

Crop Yield Prediction using Machine Learning

Mr. Telise Vinod
(Assistant Professor)
Department of Computer
Science and Engineering
MLR Institute of Technology
Hyderabad, India

Rapolu Suresh Reddy
Department of Computer
Science and Engineering
MLR Institute of Technology
Hyderabad, India

N. Chandrasena Reddy
Department of Computer
Science and Engineering
MLR Institute of Technology
Hyderabad, India

Anugu Athrey Reddy
Department of Computer
Science and Engineering
MLR Institute of Technology
Hyderabad, India

Abstract -- The greater part of India's populace depends on agribusiness as their essential kind of revenue. Horticulture's drawn-out practicality is presently genuinely compromised by climate, temperature, and other natural elements. Since it has a choice help instrument for Crop Yield Prediction (CYP), which incorporates prompting on which yields to develop and what to do all through the harvest's development season, machine learning (ML) assumes a significant part. Various methodologies created to inspect rural yield expectations utilizing computerized reasoning strategies are the focal point of the ongoing review, which is worried with a precise survey that concentrates and integrates the qualities used for CYP. The fundamental impediments of the brain network are lower crop yield gauge effectiveness and lower relative blunder. While evaluating or arranging natural products, managed learning calculations couldn't represent the nonlinear connection between information and result factors. Fully intent on making an exact and effective model for crop characterization —, for example, crop yield gauges in light of climate, crop sickness grouping, crop order in view of the development stage, etc — a ton of examination was proposed for farming turn of events. This study examines the accuracy of several ML algorithms used to estimate agricultural yields and provides a comprehensive analysis.

Keywords: Machine learning algorithms, Crop Yield Prediction (CYP).

1. INTRODUCTION

Since it is vital for the two people and creatures to exist, agriculture is the foundation of the Indian economy. By 2030, the total populace is supposed to reach 4.9 billion, up from the 2009 gauge of 1.8 billion. This will cause a significant increase in demand for agricultural products. Future human population demand for agricultural goods will necessitate effective farmland expansion and crop production increases. However, adverse weather frequently destroyed crops as a result of global warming.

Farmers are influenced when a solitary harvest comes up short for various reasons, including an absence of soil ripeness, environment variety, floods, an absence of groundwater, and others. Based on location and environmental conditions, society in some nations

encourages farmers to increase the agricultural output of particular crops. It is important to gauge and monitor farming results on the grounds that the populace is growing at a lot quicker rate. Accordingly, the affecting elements should be thought about while fostering a proficient technique for further developed crop choice considering occasional vacillations.

The essential goal of crop yield assessment is to increment rural yield, which is achieved through various deep-rooted procedures. Because of its viability in a large number of utilizations, including gauging, deformity identification, design acknowledgment, and many others, AI (ML) is popular around the world. The harvest yield creation rate is likewise helped by the ML calculations when there is a misfortune because of unfriendly conditions. Notwithstanding the presence of diverting environmental elements, the harvest choice technique utilizes AI calculations to diminish crop yield creation misfortunes. The ongoing model utilized SVM to group crop information in light of the surface, shape, and shade of examples on the debilitated surface on the grounds that it has a reasonable feeling of the blemishes.



Fig 1: Example Figure

A current methodology utilized CNN to bring down the general mistake along with crop yield forecast. Likewise, the ongoing methodology, which utilized a more modest dataset size and joined Back Propagation Neural Networks (BPNNs) with a period series model, performed

ineffectively on the grounds that fewer examples were utilized for expectation. In the space of choice strength and precision, ML techniques were used. The info yield connects in crop expectation can be resolved involving ML in various helpful ways. For yield expectation, brilliant water system frameworks, crop infection expectation, crop choice, weather conditions gauging, deciding the base help cost, and other horticultural undertakings, various machine approaches are used. Farmers' input efforts will be reduced while agricultural production will rise as a result of these strategies. Additionally, advances in technology and machines were accurate due to the significant role they played in utilizing large amounts of data.

