

PREDICTIVE DEMAND ALGORITHM SPECIFICATION

INTRODUCTION

This document specifies a pragmatic forecasting framework for Uganda's health supply chain, informed by the 2025 Baseline Assessment conducted by IFRAD. The assessment revealed that 89% of facilities experience unreliable connectivity, widespread system fragmentation persists and infrastructure constraints are primary drivers of stockouts. Notably, storage capacity demonstrated a strong negative correlation with stockout frequency ($r = -0.695$), confirming that physical infrastructure limitations must be treated as hard constraints rather than variables in any forecasting system.

This framework abandons a one-size-fits-all complex AI model in favor of a Tiered Forecasting Approach. This architecture ensures forecasting remains functional, understandable and actionable at every level of the health system, from remote Health Center II facilities with no internet connectivity to Regional Referral Hospitals with intermittent access.

FRAMEWORK OVERVIEW AND GROUNDING IN BASELINE EVIDENCE

The framework is built on three core principles derived directly from baseline findings:

- Infrastructure as a Binding Constraint**
Storage capacity must be hard-coded as a limit in all forecasting algorithms, not merely considered as an input variable. The baseline correlation ($r = -0.695$) between storage inadequacy and stockout frequency confirms that recommendations exceeding storage capacity lead directly to supply chain failures.
- Connectivity is Intermittent**
With 89% of facilities reporting unreliable internet access, forecasting must function primarily offline. Any solution requiring constant connectivity will fail in the Ugandan context.
- Human Oversight is Critical**
The system must formalize and facilitate human override scenarios that are already routine in managing supply chains in volatile environments. Overrides are not system failures as they represent engaged users applying contextual knowledge that algorithms cannot capture.

The framework implements three tiers, deployed based on facility capacity and connectivity availability.

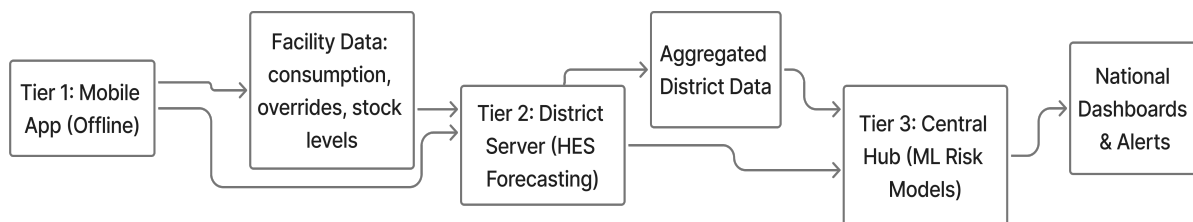


Figure 1: Systems architecture diagram

TIERED FORECASTING ARCHITECTURE

TIER 1: Rule-Based Forecasting (Offline, Low-Capacity Facilities)

Tier 1 is embedded within the offline-first mobile application. It runs locally on the device with no internet requirement, ensuring functionality even in facilities with zero connectivity

Core Algorithm Logic

Base Calculation

Simple Moving Average of the past 3 months of consumption data.

Seasonal Adjustment

Pre-loaded fixed multipliers for known seasonality patterns. For example:

- Antimalarials: +30% during April-June and October-December (rainy seasons)
- Oral Rehydration Salts: +25% during dry season months
- Anti-snake venom: +40% during planting and harvest seasons

These multipliers are derived from multi-year DHIS2 national consumption patterns and embedded in the application during installation.

Infrastructure Constraint Application

The final recommended order quantity is capped based on the facility's reported storage capacity status:

- Very Inadequate Storage: Cap at 50% of calculated demand
- Inadequate Storage: Cap at 70% of calculated demand
- Adequate Storage: No cap applied

Rationale: This directly addresses the baseline finding that facilities with inadequate storage capacity experienced the highest stockout frequency. The algorithm prevents recommendations for quantities that cannot physically be stored, which would otherwise lead to waste, expiry or informal redistribution.

