iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VIII **Month of publication:** August 2024

DOI: https://doi.org/10.22214/ijraset.2024.63933

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VIII Aug 2024- Available at www.ijraset.com

Enhancing Agricultural Sustainability through Precision Fertilizer Recommendations: A Machine Learning Approach with Comparative Ensemble Analysis

Kshitij Koyande[1], Prof. Pallavi Thakur[2]
[1]*Master of Computer Application Sardar Patel Institute of Technology, Mumbai, India*
[2]*Professor, Master of Computer Application Sardar Patel Institute of Technology, Mumbai, India*

Abstract: ***In the ever-evolving realm of agriculture, the quest for optimal crop yield is a constant challenge, intertwined with the intricate dance of environmental variables and agricultural practices. This research embarks on a journey to revolutionize the approach to fertilizer utilization, a critical factor in the cultivation equation. Leveraging the power of machine learning, specifically an advanced iteration of the random forest algorithm, the study endeavours to usher in a new era of precision agriculture. By delving into the nuanced interplay of time-series data encapsulating rainfall patterns and crop fertility, the research not only predicts nutrient requirements but also refines the predictive model using cutting-edge ensemble learning methods—XGBoost, AdaBoost, and LightGBM. The culmination of these efforts results in a sophisticated yet accessible tool, housed within a Flask Python framework, providing farmers with personalized nutrient recommendations upon inputting crop and location data. This transformative model stands as a beacon of sustainable agriculture, offering a potent means to optimize fertilization, reduce environmental impact, and bolster soil fertility. As the agricultural landscape embraces technological innovation, this research emerges as a pioneering force, empowering farmers with data-driven insights to navigate the complex tapestry of modern farming practices.***

Keywords: Precision Agriculture, Random Forest Algorithm, Ensemble Learning, Fertilizer Optimization, Sustainable Agriculture, Crop Yield Prediction.

I. INTRODUCTION

Agriculture, a linchpin of national economic growth, contributes substantially to India's GDP and holds the global rank as the second-largest producer of farm outputs. The intricate relationship between plant growth and fertilizers is paramount, with these nutrients replenishing vital elements extracted by crops from the topsoil. The absence of precise fertilization poses a severe threat, potentially resulting in a drastic reduction in crop output volume. This research introduces an advanced approach, building upon an existing project that utilizes the random forest algorithm. The refined project calculates the requisite fertilizer amount based on rainfall precipitation values, signifying a significant enhancement. Noteworthy is the comprehensive comparison conducted, assessing various ensemble learning methods, including XGBoost, AdaBoost, and LightGBM, against the established random forest algorithm. This meticulous comparison yielded a substantial increase in accuracy and a reduction in error rates, affirming the efficacy of the refined model.

To empower farmers with actionable insights, the proposed model employs a machine-learning algorithm—specifically, the random forest regression algorithm, enhanced with a k-fold cross-validation technique. This user-centric approach allows farmers to input two crucial parameters—crop type and location—enabling the model to predict the optimal amount of nutrients required and the most suitable timing for fertilizer application. The underlying technology utilizes Flask Python, a versatile web framework, ensuring access across all platforms and facilitating seamless sharing among users. This introduction encapsulates the pivotal role of agriculture in economic development, underscores the necessity of precise fertilization, and positions the improved model as a transformative tool to address this challenge. The subsequent sections will delve into the methodology, results, and implications of this enhanced model, showcasing its potential to empower farmers with accurate and actionable information for optimizing crop yield.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

A comprehensive examination of contemporary research in the field of agricultural technology and crop management reveals a myriad of approaches aimed at enhancing crop yield prediction, optimizing fertilizer usage, and improving soil health. This literature review synthesizes insights from nine research papers, providing a foundation for understanding the current landscape and challenges in the domain.

The critical issue of crop yield prediction and efficient fertilizer usage in the agrarian economy of India was addressed using various data mining techniques. The study integrated diverse attributes such as location, soil composition and nutrient values to create a precise model predicting crop yield. The proposed system not only predicts the yields but also offers fertilizer recommendations, contributing to increased agricultural productivity[1].

