Machine Learning Approaches for Crop Yield Prediction

Ms. Sudeshna Roy, Dr. Priya Chandran, Ms. Shravani Pawar
Bharati Vidyapeeth's Institute of Management & Information Technology, Navi Mumbai

Abstract: A major challenge in crop yield prediction is the factors affecting the selection of crop like environment, crop types and soil type. Different machine learning approaches have been used for the prediction of crop yield. This study aims to determine the most accurate and efficient method for predicting crop yields, which could aid farmers in making informed decisions about crop management and improve food security. This research paper explores the effectiveness of machine learning approaches for predicting crop yields. The study used logistic regression, decision tree classifier, random forest classifier, XGBoost and K-Nearest Neighbours algorithms to predict crop yields using a range of variables like N, P, K, temperature, humidity, Ph and rainfall. The results indicate that machine learning approaches have the potential to improve crop yield prediction accuracy and could be an effective tool for crop management in the future. The study shows that random forest algorithm is giving a high accuracy of 98.03 % compared to decision tree, XGBoost and KNN algorithms.

Keywords: crop yield, ML algorithms, prediction, decision tree, linear regression, XBoost, KNN, random forest

1. Introduction

The importance of agriculture to India's population as a whole cannot be underestimated. However, the agriculture industry is currently experiencing its most stressful period in the last 30 years. Numerous issues, some of which are man-made and others of which are natural, are affecting Indian agriculture. In India, the majority of families rely on rural revenues. Any technological combo, if used effectively, can make a significant difference.

In this study, a machine learning model is proposed to obtain crop recommendations with high production guarantees based on a variety of factors (such as atmospheric conditions, type of fertilizer, soil, and seed, etc.). We will display the code used for each step of the creation and assessment of our model, followed by its results. This will make our work easier to reproduce. The Python programming language and a variety of Python packages will be used in this study.

Overall, the proposed machine learning model for crop recommendation in India has the potential to significantly impact the agricultural sector by assisting farmers in making informed decisions, optimizing resource utilization, and increasing crop yields. By leveraging technology and data-driven approaches, the model aims to contribute to the overall growth and sustainability of Indian agriculture.

2. Literature Review

Given the significance of crop prediction, numerous suggestions have been proposed in the past with the goal of improving crop prediction accuracy. In this paper feed-forward back propagation Artificial Neural Network methodology has been approached to model and forecast various crop yields at rural areas based on parameters of soil (PH, nitrogen, potassium, etc.) and parameters related to the atmosphere (rainfall, humidity, etc.) [1].

This paper looks at five of India's most important crops- rice, maize, ragi, sugarcane, and tapioca during a five-year period beginning in 2005. [2]. In order to get the maximum crop productivity, various factors such as

ISSN PRINT 2319 1775 Online 2320 787
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

rainfall, groundwater, and cultivation area, and soil type were used in the analysis. K-Means technique was used for the clustering, and for the classification, the study looked at three different types of algorithms: fuzzy, KNN, and Modified KNN. After the analysis, MKNN gave the best prediction result of the three algorithms. An application for farmers can be created that will aid in the reduction of many problems in the agriculture sector [3]. In this application, farmers perform single/multiple testing by providing input such as crop name, season, and location. As soon as one provides the input, the user can choose a method and mine the outputs. The outputs will show you the crop's yield rate. The findings of the previous year's data are included in the datasets and transformed into a supported format. The machine learning models used are Naïve Bayes and KNN.

To create the dataset, information about crops over the previous ten years was gathered from a variety of sources, such as government websites. An IoT device was setup to collect the atmospheric data using the components like Soil sensors, Dht11 sensor for humidity and temperature, and Arduino Uno with Atmega as a processor. Naive Bayes, a supervised learning algorithm obtaining an accuracy of 97% was further improved by using boosting algorithm, which makes use of weak rule by an iterative process to bring higher accuracy [5]. To anticipate the yield, the study employs advanced regression techniques such as ENet, Kernel Ridge, and Lasso algorithms [4]. The three regression techniques are improved by using Stacking Regression for better prediction. However, when a comparison study is conducted between the existing system and the proposed system employing Naive Bayes and Random Forest, respectively. The proposed system comes out on top. Because it is a bagging method, the random forest algorithm has a high accuracy level, but the Naïve Bayes classifier's accuracy level is lower as the algorithm is probability based. [6].

