A Novel Approach for Crop Yield Prediction Model Validation

Moumita Goswami[1], Sanghita Bhattacharjee[2], Suvamoy Changder[3]

MCKV Institute of Engineering, India[1]

National Institute of Technology Durgapur, India[2,3]

Corresponding author: Moumita Goswami, Email: moumita02314@gmail.com

Agriculture is the backbone of the Indian economy, which is mainly dependent on nature. Nowadays, due to the random climatic changes, farmers are struggling to get a good amount of yield from the crops. Smart Agriculture is one of the crop management concepts that allows farmers to control the farm in all aspects. Crop prediction is a challenging task in the area of agriculture to increase crop productivity. Machine learning plays a vital role in crop yield prediction, including supportive decisions on what crops to produce. The main objective of this paper is to provide a double filter mechanism to focus on the validation of the model, thus predicting the suitable crops to be farmed on. Several supervised machine learning classifiers have been applied in our work for predicting an appropriate crop based on different soil and weather parameters. When the accuracy of the classifiers is almost the same, it becomes difficult to identify a proper classifier for predicting suitable crops. In our work, we have focused on the validation of the classifier model which further helps us to find the appropriate classifier to support crop yield prediction. The clustering algorithm is used for testing and validating the correctness of the proper classifier. The experimental results show that every model is not applicable in all the cases of crop prediction. Hence there is a gap between model prediction and actual implementation. This paper bridges the gap between the selection procedure.

Keywords: Agriculture, Classification algorithms, Accuracy, Validation, Clustering algorithm, Clusters.

2023. In Satyasai Jagannath Nanda & Rajendra Prasad Yadav (eds.), *Data Science and Intelligent Computing Techniques*, 103–116. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.56155/978-81-955020-2-8-10

1. Introduction

Agriculture has not only a huge aspect of the growing economy, but it is also important for humankind to survive. Previously, crop cultivation depended on farmers' hands-on expertise. However, weather change has begun to affect crop yields badly. As a result, the farmers were unable to select the proper crops based on different soil and environmental factors. They start growing the same types of crops repeatedly without trying a new variety of crops. Moreover, they started applying fertilizers in random quantities without knowing the actual quantity, which directly affects the crop yield and damages the top layer of soil. The crop management study [1] includes versatile aspects that are initiated from the combination of farming techniques in managing the biological, chemical, and physical crop environment to reach both quantitative and qualitative goals. IoT-based technologies with machine learning in the agriculture industry created revolutionary changes to existing farming methods [2-5]. Using advanced approaches to control crops, such as yield prediction, disease detection, weed detection, crop recognition, and crop quality, the growth of productivity can be controlled [6]. Furthermore, these techniques can also be applied for deciding the right amount of fertilizers for farmland, which in turn reduces human labour, improves crop cultivation, and minimizes the wastage of water in the field.

Researchers in different fields of agriculture have developed several forecasting methodologies to recognize the most suitable crop for specific tracks of land. Crop yield predictions are carried out to estimate higher crop yield which is one of the challenging issues in the agricultural sector. It is dependent on different input features such as temperature, humidity, characteristics of soil etc., and also crop yield prediction algorithms. The accuracy of crop yield can be obtained by acquiring proper inputs and models without hampering the production environment of agriculture. Input features for agriculture may differ from region to region and are intimidating to gather such information over large areas of land. Machine learning algorithms become a decision-making tool used for crop yield prediction to make decisions on what crops to grow and what to do during the growing season of the crops. It also determines patterns and establishes the correlations among the parameters of datasets. Many different models have been experimented with and tested on different crop datasets to find the appropriate crops in the past few years. This problem requires the use of numerous datasets since crop yield depends on many different parameters such as climate, soil, use of fertilizer, seed variety, etc.

