# CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHMS

Shashank C
Department of Master of Computer Applicatios
Dayananda Sagar College of Engineering
Bangalore
shashankachar72@gmail.com

Prof. Alamma B H
Department of Master of Computer Applicatios
Dayananda Sagar College of Engineering
Bangalore
alamma-mcavtu@dayanandasagar.edu

*****

**ABSTRACT:**

The Agriculture of predicting crop yield is essential to agricultural planning and decision-making. Farmers can reduce risks related to weather, pests, and diseases, increase production efficiency, and optimize resource allocation with the aid of accurate crop output predictions. Recently, a number of industries, including agriculture, have seen promising results from the use of machine learning algorithms. In this project, methods from machine learning will be used to build a cropyield forecast model. Suggested model incorporates historical data as input features, including weather patterns, soil characteristics, fertilization techniques, and crop management techniques. Artificial neural networks, random forest models, and support vectoralgorithms are just a few of the machine learning methods that are researched in search of the best- performing model. The dataset gathered for the model's training and testing comes from many agricultural locations, proving its generalizability. Utilizing suitableevaluation The model is evaluated using metrics like median total error, the root mean square erroneous and degree of determination..

Cross-validation methods are also used to verify the robustness of the model and avoid overfitting. To evaluate the generated model's superiority in terms of accuracy and predictive capacity, its performance iscontrasted with that of conventional statistical methods. The findings of this study have important ramifications providing key insights for the agricultural community so they may enhance making choices, minimize risk, and better allocate resources. In order to increase output and minimize ecological impact, landowners cantake proactive actions including altering the application of fertilizer, watering, and insect management strategies. This abstract concludes by highlighting the potential of machine learning techniques for predicting agricultural yield. The suggested method shows its efficacy in making reliable forecastsenabling producers and crop professionals to maintain sustainable and efficient techniques while making informed decisions to increase total productivity in agriculture.

**Keywords**: Feature Engineering, Ensemb1e Techniques, Hyperparameter Tuning, Resource Optimization, Crop Yie1d Prediction, Agricu1ture,Machine Learning.

**INTRODUCTION:**

Predicting crop yields is an important role for farming since it aids with the decision-making process for landowners and regulators on cultivation and security offood.

Because of their capacity to manage intricate and huge datasets, [1]. In recent years, crop yield forecasting has seen a significant increase in the use of machine learning techniques. The various factors that influence yields from agriculture, such as conditions, soil properties, historic yielding data, and plant methods of management, can be examined by machine learning algorithms. The computer programs can find connections and trends that may be used to make precise forecasts by learning from these data. Data collection is the initial stage in crop yield prediction using machine learning. On-field detectors, official the form of databases, and data from remote sensing, and other sources are used to compile historicalinformation on crop yields, weather patterns, and soil characteristics. The dataset's dependability and quality are then ensured by preprocessing the data to remove noise, missing values, and outliers.

The training set and the testing set are then created fromthe preprocessed data [2]. To ensure the dependability andquality of the dataset, this data is then preprocessed to remove vibration, values that are missing, and outliers.. For predicting crop yield, a variety of machine learning methods can be utilized, comprising decision tree structures, random forest models, artificial neural networks, and support vector algorithms and linear regression. The unique characteristics of the dataset and the goal of accuracy of predictions determine the strategy to choose.[3]. Each algorithm has strengths and disadvantages. The model predicts the crop production based on the relevant parameters that are entered, suchas weather and soil characteristics. One can assess how accurate the forecasts were by comparing them to the real yields of crops recorded in the testing set.

In a number of research, the application of algorithms based on machine learning for calculating the productivity of crops has produced positive results [4]. By giving farmers precise and timely information about crop yields, By enabling farmers to make more informed choices regarding the cultivation of crops, allocation of resources, and minimising risk, 1t has an opportunity to revolutionise agriculture.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

**LITERATURE SURVEY:**

As far as what is currently known, there is not a thorough literature study that is specifically devoted to the use of deep learning in predicting crop yields. Even while there are several SLR publications and conventional review papers on crop yield prediction, they don't directly address deep learning in this situation [3]. The authors of this work adopt a ground-breaking strategy to thoroughly review the most recent research on the creation of Deep Learning-based algorithms for agricultural production prediction.

