© January 2026 | IJIRT | Volume 12 Issue 8 | ISSN: 2349-6002

Pondicherry Crop Yield Prediction Using Machine Learning

S. Shirley[1], Vijayalakshmi A[2]
[1]*MCA., (Ph.D), Assistant Professor, Department of Master Of Computer Applications*
[2]*MCA, Christ College of Engineering and Technology, Moolakulam, Oulgaret Municipality, Puducherry-605010*

***Abstract*—Agriculture plays a vital role in the economy of India, and accurate crop yield prediction is essential for improving agricultural productivity and decision-making. In regions like Pondicherry, farmers face several challenges due to unpredictable weather conditions, variation in rainfall, temperature changes, and pesticide usage. Traditional methods of crop yield estimation are mostly based on experience and historical practices, which may not provide accurate results under changing climatic conditions.**

This project proposes a machine learning–based crop yield prediction system specifically designed for Pondicherry villages. The system analyzes environmental factors such as temperature, rainfall, and pesticide usage to predict crop yield efficiently. A trained machine learning model is integrated with a Flask-based web application that allows users to enter crop and environmental details and obtain yield predictions instantly. The system also stores prediction history and provides data visualization using graphical representations. This approach helps farmers and agricultural planners make informed decisions, reduce risk, and improve overall agricultural productivity.

***Index Terms*—Crop Yield Prediction, Machine Learning, Agriculture, Flask, Data Visualization**

I.INTRODUCTION

Agriculture is one of the most important sectors in India, supporting a large portion of the population. Crop yield prediction plays a significant role in agricultural planning, resource management, and food security. Accurate yield prediction helps farmers decide suitable crops, manage fertilizers and pesticides effectively, and reduce losses caused by adverse climatic conditions.

In recent years, climate change has introduced uncertainties in agricultural production. Variations in rainfall, temperature fluctuations, and improper pesticide usage directly affect crop yield. Traditional yield estimation techniques rely heavily on manual observation and historical trends, which are not sufficient to handle these dynamic conditions. Hence, there is a strong need for intelligent systems that can analyze multiple parameters and provide accurate predictions.

Machine learning techniques have shown promising results in analyzing large datasets and identifying complex patterns. By applying machine learning algorithms to agricultural data, crop yield can be predicted more accurately. This project focuses on developing a machine learning–based system for predicting crop yield in Pondicherry villages using environmental and agricultural data.

II. MAIN OBJECTIVES

The main objectives of the Pondicherry Crop Yield Prediction System are to design and develop a reliable machine learning–based application that can accurately predict crop yield using historical and environmental data. The system aims to assist farmers, agricultural planners, and researchers in making informed decisions by analyzing factors such as temperature, rainfall, pesticide usage, crop type, area, and year.

Another key objective is to provide a user-friendly web-based interface where users can securely register, log in, and input crop and environmental details easily. The system also focuses on visualizing agricultural data using graphs and charts to help users clearly understand trends and patterns affecting crop yield. By maintaining prediction history and ensuring data security, the project aims to deliver an efficient, accurate, and practical solution that supports better agricultural planning and improves productivity in the Pondicherry region.

III. SYSTEM OVERVIEW

The Pondicherry Crop Yield Prediction System is a web-based application developed to predict agricultural crop yield using machine learning techniques. The system analyzes historical crop data and environmental factors such as temperature, rainfall, pesticide usage, crop type, area, and year to generate accurate yield predictions. This automated approach helps farmers and agricultural planners make informed decisions and reduces dependency on traditional trial-and-error farming methods.

The system consists of an interactive frontend developed using HTML, CSS, and JavaScript, a backend implemented with the Flask web framework, and a machine learning model trained using historical crop yield data. Users can securely register and log in to the system, enter crop and environmental details, and receive predicted yield values. The system also stores prediction history in a database and provides data visualization using charts generated with Matplotlib. Overall, the system offers a reliable, user-friendly, and efficient platform for crop yield analysis and decision support in the Pondicherry region.

IV. SYSTEM ARCHITECTURE

The architecture of the Pondicherry Crop Yield Prediction System is designed using a modular and layered approach to ensure scalability, reliability, and ease of maintenance. The system follows a client–server architecture where the user interacts with the application through a web-based interface, and all processing and prediction operations are handled at the backend.