Although crop yield prediction models can evaluate the actual yield reasonably, better prediction is still a challenging task. Existing methodologies investigated crop yield prediction with machine learning algorithms, which differ from the features and the results are not evident enough to predict the best model. Most of the studies choose the models by checking their accuracy. However,

they did not focus on the validation test, which has paramount importance in selecting the proper model for predicting crops when many models give better accuracy. In this paper, we introduce a double filter mechanism that mainly focuses on the validation of the model to find the appropriate classifiers for predicting crops. We have applied several supervised learning algorithms and ensemble learning in predicting crops and then compared their accuracy. It is noticed that most of the classifiers have achieved an accuracy of more than 92%. Nevertheless, by comparing the accuracy of all the classifiers, it becomes difficult to decide the appropriate model for cultivating suitable crops. Therefore, we have further used unsupervised clustering algorithms to test the correctness of the result of each classifier that might help to find the correct model for crop yield prediction. The experimental results of validation reflect that the Logistic Regression model and Stacking Classifier give successful outcomes. However, others fail to achieve the same. Among these two, the stacking classifier provides a better result with higher accuracy than logistic regression.

The remainder of this paper is organized as follows: Section 2 discusses the related work. Section 3 describes the proposed methodology. Section 4 shows the experimental results and finally, the paper is concluded in Section 5.

2. Related Work

Crop yield prediction is dependent on different input features and also on crop yield prediction algorithms. Different machine and deep learning algorithms have been applied in prediction in the past research works. In this section, we explore some of crop prediction algorithms.

Crop recommendation using machine learning consists of several phases, like data collection, data pre-processing, data partitioning, and finally data analysis [7-9]. Classifiers like Logistic Regression, Naïve Bayes, and Random Forest, are used to find the suitable crop for particular land based on soil and weather parameters that will help the farmers to make a decision. In a work [8], the authors discussed various wrapper feature selection methods for crop prediction. The LASSO method and machine learning models are discussed in [9] to build various empirical models for wheat yield prediction in Australia. Further, different empirical and mathematical yield modeling methods have been implemented for different crops in different localities [10, 11]. In some recent works [12, 13], satellite-based remote sensing techniques were explored in predictive yield modeling. Furthermore, these techniques are also used to survey soil and crop attributes responsible for variations in crop yield which can allow real-time site-specific management of fertilizers, pesticides, or irrigation and provides data to map cropped regions for Precision Agriculture.

Apart from these studies, some recent works use deep learning models in predicting suitable crops. For example, in [14], the authors discussed different algorithms and features that have been used in crop yield prediction studies. In their analysis, the authors considered the features i.e.,

temperature, humidity, rainfall, soil type, and Artificial Neural Networks (ANN) for predicting crops. Similarly, the author in [15] applied Convolutional Neural Networks (CNNs) models for predicting crops.

Ensemble Learning (EL) also contributes to the ML models used in agricultural systems [6, 16]. EL is a concise term for methods that integrate multiple inducers for taking appropriate decisions. The key feature of EL is that via combining various models, the errors generated from a single model are likely to be compensated by other models. Accordingly, the prediction of the overall performance would be superior and more accurate as compared to a single model. In a recent work [17], deep learning is used to cultivate crops efficiently and achieve high productivity at low cost. It also helps to calculate the total predicted cost needed for cultivation. In the article [18], IoT with ML and WSN are used for Precision Agriculture that provides the mechanism for monitoring the agriculture parameters along with predicting the farmland or crop requirement such as irrigation requirement prediction. In this paper [19], different machine learning models are used to create two different services: one for recommending the suitable crop to grow based on soil and the region's weather characteristics, and another for the estimating of the hourly average air temperature.

The system in [20] comes with a model which is developed to predict crop yield and also recommending required fertilizer ratio based on atmospheric and soil parameters of the land for increasing the crop yield and farmer revenue.

