

Crop Yield Prediction Using Machine Learning Algorithm Based on Climate Variables

Parutagouda S. Khanagoudar^{1,*}, Sushma B.S.², Chandana C.N.³,
Manikanta N.³, Spoorti Suresh Awatimath³

Abstract

India's economy is based primarily on agriculture, as over 50% of the country's population depends on it for their livelihood. The long-term viability of agriculture is seriously threatened by variations in the weather, climate, and other environmental factors. Because machine learning provides tools for decision assistance in agricultural yield prediction, including guidance on which crops to plant and when to plant them during the growing season, it is essential to the process. The goal of this work is to extract and synthesise features that are used to predict agricultural productivity through a systematic review. Furthermore, a number of techniques have been created to examine crop yield prediction through the use of machine learning algorithms. The decline in crop yield forecast efficiency and the decrease in relative error are neural networks' primary drawbacks. Similarly, picking fruit for sorting and classification proved difficult due to supervised learning methods' inability to capture the non-linear connection between input and output variables. In order to produce precise and effective models for crop classification, several studies have been suggested for the advancement of agriculture. This research includes estimating crop yields based on weather, plant diseases, crop classification based on growth stage, and more. This paper investigates many machine learning (ML) approaches used to agricultural production prediction and offers a thorough evaluation of the approaches' precision.

Keywords: Decision tree regressor, knn regressor, random forest regressor, linear regressor.

INTRODUCTION

Global population support depends heavily on agriculture, and as the need for food supplies rises, it becomes more and more clear that innovative ways to increase agricultural output are required. Crop yield prediction is one of the most important parts of modern agriculture, and the application of machine learning approaches to this task has shown promise. k-Nearest Neighbours (KNN) Regressor, Decision Tree Regressor (DTR), Linear Regressor (LR), and Random Forest Regressor (RFR) are four different ML algorithms that will be utilized in this research project to develop and assess a crop production forecasting system based on historical climate variables [7–10].

The application of machine learning approaches is justified by their capacity to identify intricate patterns in sizable datasets and to offer a data-driven method for comprehending the intricate relationships that crop yields and climate have. Climate variables impact temperature, precipitation, humidity, and other key parameters that affect plant development and agricultural

*Author for Correspondence

Parutagouda S. Khanagoudar
E-mail: khanagoudar@yahoo.com

¹Professor, Department of Data Science and Engineering, Nagarjuna College of Engineering and Technology, Bangalore.

²Assistant Professor, Department of Data Science and Engineering, Nagarjuna College of Engineering and Technology, Bangalore.

³Student, Department of Data Science and Engineering, Nagarjuna College of Engineering and Technology, Bangalore.

Received Date: May 10, 2023

Accepted Date: May 14, 2023

Published Date:

Citation: Parutagouda S. Khanagoudar, Sushma B.S., Chandana C.N., Manikanta N, Spoorti Suresh Awatimath. Crop Yield Prediction Using Machine Learning Algorithm Based on Climate Variables. International Journal of Cheminformatics. 2023; 1(2): 49–52p.

yields. In order to create a model that can precisely estimate crop yields in the future, the project intends to apply machine learning skills to learn from previous climate and crop yield data [5,6].

LITERATURE SURVEY

Comparative Analysis of Crop Yield Prediction Using Machine Learning, Ajay Kumar et al, 2023, Over half of India's population depends on agriculture, which faces threats from environmental factors. Machine learning, particularly Random Forest with 99.31% accuracy, aids Crop Yield Prediction by analyzing soils, climates, and vegetation indices, addressing model uncertainty in crop growth monitoring and yield prediction.

Crop Yield Prediction using Machine Learning Algorithm, D. Reddy et al, 2021, Agriculture, vital to India's economy and supporting over 50% of the population, faces risks from environmental changes. Machine learning, notably Random Forest, enhances Crop Yield Prediction by analyzing diverse data. This study reviews various ML techniques for yield estimation, highlighting their effectiveness and accuracy in improving agricultural practices.

Location Based Prediction of Crops, Analysing the Yield and Market Demand Using R-Forest and MLR Algorithms, [1], India's economy heavily relies on agriculture, which is influenced by unpredictable rain, soil types, and climate factors like temperature and precipitation. Technological advancements can significantly benefit agriculture, enhancing crop production and improving yields for farmers amidst the challenges of global warming and soil nutrient management.

