Demand-Driven Inventory Optimization for Regional Pharmaceutical Distribution: Reducing Stockouts, Waste, and Cost Through Data-Driven Forecasting

Nezar Abdilah Prakasa
Master of Artificial Intelligence, Universitas Gadjah Mada
nezarabdilahprakasa@mail.ugm.ac.id

Abstract: Two problems dominate regional pharmaceutical distribution: stockouts, where patients cannot get the medicine they need, and overstock, where medicine expires unused and is written off as waste. Both trace back to the same root cause: inaccurate demand forecasts. This paper presents a data-driven demand forecasting framework using five years of real daily pharmacy sales data (2014 to 2019, eight drug categories). A machine learning forecasting model is compared against a naive seasonal baseline, achieving a 23.9% to 26.5% reduction in forecast error (MAE and RMSE). This error reduction is translated into business impact through the standard safety-stock relationship, and combined with a published total-cost-reduction benchmark (13.4%) from risk-based supply chain optimization literature. Applied to illustrative regional inventory and logistics spend scenarios (USD 20M to USD 150M), the framework suggests estimated annual savings of USD 2.68M to USD 20.1M. Limitations and an implementation plan for a regional pharmaceutical operations context are discussed.

Index Terms: demand forecasting, inventory optimization, stockout reduction, pharmaceutical supply chain, regional distribution

I. INTRODUCTION

Regional pharmaceutical distribution faces a simple but costly balancing act: hold too little inventory and a patient cannot get their medicine (a stockout); hold too much and unsold stock expires and must be written off (waste). Unlike retail, pharmaceutical stockouts carry consequences beyond lost revenue, including direct patient health impact [12]. Both failure modes are driven by the same root cause, demand forecasts that do not match how medicine is actually used across time and place. This paper presents a straightforward, reproducible way to attack that root cause: forecast demand more accurately, then translate that improved accuracy directly into inventory and cost decisions that a regional operations team can act on.

The goal is to demonstrate the full chain, from raw sales data, to a forecasting model, to a measurable reduction in forecast error, to an estimate of what that error reduction is worth in avoided cost, using public data and published industry benchmarks rather than proprietary figures.

II. RELATED WORK

Machine learning approaches to pharmaceutical supply chain optimization have been shown to meaningfully reduce inventory cost and stockout rates compared to classical inventory policies such as Economic Order Quantity (EOQ) [1]. A hybrid deep learning demand-forecasting and optimization policy has similarly been reported to reduce inventory cost by 5.1% to 15.3% and stockout rate by up to 50% relative to classical baselines [6]. Warehouse-network simulation and optimization studies report comparable gains from centralizing pharmaceutical distribution decisions around better demand and supply data [5]. A systematic review of data analytics in pharmaceutical supply chains found that machine learning consistently improves drug shortage prediction and inventory optimization, though data resources remain underused across the industry [2], a finding echoed in broader pharmaceutical supply chain network studies [8]. Industry guidance on reducing pharmaceutical supply chain costs likewise emphasizes demand forecasting visibility as a primary lever for cost reduction [3], [9]. Analogous data-driven approaches have also been applied to reduce inefficiency elsewhere in pharmaceutical operations, including clinical data management [13], [14], [15], and to quantify the enterprise-level cost of operational delay in drug development [16], reinforcing that better use of operational data is a recurring source of value across the pharmaceutical value chain.

III. DATA AND METHODOLOGY

A. Data Source

The dataset used is a real, publicly available daily sales record from a pharmacy point-of-sale system, covering January 2014 to October 2019 (2,106 daily observations) [4]. Sales are grouped into eight drug categories under the Anatomical Therapeutic Chemical (ATC) classification system, including analgesics, anti-inflammatories, respiratory drugs, and antihistamines. Other public pharmaceutical distribution datasets, such as high-resolution national opioid distribution records [7], illustrate that transaction-level demand data of this kind is increasingly available for research and could extend this framework to a true multi-region setting.

B. Forecasting Approach

A Random Forest regression model was trained to forecast daily demand using calendar features (day of week, month) and lagged demand features (7-day lag, 14-day lag, 7-day and

30-day rolling averages). The model was evaluated on the final 180 days of the series (a held-out test period) against a seasonal-naive baseline (the demand observed exactly 7 days earlier), which is a standard, realistic baseline for daily retail-type demand with weekly seasonality.