An algorithmic learning framework was introduced which emphasized the impact of climatic conditions and fertilizers on crop yield. Utilizing AdaBoost and Random Forest algorithms, the study predicts crop yields based on parameters like state, district, and soil conditions. The research further suggests fertilizer recommendations tailored to soil conditions, enhancing overall crop yield[2].

An enhanced genetic algorithm was proposed in one study which would offer insights into nutrient requirements for different crops. Time series data collected from sensors would be utilized in this system which focuses on soil fertility management through an evolutionary computation-based nutrient recommendation system[3].

Crop yields in India can be predicted using innovative regression techniques like kernel ridge and Stacking regression. Parameters such as state, district and season were utilized to forecast crop yields. The incorporation of advanced algorithms contributes to more accurate predictions ,aiding farmers in decision-making[4].

A systematic literature review was conducted on machine learning algorithms for crop yield prediction. The review highlighted the prevalence of Artificial Neural Networks, Convulational Neural Networks, Long Short-Term Memory and Deep Neural Networks. The study underscores the significance of features like temperature, rainfall and soil type in crop yield prediction models[5].

Algorithms like Random Forest Classifier, SVM and XGBoost can be integrated to forecast agricultural production. This could in turn help to resolve challenges faced by agriculture such as unpredictable weather conditions and soil fertility[6].

A comprehensive review of cropping systems' impact on soil health found that the inclusion of biological attributes such as microorganisms within soil contributed to the holistic understanding of soil health dynamics[7].

Manoj Kumar, Neelam Malyadri focus on the need for precise crop and fertilizer recommendations in India, a major agricultural producer. The study proposes a machine learning model that recommends suitable crops and fertilizers based on soil and weather conditions. This approach aims to diversify crops and optimize fertilizer use, contributing to increased profitability and reduced soil pollution[8].

Ishita Katiyar, Gaurav Sharma emphasize the importance of competent guidance for farmers in fertilizer use. The study proposes a new decision support system utilizing advancements in random forest algorithm. This system forecasts nutrient requirements based on time-series data, considering rainfall patterns and crop fertility, thereby enhancing soil fertility and reducing runoff potential[9].

III. PROPOSED METHODOLOGY

Prediction of optimal fertilizer amount for a given crop based on various input parameters

A. Model Selection and Comparative Analysis

In this study, we endeavoured to build a robust predictive model for determining the nutrient requirements of crops. Initial investigations involved the comparison of the widely used Random Forest model with a backpropagation model[1]. The results indicated a superior performance of the Random Forest model, prompting its selection as the primary algorithm for nutrient prediction. To ensure model robustness, we employed a k-fold cross-validation technique during Random Forest regression, leading to the identification of a model with acceptable accuracy.

B. Ensemble Learning Exploration

Recognizing the potential for further enhancement, we addressed a research gap by exploring other ensemble learning algorithms beyond Random Forest. Our experimentation revealed that AdaBoost exhibited significantly decreased performance compared to Random Forest. So we further refined the model using the XGBoost algorithm, resulting in enhanced predictive capabilities. However, the pinnacle of our findings was the implementation of the LightGBM model, demonstrating the highest accuracy, with nearly a 2 percent increase and a notably reduced error rate compared to other models.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Literature Gap and Integration of LightGBM

The literature review revealed that existing studies had employed a combination of regression, ensemble, and deep learning models for nutrient prediction. However, LightGBM, a relatively novel algorithm, had not been explored. Recognizing this gap, we introduced LightGBM into our methodology, capitalizing on its distinct advantages and a rich array of hyperparameters. This integration seeks to enhance our comprehension of nutrient prediction models while delving into the unexplored opportunities presented by LightGBM.

D. Algorithmic Workflow

As illustrated in Figure 1, our proposed algorithm initiates with user inputs, including the geographical location and crop type. Subsequently, the location information is processed through the Weather API, obtaining critical environmental characteristics such as temperature, moisture levels, and precipitation. Notably, if there is an indication of heavy rainfall, a precautionary message is displayed to the user. Otherwise, the nutrient prediction model proceeds, leveraging the insights gained from the selected ensemble learning model. This methodological approach synthesizes advancements in ensemble learning algorithms, addressing the identified research gaps and harnessing the capabilities of LightGBM for precise nutrient predictions in diverse agricultural contexts. The subsequent sections will delve into the details of data collection, feature selection, training methodologies, and model evaluation to provide a comprehensive understanding of our research methodology.