In this regard, van Klompenburg et al.'s study on the use of machine learning to estimate agricultural yields discovered that neural networks, particularly CNN, LSTM, and DNN, are the most often used approaches [5]. Depending on the particular study, different features are considered in these predictions. They also emphasised situations in which yield prediction uses image-based object counting and detection in place of conventional tabular data.

Hani et al. contrasted semi-supervised techniques with Deep Learning-based systems for fruit detection and counting in apple orchards in the area of image processing. They showed that the earlier techniques, such as Gaussian Mixture Models, outperformed the later ones, such as U-Net, Faster R-CNN, and CNN, for yield mapping.

The application of Deep Learning techniques for fruit counting and production estimation was also studied by Koirala et al. In order to estimate fruit load, they emphasised the capacity of deep learning algorithms to extract significant features and suggested strategies such CNN detectors, deep regression, and LSTM [6].

Furthermore, Lee et al. tested Deep Learning techniques to create a platform for self-predicting crop yields based on crop illnesses. For the purpose of diagnosing agricultural diseases, they discovered that CNN outperformed R-CNN and YOLO algorithms [7]. They also discovered that the artificial neural network's ReLU activation function provided the CYP module with the highest level of accuracy.

Last but not least, Zhang et al. explored Deep Learning applications in dense agricultural situations, including yield estimate, recognition, classification, detection, and counting [8]. According to their survey findings, Deep Learning generally performs better in intensive agricultural environments than other methods.

The current study represents a significant and systematic attempt to consolidate and expand the understanding of Deep Learning-based approaches for predicting crop output, even if other studies have studied various aspects of Deep Learning in this regard [9].

**METHODOLOGY:**

Collection of Data**:**

Websites, GitHub repositories, and Kaggle datasets were some of the online sources from which the data for this research project was gathered. Relevant factors The information that was acquired includes information on things like climate trends, soil features, chemical use, seedling type, and yields of crops measurements.

```
In [10]: crop.describe()
```

Out[10]:

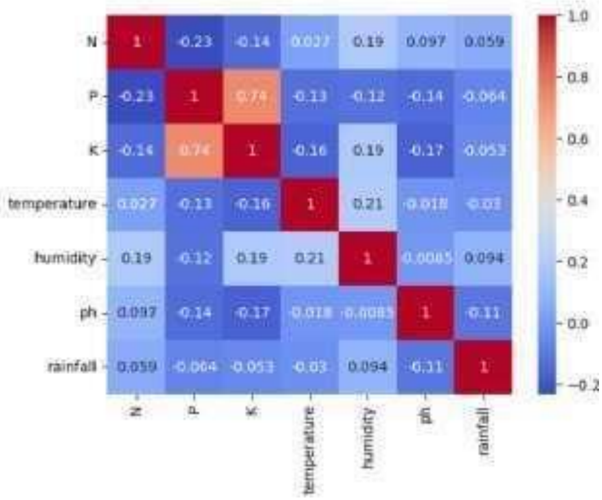
| | N | P | K | temperature | humidity | ph | rainfall |
|---|---|---|---|---|---|---|---|
| **count** | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 | 2200.000000 |
| **mean** | 50.551818 | 53.362727 | 48.149091 | 25.616244 | 71.481779 | 6.469480 | 103.463655 |
| **std** | 36.917334 | 32.985883 | 50.647931 | 5.063749 | 22.263812 | 0.773938 | 54.958389 |
| **min** | 0.000000 | 5.000000 | 5.000000 | 8.825675 | 14.258040 | 3.504752 | 20.211267 |
| **25%** | 21.000000 | 28.000000 | 20.000000 | 22.769375 | 60.261953 | 5.971693 | 64.551686 |
| **50%** | 37.000000 | 51.000000 | 32.000000 | 25.598693 | 80.473146 | 6.425045 | 94.867624 |
| **75%** | 84.250000 | 68.000000 | 49.000000 | 28.561654 | 89.948771 | 6.923643 | 124.267508 |
| **max** | 140.000000 | 145.000000 | 205.000000 | 43.675493 | 99.981876 | 9.935091 | 298.560117 |

```
In [12]: corr
```

