SETSCI conference proceedings
SETSCI Conference Proceedings
17, 40-43, 2024

Cognitive Models and Artificial Intelligence Conference
May 25-26, 2024, İstanbul-Türkiye

AICCONF 2024

https://doi.org/10.36287/setsci.17.1.0040

Crop Yield Prediction Using Machine Learning Algorithms

Dr. G. Malini Devi[1], Badhe Siri Vennela[2], Shreya Arukala[3], Sai Amogha Uppalapati[4] and Kambalapally Anudeepthi[5]

[1 2 3 4 5]*G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India*

[1]*gmalini12@gnits.ac.in,* [2]*vennela.badhe@gmail.com,* [3]*shreya.arukala@gmail.com,* [4]*usamogha@gmail.com,* [5]*anudeepthi915@gmail.com*

Abstract – The study will utilize machine learning algorithms such as Random Forest, Gradient Boosting, and Support Vector Regression (SVR) on data collected from the districts of Nalgonda, Yadadri Bhuvanagiri, and Suryapet over the past two years. The project seeks to meet the growing demand for crop yield prediction models that enhance agricultural productivity and enable informed decision-making by farmers. Model accuracy will be evaluated using metrics like mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R2). The dataset provided includes features such as soil nutrient levels, climate data, and crop yield, which will be preprocessed and subjected to feature selection. The study results are expected to contribute to the development of precise and efficient crop yield prediction models, supporting sustainable agricultural practices and empowering farmers with informed decisions regarding crop management, planting, harvesting, and overall farm management. The project's focus is to determine the best-performing algorithm and pave the way for enhanced agricultural productivity and decision-making in crop management.

Keywords – Machine Learning, Data Visualization, Predictive Modelling

I. Introduction

Agriculture, as the backbone of human civilization, constantly faces the challenge of ensuring optimal crop production for population growth, climate change, and resource scarcity. Machine learning offers solutions by predicting crop yields through historical and real-time data analysis, including climate, soil, genetics, and farming practices. Machine learning optimizes resource allocation, reducing costs and environmental impact by accurately estimating crop yields. It enables informed decisions on water, fertilizers, and pesticides, improving productivity and sustainability in agriculture.

Furthermore, machine learning improves agricultural decision-making and risk management. Accurate yield predictions enable proactive responses to market changes, support supply chain management, and prompt strategy adjustments like exploring alternative crops to reduce financial risks.

Additionally, machine learning contributes to precision agriculture by tailoring practices to field conditions. Integrating with technologies like remote sensing and IoT provides real-time data on crop health, soil moisture, and nutrients. Machine learning models offer actionable insights for precise interventions such as targeted irrigation and pest control.

Implementing machine learning in crop yield prediction does face challenges. Data availability and quality are significant issues, especially for small-scale farmers in many regions. Ensuring data reliability requires data preprocessing and quality control. Interpreting machine learning models is another challenge. While they make accurate predictions, their complexity can be a barrier. To gain trust and acceptance, interpretable models are essential as they provide insights into input-variable relationships, helping farmers understand recommendations and make informed decisions.

Scalability and accessibility of machine learning-based crop yield prediction are crucial. These systems should be user-friendly and adaptable to diverse farming contexts and technologies. Ensuring small-scale farmers can benefit from these technologies is essential for sustainable agriculture. Collaboration among researchers, policymakers, and technology providers is vital for tailoring solutions to diverse farming communities.

In summary, machine learning in crop yield prediction has the potential to revolutionize agriculture by improving accuracy and efficiency, optimizing resource allocation, enhancing decision-making, and supporting precision agriculture. Addressing challenges like data availability, interpretability, scalability, and accessibility is essential for widespread adoption and ensuring food security, sustainable agriculture, and the well-being of farming communities worldwide.

II. Literature Survey

The proposed work examines soil parameters (e.g., Nitrogen, Phosphorous, Potassium, Soil Moisture, pH), weather factors (temperature, rainfall), and crop rotation's influence on agriculture. Leveraging Machine Learning (ML) algorithms and historical weather data, the research aims to boost farm yields and crop productivity. A website is developed for monitoring fields and offering Smart Agriculture solutions. ML algorithms use weather data to recommend the most profitable crops and predict yields by considering weather, soil, and historical data. This integrated approach enhances crop yield and supports farmers' long-term profitability [7].

An integrated portal is proposed for farmers, providing crop and yield predictions, a discussion forum, news updates, and

an ecommerce platform. While the Naive Bayes algorithm predicts crops and yields, and the Apriori algorithm offers tailored recommendations, it assumes data independence and has limited representation of complex data. The system's reported accuracy is 84.71%. Nonetheless, it aims to simplify farmers' lives, enhance productivity, and encourage knowledge-sharing in agriculture [11].