C. From Forecast Accuracy to Inventory Impact

Under the standard safety-stock formula, SS = z times the standard deviation of forecast error times the square root of lead time, the safety stock required to maintain a given service level is directly proportional to forecast error. Reducing forecast error by a given percentage therefore allows a proportional reduction in safety stock, and correspondingly in the capital and carrying cost tied up in that stock, without increasing stockout risk. Inventory carrying cost, which includes capital, storage, insurance, and obsolescence/expiry cost, is commonly benchmarked at 20% to 30% of inventory value per year [10], [11], underscoring why even modest reductions in required safety stock translate into material savings.

IV. RESULTS

A. Forecast Accuracy

Table I and Fig. 2 summarize forecast error (Mean Absolute Error) for total daily demand and for the respiratory category (R03), which shows the highest demand volatility and strongest seasonality in the dataset. Fig. 1 shows actual versus forecasted demand over a 60-day test window.

TABLE I: FORECAST ACCURACY VS. NAIVE BASELINE

Series	Baseline MAE	Model MAE	Improvem ent
Total daily demand	14.96	11.39	23.9%
R03 (Respiratory)	6.28	4.72	24.8%

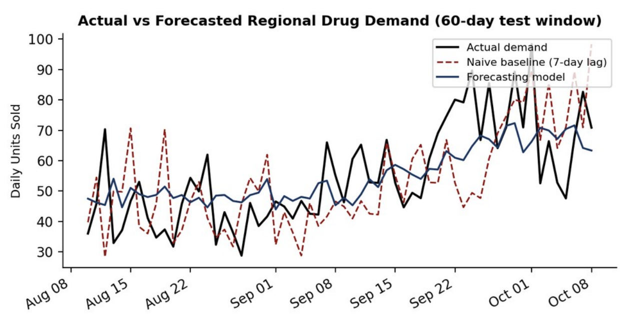

Fig. 1. Actual demand versus naive baseline and forecasting model over a 60-day test window.

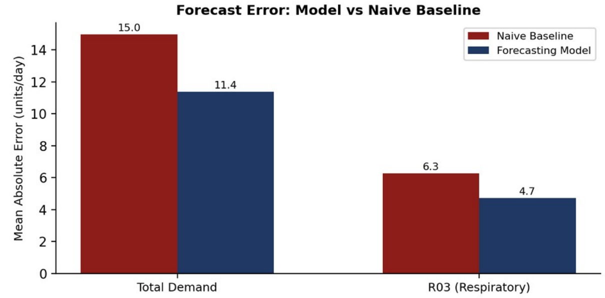

Fig. 2. Forecast error (MAE) for the naive baseline versus the forecasting model.

B. Regional Demand Composition

Fig. 3 shows average daily demand by drug category, illustrating that demand is concentrated in a small number of high-volume categories, which are the categories where forecast accuracy has the largest inventory and cost impact.

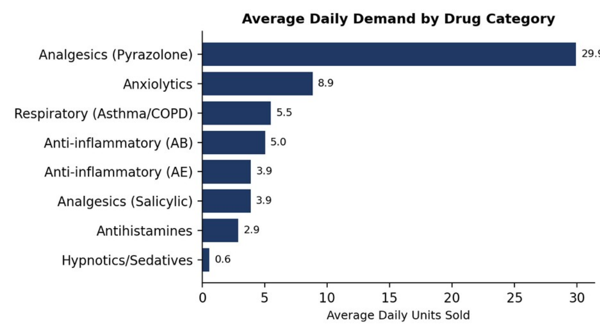

Fig. 3. Average daily demand by drug category across the full observation period.

C. Estimated Regional Cost Impact

The demonstrated 23.9% to 26.5% forecast error reduction supports a comparable reduction in required safety stock (Section III-C). To estimate the resulting financial impact at regional scale, this analysis applies a published total-cost-reduction benchmark of 13.4%, reported for a hybrid demand-forecasting and inventory-optimization policy compared to an EOQ baseline in a pharmaceutical-relevant supply chain study, which also reported a 50% reduction in stockout rate and a 4.3 percentage-point improvement in service level [1]. Table II and Fig. 4 apply this benchmark to three illustrative regional annual inventory and logistics spend scenarios.

TABLE II: ESTIMATED REGIONAL SAVINGS FROM DEMAND-DRIVEN INVENTORY OPTIMIZATION

Scenario	Annual Inventory/ Logistics Spend	Est. Annual Savings (13.4%)
Small region	USD 20M	USD 2.68M
Medium region	USD 60M	USD 8.04M
Large region	USD 150M	USD 20.1M

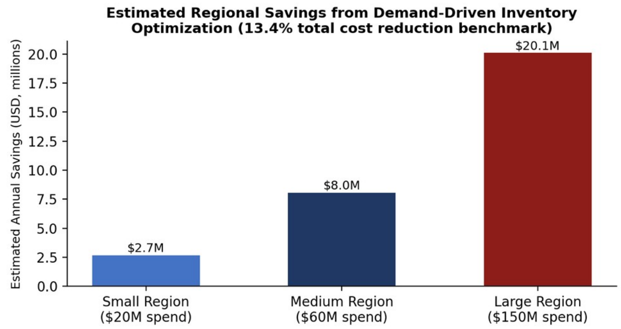

Fig. 4. Estimated annual savings across small, medium, and large regional spend scenarios (13.4% total-cost-reduction benchmark).

