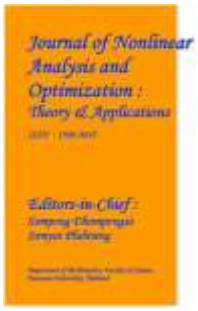
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**



CROP CLASSIFICATION AND YIELD PREDICTION USING ROBUST MACHINE LEARNING MODELS FOR AGRICULTURE SUSTAINABILITY

J.KUMARI[1], MALAMPATI BHARATH VENKATA NAGA SAI[2]

[1]Assistant Professor, Dept. of MCA, QIS College of Engineering and Technology, Ongole, Andhra Pradesh.

[2]PG Scholar, Dept. of MCA, QIS College of Engineering and Technology, Ongole, Andhra Pradesh.

ABSTRACT— A nation's economy depends heavily on agriculture since it provides a significant amount of its raw resources, jobs, and food. But problems like illnesses, degraded soil, and water scarcity still exist. Adoption of technology can solve these problems and enhance output and quality. Prediction, categorization, and automation in areas like soil pH, temperature, humidity, and nutrient levels are made possible by machine learning, a branch of artificial intelligence (AI). We recommend twenty-two distinct crops based on these inputs by utilizing machine learning classification techniques like Extra Tree Classifier (ETC), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbour (KNN), Gaussian Naive Bayes (GNB), and Support Vector Machine (SVM). We determine the best-performing model using K-fold cross-validation, Explainable AI (XAI), and feature engineering. Random Forest emerges as the top model, with an accuracy of 99.7% with precision, recall, F1 score, and confusion matrix.

Index Terms – Random forest, machine learning, and crop yield prediction.

I. INTRODUCTION

According to our assessment, which we discovered in previous research papers, everyone uses plant factors like soil type and nutrients (like potassium and nitrogen) and climatic factors like precipitation and sunlight. The problem is that we actually need to gather the data, and after a while, an

untouchable makes this assumption, which is then explained for farmer agriculture. It helps in crop management and food security decision-making by optimizing crop selection, fertilization, and irrigation. Based on different datasets, this paper suggests two reliable machine learning architectures for regression and classification. We start by looking at a crop recommendation dataset that we got from Kaggle. It has a lot of different input variables, and the farmer has to work hard to figure out the science behind them. This study employs basic elements such as the farmer's state and region, the crop, and the season (e.g., Kharif, Rabi, etc.) to work on it and determine which can be utilized directly by the farmer. Over 100 yields are spread out across the entire nation of India. These yields are sought in order to improve representation and appreciation.

The Indian Government Repository provided the data used in this evaluation. Around 2.5 lakh insights are included in the data, which includes the following properties: State, District, Crop, Season, Year, Area, and Production.

II. LITERATURE SURVEY

A. An improved crop yield prediction model using bee hive clustering approach for agricultural data sets

All throughout the world, rural experts insist on the need for an efficient tool to predict and advance crop development. The local farming community is acutely aware of the need for an integrated harvest development control system with precise predictive production management. Complex variable measurements and the unavailability of predictive displaying techniques make it extremely difficult to predict harvest production, which leads to agricultural yield tragedy. In order to predict the harvest yield and advance the dynamic in precision horticulture, this research study suggests a harvest yield prediction model (CRY) that deals with a flexible group approach over forcefully updated verifiable yield informational index. In order to deal with research and organize the harvest based on yield development design, CRY uses apiary showing. Using Clementine over the yield space information that was already available, a CRY-characterized dataset was attempted.

B. An intelligent system based on kernel methods for crop yield prediction

The effort on developing a product framework for predicting crop production from ranch and environmental data is presented in this study. This framework is based on a solo parceling of information

strategy that uses piece techniques to manage complex information in order to locate spatio-transient examples in environment information. This leads to the introduction of a potent weighted part k-implies calculation that integrates geographical limits. By breaking down the various factors influencing the yield, the computation can really handle noise, irregularities, and auto-relationships in the spatial data, allowing for a strong and fruitful information analysis. In this way, it may be used to predict oil-palm yield.

C. Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models.

In India, farming plays a vital role in the economy. Because of this, predicting crop yields is a big task that will aid in India's development. Crops are sensitive to several climatic quirks, such as precipitation and temperature. Therefore, it becomes crucial to take these factors into account when predicting a harvest's yield. Determining weather conditions is a complex cycle. ARMA (Auto Regressive Moving Average), SARIMA (Seasonal Auto Regressive Integrated Moving Average), and ARMAX (ARMA with exogenous components) are conjectured using three different methodologies in this work. In light of a fluffy reasoning model, the three are examined, and the best model is used to predict temperature and precipitation, which are then used to predict the harvest yield.

D. Crop Yield Prediction Using Data Analytics and Hybrid Approach

The creation of rural information is ongoing and enourmosly. As a result, farming information has emerged during the era of vast amounts of information. Innovative developments aid in the collection of information through the use of technological devices. In our project, we will dissect and mine this rural data to produce useful results using technologies like artificial intelligence (AI) and information analysis. Ranchers will receive these results for increased harvest production in terms of efficiency and proficiency.

III. PROPOSED SYSTEM

The overview of our proposed system is shown in the below figure.

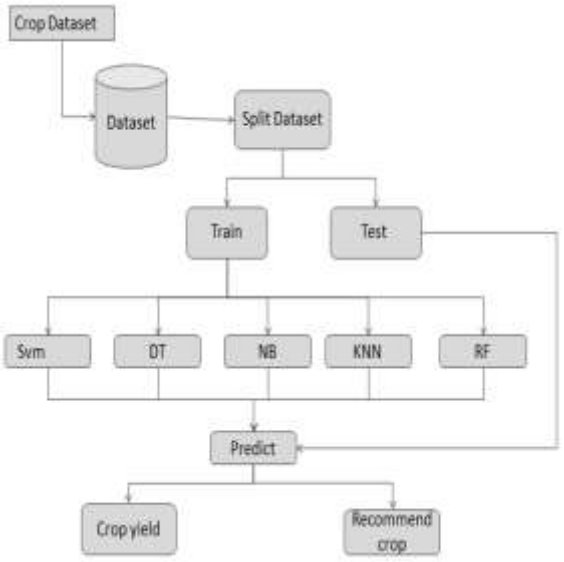


Fig. 1: System Overview

Implementation Modules

Preprocessing

Many 'NA' values are separated in Python for the provided data collection. Furthermore, as the data set includes numerical data, we used hearty scaling, which is similar to standardization but instead makes use of the interquartile range. Standardization, on the other hand, recoils data pertaining to 0 to 1.

Prediction

This module uses various AI classifiers to predict the harvest yield and creation based on the state name, crop name, number of field sections, and soil type. Additionally, recommend the harvest based on the soil type and status.

Graphical Analysis

The client can obtain a reasonable image of the cause of death investigation during this phase of implementation. Consider a variety of variables when examining the diagram. Plot the outlines at this step, such as a pie diagram, bar graph, and so on.

Implementation Algorithm

Random Forest

- It generates several option trees, each of which makes a prediction based on a test of information.
- The result that was achieved by the highest number of trees is then regarded as the final forecast.
- Irregular Backwoods is a Supervised Learning calculation that groups and relapses using the outfit learning technique. The trees in irregular woodlands run in a line with essentially no communication, and irregular woodlands are a packing operation.
- A Random Forest predicts the relative plurality of trees by creating a small number of selected trees during the preparation phase and using the mean of the classes.

Decision Tree

- Trees are created using an algorithmic approach that identifies methods to segment the informative index according to different conditions.
- It is among the most often used effective directed learning strategies.
- These non-parametric methods are used for both relapse and organization.