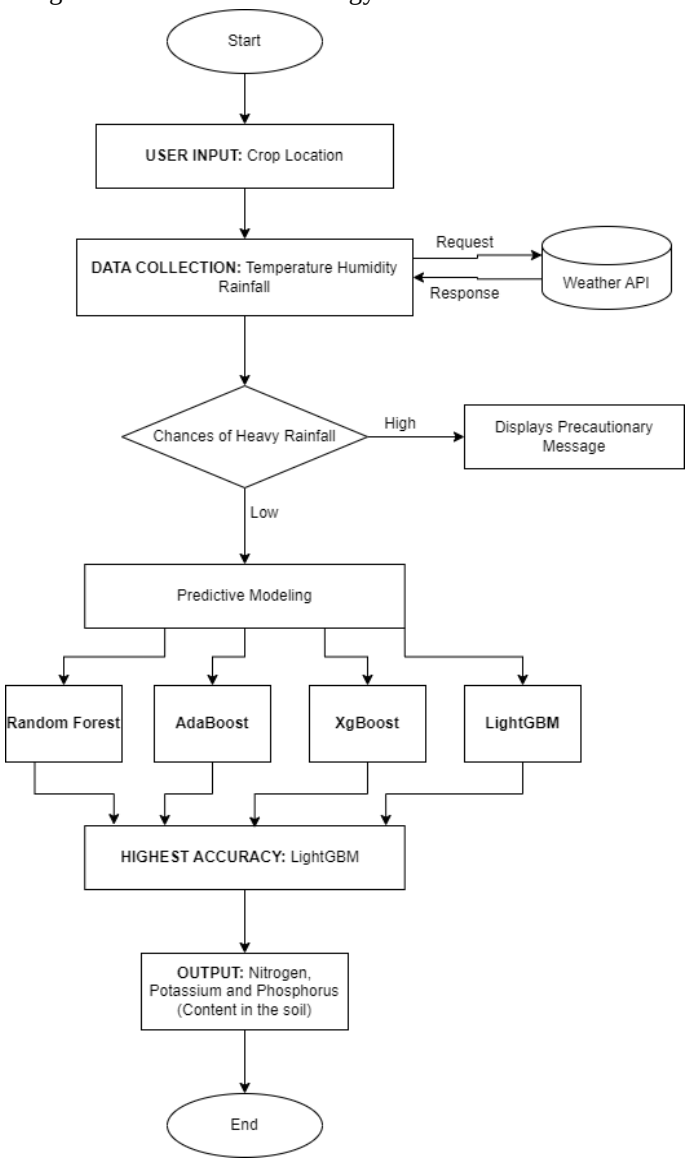


Figure 1 Proposed System Architecture

IV. DATASET

A. *Original Dataset*

The foundational dataset employed in this research encompasses a comprehensive set of fields crucial for nutrient prediction in agriculture. Key attributes included Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, rainfall, label, and information regarding the crops under consideration. Each entry in the dataset provides a snapshot of the environmental and soil conditions, as well as the corresponding nutrient levels and crop data.

B. *Data Cleaning and Operations:*

A meticulous data cleaning process was executed to ensure the integrity and quality of the dataset. Various operations were performed, including handling missing values, outlier detection and removal, and normalization of numerical features. This rigorous cleaning phase aimed to eliminate potential biases and inconsistencies in the data, fostering a robust foundation for subsequent analyses.

C. *Feature Reduction for Optimal Performance:*

In pursuit of optimal model performance, a strategic decision was taken to decrease the number of attributes belonging to the dataset. This reduction was informed by a comprehensive evaluation of feature importance and correlation analyses. The enhanced dataset, tailored for efficiency, ultimately comprised seven key fields: Crop, Temperature, Humidity, Rainfall, Label N, Label P, and Label K.

D. *Enhanced Dataset Fields*

1) *Crop:* Specifies the type of crop under consideration.
2) *Temperature:* Represents the temperature conditions recorded during the dataset entry.
3) *Humidity:* Indicates the humidity levels at the time of data collection.
4) *Rainfall:* Records the amount of rainfall corresponding to each dataset entry.
5) *Label N:* Designates the Nitrogen level associated with the entry.
6) *Label P:* Signifies the Phosphorus level recorded for the specific dataset entry.
7) *Label K:* Denotes the Potassium level corresponding to the entry.