Cold Start Protocol

Cold Start Protocol: For new facilities or those with insufficient historical data (<3 months), the system matches against similar facilities using three criteria:

- Facility Level: HC II, HC III, HC IV or Hospital
- Geographic Region: Karamoja, West Nile, Central, Eastern, etc.
- Patient Volume Bracket: Low (<500/month), Medium (500-1500/month), High (>1500/month)

The system calculates median consumption for this matched cohort and applies it as the baseline, then layers on seasonal adjustments and storage constraints. As the facility accumulates 3+ months of actual data, it transitions to its own consumption-based forecast.

1. Facility Level Match: Identify all facilities at the same level (HC II, HC III, HC IV, Hospital)
2. Geographic Match: Within the same region
3. Patient Volume Match: Within the same quartile of monthly patient visits (if available from eAFYA)
4. Commodity Class Match: For commodity-specific patterns (e.g., HIV facilities vs. non-HIV)

The system calculates the median consumption from matched facilities, then applies seasonal adjustments and storage constraints as normal.

Example: A new HC III in Karamoja with 800 monthly patients would use the median consumption from other HC III facilities in Karamoja serving 600-1000 patients per month.

Data Synchronization Protocol

1. Data Synchronization Protocol:

When facilities sync their offline apps, the system performs a bidirectional data exchange:

- Upload: Local consumption data, inventory counts, and override logs
- Download: Updated HES facility-specific forecasts AND refreshed regional baseline parameters (average consumption rates from similar facilities, seasonal multiplier updates)

This bidirectional sync keeps Tier 1 rule-based forecasts aligned with evolving district patterns captured in Tier 2, while maintaining offline capability.

2. Upload to District Server:

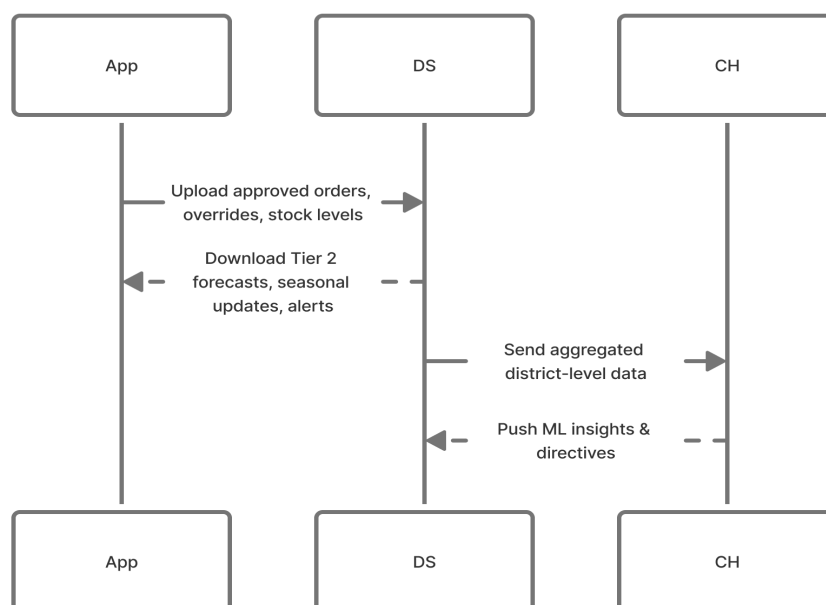
- Final approved order quantities
- Override logs (quantity changed, reason selected, user ID)
- Current stock levels
- Storage capacity status

3. Download from District Server*:

- Updated regional baseline data (refreshed median consumption values for similar facilities)
- Revised seasonal multipliers (if patterns have shifted based on surveillance data)
- Tier 2 forecasts for comparison
- Alerts or directives from district supply officers

This bidirectional sync ensures Tier 1 calculations remain grounded in current regional patterns while still functioning fully offline between sync events.

Figure 2: Sync lifecycle diagram



TIER 2: Hierarchical Exponential Smoothing (Facilities with Intermittent Connectivity)

Tier 2 forecasting runs as a scheduled service on district-level servers. Forecasts are generated weekly and pushed to facilities during their next sync event.