K Nearest Neighbour

- In light of the supervised learning approach, K-Nearest Neighbor is among the simplest machine learning computations.
- The K-NN algorithm describes another information point based on the proximity and stores all of the available information. This suggests that as new information becomes available, it is typically efficiently categorized using the K-NN algorithm.
- The K-NN calculation is mostly used for classification problems, while it can also be used for regression in relation to classification.
- ***Support Vector Machine***
- One of the most well-known supervised learning algorithms, Support Vector Machines, or SVMs, are used for both classification and regression problems. In any case, it is mostly used for machine learning classification problems.
- The SVM calculation's goal is to determine the optimal line or decision limit that can divide n-layered space into classes so that we can subsequently categorize the new data of interest. A hyperplane is the name given to this best option limit.
- SVM selects the ludicrous focuses and vectors that help create the hyperplane.

IV. RESULTS

Fig. 2: Comparison of Accuracy Different

Fig. 3: Crop Prediction

Fig. 4: Crop Identification

V. CONCLUSION

When stacking relapse is used, the results are significantly better than when those models were used separately. Currently, the outcome shown in the picture is a web application; however, our next project would be to create an application that ranchers could use to change the complete framework in their own language.

REFERENCES

[1] Arunkumar, T., Hemavathy, R., and Ananthara, M. G. (2013, February). CRY is an enhanced crop production prediction model for agricultural data sets that employs a bee hive clustering technique. Pattern Recognition, Informatics, and Mobile Engineering, International Conference, 2013 (pp. 473-478). IEEE.

[2] In April 2006, Awan, A. M., and Sap, M. N. M. An intelligent system for predicting agricultural yield that uses kernel approaches. Pacific-Asia Conference on Data Mining and Knowledge Discovery (pp. 841-846). Heidelberg, Berlin: Springer.

[3] M. A. Shaik, A. Rahim, V. Subhalakshmi, D. R. Ravi Kumar, R. Pasunuri and D. Verma, "Exploring Time Series Techniques in Production Function Modeling: ARIMA and VECM Applications," 2025 International Conference on Intelligent Computing and Control Systems (ICICCS), Erode, India, 2025, pp. 160-165, doi: 10.1109/ICICCS65191.2025.10985700.

[4] Thombare, R. A., Dhemey, P. G., Chaudhari, A. N., and Bhosale, S. V. (2018, August). Crop Yield Forecasting with a Hybrid Method and Data Analytics. (pages. 1–5) in the Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) in 2018. IEEE.

[5] M. A. Shaik, G. Rakshitha, K. Saipriya, T. Thrisha, M. Varshini and J. G. Sai, "Machine Learning for Detecting the Phishing Threats," 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal, 2025, pp.

1221-1226, doi: 10.1109/ICMCSI64620.2025.10883227.

[6] Gandhi, N., Armstrong, L. J., & Petkar, O. (2016, July). Predicting rice crop yield with artificial neural networks. IEEE Technical Advancements in ICT for Rural and Agricultural Development (TIAR), 2016 (pp. 105–110). IEEE.

[7] M. A. Shaik, V. S. Rani, A. Fatima, M. Parveen, J. Juwairiyyah and N. Fatima, "Secure Data Exchange in Cloud Computing: Enhancing Confidentiality, Integrity, and Availability Through Data Partitioning and Encryption," 2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG), Chennai - 600077, Tamil Nadu, India, 2024, pp. 1-6, doi: 10.1109/ICSTSDG61998.2024.11026651.

AUTHORS Profile

Mrs. J.Kumari is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. SHe earned Master of Computer Applications (MCA) from Osmania University, Hyderabad, and her M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). Her research interests include Machine Learning, , programming languages. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.

Mr. MALAMPATI BHARATH VENKATA NAGA SAI has received him B.SC (Computer Science) and degree from ANU 2023 and pursuing MCA in QIS College of Engineering and Technology affiliated to JNTUK in 2023-2025