This paper contributes to the following aspects- (a) Crop production prediction utilizing a range of Machine Learning approaches and a comparison of error rate and accuracy for certain regions. (b) An easy-to-use mobile app that recommends the most gainful crop. (c) A GPS-based location identifier that can be used to obtain rainfall estimates for a specific location. (d) A system that recommends the prime time to apply fertilizers [7]. On the given datasets from Karnataka and Maharashtra, different machine learning algorithms such as KNN, SVM, MLR, Random Forest, and ANN were deployed and assessed for yield to accuracy [9]. The accuracy of the above algorithms is compared. The results show that Decision Tree is the most accurate of the standard algorithms used on the given datasets, with a 99.87% accuracy rate. Regression Analysis is applied to determine the relationship between the three factors: Area Under Cultivation, Food Price Index, and Annual Rainfall and their impact on crop yield. The above three factors are taken as independent variables, and for the dependent variable, crop yield is taken into consideration. The R2 obtained after the implementation of RA shows these three factors showed slight differences indicating their impact on the crop yield [8].

In the proposed paper, the dataset is collected from the government websites such as APMC website, VC Farm Mandya, which contains data related to climatic conditions and soil nutrients [10]. Two machine learning models were used; the model was trained using the Support Vector Machine model with Radial Basis Function kernel for rainfall prediction and Decision Tree for the crop prediction. A comparative study of various machine learning can be applied on a dataset with a view to determine the best performing methodology. The prediction is found by applying the Regression Based Techniques such as Linear, Random Forest, Decision Tree, Gradient Boosting, Polynomial and Ridge on the dataset containing details about the types of crops,

ISSN PRINT 2319 1775 Online 2320 787
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

different states, and climatic conditions under different seasons [11]. The parameters used to estimate the efficiency of these techniques were mean absolute error, root mean square error, mean squared error, R-square, and cross validation. Gradient Boosting gave the best accuracy- 87.9% for the target variable 'Yield' and Random Forest- 98.9% gave the best accuracy for the target value 'Production'.

The DHT22 sensor is recommended for monitoring live temperature and humidity [12]. The surrounding air is measured with a thermistor and a capacitive humidity sensor and outputs a digital signal on the data pin to the Arduino Uno port pin. The humidity value ranges from 0-100% RH and - 40 to 80 degrees Celsius to read the temperature. The above two parameters and soil characteristics are considered as input to three different machine learning models: Support Vector Machine, Decision Tree, and KNN. The Decision Tree gave better accuracy results.

3. Methodology

3.1 Objectives of the Study

The main objectives of this study are as follows:

• To apply data pre-processing and preparation techniques in order to obtain clean data

• To build machine learning models able to recommend a crop based on crop features

• To analyse and compare models' performance in order to choose the best model which will give more throughput.

In this study, we will use a crop growth dataset. This dataset was created by enhancing rainfall, climate, and fertiliser data sets that were previously accessible for India and for the prediction of crop we are using machine earning algorithms logistic regression (LR), decision tree classifier (DT), random forest classifier (RF), XGBoost and K-Nearest Neighbours (KNN).

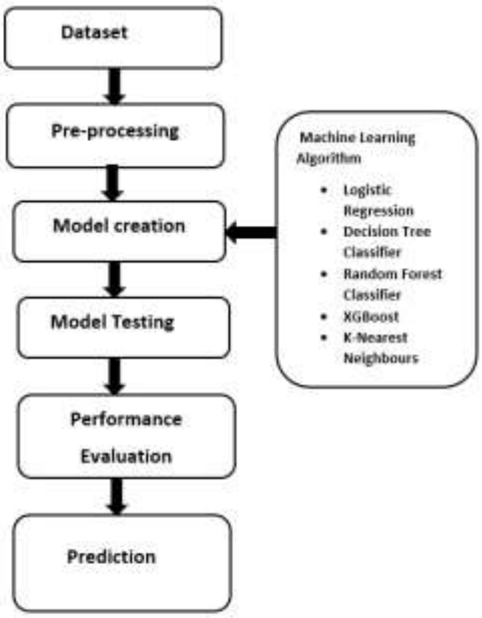


Fig 1 Flow of the system

ISSN PRINT 2319 1775 Online 2320 787
Research Paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

3.2 Model Development

1. Logistic Regression

LR is a statistical analysis model to predict a binary outcome division (1 or 0, Yes or No, True or False) for a set of independent variables. When the response variable is categorical and the log of probabilities is used as the dependent variable, LR can be thought of as a specific case of linear regression. Simply described, it determines the probability of an event happening based on how data match a logic function. Sigmoid function is a function that always has a value in the range [0, 1], continuous and easy to use.

