International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 653-657
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0090

Predictive Analytics for Soil Productivity Using Machine Learning

Mrs. K. Leelarani[1], Mohana Prakash G[2], Balaji J[3], Arun Kumar S[4]
[1]Assistant Professor, Department of CSE, Kamaraj College of Engineering and Technology, Virudhunagar, India.
[2,3,4]Student, Department of CSE, Kamaraj College of Engineering and Technology, Virudhunagar, India.
Emails: *leelaranicsecse@kamarajengg.edu.in[1], 21ucs104@kamarajengg.edu.in[2], 21ucs064@kamarajengg.edu.in[3], 21ucs126@kamarajengg.edu.in[4]*

Abstract

A major factor in agricultural production is soil productivity, which is impacted by soil fertility as well as the compatibility of the crops cultivated there. In order to improve agricultural efficiency and sustainability, farmers must make educated judgments on crop selection and fertilizer use, which requires an accurate estimate of soil productivity. This study offers a predictive analytics method that uses machine learning and K-Means clustering and crop recommendation models to evaluate soil productivity. The K-Means clustering algorithm divides soil into three classes: Fertile, Highly Fertile, and Less Fertile, depending on the amount of nitrogen (N), phosphorus (P), and potassium (K) it contains. A Random Forest Regression model is used to forecast the best crop because soil productivity is influenced by both crop selection and fertility. In order to suggest crops that maximize production potential, this model examines a variety of soil attributes and environmental factors. Through precise soil fertility prediction and crop recommendation, this study offers farmers useful information that helps them maximize soil utilization, increase crop yields, and engage in sustainable agriculture. In the end, this data-driven strategy benefits both agricultural output and environmental preservation by enhancing farm productivity and resource management.

Keywords: *Random Forest Regression, K Means Clustering, Soil Fertility, Classification, Crop Recommendation.*

1. Introduction

Agricultural productivity is heavily dependent on soil fertility, which determines the ability of soil to supply essential nutrients for plant growth. Soil fertility is influenced by key macronutrients, particularly Nitrogen (N), Phosphorus (P), and Potassium (K), along with factors such as pH levels, organic carbon content, and micronutrient availability. Maintaining an optimal NPK ratio is crucial for maximizing crop yield, there is a study done on NPK ratio how it affects the plant growth. The deficiencies or imbalances in these nutrients can lead to reduced plant growth and lower productivity. Selecting the right crop for a given soil composition is an essential aspect of sustainable agriculture. With advancements in machine learning, predictive analytics can now help in crop recommendation and soil fertility classification, enabling data-driven decision-making in agriculture. In this approach to soil productivity, we propose a machine learning-based approach that uses Random Forest Regression to predict the most suitable crop based on Nitrogen, Potassium and Phosphorous values, temperature, humidity, pH, and rainfall. The Random Forest algorithm is chosen for its ability to handle non-linear relationships and its high predictive accuracy in complex datasets. In addition to crop recommendation, K-Means Clustering is employed to classify soil into three distinct categories: Less Fertile, Fertile, and Highly Fertile. This classification is based on essential soil attributes, including macronutrient and micronutrient levels. By identifying soil fertility levels, farmers can take necessary corrective actions, such as adjusting fertilization strategies or selecting appropriate crops, to maximize yield. To be used in real-world , a web-based platform is developed where users can input soil characteristics and receive real-time crop recommendations and soil fertility classifications. This system enables precision agriculture by enhancing resource optimization, improving yield predictions, and promoting sustainable farming

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 653-657
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0090

practices. The integration of predictive modelling and clustering techniques in this study presents an efficient approach to data-driven agricultural decision-making, contributing to sustainable agriculture and plant growth [1-3].

