International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Machine Learning Approach for Yield Estimation and Crop Prediction

S.S.D.K Mahalakshmi[1], K.V.Navya Sree[2], M.Chandra Lekha[3], M. Jyothsna[4], N.Gowtham Kumar[5]

[1] *S.S.D.K Mahalakshmi, M. Tech, Assistant Professor, Department of Computer Science and Engineering, Lendi Institute of Engineering and Technology (Autonomous), Andhra Pradesh, India.*
[2,3,4,5] *B.Tech Students, Department of Computer Science and Engineering, Lendi Institute of Engineering and Technology (Autonomous), Andhra Pradesh, India.*

Abstract - Optimal crop recommendations and yield prediction are essential for improving farm productivity and food security. Conventional approaches to the prediction of crop yields are usually limited by environmental variability. The present work proposes a strong machine learning-based framework aiming to support farmers in the process of choosing the best crops and predicting yield potential more accurately. The suggested system combines three supervised learning techniques—Decision Tree, Random Forest, and XGBoost—to examine critical environmental and soil parameters like nitrogen (N), phosphorus (P), potassium (K), soil pH, rain, humidity, and temperature. These attributes enable accurate crop recommendations under given environmental conditions. For yield prediction, past agricultural information, including parameters like state, district, crop type, season, and cultivated area, is utilized. Extensive model training and evaluation demonstrate that Random Forest performs exceptionally well compared to other approaches, with 96%. Accuracy ensures farmers are provided with the correct crop yields and recommendations. To ensure maximum accessibility, the solution is implemented as a web application using Flask, providing an easy-to-use interface for farmers with minimal technical knowledge to enter farm parameters and get actionable recommendations. This system not only facilitates better planning and resource optimization but also aids in reducing risks from uncertain weather patterns and volatile market conditions.

Key Words: Crop Prediction, Yield Estimation, Random Forest, Flask Deployment, Precision Agriculture

1.INTRODUCTION

Agriculture is one of the core sectors that provide food security and contribute to the economic development of most nations. Traditional farming, nonetheless, tends to rely on the farmer's personal experience and intuition, which can sometimes not necessarily result in the best possible decisions on crop selection and estimation of yield. With increasing pressures like climate change, resource constraints, and fluctuating market prices, there is a greater need for more precise, data-based methods of agricultural planning. Machine learning has been a valuable tool for improving decision-making in agriculture by using past data to identify patterns and make predictions. Systems based on ML can assist farmers in identifying the most appropriate crops for their areas and predicting crop yields with higher accuracy, thereby producing improved productivity and profitability. Crop recommendation models evaluate parameters like soil health, rain patterns, and temperature levels to recommend the best crops, and yield estimation models help farmers plan harvests, store the produce, and negotiate good prices. This study suggests an all-encompassing ML-based solution to tackle two key areas of contemporary agriculture: yield estimation and crop recommendation. The framework employs three established supervised machine learning algorithms—Random Forest, Decision Tree, and XGBoost—to capture complex patterns among agricultural features and target variables. Pre-processed historical datasets such as soil, weather, and crop yields are utilized to train the models to predict with high accuracy. The main purpose of this project is to forecast crop yield and suggest appropriate crops through machine learning algorithms. For model training, a dataset consisting of agricultural data like state, district, season, crop, and area grown is utilized. With the use of a Random Forest-based model, the system can forecast the expected crop production or suggest the most suitable crop for a region and season. The solution is meant to help farmers and agricultural planners with data-informed decisions to help increase the yield of crops and optimize land use. The system guarantees optimal crop choice and precise yield estimation, enabling farmers to make informed decisions to reduce losses in unfavorable conditions, enhancing planning, productivity, and overall profitability in agriculture.

2. LITERATURE REVIEW

[1] Modern agriculture is a primary sector that may benefit greatly from the use of Information Technology. In the present Scenario farmer collects more than just agricultural data. The producers may benefit from the favorable soil and crop yield characteristics. In the past, farmers' years of experience with a certain yield, crop, and climate were used to make these predictions using the KNN algorithm.

