Working papers economics - Forecasting Disaggregated Food Inflation Baskets in Colombia with an XGBoost Model

Download Keep in mind
The series Working Papers on Economics is published by the Office for Economic Studies at the *Banco de la República* (Central Bank of *Colombia*). It contributes to the dissemination and promotion of the work by researchers from the institution. This series is indexed at Research Papers in Economics (RePEc).

On multiple occasions, these works have been the result of collaborative work with individuals from other national or international institutions. The works published are provisional, and their authors are fully responsible for the opinions expressed in them, as well as for possible mistakes. The opinions expressed herein are those of the authors and do not necessarily reflect the views of Banco de la República or its Board of Directors.

AUTHORS AND/OR EDITORS Anzola-Bravo, César Poveda-Olarte, Andrea Paola
XGBoost models are more accurate in forecasting food inflation than linear models for most of the 33 baskets studied, especially for longer forecast horizons.

Publication Date: Thursday, 11 of December 2025 **Approach**
This article aims to develop statistical models to forecast monthly inflation for the next 12 months for 33 baskets that make up the Consumer Price Index (CPI) for food. To this end, both traditional time series models and Machine Learning approaches are employed. Each basket is modeled independently, incorporating four groups of explanatory variables relevant to food supply: climate variables, the nominal exchange rate, commodity prices, and transportation and energy costs. Commodity and energy prices influence food inflation by affecting production, transportation, and processing costs. Additionally, climate impacts agricultural production by altering crop growth, soil quality, and livestock health. Furthermore, the document seeks to interpret the forecasts using SHAP values (SHapley Additive exPlanations), a widely used tool to explain machine learning model predictions.

Contribution
This research contributes to the literature on food inflation forecasting by demonstrating that tree-based models such as XGBoost produce more accurate forecasts than linear time series models. Moreover, we show that it is possible to break down the predictions of this model into contributions from fundamental explanatory variables. The dynamics of each food basket respond to variables such as climate, commodity prices, transportation costs, and exchange rate behavior.

The interpretation of these model predictions is achieved through an exhaustive process of selecting explanatory variables and an algorithm we propose to select the optimal lags of these variables. This algorithm reduces the number of variables, simplifying interpretation and lowering computational costs.

Results
XGBoost models are more accurate in forecasting food inflation than linear models for most of the 33 baskets studied, especially for longer forecast horizons. Forecast errors from the XGBoost model were between 5% and 60% lower than those of linear models, depending on the basket and forecast horizon, and for the aggregate food inflation basket, errors were on average 25% lower.

There is a high degree of heterogeneity in the explanation of forecasts depending on the basket analyzed. For certain food groups, such as perishables, climate and inflation persistence are relevant factors. In contrast, other foods, such as industrial products, are mainly explained by commodity costs and inflation persistence.

Fuente: https://www.banrep.gov.co/en/publications-research/working-papers-economics/forecasting-disaggregated-food-inflation