2. Methodology

The development of a web application that predicts the soil fertility and crop suitable for the soil involves a systematic approach combining machine learning technique for prediction and classification. The methodology is structured into the following key stages:

- **Data Collection and Preprocessing:** The foundation of this study relies on two datasets: one for crop recommendation and another for soil fertility classification. The crop recommendation dataset includes key soil attributes such as Nitrogen (N), Phosphorus (P), and Potassium (K) levels, along with environmental factors like temperature, humidity, pH, and rainfall. The soil fertility dataset consists of essential macronutrients and micronutrients, including Sulfur (S), Zinc (Zn), Iron (Fe), Copper (Cu), Manganese (Mn), and Boron (B), which are crucial for determining soil productivity. Before training the models, preprocessing steps were applied to ensure data consistency and quality. This included handling missing values, normalizing numerical features, and encoding categorical data where necessary. Feature scaling was particularly important for algorithms like K-Means clustering, ensuring that all attributes contributed equally to the classification.
- **Crop Recommendation using Random Forest Regression:** To predict the most suitable crop for a given set of soil and environmental conditions, we employed Random Forest Regression. This algorithm is an ensemble learning method that builds multiple decision trees and combines their outputs to enhance prediction accuracy and reduce overfitting. During model training, hyperparameters such as the number of decision trees, maximum depth, and minimum samples per leaf were fine-tuned to optimize performance. The model was trained on historical data that maps soil conditions to optimal crop choices, learning the complex relationships between nutrient levels and agricultural productivity. The trained model was evaluated using metrics like Mean Squared Error (MSE) and R-squared (R^2) scores, ensuring that it generalizes well to unseen data. Once validated, the model was integrated into the web application for real-time crop recommendations.
- **Soil Fertility Classification:** For classifying soil fertility, we utilized K-Means Clustering, an unsupervised learning algorithm that groups similar data points into clusters based on their feature similarities. The primary objective was to classify soil into three categories: Less Fertile, Fertile, Highly Fertile. To determine the optimal number of clusters (K), we used the Elbow Method, which analyzes the variance within clusters to identify the most appropriate value for K. The model was then trained on soil nutrient data, allowing it to automatically categorize soil samples based on their fertility levels. Once clustering was completed, each soil sample was labelled according to its assigned cluster, providing an easy-to-interpret classification for farmers and agricultural experts [4-7].
- **Web Implementation:** The web application, built using Flask, provides an intuitive interface for users to input soil parameters such as N, P, K, pH, temperature, and humidity. The backend processes these inputs using trained Random Forest Regression for crop recommendation and K-Means Clustering for soil fertility classification. Predictions are displayed along with relevant crop images, retrieved dynamically. The system ensures seamless interaction through API calls, enabling real-time results. This web-based approach enhances precision agriculture by offering data-driven insights for optimal crop selection and soil fertility assessment.
- **Testing and Deployment:** The system was thoroughly tested to ensure accurate predictions and smooth functionality. The Random Forest Regression model for crop recommendation was

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 653-657
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0090

evaluated using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to measure its accuracy. Similarly, the K-Means Clustering model for soil fertility classification was assessed using the Silhouette Score to validate its clustering performance. After testing, the web application was deployed on a cloud platform such as AWS, Heroku, or PythonAnywhere, making it easily accessible to users. The models were saved using Joblib, allowing real-time predictions when users input soil parameters. This deployment ensures that farmers and agricultural experts can quickly get reliable crop recommendations and soil fertility classifications, ultimately helping improve productivity, Figure 1.

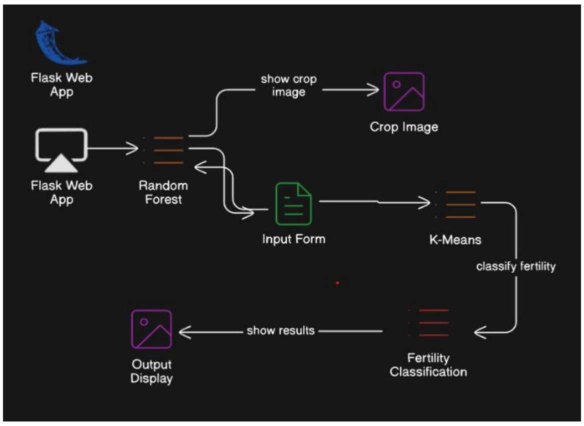

Figure 1 Methodology for Predictive Analytics for Soil Productivity Using Machine Learning