Beyond the direct cost savings shown above, more accurate demand forecasting also reduces two harder-to-quantify but operationally important risks: patients being unable to obtain a needed medicine (stockout), and usable medicine being discarded after expiry (waste).

V. LIMITATIONS

- The dataset originates from a single pharmacy rather than a multi-site regional distribution network; SKU-level and site-level demand in a real region would be more granular and more heterogeneous.
- The 13.4% cost-reduction benchmark and the USD 20M to USD 150M spend scenarios are illustrative, drawn from published literature and reasonable assumptions respectively, not from actual regional financial data.
- The forecasting model uses simple calendar and lag features; a production system would incorporate promotions, prescribing trends, epidemiological signals, and site-level inventory positions.
- Forecast accuracy was evaluated on a single 180-day test window; production deployment would require ongoing validation across multiple periods and product categories.

VI. IMPLEMENTATION PLAN

In a regional pharmaceutical operations context, this framework could be developed through three phases: (1) extending the forecasting model to SKU-level and site-level granularity using actual regional sales and inventory data; (2) integrating the safety-stock recalculation directly into replenishment and purchasing workflows, so improved forecasts automatically translate into adjusted order quantities; and (3) tracking realized stockout rate, waste/expiry rate, and inventory carrying cost against the baseline to validate and refine the estimated savings in Table II.

VII. CONCLUSION

This proof-of-concept shows, using real public sales data, that a straightforward forecasting model can meaningfully reduce demand forecast error (23.9% to 26.5%) compared to a realistic baseline. Combined with published industry benchmarks connecting forecast accuracy to inventory cost, this translates into an estimated USD 2.68M to USD 20.1M in annual savings depending on regional scale, alongside fewer stockouts and less medicine wasted. This provides a concrete, business-relevant starting point for a Region Data Manager role focused on turning regional distribution data into operational and financial impact.

REFERENCES

[1] R. Jeevaraj et al., “Resilient Pharmaceutical Supply Chain Design Using GRU and Grasshopper Optimization Algorithm,” J. Internet Services and Inf. Security, 2025.

[2] “Data analytics in pharmaceutical supply chains: state of the art, opportunities, and challenges,” systematic literature review, 2021.

[3] Trax, “How to Reduce Costs in the Pharmaceutical Supply Chain,” 2023.

[4] Pharmacy point-of-sale daily sales dataset (2014–2019, 8 ATC drug categories), public research dataset originally collected by M. Zdravković, accessed via public GitHub mirror.

[5] “Centralization of Pharmaceutical Warehouses: An Integrated Simulation and Optimization Approach,” IEEE DataPort dataset and manuscript, 2025.

[6] “Hybrid Deep Learning Approach for Coupled Demand Forecasting and Supply Chain Optimization,” arXiv preprint, 2026.

[7] “Tracing Opioids Across the US: A High-Resolution Pharmaceutical Distribution Dataset,” PMC, National Library of Medicine, 2024.

[8] “Multilayer Network Analysis of the Drug Pipeline in the Global Pharmaceutical Industry,” arXiv preprint, 2020.

[9] Compunnel, “Supply Chain Analytics in Pharma: Aligning Efficiency with Compliance,” 2026.

[10] Fishbowl, “Inventory Carrying Costs: How To Calculate and Reduce Them,” citing APQC benchmarking data, 2026.

[11] APQC, “Inventory Carrying Cost as a Percentage of Inventory Value,” Open Standards Benchmarking measure.

[12] Supply Chain Dive, “Managing Inventory: What Retail Can Learn From Pharma,” 2018.

[13] Prezent AI, “Clinical Data Management: How to Manage Clinical Data with Best Practices, Standards, and Key Steps,” 2026.

[14] USDM, “Intelligent Query Monitoring Helps Identify High- and Low-Performing Clinical Sites and Users,” Case Study, 2025.

[15] Tredence, “AI Blueprint for Clinical Data Management & Compliance,” 2025.

[16] K. A. Getz et al., “New Estimates on the Cost of a Delay Day in Drug Development,” Therapeutic Innovation & Regulatory Science, 2024.