2. LITERATURE SURVEY

PREDICTING THE YIELD OF THE CROP USING MACHINE LEARNING ALGORITHM

The prosperity of the country's economy is largely due to agriculture. A major threat to agriculture has emerged as a result of environmental changes and climate change. The use of machine learning, or ML, is a crucial method for coming up with practical and efficient solutions to this problem. Crop Yield Prediction is the process of predicting crop production using data from the past, such as weather conditions, soil parameters, and crop yield. In this study, crop production estimates based on existing data are estimated using the Random Forest method. The models were developed using actual Tamil data, and samples were used to evaluate them. Before cultivating on agricultural land, the farmer will use the forecast to predict crop yield. To accurately estimate future agricultural productivity, Random Forest, a sophisticated and widely used supervised machine learning technique, is utilized.

Applications of machine learning techniques in agricultural crop production: a review

This report was written trying to look at research discoveries on the utilization of AI techniques in horticultural harvest creation. Methods and Analysis of the Data: A novel approach to agricultural crop management production is this strategy. The Directorate of Financial aspects and Insights gives precise and opportune crop production projections for critical strategy choices like import-trade, cost, promoting conveyance, etc. In any case, it is vital to remember that these past evaluations are not exactly on the grounds that they require broad elucidating examination in light of different subjective standards. Therefore, an objective and statistically sound crop production forecast are required. A huge amount of data has been created as a result of advances in computers and data storage. Findings: New approaches and methods, such as machine learning, have been developed to combine data knowledge with agricultural production assessment due to the difficulty of extracting complex information from raw data. The reason for this study was to look at these original ways to deal with tracking down huge associations among them and the different

variables in the data set. Application/Improvement: Regression analysis, decision trees, artificial neural networks, information fuzzy networks, and Bayesian belief networks are a few methods. With regards to agribusiness, the Markov chain model, k-means clustering, k-closest neighbor, and support vector machine were completely illustrated.

A Model for Prediction of Crop Yield.

Another field of exploration in agrarian creation examination is information mining. In agriculture, yield expectation is an urgent point. Each farmer would need to know the amount he could hope to gather. Previously, yield expectation depended on the rancher's experience with a specific harvest and field. In view of the information that is now accessible, it is as yet impractical to tackle the basic issue of anticipating yield. Data mining methods are the most ideal choice for this objective. Various Data Mining methods are used and analyzed in agribusiness to estimate crop yield for the next year. A strategy for utilizing verifiable information to foresee rural efficiency is created and tried in this review. Affiliation rule mining on agrarian information is utilized to accomplish this. The objective of this examination is to make a forecast model that can be utilized to foresee the future agricultural result. The region of Tamil Nadu in India is the focal point of this concise examination of farming creation forecast utilizing an information mining procedure in light of affiliation rules. The consequences of the analyses show that the proposed technique precisely predicts rural yield.

Agricultural crop yield prediction using an artificial neural network approach

by considering various climatological factors that affect the climate in various areas of the planet. These weather patterns straightforwardly affect crop creation. The connections between large-scale climatological events and agricultural productivity have been the subject of numerous studies. To increase their efficiency, it has been demonstrated that artificial neural networks are effective modeling and prediction tools. The crop prediction approach uses a variety of soil and atmospheric parameters to predict the right crop. Thought ought to be given to the dirt sort, pH, nitrogen, phosphate, potassium, natural carbon, calcium, magnesium, sulfur, manganese, copper, iron, depth, temperature, precipitation, and dampness. For this reason, we utilized an artificial neural network (ANN).

Predictive ability of machine learning methods for massive crop yield prediction.

For farming arranging reasons, the legitimate yield computation for the different crops remembered for the arranging is a huge issue. The use of machine learning, or ML, is a crucial method for coming up with practical and efficient solutions to this problem. In order to find the most accurate strategy, several ML approaches for yield prediction have been compared. In general, the amount of

crops and techniques examined is insufficient, and agricultural planning information is lacking. This study looks at how well ML and linear regression algorithms can predict crop production in ten agricultural datasets. Support vector relapse, k-closest neighbor, M5-Prime relapse trees, and perceptron multi-facet brain networks were assessed. Four exactness measurements were filled in as approval standards for the models: standardized mean absolute error (MAE), correlation factor (R), root mean square error (RMS), and root relative square error (RRSE). The genuine information from a water system zone in Mexico was utilized to construct the models. Tests from two sequential years were utilized to assess the models. The M5-Prime and k-nearest neighbor approaches have the most noteworthy typical relationship factors (0.41 and 0.42), the least typical RMSE mistakes (5.14 and 4.91), and the most reduced typical RRSE blunders (79.46% and 79.78%). Since it furnishes the most crop yield models with the least mistakes, M5-Prime is a magnificent device for crop yield expectation in rural preparation.

