https://doi.org/ 10.33472/AFJBS.6.10.2024.5185-5194

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

CROP YIELD ESTIMATION BASED ON PREDICTIVE ANALYTIC APPROACH

1. *Dr.T. Sundaravadivel, Assistant Professor/ Programmer, Department of Information Technology, Annamalai University*
2. *Dr.M. Arasakumar, Assistant Professor, Department of Information Technology, Annamalai University*

Article History

Volume 6,Issue 10, 2024

Received:30 Apr 2024

Accepted : 28 May 2024

doi: 10.33472/AFJBS.6.10.2024.5185-5194

ABSTRACT

Crop yield prediction is a difficult task in agriculture since it depends on a variety of factors, including weather patterns, pest and disease activity, soil quality, and management techniques. Although conventional techniques for predicting agricultural production were originally thought to be adequate, the environment of today is much more complicated, and seasonal changes are no longer enough to explain variances in accuracy. Even the most seasoned specialists are challenged by the dizzying number of variables that climate change has introduced to the equation. The fine balance between sowing and reaping feels more fragile due to fluctuating weather patterns, severe occurrences, changed growth seasons, and changing pest pressures. In spite of this uncertainty, however, state-of-the-art instruments and quick thinking have led to the development of the novel yield sampling technique as a stopgap measure to raise crop yield forecast accuracy. In order to provide more accurate yield calculations, this paper suggests a machine learning based method that combines rainfall, fertilizer, and historical data. The raw data captured from different sources are preprocessed using feature scaling and feature selection techniques before used for training the machine learning algorithms. Predictive analytics based estimation has been proved as a robust technique in the field of data mining.

Keywords: crop yield prediction, machine learning algorithms, predictive analytics, feature scaling, feature selection.

1. INTRODUCTION

India's economy is based mostly on agriculture because it is essential to the existence of both humans and animals in the country [1]. With an anticipated 1.8 billion people on the planet in 2009 and 4.9 billion expected by 2030, there will be a dramatic rise in the demand for agricultural goods. The human population's need for agricultural goods will increase in the future, necessitating the effective development of farmlands and an increase in crop yields. Meanwhile, crops were often ruined by unfavorable weather conditions as a result of global warming [2]. A single crop failure brought on by unpredictable weather patterns, floods, poor soil fertility, a shortage of groundwater, or other similar causes destroys the crops, which in turn has an impact on the farmers. In some countries, farmers are advised by society to raise agricultural yields unique to the region and surrounding conditions [3]. Crop yield has to be estimated and monitored since the population is growing at a much faster rate [4]. Consequently, in order to pick

crops more effectively with regard to seasonal variance, a suitable approach must be created that takes into account the influencing factors [5].

The method of precisely projecting a crop's potential yield for a given season and location is known as crop yield prediction. Knowledge of the area, the crops cultivated there, and scientific competence are necessary for accurate yield prediction. Numerous data sets are used to estimate it, including those related to weather patterns, seed genetics, soil characteristics, farm management, and statistics. The yield estimations are computed using a multi-tier approach that has been tested in the field. Although yield prediction is extremely important for striking the correct balance between profitability and sustainability for the agriculture industry, its inherent uncertainties resulting from a multitude of contributing variables frequently cause errors. Regardless of technique, the precision of field observations determines the accuracy of yield forecasting. The foundation of yield prediction models is data. The more accurate and thorough the data, the more reliable and strong the predictions will be. Manual random sampling is the procedure used in traditional yield sampling techniques. This approach uses visual evaluations and historical yield data, which are then extrapolated to provide production estimates. It fails to take into consideration the field's heterogeneity, which leads to imprecise sample gathering and distorted forecasts. Fortunately, a new age of more reliable and efficient prediction models is being ushered in by the development of technology, which includes machine learning techniques, remote sensing, and satellite monitoring.

Engineers and scientists create special machine-learning algorithms to forecast agricultural output ranges. In order to predict yield at significant crop stages throughout the cycle, Hybrid and Advanced Yield Estimation Model combines data from fields, satellite imaging, remote sensors, and meteorological data. These models have been fine-tuned for a variety of crops, geographies, and types. Hybrid methodology that includes weather, crop science, evapotranspiration, crop knowledge graphs, and statistics from agricultural operations was adopted in different literature. Re-estimating crop production entails making changes to the original estimate of crop yield in light of new information, guaranteeing a more precise harvest forecast. Re-estimation is needed in cases when early-season projections are affected by unanticipated occurrences that affect actual yield. New data inputs gathered during the season are used to assist account for these variations and enhance yield forecasting accuracy.

