Crop Yield Prediction for Cereals using Machine Learning

Shweta Koparde, Akanksha Behare, Simran Kasare, Janhavi Patil, Kripa Nadar

Computer Engineering, Dr. D. Y. Patil Institute of Technology, Sant Tukaram Nagar, Pimpri, Pune, Maharashtra, India

Corresponding author: Shweta Koparde, Email: shweta.koparde@dypvp.edu.in

Agriculture is a major contributor to India's financial well-being. However, challenges like population growth and climate change impact crop production. Machine learning is vital for crop forecasting and decision-making, assisting in selecting crops and optimizing farming practices According to our analysis, the most used features are humidity, temperature, soil type, rainfall, pH, area, production and the applied algorithm is Decision Tree, Support Vector Machine (SVM), Random Forest & Gradient Boosting to cultivate. This prediction helps identify the best cereal crops based on weather conditions. In recent years, farmers have faced issues like reduced rainfall and poor soil quality, leading to crop failures. Precision farming helps adapt crop management to changing environmental conditions, promoting smart farming. The study aims to help people grow high-yield cereal crops, plan their activities, and find solutions to agricultural challenges.

Keywords: Crop yield prediction, Agriculture, temperature, rainfall, soil, Machine Learning, Cereals.

in: *Advancements in Communication and Systems.* Ed. by Ashish Kumar Tripathi and Vivek Shrivastava. Computing and Intelligent Systems, SCRS, India., 2023, pp. 407–418. DOI: https://doi.org/10.56155/978-81-955020-7-3-36

1 Introduction

Farming is the soul of India, filling in as the essential wellspring of business for a huge number of individuals and assuming a vital part in the country's economy. In a country with such a vast and diverse agricultural landscape, the ability to accurately predict crop yields is of paramount importance. Precise harvest yield forecast supports food security as well as enables ranchers and policymakers with basic data to streamline asset designation and work on horticultural practices.

Traditionally, crop yield prediction has been a challenging endeavour, often relying on historical data, weather patterns, and expert knowledge. However, with the advent of modern data analytics and machine learning techniques, there has been a paradigm shift in the way we approach this problem. In recent years, supervised learning algorithms, particularly classification algorithms, have emerged as powerful tools for crop yield prediction in India.

ML offers a practical appeal to enhance crop yield prediction by leveraging multiple features. AI, specifically machine learning, can identify patterns and relationships within datasets, extracting valuable insights from the information available. To build predictive models, historical datasets are used for training, where past outcomes serve as the basis for learning. During the testing phase, a portion of the historical data not involved in training is employed to assess model performance.Our research is dedicated to developing a crop prediction system specifically tailored for cereals in India. To enhance the system's accuracy, we have incorporated real-time datasets, with a primary focus on key cereal crops such as rice, wheat, and maize. The study aims to optimize recommendations by leveraging up-to-date information and fine-tuning the model's performance based on the dynamic conditions of cereal cultivation in India.

This research paper delves into the application of machine learning techniques for predicting crop yields, with a specific focus on cereals. The study aims to not only forecast crop yields accurately but also recommend the most suitable crop based on the environmental conditions. Employing algorithms such as Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Decision Trees which requires less computational complexity & small training dataset in comparison to alternative algorithms like Artificial Neural Networks (ANN), K Nearest Neighbour, Naïve Bayes. This paper elucidates the unique attributes of our algorithm, highlighting its efficiency and efficacy in crop yield prediction for cereals, ultimately contributing to informed decision-making in agriculture and promoting sustainable farming practices.

Contributions

In this review, we present a complete way to deal with address the basic issue of harvest yield expectation in India, a country profoundly subject to farming for its food and financial prosperity. Our research centres on maize, rice and wheat. Our endeavour seeks to revolutionize our approach to agriculture in India, making significant contributions to the country's agricultural sustainability. These picked crops, we have tracked down through broad factual examination, structure the foundation of India's farming scene. They not just contribute considerably to the country's food supply yet additionally assume a vital part in the economy.

All in all, this survey fills in as a compass directing specialists, policymakers, and rural partners toward the promising fate of harvest yield forecast in India. By understanding the collaboration between cutting edge preprocessing procedures and regulated learning calculations, we are better prepared to graph a course toward a more prosperous rural scene in India.