3. METHODOLOGY

Most of the existing models for CYP utilized neural networks, random forests, and KNN regression procedures for ideal expectations. Several machine learning (ML) methods were also utilized.

The current research on agricultural yield prediction using machine learning encounters the following issues:

1. Due to their complexity, ML algorithms were very expensive to create, repair, and maintain.
2. The ML approach used to foresee mustard and wheat crop creation included information and result in information yet bombed measurably to deliver improved results.
3. The relapse model neglected to precisely conjecture complex information, like outrageous worth and nonlinear information, in view of the direct connection between the boundaries.
4. For yield expectation and arrangement, existing K-NN models were utilized, however, their presentation was hampered by KNN's nonlinear and profoundly versatile issues. They were utilized in a territory model that made the information vector more layered and made it harder to order things.
5. Because there was insufficient data to estimate crop output, no appropriate judgment was made during the categorization process.

Disadvantages:

- The current K-NN models were utilized for yield prediction and classification, but their performance was hampered by KNN's nonlinear and highly adaptive issues.

The practical application and quantification of machine learning methods are the primary focus of this research. To create a steady pattern, the strategy introduced here likewise considers the disconnected information from the temperature and precipitation datasets. The crop yield forecast is made by considering the elements in general, rather than the standard act of working out each perspective in turn.

Advantages:

- The investigation discovered that CNN, LSTM, and DNN were the most regularly utilized calculations; nonetheless, CYP actually required improvement.
- The current study demonstrates a number of existing models that perform models for the best crop yield forecast and take into account factors like temperature and weather conditions.
- In conclusion, the experimental investigation demonstrated that crop forecast progress was enhanced when ML and the agricultural domain field were combined.

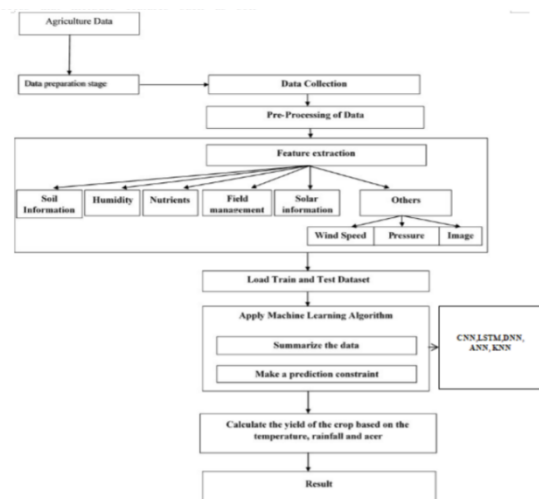


Fig 2: Proposed Architecture

Modules:

Upload Crop Dataset

Arrangement and relapse calculations utilize the crop production information to foresee the yield's name and yield.

Preprocess Dataset

Using datasets from the Indian government, tests were conducted, and the Random Forest Regressor was found to have the highest yield forecast accuracy. In terms of rainfall prediction, the sequential model, which is a Simple Recurrent Neural Network, performs better than the LSTM. A yield projection for a particular locale can be made by considering temperature, precipitation, and other data like the season and the region.

Train Machine Learning

Based on the crops grown in the district, this focuses on yield projections by the district. The crop with the highest yield is used to estimate the yield for each crop district.

