

Smart Crop Recommendation System Using Machine Learning Models

R.K.S.D.C. Senevinayaka, B.M.P.P. Amarakoon, and J. Tuvensha

Department of ICT, Faculty of Technological Studies, University of Vavuniya, Sri Lanka

Abstract

Crop recommendation systems have become essential tools in modern agriculture, assisting farmers in selecting the most suitable crops based on a combination of environmental and soil factors to optimize yield and resource efficiency. In this study, we developed a crop recommendation system capable of suggesting the best crop from 22 options tailored to the specific land and environmental conditions of a given area. Our dataset was sourced from Kaggle, a widely recognized open data platform, and underwent thorough preprocessing to clean and prepare the data. Additionally, we employed the Variance Inflation Factor (VIF) method to detect and eliminate multicollinearity among features, thereby refining the model's focus on the most impactful variables. While previous models in the literature have utilized various machine learning approaches such as logistic regression, decision trees, and support vector machines to predict suitable crops, many have reported moderate accuracies and often relied on fewer crop options or less comprehensive datasets. To enhance predictive performance, we experimented with multiple classifiers, including logistic regression, decision trees, random forests, k-nearest neighbors, support vector machines, and XGBoost and found that the random forest model significantly outperformed others, achieving an accuracy of 99.32%. This is notably higher than the accuracy levels reported in several prior studies, which typically range between 85% and 95%. Our findings highlight that combining rigorous data preprocessing techniques, feature selection through VIF, and advanced machine learning algorithms like random forests can lead to exceedingly reliable crop recommendations. This improved predictive capability underscores the potential of data-driven approaches to empower farmers with better decision-making tools, ultimately fostering higher yields and more sustainable agricultural practices.

Keywords: Variation Inflation Factor, random forests, k-nearest neighbors, support vector machines, and XGBoost