MATEC Web of Conferences **392**, 01086 (2024) https://doi.org/10.1051/matecconf/202439201086
ICMED 2024

Zone-II & III: Machine Learning based Rice Yield Prediction in Andhra Pradesh

Sowjanya Ramisetty[1*], *Devdas* Bansode[2], *Vijay Kumar* Atmakur[3], *and Gnanasudha Pradeep* Ghantasala[4], D. Ushasree[5], *Manish* Kumar[6]

[1]Department of Computer Science and Engineering, KG Reddy College of Engineering & Technology, Hyderabad, Telangana,501504, India
[2]Department of Computer Science and Engineering (AI&ML), KG Reddy College of Engineering & Technology, Hyderabad, Telangana,501504, India
[3]Department of Artificial Intelligence, Vidya Jyothi Institute of Technology, Hyderabad, Telangana,500075, India
[4]Department of Computer Science and Engineering, Alliance University, Bengaluru, Karnataka, 562106, India
[5]Department of CSE, GRIET, Hyderabad, Telangana, India
[6]Lovely Professional University, Phagwara, Punjab, India.

Abstract. In a unique approach, this research predicts rice yield in Zones II and III of Andhra Pradesh using machine learning algorithms. Because food security and agricultural sustainability are becoming more and more important, an accurate assessment of crop yields is necessary for agricultural planning and decision-making. By using machine learning approaches, we hope to develop robust prediction models that can forecast rice yields depending on a variety of agronomic and environmental conditions. Our plan includes collecting a lot of data on variables such as soil qualities, weather patterns, crop management practices, and historical yield records from the target regions. Using careful feature engineering and data preprocessing, we find relevant predictors and derive significant insights for rice yield estimation. Next, we employ state-of-the-art machine learning methods, like Support Vector Machines, Random Forest, and Gradient Boosting, to build trustworthy prediction models. These models are trained on historical data and then cross-validated to guarantee accuracy and generalizability. Our goal is to combine domain expertise in agriculture with cutting-edge machine learning techniques to provide farmers, policymakers, and other agricultural stakeholders with insightful understandings of the dynamics of rice production. In Zones II and III of Andhra Pradesh, this will make it easier to allocate resources and make well-informed decisions that will improve agricultural output and sustainability.

1 Introduction

* Corresponding author: drsowjanyaramisetty@gmail.com

India's economy is mostly dependent on agriculture, and rice is one of the most important staple crops. In the southern state of Andhra Pradesh, zones II and III are crucial for rice production, which contributes significantly to the state's agricultural output. However, the unpredictability of rice harvests presents challenges for legislators, farmers, and agricultural planners. Due to fluctuations in temperature, soil conditions, and agricultural practices, variations in rice yields can affect both food security and economic stability. To solve these problems, there is growing interest in applying machine learning approaches to reliably anticipate rice yields. We can give Andhra Pradesh's farmers practical insights and decision-making tools through the use of data-driven models, empowering them to optimize crop management techniques and increase agricultural productivity.

Using machine learning algorithms into agriculture could transform traditional farming practices and increase agricultural resilience to environmental uncertainties and climate change. Machine learning-based rice yield prediction models offer a data-driven method to understanding the complex interrelationships among crop yields, agronomic practices, and environmental conditions. By analyzing historical yield data, weather patterns, soil qualities, and other relevant variables, these models might identify patterns and connections that traditional statistical methods would overlook. Farmers and other agricultural stakeholders can use this comprehensive approach to yield prediction to make well-informed decisions about planting schedules, irrigation management, fertilization techniques, and pest control methods, thereby creating more sustainable and effective agricultural systems.

In Andhra Pradesh's Zones II and III, we look into the potential applications of machine learning approaches for yield predictions. We leverage a comprehensive dataset comprising historical yield records, meteorological data, soil information, and satellite imagery to create and evaluate machine learning models capable of consistently forecasting rice yields at the regional level. Through providing practical insights into agricultural production dynamics and yield variability, our research aims to advance sustainable farming practices and precision agriculture for farmers and policymakers. Agricultural communities in Zones II and III may make better use of machine learning's predictive powers to better adapt to changing environmental circumstances, lower risks, and allocate resources in an efficient manner to improve livelihoods and food security.

2 Literature Survey

This article presents an overview of the application of these machine learning techniques in rice-growing regions of India. This work investigates the experimental outcomes of applying the SMO classifier [1] with the WEKA tool, using a dataset that includes 27 districts in the Indian state of Maharashtra.The concept put out in this work is the use of an improved extreme linear machine to estimate agricultural yield. We employed the Kalman Filter Algorithm [2] to prepare the data, Linear Discriminant Analysis to extract particular characteristics, and the upgraded machine to conduct crop prediction. Predictions of rice crop yield have been made using factors such as cultivation area, season, and geography.