Utilizing several well-established models to boost crop production yield is the main goal of crop yield estimation, which aims to raise agricultural crop output. Due to its effectiveness in a variety of fields, including forecasting, defect detection, pattern recognition, etc., machine learning (ML) is being employed globally these days. When there is a loss due to adverse conditions, the ML algorithms also assist in increasing the rates of agricultural yield output. ML algorithms are used in crop selection to minimize crop yield production losses regardless of distracting environmental factors. Large volumes of data from IoT sensors and other sources may be analyzed by machine learning algorithms. It is a fast expanding discipline that might revolutionize agricultural growth and yield prediction and analysis. Without explicit programming, computer systems may learn from experience and become more intelligent through the use of machine learning algorithms, which employ statistical and mathematical models and algorithms to evaluate data and provide predictions [6]. Machine learning algorithms may be taught on extensive farm data, including weather patterns, soil characteristics, crop growth phases, and pest and disease outbreaks, to be used in agriculture, particularly in the cultivating area. Machine learning algorithms are able to anticipate crop growth, production, and quality with great accuracy by analyzing the gathered data [7].

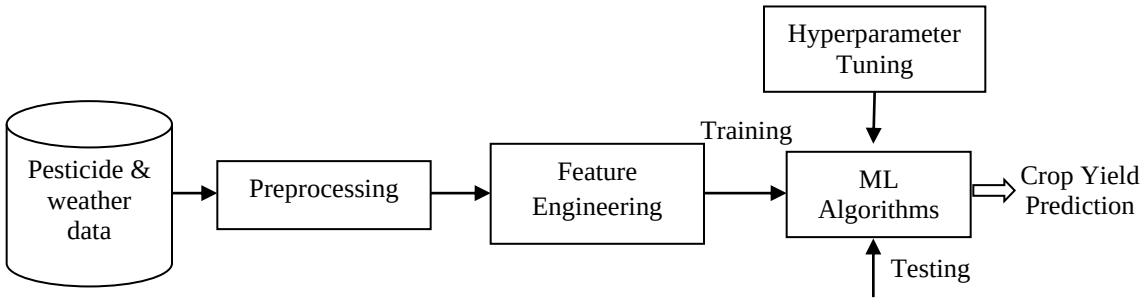

Fig. 1 Flow of process in ML based Crop Yield Prediction

This paper proposes a ML based approach for estimating the crop yield based on the weather and fertilizer data. The correlation between the different variables was assessed and pre-processing techniques were adopted. Feature scaling was performed to rescale the values of the variable to a similar scale. The name of the country was encoded

using one-hot encoding and initially the model training was started with all variables and using backward elimination algorithm, the variables that are not relevant to the prediction was eliminated. Various regression algorithms including linear regression, gradient boosting regression, decision tree regression, and SGD regression were analyzed to evaluate their effectiveness in predicting the crop yield.

2. LITERATURE SURVEY

Crop yield prediction techniques make use of machine learning techniques and algorithms to enhance crop quality and optimize farmer profits. The economy as a whole benefits from the agriculture sector's higher quality. The literature has covered this topic in great depth [8, 9]. A overview of machine learning techniques for predicting the production of palm oil is covered in [8]. The advantages, drawbacks, and limits of the proposed methodologies were the main topics of the authors' comparative examination of the relevant work. Furthermore, the authors offered a new architecture based on machine learning techniques to forecast palm oil production based on the analysis and assessment of the previous research. The authors of [9] examined soil quality with an emphasis on crop prediction and yield, as soil characteristics have a significant impact on agricultural productivity. The scientists looked at a variety of soil parameters, including temperature, rainfall, moisture content, pH value, humidity, and NPK (nitrogen, phosphorus, and potassium) levels. Three machine learning algorithms are the subject of comparative analyses: Random Forest, Logistic Regression, and Naïve Bayes. Additionally, the authors performed an accuracy comparison of these algorithms.

