

A DATA-DRIVEN APPROACH FOR RICE YIELD ANALYSIS USING RANDOM FOREST

SREEKUMAR BISWAS^{1*}, SUPARNA SEN¹, RAHUL PRASAD¹, MITRA MONDAL¹,
SUDIPTO BISWAS¹, DHRUP KUMAR SINHA¹, AND PRAKASH DEY¹

This study presents an evidence-based approach for prediction of rice yield in India using Random Forest model trained on the UPAg dataset. The workflow includes data cleaning, restructuring, and feature engineering, followed by exploratory data analysis and hyperparameter tuning using GridSearchCV. The optimized model achieved R^2 value of 0.86, an RMSE of 288.13, and an MAE of 76.68, thus reflecting its strong predictive performance. Feature importance analysis revealed that the production-to-area ratio and previous-year yield (Yield_Lag1) were dominant predictors. This research demonstrates the effectiveness of ensemble learning in agricultural analytics, providing a scalable framework for data-driven crop yield forecasting.