3. Experimental Testing

- **Experimental Setup:** The experimental setup involved preprocessing soil and environmental data to train machine learning models for crop recommendation and soil fertility classification. The Random Forest Regression model was trained using key soil nutrients (N, P, K) and environmental factors, while K-Means clustering classified soil fertility levels. The models were integrated into a Flask-based web application, enabling real-time predictions. Evaluation methods like the Elbow Method and Silhouette Score optimized clustering, and the system was tested on various soil samples to ensure accuracy and reliability.
- **Model Evaluation:** Model evaluation was conducted to assess the accuracy and reliability of both the Random Forest Regression model for crop recommendation and the K-Means clustering model for soil fertility classification. The Random Forest model was evaluated using metrics such as Mean Squared Error (MSE) and R^2 score, ensuring precise crop predictions. The K-Means model's clustering performance was analyzed using the Elbow Method and Silhouette Score to determine the optimal number of clusters. The system was tested on real-world soil data, validating its effectiveness in providing accurate recommendations for improved agricultural productivity [8-10].

3.1 Comparison of Regression Techniques in Crop Recommendation

The Figure 2 model is trained and the output of the regression techniques are compared.

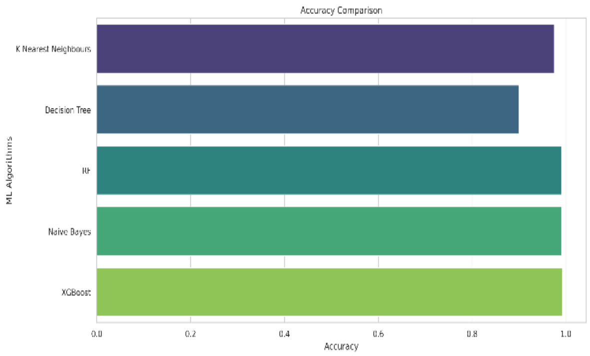

Figure 2 This Chart Compares the Model Accuracy with Different Regression Techniques

- **Key Results:** The experimental technique involved training and testing the models on real-world soil and crop datasets to ensure accurate predictions. The Random Forest Regression model effectively recommends crops based on soil properties, while the K-Means clustering model classifies soil fertility into three distinct categories. The system provides real-time predictions through a web application, allowing users to input soil parameters and receive both crop suggestions and fertility classification. This approach enhances decision-making for farmers by offering data-driven insights, ultimately

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 653-657
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0090

improving soil productivity and crop yield.

4. Result and Discussion

The result of the proposed sytem are given in orderly manner and explained.

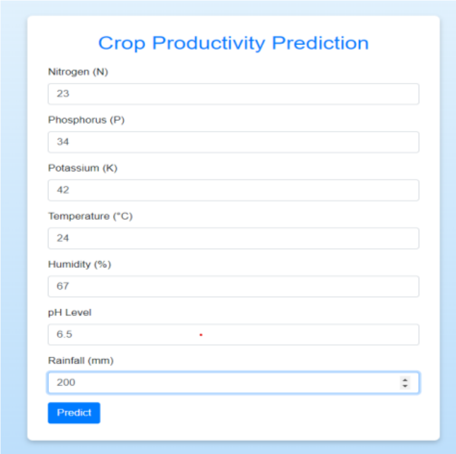

Figure 3 This Is the Page in Web Application Where the Input for The Model is Given

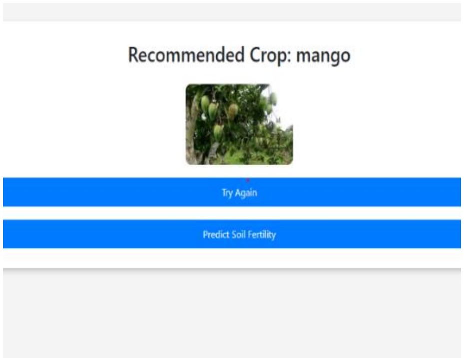

Figure 4 This is the Result Page That Shows the Predicted Crop and Its Image. From Here the Application Can Be Navigated to Find the Soil Fertility

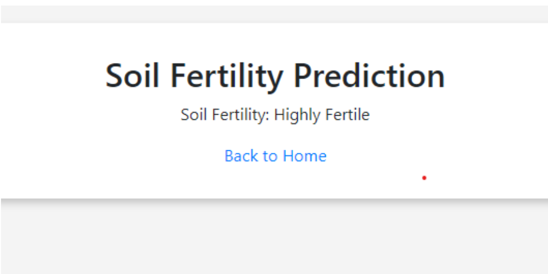

Figure 5 The Page That Shows the Fertility of The Soil Using K Means Algorithm and Classifies as Less Fertile, Fertile and Highly Fertile

