

SUBMITTED BY

Team GenomeX

Members:

Sahas Eashan

Nuwan Dhananjaya

Sanugi Wickramasinghe

EVENT

BioFusion Hackathon

Submission

MedFollow

Agentic Decision-Support for
Medication Follow-Up

January 3, 2026

Deliverables:

Notebook: GenomeX_Notebook.ipynb

Report: GenomeX_Report.pdf

1. Problem

Sri Lanka faces a sharp affordability and availability gap for medicines. Patients routinely buy only 2–3 days of a 7–14 day course or halt treatment early when they “feel okay.” That wastes scarce money, leaves unused doses that others could have consumed, and forces costly repeat doctor visits just to check progress. When symptoms improve only partially, people are unsure whether to continue, switch, or seek urgent care; the result is delayed recovery, avoidable complications, and financial strain on already tight household budgets. The guidance gap is widest for those juggling work, travel, and the cost of re-channeling a clinician every few days.

- **Affordability + supply:** High prices and shortages push patients to partial-course purchasing, creating incomplete treatments and inconsistent adherence.
- **Access burden:** Re-channeling a doctor every few days is expensive, time-consuming, and often infeasible outside major cities.
- **Outcome risk:** Stopping an effective drug too early or continuing an ineffective one delays recovery and can trigger complications.
- **Waste:** Partially used packs mean money lost and doses others could have used; disposal also raises safety and environmental concerns.
- **Uncertainty:** Patients cannot easily tell if “somewhat better” means continue, switch, or see a doctor immediately.

2. Solution

MedFollow is an agentic, transparent follow-up assistant designed for this reality. It reads patient-reported baseline and current symptoms, combines them with evidence from drug-review data, and returns a conservative recommendation: continue medication, schedule follow-up, or seek urgent care. The goal is to cut unnecessary clinic visits while protecting safety and explaining why a decision was made. Each decision is paired with rationale, thresholds, and similar-case evidence so patients and clinicians can see the basis for the recommendation.

- **Inputs:** Baseline symptoms, current symptoms, medication, condition, adherence, and red-flag terms (e.g., chest pain, swelling).
- **Decision path:** Symptom change signals + similarity search + class-weighted model + stability and safety checks, tuned to avoid harmful misses on the `not_effective` class.
- **Outputs:** A clear action (continue / follow-up / urgent), calibrated class probabilities, similar cases, and integrity flags that highlight any uncertainty.
- **Safety posture:** Conservative defaults; if uncertainty, instability, OOD, or red flags appear, route to follow-up or urgent care rather than over-confident guidance.
- **Access benefit:** Reduces the need for immediate re-channeling when symptoms are clearly improving, saving cost and time while still surfacing red flags quickly.

3. Data

- **Source:** Drug Reviews (Druglib.com), UCI ML Repository (CC BY 4.0).
- **Access:** Public URL "<https://archive.ics.uci.edu/static/public/461/data.csv>".
- **Fields:** reviewID, urlDrugName, rating, effectiveness (target), sideEffects, condition, benefitsReview, sideEffectsReview, commentsReview.
- **Repo:** <https://github.com/sahas-eashan/bio-FUSION>

4. Method

- **Model:** Word + char TF-IDF features feeding class-weighted logistic regression; thresholds tuned (e.g., $t_{\text{not}} = 0.25$, $t_{\text{eff}} = 0.55$) to protect the `not_effective` class. Class weights are critical in a medical setting: we strongly penalize predicting “effective” when the true label is “not_effective,” because that mistake could keep patients on a failing therapy.
- **Parsing:** Gemini (`models/gemini-flash-latest`) used only to structure user text (baseline/current symptoms, med, condition, adherence, severe flags); classification remains local.
- **Explainability:** Class probabilities, similar-case retrieval (TF-IDF cosine), and safety prompts shown with each decision.
- **Reproducibility:** Seeds set; train/val/test split stratified on labels; environment captured in notebook.

5. Agentic Safety Stack

- **Consistency checks:** Ask overlapping questions (start date, first dose date, total doses) and down-weight confidence if answers conflict.
- **Plausibility / range:** Flag impossible vitals or logical conflicts (e.g., $\text{temp} > 45\text{C}$, “fully recovered” + pain 9/10, “0 doses” + “side effects after taking it”).
- **Symptom progression logic:** Improved + new severe symptoms \Rightarrow urgent review; worsening on-meds \Rightarrow follow-up/clarify.
- **Stability / robustness:** Check if small text changes flip the decision; if unstable, route to follow-up.
- **Anomaly / OOD:** If input is far from training distribution, lower confidence and recommend follow-up.
- **Red flags:** Chest pain, dyspnea, rash with swelling, bleeding, fainting, high fever \Rightarrow seek urgent care.