It is evident from the above discussion that most of the algorithms used a single model to predict the crops. In our work, we have considered both single-level models and ensemble learning, namely voting classifier and stacking classifier to predict the accurate crop and also compared their performance. Furthermore, validation of each model is tested and verified to identify the proper classifier beyond the accuracy of each classification algorithm.

3. Methodology

We followed the supervised learning approaches in crop prediction and validate each of them for correctness. To predict the suitable crops, we considered the dataset comprised of seven attributes, discussed in the following section. The proposed framework is divided into four steps: Data acquisition, preprocessing, crop prediction, and model validation as shown in Fig. 1.

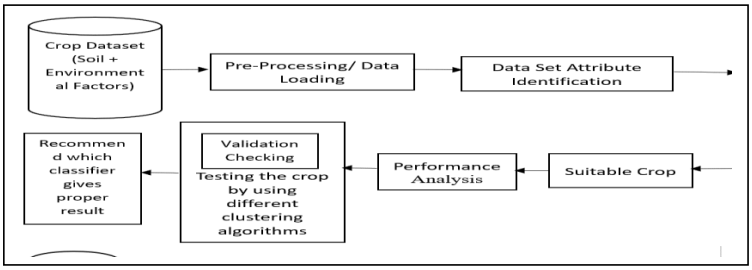

Fig. 1: Process flow of proposed system

3.1. Dataset Preparation

Data acquisition is the first step in recommendation system. We obtained the dataset of crop recommendation from the Kaggle site [21]. The dataset includes the features like Nitrogen (N), Phosphorus (P) and Potassium (K), Temperature, Humidity, PH, Rainfall etc. Fig.2 shows the sample labelled dataset with feature values. The relationship between features in the dataset is shown using the correlation matrix as given in Fig. 3. This matrix helps to determine the dependency between multiple variables at the same time. The correlation matrix shows that feature 'P' is highly correlated with feature 'K' whereas it is poorly correlated with 'N'. Similarly, 'Temperature' is highly correlated with 'Humidity' but poorly correlated with 'K' and 'P'.

3.2. Preprocessing

Once the data is collected, the first task is data cleaning or preprocessing. Preprocessing helps to remove missing values, duplication, and null values present in the dataset. It is noted that our data set is already pre-processed and a clean data set containing no missing or duplicate values. The final data is used for the prediction process.

N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.8797437	82.00274423	6.502985	202.9355	rice
85	58	41	21.7704617	80.31964408	7.038096	226.6555	rice
60	55	44	23.0044592	82.3207629	7.840207	263.9642	rice
74	35	40	26.4910964	80.15836264	6.980401	242.864	rice
78	42	42	20.1301748	81.60487287	7.628473	262.7173	rice
69	37	42	23.0580487	83.37011772	7.073454	251.055	rice
69	55	38	22.708838	82.63941394	5.700806	271.3249	rice
94	53	40	20.2777436	82.89408619	5.718627	241.9742	rice
89	54	38	24.5158807	83.5352163	6.685346	230.4462	rice
68	58	38	23.2239739	83.03322691	6.336254	221.2092	rice
91	53	40	26.5272351	81.41753846	5.386168	264.6149	rice
90	46	42	23.9789822	81.45061596	7.502834	250.0832	rice
78	58	44	26.800796	80.88684822	5.108682	284.4365	rice
93	56	36	24.0149762	82.05687182	6.984354	185.2773	rice

Fig.2: Characteristics of Dataset

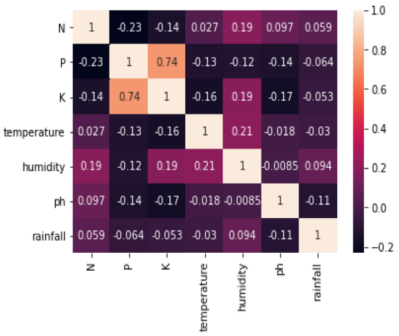

Fig.3: Correlation among attributes

3.3. Machine Learning Models for Crop Prediction

Different algorithms and techniques are used in machine learning for agricultural data analysis. In this paper, we have comprehensively compared the performance of various machine learning models including Logistic Regression, Support Vector Machine, Random Forest, etc, and Ensemble Learning like Voting Classifier, Stacking Classifier in crop yield prediction briefly described below and then validated the models to find a suitable one which will recommend the most accurate crop for the farmers.