[2] This paper aims to examine the characteristics of agricultural information, data mining, and analysis of the user query information in the agricultural information to design a data mining and statistical model on precision agriculture of time series representation and time series measurement of the ESU algorithm.

[3] The factors of palm oil crop yield prediction to use machine learning models, linear regression (LR) and multiple linear regression (MLR), multivariate adaptive regression splines (MARS), k-nearest neighbors (K-NN), support vector machine (SVM) and support vector regression (SVR), decision tree (DT), random forest (RF), extremely randomized trees (extra tree)

(ERT), artificial neural network (ANN), deep neural network (DNN), convolution neural network (CNN) and long short-term memory (LSTM).

[4] The Soil is an important dynamic resource that supports plant life. Soil data analysis can provide information that is important for maximizing nutrient use efficiency and agricultural productivity of Prototype-based clustering algorithms are most commonly used, including FCM, Gustafson-Kessel, SOM, EM, and other fuzzy clustering methods. Especially, FCM is dominantly implemented in the applications of agricultural production management, zoning, and soil-land classification.

[5] Agricultural production is mostly affected by environmental factors. Weather influences crop growth and development, causing large intra-seasonal yield variability. Nowadays, a growing number of applications of machine learning techniques are being used in agriculture. This is an advanced research field and is expected to grow in the future. The integration of computer science with agriculture helps in forecasting crops to Some machine learning techniques are Nonlinear regression for forecasting corn yields, Markov chain approach for forecasting cotton yields, linear regression for estimating grain yield of maturing rice, and FINKNN: a fuzzy interval number k-nearest neighbor classifier for prediction of sugar production.

[6] This model integrates various environmental data obtained from sugar mill surveys and government agencies with the analysis of aerial images of sugarcane fields obtained with drones. The drone images used to identify Defective cane can result from adverse weather or other cultivation issues. The model utilizes data mining to recognize and classify the dataset from the sugarcane field.

[7] Soil is an important component of agriculture to predict soil to used the data mining classification techniques, are k-Nearest-Neighbor (k-NN), Random-Forest (RF), Decision Tree (DT), and Naïve-Bayes (NB) to used all these algorithms in weak tools.

[8] Crop portraits have a development perspective it will play an important role in today's digital agriculture of crops, pesticides and pests and diseases to measure Dv vertex of the construct to based on a graph database to successfully predict to root rot diseases.

[9] This paper provides a detailed approach for farmers for crop selection based on the soil requirements to boost the productivity of the crop using data mining techniques, are classification and regression trees (CART) Algorithm and random forest

[10] This system to various cities of Maharashtra state focused on on predicting the yield of the crop by using a different machine learning algorithm. The application of (Smart Farm developed in this research helps users to predict the crop yield using different climatic parameters, the Random Forest algorithm, which we decided to use to train our model to give high accuracy and best prediction.

[11] This paper study of critical global issue of food security of climate change on agricultural land sustainability our research is to predict the risks associated with land suitability degradation and changes in irrigation patterns, directly impacting food security using machine learning techniques to use LSTM model to analyze the impact of climate change on agricultural land suitability for food security.

[12]This paper study to optimize and improve the quality of agricultural productivity through early detection and accurate classification of pests and diseases in agricultural sector to use machine learning techniques of logistic regression, extreme gradient boosting (XGBoost), and convolutional neural network (CNN) to analyze spectral data and classify the presence of pests and diseases in satellite images to significant breakthroughs in the agricultural sector, helping to improve the efficiency, sustainability, and quality of crop production.

3. PROPOSED METHODOLOGY

Over the last several decades, the digital information system has grown more ambiguous. The efficiency of the modern agriculture sector is more beneficial to greatly benefited from the use of information technology. Our aim of this project is to develop a yield estimation and crop recommendation system using machine learning models. The machine learning model accurately predicts crop production based on real-world agricultural data, helping farmers to make better decisions. The proposed system of our project is mainly focused on crop recommendation and yield estimation. Two modules are developed using machine learning modules to designed for assist farmers in making data-driven decisions regarding optimal crop selection and expected yield output in Figure 1 shows the proposed system architecture.

Figure 1: System architecture of proposed methodology.