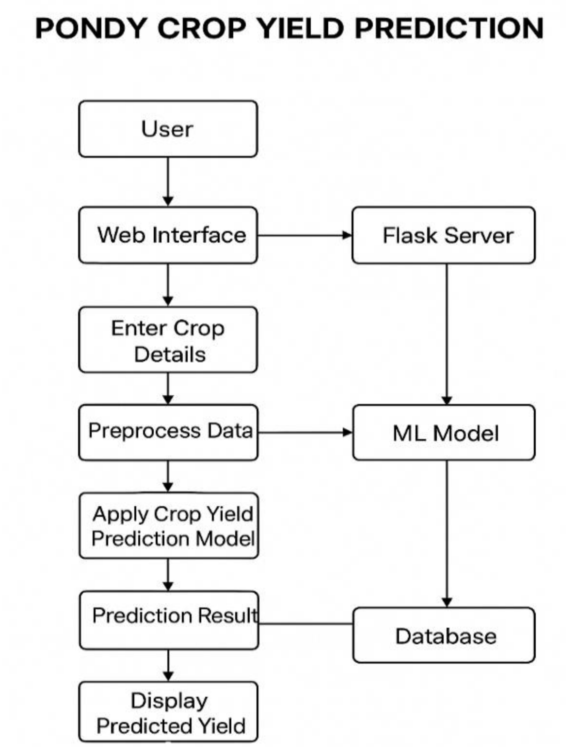
The User Interface Layer is developed using HTML, CSS, and JavaScript. This layer allows users to register, log in, enter crop details such as year, crop type, area, temperature, rainfall, and pesticide usage, and view prediction results. It also provides access to visualization features and prediction history, ensuring smooth and user-friendly interaction.

The Application Layer is implemented using the Flask web framework. This layer acts as a bridge between the user interface and the machine learning model. It handles user requests, input validation, session management, and communication with the database. Once valid inputs are received, the application layer forwards the data to the prediction module and returns the results to the user interface.

The Machine Learning Layer contains the trained crop yield prediction model. Historical crop data is pre-processed and used to train the model using suitable machine learning algorithms. During prediction, the model processes environmental and crop-related inputs and generates accurate yield predictions. The trained model is stored and loaded during runtime to ensure fast prediction.

The Database Layer uses SQLite to store user details, prediction history, and related application data. This layer ensures secure storage and efficient retrieval of information. All database operations are performed through controlled backend queries, maintaining data integrity and security.

Overall, this architecture ensures clear separation of responsibilities between layers, smooth data flow, secure data handling, and efficient prediction performance, making the system reliable and suitable for real-world agricultural applications.



V. ALGORITHM

RANDOM FOREST REGRESSION

Random Forest Regression is a supervised machine learning algorithm used in the Pondicherry Crop Yield Prediction system to estimate crop yield based on environmental and agricultural factors. It works by constructing multiple decision trees during training and combining their outputs to produce a more accurate and stable prediction. By using input parameters such as temperature, rainfall, pesticide usage, crop type, and cultivation area, the model learns complex and non-linear relationships present in historical crop data. Random Forest Regression reduces overfitting, handles large datasets efficiently, and provides reliable yield predictions, making it well suited for agricultural forecasting and decision support systems.

VI. RESULT AND DISCUSSION

The Pondicherry Crop Yield Prediction system was evaluated using historical agricultural data containing parameters such as year, crop type, cultivation area, temperature, rainfall, and pesticide usage. The trained Random Forest Regression model produced accurate and consistent yield predictions, demonstrating its ability to learn complex relationships between environmental factors and crop yield. The predicted values closely matched the actual yield values, indicating good generalization performance of the model on unseen data.

The results show that rainfall and pesticide usage have a significant influence on crop yield, while temperature and crop type also contribute to variations in production. The model effectively handled non-linear relationships and reduced overfitting due to the ensemble nature of Random Forest. Graphical visualizations generated using Matplotlib further helped in understanding yield trends across different years and regions in Pondicherry. Overall, the experimental results confirm that the proposed system is reliable, efficient, and suitable for real-time crop yield prediction, supporting farmers and agricultural planners in making informed decisions.

VII. BENEFITS

The Pondicherry Crop Yield Prediction system provides several benefits to farmers, agricultural planners, and researchers by using machine learning techniques for accurate yield estimation. The system helps in predicting crop yield in advance based on environmental and agricultural factors such as rainfall, temperature, and pesticide usage, enabling better planning and decision-making. Early prediction reduces uncertainty and supports efficient crop management.