A soil prediction dataset was gathered from Tamil Nadu Agricultural University in India in [10]. The authors asserted that their comparative examination of several machine learning algorithms, including Naïve Bayes and Instance-Based Learner (IBK) algorithms, assisted farmers in making informed decisions about crop selection and production. 22 crop varieties were suggested by the authors in [11], who also suggested a three-step approach that included data preprocessing and feature extraction, classification, and performance evaluation. Based on the comparison, the authors concluded that Naïve Bayes, with its accuracy of 99.45%, is the best classifier for this case. The authors in [12] provided an extensive review on crop monitoring methods regarding estimation of yield and disease identification using deep learning models, since crop diseases are the primary cause of yield losses, particularly in developing countries, and crop monitoring is thought to be the primary focus in the smart-farming process. The authors asserted that agricultural monitoring strategies employing deep learning techniques are more potent and accurate than certain older methods used in poor nations, based on the comparison and findings. The authors of [13] put out a brand-new method for auxiliary data on practical applications in the agriculture industry that is based on support vector machines. In comparison to the apps that were already in use, the authors reported that they achieved an accuracy of 91%. This suggested approach may be used by farmers to increase agricultural yields, and it can also be used by various governmental sectors to boost crop production. The authors did not, however, offer any suggestions for fertilizer systems that might enhance crop management.

The authors of [14] have proposed the use of machine learning and Multilayer Perception in conjunction with rainfall data to assist farmers in making informed decisions about harvests even prior to planting. Moreover, the recommended approach concentrates on the best procedure for crop sale and storage. The authors asserted that the suggested strategy would help farmers achieve better agricultural output outcomes based on the findings they had supplied. The primary goal of the study in [15] was to forecast crop productivity loss using linear regression techniques using information from the statistics from the prior year. Samples were used for the assessment process, and real-world data was used to generate the models. In this assessment, the authors used both the Naïve Bayes and Decision Tree methods. They asserted that the suggested approach increases output and optimizes farmer profit. Farmers can utilize a web-based program created by the authors in [16] to anticipate crop yields. This program gives farmers a list of different crops that have been planted in the past so they may forecast and discover which crop will be most beneficial to grow in the future. In addition, the application may give farmers climatic data and statistics to assist them in making the best choices for market pricing and demand.

The primary goal of the author's study in [17] was to assess how well various publications that attempted to forecast fungal diseases on crops performed between 2016 and 2020. The various machine learning methods used in the literature were assessed by the writers. The authors found that SVM, decision tree variants, and Naïve Bayes had the greatest performance among all machine learning models and were frequently used to get the best outcomes for crop disease prediction annually based on the comparison data that were supplied. In [18], the authors suggested a machine learning-based system that uses the Convolutional Neural Network (CNN) approach to predict agricultural illnesses in plants early on. A village's dataset is used for testing and training. A database containing many diseases is gathered, and the classifier is trained to compare the accuracy and choose the disease with the highest accuracy. Farmers can forecast plant illnesses and choose the best crop to grow with the aid of the model that is offered.

The authors of [19] gathered statistics that were made available online after realizing the difficulties Indian farmers had in predicting agricultural production. The dataset was grouped using the K-Means Clustering technique to enable data analysis and research, and the optimal crop to plant was identified using the Naïve Bayes algorithm. The study and findings presented indicate that farmers can benefit from the suggested approach not only in terms of early crop production prediction but also in terms of choosing the most suitable crop for their plans. It is evident from the literature study above that this topic has been covered in the literature from a variety of angles. Machine learning algorithms have been used in the majority of this field's work to assist farmers in crop prediction, increasing yield and boosting overall productivity. Many study investigations, however, failed to take into account an actual issue for which categorization and performance evaluation should be done. Even though machine learning (ML) in agriculture has a bright future, like every ML problem, the quality of the input data has a major impact on the quality of the output. Crop yield forecast accuracy is largely dependent on the availability and quality of data. A wide range of variables, including weather, soil, past yield, and satellite imaging, are needed for crop forecast. In model construction, ensuring data quality via effective feature selection, preprocessing, and collecting is essential.

3. DATASET DESCRIPTION

Supply chain management, food security, and agricultural productivity may all be enhanced by precise yield forecasts. This study made use of a synthetic dataset [20] that was designed to mimic actual agricultural data, replete with outliers included to test and strengthen the predictive models. The 20,000 records in the dataset include the following characteristics: yield, rainfall, irrigation, sunny days, fertilizer volume, seed kind, and soil quality. The rating of soil quality falls between 50 and 100. A high-yield variety is indicated by a binary indication, 1 for seed variety. The amount of fertilizer applied in kilograms per hectare is indicated by the symbol "fertilizer (kg/hectare)." Sunny Days is a measure of how many sunny days there are in a growing season. Rainfall: The total amount of millimeters of precipitation received throughout the growing season. A growing season's worth of irrigations is determined by the irrigation schedule. Yield (kg/ha): The goal variable for prediction, representing the agricultural yield in kilograms per hectare. The Fertilizer (kg/hectare) and Yield (kg/hectare) columns include purposefully placed outliers to mimic real-world anomalies and serve as a testbed for robust regression algorithms and outlier identification. Fig. 2 presents the plot of relation between various parameters available in the dataset. There is a strong association between yield and seed variety, and irrigation. Also there is a moderate relation between fertilizer volume and yield.