Algorithm:

Step 1: Initialize the parameter w1, w2, …wn and β.

Step 2: Use sigmoid function to convert the result into rank [0,1].

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Step 3: Evaluate the weight vector w.

$$J = -\sum_{i=1}^{N} (y_i * log(\hat{y}_i) + (1-y_i) * log(1-\hat{y}_i))$$

Step 4: Calculate the cost function.

$$J(w) = \frac{1}{n}\sum_{i=1}^{n} L^{(i)}(w)$$

Step 5: Using gradient descent to minimize. If the value of cost function is small enough, break the iteration, else repeat the process.

Step 6: End the algorithm.

2. Decision Tree

DT is a supervised learning technique used for both classification and regression problems where each path is a set of decisions leading to a class. A sequence of questions is asked by taking an instance from the training set. The non-terminal node such as root and internal nodes has decision attributes. In this research, we have used ID3 DT for the study of crop prediction.

ID3 Algorithm:

The criteria for measuring Information Gain are the Gini index and Entropy. Entropy and Gini Index are the metrics that measure the impurity of the nodes. A node is considered impure if it has multiple classes else, it is considered pure.

Formula for calculating Entropy

$$E(S) = \sum_{i-=1}^{c} -p_i \log_2 p_i$$

ISSN PRINT 2319 1775 Online 2320 787
Research Paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

Formula for calculating Gini index

$$\text{Gini}=1-\sum_{i=1}^{C}(p_i)^2$$

3. Random forest algorithm

The RF method consists of multiple decision tree classifiers to enhance the model's performance. It is a supervised learning algorithm. DT are created at random using the instances from the training set. Each of the decision trees gives out predictions as their outcome. The final prediction for the model is decided by majority voting. One of the reasons for its popularity as a machine learning approach is that it can handle the issue of overfitting, and accuracy can be increased by using more trees.

Algorithm:

Step 1: K instances are chosen at random from the given training dataset.
Step 2: Decision trees are created for the chosen instances.
Step 3: The N is selected for the number of estimators to be created.
Step 4: Step 1 & Step 2 is repeated.
Step 5: For the new instance, the predictions of each estimator is determined, the category with the highest vote is assigned.

4. XGBoost

XGBoost regression is abbreviation form for extreme gradient boost regression. Loss function is present in objective function which shows the difference between actual values and predicted values whereas regularisation term is used for showing how far is actual value away from predicted value. To limit mistakes from previous models, boosting weights the newly added models based on different optimization methods.

Algorithm:

Step 1: Initialize pseudo-residuals to be equal for each data point
Step 2: In the loop i, new train model will be added to fit existing pseudo-residuals values
Calculate confidence score ***ci*** of the model just trained

$$\text{Update main model:} \quad W = W + c_i * w_i$$

Finally, calculate the value pseudo-residuals to make the label for the next model

Step 3: Then repeat with **(**i + 1) loop.

5. K-Nearest Neighbours

Nearest Neighbours is a type of instance-based learning. For this technique, the model tries to find a number (k) of training examples closest in distance to a new point, and predict the output for this new point from these closest neighbours. k can be a user-defined number (k-nearest neighbours) or vary based on the local density of

ISSN PRINT 2319 1775 Online 2320 787
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

points (radius-based neighbours). The distance metric used to measure the closeness is mostly the Euclidean distance.

4 Implementation using Machine Learning Techniques

This project is implemented in python. Dataset contain 75% training data and 25% data.

Data Description

For this study, we obtained dataset in a .csv format Kaggle[13]. There are 8 characteristics (columns) and 2200 rows in the dataset. The attributes of dataset are given in Table1.