The agriculture department[3] may use the project's results to determine crop yield. The best result, with a mean absolute error of 0.3133, is produced by the neural network algorithm.Numerous elements that are crucial to agriculture, including soil, climate, fertilizer, and irrigation, affect the crop's yield[4]. After examining the production and yield of rice in Tamil Nadu, a paradigm is created to test the prediction of rice yield based on the elements impacting the rice yield.

In AP, time-series data analytics are still utilized to forecast rice prices, such as machine learning algorithms. Prototypes of SARIMA, Holt-Winter's Periodic technique [5], and Long Short-Term Memory Neural Networks were implemented. Based on the RMSE value on the rice database, which included expenses from 2001 to 2020, their performance was assessed. In order to forecast rice production in the Bangladesh region, we used weather forecast scenarios in this paper [6]. The rice yield production amount is predicted with 95.89% accuracy using Support Vector Regression , which is used for target prediction.
Three steps of initial pre-processing are carried out by the suggested HTSAE-RCYP[7] technique: data conversion, null value removal, and data normalization. The proposed HTSAE-RCYP technique uses the SAE model, which predicts rice yield levels with accuracy, for the yield prediction process.In this paper, a prototype decision support system for projecting rice crop yield for the Indian state of Maharashtra is proposed.To predict crop yield, a larger dataset and a larger study area could be utilised with the suggested prototype. This will give the farmer a guide [8] to help with decision-making regarding the possible crop yield for a given climate scenario.

In order to predict rice production yield and explore the factors influencing rice crop productivity for multiple districts in the Indian state of Maharashtra, this study used neural networks [9]. The accuracy rate was 97.5 percent, with a sensitivity of 96.3 and a specificity of 98.1, according to the results. In order to forecast agricultural productivity, this study looked at a number of machine learning techniques, including supervised and unsupervised learning procedures. Classification strategies have been demonstrated to give very accurate estimates of multiple crop yields with a variety of features when compared to regression[10] and unsupervised learning approaches.

We have examined four classes in this paper for the leaf category classification. To identify the true disease affecting the afflicted plants, we employed deep learning techniques [11]. We put three different architectures into practice: VGGNet16, RenNet101, and AlexNet. The most accurate of these three is Alexnet. In our dataset, the training and testing accuracy of the AlexNet model is 92.35% and 85.27%[12] respectively. In order to identify plant diseases, this study reviews the most recent and cutting-edge methods currently in use.

3 Methodology

Using machine learning techniques, the methodology forecasts rice yields in Andhra Pradesh's Zones II and III. We initially preprocess the dataset by encoding categorical variables and handling missing values to make sure it is prepared for model training. All columns other than "Dist Code," "Year," "State Name," and "Dist Name" are defined as features (X), and "RICE YIELD (Kg per ha)" is the target variable (y).

The train_test_split method splits the dataset into training and testing sets to facilitate model evaluation. The three machine learning models that are defined are the Random Forest Regressor, Gradient Boosting Regressor, and Support Vector Machine Regressor with a radial basis function kernel. The Mean Squared Error between the expected and actual values is used to assess each model's performance on the test set after it has been trained using the training data.

Each model is also subjected to 5-fold cross-validation in order to assess its robustness and generalizability. The cross-validation metrics, such as the negative mean squared error (neg_mean_squared_error), are calculated. The accompanying Figure 1 provides a detailed explanation of the method used to forecast the yield of rice crops in Andhra Pradesh's Zones II and III.

