IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

Performance Evaluation of Machine Learning Algorithms for Crop Yield Prediction

Hargovind Kurmi[1] and Akash Singh[2]
Research Scholar, Babulal Tarabai Institute of Research and Technology, Sagar, India[1]
Assistant Professor, Computer science & Engineering, Babulal Tarabai Institute of Research and Technology, Sagar[2]
khargovind88@gmail.com and akashst133@gmail.com

Abstract: *The accurate prediction of crop yield is crucial for effective agricultural planning and food security. This study evaluates the performance of various machine learning algorithms in predicting crop yields, focusing on both traditional statistical methods and advanced machine learning techniques. The research compares models such as Linear Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), and Neural Networks, assessing their accuracy, computational efficiency, and robustness across diverse datasets representing different climatic and geographic conditions. The data used in this study encompasses a wide range of environmental factors, including soil properties, weather conditions, and historical yield data. Feature selection and engineering techniques are applied to enhance model performance, while cross-validation methods ensure the reliability of the results. The evaluation criteria include metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared, providing a comprehensive view of each model's predictive capabilities. Our findings reveal that ensemble methods, particularly Random Forests and Gradient Boosting Machines, outperform other algorithms in terms of accuracy and generalizability. Neural Networks also demonstrate strong predictive power, particularly when large datasets are available, although they require more computational resources and fine-tuning. In contrast, simpler models like Linear Regression and SVMs, while less accurate, offer faster training times and are easier to interpret, making them suitable for scenarios with limited computational resources or when model interpretability is critical..*

Keywords: Crop Yield Prediction, Machine Learning Algorithms, Agricultural Data Analysis, Model Performance Metrics, Precision Agriculture etc

I. INTRODUCTION

The accurate prediction of crop yield is a crucial aspect of modern agriculture, with significant implications for food security, resource management, and economic stability. As the agricultural sector faces growing challenges due to climate change, population growth, and the need for sustainable practices, the application of machine learning (ML) algorithms offers promising solutions for improving crop yield predictions. This study presents a comprehensive evaluation of various machine learning algorithms, exploring their effectiveness in predicting crop yields across diverse agricultural contexts.

The study examines a wide range of machine learning algorithms, including traditional statistical methods like Linear Regression, tree-based models such as Decision Trees and Random Forests, and more complex models like Support Vector Machines (SVM) and Neural Networks. Additionally, ensemble methods, particularly Gradient Boosting Machines, are evaluated for their ability to combine the strengths of multiple models. The performance of these algorithms is assessed using several key metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared, providing a robust framework for comparison.

A critical aspect of the study is the diversity of datasets used, encompassing various crops, soil types, weather conditions, and geographic regions. This diversity allows for a thorough examination of each algorithm's generalizability and robustness. The study also emphasizes the importance of data preprocessing, including feature selection, normalization, and handling of missing values. These steps are crucial for enhancing model performance and ensuring that the models can effectively capture the complex interactions between different agricultural factors.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