Our study became feasible through the collection of climates, agricultural, pesticide, and chemical data from various regions across India. The input data underwent preprocessing to address missing values, standardize or scale attributes, and convert categorical variables. We utilized a variety of ML algorithms, including random forests, decision trees, and gradient boosting techniques, support vector machines (SVM)to construct predictive models. The models are prepared on verifiable information and assessed utilizing proper execution measurements to guarantee precision and dependability. In this work, we have been gone up against with the accessibility of environment and rural information for India. Regardless of this impediment, we were consoled that the information we had the option to gather permitted us to plan an expectation model while staying away from overfitting or underfitting cases.

2 Motivation

The motivation behind developing a machine learning-based crop prediction system for cereals lies in the urgent need to address the complexities faced by modern agriculture. With the global challenges of climate change, population growth, and the imperative for sustainable practices, providing farmers with a precise and adaptable tool for optimal crop selection becomes paramount. Traditional methods often fall short in considering the multifaceted factors influencing agricultural productivity. Leveraging machine learning algorithms aims to revolutionize crop recommendations by honing in on data-driven insights, offering farmers a reliable decision support tool. This research endeavours to empower farmers, reduce uncertainties in crop selection, and enhance yields. Furthermore, the project aligns with the broader objective of promoting sustainability in agriculture, contributing to resource efficiency and resilience against climate-related challenges.

3 Literature Review

Thomas van Klompenburga, Ayalew Kassahuna, and Cagatay Catalb [1] employ machine learning, notably Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), for accurate crop yield prediction utilizing variables like temperature, rainfall, and soil type. One advantage is the enhanced accuracy in yield prediction, while a disadvantage lies in the requirement for extensive, precise data for effective model training, which can be challenging to obtain universally.

The paper "Crop Yield Prediction using Machine Learning and Deep Learning Techniques" by Kavita Jhajharia et al. [2] employs various machine learning methods to forecast crop yield in Rajasthan, India. While enhancing prediction accuracy, computational complexity poses a challenge, potentially hindering widespread adoption in agricultural settings.

In "Advances in Engineering Software 175 103326," authors S. Iniyan, V. Akhil Varma, and C. Teja Naidu [3] utilize various regression models, including LSTM, to forecast crop yields based on climate and soil parameters. Despite highlighting LSTM's superior accuracy and providing a user-friendly web application for farmers, scalability challenges persist due to computational complexities associated with implementing such models broadly.

Bharati Panigrahi, Krishna Chaitanya Rao Kathala, and M. Sujatha's [4] it highlighting the potential of deep learning and computer vision for disease detection, the study underscores technology's pivotal role in precise forecasting, despite potential challenges in implementation.Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, and B. Jabber [5] simplify crop yield prediction, utilizing advanced regression techniques to enhance accuracy. However, reliance on third-party data sources and language accessibility may present implementation challenges.Pallavi Kamath, Pallavi Patil, Shrilatha S, Sushma, and Sowmya S [6] explore crop yield forecasting using data mining to enhance Indian agriculture, leveraging machine learning for accuracy. However, scalability limitations to larger areas and districts pose challenges, limiting widespread applicability.In "Decision Analytics Journal 3 100041," S.K.S. Durai and M.D. Shamili [7] review crop cultivation, weed detection, and insect identification using machine learning. They suggest integrating various algorithms for improved accuracy, providing comprehensive agricultural solutions, yet complexity in implementation may pose challenges.

Dr. Jayanarayan Reddy and Dr. M. Rudra Kumar [8] explore crop yield prediction using machine learning, aiming to bolster agricultural productivity in the face of environmental uncertainties. While machine learning algorithms facilitate efficient crop selection and yield enhancement, limitations include challenges in capturing nonlinear relationships and reduced prediction efficiency in specific scenarios.Mamunur Rashid, Bifta Sama Bari, Yusri Yusup, and Mohamad Anuar Kamaruddin [9] meticulously review crop yield prediction, particularly focusing on palm oil, employing machine learning algorithms. Advantages include the high accuracy and efficiency of machine learning models compared to traditional statistical methods. However, challenges may arise in addressing complex agricultural structures and ensuring scalability across diverse environmental conditions.In "Analysis of agricultural crop yield prediction using statistical techniques of machine learning" by Janmejay Pant, Rajendra Pant, Manoj Kumar Singh, et al., [10] the authors explore crop yield prediction in India, leveraging machine learning to improve productivity. While machine learning enables accurate predictions and identifies key features like rainfall and temperature, challenges may include model selection complexity and the need for robust generalization across diverse crop types.In a research study led by S. Deepa, Alli, and Sheetal [11] the focus was on cotton yield. The study harnessed a dataset containing cotton prices sourced from various states in India, specifically focusing on cotton price data for the year 2019. The predictive models, developed using this dataset, were subsequently compared with data from 2020, revealing a remarkable consistency between predicted and actual prices.Nigam A., Garg S., Agrawal A., and Agrawal P.'s [12] research on crop yield prediction via machine learning presents a promising solution to optimize crop selection in India, potentially enhancing agricultural productivity and farmer income. However, the complexity of algorithms and data availability might hinder widespread adoption among farmers, posing a challenge for practical implementation.

