NiñoSalud AI

Climate, surveillance and reproducible evidence in northern Peru

Cristian Daniel Medina Gomez · Independent undergraduate-level project
Background in Economics and Management at Bocconi; interest in data science and health sciences.

Problem and research question

Public-health records and climate observations are useful only when their definitions and limitations are understood. NiñoSalud asks whether observed weather improves short-term prediction of reported diarrheal episodes in Piura, and how source-quality assumptions affect regional dengue indicators.

Prototype and method

An offline web dashboard combines official-source surveillance, weather and territorial context. The main experiment predicts one future week's all-age EDA count four weeks ahead in eight provinces. Two matched Random Forests and three simple baselines use 2019–2022 for training and 992 eligible predictions in 2023–2025 for exploratory evaluation. Python scripts and aggregate inputs are included.

Matched method	Mean absolute error (episodes / province-week)
Last observed week (persistence)	45.088
Random Forest + observed climate	50.294
Same Random Forest, without climate	52.763

Finding: climate reduces the forest's error by about 4.7% relative to the no-climate forest, but does not beat persistence. The approximate cluster-bootstrap interval for climate forest minus persistence MAE is −0.474 to +13.349 episodes; no superiority or equivalence claim is justified.

Source audit and scope

An older EDA extract conflicts with the newer CDC annual totals. In a separate 2025 regional dengue audit, Huancabamba's median recorded date interval ranges from 5 to 19 days across inclusion cutoffs; Piura remains at 3 days. This is sensitivity evidence, not proof of delayed care.

Limits: retrospective evaluation already inspected during development; missing climate fields; changing surveillance; unverified timing definitions; no prospective or clinical validation. Territorial snapshots are contextual, not historical predictors. No live alerts or 2027 epidemic forecast are generated.

Sources: CDC Perú/MINSA and GORE Piura supplied snapshots; CENEPRED, INEI and SUSALUD territorial context. Full provenance, methods and links are in the repository. **Reproduce: research/eda_climate_ablation.py** and the documented audit scripts.

Project direction, source collection and interface testing by the project owner; generative-AI assistance in code, analysis and writing. No institutional endorsement is claimed.