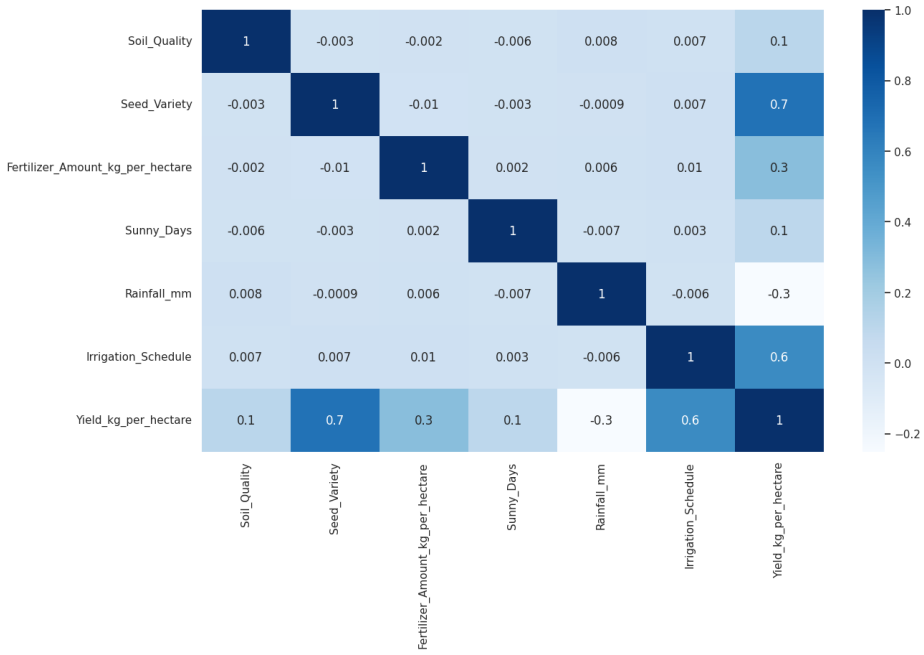

Fig. 2 Plot of Pearson Correlation Coefficient

4. METHODOLOGY

The problem of predicting the crop yield has been formulated as a regression analysis. A statistical method for establishing the link between one or more independent (predictor) variables and a single dependent (criterion) variable is regression analysis. A linear combination of the predictors produces a projected value for the criteria in the analysis. A machine learning approach called gradient boosting is applied to regression and classification issues. Based on the idea that several incompetent students can combine to create a more precise forecaster, it functions. Gradient boosting is a technique that builds weaker prediction models one after the other, each of which attempts to anticipate the mistake left by the preceding model. As a result, the algorithm has a tendency to quickly overfit. By improving on some of Adaboost's characteristics, gradient boosting produced a more powerful and effective algorithm. The gradient boosting algorithm differs somewhat from the Adaboost technique. It utilizes a loss function to minimize loss and converge onto a final output value, as opposed to taking the weighted average of each individual output as the final output. Gradient boosting gets its name from the fact that gradient descent is used for the loss function optimization. Moreover, gradient boosting substitute's decision stumps with shorter, simpler decision trees. As the model gets better, the weak learners are fitted so that every new learner fits into the residuals of the preceding stage. By combining the outcomes of each stage, the final model produces a powerful learner. The residuals are found using a loss function.

Hyperparametes are essential components of learning algorithms that influence a model's accuracy and performance. Two essential hyperparameters for gradient boosting decision trees are learning rate and *n_estimators*. The learning rate, represented by α, indicates the model's learning speed. Every additional tree changes the model as a whole. Learning rate determines how much of a change occurs. The model learns more slowly the lower the learning rate. A slower learning rate has the benefit of making the model more reliable and effective. Slow-learning models outperform fast-learning models in statistical learning.