Bang S., Bishnoi R., Chauhan A.S., Dixit A.K., Chawla's [13] research on crop yield prediction integrates ARMA, SARIMA, and ARMAX models with fuzzy logic, aiding Indian agriculture by enabling proactive planning. Despite challenges in accurately forecasting rainfall, the fuzzy logic integration ensures reasonable yield estimations within defined ranges, mitigating weather forecast uncertainties.Bhosale, S. V., Thombare, R.A., Dhemey, P.G., and Chaudhari, A.N.'s [14] study on crop yield prediction in India through data analytics and hybrid approaches facilitates informed decision-making for farmers, utilizing techniques like K-means clustering and Apriori algorithm. However, challenges such as data availability and algorithm complexity may impede widespread adoption, despite the potential to optimize agricultural productivity.Kavita and Pratistha Mathur's [15] research demonstrates the potential of machine learning in predicting crop yields for Indian farmers, offering insights for optimized agricultural production. However, challenges like data availability and algorithm complexity may impede widespread adoption, despite the benefits of informed decision-making and improved yields for small-scale farmers.Kavita and Pratistha Mathur's [16] investigation into satellite-based crop yield prediction using machine learning recognizes its potential benefits for informed agricultural decision-making. However, challenges such as data standardization and model optimization persist, even though advanced techniques like deep learning show promise in enhancing forecasting accuracy, suggesting avenues for further research and integration with farmer expertise to maximize agricultural outcomes.Bali, Nishu, and Anshu Singh's [17] exploration of machine learning for crop yield prediction recognizes its potential advantages for addressing agricultural sustainability but underscores challenges like parameter standardization and model optimization. Despite these hurdles, advanced techniques like neural networks show promise, signaling opportunities for further research in hybridized and deep learning approaches to improve crop yield forecasting.Ananthara, M. G., Arunkumar, T., and Hemavathy, R.'s [18] (2013) CRY algorithm employs bee hive clustering for crop yield prediction, showing superior accuracy compared to the C&R tree algorithm. However, its applicability to diverse crops and regions may be limited, despite success in specific agricultural contexts.

Awan, A. M., & Sap, M. N. M. [20] developed an intelligent system for crop yield prediction, employing kernel methods and considering parameters like plantation, latitude, temperature, and rainfall precipitation. Their experiment with weighted k-means kernel method with spatial constraints showed promise in analyzing oil palm fields, although limitations may arise in its generalizability to other crops and regions.Anakha Venugopal, Aparna S, Jinsu Mani, and Rima Mathew's [20] study elucidates machine learning's utility in agricultural decision-making via crop yield prediction, though challenges like historical data reliance and parameter precision persist. Despite these obstacles, classifiers like Random Forest exhibit promising accuracy, offering farmers informed crop selection.

4 Material and Methods

4.1 Study Area

The research area in India covers a vast expanse, and India covers a landmass of approximately 3.287 million square kilometres and the population of the India was estimated to be over 1.3 billion peoples. India is the seventh largest country in the world by land area. It is vast and diverse country, is an agricultural powerhouse with a rich mosaic of climate and ecosystems. In this research, we selected cereals like Rice, Maize, and Wheat.

Rice is vital crop in India, grown in regions like Punjab, West Bengal, Uttar Pradesh, Andhra Pradesh and Tamil Nadu. India is one of the largest rice producers globally. Rice has two main types: Basmati (It is long grain) and Non-Basmati (It includes various kinds that we used in everyday dishes). Maize it also called as corn, and it is widely grown in states likes Karnataka and Andhra Pradesh. It is versatile crop for human food, animal feed and industrial products.

In India, the kind of crops we grown in each state depends on the weather there. For example, places in the north where it gets cold are good for wheat while states in the south with warm weather are great for growing rice.

4.2 Harvested Crops

In India, Rapid population growth and the need for more food have let to expanding agriculture. Different regions of India have diverse agricultural systems because of their unique environments. Each region specializes in crops that align with its climate, soil and geographical conditions, resulting in a diverse and abundant agricultural landscape across the country.