This streamlined dataset not only optimizes computational efficiency but also retains the essential information required for accurate nutrient predictions. The later segments will delve into the techniques used for training and testing the predictive models using this refined dataset, providing a transparent and detailed account of the research process.

V. RESULT AND DISCUSSION

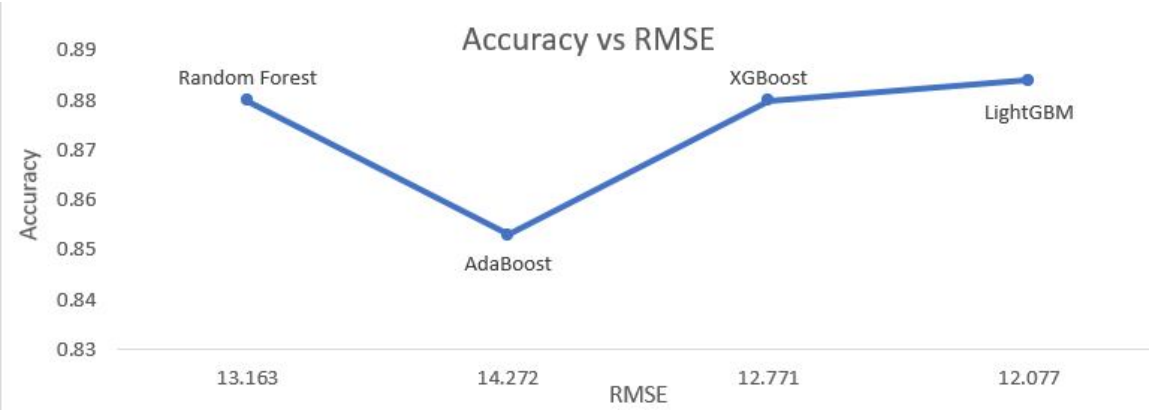


Figure 2 Accuracy Vs. RMSE

The implementation of advanced machine learning algorithms, specifically the random forest algorithm, in predicting and optimizing fertilizer use has yielded promising outcomes. The model was trained and tested on a comprehensive dataset, encompassing diverse crops, geographical locations, and time-series data on rainfall and soil fertility.

Algorithm	Training Accuracy	Overall Accuracy	RMSE	R2	MAE
Random Forest	0.88	0.88	13.163	0.8787	8.880
AdaBoost	0.857	0.853	14.272	0.92	9.621
XGBoost	0.887	0.88	12.771	0.9429	8.613
LightGBM	0.884	0.884	12.077	0.9490	8.148

TABLE 1 Performance Metrics of Various Machine Learning Algorithms

The comparison of ensemble learning methods, including XGBoost, AdaBoost, and LightGBM, against the random forest algorithm revealed a notable enhancement in accuracy and a reduction in error rates. The LightGBM model emerged as the most effective in predicting nutrient requirements for various crops based on rainfall patterns.

The tool's user interface, developed using Flask Python, provides a seamless experience for farmers, allowing them to input crop and location data. The model then generates precise recommendations for nutrient quantities and optimal timing for fertilizer application, contributing to improved crop yields. The integration of time-series data analysis and ensemble learning not only enhances predictive accuracy but also addresses the dynamic nature of agricultural systems. The model's ability to adapt to changing climatic conditions and evolving soil fertility over time positions it as a valuable tool for sustainable farming practices.

The discussion extends to the broader implications of the research, emphasizing the potential for widespread adoption of data-driven approaches in agriculture. By minimizing fertilizer waste and runoff, the model contributes to environmental sustainability while empowering farmers with actionable insights. The research conducted a thorough analysis of ensemble learning algorithms for optimizing nutrient predictions in agriculture. Random Forest demonstrated a training accuracy of 0.88 but exhibited a high error rate with a Root Mean Squared Error (RMSE) value of 13.163 and R-squared value of 0.8787 and MAE value of 8.880. Adaboost, with a training accuracy of 0.857 and an RMSE of 14.272, fell slightly short of Random Forest. It had R-squared value of 0.92 and MAE value of 9.621 XGBoost emerged as a superior alternative, surpassing both Random Forest and Adaboost with a lower RMSE value of 12.771, coupled with a training accuracy of 0.887 , an overall accuracy of 0.88 , R-squared value of 0.9429 and MAE value of 8.613. The standout performer, however, was LightGBM, boasting the lowest RMSE value of 12.077. With a training model accuracy of 0.884 aligning with an overall accuracy of 0.884, R-squared value of 0.9490 and MAE value of 8.148, LightGBM demonstrated unparalleled precision in nutrient predictions.