Upload Test Data & Predict Yield

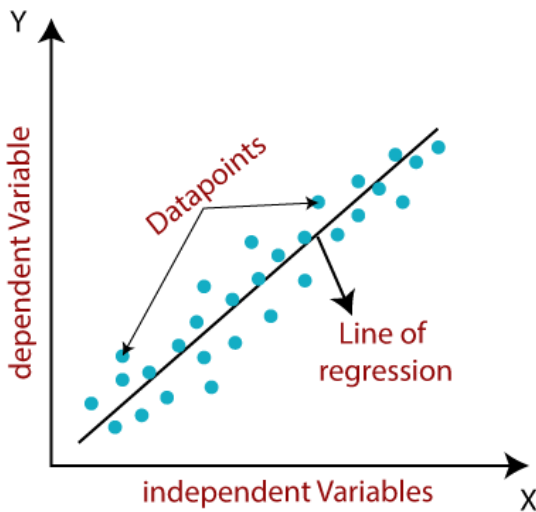
The outcomes show that Random Forest is the best classifier when all boundaries are pooled. In addition to assisting farmers in selecting the most suitable crop for cultivation in the upcoming season, this will help to bridge the technological and agricultural industries.

4. IMPLEMENTATION

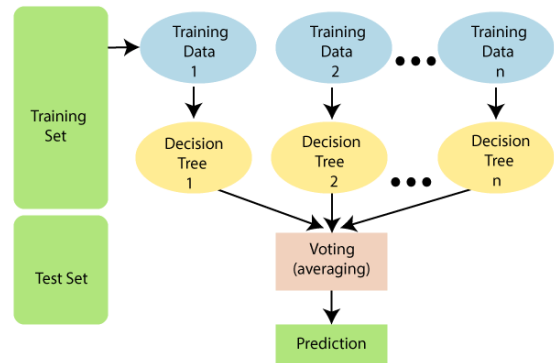
Algorithms:

Linear Regression: An order strategy known as logistic regression utilizes managed figuring out how to gauge the probability of an objective variable. There are just two feasible classes in light of the fact that the objective or ward variable has a dichotomous nature. When applied to our dataset, the Linear Regression technique accomplishes a precision of 87.8%.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:



Random Forest: Crop improvement can be analyzed involving Random Forests comparable to current biophysical change and meteorological circumstances. The Random Forest strategy predicts the information from every subset, fabricates choice trees from various information tests, and afterward votes to offer the framework a superior response. The sacking technique is utilized to prepare the information in Random Forest, and the outcome is more exact. For our data, RF provides an accuracy of 92.81 percent.



Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

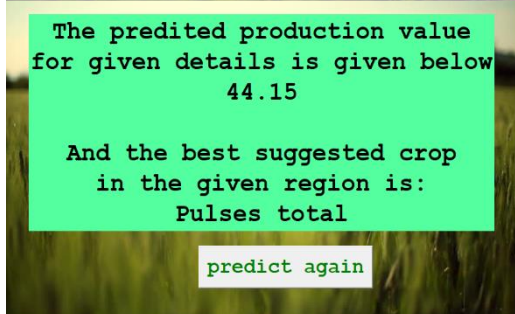
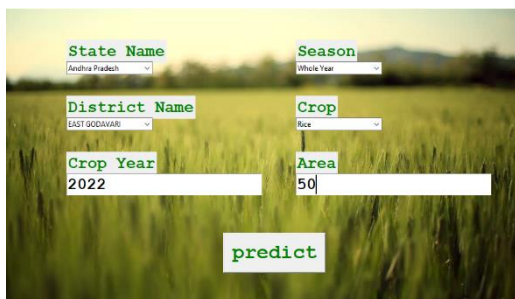
As the name suggests, “Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.” Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output