Out[12]:

| | N | P | K | temperature | humidity | ph | rainfall |
|---|---|---|---|---|---|---|---|
| **N** | 1.000000 | -0.231460 | -0.140512 | 0.026504 | 0.190688 | 0.096683 | 0.059020 |
| **P** | -0.231460 | 1.000000 | 0.736232 | -0.127541 | -0.118734 | -0.138019 | -0.063839 |
| **K** | -0.140512 | 0.736232 | 1.000000 | -0.160387 | 0.190859 | -0.169503 | -0.053461 |
| **temperature** | 0.026504 | -0.127541 | -0.160387 | 1.000000 | 0.205320 | -0.017795 | -0.030084 |
| **humidity** | 0.190688 | -0.118734 | 0.190859 | 0.205320 | 1.000000 | -0.008483 | 0.094423 |
| **ph** | 0.096683 | -0.138019 | -0.169503 | -0.017795 | -0.008483 | 1.000000 | -0.109069 |
| **rainfall** | 0.059020 | -0.063839 | -0.053461 | -0.030084 | 0.094423 | -0.109069 | 1.000000 |

```
In [13]: sns.heatmap(corr,annot=True,cbar=True, cmap='coolwarm', fmt='.2g')
Out[13]: <Axes: >
```



Creating data:
To guarantee data quality and consistency,preprocessing was applied to the collected data. This required handling both missing numbers and the elimination of noisy data, such as outliers. To prevent redundancy and preserve data integrity, duplicate rowswere also eliminated.

**Extracting Specifications:**
Feature extraction was carried out on the preprocessed data to obtain the pertinent variables for crop yield prediction. In this stage, the available variables were chosen and converted into useful features. For feature extraction, factors including soil details, Vitamins and Minerals cultivation methods, and past weather information were considered.

**Algorithms for Supervised Learning Selection:**
Several supervised learning algorithms were used in this study to forecast crop yield. The chosen techniques included the use of the Gradient Booster Regressor, a Random Forest Regressor, Assistance Vector Regressor (SVR), and the Choice Tree Regressor. These algorithms were chosen because they have a track record of success with regression problems and can handle very big and varied datasets.

**Training and Evaluation:**
The preprocessed and feature-extracted data were used to train the chosen algorithms. The set of data was split into sets for training and validation during the training phase in order to evaluate the efficacy of the models. Forthe evaluation, metrics for performance including the root mean square error (RMSE) and degree of estimation(R2) were used. These measures gave information on theprecision and goodness-of-fit of the algorithms forforecasting the growth of crops.

**The best performing model is choose:**
The model with the greatest efficacy for predicting the yield of crops was found to have the greatest degree of precision and performance. based on the evaluation findings. The results would be further analyzed and interpreted using the chosen model.

**Simulated crop models are compared:**
In order to investigate the impact of doing so, the selected ML models were trained and evaluated using both observed data and generated cropvariables from the model (APSIM).. The comparison between the models integrating simulation crop model variables and thoseusing only observed data permitted for evaluating theaccuracy of the increase in yield prediction.

**Statistical Analysis:**
The significance of the findings and the correlations between the input factors and the anticipated crop output were assessed by statistical analysis. The analysis of the data revealed the main variables affecting crop yield and their relative importance in the ML models.