A farmer who is curious about his harvest and wants to find out as soon as possible in the growing season, how much yield estimation to a particular crop. Our system design has two modules: a Random Forest Classification Model for crop recommendation and another module is Random Forest Regression Model for yield prediction. The datasets used for training these models include essential agricultural parameters such as soil type, pH, temperature, humidity, rainfall, and other location-specific environmental conditions, sourced from agricultural repositories. This data flows through two independent pipelines based on the user's intent. One of the pipelines is the crop recommendation pipeline, the input features are pre-processed—missing values handled, categorical variables encoded, and numerical values normalized—to use the Random Forest Classification model. The yield estimation pipeline of input data uses performed pre-

processing technique to train a random forest regression model to predict the expected crop yield based on historical yield records, weather data, and soil characteristics.

3.1 Accuracy of crop recommendation

The performance of classification models for crop recommendation is evaluated using three machine learning models: Decision Tree, Random Forest, and Gradient Boosting. The model analyzes the patterns and relationships in the dataset to suggest the most suitable crop for the given conditions based on their training model their accuracy. The Random Forest classifier achieved the highest accuracy of approximately 96.6%.

The Decision Tree model, though simpler and faster, achieved a lower accuracy of around 92%, suggesting it may be more prone to overfitting or missing complex feature interactions. And the Gradient Boosting with about 96.2%. These results highlight the effectiveness of ensemble methods, of Random Forest, in generating accurate crop recommendations, reinforcing its suitability for practical agricultural applications, for useful to farmers.

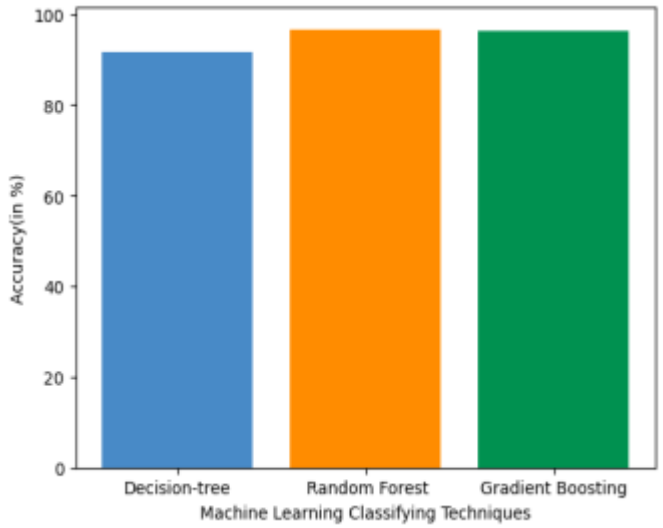

Figure 2: Accuracy Comparison Graph of crop recommendation

3.2 Crop Yield Estimation Accuracy:

The performance of regression models for crop yield estimation is the three machine learning techniques: Decision Tree, Random Forest, and XG Boosting. The models were assessed using the R-squared (R^2) metric, which measures how well the predicted values match the actual values and predicts the amount of yield based on given conditions, on their training model, of their accuracy. The Random Forest model outperformed the others with an R^2 score is 0.96.
The Decision Tree model, although simpler and more interpretable, lagged significantly with an R^2 score of approximately 0.44, reflecting its limitations in handling feature interactions and overfitting. And. XGBoost followed with an R^2 score around 0.88. These results highlight the robustness and accuracy of ensemble methods like the Random Forest model, which predicts the amount of yield that can be achieved by the farmer. It is helpful their cultivate strategies more effectively, optimize resource utilization, and increase productivity.