In India we growRice in Haryana, Punjab,Assam, West Bengal, coastal regions of Odisha, Andhra Pradesh, Telangana, Tamil Nadu, Kerala and Maharashtra.Maize is primarily grown in states like Karnataka, Andhra Pradesh, and Maharashtra, while wheat cultivation is concentrated in states such as Punjab, Haryana, and Uttar Pradesh in India.

4.3 Data Sources

This study is based on agricultural data, Rainfall data, Temperature data, pH, Nitrogen, Phosphorous, Potassium, Humidity, area and production. Agriculture data is real data which is collected from Kaggle and real data. For the real time data, we visited some areas in neighbour villages and collected the necessary information according to the parameters. We have chosen reliable data in particular crop yield: Rice, Maize& Wheat. The parameters of our prediction models are famous and easily available and the parameters are:

1 **Temperature:** The average annual temperature in India typically varies from around 5°C in coldest months to approximately 50°C int the Warmest months. Temperature variations depending on the time of year and the specific location within the country.
2 **Nitrogen:** The Nitrogen contents in soil varies, with an average ranging from 0.1% to.5%, but can it be higher in certain regions or due to fertilization.
3 **Phosphorous:** The phosphorous content in soil typically varies, with an average ranging from 0.01% to 0.5%. It used for strong root growth, better flowering and fruiting and resistance to stress, all of which enhance crop yields and quality in agriculture.

4. **Potassium:** Potassium contents in soil commonly falls within the range of 0.1% to 2%. Potassium is crucial for plant health and plays a significant role in soil fertility.
5. **pH:** The pH level in soil can range from acidic to alkaline. Soil pH levels range from 5.5 to 7.5.
6. **Rainfall:** Rainfall in India varies widely by region and season. On average, India receives around 1,170 millimetres of annual rainfall.
7. **Area:** The area that corresponds to the total sown area.
8. **Production:** The amount of crop produced.
9. **Humidity:** Humidity in India varies with the climate and region. Coastal areas experience higher humidity levels, often exceeding 70% or more, while drier inland regions may have humidity levels below 50%.

5 Proposed Method

We have proposed a conceptual system centered around machine learning models, comprising five crucial steps: ETL (Extract, Transform, Load), feature engineering, model training, evaluation, and model deployment. In the ETL phase, we initially collect crop data from various sources, followed by a series of transformations, cleaning, and processing steps specific to each data source. The refined dataset is ultimately merged and loaded into a centralized storage system. The feature engineering phase involves the application of data analysis techniques to unveil hidden insights within the dataset, with the primary goal of preparing the crop data for machine learning models.

In the model training step, we utilize three distinct machine learning algorithms to train on the combined dataset, aiming to create a predictive model for crop yield estimation. These algorithms include decision trees, random forests, support vector machines (SVM), and gradient boosting techniques. The overarching principle of machine learning is to develop a predictive crop yield model capable of generalizing well to new data, minimizing predictive errors. This machine learning model can be conceptualized as a complex function 'h' within the realm of supervised learning. It takes a crop data matrix 'X' as input, along with optimization parameters denoted as 'Theta', producing an output 'Y.' Equation (1) provides a generalized mathematical representation of this process.

The crop dataset can be viewed as a pair (X, Y), where $X = [x1, x2, \ldots, xn]$, with $xi \in \mathbb{R}1 \times m$, and $Y \in \mathbb{R}m$. X is a matrix of input data where the columns represent crop features, and the rows represent time-series data collection in each country. Y represents the predicted crop yield. m is the number of instances in the dataset, and n is the number of features considered.

$$X = [\ 1, \ldots, xn] \Longrightarrow h\ (X, \Theta) \Longrightarrow \mathrm{Y} \qquad (1)$$

5.1 Decision Tree:

A decision tree is a visual tool used to make decisions or solve problems. It's like a flowchart with branches and nodes, where each node represents a choice, and each branch represents an outcome or decision. We should start at the top (the root), we follow the branches based on our choices, and we reach to the conclusion at the end (the leaf).

Steps to calculate decision tree:

1. Identify the Decision or Problem
2. List Choices
3. Determine Outcomes
4. Assign Values or Probabilities
5. Create the Tree
6. Analyse the Tree

5.2 Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression tasks. It seeks to find a hyperplane that maximizes the margin between different classes, making it effective when data is linearly separable. SVMs can also handle non-linear data through kernel functions. They work well in high-dimensional spaces, making them valuable for various applications, but their performance can be sensitive to outliers and requires careful selection of parameters. SVMs are widely used for tasks like image and text classification.