Table 1: Dataset attributes

Feature	**Description**
N	ratio of Nitrogen content in soil
P	ratio of Phosphorous content in soil
K	ratio of Potassium content in soil
temperature	temperature in degree Celsius
humidity	relative humidity in %
Ph	PH value of the soil
Rainfall	rainfall in mm

Evaluation Parameters

Classification report is a metric used for evaluating the performance of a classification algorithm's predictions. The evaluation parameters used in this research are given below:

1. **Precision** refers to a classifier's ability to identify the number of positive predictions which are relatively correct. It is calculated as the ratio of true positives to the sum of true and false positives for each class.

   Precision = $\frac{TP}{TP+FP}$ where TP is true positive & FP is false positive.

2. **Recall** is the capability of a classifier to discover all positive cases from the confusion matrix. It is calculated as the ratio of true positives to the sum of true positives and false negatives for each class.

   Recall = $\frac{TP}{TP+FN}$ where TP is true positive & FN is false negative.

3. **F1 score** is a weighted harmonic mean of precision and recall, with 0.0 being the worst and 1.0 being the best. Since precision and recall are used in the computation, F1 scores are often lower than accuracy measurements.

   F1 score = $\frac{2*PR}{P+R}$ Where P-Precision; R-Recall

ISSN PRINT 2319 1775 Online 2320 787
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

4. Accuracy

Accuracy is defined as Number of correct predictions Total number of predictions. For binary classification, accuracy can also be calculated in terms of positives and negatives

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$ where TP is true positive, TN true negative, & FP is false positive, FN is false negative

5. Results and Discussions

The detailed results of LR, DT, RF, XBOOST and KNN algorithm are shown in Table 2, 3,4, 5 and 6 respectively. Heat map for correlation matrix of LR, DT, RF, XBOOST and KNN algorithm are shown in Fig 2(a),(b),(c),(d) and Fig 2(e) respectively.

Table 2: Results of LR algorithm for one epoch

Crop	Precision	Recall	F1-score	Support
Apple	0.97	0.78	0.86	36
Banana	0.97	1.00	0.98	28
Black gram	0.93	0.74	0.82	34
Chickpea	0.97	0.97	0.97	29
Coconut	1.00	0.94	0.97	33
Coffee	0.97	0.97	0.97	36
Cotton	0.90	0.97	0.93	36
grapes	1.00	1.00	1.00	23
Jute	0.79	0.88	0.83	25
Kidney beans	0.97	1.00	0.99	33
Lentil	0.90	0.93	0.92	41
Maze	0.96	0.87	0.91	30
Mango	1.00	0.97	0.98	29
Moth beans	0.76	1.00	0.86	28
Mung bean	0.96	1.00	0.98	25
Muskmelon	1.00	0.96	0.98	28
Orange	0.97	1.00	0.99	33
Papaya	1.00	0.96	0.98	26
Pigeon peas	1.00	0.96	0.98	23
Pomegranate	0.90	0.96	0.93	28
Rice	0.91	0.91	0.91	34
Watermelon	0.96	1.00	0.98	22
Accuracy			0.94	660
Macro avg	0.94	0.94	0.94	660
Weighted avg	0.94	0.94	0.94	660

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 787
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

