they concluded that deep neural network works effectively.

S. Veenadhari et al 2014 [8] had developed a crop production model in the selected regions of Madhya Pradesh. A website called crop advisor has been developed for user access. They collected 20 years of agricultural data for their prediction. Crop prediction analysis is mainly based on four major crops: Soybean, Maize, Paddy, and Wheat. The prediction has fully relied on climatic factors and the parameters included are minimum temperature, maximum temperature, rainfall, cloud cover. C4.5 decision tree algorithms were used for crop production.

Potnuru Sai Nishant et al 2020 [9] build a machine learning model for crop yield prediction. The system was very simple and can be directly used by farmers. They have used regression models- Lasso, ENet, Kernel Ridge to predict the crop yield. Stacking regression was used to enhance the regression model.

Aruvansh Nigam et al 2019 [10] have analyzed the best model to predict crop yield. The parameters used are temperature and rainfall. They compared four machine learning algorithms- XGBoost, Random Forest, KNN classifier, Logistic classifier. Finally concluded that the random forest classifier has the best accuracy compared to other algorithms.

Namgiri Suresh et al 2021 [11] have proposed a crop yield prediction system using the Random Forest algorithm. The parameters included are- temperature, rainfall, humidity.

Twenty years of agricultural data have been analyzed for the prediction. Crop selection methods have been used to predict crop yield. CSM allows the farmer to analyze the previously cultivated crops in a specific area. So that farmers can have an idea about what to cultivate next time

III. PROPOSED SYSTEM METHODOLOGY

The proposed system includes two machine learning algorithms. The system aims to perform two tasks- predicting the yield of crops and recommending suitable fertilizer. For the prediction of crop yield AdaBoost algorithm and for fertilizer recommendation Random Forest algorithm were used. All the necessary data are collected through the government web portal. Based on the requirement of the model the raw datasets are altered, preprocessed, and null values are removed based on the rows values. Based on the user choice either he/she can go for crop yield prediction or fertilizer recommendation or both. The user is asked to enter the details of necessary parameters like crop name, season, state, district, area. The values are passed into a machine learning model-AdaBoost. From the above input values yield of a crop is predicted and shown in tonnes. Coming to the fertilizer recommendation model, the user is asked to enter the details like state, district, crop name, soil type, NPK values-Nitrogen, Phosphorous, Potassium. The values entered are passed into a random forest model and the output fertilizer name is displayed.

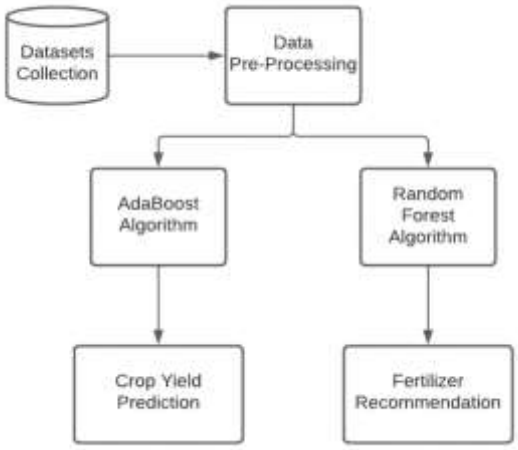

Fig. 1 Process of model

A. Datasets Description

We have collected the past twenty years of agricultural data across India to function this system. So that the model can be used in any region of India. The prediction of crop yield is based on the six factors and the recommendation of fertilizer is based on six factors mentioned below.

The parameters used for this model are:

For crop yield prediction:

- State
- District
- Season
- Crop
- Area
- Production

For fertilizer recommendation:

- Temperature
- Humidity
- Moisture
- Soil type
- Crop type
- NPK

The datasets include crop data which are across India. In India, large varieties of crops are cultivated. It is difficult to analyze all varieties of the crop. So that we have taken major 10 crops which are widely farmed. Coming to the pre-processing, all the data are converted to a numerical form of data. The same has been followed to fertilizer data.