VI. CONCLUSION

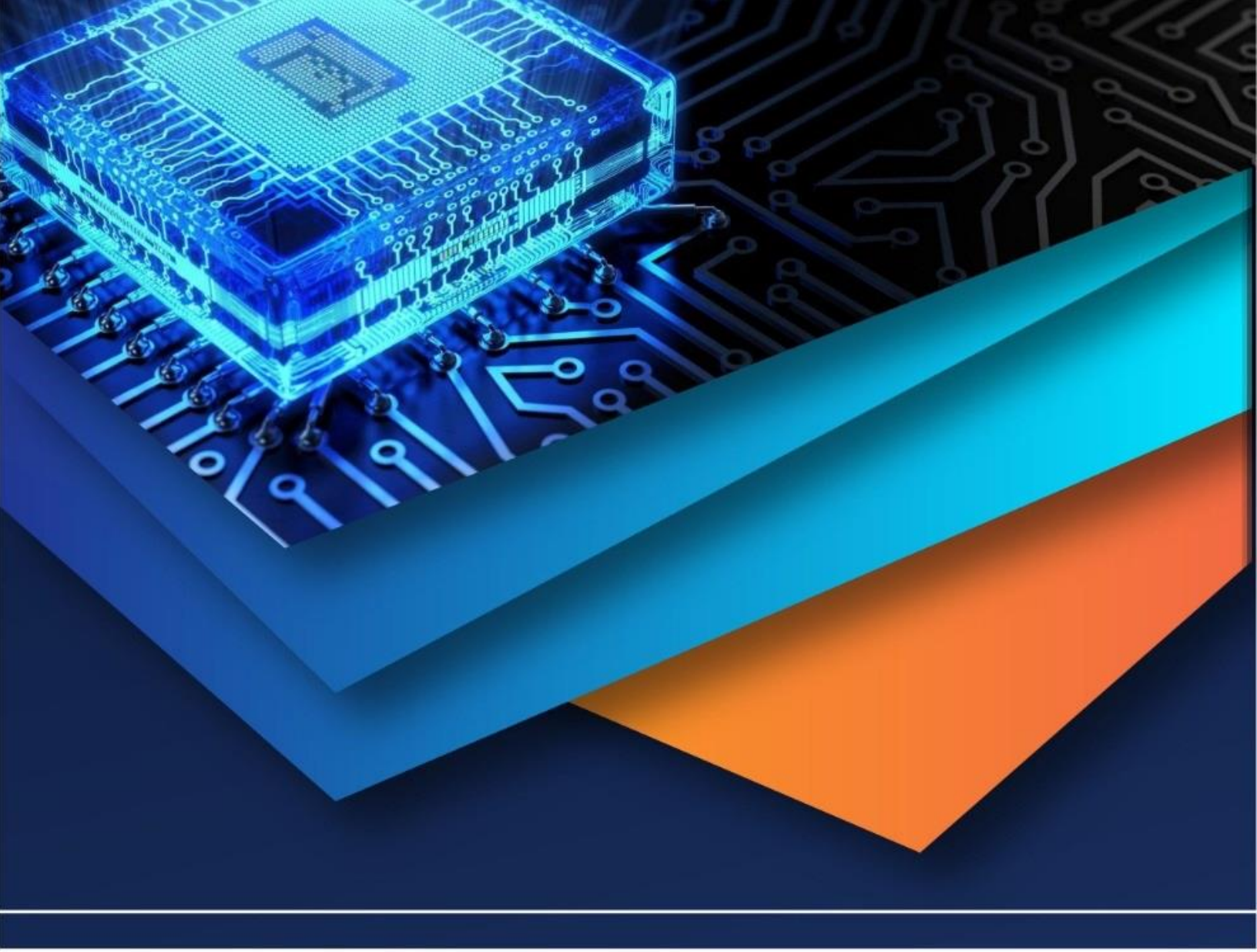
In summary, while Random Forest, Adaboost, and XGBoost offered varying levels of competence, LightGBM emerged as the optimal choice for accurate and efficient nutrient predictions in diverse agricultural contexts. This research contributes to the understanding of ensemble learning applications in agriculture and emphasizes the importance of adopting advanced models like LightGBM for sustainable and precise farming practices. As agriculture faces evolving environmental challenges, leveraging innovative predictive models becomes imperative for ensuring security of food across the globe and the promotion of sustainable agricultural practices.