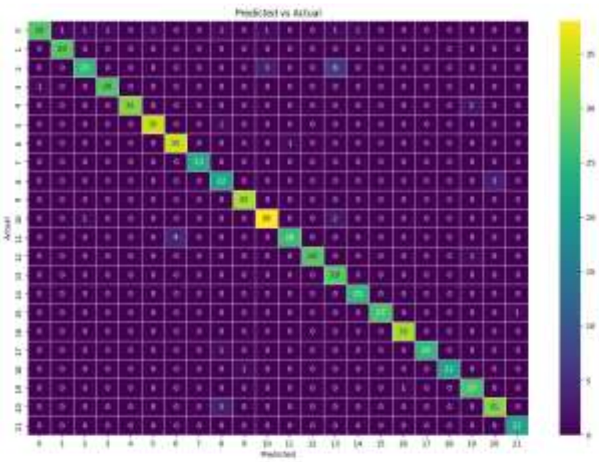


Fig. 2 (a): Correlation Matrix of LR algorithm

Table 3: Results of DT algorithm for one epoch

Crop	Precision	Recall	F1-score	Support
Apple	0.97	0.78	0.86	36
Banana	0.97	1.00	0.98	28
Black gram	0.93	0.74	0.82	34
Chickpea	0.97	0.97	0.97	29
Coconut	1.00	0.94	0.97	33
Coffee	0.97	0.97	0.97	36
Cotton	0.90	0.97	0.93	36
Grapes	1.00	1.00	1.00	23
Jute	0.79	0.88	0.83	25
Kidney beans	0.97	1.00	0.99	33
Lentil	0.90	0.93	0.92	41
Maze	0.96	0.87	0.91	30
Mango	1.00	0.97	0.98	29
Moth beans	0.76	1.00	0.86	28
Mung bean	0.96	1.00	0.98	25
Muskmelon	1.00	0.96	0.98	28
Orange	0.97	1.00	0.99	33
Papaya	1.00	0.96	0.98	26
Pigeon peas	1.00	0.96	0.98	23
Pomegranate	0.90	0.96	0.93	28
Rice	0.91	0.91	0.91	34
Watermelon	0.96	1.00	0.98	22
Accuracy			0.94	660
Macro avg	0.94	0.94	0.94	660
Weighted avg	0.94	0.94	0.94	660

ISSN PRINT 2319 1775 Online 2320 787
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

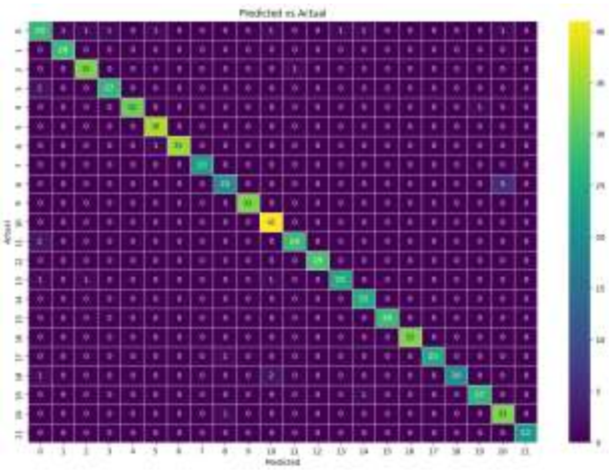


Fig. 2 (b): Correlation Matrix of DT algorithm

Table 4: Results of RF algorithm for one epoch

Crop	Precision	Recall	F1-score	Support
Apple	0.97	0.78	0.86	36
Banana	0.97	1.00	0.98	28
Black gram	0.93	0.74	0.82	34
Chickpea	0.97	0.97	0.97	29
Coconut	1.00	0.94	0.97	33
Coffee	0.97	0.97	0.97	36
Cotton	0.90	0.97	0.93	36
grapes	1.00	1.00	1.00	23
Jute	0.79	0.88	0.83	25
Kidney beans	0.97	1.00	0.99	33
Lentil	0.90	0.93	0.92	41
Maze	0.96	0.87	0.91	30
Mango	1.00	0.97	0.98	29
Moth beans	0.76	1.00	0.86	28
Mung bean	0.96	1.00	0.98	25
Muskmelon	1.00	0.96	0.98	28
Orange	0.97	1.00	0.99	33
Papaya	1.00	0.96	0.98	26
Pigeon peas	1.00	0.96	0.98	23
Pomegranate	0.90	0.96	0.93	28
Rice	0.91	0.91	0.91	34
Watermelon	0.96	1.00	0.98	22
Accuracy			0.94	660
Macro avg	0.94	0.94	0.94	660
Weighted avg	0.94	0.94	0.94	660

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 787
Research Paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