(1) *Logistic Regression (LR)* - Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. It is most helpful for understanding the effect of several independent variables on a single target variable.
(2) *Naive Bayes (NB)* - Naive Bayes classifier is also a supervised learning classification algorithm based on the Bayes theorem which is used to make quick predictions. The Naïve Bayes model is one of the simple and most effective classification algorithms which is easy to build and particularly useful for very large datasets.
(3) *Random Forest (RF)* - Random Forest is a famous machine learning algorithm that can be used in both classification and regression problems in ML. It can examine crop growth related to the current climatic and soil conditions. The Random forest algorithm consists of many decisions trees on various subsets of the given data samples and then predicts the data from each subset and then by voting gives better predictive accuracy and final output for the system. It uses bagging and feature randomness to train the data, increasing the accuracy of the result and preventing the overfitting problem.
(4) *Decision tree (DT)* - The Decision Tree is a popular predictive model in the form of a tree structure that breaks a dataset into smaller and smaller subsets. It is used to test the conditions at each tree level and move down the tree where different decisions are recognized [34]. So, it produces a sequence of rules that can be used to classify the data and provide solutions based on given conditions. The root node is considered the topmost decision node in a tree that corresponds to the best predictor.
(5) *Support vector machine (SVM)* – The Support Vector Machine is one of the most popular Supervised Learning algorithms, which is used for classification, regression, and outliers detection. It breaks data into different categories, which further separates the data into two hyperplane groups. Training points specify the vector which helps in creating the hyperplane.
(6) *Ensemble Learning* - The ensemble model is a predictive model that is used in this study to combine decisions from multiple models to improve the overall performance. In this study, three diverse algorithms such as logistic regression, decision tree, and random forest were combined to improve model performance. Many predictions are used from two or more models to produce one optimal predictive model. In this approach, voting and stacking classifier are used to make the final predictions as shown in.
Voting classifier (VC) - Voting Classifier is an ensemble learning technique mainly used for classification problems. This method consists of building multiple models independently and getting their individual output based on their highest probability of chosen class called 'vote'. In this classification model, the predictions for each label are summed and the label with the majority vote is predicted. It is a technique that may be used to improve model performance than any single model used in the ensemble.
In our model three classification models (logistic regression, decision tree, and random forest) are combined using sklearn.VotingClassifier. Then the model is trained and the class with maximum votes is returned as output.
Stacking classifier (SC) - Stacking is an ensemble method that combines multiple models (classification or regression) via meta-model (meta-classifier or meta-regression). The base models are trained on the complete dataset, then the meta-model is trained on values returned (as output) from base models. The base models in stacking are typically different and the meta-model helps to find the features from base models to achieve the best accuracy. Stacking features are first extracted by training the dataset with all the first-level models. In our model, three classification models (logistic regression, decision tree, and random forest) are used by using sklearn.StackingClassifier.

3.4. Model Validation

To validate the correctness of each model in crop prediction, we applied unsupervised clustering algorithms. Clustering divides the data into multiple groups which are meaningful and useful for data summarization. In our study, we have used K-Means and Fuzzy C-Means algorithms. Fig. 4 displays the sequential steps of the validation method. These clustering algorithms are applied to data to form the clusters of different types of crops based on the seven features. The formation of four clusters by K-Means is depicted in Fig 5. In each cluster, we find the centroid values in order to validate the correctness of the machine learning algorithms.