The objective of the support vector machine (SVM) algorithm involves identifying a hyperplane within an N-dimensional space, where N represents the number of features within the dataset, to effectively separate different classes of data points. Multiple potential hyperplanes could be considered to distinguish between these classes. The aim is to select a plane that offers the best separation, which means having the maximum distance between data points of the two classes [12]. Maximizing this distance, known as the margin, provides a buffer that aids in confidently organizing future data points. To achieve this, the SVM algorithm utilizes the Hinge function. The Hinge function is instrumental in increasing and maximizing the margin between the data points and the hyperplane, contributing to better classification and improved generalization of the model.

$$c\,(x, y, (x)) = \{0,\ if\ y{*}(x) \geq 1 \\ 1 - y{*}(x),\ \mathrm{else}\} \tag{2}$$

Hinge loss function

When the predicted value aligns with the actual value, the cost amounts to 0. However, if there's a disparity in signs between the predicted and actual values, a loss value is computed. To ensure a balanced regularization effect, the regularization function is incorporated into the Hinge loss function. This combined function helps manage the trade-off between maximizing the margin (the Hinge loss function) and preventing overfitting or excessive complexity in the model (achieved through regularization).

$$m\mathrm{in}\omega\lambda\ ||\omega||\ 2 + \textstyle\sum_{\mathrm{i=1}}^{\mathrm{n}}(1 - \mathrm{Yi}\ (\mathrm{xi}, \omega)) + \tag{3}$$

The loss function in SVM sets the stage for the partial differentiation of relevant weights to determine the gradients. These gradients are instrumental in updating the {weights. Post differentiation, the addition of a regularization parameter ensures the prevention of misclassification. Even after this adjustment, if any misclassification persists, the model might err in predicting the class of data points. To address this possibility, the loss, coupled with the regularization parameter, plays a crucial role in executing the final gradient update, striving to minimize errors and refine the model's predictive accuracy.

$$\frac{\delta}{\delta\omega_{\mathrm{k}}}\lambda\ ||\omega||\ 2 = 2\lambda\omega\mathrm{k} \tag{4}$$

$$\frac{\delta}{\delta\omega_{\mathrm{k}}}(1\text{-yi (xi}, \omega)) + = \begin{cases} 0, & \mathrm{if}\ \mathrm{y} * \mathrm{x}\ \mathrm{f(x)} > 1 \\ -\mathrm{y_i x_{ik}}\ , & \mathrm{else} \end{cases} \tag{5}$$

Gradients

$$\omega = \omega - \alpha 2\lambda\omega \tag{6}$$

Gradient Update – No misclassification

$$\omega = \omega + \alpha\ (yixi - 2\lambda\omega) \tag{7}$$

Gradient Update – Misclassification

5.3 Random Forest

Random Forest is a machine learning approach that employs an ensemble of decision trees to enhance predictive accuracy. Its strategy involves addressing challenges such as overfitting and data noise by aggregating outcomes from multiple trees. Known for its versatility, Random Forest is well-suited for both classification and regression tasks. Its strengths lie in effectively managing extensive datasets, dealing with high-dimensional features, and evaluating feature importance.

5.4 Gradient Boosting

In machine learning, Gradient Boosting is a powerful technique that builds a predictive model by combining multiple weak models, usually decision trees. It iteratively corrects errors made by the previous models, resulting in a highly accurate and robust ensemble. Gradient Boosting is known for its ability to handle complex patterns in data and is widely used in applications like classification, regression, and ranking. However, it requires careful hyperparameter tuning and can be sensitive to overfitting.

System Architecture (see Figure 1)