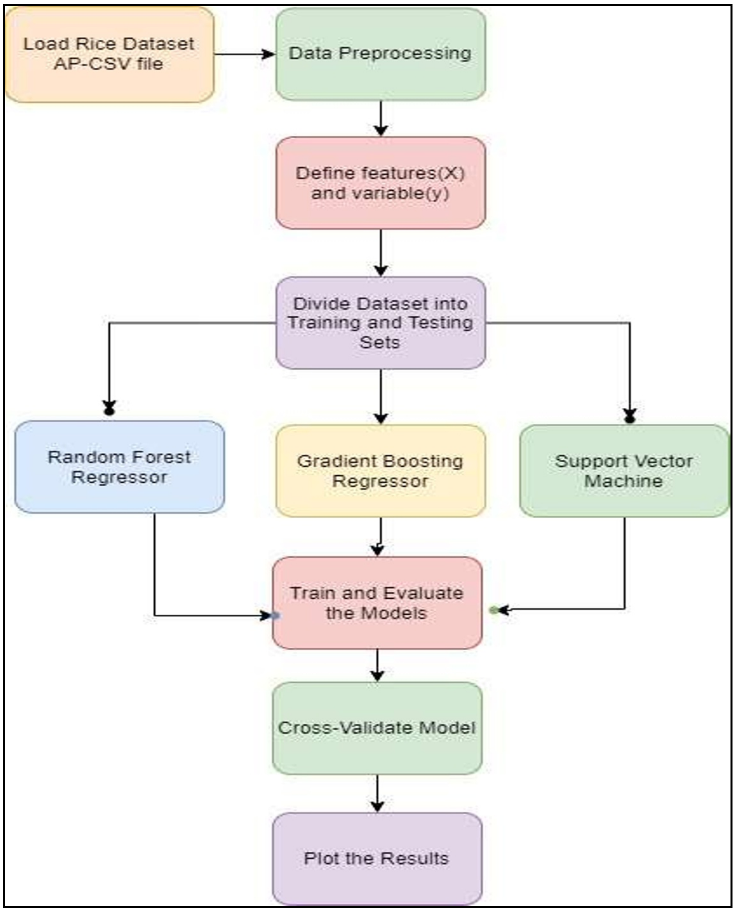

Fig. 1. Process of predicting crop yield prediction using machine learning models

4 Results

Plots showing the MSE for each model and the cross-validation MSE for comparison are used to visualize the results. By integrating machine learning methodologies with domain knowledge in agriculture, this research aims to provide valuable insights for farmers, policymakers, and agricultural stakeholders, facilitating informed decision-making and resource allocation to enhance agricultural productivity and sustainability in Andhra Pradesh's Zones II and III.

These measurements The machine learning models' performance in forecasting rice yields in Andhra Pradesh's Zones II and III is shown by Mean Squared Error (MSE) and Cross Validation MSE. The average squared difference between the target variable's (in this case, rice yields) expected and actual values is measured by MSE in this situation. The prediction performance is better when the MSE values are lower. A comparison of MSE in machine learning models is presented in Figure 2.

- Random Forest: MSE of 48.71 indicates relatively low error, suggesting that the Random Forest model performs well in predicting rice yields.

- Gradient Boosting: MSE of 98.77 indicates a slightly higher error compared to Random Forest but still suggests reasonable predictive performance.
- Support Vector Machine: A very high MSE of 542076.41 suggests poor predictive performance of the SVM model in this context.

Figure 3 shows Cross-validation MSE provides an estimate of the model's performance on unseen data and helps assess its generalization capability.

- Random Forest: Cross-validation MSE of 434.65 is lower than the test MSE, indicating good generalization performance.
- Gradient Boosting: Cross-validation MSE of 397.94 is also lower than the test MSE, suggesting good generalization.
- Support Vector Machine : The extremely high cross-validation MSE of 613584.29 further confirms poor generalization and predictive performance of the SVM model.