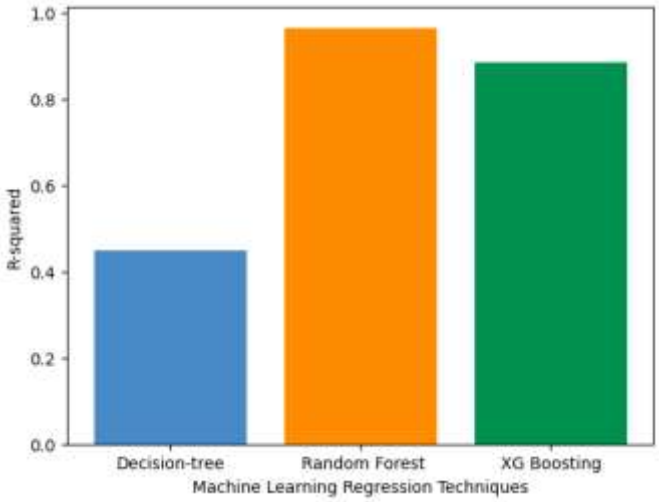

Figure 2: Accuracy Comparison Graph of crop yield estimation

4. RESULTS & DISCUSSION

In this study, our project aims to explore the most common machine learning models used in the agriculture sector. The machine learning techniques are broken down into 2 modules: Classification is used for crop recommendation, and Regression is used for crop yield estimation. Classification techniques are developed to categorize unknown samples with the help of identified samples to recommend a crop for farmers. Classification methods like Random Forest, Decision Tree, and Gradient Boosting. The results are shown in Figure.