**Model choice and completion:**
The most appropriate ML model was chosen as the final model for crop yield prediction based on the evaluation findings, including accuracy, interpretability, and computing efficiency. The finished model would be put to use in agriculture management studies and in real1world applications. The tests and analysis in this research project, which aimed to improve crop yield prediction accuracy by integrating ML models with simulation crop models, were carried out using the methodology that was outlined.

## MODELING AND ANALYSIS

We outline the models and resources used in our research effort for agricultural production prediction in this part. We provide details regarding the instructional as well as assessment datasets as a description of the supervised learning techniques used.

1. **Models used:**
For agricultural yield prediction, the following supervised learning algorithms were used.
    a) Gradient Boosting Regressor:
    An ensemble learning approach called the Gradient Boosting Regressor combines a number of weak learners (decision trees) to produce a robust prediction model. It uses gradient descent optimisationto reduce mistakes and raise prediction precision.

b) A forest-based random model of regression Another ensemble learning system that makesuse of decision trees is the Random Forest
Regressor. It builds several Trees of choices producea typical of the projections from decision makers all over training. each tree, increasing accuracy.

b) Support Vector Regressor (SVR):
A machine learning algorithm for regressiontasks is called the Support Vector Regressor. While taking into account a margin around the actual output values, it seeks to reduce the difference between projected and actual output values.

SVR uses a kernel function to translate the input data into a higher-dimensional space and then identifies the hyperplane that best fits the data.

c) Decision Tree Regressor:
A popular algorithm for regression tasks is the Decision Tree Regressor. The input data is divided according on the feature values, and a tree structure is iteratively created to forecast the values of fresh data. To maximise variance reduction, the algorithm chooses the optimum feature and threshold at each internal node.

## RESULTS AND DISCUSSION

Table 1. Summary of Performance Metrics for Different Model

| SN. | Model | Accuracy |
|---|---|---|
| 1 | Gradient Boosting Regressor | 0.98 |
| 2 | Random Forest Regressor | 0.99 |
| 3 | Support Vector Regressor | 0.96 |
| 4 | Decision Tree Regressor | 0.98 |

## CONCLUSION:

In summary, using algorithms that use machine learningto estimate yields from crops has shown a lot of potential for enhancing farming methods and increasing productivity.. These algorithms can effectively predict crop yields based on a variety of variables, including weather conditions, soil quality, andother characteristics.

and techniques for managing crops. With the use of this technology, farmers and agricultural specialists may decide on irrigation, fertilization, and pest management in an informed manner, maximizing resource allocationand reducing waste. Furthermore, it helps them to foresee probable yield losses and take preventative action beforehand, minimizing the impact of unfavorable conditions on crop production. Precision agriculture has been made easier by the application of machine learning algorithms in crop yield prediction.
Farmers may now adjust their operations to particularfields based on the anticipated yields.

## REFERENCES:

[1] P. Priya, U. Muthaiah, M. Balamurugan, predicted yield of the crop using machine learning algorithm. International Journal of Engineering Science ResearchTechnology, 2018.

[2] Aruvansh Nigam, Saksham Garg, Archit AgrawalCrop Yield Prediction using ML Algorithms, 2019.

[3] Dr. Y Jeevan Kumar, has used different Supervised Learning Approach for Crop Production, 2020.

[4] Dr. Jayanarayana Reddy, Dr. M Rudra Kumar, CropYield production using ML Algorithms [IEEE], 2022.

[5] J. Jeong, J. Resop, N. Mueller and team. Random forests for global and regional crop yield prediction, 2020.

[6] Narayanan Balakrishnan and Dr. Govindarajan Muthukumarasamy. Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE), 2018.

[7] V. Pandith, H. Kour, S. Singh, J. Manhas, and V. Sharma, "Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from SoilAnalysis," Journal of Scientific Research, vol. 64, no. 2, 2020.

[8] 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.

[9] F. H. Tseng, H. H. Cho, and H. T. Wu, "Applyingbig data for intelligent agriculture-based crop selection