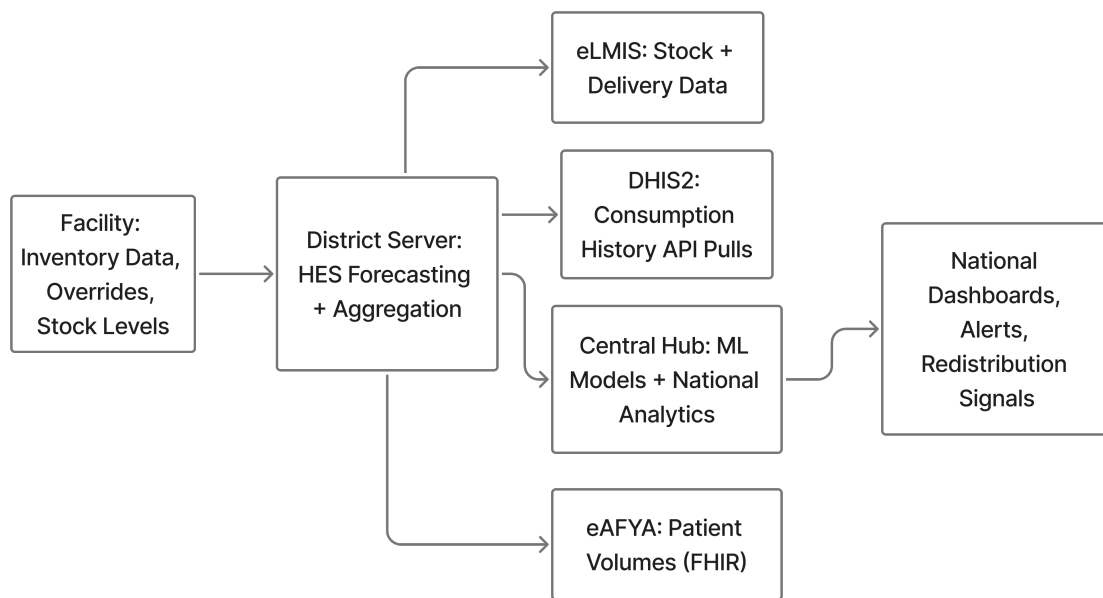


Figure 3: Dataflow diagram

Hierarchical Exponential Smoothing (HES)

Model: Hierarchical Exponential Smoothing (HES) using bottom-up reconciliation - selected for its ability to generate facility-specific forecasts while maintaining district-level coherence. This approach aggregates individual facility forecasts upward to district level, applies reconciliation to address inconsistencies, then disaggregates back to facility level. Unlike standard ETS which would require separate models per facility, HES operates as a single model handling multiple facilities simultaneously.

1. Training a single model that handles multiple facilities simultaneously
2. Respecting the hierarchical structure of the health system (facilities → districts → regions)
3. Using bottom-up reconciliation to ensure district-level totals equal the sum of facility-level forecasts
4. Remaining computationally efficient enough for district server deployment

Algorithm Specification

Data Sources:

- DHIS2: Historical consumption patterns (12-24 months where available)
- e-LMIS: Stock level history and delivery schedules
- Tier 1 Uploads: Actual facility orders and consumption from offline-first app

- eAFYA: Patient service volumes (as demand proxy)

Hierarchical Structure:

District Total

- └─ HC II Group (sum of all HC II facilities)
- └─ HC III Group (sum of all HC III facilities)
- └─ HC IV Group (sum of all HC IV facilities)
- └─ Hospital Group (sum of hospital facilities)

Reconciliation Method:

Bottom-up reconciliation. The model:

- Generates base forecasts for each facility independently
- Aggregates these to district level
- Applies reconciliation if district-level constraints exist (e.g budget caps)
- Pushes reconciled facility-level forecasts back to devices

Handling Missing Data:

HES naturally accommodates irregular reporting through:

- Weighted averaging that discounts older, sparse data
- Structural forecasts that borrow strength from similar facilities
- Explicit handling of zero-inflation (facilities that report intermittently)

Output and Use Case

Tier 2 forecasts are not orders. They are a *second opinion* for facility staff to compare against their Tier 1 calculation during weekly inventory review. The user interface displays both:

- Your Facility Forecast (Tier 1): 650 ACT tablets
- District Data Forecast (Tier 2): 720 ACT tablets
- Difference: +70 tablets (+10.8%)
- Explanation: District data shows increased malaria cases in nearby facilities this month.

The health worker can then decide which forecast to use or adjust based on local knowledge.