The system improves agricultural productivity by assisting farmers in selecting suitable crops and optimizing resource usage. By analysing historical data, the model identifies important patterns that influence yield, helping to reduce losses caused by improper planning or adverse climatic conditions. Visualization features further enhance understanding by presenting trends and comparisons in an easy-to-interpret graphical format.

Additionally, the system reduces dependency on manual estimation methods, which are often time-consuming and inaccurate. Its user-friendly web interface makes it accessible even to users with minimal technical knowledge. Overall, the system promotes data-driven agriculture, supports sustainable farming practices, and contributes to improved food security and economic stability in the Pondicherry region.

VIII. DIFFICULTIES AND CHALLENGES FACED

One of the major challenges faced during the development of the Pondicherry Crop Yield Prediction system was the availability and quality of agricultural data. The dataset contained missing values, inconsistent entries, and variations across different years and regions, which required extensive pre-processing. Cleaning and preparing the data for machine learning analysis was time-consuming and critical to ensure accurate prediction results.

IX. CONCLUSION

The Pondicherry Crop Yield Prediction system successfully demonstrates the application of machine learning techniques to predict agricultural yield based on key environmental and crop-related parameters. By analyzing factors such as rainfall, temperature, pesticide usage, crop type, and area, the system provides accurate and reliable yield predictions that can support farmers and agricultural planners in decision-making. The integration of a user-friendly web interface, data visualization, and prediction history enhances usability and practical relevance. Overall, the system proves to be an effective, scalable, and efficient solution for improving agricultural planning and promoting data-driven farming practices.

X. FUTURE ENHANCEMENTS

The system can be further enhanced by incorporating real-time weather data from government or meteorological APIs to improve prediction accuracy. Including additional parameters such as soil type, humidity, fertilizer usage, and irrigation methods can make the model more robust and realistic. Advanced machine learning and deep learning models can be explored to achieve better performance with large datasets. The system can also be expanded to support mobile applications and multilingual interfaces to benefit farmers more effectively. Furthermore, continuous model retraining with updated agricultural data will help the system adapt to changing climate conditions and improve long-term reliability.

REFERENCES

[1] Mitchell, T. M., *Machine Learning*, McGraw-Hill Education, 1997.

[2] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 2006.

[3] Han, J., Kamber, M., and Pei, J., *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2012.

[4] Jain, A., Kumar, A., and Pandey, S., "Crop Yield Prediction Using Machine Learning Techniques," *International Journal of Computer Applications*, Vol. 182, No. 7, 2018.

[5] Pant, J., Sharma, R., and Singh, P.,"Analysis of Agricultural Data Using Data Mining Techniques for Crop Yield Prediction," *IEEE International Conference on Computing*, 2019.

[6] Khosla, E., Dharavath, R., and Priya, R., "Crop Yield Prediction Using Machine Learning Algorithms," *International Journal of Engineering Research & Technology (IJERT)*, Vol. 7, No. 6, 2018.

[7] Jeong, J. H., Resop, J. P., Mueller, N. D., et al., "Random Forests for Global and Regional Crop Yield Predictions," *PLOS ONE*, 2016.

[8] Lobell, D. B., Schlenker, W., and Costa-Roberts, J., "Climate Trends and Global Crop Production," *Science*, Vol. 333, 2011.

[9] Pedregosa, F., Varoquaux, G., Gramfort, A., et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, 2011.

[10] Breiman, L., "Random Forests," *Machine Learning*, Vol. 45, No. 1, 2001.

[11] Python Software Foundation, *Python 3 Documentation*, 2024.

[12] Scikit-learn Developers, *Scikit-learn Documentation*, 2024.

[13] Pandas Development Team, *Pandas Documentation*, 2024.

[14] NumPy Developers, *NumPy Documentation*, 2024.

[15] Hunter, J. D., "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, IEEE, 2007.

[16] Flask Development Team, *Flask Web Framework Documentation 2024.*

[17] Quinlan, J. R., "Induction of Decision Trees," *Machine Learning*, Vol. 1, No. 1, 1986.

[18] Chakraborty, M., and Joseph, A., "Machine Learning Techniques for Agricultural Yield Prediction," *International Journal of Advanced Research in Computer Science*, 2020.

[19] FAO, *Food and Agriculture Organization Statistical Database*, United Nations, 2023

[20] Government of India, *Agricultural Statistics at a Glance*, Ministry of Agriculture & Farmers Welfare, 2022.