Large-scale crop mapping is crucial for resource monitoring, but traditional methods are time-consuming due to labeling and complex feature design. A deep one-class crop extraction framework is proposed in this paper, automating feature extraction, and overcoming the lack of a negative class. It supports various remote sensing data types. Although it has a high cost and limited adaptability, it promises accurate and efficient crop mapping, reducing labeling efforts and feature design complexity [5].

This research uses MLP neural networks for district-level wheat crop yield forecasting. It introduces a new activation function, DharaSig, providing accurate results faster than existing methods. The study evaluates various activation functions, including DharaSig, DharaSigm, and SHBSig, achieving 93.4% accuracy. While it has limited flexibility, these new functions outperform the default 'sigmoid' activation function for agriculture datasets [1].

In low-lying countries like Bangladesh, accurate crop yield estimation is crucial due to climate change's impact. Rice, with an annual production exceeding forty million tons, is particularly affected. The paper presents the WPSRY (Weather-based Prediction System for Rice Yield) approach, using Neural Networks to predict weather parameters and Support Vector Regression to estimate rice yields. It achieves an impressive accuracy of 93.71%, aiding farmers and policymakers in addressing climate change challenges in rice production [6].

Another study focuses on using machine learning algorithms to predict crop yields based on NPK (nitrogen, phosphorus, potassium) values. By integrating these nutrients, it aims to improve prediction accuracy. The research explores various machine learning techniques for crop yield prediction, offering insights into their effectiveness in agriculture and contributing to the field's knowledge [13].

III. Methodology

The objectives of this project are focused on enhancing crop prediction and user experience in agriculture:

Data Enhancement: Our first objective is to improve the quality of the dataset by employing preprocessing techniques. This includes handling missing values, outlier detection, and normalization. We also address class imbalances in the data while splitting it into training and testing sets to ensure effective modeling.

Attribute Selection and Optimization: To enhance model accuracy, we aim to select ideal attributes and optimize their use. We will leverage feature importance and dimensionality reduction techniques to achieve this objective.

Algorithm Selection and Comparison: We plan to implement a range of algorithms, including high-accuracy ones like Gradient Boosting and Random Forest, as well as lower-accuracy alternatives like KNN. The goal is to compare their performance and computational complexity to make informed choices.

Performance Evaluation: The fourth objective involves assessing the model's performance using various metrics such as accuracy, precision, recall, F1-score, and the confusion matrix. Additionally, we will fine-tune hyperparameters to optimize the model's performance.

User-Friendly UI Development: We aim to create an intuitive user interface that allows users to input soil nutrient values and receive instant crop suggestions along with confidence levels. The interface will also support multiple queries for a seamless user experience.

Error Handling and Optimization: To ensure the system's reliability, we will implement robust error handling mechanisms. We will introduce caching for faster responses and consider data visualization techniques to aid in data exploration and model interpretation.

Continuous Improvement and Monitoring: Our final objective is to continuously monitor and update the system with new data. This ongoing refinement process will contribute to sustainable agriculture development and ensure the system's accuracy over time.

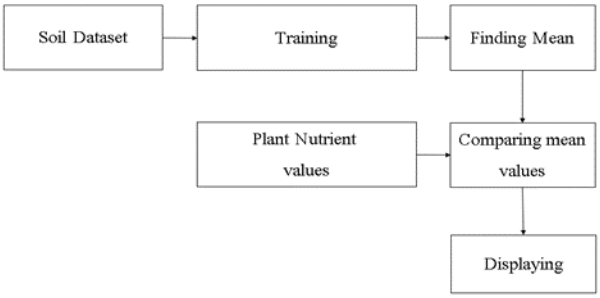

Fig. 1. Methodology Diagram

IV. Implementation

In the proposed system, crop suggestion and yield prediction rely on soil macro-nutrients (NPK: Nitrogen, Phosphorus, Potassium) obtained from a Soil Health Card. This card assesses soil health and provides descriptive indicators, often based on farmers' knowledge. The system offers two ways to acquire NPK values. The first method employs image processing to extract NPK values directly from the image, eliminating manual intervention and simplifying the process for users.

NPK values in the soil health card are obtained using image processing, ensuring precise and reliable results. This streamlined data collection method saves time and resources, allowing users to rapidly retrieve NPK values with automation, reducing human error, and enhancing system usability for all users.

