

Crop Yield Prediction Based on Indian Agriculture Using Machine Learning

Ayush Aravind

Department of Computer Science and Engineering, PES University, Bengaluru, India

Pooja M. N.

Department of Computer Science and Engineering, PES University, Bengaluru, India

Suhas P. H.

Department of Computer Science and Engineering, PES University, Bengaluru, India

H. N. Gagan

Department of Computer Science and Engineering, PES University, Bengaluru, India

Abstract:

This paper develops a machine-learning framework to estimate crop yield per hectare in different Indian districts based on open-agricultural data. The system combines data preprocessing, correlation-based and LASSO feature selection methods, and the performance evaluation of six regression algorithms: Ridge, Decision Tree, Random Forest, XGBoost, LightGBM, and CatBoost. The CatBoost model was able to perform well with an R^2 close to 0.93 and a low RMSE, demonstrating accurate yield prediction can be achieved using agro-economic and spatial features without explicit soil and weather data. Furthermore, the model is made available through a FastAPI backend along with a React-based multilingual web interface, which collectively offers live, easy-to-reach, and data-driven insights for farmers and policymakers.

Keywords:

Crop yield prediction, CatBoost, LightGBM, Machine learning, XGBoost.