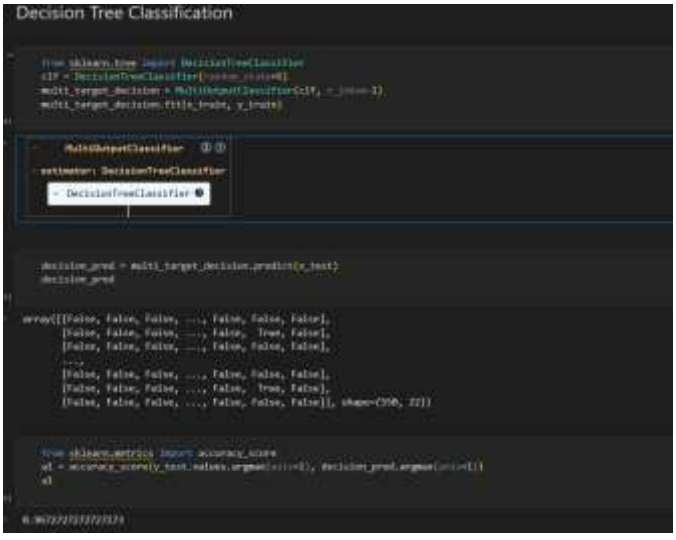

Figure 4: Decision Tree Classification for Crop Recommendation

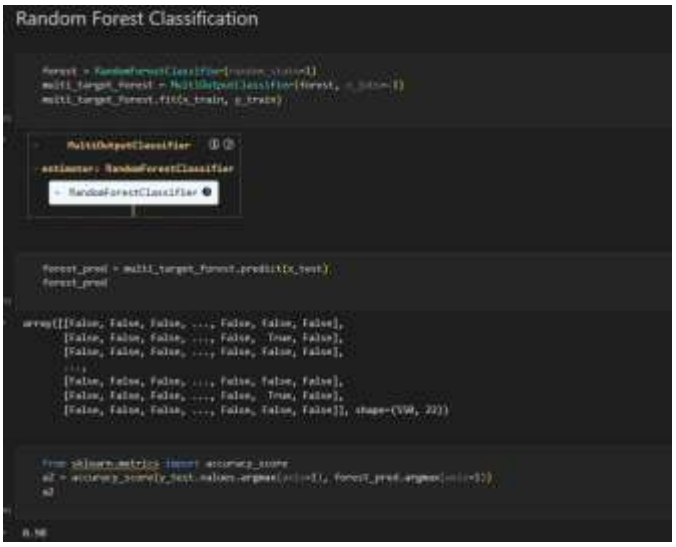

Figure 4.1 Random Forest Classification for Crop Recommendation

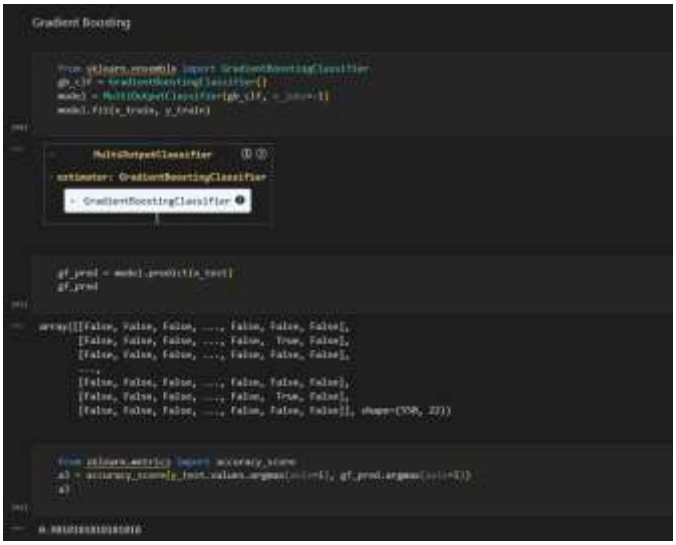

Figure 4.2 Gradient Boosting Classification for Crop Recommendation.

The regression techniques are developed for the purpose of crop yield estimation with the help of identifying to predict the amount of yield that a farmer can achieve. So, the Regression models are Random Forest regression, Decision Tree regression, and XG Boosting regression. Based on these Models, to measure the R squared values. The results are shown in the figure.