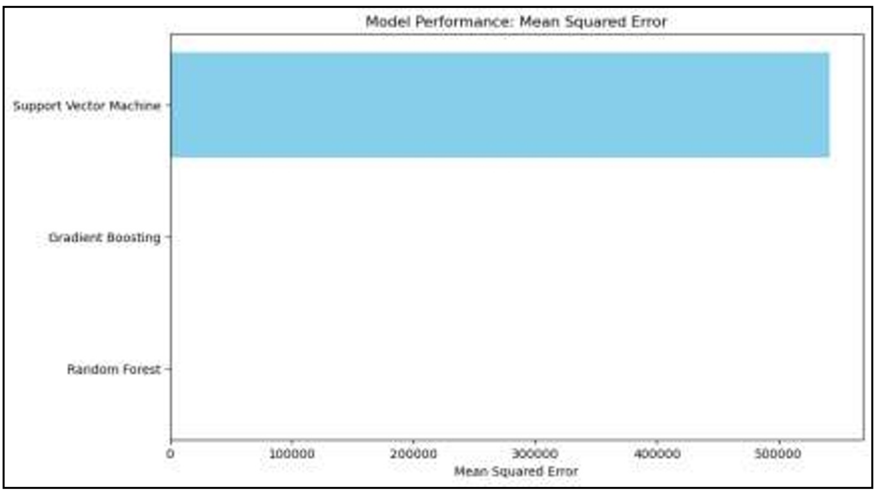

Fig. 2. Model Performance: Mean Squared Error

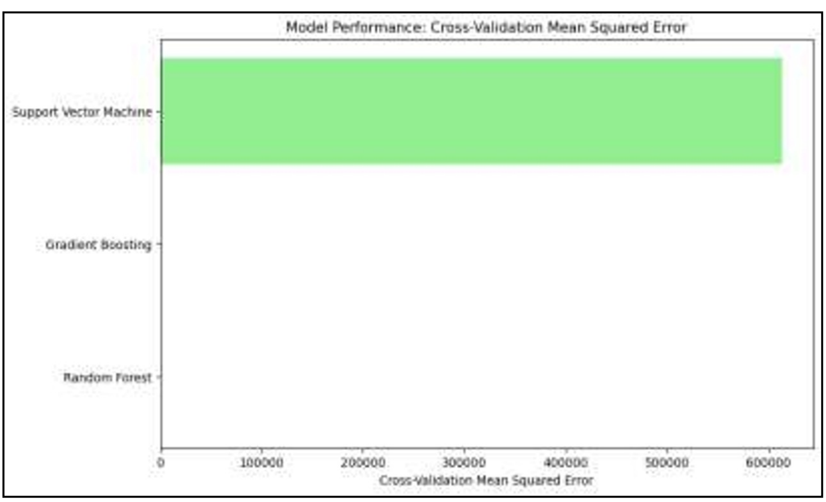

Fig. 3. Model Performance: Cross-validation Mean Squared Error

5 Conclusion and Future Work

Our research concludes by showing how well machine learning models can predict rice yield in Andhra Pradesh's Zones II and III. We have combined complex machine learning methods, such as Random Forest, Gradient Boosting, and Support Vector Machines, with copious amounts of environmental and agronomic data to create our robust predictive models. Our results demonstrate the possibility of making well-informed decisions and allocating resources in agricultural planning, which can enhance productivity and sustainability in the production of rice. To further help farmers and policymakers, future research could investigate the integration of real-time data and remote sensing technologies to improve the precision and timeliness of yield predictions.

References

1. N. Gandhi, L. J. Armstrong, O. Petkar, A. K. Tripathy, Rice crop yield prediction in India using support vector machines, *in Proceedings of the 13th International Joint Conference on Computer Science and Software Engineering*, IJCSSE,1-5, Khon Kaen, Thailand (2016)
2. S. Vashisht, P. Kumar, M. C. Trivedi, Improvised Extreme Learning Machine for Crop Yield Prediction, *in Proceedings of the 3rd International Conference on Intelligent Engineering and Management*, ICIEM, 754-757, London, United Kingdom (2022)
3. L. B. Rananavare, S. Chitnis, Crop Yield Prediction Using Temporal Data, *in Proceedings of the IEEE International Conference on Electronics, Computing and Communication Technologies*, CONECCT,1-5, Bangalore, India (2022)
4. K. Mariappan,J. A. Ben Das, A paradigm for rice yield prediction in Tamilnadu, *in Proceedings of the IEEE Technological Innovations in ICT for Agriculture and Rural Development*,TIAR,18-21, Chennai, India (2017)
5. P. C. Shaker Reddy, G. Suryanarayana, L. P. K,S. Yadala, Data Analytics in Farming: Rice price prediction in Andhra Pradesh, *in Proceedings of the 5th International Conference on Multimedia, Signal Processing and Communication Technologies*,IMPACT,1-5,Aligarh, India (2022)
6. T. Rahman, S. Aktar, Machine Learning Approaches to Predict Rice Yield of Bangladesh, *in Proceedings of the International Conference on Innovations in Science, Engineering and Technology*, ICISET, 329-333, Chittagong, Bangladesh (2022)
7. S. Thirumal, R. Latha, Automated Hyper parameter Tuned Stacked Autoencoder based Rice Crop Yield Prediction Model, *in Proceedings of the 7th International Conference on Trends in Electronics and Informatics*, ICOEI,14-18, Tirunelveli, India (2023)
8. N. Gandhi, L. J. Armstrong, O. Petkar, Proposed decision support system for Indian rice crop yield prediction, *in Proceedings of the IEEE Technological Innovations in ICT for Agriculture and Rural Development*, TIAR, 13-18, Chennai, India (2016)
9. N. Gandhi, O. Petkar, L. J. Armstrong, Rice crop yield prediction using artificial neural networks, *in Proceedings of the IEEE Technological Innovations in ICT for Agriculture and Rural Development*, TIAR, 105-110, Chennai, India (2016)

10. G.L. Sunil, V. Nagaveni, U. Shruthi, A Review on Prediction of Crop Yield using Machine Learning Techniques, *in Proceedings of the 2022 IEEE Region 10 Symposium*, TENSYMP, Mumbai,1-5, India (2022)
11. Bajpai, N. K. Tiwari, A. K. Tripathi, V. Tripathi, D. Katiyar, Early leaf diseases prediction in Paddy crop using Deep learning model, *in the Proceedings of the 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing*, PCEMS,1-6, Nagpur, India, (2023)
12. G. Singh , R. Singh, Rice Leaf Disease Prediction: A Survey, *in Proceedings of the 2023 International Conference on Inventive Computation Technologies*, ICICT,582-587, Lalitpur, Nepal (2023)