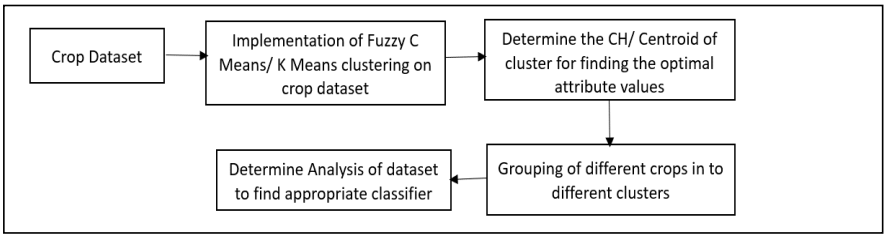

Fig. 4: Steps of the validation test to find correctness of classifier

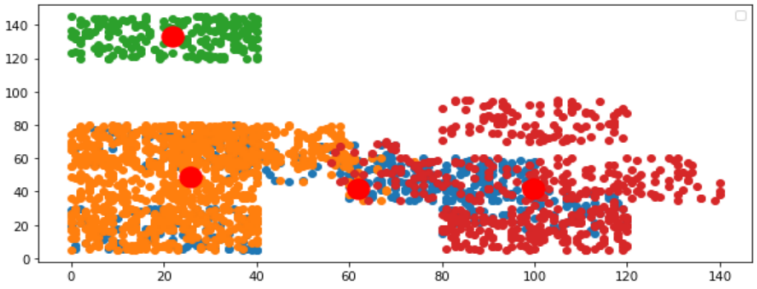

Fig. 5: Clusters by K-Means

4. Experimental Results

4.1. Model Performance Analysis

For the experimental analysis, we have developed all algorithms as mentioned in section 3.2 by using Python libraries such as pandas, numpy, matplotlib, sklearn etc. To test the performance, the dataset is divided into two parts (training and testing) with 80:20 ratio. We have collected a total of 2200 data samples. Among 2200 data samples, 1760 (80%) samples are used to train the machine learning models and 440 (20%) data samples are used to test how accurately the models predict the crop. Table 1 shows the predicted crop on the basis of input features. The accuracy, precision, recall and F1-score of all the classifiers are shown in Table 2. The results reveal the DT has obtained a low performance in our work. The SVM and LR have shown better performance than DT. But RF, Naive Bayes, voting, and stacking classifiers have obtained very good performance with an accuracy of more than 99 %. Since many classifiers have given better results in prediction, it is more important to validate the correctness of each classifier.

Table 1. Execution result of the different classifiers

Classifier	N	P	K	Temperature	Humidity	Ph	Rainfall	Predicted Crop
Logistic Regression (LR)	89	46	61	28	70.3	7.0	150.9	jute
	104	18	30	23.6	60.3	6.7	140.91	coffee
	22.67	131.59	196.55	23.25	5.99	86.44	91.55	grapes
Support Vector Machine (SVM)	89	46	61	28	70.3	7.0	150.9	jute
	104	18	30	23.6	60.3	6.7	140.91	coffee
	22.67	131.59	196.55	23.25	5.99	86.44	91.55	grapes
Naïve Bayes Algorithm (NB)	89	46	61	28	70.3	7.0	150.9	jute
	104	18	30	23.6	60.3	6.7	140.91	coffee
	22.67	131.59	196.55	23.25	5.99	86.44	91.55	mothbeans
Random Forest (RF)	89	46	61	28	70.3	7.0	150.9	jute
	104	18	30	23.6	60.3	6.7	140.91	coffee
	22.67	131.59	196.55	23.25	5.99	86.44	91.55	chickpea
Decision Tree (DT)	89	46	61	28	70.3	7.0	150.9	coffee
	104	18	30	23.6	60.3	6.7	140.91	coffee
	22.67	131.59	196.55	23.25	5.99	86.44	91.55	chickpea
Voting Classifier [DT+ RF + SVM]	89	46	61	28	70.3	7.0	150.9	coffee
	104	18	30	23.6	60.3	6.7	140.91	coffee
	22.67	131.59	196.55	23.25	5.99	86.44	91.55	apple
Voting Classifier [LR + DT+ RF]	89	46	61	28	70.3	7.0	150.9	coffee
	104	18	30	23.6	60.3	6.7	140.91	coffee
	22.67	131.59	196.55	23.25	5.99	86.44	91.55	grapes
Stacking classifier [LR + DT+ RF]	89	46	61	28	70.3	7.0	150.9	coffee
	104	18	30	23.6	60.3	6.7	140.91	coffee
	22.67	131.59	196.55	23.25	5.99	86.44	91.55	apple