The last method involves manual entry of NPK values for users who prefer or when image quality is poor. It's rarely needed but provides an alternative when necessary.

A. Machine Learning Algorithms

Support Vector Machines (SVM)

Support Vector Regression (SVR) combines Support Vector Machines with regression to find an optimal hyperplane while accommodating an epsilon-insensitive zone, using kernel functions for nonlinearity, and incorporating regularization to prevent overfitting. It's versatile, applicable in banking, healthcare, and environmental science, for tasks like stock price forecasting and patient outcome estimation. SVR effectively models complex interactions, balances model complexity for accurate predictions, and works well with small to medium-sized datasets, making it a robust and widely

applicable regression method.

Random Forest Regressor

Random Forest is a popular ensemble learning method that combines decision trees to create a robust model. It uses bootstrap sampling and random subsets of features for each tree, reducing overfitting and improving generalization. It's effective for large datasets, excels in both classification and regression, handles high-dimensional data, complex interactions, and outliers. It also provides valuable insights into feature importance, making it widely used across industries.

Gradient Boosting Regressor

Gradient Boosting is a strong machine learning method that combines weak models (often decision trees) to create a precise and robust ensemble model. It corrects mistakes made by earlier trees, minimizes loss using gradient descent, and focuses on challenging cases. It's versatile, resistant to overfitting, and excels at complex data patterns. Industries like finance, healthcare, and online advertising use it for regression, classification, and ranking, making it popular in competitions.

V. Results and Discussions

We evaluated the performance of various decision tree algorithms and examined their accuracies using the best test split.

The algorithms we experimented with included the k-nearest neighbours, support vector regressor, basic decision tree, random forest, and gradient boosting. For each algorithm, we trained the models using a comprehensive dataset that consisted of crop yield samples along with corresponding N, P, and K values.

Table 5.1. Comparison of accuracies for all algorithms – Rice

Algorithm	Accuracy Percentage
KNN	94.50
SVR	96.00
Decision tree	88.33
Random forest	95.51

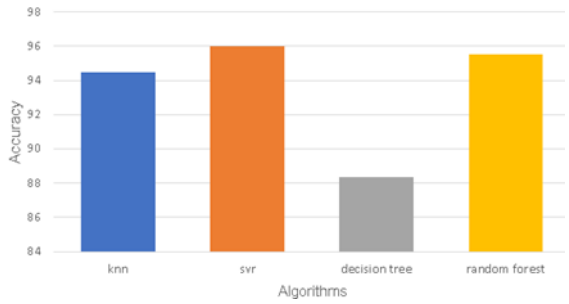

Fig. 2. Histogram Comparing accuracies of all algorithms on the rice dataset.

Table 2. Comparison of accuracies for all algorithms -Groundnut

Algorithm	Accuracy Percentage
KNN	88.45
SVR	94.29
Decision tree	89.50
Random forest	90.66

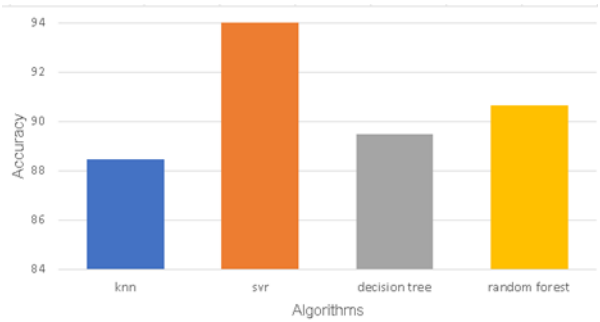

Fig. 3. Comparing accuracies of all algorithms on the groundnut dataset

Table 3. Comparison of accuracies for all algorithms - red gram

Algorithm	Accuracy Percentage
KNN	96.35
SVR	82.49
Decision tree	91.03
Random forest	96.98

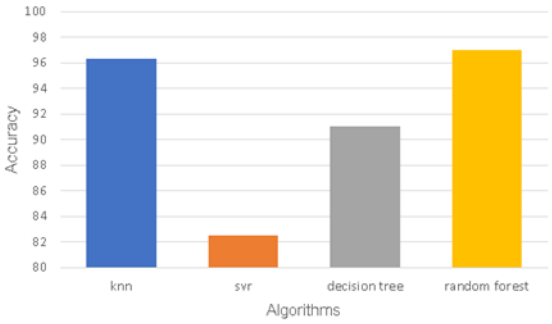

Fig. 4. Comparing accuracies of all algorithms on the red gram dataset

Table 4. Comparison of accuracies for all algorithms - Maize

Algorithm	Accuracy Percentage
KNN	72.36
SVR	83.74
Gradient Boosting Regressor	93.48
Random forest	88.57