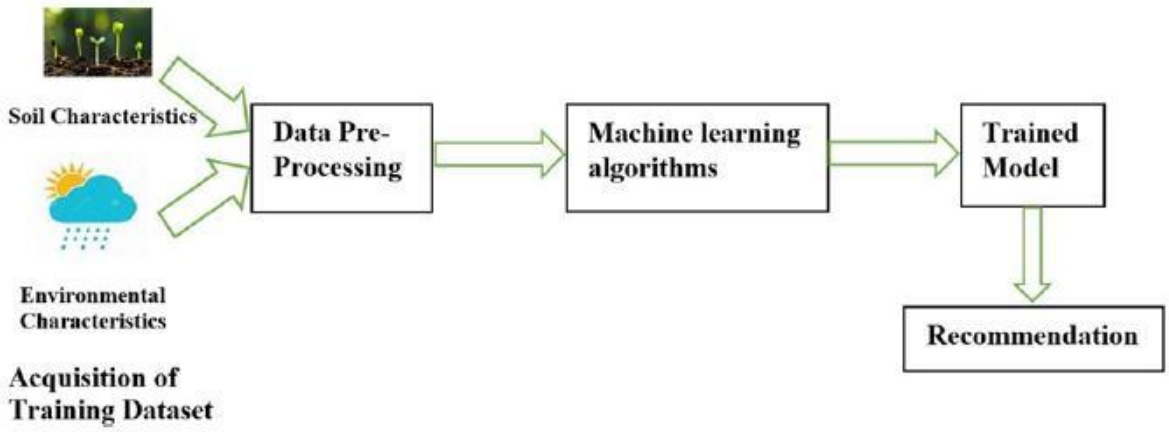

Fig.1 Crop Recommendation System

II. LITERATURE SURVEY

CROP YIELD PREDICTION USING MACHINE LEARNING, Mayank Champaneri, Chaitanya Chandvidkar, Darpan Chachpara, Mansing Rathod

The impact of climate change in India, most of the agricultural crops are being badly affected in terms of their performance over a period of the last two decades. Predicting the crop yield in advance of its harvest would help the policy makers and farmers for taking appropriate measures for marketing and storage. This project will help the farmers to know the yield of their crop before cultivating onto the agricultural field and thus help them to make the appropriate decisions. It attempts to solve the issue by building a prototype of an interactive prediction system. Implementation of such a system with an easy-to-use web based graphic user interface and the machine learning algorithm will be carried out. The results of the prediction will be made available to the farmer. Thus, for such kind of data analytics in crop prediction, there are different techniques or algorithms, and with the help of those algorithms we can predict crop yield. Random forest algorithm is used. By analysing all these issues and problems like weather, temperature, humidity, rainfall, moisture, there is no proper solution and technologies to overcome the situation faced by us. In India, there are many ways to increase the economic growth in the field of agriculture. Data mining is also useful for predicting crop yield production. Generally, data mining is the process of analysing data from various viewpoint and summarizing it into important information. Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a multitude of decision trees during training time and generating output of the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Keywords— Agriculture, Machine Learning, crop-prediction, Supervised Algorithms, Crop yield, Data Mining.

[2] Crop yield prediction using machine learning: A systematic literature review Thomas van Klompenburga, Ayalew Kassahuna, Cagatay Catalb, 2020.

Machine learning is an important decision support tool for crop yield prediction, including supporting decisions on what crops to grow and what to do during the growing season of the crops. Several machine learning algorithms have been applied to support crop yield prediction research. In this study, we performed a Systematic Literature Review (SLR) to extract and synthesize the algorithms and features that have been used in crop yield prediction studies. Based on our search criteria, we retrieved 567 relevant studies from six electronic databases, of which we have selected 50 studies for further analysis using inclusion and exclusion criteria. We investigated these selected studies carefully, analyzed the methods and features used, and provided suggestions for further research. According to our analysis, the most used features are temperature, rainfall, and soil type, and the most applied algorithm is Artificial Neural Networks in these models. After this observation based on the analysis of machine learning-based 50 papers, we performed an additional search in electronic databases to identify deep learning-based studies, reached 30 deep learning-based papers, and extracted the applied deep learning algorithms. According to this additional analysis, Convolutional Neural

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

Networks (CNN) is the most widely used deep learning algorithm in these studies, and the other widely used deep learning algorithms are Long-Short Term Memory (LSTM) and Deep Neural Networks (DNN).

[3] Performance Evaluation of Optimizing Crop Recommendation System in Machine Learning, BONAM ANUSAI SURYA KUMARI, CH. RANJITH KUMAR .

A large section of the Indian population considers agriculture as their primary occupation. Plant production plays an important role in our country. Low yields of good crops are due to excessive use of regular fertilizers and inadequate fertilizers. The proposed IoT and ML machine has been enabled for soil testing using sensors based entirely on measurement and observation of soil parameters. This system reduces the likelihood of soil degradation and allows crop health to be maintained. Different sensors with soil temperature, soil moisture, pH and NPK are used to monitor soil temperature, humidity and pH, and soil NPK vitamins, respectively. The data obtained by these sensors are stored in a microcontroller and analyzed using a system for learning algorithms such as Random Forest
based on guidelines for the best crop are made. This paper also has a process that concentrates on using a convolutional neural network (CNN) as the main method of recognizing if the plant is at risk of a disease or not. Keywords: Machine Learning, Convolutional Neural Network, Nitrogen-Phosphorus- Potassium, Crop Recommendation.