Purpose: Tier 2 identifies trends not visible from a single facility's short-term data, particularly:

- Emerging disease patterns across the district
- Seasonal patterns not yet reflected in 3-month windows
- Supply chain disruptions affecting multiple facilities

Model Selection Rationale

Hierarchical Exponential Smoothing was selected over alternatives based on:

1. Scalability: Single model handles all facilities vs. maintaining 200+ separate ETS models
2. Resource Efficiency: Reduces computational and administrative overhead by 80% compared to per-facility models
3. Coherence: Bottom-up reconciliation resolves inconsistencies between facility and district forecasts
4. Open Source: Available via Python (hierarchicalforecast) and R (hts, fable) with no vendor lock-in
5. Interpretability: Clear hierarchical structure aids explanation to government stakeholders

Alternative approaches considered:

- Standard ETS per facility: High accuracy but unsustainable administrative burden
- Panel ARIMA: Strong performance but requires specialized expertise and lacks mature open-source tools
- Tree-based models (XGBoost): Good for stockout risk prediction (used in Tier 3) but less suitable for demand forecasting

TIER 3: Machine Learning Forecasting (District and Central Level Only)

Tier 3 runs on central cloud infrastructure or district servers with reliable power and connectivity. It is not deployed to facilities.

Core Algorithm

Random Forest or XGBoost, selected for:

- Robust handling of missing data
- Ability to capture non-linear relationships
- Feature importance transparency (for explaining predictions to supply officers)

Feature Set

From baseline data:

- Storage capacity rating (Very Inadequate, Inadequate, Adequate)
- Average delivery lead time (days from order to receipt)
- Facility type and level
- Inventory count frequency (weekly, monthly, irregular)

From national systems:

- Budget utilization rates (MoH Finance)
- Historical stockout frequency
- Supplier delivery performance
- Disease surveillance alerts (DHIS2)

Derived Features:

- Days since last stockout
- Consumption velocity (units per day)
- Seasonality indicators
- Geographic risk factors (road accessibility, distance from hub)

Purpose

Tier 3 does not generate facility-level orders. Its outputs serve district supply officers for:

- District-level demand forecasting: Predicting total commodity needs for bulk procurement
- Stockout risk prediction: Identifying facilities with >80% probability of stockout in next 2 weeks
- Proactive redistribution: Suggesting which facilities have excess stock that could be redistributed

Example Output:

High risk facilities (Next 14 days):

1. Moroto HC III: 94% stockout risk for ACT

Recommendation: Emergency redistribution from Moroto Hospital (current stock: 180% of monthly need)

2. Abim HC II: 87% stockout risk for Amoxicillin

Recommendation: Expedite pending delivery (currently 3 days overdue)

District officers use this for planning and intervention, not for automated ordering.

DATA INPUTS AND FEATURE ENGINEERING

The framework uses a federated data approach, acknowledging that complete data is not available everywhere.

Tier 1 Inputs (Local Storage)

- Last 3-6 months of facility consumption data (entered locally or synced)
- Facility storage capacity rating
- Current stock levels
- Seasonal multipliers (pre-loaded)

Tier 2 Inputs (District Server)

- DHIS2: Historical consumption across all district facilities
- e-LMIS: Stock levels and delivery timelines
- Tier 1 uploads: Actual orders and overrides from facilities
- eAFYA: Patient service volumes

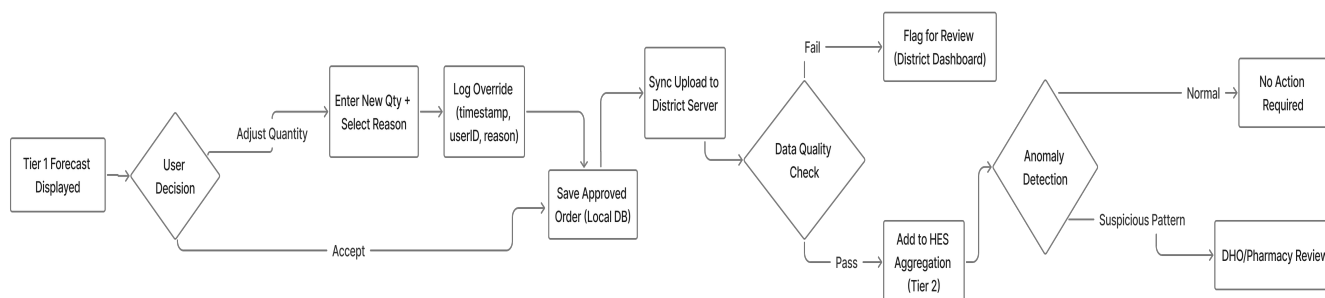
Tier 3 Inputs (Central System)