5. RESULTS

Loading Dataset

| | State_Name | District_Name | Crop_Year | Season | Crop | Area | Production |
|--------|-----------------------------|---------------|-----------|------------|---------------------|----------|------------|
| 0 | Andaman and Nicobar Islands | NICOBARS | 2000 | Kharif | Areca nut | 1254.0 | 2000 |
| 1 | Andaman and Nicobar Islands | NICOBARS | 2000 | Kharif | Other Kharif pulses | 2.0 | 1 |
| 2 | Andaman and Nicobar Islands | NICOBARS | 2000 | Kharif | Rice | 102.0 | 321 |
| 3 | Andaman and Nicobar Islands | NICOBARS | 2000 | Whole Year | Banana | 176.0 | 641 |
| 4 | Andaman and Nicobar Islands | NICOBARS | 2000 | Whole Year | Cashewnut | 720.0 | 165 |
| -- | -- | -- | -- | -- | -- | -- | -- |
| 246086 | West Bengal | PURULIA | 2014 | Summer | Rice | 306.0 | 801 |
| 246087 | West Bengal | PURULIA | 2014 | Summer | Sesamum | 627.0 | 463 |
| 246088 | West Bengal | PURULIA | 2014 | Whole Year | Sugarcane | 324.0 | 16250 |
| 246089 | West Bengal | PURULIA | 2014 | Winter | Rice | 279151.0 | 597899 |
| 246090 | West Bengal | PURULIA | 2014 | Winter | Sesamum | 175.0 | 88 |

Output Screen



6. CONCLUSION

Each study would investigate CYP utilizing ML methods unmistakable from the highlights, and the ongoing review covered various elements that are basically reliant upon information accessibility. The determination of the highlights was for the most part affected by the informational collection's accessibility; in any case, utilizing more elements didn't necessarily create improved results. The choice of the elements depended on land area, size, and harvest attributes. Thus, hands down the least best-performing attributes were explored and used in the examinations. Neural networks, random forests, and KNN relapse strategies were involved by most of the current models for CYP; in any case, an assortment of AI methods was likewise used for ideal expectation. The exploration tracked down that CNN, LSTM, and DNN were the most generally utilized calculations; nonetheless, CYP actually required a further turn of events. The ongoing review exhibits various existing models that take temperature, meteorological circumstances, and crop yield forecasts into account. Finally, the trial examination exhibited that crop forecast progress might be improved by joining ML and the agricultural space field. Nonetheless, further improvement in highlight choice was expected considering the impacts of temperature change on agribusiness. Extra express treatment was expected for the significant potential that ought to be the focal point of resulting examinations, for example, the underlying deferral to line geographical locales. An ML method is utilized to make a nonparametric part of the

model, and afterward, qualities from deterministic crop models are utilized to accomplish wonderful measurable CO₂ treatment. By sticking to the previously mentioned targets, extra exploration would further develop assessments of farming yields. Furthermore, compost should be integrated into crop yield assessment to perform soil figures with the goal that agriculturalists can pursue better choices considering low crop yield assessment. A DL-based model for CYP should be made and created based on the exploration discoveries.

6. REFERENCES

- [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, "Prediction of Crop Yield Using SVM Approach with the Facility of E-MART System" *Easychair* 2020.
- [6] S. D. Kumar, S. Esakkirajan, S. Bama, and B. Keerthiveena, "A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier," *Microprocessors and Microsystems*, vol. 76, pp.103090, 2020
- [7] P. Tiwari, and P. Shukla, "Crop yield prediction by modified convolutional neural network and geographical indexes," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 8, pp. 503-513, 2018.
- [8] P. Sivanandhini, and J. Prakash, "Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 5, pp. 1092-1096, 2020.
- [9] N. Nandhini, and J. G. Shankar, "Prediction of crop growth using machine learning based on seed," *Ictact journal on soft computing*, vol. 11, no. 01, 2020
- [10] A. A. Alif, I. F. Shukanya, and T. N. Afee, "Crop prediction based on geographical and climatic data using machine learning and deep learning", *Doctoral dissertation, BRAC University* 2018.
- [11] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deeplearning-based detector for real-time tomato plant diseases and pests' recognition," *Sensors*, vol. 17, no. 9, pp. 2022, 2017.
- [12] J. Sun, L. Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model," *Sensors*, vol. 19, no. 20, pp. 4363, 2019.
- [13] K. A. Shastry, and H. A. Sanjay, "Hybrid prediction strategy to predict agricultural information," *Applied Soft Computing*, vol. 98, pp. 106811, 2021.
- [14] D. A. Bondre, and S. Mahagaonkar, "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms," *International Journal of Engineering Applied Sciences and Technology*, vol. 4, no. 5, pp. 371-376, 2019.