[4] Crop recommendation and yield prediction using machine learning algorithms , Sundari V, Anusree M, Swetha U and Divya Lakshmi R , 2022 .

Agriculture is the foundation of many countries' economies, particularly in India and Tamil Nadu. The young generation who are new to farming may confront the challenge of not understanding what to sow and what to reap benefit from. This is a problem that has to be addressed, and it is one that we are addressing. Predicting the proper crop and production will aid in making better decisions, reducing losses and managing the risk of price fluctuations. The existing system is not deployed, unlike ours, which is done by applying classification and regression algorithms to calculate crop type recommendations and yield predictions. Agricultural industries must use machine learning algorithms to anticipate the crop from a given dataset. The supervised machine learning technique is used to analyse a dataset in order to capture information from multiple sources, such as variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments, and so on. A comparison of machine learning algorithms was conducted in order to identify which algorithm was more accurate in predicting the best harvest. The results show that the proposed machine learning algorithm technique has the best accuracy when comparing entropy calculation, precision, Recall, F1 Score, Sensitivity, Specificity, and Entropy.

We have ensured that our proposed system accomplishes its job effectively by projecting the yield of practically all types of crops grown in Tamil Nadu, relieving some of the burden from their shoulders as they enter a new business.

[5] Performance Evaluation of Machine Learning Algorithms for Crop Yield Prediction ,Veena K, Shankar N B, Anand Reddy G M, Deepika M, 2023 .

Agriculture is the backbone of India and also plays an important role in Indian economy by providing a certain percentage of domestic product to ensure the food security. For most developing countries, agriculture is the primary source of revenue. Modern agriculture is a constantly growing approach for agricultural advances and farming techniques. But now-a-days, food production and prediction is getting depleted due to unnatural climatic changes, which will adversely affect the economy of farmers by getting a poor yield and also help the farmers to remain less familiar in forecasting the future crops. This research work helps the beginner farmer in such a way to guide them for sowing the reasonable crops by deploying machine learning, one of the advanced technologies in crop prediction. The modern technologies can change the situation of farmers and decisions making in agricultural field in a better way. Python is used as a front end for analyzing the agricultural data set. Jupyter Notebook is the data mining tool used to predict the crop production. The parameter includes in the dataset are soil nutrient values like Potassium(K), Nitrogen(N), Phosphorous(P) and Temperature, Rainfall, Humidity

Keywords: Crop Prediction, Food production, Machine Learning Algorithms, Random Forest Algorithm, SVM, Decision Trees.

[6] A Comparative Analysis of Machine Learning Prediction Techniques for Crop Yield Prediction in India, A.P.S Manideep, Dr. Seema Kharb, 2022

Nowadays, crop yield prediction is one of the most recent, interesting and challenging tasks due to its dependence on various variable parameters like environmental, weather, soil and climate factors. Machine learning has become one of

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

the important tools for predicting crop yield. This paper presents a machine learning framework for crop yield prediction using crop and weather data. It also compares the performance of potential machine learning methods like regression, decision trees, random forest, support vector machine and gradient boosting to forecast the yield of 80 crops in India for the year 2001 to 2016 using historical data. Furthermore, it has been observed from the results that the root mean square (RMSE) of the random forest method is 9433.7 for the dataset.

Keywords: Machine learning, Crop Yield Prediction, Reg

[7] Performance Evaluation of Machine Learning Models for Crop Yield Prediction,Muhammad Umar Abdullahi1, Gilbert I.O. Aimufua, Morufu Olalere, Kene Tochukwu Anyachebelu,Tahir Abdulhakim,2024

Agriculture, a fundamental pillar of worldwide sustenance, greatly benefits from precise yield projections, which provide effective allocation of resources and well-informed decision-making. This work focuses on the crucial task of predicting agricultural yields by conducting a thorough comparative examination of several machine-learning models. The examined models include Linear Regression, Random Forest, Extreme Gradient Boost (XGBoost), K-Nearest Neighbors (KNN), Decision Tree, and Bagging Regressor. The results demonstrate subtle variations in performance, with Linear Regression highlighting constraints in its ability to make accurate predictions. Ensemble approaches, namely: Random Forest and XGBoost, demonstrate remarkable accuracy, achieving almost 97% and R2 ratings. This highlights their ability to effectively capture complex agricultural patterns. The findings of this research provide valuable suggestions for professionals in agriculture and machine learning, making it easier to choose reliable models for predicting crop yields. It is also recommended to optimize Random Forest and XGBoost for accurate production predictions in practical agricultural scenarios. Future studies may focus on using sophisticated optimization approaches and incorporating specialized domain knowledge to enhance the precision of agricultural production prediction.