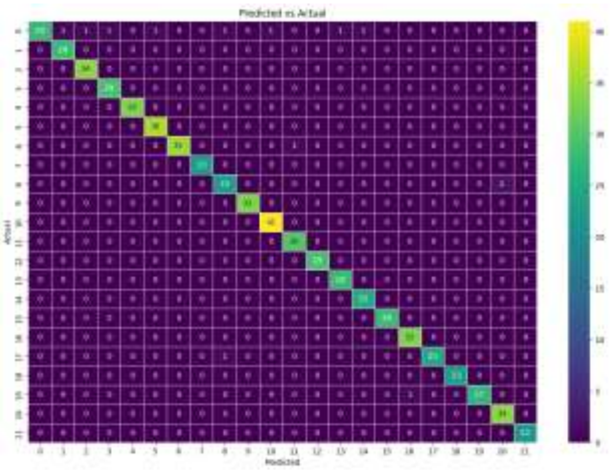


Fig. 2Correlation Matrix of RF algorithm

Table 5: Results of XBOOST algorithm for one epoch

Crop	Precision	Recall	F1-score	Support
Apple	0.97	0.78	0.86	36
Banana	0.97	1.00	0.98	28
Black gram	0.93	0.74	0.82	34
Chickpea	0.97	0.97	0.97	29
Coconut	1.00	0.94	0.97	33
Coffee	0.97	0.97	0.97	36
Cotton	0.90	0.97	0.93	36
grapes	1.00	1.00	1.00	23
Jute	0.79	0.88	0.83	25
Kidney beans	0.97	1.00	0.99	33
Lentil	0.90	0.93	0.92	41
Maze	0.96	0.87	0.91	30
Mango	1.00	0.97	0.98	29
Moth beans	0.76	1.00	0.86	28
Mung bean	0.96	1.00	0.98	25
Muskmelon	1.00	0.96	0.98	28
Orange	0.97	1.00	0.99	33
Papaya	1.00	0.96	0.98	26
Pigeon peas	1.00	0.96	0.98	23
Pomegranate	0.90	0.96	0.93	28
Rice	0.91	0.91	0.91	34
Watermelon	0.96	1.00	0.98	22
Accuracy			0.94	660
Macro avg	0.94	0.94	0.94	660
Weighted avg	0.94	0.94	0.94	660

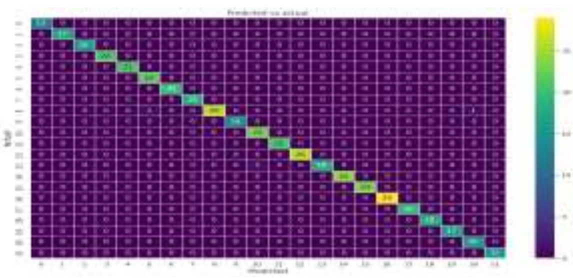


Fig. 2Correlation Matrix of XBOOST algorithm

Table 6: Results of KNN algorithm for one epoch

Crop	Precision	Recall	F1-score	Support
Apple	0.97	0.78	0.86	36
Banana	0.97	1.00	0.98	28
Black gram	0.93	0.74	0.82	34
Chickpea	0.97	0.97	0.97	29
Coconut	1.00	0.94	0.97	33
Coffee	0.97	0.97	0.97	36
Cotton	0.90	0.97	0.93	36
grapes	1.00	1.00	1.00	23
Jute	0.79	0.88	0.83	25
Kidney beans	0.97	1.00	0.99	33
Lentil	0.90	0.93	0.92	41
Maze	0.96	0.87	0.91	30
Mango	1.00	0.97	0.98	29
Moth beans	0.76	1.00	0.86	28
Mung bean	0.96	1.00	0.98	25
Muskmelon	1.00	0.96	0.98	28
Orange	0.97	1.00	0.99	33
Papaya	1.00	0.96	0.98	26
Pigeon peas	1.00	0.96	0.98	23
Pomegranate	0.90	0.96	0.93	28
Rice	0.91	0.91	0.91	34
Watermelon	0.96	1.00	0.98	22
Accuracy			0.94	660
Macro avg	0.94	0.94	0.94	660
Weighted avg	0.94	0.94	0.94	660