List of crops considered:

- Rice
- Maize
- Blackgram
- Lentil
- Banana
- Mango
- Grapes
- Apple
- Orange
- Papaya

B. AdaBoost Algorithm

AdaBoost is also known as “Adaptive Boosting” algorithm which is suitable for both regression and classification problems. AdaBoost is one of the boosting techniques, itself an ensemble model and usually composed of a group of decision trees. In simple words, AdaBoost combines weak learners into strong learners. Initially, the preprocessed original datasets are passed into a machine learning model-AdaBoost, based on the dataset passed the prediction starts from the first iteration. In this dataset passed are separated into weak learners and strong learners. The weight of the weak learners is increased and combined with strong learners, then made into a newly updated dataset. This updated dataset is passed to the next iteration of the model. The process is done till the error gets minimum. This is known as boosting; the main aim is to improve the accuracy by minimizing the rate of error. The main reason to choose AdaBoost is, it combines multiple classifiers to increase the accuracy. In other words, it combines multiple weak classifiers to build a strong classifier model so that the accuracy of the model gets increased. AdaBoost is an ensemble model. An ensemble model makes better predictions and better performance compared to the activity of a single model. It also works on fewer parameters and reduces the overfitting problem.

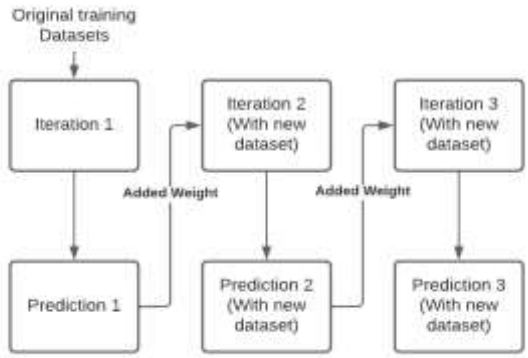

Fig. 2 AdaBoost flowchart

C. *Random Forest Algorithm*

Random Forest is a supervised machine learning algorithm that is used for both regression and classification problems. Random Forest itself is an ensemble method of problem-solving. It combines a cluster of decision trees and the prediction process goes by, taking the average and voting of other models. The final prediction of the regression problem is done by taking the average of other models and for the classification problem majority voting-based prediction process is undertaken. It works on the concept of bagging. Initially, the original datasets are passed as a random sample to each decision tree with a replacement called sampling. Hence each decision tree will produce its individual prediction. Based on the majority voting the output is finalized. Each subset of data is shared with each decision tree, so that the results produced have high accuracy. Comparing both regression and classification problems, the algorithm works effectively in classification. The reason for choosing Random Forest is, also an ensemble model, it reduces overfitting problems and reduces variance which tends to increase the accuracy. It can handle missing values automatically and is very stable that is any change in one decision tree will not affect another one. Also, have less impact over the noise.

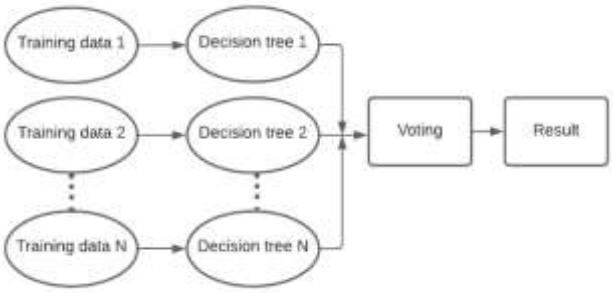

Fig. 3 Random Forest flowchart

IV. Result and Analysis

Crop yield prediction:

```
#SAMPLE OUTPUT
#State_Name,  District_Name,  Season, Crop, Area

X1=[[ 0,  423,  1,  10, 102.0]]
p=AdaBoost.predict(X1)
A=p[0]/1000
print("Yield: {0} tonnes".format(A))
```

```
Yield: 685.862623046875 tonnes
```

Fig. 4 Crop yield prediction-Output

```
print("Accuracy:%s"%(AdaBoost.score(X_test,Y_test)*100))
```

```
Accuracy:82.22138262028261
```

Fig. 5 Crop yield prediction- Accuracy

From the given datasets random values are passed as input to the proposed model. The above image Fig.4 and Fig.5 shows the performance of the model. Based on the input values it throws an output: Yield: 685 tonnes and has an accuracy of 82%.

```
#SAMPLE OUTPUT
#[Temparature,  Humidity, Moisture, Soil Type, Crop Type, N,P,K

X1= [[26, 52, 38, 4,  3,  37, 0,  0]]
p=model.predict(X1)
print("Fertilizer Name:",p[0])
```

```
Fertilizer Name: 6
```

