Milestone 5 - Task 2: Monitoring and Maintenance Plan

CRISP-DM Phase 6 - Plan Monitoring and Maintenance

Group A

YYYY-MM-DD

Table of Contents

# Executive Summary

This Monitoring and Maintenance Plan defines how the deployed model will be observed, evaluated, and updated to ensure reliable and ethical performance in day‑to‑day use. It covers input validation, performance monitoring, drift detection, fairness checks, retraining triggers, rollback, governance, dynamic aspects, sunset criteria, and alignment to the project’s original business objectives.

Scope alignment: - Inference uses the engineered feature schema from training (Milestone 3) and does not process PII. - Monitoring references the hold‑out and engineered datasets stored under 02\_Project/Data/ and the model artifact under 02\_Project/Milestone\_3/Task\_03/outputs/. - Cadence and thresholds follow the Evaluation decisions from Milestone 4 and acceptance criteria in Milestone 4 Task 3 (Next Steps).

Key outcomes: - A monthly performance and drift review with clear triggers for action. - Documented retraining and rollback procedures. - Explicit criteria for when to pause or retire the model.

# Data Input Checks

Objective: Prevent invalid or out‑of‑scope inputs from reaching the model and ensure inference schema parity with training.

Authoritative sources: - Training schema: 02\_Project/Data/04\_Split/train\_data.csv - Feature roles: 02\_Project/Data/03\_Scaled/feature\_metadata.csv

Checks (at app startup and per request where applicable): - Schema parity: All required feature columns exist; no unexpected columns used for prediction. - Data type checks: Numeric vs categorical fields conform to expected types. - Range checks: Numeric inputs fall within robust training bounds (1st–99th percentile by feature) derived from train\_data.csv. - Allowed values: Categorical fields (e.g., sample\_size\_tier) come from the set present in train\_data.csv. - Missingness: No NA values passed to prediction; defaults applied for any optional fields.

Actions on failure: - Reject with a user‑friendly message and log a validation error event. - Do not attempt silent auto‑correction beyond the documented defaults.

# Performance Monitoring

Objective: Track predictive quality over time and detect degradation early.

Metrics (log‑scaled target): - RMSE, MAE, R² for value\_log\_scaled.

Baselines and datasets: - Hold‑out baseline: 02\_Project/Data/04\_Split/test\_data.csv (kept untouched for comparison). - Live/monitoring dataset: append periodic samples from new usage or refreshed data to a dated file (e.g., 02\_Project/Data/05\_Monitoring/monitoring\_YYYY\_MM.csv).

Cadence and targets (aligned to Milestone 4 Task 3 acceptance criteria): - Monthly review or after ≥ 100 new scored records (whichever comes first). - Targets: RMSE ≤ 0.06, MAE ≤ 0.04, R² ≥ 0.95 on a representative monitoring sample.

Actions on breach: - Single breach (minor): Investigate data quality and drift; increase sampling; prepare retraining plan. - Consecutive breaches (2 cycles) or large deviation: Trigger retraining workflow and consider temporary rollback.

# Drift Detection

Objective: Detect changes in input distributions and output behavior that may invalidate prior model assumptions.

Input drift (feature‑wise): - Compare monitoring sample vs. train\_data.csv using distributional tests (e.g., PSI). - Rules of thumb: PSI 0.1–0.2 (small), 0.2–0.3 (moderate), > 0.3 (significant).

Output drift: - Track prediction mean/variance and calibration vs. recent ground truth (when available). - Sudden shifts or systematic bias indicate potential concept drift.

Actions on drift: - Moderate input drift or stable output drift: watch and schedule ablation checks. - Significant input drift (PSI > 0.3) or deteriorating outputs: trigger retraining workflow.

# Fairness and Ethics

Objective: Ensure performance consistency across key subgroups and uphold ethical use.

Candidate slices (if available in engineered features or linked metadata): - Geographic or survey categories (e.g., survey\_cohort, dataset\_source\_encoded, or a geographic proxy field). - Sample size tiers (e.g., sample\_size\_tier).