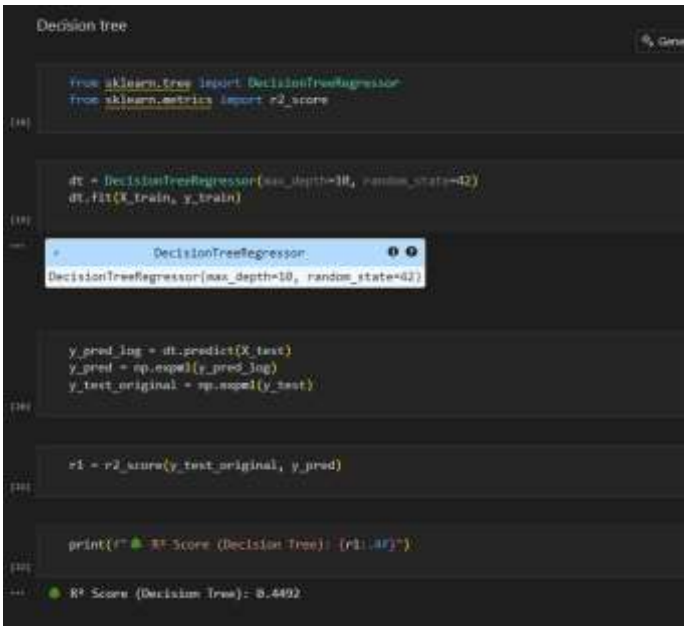

Figure 4.3 Decision Tree Regression for Crop Yield Estimation

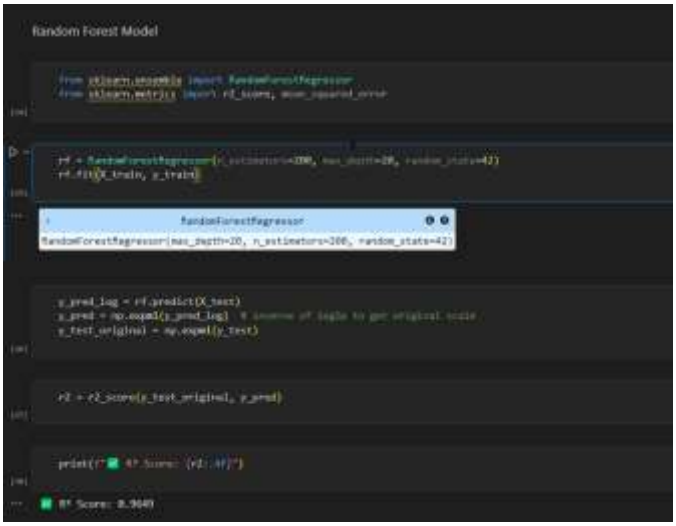

Figure 4.4 Random Forest Regression for Crop Yield Estimation

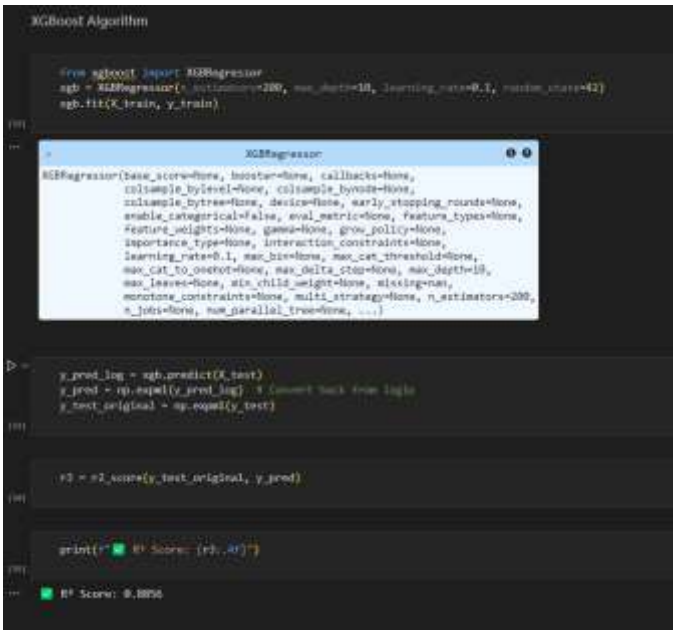

Figure 4.5 XG Boosting Regression for Crop Yield Estimation

The entire system is deployed using Flask, making it accessible as a web-based application. It consists developed two modules of crop recommendation and crop yield estimation. It provides real-time time accurate crop planning and productivity forecasting. The tool helpful to farmers to plan their cultivation strategies more effectively, optimize resource utilization, increase productivity, and mitigate risks associated with uncertain weather and market conditions. The results are shown in figures.

