ABSTRACT

Chili (Capsicum annuum L.) is a strategic horticultural commodity, where fruit quality determines both economic value and cultivation success. However, pre-harvest quality assessment is still manual and subjective. This study aims to develop a pre-harvest chili fruit quality classification system based on vegetative plant data, including plant height, number of leaves, stem diameter, and observation week. The analysis follows the CRISP-DM methodology, which consists of six main stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The process includes data cleaning, feature engineering, and classification using machine learning algorithms. The main model applies a Two-stage XGBoost approach—a hierarchical classification system that first separates grade C from non-C, then distinguishes between grades A and B. As benchmarks, XGBoost Baseline and MLP Baseline models are also used. Evaluation results show that the Two-stage XGBoost outperforms with an accuracy of 0.83, macro F1-score of 0.82, and macro AUC of 0.86. The model was then implemented into a web application using Flask to facilitate real-time quality prediction. This study demonstrates that vegetative data-based classification effectively improves the accuracy of quality assessment and supports the digitalization of precision agriculture.

***Key words**: chili quality classification, machine learning, precision agriculture, vegetative features, XGBoost.*