Slice metrics: - Report RMSE/MAE by slice; flag gaps where |MAE\_slice − MAE\_overall| > 0.02 (tunable threshold).

Actions on disparity: - Investigate data representation; consider reweighting or feature updates. - Escalate to stakeholders if disparities persist for 2 cycles; consider additional model constraints or usage guidance.

Privacy: - No PII; only aggregate slice metrics reported externally.

# Retraining and Rollback

Objective: Provide a controlled, auditable path to refresh or revert the model.

Retraining triggers (any of): - Performance criteria breached in 2 consecutive cycles. - Significant input drift (PSI > 0.3) for key features. - Material data or domain change (see Dynamic Aspects). - Availability of substantial new high‑quality data.

Retraining workflow (high‑level): - Freeze current monitoring period; snapshot new training data partition. - Rebuild pipeline consistent with Milestone 3 (train only fit for transformations). - Use cross‑validation for hyperparameter tuning; reserve the test set for final evaluation only. - Compare against baseline/simple models to confirm value‑add. - Produce an evaluation report and update the model registry (artifact + metadata).

Versioning and promotion: - Store artifacts under 02\_Project/Milestone\_3/Task\_03/outputs/ with semantic versioning (e.g., final\_random\_forest\_model\_vX.Y.Z.rds). - Promote to production only after acceptance tests and stakeholder sign‑off.

Rollback procedure: - Keep N‑1 stable artifact and a simple switch (configuration or environment variable) in the app. - If severe degradation or defects occur, revert to N‑1 and log a rollback incident report.

# Governance and Audit

Objective: Ensure transparency, accountability, and reproducibility.

Record‑keeping: - Decision log: key changes, rationales, and approvals (e.g., 02\_Project/Logs/decision\_log.md). - Changelog: app and model updates (e.g., 02\_Project/Logs/changelog.md). - Monitoring reports: monthly summaries under 02\_Project/Data/05\_Monitoring/Reports/.

Ownership: - Product owner, data lead, and ops contact documented on the first page of each monitoring report.

Approvals: - For promotion or rollback, record approvals from the product owner and data lead.

# Dynamic Aspects (What Can Change)

Anticipated changes and hooks: - Data refresh cycles and coverage (e.g., new survey waves). - Indicator definitions or feature engineering logic. - Geographic boundaries or population shifts. - Stakeholder priorities or business objectives.

Mitigations: - Document each change; assess impact on schema and metrics; schedule retraining if impact is material.

# Sunset Criteria (When Not To Use)

Stop‑use rules (any of): - Performance below thresholds for 2 consecutive cycles despite remediation. - Major domain or indicator definition change invalidates prior modeling assumptions. - Ethical concerns (material, unresolved subgroup disparities).

Actions on sunset: - Notify stakeholders; disable access to the app; document reasons; open a new CRISP‑DM cycle if appropriate.

# Business Objectives Evolution

Restate initial problem (Milestone 1): Predict health outcomes to support policy intervention with interpretable, robust models under data constraints.

Review cadence: - Quarterly review of objectives and thresholds with stakeholders. - Update acceptance and monitoring thresholds as needed; document changes in the decision log.

# Implementation Plan (Task 2 Scope)

1. Create monitoring folders (02\_Project/Data/05\_Monitoring/ and .../Reports/).
2. Define a monthly schedule for metric computation and drift checks; assign ownership.
3. Produce the first monitoring baseline report using the existing test set.
4. Draft retraining SOP and rollback checklist.
5. Set up the decision log and changelog locations.

# References

* Milestone 5 brief: 02\_Project/Milestone\_5/Milestone\_5.md
* Milestone 4 evaluation: 02\_Project/Milestone\_4/Task\_2.md, 02\_Project/Milestone\_4/Task\_3.md
* Data folders: 02\_Project/Data/03\_Scaled/feature\_metadata.csv, 02\_Project/Data/04\_Split/train\_data.csv, .../test\_data.csv