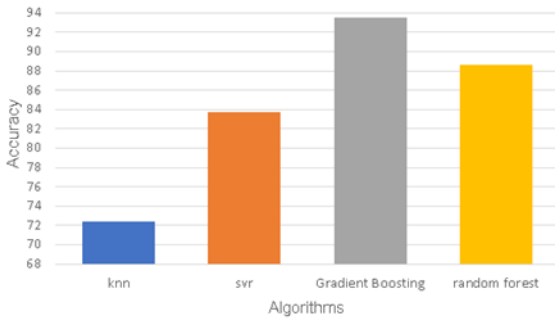

Fig. 5. Comparing accuracies of all algorithms on the maize dataset

VI. Conclusion

In conclusion, SVR, Random Forest, and Gradient Boosting algorithms hold promise for precise crop yield prediction by effectively handling complex, high-dimensional datasets and capturing intricate relationships among factors influencing crop yield. SVR's capacity to model non-linear relationships, Random Forest's proficiency with diverse and noisy data, and Gradient Boosting's ability to uncover complex patterns make them suitable choices. These algorithms are well-matched to a dataset comprising soil data, weather conditions, crop types, and planting dates, offering accurate predictions by accommodating non-linear relationships in agricultural

factors. Leveraging these algorithms can empower farmers with informed decisions for crop management, promoting sustainability and productivity in agriculture, although further research and development are crucial for practical implementation and enhanced effectiveness.

REFERENCES

[1] Bhojani, S.H., Bhatt, N. Wheat crop yield prediction using new activation functions in neural network. Neural Computing & Applications 32, 13941–13951 - 2020

[2] Choudhury, P. R. D., & Patil, P. M., A review of machine learning techniques for crop yield prediction. In Proceedings of the 2nd International Conference on Energy, Environment, and Sustainable Development (pp. 285-294), Springer - 2020

[3] Juhi Reashma S R K, & Pillai, A. S. Edaphic factors and crop growth using Machine learning – A Review. In: Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS) - 2017

[4] Kumar, A., & Panigrahi, B. K. (2020). Crop yield prediction using machine learning: A review. Computers and Electronics in Agriculture, 174, 105507 - 2020

[5] L. Lei, X. Wang, H. Zhao, X. Hu, C. Luo, and Y. Zhong, Deep One Class Crop Extraction Framework for multimodal remote sensing imagery, IEEE International Geoscience and Remote Sensing Symposium IGARSS - 2021

[6] M. A. Hossain, M. N. Uddin, and Y. M. Jan, Predicting rice yield for Bangladesh by exploiting weather conditions, International Conference on Information and Communication Technology Convergence (ICTC) - 2017

[7] M. S. Teja, T. S. Preetham, L. Sujihelen, Christy, S. Jancy and M. P. Selva, Crop Recommendation and Yield Production using SVM algorithm, 6th International Conference on Intelligent Computing and Con trol Systems (ICICCS) - 2022

[8] Pradhan, P., Kumar, V., & Ramachandran, K. I. Crop yield prediction models: Recent advances and future directions. Computers and Electronics in Agriculture, 162, 967-982 - 2019

[9] Rastogi, R., & Patel, S., Crop Yield Prediction Using Machine Learning Techniques: A Comprehensive Review. International Journal of Computer Science Trends and Technology, 6(3), 96-101 – 2018

[10] Saurabh, N., Kumari, S., & Suri, B. M. Crop Yield Prediction using Machine Learning Techniques with NPK Values. In: Gupta, V., Bhatnagar, V., & Bhateja, V. (Eds.), Proceedings of International Conference on Data Engineering and Communication Technology (ICDECT), 23-33 - 2020

[11] Shreya S, Sushmita S R, Vaanathe L R and Madhumati R, Agro World: A Naive Bayes based System for Providing Agriculture as a Service, 6th International Conference on Intelligent Computing and Control Systems (ICICCS) - 2022

[12] Teja, M. S., Preetham, T. S., Sujihelen, L., Christy, & Jancy, S., Selva, M. P. Crop Recommendation and Yield Production using SVM algorithm, IEEE - 2021

[13] Tiwari, N., & Bhatt, R. Crop Yield Prediction using Machine Learning Algorithms with NPK Values. In: Kumar, N., Verma, H., & Bhatnagar, V. (Eds.), Proceedings of International Conference on Machine Intelligence and Data Science (ICMIDS), 79-89 - 2020