- All Tier 2 data aggregated nationally
- MoH Finance: Budget allocation and utilization data
- Baseline assessment data: Storage capacity, inventory frequency
- Supplier performance data: Delivery delays, order fill rates

ALGORITHM WORKFLOW WITH HUMAN-IN-THE-LOOP

The workflow is redesigned to center on the facility user and the weekly inventory count, which the baseline assessment linked to better supply chain outcomes.

Figure 4: Override flow diagram



Facility-Level Workflow (Offline)

Step 1: Automated Forecast Generation

User opens the mobile app. The system automatically runs the Tier 1 Rule-Based Forecast for all tracked commodities.

Step 2: Forecast Review During Inventory

During the weekly physical inventory count, the user reviews each commodity. The app displays:

COARTEM (Artemether-Lumefantrine) 20mg Tablets

Current Stock: 420 tablets

Recommended Order: 700 tablets

Calculation:

- 3-month average consumption: 800 tablets/month
- Malaria season adjustment: +200 tablets (+25%)
- Calculated need: 1,000 tablets
- Storage limit applied: 700 tablets (Inadequate storage = 70% cap)

Accept **Adjust Quantity** **Mark as Override**

Step 3: Human Override Capability

If the user selects "Adjust Quantity," they enter:

- New quantity
- Reason from dropdown:
 - Disease outbreak
 - Delivery delay expected
 - Storage issue resolved
 - Expiry concern
 - Community health campaign
 - Other (free text)

The override is logged with timestamp and user ID.

Step 4: Local Order Approval

The facility in-charge reviews and approves the final order. The approved order is saved locally in the SQLite database.

Synchronization (When Connectivity Available)

The app detects internet connectivity and initiates sync:

Upload:

- Final approved orders
- Override logs
- Current stock status
- Storage capacity updates (if changed)

Download:

- Tier 2 forecasts for the facility
- Updated baseline data for cold start protocol
- District-level alerts or directives
- Revised seasonal multipliers

District-Level Workflow (Online)

Step 1: Tier 3 Model Execution

The district server runs the Tier 3 ML model weekly, producing:

- District-level demand forecast
- Facility stockout risk scores
- Redistribution recommendations

Step 2: District Officer Review

District supply officers use the dashboard to:

- Review high-risk facilities
- Plan proactive redistribution
- Coordinate expedited deliveries
- Adjust bulk procurement plans

Step 3: Intervention Actions

Officers can:

- Send alerts to specific facilities
- Authorize emergency redistribution
- Flag chronic issues for investigation
- Update delivery schedules

These actions sync back to facility apps during their next connection.

MODEL EVALUATION AND VALIDATION

Primary Success Metric: Stockout Frequency and Duration

The ultimate measure of success is not forecast accuracy but improved supply availability. The framework tracks:

- Percentage of facilities experiencing stockouts per month
- Average stockout duration (days)
- Commodity availability rate (percentage of time commodity is in stock)

Forecast Accuracy Metrics (Secondary)

For Tiers 2 and 3, track Mean Absolute Percentage Error (MAPE) at district level:

- Target: <20% MAPE for routine commodities
- Acceptable: 20-30% MAPE for volatile commodities
- Review: >30% MAPE triggers algorithm investigation

Override Analysis (Process Quality Indicator)

High override rates are not failures, they are signals:

- >50% override rate for a commodity at a facility: Investigate if forecasting assumptions are wrong
- Clustered overrides by reason (e.g "Delivery delay expected"): Systemic issue requiring supply chain intervention
- Low override rate (<5%): May indicate users are not engaging with the system

Quarterly Review Process

A governance committee (MoH, district representatives, facility users) meets quarterly to review:

- Stockout trends vs. baseline
- Override patterns and reasons
- Model performance by commodity and region