Fig. 6 Fertilizer Recommendation-Output

Table-1: Fertilizer recommendation accuracy score

	Precision	*Recall*	*F1-Score*
0	1.00	0.75	0.86
1	1.00	0.89	0.94
2	0.50	1.00	0.67
3	1.00	1.00	1.00
4	1.00	1.00	1.00
5	1.00	1.00	1.00
6	1.00	1.00	1.00
Macro Avg	0.93	0.95	0.92
Weighted Avg	0.97	0.95	0.96
Accuracy		0.95	

Next to the yield prediction here comes fertilizer recommendation. Fig.6 shows a sample output of the recommendation model. From the given datasets random values are passed as input. In response to that, the output of the fertilizer name is shown in terms of numbers. Table-1 shows the performance of the model and throws an accuracy of 95%, the classification report state that macro average of precision: 93%, macro average of recall: 95%, macro average of f1 score: 92%, and weighted average of precision: 97%, weighted average of recall: 95%, weighted average of f1 score: 95%. The error metric used in this

model is mean absolute error and the error rate was 3530.417713.

V. CONCLUSION

The proposed system suggested a machine learning algorithm for crop yield prediction and for fertilizer recommendation. The proposed machine learning model successfully produces the output and both the algorithm AdaBoost, Random Forest perform well. In the future, a user-friendly web application is made so that every user can use this and get the yield of a crop and fertilizer recommended. Apart from AdaBoost, the gradient Boost algorithm can be used for this prediction model. Also for random forest, SVM and decision tree algorithm can be used for this prediction model.

VI. REFERENCES

[1] D. Elavarasan and P. M. D. Vincent, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications," in IEEE Access, vol. 8, pp. 86886-86901, 2020, doi: 10.1109/ACCESS.2020

[2] S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0769-0773, doi: 10.1109/ICCSP.2019.8698087.

[3] S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0769-0773, doi: 10.1109/ICCSP.2019.8698087.

[4] N. Gandhi and L. J. Armstrong, "Rice crop yield forecasting of the tropical wet and dry climatic zone of India using data mining techniques," 2016 IE EE International Conference on Advances in Computer Applications (ICACA), 2016, pp. 357-363, doi: 10.1109/ICACA.2016.7887981.

[5] S. Bang, R. Bishnoi, A. S. Chauhan, A. K. Dixit and I. Chawla, "Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models," 2019 Twelfth International Conference on Contemporary Computing (IC3), 2019, pp. 1-6, doi: 10.1109/IC3.2019.8844901.

[6] S. V. Bhosale, R. A. Thombare, P. G. Dhemey and A. N. Chaudhari, "Crop Yield Prediction Using Data Analytics and Hybrid Approach," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697806.

[7] 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), 2018, pp. 796-802, doi: 10.1109/ICISC.2018.8398908

[8] A. Suresh, P. Ganesh Kumar and M. Ramalatha, "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), 2018, pp. 88-93, doi: 10.1109/CESYS.2018.8723956.

[9] T. Islam, T. A. Chisty and A. Chakrabarty, "A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh,"2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2018, pp. 1-6, doi: 10.1109/R10-HTC.2018.8629828.

[10] S. Veenadhari, B. Misra and C. Singh, "Machine learning approach for forecasting crop yield based on climatic parameters," 2014 International Conference on Computer Communication and Informatics, 2014, pp. 1-5, doi: 10.1109/ICCCI.2014.6921718.

[11] P. S. Nishant, P. Sai Venkat, B. L. Avinash and B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-4, doi: 10.1109/INCET49848.2020.9154036.

[12] A. Nigam, S. Garg, A. Agrawal and P. Agrawal, "Crop Yield Prediction Using Machine Learning Algorithms," 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 125-130, doi: 10.1109/ICIIP47207.2019.8985951.

[13] N. Suresh et al., "Crop Yield Prediction Using Random Forest Algorithm," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 279-282, doi: 10.1109/ICACCS51430.2021.9441871.

[14] Y. Gandge and Sandhya, "A study on various data mining techniques for crop yield prediction," 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017, pp. 420-423, doi: 10.1109/ICEECCOT.2017.8284541.

[15] Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha and V. G. R. R. Devi, "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 736-741, doi: 10.1109/ICCES48766.2020.9137868.

[16] R. Medar, V. S. Rajpurohit and S. Shweta, "Crop Yield Prediction using Machine Learning Techniques," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-5, doi: 10.1109/I2CT45611.2019.9033611.