The model requires more time to train, which leads us to the second important hyperparameter. The number of trees in the model is denoted by *n_estimator*. To train the model, we need additional trees if the learning rate is poor. Nonetheless, we must use extreme caution when deciding how many trees to plant. Using an excessive number of

trees increases the danger of overfitting. Overfitting from the inclusion of too many trees is a problem with gradient boosting decision trees, but not random forests. Adding too many trees won't lead to overfitting in random forests. After a while, the model's accuracy doesn't become much better, but overfitting isn't an issue either. However, while choosing gradient boosting decision trees, we must exercise caution in the amount of trees we choose, as an excessive number of weak learners in the model may induce overfitting of the data. As such, gradient boosting decision trees need extremely cautious hyperparameter tweaking.

Gradient boosting methods sometimes perform poorly on test datasets because to overfitting. Thus, in order to enhance the gradient boosting algorithm's performance, the following strategies were used.

- **Stochastic Gradient Boosting**

In stochastic gradient boosting, the training dataset is subsampled, and individual learners are trained on random samples produced by this subsampling. This lowers the correlation between the outcomes from different students, and merging low-correlated findings yields a superior final product.

- **Shrinkage**

Each tree's forecasts are cumulatively added one after the other. Alternatively, the algorithm's learning can be slowed down by weighting each tree's contribution to this total. This weighting is referred to as a learning rate or shrinkage. Your gradient boosting model's performance can be significantly enhanced by using a low learning rate. Generally speaking, the best outcomes come from a learning rate between 0.1 and 0.3. The training time can be greatly increased by a low learning rate since it will take the model more iterations to converge on a final loss value.

- **L1 and L2 Regularization**

Leaf weight values were subject to L1 and L2 regularization penalties in order to impede learning and avoid overfitting. Regularization is frequently used in gradient tree boosting implementations by restricting the minimum number of observations in the terminal nodes of the trees.

4.1 Outlier Elimination

The Z score, which is often referred to as the standard score, is employed in machine learning model training to scale the features inside a dataset. It was applied to identify anomalies. The dataset's variables and characteristics have varying ranges of values. Positive and negative Z-scores are also possible. The likelihood of a particular data point being an outlier increases with its distance from zero.

$$Z_{score} = (X - \bar{X})/\sigma$$

where $\bar{\bar{X}}$ denotes the mean, and σ denotes the standard deviation.

4.2 Feature Scaling

The technique of normalizing a dataset's feature range is known as feature scaling. Features in real-world datasets frequently differ in terms of size, range, and unit of measurement. Consequently, feature scaling must be done in order for machine learning models to comprehend these characteristics on the same scale. In the absence of feature scaling, a machine learning algorithm would typically treat all values, regardless of unit, as higher and weigh larger values as such. Scaling ensures that every feature has a comparable range and is on a comparable scale. This is important since a lot of machine learning approaches depend on the size of the features. Larger scale elements could control the learning process and significantly affect the results. Scaling the features can improve the algorithm's performance by accelerating the algorithm's convergence to the desired result. It is possible to prevent numerical instability by limiting the size of scale differences between features. Ensuring that every characteristic receives equal consideration during the learning process is made possible by scaling features. Without scaling, learning might be dominated by features at a larger scale, leading to biased results. Scaling ensures that each feature contributes equally to model predictions while also eliminating this bias.

4.3 Feature Selection

In machine learning, the backward elimination strategy is used to determine which subset of a given collection of characteristics is the best. It eliminates characteristics that have the least capacity for forecasting or are not descriptive of the target parameter repeatedly. A quick and efficient method for choosing a subset of parameters for a linear regression model is backward elimination. It can be automated and is simple to apply. Fitting a multivariate linear regression model with all of the independent variables is the first step in the backward elimination procedure. A new model fits once the parameter with the greatest p-value is eliminated from the model. Until every variable in the model has a p-value below a predetermined threshold, usually 0.05, this process is repeated. A more methodical technique called backward elimination begins with a whole feature set and proceeds to eliminate each feature one at a time until the model's performance achieves its maximum. Although this approach is more computationally efficient, it might not also identify the best collection of attributes.

5. EXPERIMENTS AND RESULTS

Gradient boosting necessitates preprocessing measures like feature scaling since it is susceptible to outliers and noise in the data. GBM computes the loss function's negative gradient in relation to its predictions at each iteration. The new tree is guided to better match these residuals by this gradient, which symbolizes the direction of learning. The effect of each individual tree varies depending on the learning rate, which also determines the step size along this gradient. To avoid overfitting, the objective function has a regularization term along with the loss function. The objective function aids the model in striking a compromise between accurately fitting the training set and extrapolating results to unobserved data. The Gradient Boosting Regression's hyperparameters are tuned via grid search.