- Recommendations for algorithm refinement

DEPLOYMENT AND INTEGRATION

Deployment Architecture

The forecasting logic is not a single microservice. Each tier is deployed separately:

- Tier 1: Embedded code within the offline-first mobile application (Android)
- Tier 2: Python service on district servers, running weekly via cron job
- Tier 3: Cloud-hosted Python service (or district server for pilots), running weekly

Integration Approach

The system follows the Integration Maturity Model, primarily functioning at Level 2 (Opportunistic Sync):

Data Pull (from existing systems):

- DHIS2: Consumption data via REST API
- e-LMIS: Stock levels and delivery data via API
- eAFYA: Patient volumes via OpenMRS FHIR endpoints
- MoH Finance: Budget data via batch file transfer (CSV)

Data Push (to existing systems):

- Forecasts published to DHIS2 as analytics
- High-risk facility alerts sent to e-LMIS
- Order logs archived to central warehouse

For facilities with more consistent connectivity, the system can advance to Level 3 (Real-Time Interoperability) using FHIR resources for immediate data exchange.

Technology Stack (Open Source)

- Tier 1 Mobile App: React Native, SQLite
- Tier 2 Forecasting: Python, `hierarchicalforecast` library, pandas
- Tier 3 ML: Python, scikit-learn (Random Forest), XGBoost
- Data Sync: RESTful APIs over HTTPS with TLS 1.2+
- Database: PostgreSQL (district/central), SQLite (mobile)

ETHICAL, GOVERNANCE AND TRANSPARENCY REQUIREMENTS

Transparency and Explainability

Every forecast must be explainable in simple terms:

Tier 1 Example:

"Based on your last 3 months of usage, adjusted for malaria season (+30%), limited by your storage capacity."

Tier 3 Example:

"High stockout risk due to: delivery delays (40% contribution), low recent stock levels (35%), nearby facility outbreaks (25%)."

The reasoning must be available to facility staff (Tier 1) and district officers (Tier 3) in plain language.

Governance Structure

A Forecasting Governance Committee will be established:

Membership:

- Ministry of Health (Pharmacy Division)
- District Health Officers (rotating representatives)
- Facility users (nominated by peers)
- Technical partner (IFRAD or successor)
- Academic validator (Kyambogo University or similar)

Responsibilities:

- Quarterly review of model performance
- Analysis of override patterns
- Approval of algorithm changes
- Recommendations for refinement

Human Agency Protection

The system explicitly logs that overrides are a sign of an engaged user, not a system failure. There are no penalties for appropriate overrides. However, the system does flag:

- Suspicious patterns (consistent over-ordering followed by unexplained stock disappearance)
- Safety issues (reordering quantities that violate clinical protocols)

These flags are for supervisory review, not automated punishment.

Data Governance Compliance

All data handling aligns with:

- Uganda's Data Protection and Privacy Act (2019)
- Ministry of Health National eHealth Policy
- WHO Digital Health Guidelines

Patient-level data is never used. All forecasting operates on aggregated, anonymized commodity consumption data.

RISKS AND MITIGATION STRATEGIES

Risk 1: Model Drift

As consumption patterns change, forecast accuracy may degrade.

Mitigation:

- Monthly automated model retraining (Tiers 2 and 3)
- Quarterly governance review of performance trends
- Automated alerts if MAPE exceeds thresholds

Risk 2: Data Quality Issues

Garbage in, garbage out. Poor data will produce poor forecasts.

Mitigation:

- Data quality dashboards for district officers
- Validation rules in mobile app (consumption cannot exceed 3x historical max)
- Facility-level feedback loop (app shows recent data quality score)

Risk 3: User Trust and Adoption

If users do not trust the forecasts, they will ignore them.

Mitigation:

- Transparent explanations with every forecast

- User override capability with no penalties
- Participatory design workshops before deployment
- Continuous feedback mechanism via in-app surveys

Risk 4: Infrastructure Failure

Servers crash, connectivity fails, devices break.

Mitigation:

- Tier 1 functions fully offline-no dependency on infrastructure
- District server redundancy (backup power, RAID storage)
- Cloud failover for Tier 3 if district servers fail
- CSV export capability as ultimate fallback

Risk 5: Misuse or Gaming

Users might attempt to manipulate forecasts for personal gain.

