Name: Seowoo Kim

Title: Dengue Fever Forecast in Peru - Insights for Public Health Policy

Introduction

Dengue fever is a growing public health concern in Peru, particularly in areas prone to outbreaks. This report aims to provide policymakers with a comprehensive analysis using a vector autoregression (VAR) model to forecast dengue case counts across Peruvian districts. The objective is to inform public health interventions and resource allocation for effective disease management.

Data

The dataset comprises 62,752 observations across 12 variables, including year, epidemiological week, dengue cases, and various environmental factors. It spans from the first week of 2001 to the 36th week of 2023 and is divided into 53 districts for localized analysis. There are no missing values in the dataset.

Method

A new variable was created to forecast dengue cases by combining the year and epidemiological week to establish a temporal pattern. The number of dengue cases was normalized using population data to calculate incidence rates. The training data covered weeks from the first week of 2001 to the 16th week of 2023. The dataset was examined for seasonal trends and adjusted where necessary. The VAR model was then fitted for each district, with a significance level of 0.05, to select the most optimal parameters and produce robust forecasts. The predicted incidence rates were converted to dengue case numbers and visualized on maps to highlight district-level trends. Model accuracy was evaluated using the Root Mean Square Percentage Error (RMSPE), excluding districts with zero cases.
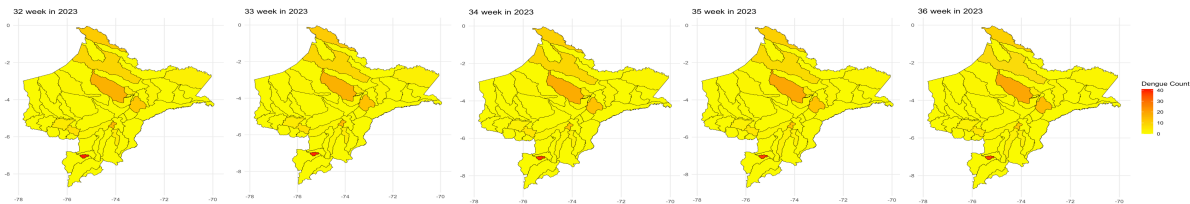
Result

Figure 1: Dengue Fever Count from the 32nd week (Left) to the 36th week (Right). Red indicates a higher count, and yellow indicates a lower one in each area. The forecasting model predicts that districts like 'Inahuaya,' 'Alto Nanay,' and 'Capelo' will remain hotspots for dengue fever, indicating that prioritized public health measures are necessary in these regions. Many other districts display low dengue counts, suggesting effective prevention or low risk.

District Name	RMSPE
Teniente Manuel Clavero	9.59
Mazan	3.57
Jeberos	0.15
Alto Nanay	0.02

Table 1: RMSPE values of districts (Partial List of 53 Districts). The average RMSPE across districts with non-zero cases is 1.19, which shows relatively more robust accuracy than other models in relevant papers. Anomalies were detected in the 'Napo' district, where the RMSPE was significantly higher due to post-2020 outliers. Adjusting the training data improved this district's forecast accuracy.

Discussion

These forecasting results reveal that immediate public health interventions should be prioritized in high-incidence areas like 'Inahuaya' to contain potential outbreaks. Although some districts show low dengue counts, preventive measures should be maintained to sustain their status. Further data collection may reveal additional factors influencing dengue incidence, providing a deeper understanding of the disease. RMSPE may not be suitable for districts with zero cases, and additional evaluation criteria should be considered. Future forecasting models could incorporate Conditional Auto-Regressive (CAR) models or neural networks to refine predictions and inform intervention strategies.