The technique methodically looks for the set of hyperparameters that produces the best accuracy by establishing a parameter grid with different values for parameters like the maximum depth of trees, learning rate, and number of estimators. Through cross-validation, the Grid Search iteratively trains and assesses the model with various combinations of hyperparameters. Ultimately, the optimal model and its optimum parameters are found, and the optimized model is then used to make predictions on the test set. The procedure seeks to effectively investigate a broad variety of hyperparameter values in the hopes of maybe identifying ideal configurations that optimize model accuracy.

The more trees there are in a model, the better the data can be learned. Finding the optimal value for a parameter requires parameter searching since adding a large number of trees can significantly slow down the training process. The more trees there are in a model, the better the data can be learned. Finding the optimal value for a parameter requires parameter searching since adding a large number of trees may significantly slow down the training process. One sample at a time or all of the samples at each node may be taken into consideration in order to determine the minimal number of samples needed to divide an internal node. The tree becomes more limited when this value is increased since it must take into account more samples at each node. xgboost will fit training data far better than linear regression; however, this also makes it more difficult to understand and prone to overfitting. Depending on needs and facts, either option could turn out to be superior.

Table 5.1 . Performance Analysis of Regression algorithms

Model	R^2	MAE	MSE	RMSE	MAX	MAPE
Linear Regression	0.94	41	2551	51	180	6.45
Decision Tree	0.85	62	6035	78	301	9.77
Stochastic Gradient	0.94	41	2551	51	181	6.45

Descent						
Gradient Boosting	0.93	42	2702	52	174	6.67
k-Nearest Neighbor	0.91	48	3674	61	302	7.75
Random Forest Regression	0.93	44	2983	55	206	7.01