4.1 Discussion

The results from Figure 3, Figure 4 & Figure 5, our predictive system highlight the effectiveness of machine learning in agricultural decision-making. The Random Forest Regression model demonstrated high accuracy in recommending suitable crops based on soil attributes, ensuring that farmers can make informed choices to maximize yield. Similarly, K-Means clustering successfully categorized soil fertility, providing a clear understanding of soil health. One of the key advantages of this approach is its ability to analyze multiple soil factors simultaneously, which traditional methods often overlook. By leveraging machine learning, we can move beyond generalized recommendations and provide personalized insights based on specific soil conditions. However, the accuracy of predictions depends on the quality and variety of data used for training. Regional variations in soil composition and changing environmental factors may require periodic model updates to maintain precision. Additionally, integrating real-time sensor data could further enhance the model's performance, making it even more adaptive to dynamic agricultural conditions. Overall, this system offers a practical and scalable solution for farmers and agricultural experts, bridging the gap between data science and real-world farming practices.

Conclusion

This study highlights how machine learning can improve agricultural productivity by recommending suitable crops and classifying soil fertility. Using Random Forest Regression for crop prediction and K-Means clustering for soil classification, our system provides data-driven insights to help farmers make better decisions. In the future, this approach can be enhanced by integrating real-time soil data using IoT-enabled sensors, allowing continuous monitoring and adaptive recommendations. This would pave the way for smart farming solutions, optimizing resource use and ensuring sustainable agriculture.

Reference

[1]. Amalu, U.C., Isong, I.A. Status and spatial variability of soil properties in relation to fertilizer placement for intercrops in an oil palm plantation in Calabar, Nigeria. Niger. J.

Crop Sci. 2018, 5, 58– 72.
[2]. Prakash Kanade, Fortune David, Sunay Kanade, Convolutional Neural Networks (CNN) based Eye-Gaze Tracking System using Machine Learning Algorithm, European Journal of Electrical Engineering and Computer Science, Volume 5, Issue 2, pp 36-40, 2021.
[3]. D. Zhang, P. Gao, Y. Zhou, Y. Wang and G. Zhou, "An experimental and numerical investigation on temperature profile of underground soil in the process of heat storage," Renewable Energy, vol. 148, pp. 1–21, 2020.
[4]. Bhattacharya, B., & Solomatine, D. P. Machine learning in soil classification. Neural Networks, Vol 19, pp.186-195, 2006.
[5]. T. Amsalu and W. Cornelis, "The potential of termite mound spreading for soil fertility management under low input subsistence agriculture," Journal of Agriculture, vol. 11, pp. 1–16, 2021.
[6]. Armstrong, L. J., Paul, M., Vishwakarma, S. K., & Verma, A. (2015, December).Analysis of Soil Behavior and Prediction of Crop Yield Using Data Mining Approach. In Computational Intelligence and Communication Networks (CICN), 2015 International Conference on (pp. 766-771) IEEE.
[7]. M. Koch, M. Naumann, E. Pawelzik, A. Gransee, and H. Thiel, ''The importance of nutrient management for potato production part I: Plant nutrition and yield,'' Potato Res., vol. 63, no. 1, pp. 97–119, Mar. 2020.
[8]. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, ''From local explanations to global understanding with explainable AI for trees,'' Nature Mach. Intell., vol. 2, no. 1, pp. 56–67, Jan. 2020.
[9]. G. A. Helfer, J. L. V. Barbosa, R. dos Santos, and A. B. da Costa, ''A computational model for soil fertility prediction in ubiquitous agriculture,'' Comput. Electron. Agric., vol. 175, Aug. 2020, Art. no. 105602.
[10]. Senol, H.; Alaboz, P.; Demir, S.; Dengiz, O. Computational intelligence applied to soil quality index using GIS and geostatistical approaches in semiarid ecosystem. Arab. J. Geosci. 2020, 13, 1–20.