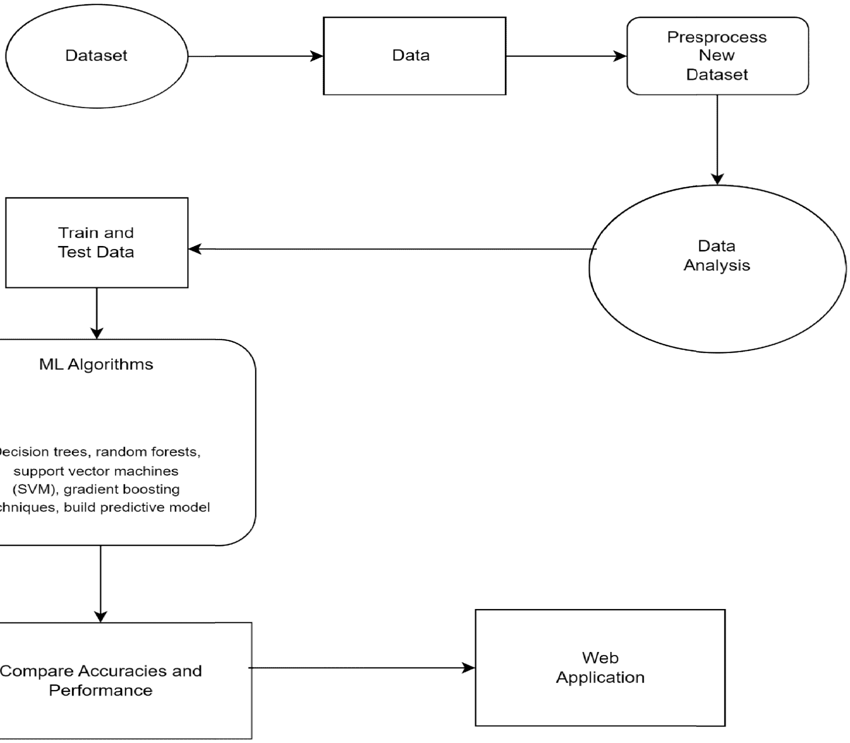

Figure 1. System Architecture

6 Results & Discussions

The analysis reveals intriguing trends in crop production and its correlation with rainfall patterns from 1997 to 2019. Over this period, while overall crop production has shown a steady increase, Rice and Maize display notable exceptions due to consistent changes in rainfall patterns. This discrepancy suggests a shift towards modern irrigation techniques for most crops, while rice, a water-intensive crop, continues to rely primarily on natural water sources like rain. Let's delve into each crop separately.(Figure 2) The graph Annual rainfall vs. Year (Maize) shows a fluctuating trend in annual rainfall for maize over the years. The rainfall seems to have been highest around the year 2000 and has been generally decreasing since then, with some fluctuations.(Figure 3) Similar to maize, the graph shows a fluctuating trend in annual rainfall for rice over the years. The rainfall seems to have been highest around the year 2000 and has been generally decreasing since then, with some fluctuations. (Figure 4) For Maize, the area under irrigation seems relatively consistent, experiencing a slight dip between 1999 and 2002 for reasons unknown. Yet, from 2003 to 2006 and 2013 to 2019, the area appears constant. This constancy could be attributed to better seasonal rainfall, improved irrigation methods, and advancements in farming practices likeenhanced fertilizers and pesticides, contributing to more stable production. In contrast, (Figure 5) examining Rice reveals fluctuations in the area under irrigation and production. There was a decline in the area used between 2000 and 2003, followed by a gradual increase until 2018.For Maize (Figure 6) the production remained relatively consistent over the years. Despite this, (Figure 7) the production for Rice remained relatively consistent over the years, with a notable increase in 2018. Several factors might explain this trend, including a growing necessity to supply other regions with produce. These observations suggest that land availability, reflected in the area under cultivation, has remained relatively static over time. This limitation indicates a challenge in expanding agricultural land due to competing demands for land in various sectors. However, the escalating demand for crops due to population growth requires increased production. The consistent rise in yield despite limited land availability underscores the significance of modernized irrigation, production techniques, and improved agricultural inputs like fertilizers, contributing to enhanced crop yield. In essence, while land for cultivation remains finite, the application of advanced agricultural methodologies has facilitated increased crop yields, meeting the growing demands despite land constraints.

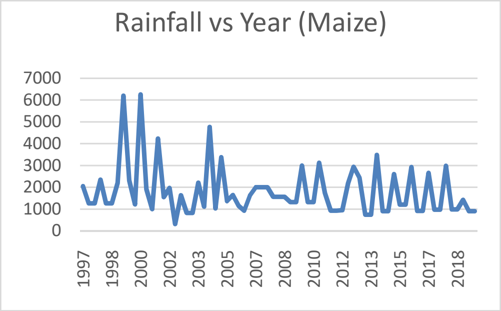

Figure 2. Annual rainfall vs. Year (Maize)

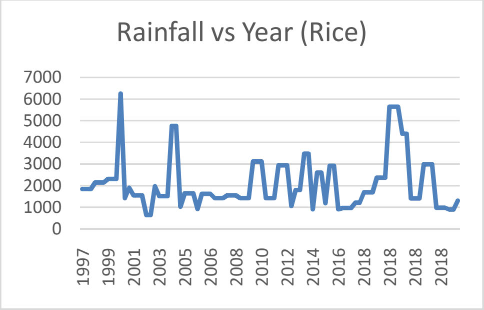

Figure 3. Annual rainfall vs. Year (Rice)

