Extreme gradient boosting (XGBoost) regressor and shapley additive explanation for crop yield prediction in agriculture

ABSTRACT

The primary purpose of precision agriculture is to maximize crop yields while utilizing a limited amount of land resources. Apart from industrialization, which fuelled Malaysia's significant economy and development, the country's agriculture industry performs a major role in guaranteeing food security and safety, as well as long-term development and wealth creation. To increase the nation's food security, policymakers must rely on accurate crop yield predictions in order to easily obtain trade - related evaluations. Machine Learning can help anticipate yields more accurately. This paper proposes to use the XGBoost model for annual crop yield prediction in Malaysia. Experiments on the generated yield dataset show promising results with 0.98 R-Squared value and outperformed the current models. The implementation of the suggested model is extensively evaluated using the Shapley Additive Explanation (SHAP) to discover the essential features such as average temperature, average rainfall, and pesticide in the crop yield prediction. The estimates provided by machine learning algorithms will aid farmers in deciding what to grow because of this research.