

## Crop Yield Prediction

Abhishek Sahani<sup>1</sup>, Deepak Kumar Mishra<sup>2</sup>, Sumit Kumar Gupta<sup>3</sup>, Livingston Jeeva<sup>4</sup>

<sup>1,2,3</sup>Students, Department of Computer Science and Engineering, School of Management Sciences, Lucknow, Uttar Pradesh, India.

<sup>4</sup>Assistant Professor, Department of Computer Science and Engineering, School of Management Sciences, Lucknow, Uttar Pradesh, India.

### How to cite this paper:

Abhishek Sahani<sup>1</sup>, Deepak Kumar Mishra<sup>2</sup>,  
Sumit Kumar Gupta<sup>3</sup>, Livingston Jeeva<sup>4</sup>,  
"Crop Yield Prediction", IJIRE-V5I02-186-  
188.

Copyright © 2024 by author(s) and  
5<sup>th</sup> Dimension Research Publication.  
This work is licensed under the Creative  
Commons Attribution International License  
(CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>

**Abstract:** Everyone in India is aware that agriculture is the foundation of our nation. This article forecasts the output of nearly every crop cultivated or farmed in India. Using basic variables like State, District, Season, and Area, the user of this script can forecast the crop's yield for the year they wish to grow or produce it. The Random Forest Algorithm (RFA) of Supervised Learning and the idea of regression to enhance the algorithms are two examples of sophisticated regression approaches used in the study to forecast the yield. While there is still much to be done in many disciplines, technological breakthroughs have already shown improvements in numerous areas. It will show the ratio of the soil's phosphorus, potassium, and nitrogen levels.

**Keyword:** Farming, Machine Learning Techniques, Indian Agriculture, Random Forest Algorithm, Proper Yield

## I. INTRODUCTION

In our research work, we found that all climate factors like soil type, rainfall, solar nutrients (nitrogen, phosphorus, etc.) are used by humans, but the problem is that the data needs to be collected. A third party makes predictions and explains them to the farmer. Farmers and growers who do not understand the science behind these methods require a lot of time and effort. In this study, we use simple variables like farmer's personality, residential area, crops and weather to ensure that it is fair and suitable for farmers. In India, more than hundreds of fruits are grown all over our country. The cut is broken for better understanding and visualization. The data for this study was obtained from the archives of the Government of India. With around 4.5 lakh views, the data includes many parameters including state, region, product, season, year, location and production. Many states and regions in India feature the most recognizable product categories and seasons. To reduce errors and generate accurate predictions for different types of products, we use a forward regression approach, including a Random Forest Algorithm (RFA) based on supervised learning. It is the largest and most important community in India. Because ancient peoples grew food on their own agricultural lands and could meet their own needs. As a result, natural foods are grown and used by many species, including humans, animals and birds. Living things enjoy life and prosperity by eating the green things produced from the soil. The agricultural sector is deteriorating since the appearance of new technologies and methods.

## II. LITERATURE SURVEY

Crop yield prediction studies are conducted using a variety of machine learning techniques and algorithms available to agricultural institutions. This paper reviews some of these studies conducted in research papers with the techniques used and results.

1. Anantara, M.G. et al. (February 2013) In "Survey On Crop Yield Prediction", a predictive model for agricultural datasets called his SVM algorithm for crop yield using regenerative techniques is proposed. They considered parameters such as plant species, soil and soil pH, humidity, and plant sensitivity. Their analysis was mainly concerned with wet rice and sugarcane yields in India. The proposed algorithm is a support vector machine SVM algorithm, which performs well with an accuracy of 85%.
2. Awan, A.M. et al. (April 2006) "Survey on Crop Yield Prediction", focuses on the clustering decision tree method for predicting crop yields and analyzes parameters such as plantation, temperature, and rainfall. We've created a new, intelligent framework that takes you into account. their different regions. Experimental weighting of decision tree methodology with spatial constraints for analysis of palm fields.
3. Chawla, L et al. (August 2019) "Survey on Crop Yield Prediction", Fuzzy logic was used to predict crop yield using a statistical time series model. They considered various parameters
4. Chaudhari, A.N. et al. (August 2018) "Online Survey on Crop Yield Prediction", two algorithms are used to improve crop yield prediction: clustering K-means and bias algorithm. I have used many other algorithms to use it effectively. It was also found that parameters such as pH levels of different regions, rainfall, and soil types were also taken into account, allowing the system to identify which crops are highly suitable for cultivation based on the aforementioned characteristics.