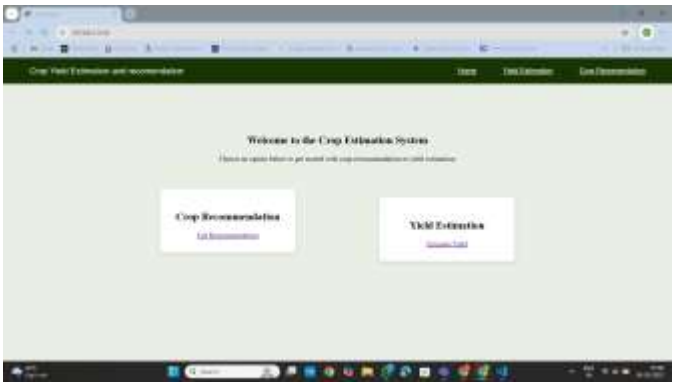

Figure 4.6 User Interface

Figure 4.7 Crop Recommendation Input Interface

Figure 4.8 Crop Recommendation Output Interface

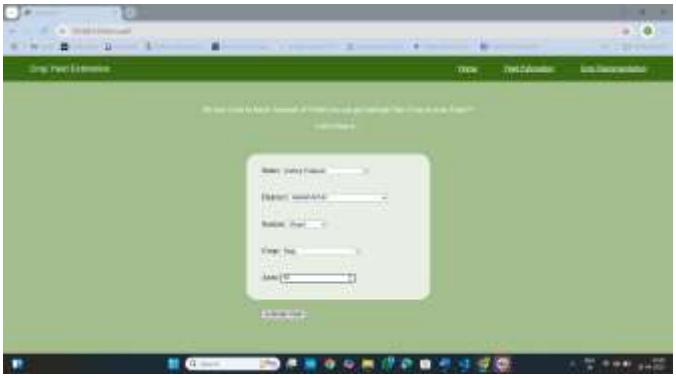

Figure 4.9 Crop Yield Estimation Input Interface

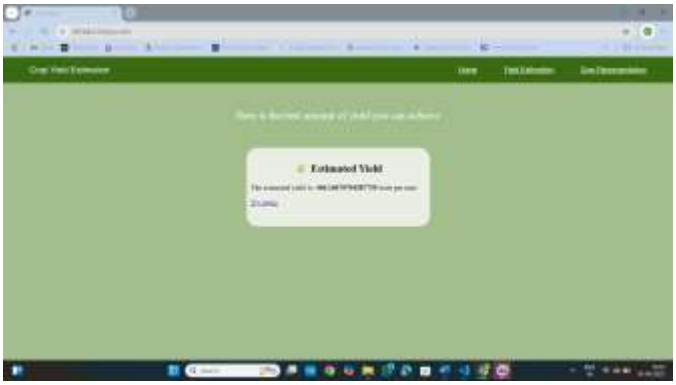

Figure 4.10 Crop Yield Estimation Output Interface

5. CONCLUSION

This project demonstrates the potential of machine learning in transforming traditional agriculture by providing accurate crop recommendations and yield estimations. By leveraging algorithms like Random Forest, Decision Tree, and XGBoost, and utilizing key agricultural inputs such as soil nutrients, climate conditions, and historical production data, the system delivers reliable, data-driven insights to farmers. The development of a web application ensures accessibility and practical utility, empowering even non-technical users to make informed farming decisions. Ultimately, this solution supports increased productivity, optimized resource use, and better

resilience against environmental and market uncertainties, contributing to sustainable agricultural development.

6. FUTURE SCOPE

We can extend the web application by using a Multilingual Interface, which will broaden the system's accessibility by supporting regional languages such as Hindi, Telugu, Tamil, and Bengali. This feature ensures that farmers across diverse linguistic backgrounds can comfortably use the system, increasing trust, engagement, and effective communication in the agricultural sector.

7. REFERENCES

[1]. Raman, R., Kantari, H., Gokhale, A. A., Elangovan, K., Meenakshi, B., & Srinivasan, S. (2024, March). Agriculture yield estimation using machine learning algorithms. In 2024 International Conference on Automation and Computation (AUTOCOM) (pp. 187-191). IEEE.

[2]. Z. Rao and J. Yuan, "Data mining and statistics issues of precision and intelligent agriculture based on big data analysis," Acta Agriculture Scandinavica, Section B—Soil & Plant Science, vol. 71, no. 9, pp. 870–883, 2021

[3]. Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE Access, 9, 63406-63439.

[4]. H. Gao, "Agricultural Soil Data Analysis Using Spatial Clustering Data Mining Techniques," in IEEE 13th International Conference on Computer Research and Development, pp. 83–90, 2021.

[5] Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: a review paper. Indian J. Sci. Technol, 9(38), 1-14.

[6]. B. Tanut, R. Waranusast, and P. Riyamongkol, "High accuracy pre-harvest sugarcane yield forecasting model utilizing drone image analysis, data mining, and reverse design method," Agriculture, vol. 11, no. 7, pp. 1–6, 2021.

[7]. K. I. Taher, A. M. Abdulazeez, and D. A. Zebari, "Data mining classification algorithms for analyzing soil data," Asian Journal of Research in Computer Science, vol. 8, no. 2, pp. 17–28, 2021.

[8]. Y. X. Shi, B. K. Zhang, Y. X. Wang, H. Q. Luo, and X. Li, "Constructing crop portraits based on graph databases is essential to agricultural data mining," Information, vol. 12, no. 6, pp. 1–7, 2021.

[9]. J. Suarez, B. Singh, F. H. Almukhtar, R. Kler, S. Vyas, and K. Kaliyaperumal, "Identifying smart strategies for effective agriculture solution using data mining techniques," Journal of Food Quality, pp. 1–9, 2022.

[10] Moraye, K., Pavate, A., Nikam, S., & Thakkar, S. (2021). Crop yield prediction using a random forest algorithm for major cities in Maharashtra state. International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN, 2347-5552.

[11] Shevchenko, V., Lukashevich, A., Taniushkina, D., Bulkin, A., Grinis, R., Kovalev, K., ... & Maximov, Y. (2024). Climate change impact on agricultural land suitability: An interpretable machine learning-based Eurasia case study. IEEE Access, 12, 15748-15763.

[12] Tussupov, J., Yessenova, M., Abdikerimova, G., Aimbetov, A., Baktybekov, K., Murzabekova, G., & Aitimova, U. (2024). Analysis of formal concepts for verification of pests and diseases of crops using machine learning methods. IEEE Access, 12, 19902-19910.