6. Results

- **Metrics:** Confusion matrices (Figures 1, 2) show baseline recall for `not_effective` around 0.43, improved to ≈ 0.88 after class-weight tuning and thresholding; trade-offs appear as lower `effective`/`mixed` accuracy, but safety risk is reduced for missed failures. Fill in accuracy/F1/ROC-AUC and calibration from the notebook run.
- **Thresholds:** Notebook uses $t_{\text{not}} = 0.25$ and $t_{\text{eff}} = 0.55$ to protect against false “effective” on truly `not_effective` cases.
- **Error analysis:** Residual confusions are mainly `effective` vs `mixed`; add calibration and richer symptom-change features to improve without hurting `not_effective`.

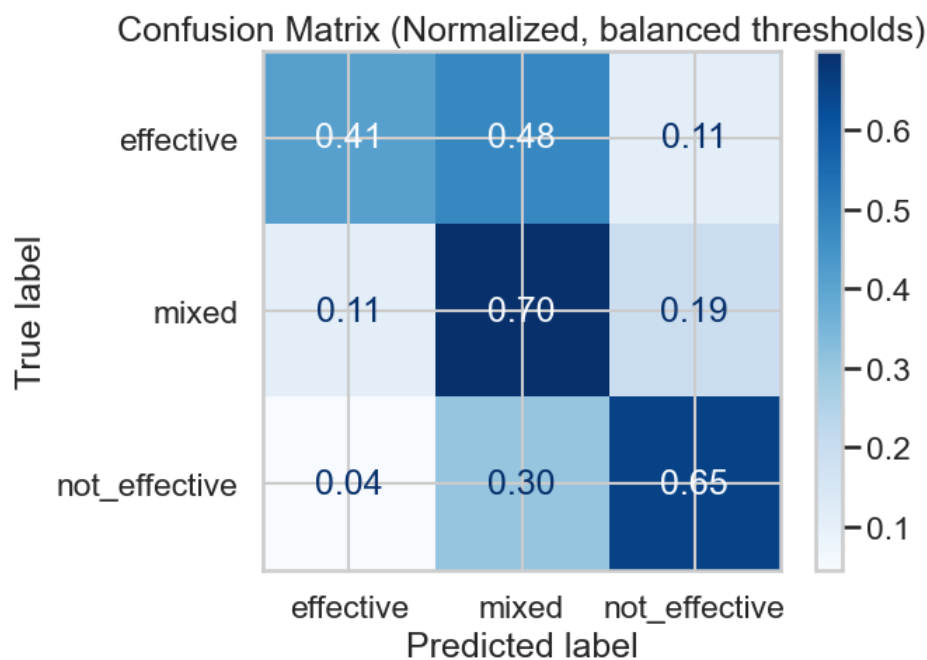


Figure 1: Normalized confusion matrix after class-weight tuning; note improved recall for `not_effective`.

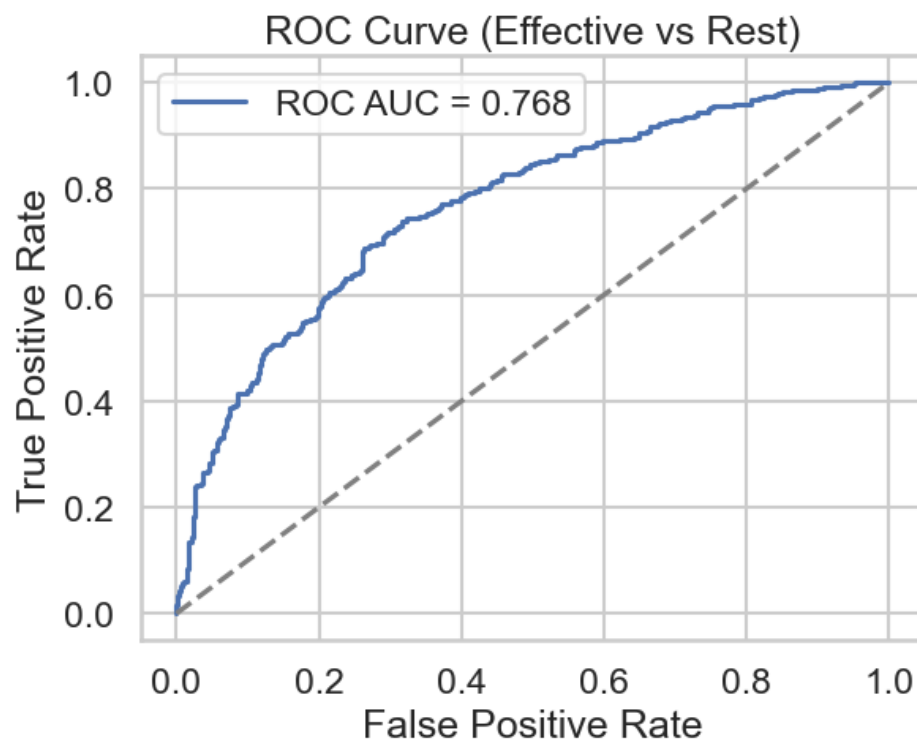


Figure 2: Alternate thresholding view; balances gains in `not_effective` while keeping `effective` and `mixed` reasonable.