Dengue Fever Weekly Risk Alert (Thailand)

Team: *Group 8*
Course: Capstone FTL — Idea Proposal Submission

1) Project Idea

We will build a small system that gives weekly dengue risk alerts for each district/province in Thailand.
The tool looks at:

- **Weather**: rain, temperature, humidity, wet-spell days.
- **Behavior signals**: Google searches for “dengue,” “ไข้เลือดออก,” “ยุงลาย,” etc.
- **Past dengue cases** to check accuracy.

It outputs:

- A risk score (0–1) and class (Low / Medium / High) per area/week.
- A map + table with the Top-K high-risk districts.
- A short why-this-area note using simple explanations (top 3 drivers from SHAP).

Goal: help local teams act earlier (larval control, community notice, school clean-ups) with a clear, easy-to-read dashboard.

2) Relevance to SDGs

- **SDG 3 – Good Health & Well-Being:** early warning can cut cases and hospital load.
- **SDG 11 – Sustainable Cities & Communities:** supports city planning for vector control.
- **SDG 13 – Climate Action:** links climate patterns (rain/heat) to disease risk and helps climate-health adaptation.

3) Literature Examples (brief)

- **Climate and weather drivers.** Temperature, rainfall, and humidity consistently shape dengue transmission by affecting mosquito survival, biting rates, and viral replication. Recent regional studies in Southeast Asia show risk increases with warm temperatures and moderate rainfall (but can drop with extreme rain that flushes larvae), supporting week-scale warning horizons.

- **Large-scale climate signals (ENSO).** El Niño–Southern Oscillation (ENSO) modulates dengue mainly via temperature anomalies. Multicountry analyses and re-analyses (Americas + Asia) find higher dengue risk during El Niño-related warming and region-wide epidemic synchrony, making ENSO a useful seasonal leading indicator.

- **Search behavior as an early signal.** Internet search activity (e.g., Google Trends) correlates with dengue incidence and improves short-term forecasts, especially in urban contexts. Thailand-focused work has already combined Google Trends (Thai keywords) with meteorology to forecast weekly cases.

- **Thailand & regional dynamics.** Thailand shows province-level spatiotemporal synchronization of dengue over decades, which argues for province/district-granular models and time-aware validation. Province case series commonly come from the Ministry of Public Health’s Report 506 (weekly), with Thailand modernizing to a Digital Disease Surveillance (DDS) API in 2024.

4) Describe Your Data

Sources (open):

- **ERA5 / GPM (satellite) / TRMM**: daily rain, temp, humidity → aggregated to **weekly by district**.

- **Google Trends**: weekly relative search volume for dengue-related keywords (Thai + English).

- **Historical dengue cases**: provincial/district time series from Thai MOPH/WHO (where public).

Format & size: CSV/Parquet tables with columns: `date_week, province/district_id, rain_mm, temp_C, humidity_%, wetspell_days, search_dengue, …, cases` (if available). About **3–10 years** of weekly rows × ~**77 provinces** (or districts when possible).

Pre-processing:

- Align to **ISO week**; fill small gaps; create **lags** (e.g., rain at t-1..t-4), **rolling means/sums**, **wet-spell counts**, and **seasonality features** (week-of-year, holidays, ENSO index if used).
- Normalise features; train/validation **time-split** (walk-forward).

5) Approach (Machine Learning)

Why ML (not deep learning first): our data are **tabular + short time series**; **tree ensembles** are strong, fast, and **interpretable**.

Baseline pipeline:

1. **Models:** Logistic Regression (simple), **XGBoost/LightGBM** (main), Random Forest (backup).
2. **Target:** High-risk label per area/week (e.g., top 20% of historical cases or exceed a moving threshold).
3. **Training:** walk-forward cross-validation by week; province-aware splits to avoid leakage.
4. **Explainability: SHAP** to show top drivers each week/area.
5. **Evaluation:** AUROC/PR-AUC, F1 for High-risk, **Brier score**, and a **Top-K hit rate** ("Did we flag the districts that later spiked?").
6. **Output:** a **weekly map and table** plus a short reason line (e.g., "High rain last 2 weeks + high humidity + search spike").

Stretch goals:

- Add **LSTM/Temporal Fusion Transformer** if longer series become available.
- Add **entomology proxies** (NDVI, standing-water indices) from satellite images.
- Simple **SMS/LINE** alert for local partners.

Deliverables (MVP):

- One-page **dashboard** (map + Top-K list + reasons).
- Clean **CSV features** and code to re-run **weekly**.
- Short **user guide** for public-health staff.

References

1. Soneja, S., et al. (2021). *A review of dengue's historical and future health risk from a changing climate*. **Int J Environ Res Public Health**. https://pmc.ncbi.nlm.nih.gov/articles/PMC8416809/ PMC

2. Sugeno, M., et al. (2023). *Association between environmental factors and dengue in Lao PDR*. **BMC Public Health**. https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-023-17277-0 BioMed Central

3. Tian, Y., et al. (2025). *Rising dengue risk with increasing ENSO*. **Nat Commun**. https://www.nature.com/articles/s41467-025-63655-0 Nature

4. van Panhuis, W. G., et al. (2015). *Region-wide synchrony & traveling waves of dengue*. **Proc. R. Soc. B**. https://pmc.ncbi.nlm.nih.gov/articles/PMC4620875/ PMC

5. Johansson, M. A., et al. (2009). *Multiyear climate variability & dengue—ENSO*. **PLOS Medicine**. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000168 PLOS

6. Li, Z., et al. (2022). *Forecasting dengue by integrating Google Earth Engine, AI & GT*. **Int J Appl Earth Obs Geoinf** (open-access preprint). https://pmc.ncbi.nlm.nih.gov/articles/PMC9603269/ PMC

7. Puengpreeda, A., et al. (2020). *Weekly Forecasting Model for DHF in Thailand using Google Trends & meteorology*. **Engineering Journal**. https://engj.org/index.php/ej/article/view/3498