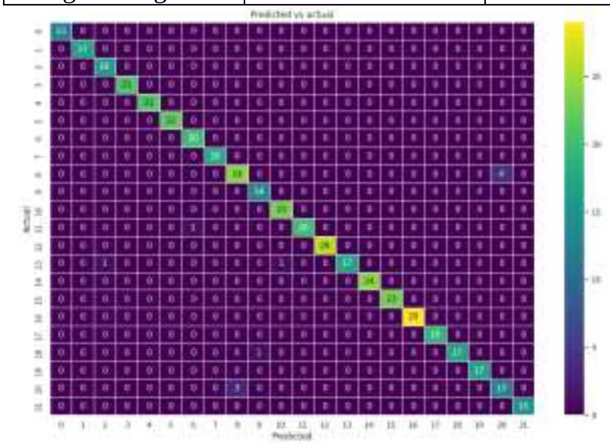


Fig. 2Correlation Matrix of KNN algorithm

We created and tested five machine learning models: for each model, we trained (fitted) the model to our training data (x_train and y_train), then tested the model on our test data (x_test), and finally, we assessed the model performance by comparing the model predictions with the true values in y_test. We used accuracy score to evaluate model performance.

ISSN PRINT 2319 1775 Online 2320 787
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

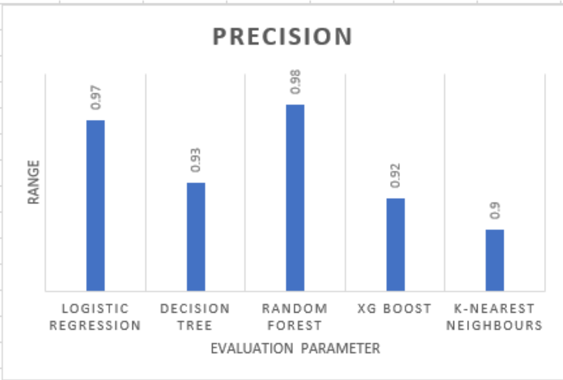
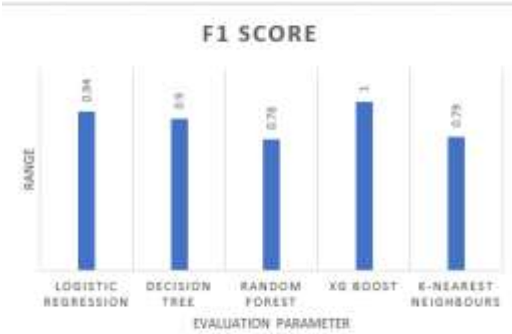


Fig 2 (f) Precision of ML algorithms

Fig 2 (g) Recall of ML algorithm



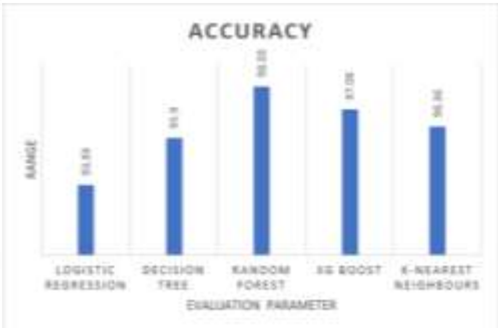


Fig (h) F1 score ML algorithms

Fig 2(i) Accuracy of ML algorithms

Using the results, we got in the previous section, we present a table that shows the Accuracy Score for each model when applied to the test set x_test.

Model	Precision	Recall	F1-Score	Training Score (%)	Accuracy Score(%)
Logistic Regression	0.97	0.96	0.9398	93.98	93.93
Decision Tree	0.93	0.92	0.9249	92.49	93.21
Random Forest	0.98	0.9	0.9749	97.49	98.03
XG Boost	0.92	0.92	0.92	92.1	92.87
K-Nearest Neighbours	0.9	0.98	0.9382	93.82	95.31

After calculating and analysing all the accuracy scores we found Random Forest model gives the highest value. Logistic Regression, Decision Tree, Random forest, XG Boost, KNN gave values as 93.93, 93.21, 98.03, 92.87, 95.31 respectively.

5. Conclusion

In this research paper we have used different machine learning algorithms to study crop yield prediction. In this paper, we built several regression models to predict the best-suited crop for the particular land that can help farmers to grow crops more efficiently. In order to figure out which model performed the best, we compared and assessed each one. We also evaluated the significance of the features as ranked by several models. In this paper, we followed the data science process starting with getting the data, then cleaning and pre-processing the data, followed by exploring the data and building models, then evaluating the results and communicating them with visualizations.