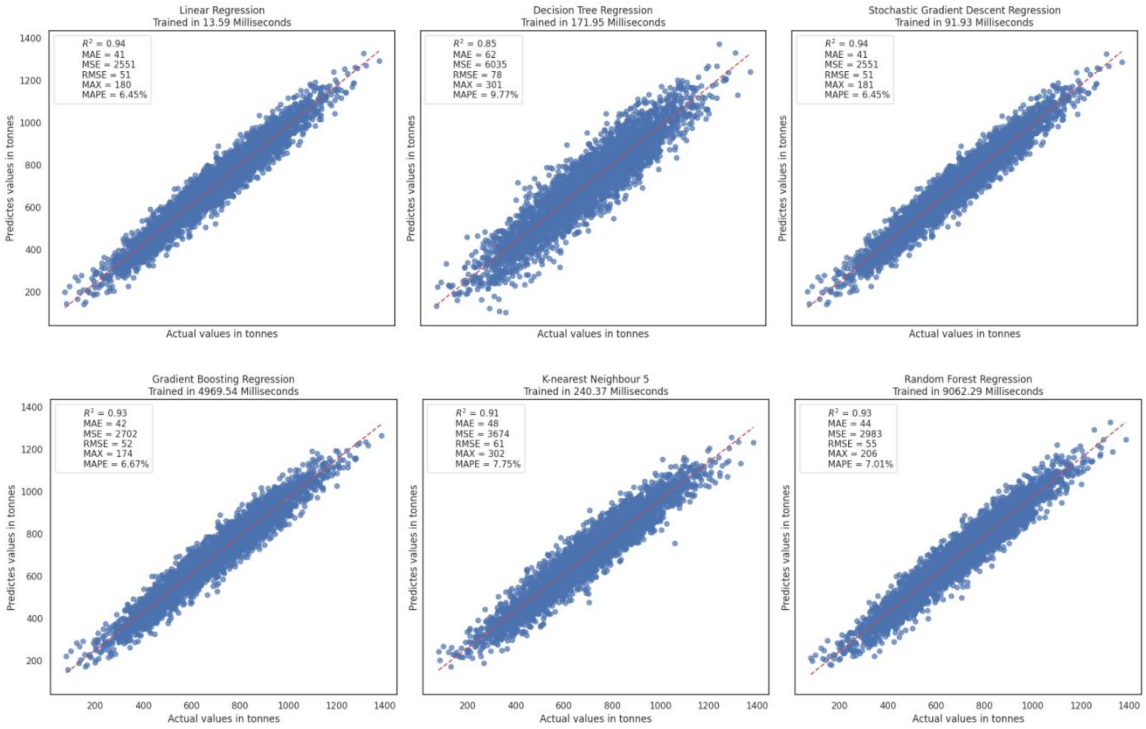

Fig. 3 Comparison of Predicted vs Actual Yield

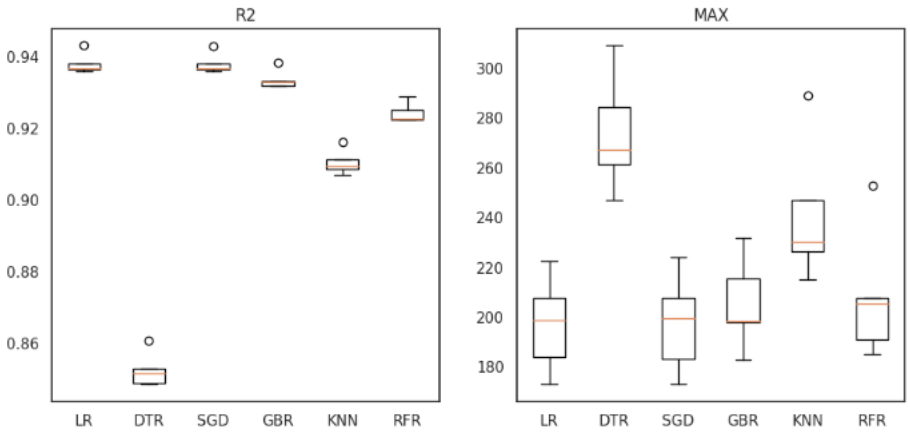

Fig. 4 Comparison of R^2 and Maximum Error

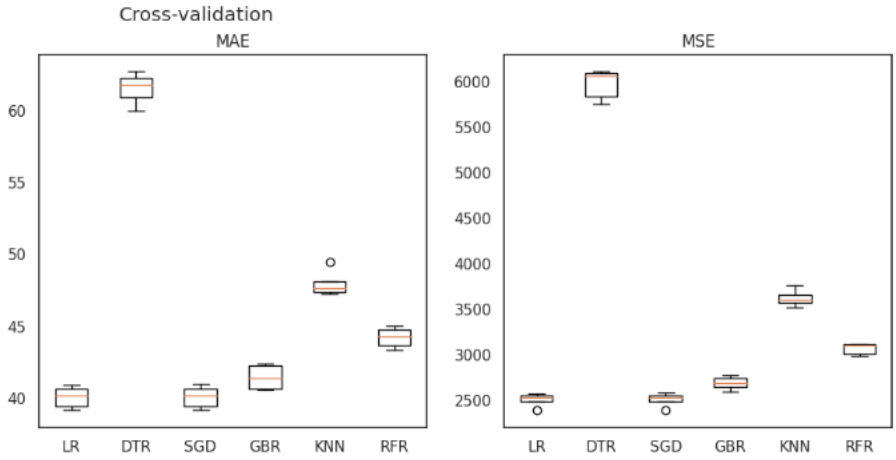

Fig. 5 Comparison of Absolute and Squared Error

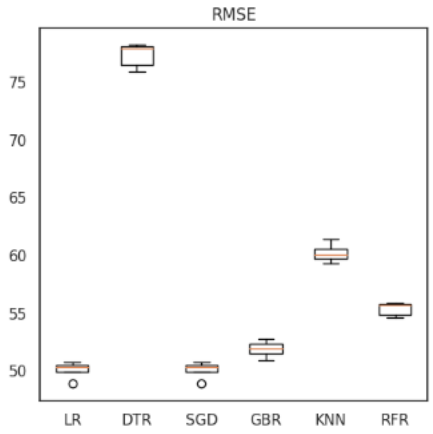

Fig. 6 Comparison of Root Mean Square Error

CONCLUSION

This study examined a range of characteristics that have a strong correlation with crop output and investigated many regression methods to see how successful they were. There is a direct correlation between soil nutrients and crop production, or growth. It is well known that low soil nutrient levels can hinder crop productivity. Crop quality and output can be directly impacted by soil nutrition. Among the many ML categories, agricultural yield prediction benefits from supervised learning approaches. This is because of their strong prediction powers and adaptability to various attribute kinds. One strategy in machine learning to increase prediction accuracy is the backward elimination procedure. Features that are not statistically significant or not predictive of the target variable are eliminated using this strategy. One effective method to increase prediction accuracy and create stronger machine learning models is backward elimination. Crop overfertilization exacerbates bug problems.