Enhancing crop yield prediction in india: A Comparative Analysis of Machine Learning Models, Kothakonda Chandhar et al, 2023, Farming is crucial for India's economy, but changing weather and climate make it challenging. We proposed a crop yield prediction mechanism using machine learning models like Gradient Boosting, Random Forest, Support Vector, and Decision Tree Regressors. Decision Tree Regressor proved most accurate, aiding farmers in optimizing crop production.

An Efficient Crop Yield Prediction Framework Using Hybrid Machine Learning Model, [2], India's food security relies on accurate crop yield predictions, complicated by varied factors. This study used ML algorithms (Random Forest, XGBoost, Decision Tree, SVM, Linear Regression) on Kaggle data, creating a hybrid model with 98.6% accuracy. The resulting 'Crop Yield Predictor' tool offers actionable insights for practical agricultural use.

Methodology

Start: The process begins with the initiation of a study or project focused on crop yield prediction. This includes defining the goals, scope, and overall structure of the study.

Dataset: We collect a dataset containing relevant information about climate variables and their corresponding crop yields. This dataset serves as the basis for training and testing machine learning models.

Pre-processing and feature engineering: Clean and preprocess your dataset to handle missing values, outliers, or inconsistencies. You can perform feature engineering to extract relevant information or create new features that may improve the predictive power of your model.

Split data: Split the dataset into two subsets. One for training the machine learning model and one for testing its performance. This ensures a fair evaluation of the model's generalization capabilities [3].

Initialize model parameters: Select an appropriate machine learning algorithm to predict crop yields based on climate variables. Initialize model parameters based on the nature of the problem and the characteristics of the data set.

Train and test the model: Using the given climatic variables and accompanying crop yields, train a machine learning model on the training subset. To assess the model's performance on untested data, test it on a batch of data once it has been trained.

Evaluate model: Assess the performance of the model using appropriate measures, such as recall, accuracy, and precision, or other metrics pertinent to the objectives of the project. This stage contributes to our understanding of the model's prediction accuracy for crop yields [4].

Consider the model to be the best one to forecast crop yields based on climate variables if it fits the predetermined criteria and performs satisfactorily. Make predictions using the best-trained model based on upcoming or fresh data. Crop production forecasts are produced by feeding the model with climate factors. Predictions along with the associated data should be saved in a database for future reference or to make the integration of the model into decision support systems easier (Figure 1).

RESULT

Crop yields must be forecast by machine learning algorithms based on data on the climate in order to obtain analytical information about agricultural output. These algorithms forecast agricultural yields in certain areas by utilising historical climate data, which includes temperature, humidity, precipitation, and other pertinent variables. But in this case, dealing with unstructured data can lead to issues with comprehensibility (Table 1).