5. Manjula, A. et al. "Agro-meteorological yield prediction using machine learning", create crop selections and predict yields by considering various indicators such as vegetation, temperature, humidity, and various vegetation factors. Then, to improve understanding, climatic and agricultural factors and other disturbances caused by projections are distinguished.
6. Tripathy, A. K. et al. (July 2016) "Survey On Crop Yield Prediction", Some of the authors use support vector machine (SVM) algorithm to predict high crop yield with the same feature. Predicted yield, states that a novel decision-making system is proposed that provides inputs and an interface to achieve better results when applying SVM and neural networks to predict high crop yields. We also considered climatic factors such as temperature, precipitation, and evapotranspiration, primarily (taking into account the range) of the reference plants.
7. Ishwar Chawla et al. Forecasting Temperature and Rainfall Factors Using ARMA, SARIMA, ARMAX Models, Fuzzy Logic Based Crop Yield Forecasting: Agriculture is of great importance in the Indian economy. Predicting crop yields is critical to India's prosperity. Various weather factors such as temperature and precipitation affect crops. Therefore, it is important to consider these aspects when estimating crop yield. It's difficult to predict the weather. In this study, ARMA, SARIMA, and ARMAX (autoregressive integrated moving average) are used as three forecasting methods (ARMA with exogenous variables). The results of the three models are compared and the most accurate model is used to predict temperature and precipitation, which are then used to predict crop yields.

### III. METHODOLOGY

The algorithms used in our project are Random Forest Algorithm (RFA) and Neural Network. The measured results of our experiments show that the RFA algorithm outperforms other algorithms and not only achieves the highest accuracy in weather forecasting, but also helps in achieving high crop yields. Depending on the soil quality, the system estimates crop yield and recommends fertilizers if necessary. The results are processed by two controllers. The address is used as input to the controller in addition to using third-party applications such as APIs for weather and temperature, soil type, local soil nutritional value, local precipitation, soil composition, etc. calculated. It then takes the pH value as input and uses that value to determine the alkalinity of the soil. Also, the proportion of nutrients such as nitrogen (N), phosphorus (P), and potassium (K).

#### Neural Networks

Neural networks, a subset of machine learning, often referred to as artificial neural networks (ANNs) or simulated neural networks (SNNs), are the basis of deep learning approaches. Their organization and terminology are modeled after the human brain and mimic the communication of real neurons. The input layer, one or more hidden layers, and the output layer form the node layers of an artificial neural network (ANN). Each node or artificial neuron is connected to other nodes and has weights and thresholds. A node whose output exceeds a specified threshold becomes active and begins sending information to the top layer of the network.

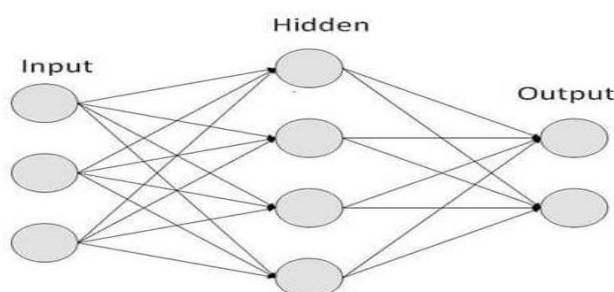


Figure-1 [Neural Network]

#### Random Forest Algorithm

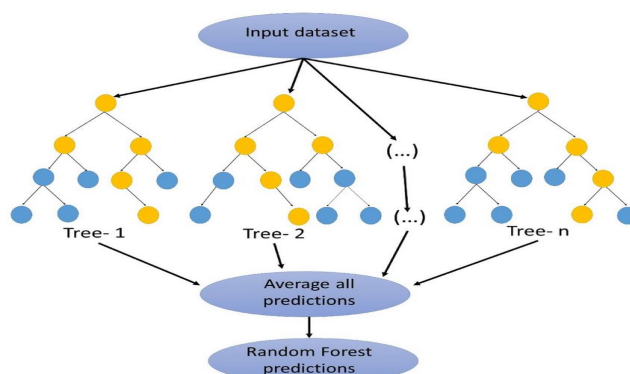


Figure-2 [Random Forest Algorithm]

Random Forest is a technique used in supervised learning methodologies. You can use it to solve ML problems involving both classification and regression. It is based on the concept of ensemble learning, which is a way to combine different classifiers to tackle difficult problems and improve model performance. Random Forest, as the name suggests, is a classifier that takes a series of decision trees on different subsets of a provided dataset and averages them to improve the predictive accuracy of the dataset. Rather than relying on a single decision tree, a random tree uses predictions from each tree and predicts the outcome based on a large proportion of the predicted votes.