REFERENCE

[1] R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. L. Priya, “Prediction of crop yield using machine learning,” Int. Res. J. Eng. Technolgy, vol. 5, 2018.
[2] F. H. Tseng, H. H.Cho, and H. T. Wu, “Applying big data for intelligent agriculture-based crop selection analysis,” IEEE Access, vol. 7, pp. 116965-116974, 2019.

[3] A. Suresh, N. Manjunathan, P. Rajesh, and E. Thangadurai, "Crop Yield Prediction Using Linear Support Vector Machine," European Journal of Molecular & Clinical Medicine, vol. 7, no. 6, pp. 2189- 2195, 2020.
[4] M. Alagurajan, and C. Vijayakumaran, "ML Methods for Crop Yield Prediction and Estimation: An Exploration," International Journal of Engineering and Advanced Technology, vol. 9 no. 3, 2020
[5] P. Kumari, S. Rathore, A. Kalamkar, and T. Kambale, "Predicition of Crop Yeild Using SVM Approch with the Facility of E-MART System" Easychair 2020.
[6] Shahrin, F.; Zahin, L.; Rahman, R.; Hossain, A.J.; Kaf, A.H.; Abdul Malek Azad, A.K.M. Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning. In Proceedings of the 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 17–19 December 2020; pp. 21–24.
[7] Tawseef, A.S.; Tabasum, R.; Faisal, R.L. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Comput. Electron. Agric.* **2022**, *198*, 107119.
[8] Babber, J.; Malik, P.; Mittal, V.; Purohit, K.C. Analyzing Supervised Learning Algorithms for Crop Prediction and Soil Quality. In Proceedings of the 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 29–31 March 2022; pp. 969–973.
[9] Ishak, M.; Rahaman, M.S.; Mahmud, T. FarmEasy: An Intelligent Platform to Empower Crops Prediction and Crops Marketing. In Proceedings of the 13th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 17–20 April 2021; pp. 224–229.
[10] Chandraprabha, M.; Dhanaraj, R.K. Soil Based Prediction for Crop Yield using Predictive Analytics. In Proceedings of the 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 17–18 December 2021; pp. 265–270.
[11] Ray, R.K.; Das, S.K.; Chakravarty, S. Smart Crop Recommender System-A Machine Learning Approach. In Proceedings of the 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 27–28 January 2022; pp. 494–499.
[12] Priyadharshini, K.; Prabavathi, R.; Devi, V.B.; Subha, P.; Saranya, S.M.; Kiruthika, K. An Enhanced Approach for Crop Yield Prediction System Using Linear Support Vector Machine Model. In Proceedings of the International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 10–11 March 2022; pp. 1–5.
[13] Malathy, S.; Vanitha, C.N.; Kotteswari, S.; Mohankkanth, E. Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques. In Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 9–11 May 2022; pp. 437–442.
[14] Chowdary, V.T.; Robinson Joel, M.; Ebenezer, V.; Edwin, B.; Thanka, R.; Jeyaraj, A. A Novel Approach for Effective Crop Production Using Machine Learning. In Proceedings of the International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 16–18 March 2022; pp. 1143–1147.
[15] Yamparla, R.; Shaik, H.S.; Guntaka, N.; Marri, P.; Nallamothu, S. Crop Yield Prediction using Random Forest Algorithm. In Proceedings of the 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 22–24 June 2022; pp. 1538–1543.
[16] Apeksha, R.G.; Swati, S.S. A brief study on the prediction of crop disease using machine learning approaches. In Proceedings of the 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, 18–19 June 2021; pp. 1–6.
[17] Kumar, R.; Shukla, N.; Princee. Plant Disease Detection and Crop Recommendation Using CNN and Machine Learning. In Proceedings of the International Mobile and Embedded Technology Conference (MECON), Noida, India, 10–11 March 2022; pp. 168–172.
[18] Bhosale, S.V.; Thombare, R.A.; Dhemey, P.G.; Chaudhari, A.N. Crop Yield Prediction Using Data Analytics and Hybrid Approach. In Proceedings of the Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018.
[19] Alwis, S.D.; Hou, Z.; Zhang, Y.; Na, M.H.; Ofoghi, B.; Sajjanhar, A. A survey on smart farming data, applications and techniques. *Comput. Ind.* **2022**, *138*, 103624.
[20] https://www.kaggle.com/datasets/blueloki/synthetic-agricultural-yield-prediction-dataset/code