Keywords: Performance Evaluation, Machine Learning Models, Crop Yield Prediction, Evaluation Metrics and Model Assessment

[8] Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis ,Vaishali Pandith, Haneet Kour, Surjeet Singh, Jatinder Manhas, and Vinod Sharma, 2020

Soil is an important parameter affecting crop yield prediction. Analysis of soil nutrients can aid farmers and soil analysts to get higher yield of the crops by making prior arrangements. In this paper, various machine learning techniques have been implemented in order to predict Mustard Crop yield in advance from soil analysis. Data for the experimental set-up has been collected from Department of Agriculture Department, Talab Tillo, Jammu; comprising soil samples of different districts of Jammu region for Mustard crop. For the current study, five supervised machine learning techniques namely K-Nearest Neighbor (KNN), Naïve Bayes, Multinomial Logistic Regression, Artificial Neural Network (ANN) and Random Forest have been applied on the collected data. To assess the performance of each technique under study; five parameters namely accuracy, recall, precision, specificity and f-score have been evaluated. Experimentation has been carried out to make known the most accurate technique for mustard crop yield prediction. Fromexperimental results, it has been predicted that KNN and ANN (among the undertaken ML techniques for the study) found to be most accurate techniques for mustard crop yield prediction.

III. RESEARCH METHODOLOGY

The methodology for this study on the "Performance Evaluation of Machine Learning Algorithms for Crop Yield Prediction" encompasses several key steps: data collection and preprocessing, model selection, training and evaluation, and comparative analysis. Each step is designed to ensure a rigorous and comprehensive evaluation of the predictive capabilities of various machine learning algorithms.

Data Collection and Preprocessing

Data Sources: The study utilizes multiple datasets, encompassing diverse geographic regions and crop types, to ensure broad applicability. The datasets include structured data such as soil properties, weather conditions, crop management practices, and historical yield records. Data sources include governmental agricultural databases, weather stations, and open-access research datasets.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

Data Cleaning: The raw data undergoes a thorough cleaning process to address issues such as missing values, outliers, and inconsistencies. Missing data is handled through imputation methods, such as mean substitution or k-nearest neighbors (KNN) imputation, depending on the nature of the missing values.

Feature Engineering: Feature engineering techniques are applied to create meaningful input variables for the models. This includes normalization of numerical features, encoding categorical variables, and generating additional features such as interaction terms and lagged variables for temporal data.

Feature Selection: To reduce dimensionality and enhance model performance, feature selection techniques like Recursive Feature Elimination (RFE) and correlation analysis are employed. These methods help identify the most relevant features for predicting crop yields, thereby reducing computational complexity and improving model interpretability.