### Existing Systems and Disadvantages

The current system does not have a computerized system for identifying recommended culture or behavioral analysis. First, it is only suitable for instance-level techniques that require an instance classifier. As mentioned previously, current common neural network techniques take discrete instances as input and transform them into an embedding space using a deep neural network.

### Advanced Systems

Proposed Method as environmental variables vary from region to region, a machine learning model is used to estimate the ideal plant variety for a selected plot. It uses data from standard datasets to train crop suggestion models and uses machine learning techniques to select the best crops with the highest growth probability for cultivation. Naive bias and random forestry machine technology support determines the optimal crop type. Based on this model, it was determined which types of crops farmers should grow. Aspects to consider include humidity, temperature, soil moisture, and rainfall.

### Advantages

1. Efficient and safe system.
2. The advantage of this approach to farming is that it allows selection of appropriate crops based on factors such as climate and temperature.
3. Machine learning helps avoid assumptions, challenges in using larger sample sizes, and complications and makes predictions based on available data.

## IV. CONCLUSION

This project leverages machine learning and uses random forests, polynomial regression, and decision tree techniques to evaluate performance. Random Forest provides the best yield prediction among the three algorithms of the proposed model compared to other methods. The output showing improvements in the dataset is classified using random forest regression techniques and neural network models. After the analysis, we concluded that the proposed model is more effective than current approaches to determine crop yield. Adopting the above methods will help in improving the agricultural practices of our country. Additionally, it can also be used to reduce farmers' losses and increase crop yields, thereby increasing agricultural capital. Models can be improved through integration. Using stacked regression results in less predictable results than using each model individually.

### Reference

- [1]. Paul, M., S. K. Vishwakarma, and A. Verma (2015, December). *Data mining approach for soil behavior analysis and crop yield prediction*. 2015's Computational Intelligence and Communication Networks International Conference (CICN) (pp. 766-771). IEEE.
- [2]. R. Jaikla, S. Auephanwiriyakul, & A. Jintrawet (2008, May). *Using support vector regression, estimate rice yield*. The Fifth International Conference on Electrical Engineering/Electronics, Computer, Telecommunication, and Information Technology was held in 2008.
- [3]. Ramalatha, M., Suresh, A., and Kumar, P. G. (2018, October). *Tamil Nadu's main agricultural yields are predicted using K-means and Modified KNN*. The third International Conference on Communication and Electronics Systems (ICCES) was held in 2018. (pp. 88-93). IEEE.
- [4]. Ananthara, M. G., Hemavathy, R., & Arunkumar, T. (2013, February). *CRY is a more accurate model for predicting crop yields from agricultural data sets utilizing bee hive clustering*. In 2013 International Conference on Informatics, Mobile Engineering and Pattern Recognition.
- [5]. KM Meheta (January 2015). *XCYPF: A adaptable and scalable framework for crop yield prediction in agriculture*. 9th International Conference on Intelligent Systems and Control (ISCO) of the IEEE was held in 2015.
- [6]. Misra, B., Veenadhari, S., and Singh, C. C. (2014, January). *A method using machine learning for predicting crop yield based on climate factors*. Global Conference on Computer Communication and Informatics in 2014.
- [7]. S. Bang, R. Bishnoi, A. S. Chauhan, A. K. Dixit, & I. Chawla (2019, August). *Crop Yield Prediction Based on Fuzzy Logic Using Predicted Temperature and Rainfall Parameters from ARMA, SAR/MA, and ARMAX Models*. The 12th International Conference on Contemporary Computing (IC3) will be held in 2019.
- [8]. Shah A, Dubey A, Hemnani V, Gala D, and Kalbande D. R. (2018). *Crop Yield Prediction Using Regression Techniques in a Smart Farming System*. International Conference on Wireless Communication Proceedings (pp. 49-56). Singapore's Springer.
- [9]. Sap, M. N. M., and Awan, A. M. (2006, April). *a kernel-based intelligent system for forecasting crop yields*, Conference on knowledge Discovery and Data Mining in Pacific-Asia.
- [10]. Miniappan, N., Kadir, M. K. A., and Ayob, M. Z. (August 2014). *Prediction of wheat yield using artificial neural networks*. The fourth International Conference on Engineering, Technology, and Entrepreneurship (ICE2T) was held in 2014.