Mitigation:

- Audit trails for all orders and overrides
- Anomaly detection for unusual patterns
- Supervisory review dashboards
- Separation of forecasting and approval authority

PROPOSED POST-GRANT IMPLEMENTATION ROADMAP (2026+)

Phase 1: Pilot (Months 1-3)

- Deploy Tier 1 in 10 facilities across 2 districts
- Validate data sync functionality
- Collect user feedback on interface and explanations
- Measure baseline stockout rates for comparison

Phase 2: Tier 2 Integration (Months 4-6)

- Deploy district-level servers
- Implement HES forecasting service
- Train district officers on dashboard use
- Begin comparative analysis (Tier 1 vs. Tier 2 accuracy)

Phase 3: Tier 3 Deployment (Months 7-9)

- Deploy ML forecasting for stockout risk prediction
- Pilot proactive redistribution protocols
- Refine feature set based on district feedback
- Conduct first governance committee review

Phase 4: Scale (Months 10-12)

- Expand to 50+ facilities
- Integrate with national procurement planning
- Publish technical specifications for replication
- Transfer ownership to Ministry of Health

SUCCESS CRITERIA

The framework will be considered successful if, after 12 months of deployment:

1. Stockout Reduction: 30% reduction in stockout frequency vs. baseline
2. Duration Improvement: 40% reduction in average stockout duration

3. User Adoption: >80% of facilities actively using and trusting forecasts
4. Override Appropriateness: <20% of overrides flagged as anomalous
5. System Reliability: >95% uptime for Tier 1 offline functionality
6. Data Quality: >70% of facility data passing quality thresholds

These criteria focus on outcomes (supply availability) and process quality (user engagement), not just forecast accuracy.

CONCLUSION

This Predictive Demand Algorithm Specification presents a pragmatic, tiered approach to forecasting that respects Uganda's health system realities. By prioritizing offline functionality, human oversight and infrastructure constraints, the framework provides guidance to facility staff while building toward more sophisticated district and national-level analytics.

The technical validation by Kyambogo University confirmed the feasibility and appropriateness of this approach, with refinements to ensure scalability (Hierarchical Exponential Smoothing for Tier 2) and operational clarity (cold start protocol, data synchronization specifications). The framework is designed for government ownership, open-source implementation and replication across similar humanitarian contexts.

Document Version: Final (post-technical validation)

Date: November 2025

Authors: Gideon Abako, Timothy Kavuma, International Foundation for Recovery & Development

Validator: Ojok Ivan, Kyambogo University Department of Data Science, Networks & Artificial Intelligence

License: Creative Commons Attribution 4.0 International (CC BY 4.0)

DOCUMENT REVISION HISTORY

| Version | Date | Changes |
|---------|---------------|--|
| 1.0 | October 2025 | Initial predictive demand algorithm specification based on 2025 baseline assessment. Original Tier 2 model specified as Exponential Triple Smoothing (ETS). |
| 2.0 | November 2025 | <p>Post-validation revision incorporating Kyambogo University technical review findings:</p> <p>Tier 2 Model Change: Replaced standard ETS with Hierarchical Exponential Smoothing (HES) using bottom-up reconciliation. Rationale: Single model handles multiple facilities simultaneously, reducing computational and administrative overhead by 80% compared to per-facility ETS models while maintaining forecast coherence.</p> <p>Cold Start Protocol Enhancement: Added detailed facility matching criteria (facility level, geographic region, patient volume bracket, commodity class) for new facilities with <3 months historical data.</p> <p>Data Synchronization Protocol: Specified bidirectional sync workflow with explicit upload/download payloads including override logs, Tier 2 forecasts, updated baseline parameters, and seasonal multiplier refreshes.</p> <p>Model Selection Rationale Section: Added comprehensive comparison of HES vs. alternatives (standard ETS, Panel ARIMA, tree-based models) with scalability, resource efficiency, and interpretability justifications.</p> <p>Bottom-Up Reconciliation Method: Detailed hierarchical structure (district → facility type groups → individual facilities) with reconciliation process for budget constraints.</p> <p>Open Source Implementation: Specified Python libraries (hierarchicalforecast, scikit-learn, XGBoost) for government ownership and replication</p> |