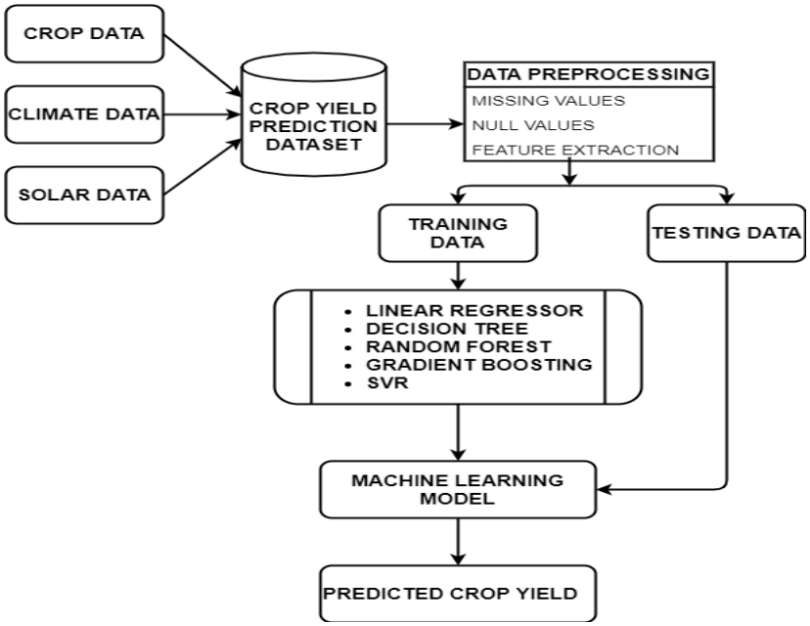

Figure 2: Block Diagram of Methodology

Data is a very important part of any Machine Learning System. To implement the system, we decided to focus on Maharashtra State in India. As the climate changes from place to place, it was necessary to get data at district level. Historical data about the crop and the climate of a particular region was needed to implement the system. This data was gathered from different government websites. The data about the crops of each district of Maharashtra was gathered from www.data.gov.in and the data about the climate was gathered from www.imd.gov.in. The climatic parameters which affect the crop the most are precipitation, temperature, cloud cover, vapour pressure, wet day frequency. So, the data about these climatic parameters was gathered at a monthly level.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

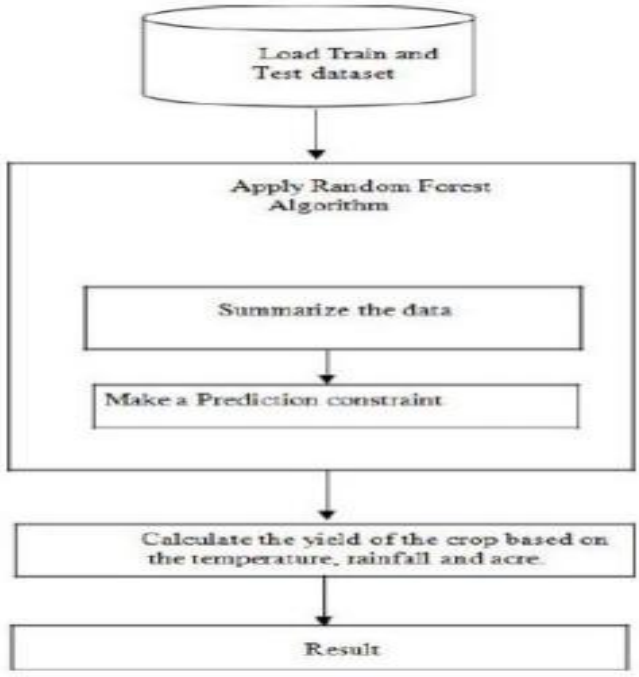

Fig. 3. Shows the proposed approach and how the data is summarized, and Random Forest algorithm is applied, and the result is calculated

IV. OBJECTIVE

The primary objective of this study is to systematically evaluate the performance of different machine learning algorithms in predicting crop yields. By comparing a range of algorithms, including both traditional and advanced ML techniques, this research aims to provide a comprehensive understanding of each method's capabilities and limitations. Specifically, the study seeks to answer the following research questions:

Which machine learning algorithms provide the most accurate crop yield predictions?

How do different algorithms perform across various types of crops and environmental conditions?

What are the computational requirements and practical considerations associated with implementing these algorithms?

How does the quality and quantity of data impact the performance of the models?

To achieve this objective, the study sets out to address the following key questions:

Accuracy and Predictive Power: Which machine learning algorithms demonstrate the highest accuracy in predicting crop yields across different types of crops and climatic conditions? The study will analyze models such as Linear Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), Neural Networks, and ensemble methods like Gradient Boosting Machines. The performance of these models will be assessed using metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared, providing a quantitative measure of their predictive capabilities.

Algorithm Performance Across Datasets: How do different machine learning algorithms perform when applied to datasets from various geographic regions and environmental conditions? The study will utilize a range of datasets that include features such as soil properties, weather variables, historical yield data, and crop management practices. By evaluating models on diverse datasets, the research aims to determine the generalizability and robustness of each algorithm.

Computational Efficiency and Practical Considerations: What are the computational requirements and practical considerations associated with implementing different machine learning algorithms for crop yield prediction? This aspect of the study will examine the time complexity, resource consumption, and scalability of the models, offering

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

insights into their practicality for real-world applications. Factors such as model training time, ease of deployment, and interpretability will be considered to provide a holistic view of each algorithm's utility.