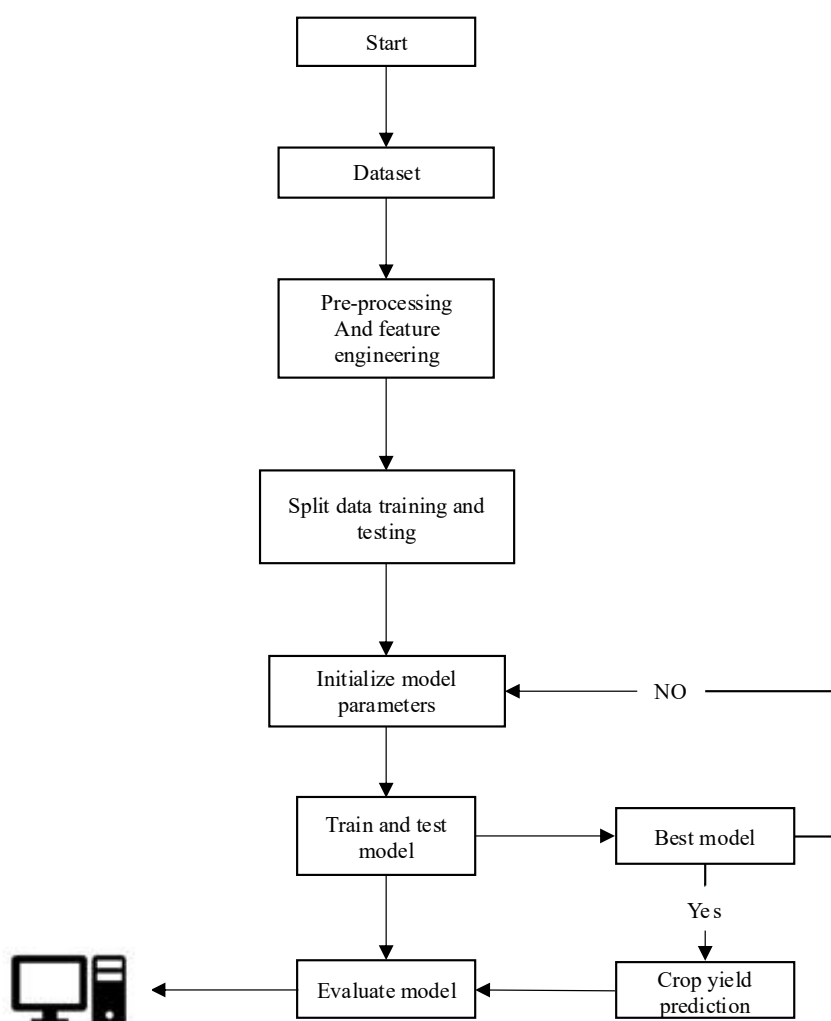


Figure 1. Methodology of crop yield prediction System.

Table 1. Challenges faced through the years.

Year of Publication	Techniques Used	Challenges
2023	Linear model, Random Forest, Neural networks.	Linear regression is a useful tool for analysis, it does have its disadvantages, including its sensitivity to outliers and more.
2023	Correlation, Pearson's correlation coefficient.	The data that gets gathered from this work is typically more applicable to everyday encounters. You do not need to attempt to extrapolate the findings from a laboratory setting into something that works into the routine.
2022	Support vector machine, Linear regression.	SVM algorithm is not suitable for large data sets. SVM does not perform very well when the data set has more noise i.e. target classes are overlapping.
2022	Logistic regression, Decision tree, k-Nearest Neighbour.	KNN is sensitive to noise in the dataset. We need to manually impute missing values and remove outliers.
2020	Neural Networks, Gradient Boosting Tree.	hardware dependency, depending on the type of applications being use.

CONCLUSION

The outcomes of this study have significant implications for precision agriculture, providing farmers and policymakers with valuable tools for informed decision-making. By understanding the intricate relationships between climate variables and crop yields, these models contribute to optimizing resource allocation, improving crop planning, and managing agricultural risks. Furthermore, the adaptability of the models to changing climate patterns positions them as crucial assets for promoting sustainable and resilient agricultural practices. As technology continues to advance, integrating machine learning into agriculture not only enhances productivity but also fosters a more sustainable and efficient approach to global food production. The conclusions of the current research contribute to the ongoing dialogue on leveraging technology to address the challenges of feeding a growing world population in the face of evolving climatic conditions.

REFERENCES

1. M Alagurajan, C Vijayakumaran, 2020. ML Methods for Crop Yield Prediction and Estimation: An Exploration, *International Journal of Engineering and Advanced Technology*. 9(3):2249–8958.
2. Jasmin Bharadiya, Nikolaos Tzenios, Manjunath Reddy, 2023. Forecasting of Crop Yield using Remote Sensing Data, Agrarian Factors and Machine Learning Approaches, *Article in Journal of Engineering Research and Reports*, 24(12):29–44.
3. Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. jabber, 2020. Crop Yield Prediction based on Indian Agriculture using Machine Learning, *International Conference for Emerging Technology (INCET)*, pp.1-4, doi: 10.1109/INCET49848.2020.9154036.
4. Ersin Elbasi, Chamseddine Zaki, Ahmet E. Topcu, Wiem Abdelbaki, 2023. Crop Prediction Model Using Machine Learning Algorithms, *Article in Advances in Technology Applied in Agricultural Engineering, Appl.Sci.*2023, 13(16),9288.
5. Dhivya Elavarasan, P. M. Durairaj Vincent, 2020. “Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications”, *Article in IEEE Access*, vol. 8, pp. 86886-86901, doi:10.1109/ACCESS.2020.2992480.
6. Srinivasan S, et al. Deep convolutional neural network-based image spam classification. In 2020 6th conference on data science and machine learning applications (CDMA). IEEE; 2020.
7. Archontoulis S V, et al. Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt. *Crop Science*. 2020; 60(2):721–738.
8. Ersoz ES, Martin NF, Stapleton AE. On to the next chapter for crop breeding: convergence with data science. *Crop Science*. 2020;60(2):639–655.
9. Washburn JD, Burch MB, Franco JAV. Predictive breeding for maize: Making use of molecular phenotypes, machine learning, and physiological crop models. *Crop Science*. 2020;60(2):622–638.
10. Sarker IH. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Sci*. 2021;2(6):420.