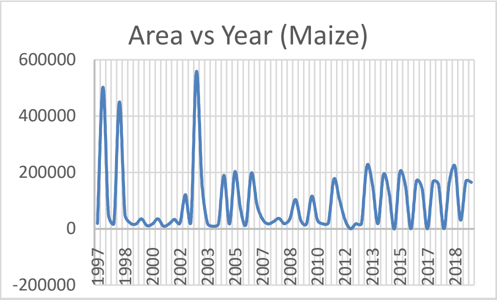

Figure 4. Area under irrigation vs. Year (Maize)

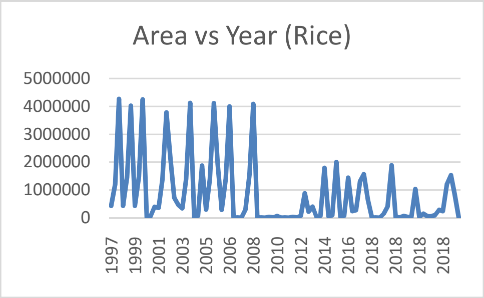

Figure 5. Area under irrigation vs. Year (Rice)

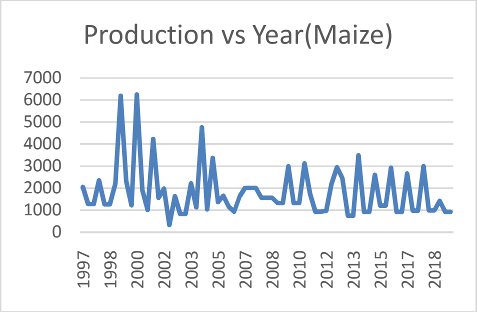

Figure 6. Production vs. Year (Maize).

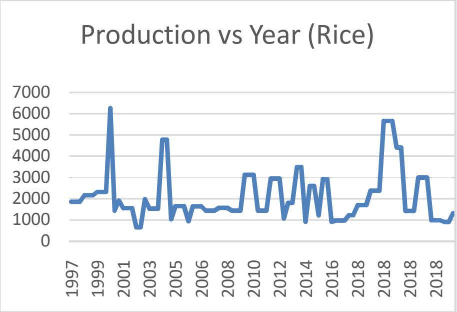

Figure 7. Production vs. Year (Rice)

Table 1 shows the comparison of all models

Table 1. Model accuracies (R2 Score)

MODELS	ACCURACY (R2 SCORE)
Random Forest	0.9931
Decision Tree	0.9818
Support Vector Machine	0.9681
Gradient Boosting	0.9818

In this research investigation, we conducted a thorough assessment of various machine learning algorithms aimed at predicting crop outcomes, with a specific focus on cereals such as maize, wheat, and rice. The effectiveness of each algorithm was measured using the R2 score, a metric indicating the degree of fit between the models and the actual crop yield data. The results (Table 1) underscore the exceptional accuracy of the Random Forest algorithm, exhibiting an impressive R2 score of 0.9931. This result signifies the algorithm's ability to predict crop yields for the cereals under consideration with remarkable precision.

In close pursuit, both the Decision Tree and Gradient Boosting algorithms demonstrated robust predictive capabilities, each achieving R2 scores of 0.9818. These findings highlight the suitability of these algorithms for precise crop yield predictions, providing valuable insights for agricultural planning and resource allocation. Additionally, the Support Vector Machine algorithm showcased commendable performance with an R2 score of 0.9681, further emphasizing its potential as a reliable tool in crop prediction systems. The consistently high R2 scores across all algorithms affirm the importance of incorporating machine learning techniques to enhance accuracy and efficiency in agricultural practices.

In summary, our research underscores the pivotal role of machine learning algorithms in advancing crop prediction systems, specifically for crucial cereals such as maize, wheat, and rice. The results validate the effectiveness of Random Forest, Decision Tree, Support Vector Machine, and Gradient Boosting algorithms in accurately forecasting crop yields. These findings contribute significantly to the field of precision agriculture, offering a robust foundation for implementing data-driven decision-making processes to optimize crop production and ensure food security.