References

1. Dahikar S and Rode S V 2014 Agricultural crop yield prediction using artificial neural network approach International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering vol 2 Issue 1 pp 683-6.
2. Suresh A, Ganesh P and Ramalatha M 2018 Prediction of major crop yields of Tamilnadu using K-means and Modified KNN 2018 3rd International Conference on Communication and Electronics Systems (ICCES) pp 88-93 doi:10.1109/CESYS.2018.8723956.
3. Medar R, Rajpurohit V S and Shweta S 2019 Crop yield prediction using machine learning techniques IEEE 5th International Conference for Convergence in Technology (I2CT) pp 1-5 doi:10.1109/I2CT45611.2019.9033611.
4. Nishant P S, Venkat P S, Avinash B L and Jabber B 2020 Crop yield prediction based on Indian agriculture using machine learning 2020 International Conference for Emerging Technology (INCET) pp 1-4 doi: 10.1109/INCET49848.2020.9154036.
5. Kalimuthu M, Vaishnavi P and Kishore M 2020 Crop prediction using machine learning 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) pp 926-32 doi:10.1109/ICSSIT48917.2020.9214190.
6. Geetha V, Punitha A, Abarna M, Akshaya M, Illakiya S and Janani A P 2020 An effective crop prediction using random forest algorithm 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) pp 1-5 doi: 10.1109/ICSCAN49426.2020.9262311.
7. Pande S M, Ramesh P K, Anmol A, Aishwaraya B R, Rohilla K and Shaurya K 2021 Crop recommender system using machine learning approach 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) pp 1066-71 doi:10.1109/ICCMC51019.2021.9418351.
8. Sellam V, and Poovammal E 2016 Prediction of crop yield using regression analysis Indian Journal of Science and Technology vol 9(38) pp 1-5.
9. Bharath S, Yeshwanth S, Yashas B L and Vidyaranya R Javalagi 2020 Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop Yield Prediction International Journal of Engineering Research & Technology (IJERT) NCETESFT – 2020 vol 8 Issue 14.
10. Mahendra N, Vishwakarma D, Nischitha K, Ashwini and Manjuraju M. R 2020 Crop prediction using machine learning approaches, International Journal of Engineering Research & Technology (IJERT) vol 9 Issue 8 (August 2020).
11. Gulati P and Jha S K 2020 Efficient crop yield prediction in India using machine learning techniques International Journal of Engineering Research & Technology (IJERT) ENCADEMS-2020 vol 8 Issue 10.
12. Gupta A, Nagda D, Nikhare P, Sandbhor A, 2021, Smart crop prediction using IoT and machine learning International Journal of Engineering Research & Technology (IJERT) NTASU – 2020 vol 9 Issue 3.
13. S. Batool, J. Rashid, M. W. Nisar, J. Kim, T. Mahmood, and A.Hussain, ―A random forest students' performance prediction (RFSPP) model based on students' demographic features,‖ presented at 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), Pakistan, 2021.
14. F. Balahadia, M. Villarica, G. M. Catedrilla, and J. Asor, ―Building model for crime pattern analysis through machine learning using predictive analytics, ‖ International Journal of Science, Technology, Engineering and Mathematics, vol. 2, no. 1, pp. 61–73, 2022.
15. S. Bum, I. B. Iorliam, E. O. Okube, and A. Iorliam, ―Prediction of student's academic performance using linear regression, ‖ Nigerian Annals of Pure and Applied Science, vol. 2, pp. 259–264, 2019.
16. E. Y. Obsie and S. A. Adem, ―Prediction of student academic performance using neural network, linear regression and support vector regression: A case study, ‖ International Journal of Computer
17. Applications, vol. 180, no. 40, pp. 39–47, 2018.
18. M. A. Umar, ―Student academic performance prediction using artificial neural networks: A case study, ‖ International Journal of Computer Applications, vol. 178, no. 48, pp. 24–29, 2019.
19. P. M. Arsad, N. Buniyamin, and J.-L. Ab Manan, ―A neural network students' performance prediction model (NNSPPM),‖ presented at 2013 IEEE International Conference on Smart Instrumentation,
20. Measurement and Applications (ICSIMA), Malaysia, 2013.
21. Dataset: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset.