
Identifying Key Predictors of Food Security in Mexico using Machine Learning Models

Abstract

The purpose of this paper is to measure the prevalence of food insecurity in Mexico by employing the National Survey of Household Income and Expenditure 2022 (in Spanish, ENIGH 2022) and the experience-based, Latin American and Caribbean Food Security Scale (in Spanish, ELCSA). It also draws on Perignon et al. (2023) revised Health Purchase Index (r-HPI), which evaluates the dietary diversity and nutritional quality of household food purchases. This study analysed 90,102 household records, providing a nationally representative sample. The Rasch Model was used to validate the ELCSA survey as a valid measurement tool. A logit regression model and three machine learning approaches — Least Absolute Shrinkage and Selection Operator (LASSO), Extreme Gradient Boosting (XGBoost), and Random Forest (RF) — were used to identify the predictors of food insecurity. The results indicate that the prevalence of moderate to severe food insecurity was 20.86%, or roughly 2.6 million people, and severe food insecurity was 3.87% or 497,682 people, based on Mexico's population in 2022 (128.6 million people). Statistical validation recommended by the Food and Agriculture Organization of the United Nations (FAO) showed that the validation tests of the ELCSA data were within the expected range for infit, outfit, reliability, and residual correlation. The tuned XGBoost model was selected as the preferred machine learning algorithm for predictions because it had the highest accuracy, F1 score, and AUC, and strong performance in precision and recall. The SHAP value analysis based on the tuned XGBoost model revealed that the main predictors influencing food security were log total income, education, household size, income earners, and age. The most vulnerable groups susceptible to food insecurity included household heads between the ages of 20 and 25, those over 55, those with low dietary quality, and those with a high ratio of food expenditure. It is concluded that ELCSA data is a valid measurement tool for assessing the prevalence of food insecurity in Mexico, and the tuned XGBoost can be used as a predictive machine learning (ML) model for food security. And therefore, based on the results and the supporting evidence from the literature, it is recommended that social policies related to health, education, and career development, particularly those involving cash transfers to the most vulnerable groups, be improved for their efficacy and quality in reducing food insecurity, in addition to their coverage.

Keywords: Food insecurity, Mexico, ELCSA, Household characteristics, Machine learning, XGBoost

Table 1. Models' performance statistics

Model	Accuracy	Precision	Recall	F1 Score	Area Under Curve
Initial RF Model	0.8157	0.5000	0.0230	0.0440	0.6879
Initial XGBoost Model	0.8119	0.8204	0.9851	0.8952	0.6619
Tuned RF Model	0.8158	0.8158	1.0000	0.8985	0.6961
Tuned XGBoost Model	0.8168	0.8177	0.9978	0.8988	0.7094
Tuned LASSO model	0.8149	0.8197	0.9910	0.8973	0.7017
Logistic Econometric model	0.8111	0.8164	0.9900	0.8949	0.7008

Source: ENIGH 2022 (INEGI). The tuned XGBoost model was selected as the preferred machine learning algorithm for predictions because it had the highest accuracy, F1 score, and area under curve, and strong performance in precision and recall.