7 Conclusion & Future Scope

In this research study, diverse machine learning algorithms were employed to forecast crop yields in India, utilizing a dataset focused on four key crops: rice, wheat and maize. Notably, the Decision Tree demonstrated the highest accuracy in predicting crop yields. Looking ahead, potential improvements to the model's predictive capabilities can be achieved by incorporating additional pertinent features in future iterations of the research.This comprehensive outlook provides a more expansive and representative perspective of the country's agricultural landscape, fostering a holistic comprehension of crop yield dynamics. Furthermore, the integration of real-time weather data into predictive models marks the onset of an era characterized by heightened precision and dependability. By harnessing state-of-the-art technology and advanced analytical techniques, these inventive approaches augment the accuracy of yield projections, empowering farmers, policymakers, and agribusinesses with valuable, actionable insights. It is essential to emphasize that the progress in this field has been achieved through original research and ethical research practices, placing a paramount importance on the authenticity and integrity of findings. This dedication ensures the ongoing development and sustainability of this pivotal domain within the agricultural sciences.

Reference

[1] Thomas van Klompenburg, Ayalew Kassahun Cagatay Catal Crop yield using machine learning by Information Technology in 2020 on Computers and Electronics in Agriculture.

[2] Kavita Jhajharia, Pratistha Mathur, Sanchit Jain, Sukriti Nijhawan in 2023 in International Conference on Machine Learning and Data Engineering. 2

[3] S Iniyan, V Akhil Varma, Ch Teja Naidu Crop yield prediction using machine learning techniques from department of computer technologies in 2023 on science direct. 3

[4] Bharati Panigrahi, Krishna Chaitanya Rao Kathala, M. Sujatha, A Machine Learning-Based Comparative Approach to Predict the Crop Yield Using Supervised Learning with Regression Models by Procedia Computer Science 218 (2023). 4

[5] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber, Crop Yield Prediction based on Indian Agriculture using Machine Learning in 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020. 5

[6] Pallavi Kamath, Pallavi Patil, Shrilatha S, Sushma, Sowmya S, Crop yield forecasting using data mining in Global Transition Proceedings 2(2021).402-407. 6

[7] Senthil Kumar Swami Durai, Mary Divya Shamili, Smart farming using Machine Learning and Deep Learning techniques in Decision Analytics Journal 2022 on Science Direct. 7

[8] Dr. Jayanarayan Reddy, Dr. M. Rudra Kumar, Crop Yield Prediction using machine learning algorithm in in International Conference on Intelligent Computing and Control System in 2021. 8

[9] Mamunur Rashid, Bifta Sama Bari, Yusri Yusup, Mohamad Anuar Kamaruddin, Nuzat Khan, Comprehensive Review of crop yield prediction using machine learning approach with special emphasis on palm oil yield prediction in 2021. 9

[10] Janmeja Pant, R.P. Pant, Manoj Kumar Singh, Devesh Pratap Singh, Himanshu Pant, Analysis of agricultural crop yield prediction using statistical techniques of machine learning in 2021. 10

[11] S. Deepa, A. Alli, Sheetac, S. Gokila, Machine learning regression model for material synthesis prices prediction in agriculture in 2020. 11

[12] Nigam A., Garg S., Agrawal A., Agrawal P., (2019) Crop yield prediction using machine learning algorithm in fifth international conference on Image Information Processing Nov 15 IEEE.

[13] Bang S., Bishnoi R., Chauhan, A.S., Dixit, A.K., Chawla (2019) in twelfth International Conference on Contemporary Computing.

[14] Bhosale, S. V., Thombare, R.A., Dhemey, P.G., Chaudhari, A.N (2018). Crop Yield Prediction suing Data analytics and hybrid approach in fourth International Conference on computing Communication Control and Automation.

[15] Kavita, Ms and Pratistha Mathur (2020), Crop Yield Estimation in India using Machine learning in 5th international Conference on computing Communication Control and Automation.

[16] Kavita, Pratistha Mathur (2021) Satellite- based Crop Yield Prediction using machine learning algorithm in Asian Conference on Innovation in technology.

[17] Bali, Nishu, Anshu Singh (2022) Emerging trends in machine learning to predict crop yield and study its influential factors.

[18] Ananthara, M. G., Arunkumar, T., & Hemavathy, R. (2013, February). CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (pp. 473-478). IEEE

[19] Awan, A. M., & Sap, M. N. M. (2006, April). An intelligent system based on kernel methods for crop yield prediction. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 841-846). Springer, Berlin, Heidelberg.

[20] Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Vinu Williams, 2021, Crop yield prediction using machine learning International Journal of Engineering Research and technology.