Table 2. Accuracy comparison of classifiers

Classification Algorithms	Accuracy (%)	Precision	Recall	F1-score
Logistic Regression	96.81	0.97	0.97	0.97
Support Vector Machine	98.86	0.99	0.99	0.99
Naïve Bayes Algorithm	99.31	0.99	0.99	0.99
Random Forest Algorithm	99.55	1	1	1
Decision Tree	92.5	0.95	0.93	0.92
Voting Classifier [Decision Tree+ Random Forest + SVM]	99.77	1	1	1
Voting Classifier [Logistic Regression + Decision Tree+ Random Forest]	99.54	1	1	1
Stacking classifier [Logistic Regression + Decision Tree+ Random Forest]	99.31	0.99	0.99	0.99

4.2. Validation Test of Different Classifiers

In order to validate the correctness of the classifiers, centroid values are calculated for each cluster using K-Means and Fuzzy-C Means algorithm and the result is shown in Table 3. It is also noticed that both the algorithms give almost the same output. Table 4 shows the predicted crop of each classifier residing within the appropriate cluster based on the central values of each independent attribute of clusters. Table 5 tabulates the validation result of each classifier where we can take the decision that which classification algorithm gives a better result concerning seven centroid values of each attribute of the dataset. However, the RF, Naive Bayes, voting, and stacking classifiers achieve accuracy above 99% as given in Table 2 but fail in the validation test in our work. However, LR with an accuracy of 96.81 and stacking classifier with an accuracy of 99.31 give successful results for all four clusters. Therefore, the stacking classifier best option for predicting crops for our data sample.

Table 3. Different cluster values with the centroid of two Algorithms

Name of Clustering Algorithm		Clusters	Centroid Values
K-Means Clustering	Cluster 1	['rice' 'pigeonpeas' 'papaya' 'coconut' 'jute' 'coffee']	The centers of the four clusters are: [61.79957806 42.01054852 35.25738397 26.50844776 6.43627149 77.56013583 190.14025778]
	Cluster 2	['maize' 'chickpea' 'kidneybeans' 'pigeonpeas' 'mothbeans' 'mungbean' 'blackgram' 'lentil' 'pomegranate' 'mango' 'orange' 'papaya' 'coconut']	[25.68384539 48.60356789 28.79682854 25.44479187 6.60337505 60.73631456 81.84376106]
	Cluster 3	['grapes' 'apple']	[21.99 133.375 200. 23.24025877 5.97779981 87.1043052 91.13330408]
	Cluster 4	['maize' 'banana' 'watermelon' 'muskmelon' 'papaya' 'cotton' 'coffee']	[99.82205029 42.10638298 38.99419729 26.05200288 6.42881639 80.83676492 70.96043328]
Fuzzy C means Clustering	Cluster 1	['rice' 'kidneybeans' 'pigeonpeas' 'papaya' 'coconut' 'jute' 'coffee']	The four centers of the clusters are: [63.43988311 43.41180703 36.4620432 26.11075124 6.47474181 77.67972482 184.320222]
	Cluster 2	['maize' 'banana' 'watermelon' 'muskmelon' 'papaya' 'cotton' 'coffee']	[87.82521493 40.41315515 36.53753067 25.46764207 6.51601581 77.41401134 73.29941946]
	Cluster 3	['maize' 'chickpea' 'kidneybeans' 'pigeonpeas' 'mothbeans' 'mungbean' 'blackgram' 'lentil' 'pomegranate' 'mango' 'orange' 'papaya' 'coconut']	[27.32206712 49.88992144 26.10604285 26.56231263 6.59039967 62.18783152 76.64530518]
	Cluster 4	['grapes' 'apple']	[22.66943217 131.58878695 196.55546008 23.24713944 5.99146543 86.43632253 91.55179037]