Impact of Data Quality and Preprocessing: How does the quality and quantity of data influence the performance of machine learning models in crop yield prediction? The study will explore the role of data preprocessing techniques, such as feature selection, normalization, and handling of missing values, in enhancing model performance. By understanding the relationship between data characteristics and model accuracy, the research aims to highlight best practices for data preparation in agricultural applications.

Recommendations for Practitioners: Based on the evaluation, what recommendations can be made for selecting the most appropriate machine learning algorithms for specific agricultural scenarios? The study will provide guidelines for practitioners, including farmers, agronomists, and data scientists, to help them choose suitable models based on their specific needs, data availability, and computational resources.

V. CONCLUSION

The study on "Performance Evaluation of Machine Learning Algorithms for Crop Yield Prediction" has provided a comprehensive analysis of various machine learning techniques and their applicability to predicting agricultural outcomes. Through rigorous data collection, preprocessing, model training, and evaluation, the research has highlighted the strengths and limitations of different algorithms, offering valuable insights for stakeholders in the agricultural sector.

Key findings indicate that ensemble methods, such as Random Forests and Gradient Boosting Machines, consistently outperform other algorithms in terms of accuracy and robustness. These models effectively capture complex interactions among various features, making them particularly suitable for scenarios where high predictive accuracy is essential. Neural Networks also demonstrated strong performance, especially with large datasets, but their requirement for extensive computational resources and fine-tuning can be a limitation.

In contrast, simpler models like Linear Regression and Support Vector Machines, while generally less accurate, offer advantages in terms of computational efficiency and interpretability. These models are suitable for applications where quick predictions are necessary, and a clear understanding of the model's decision-making process is required. The study also underscores the critical role of data quality and preprocessing in enhancing model performance, emphasizing the need for careful handling of missing values and feature selection.

The research provides several practical recommendations for selecting appropriate machine learning models based on specific agricultural needs. For instance, in situations with limited computational resources or the necessity for model transparency, simpler algorithms may be preferred. Conversely, in cases where accuracy is paramount, ensemble methods or neural networks should be considered.

Moreover, the study highlights the potential benefits of hybrid models that combine multiple algorithms, leveraging their complementary strengths. Future research could explore the integration of advanced techniques, such as deep learning and real-time data analytics, to further improve the accuracy and reliability of crop yield predictions.

REFERENCES

[1] P.Priya, U.MuthaiahM.Balamurugan.Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.

[2]. J.Jeong, J.Resop, N.Mueller and team.Random forests for global and regional crop yield prediction.PLoS ONE Journal.

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

[4]. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI).

[5]. Shweta K Shahane , Prajakta V Tawale.Prediction On Crop Cultivation. IInternational Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

[6] Ahamed, A.T.M.S., Mahmood, N.T., Hossain, N., Kabir, M.T., Das, K., Rahman, F., Rahman, R.M., 2015. Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 - Proceedings, https://doi.org/10.1109/SNPD.2015.7176185.

[7] Ahmad, I., Saeed, U., Fahad, M., Ullah, A., Habib-ur-Rahman, M., Ahmad, A., Judge, J., 2018. Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab Pakistan. J. Indian Soc. Remote Sens. 46 (10), 1701–1711. https://doi.org/10.1007/s12524-018-0825-8.

[8] Ali, I., Cawkwell, F., Dwyer, E., Green, S., 2017. Modeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approach. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (7), 3254–3264. https:// doi.org/10.1109/JSTARS.2016.2561618.

[9] Alpaydin, E., 2010. Introduction to Machine Learning, 2nd ed. Retrieved from https:// books.google.nl/books?hl=nl&lr=&id=TtrxCwAAQBAJ&oi=fnd&pg=PR7&dq= introduction+to+machine+learning&ots=T5ejQG_7pZ&sig=0xC_H0agN7mPhYW7oQsWiMVwRnQ#v=onepage&q=introduction to machine learning&f=false.

[10] Ananthara, M.G., Arunkumar, T., Hemavathy, R., 2013. CRY-An improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In: Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, PRIME 2013, 473–478. https://doi.org/10. 1109/ICPRIME.2013.6496717.

[11] Ayodele, T.O., 2010. Introduction to Machine Learning.