REFERENCES

[1] S.Bhanumathi, M.Vineeth and N.Rohit, "Crop Yield Prediction and Efficient use of Fertilizers", International Conference on Communication and Signal Processing, April 4-6, 2019, India
[2] Jeevaganesh R , Harish D , Priya B , "A Machine Learning-based Approach for Crop Yield Prediction and Fertilizer Recommendation", Proceedings of the Sixth International Conference on Trends in Electronics and Informatics ICOEI 2022
[3] Usman Ahmed , Jerry Chun-Wei Lin , Gautam Srivastava , Youcef Djenouri ,"A nutrient recommendation system for soil fertilization based on evolutionary computation", Computers and Electronics in Agriculture Volume 189, October 2021, 106407
[4] Potnuru Sai Nishant,Pinapa Sai Venkat, Bollu Lakshmi Avinash,B.Jabber,"Crop yield prediction based on Indian agriculture using Machine learning", 2020 International Conference for Emerging Technology(INCET) Belgaum,India Jun 5-7,2020
[5] Thomas van Klompenburga , Ayalew Kassahuna , Cagatay Catalb,"Crop yield prediction using machine learning: A systematic literature review", Computers and Electronics in Agriculture Volume 177, October 2020, 105709
[6] Avi Ajmera1 , Mudit Bhandari2 , Harshit Kumar Jain3 , Supriya Agarwal4, "Crop, Fertilizer, Irrigation Recommendation using Machine Learning Techniques", International Journal for Research in Applied Science and Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue XII Dec 2022
[7] Tony Yang a, Kadambot H.M. Siddique b , Kui Liu,"Cropping systems in agriculture and their impact on soil health-A review", Global Ecology and Conservation Volume 23, September 2020, e01118
[8] Manoj Kumar, Neelam Malyadri,Srikanth MS , Dr Ananda Babu J, "A machine learning model for crop and fertilizer recommendation", Nat Volatiles and Essent.Oils,2021:8(5)

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[9] Ishita Katiyar1, Gaurav Sharma2, Sumukha Hegde3, Arpit Chakraborty4,Manash Sarkar5 ,"An Efficient and Intelligent Decision Making for Eco-Fertilization", 2023 IJCRT — Volume 11, Issue 5 May 2023 — ISSN: 2320-2882
[10] Ramesh A. Medar, "A survey on data mining techniques for crop yield prediction," International Journal of Advance in Computer Science and Management Studies, vol. 2, no. 9, 2014, ISSN: 2231-7782.
[11] Pooja M. C. et al., "Implementation of Crop Yield Forecasting Using Data Mining," International Research Journal of Engineering and Tech nology (IRJET), 2018, p-ISSN: 2395-0072.
[12] Jig Han Jeong et al., "Random Forests for Global and Regional Crop Yield Predictions," PLOS ONE, DOI: 10.1371/journal.pone.0156571, June 3, 2016.
[13] Jharna Majumdar, Sneha Naraseeyappa, and Shilpa Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," Springer Journal, 2017.
[14] S. Kanaga Subba Raju et al., "Demand based crop recommender system for farmers," in International Conference on Technological Innovations in ICT for Agriculture and Rural Development, 2017.
[15] Shriya Sahu et al., "An Efficient Analysis of Crop Yield Prediction Using Hadoop framework based on random Forest approach," in International Conference on Computing, Communication and Automation, 2017.
[16] E. Manjula and S. Djodkitachoumy, "International journal of computation Intelligence and informatics," vol. 6, no. 4, March 2017



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)