Table 4. Validation by checking the result of each classifier

Cluster No.	Cluster points (N,P,K,Temp, Hum,Ph,RH)	Crops in a Cluster	LR Result	SVM Result	NB Result	RF Result	DT Result	VC1 Result	VC2 Result	SC Result
Cluster1	[63.43988311 43.41180703 36.46204326 26.11075124 6.47474181 77.67972482 184.320222]	['rice' 'kidneybeans' 'pigeonpeas' 'papaya' 'coconut' 'jute' 'coffee']	coffee	kidney beans	mothbeans	coffee	coffee	coffee	chickpea	kidney beans
Cluster2	[87.82521493 40.41315515 36.53753067 25.46764207 6.51601581 77.41401134 73.29941946]	['maize' 'banana' 'watermelon' 'muskmelon' 'papaya' 'cotton' 'coffee']	coffee	chickpea	mothbeans	coffee	mango	chickpea	coffee	maize
Cluster3	[27.32206712 49.88992144 26.10604285 26.56231263 6.590399676 2.187831527 6.645305178]	['maize' 'chickpea' 'kidneybeans' 'pigeonpeas' 'mothbeans' 'mungbean' 'blackgram' 'lentil' 'pomegranate' 'mango' 'orange' 'papaya' 'coconut']	mothbeans	kidney beans	mothbeans	chickpea	mango	kidney beans	chickpea	kidney beans

Cluster4	[22.66943217 131.58878695 196.55546008 23.24713944 5.99146543 86.43632253 91.55179037]	['grapes' 'apple']	pes	es	bean s	pea	pea	e	s	e

5. Conclusion

This paper highlights the use of machine learning algorithms in predicting crops. We have selected a dataset of seven attributes and applied eight machine learning algorithms to the data set. The random forest, naïve bayes, voting classifier, and stacking classifier have achieved higher accuracy in prediction. To find the best model in crop prediction, we further validate the models for correctness. We have noticed that through voting classifiers, naïve bayes and random forest have obtained a very good accuracy (> 99%) in prediction, but they failed in the validation test. The results also confirm that the stacking classifier and logistic regression give successful outcomes for all the four clusters based on seven centroid values of each attribute of the dataset. However, the stacking classifier has obtained higher accuracy (> 99%) than logistic regression and becomes the best model for crop prediction in our work. In our future work, we can use this validation model in the prediction of the appropriate amount of fertilizer for the farmland.

Table 5. Correctness checking of each classifier

SL. No	Classification Algorithms	Cluster1	Cluster2	Cluster3	Cluster4
1	Logistic Regression	Successful	Successful	Successful	Successful
2	Support Vector Machine	Successful	X	Successful	Successful
3	Naïve Bayes Algorithm	X	X	Successful	X
4	Random Forest Algorithm	Successful	Successful	Successful	X
5	Decision Tree	Successful	X	Successful	X
6	Voting Classifier [Decision Tree+ Random Forest + SVM]	Successful	X	Successful	Successful
7	Voting Classifier [Logistic Regression + Decision Tree + Random Forest]	X	Successful	Successful	Successful
8	Stacking classifier [Logistic Regression + Decision Tree + Random Forest]	Successful	Successful	Successful	Successful

References

1. Yvoz, S.; Petit, S.; Biju-Duval, L.; Cordeau, S.: A framework to type crop management strategies within a production situation to improve the comprehension of weed communities. European Journal of Agronomy 115, 1-23 (2020).

2. Muhammad, S. F., Samyla, R., Adnan, A., Kamran, A., Muhammad, A. N.: A survey on the role of iot in agriculture for the implementation of smart farming. IEEE ACCESS 7, 156237 – 156271(2019)

3. Muhammad, A.; Mohammad, A. U.; Zubair, S.; Mansour, A. ; El-Hadi M. A..: Internet-of-Things (IoT) based smart agriculture: toward making the fields talk. IEEE ACCESS 7, 129551 – 129583(2019).

4. Devi K, Komal.; Premkumar, Josephine.; Kavitha, K.; Anitha, P.; Kumar, M. Sathish.; Mahaveerakannan, R.: A review: smart farming using IoT in the area of crop monitoring. Annals of the omanian Society for Cell Biology 25, 3887–3896 (2021).

5. Kour, Vippon. Preet.; Arora, Sakshi.: Recent developments of the internet of things in agriculture: A survey. IEEE ACCESS 8, 129924 - 129957 (2020).

6. Benos, L.; Tagarakis, Aristotelis C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D.; Machine learning in agriculture: A comprehensive updated review. Sensors 21(11), 3758 (2021).

7. Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V.: Machine learning applications for precision agriculture: a comprehensive review. IEEE ACCESS 9, 4843 – 4873 (2020).

8. Suruliandia, A.; Mariammal, G.; Raja, S.P.: Crop prediction based on soil and environmental characteristics using feature selection technique. Mathematical and Computer Modelling of Dynamical Systems, 27, 117–140 (2021).

9. Cai, Yaping et al.: Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology, 274, 144-159 (2019).

10. Cousens, R.: An empirical model relating crop yield to weed and crop density and a statistical comparison with other models. The Journal of Agricultural Science 105, 513–521(1985).

11. Dourado-Neto, D et al.: Principles of crop modeling and simulation: I. uses of mathematical models in agricultural science. Scientia Agricola 55, 46–50 (1998).

12. Doraiswamy, P. C.; Moulin, S.; Cook, P.W.; Stern, A.: Crop yield assessment from remote sensing. Photogrammetric, Engineering and Remote Sensing 69, 665–674 (2003).

13. Prasad, A. K.; Chai, L.; Singh, R. P.; Kafatos, M.: Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 8, 26–33 (2006).

14. Van Klompenburg, T.; Kassahun, A.; Catal, C.: Crop yield prediction using machine learning: A systematic literature review. Computer and Electronics in Agriculture. 177, 105709 (2020).

15. Nevavuori, P.; Narra, N.; Lipping, Tarmo.: Crop yield prediction with deep convolutional neural networks. Computers and Electronics in Agriculture 163, 104859 (2019).

16. Sagi, O.; Rokach, L.: Ensemble learning: a survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8, e1249 (2018).

17. Durai , Senthil Kumar Swami.; Shamili, Mary Divya.: Smart farming using Machine Learning and Deep Learning techniques. Decision Analytics Journal 3, 100041 (2022)

18. Singh, Dushyant Kumar.; Sobti, Rajeev.: Role of Internet of Things and Machine Learning in Precision Agriculture: A Short Review. 6th IEEE International Conference on Signal Processing, Computing and Control (ISPCC), 2021.

19. Chouaib, El Hachimi.; Salwa, Belaqziz.; Saidi, Khabba.; Abdelghani, Chehbouni.: Towards precision agriculture in Morocco: A machine learning approach for recommending crops and forecasting weather. International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA), 2021.

20. A, Palaniraj.; S, Balamurugan A.; R, Durga Prasad.; P, Pradeep.: Crop and Fertilizer Recommendation System using Machine Learning. International Research Journal of Engineering and Technology (IRJET) 8, (2021).

21. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset.