**Fake Drug Detection via Pharmacy Sales Analytics**

## EXECUTIVE SUMMARY

Counterfeit and substandard medicines remain a critical public health and economic problem in Nigeria. Manual inspections by regulators such as NAFDAC are too slow to keep pace with the scale of distribution.

This project demonstrates how **pharmacy sales analytics combined with anomaly detection models** can provide an early warning system for counterfeit detection. Using a dataset of **10,000 pharmacy transactions across urban and rural areas**, we cleaned, engineered features, applied multiple anomaly detection algorithms, and developed a risk-scoring dashboard.

Key findings include:

* Pharmacies such as **Rodriguez-Graham** and **Gracia-James** showed consistently low prices and elevated risk scores.
* Suppliers such as **Novak PLC** and **Hansley** were repeatedly linked to anomalies.
* Around **31% of flagged drugs were near expiry**, suggesting manipulation of expiry dates or clearance of unsafe stock.
* Ensemble anomaly detection identified **1.3% of transactions as high-risk**, balancing sensitivity with reduced false positives.

This work provides a blueprint for scaling **data-driven monitoring** into Nigeria’s regulatory system.

1. BACKGROUND AND CASE STUDY

Nigeria’s counterfeit drug crisis is long-standing, with tragic historical cases such as the **1990 paracetamol syrup disaster (109 child deaths)**. Investigations continue to uncover counterfeit **antimalarials, antibiotics, and antihypertensives**, particularly in **rural and semi-urban pharmacies**. This includes the recent discovery of fake antimalarials worth over #1.2 billion.

Key characteristics of counterfeit circulation include:

* **Suspicious pricing**: abnormally low compared to genuine pharmacies.
* **Unregistered brands**: “new” brands not on NAFDAC’s list.
* **Sales spikes**: abnormal seasonal surges in malaria treatment drugs.
* **Forged or missing expiry dates.**

This project focuses on **pharmacy sales analytics** as a scalable way to flag such risk

1. PROBLEM STATEMENT

* **Context:** Fake and substandard drugs are easily distributed through informal channels such as open markets, rural pharmacies, and mobile vans. Expiry dates are often falsified, and new “unknown” brands emerge frequently.
* **Issue:** Manual inspections are insufficient for scale.
* **Goal:** Build a **data-driven anomaly detection system** that flags pharmacies, suppliers, and brands with suspicious sales behaviour.
* **Impact:** Enable regulators and pharmacy chains to act earlier, protect patients, and rebuild trust.

1. PROJECT OBJECTIVES

PRIMARY OBJECTIVE:

Develop an analytics system that detects suspicious pharmacies, suppliers and brands using sales data.

SECONDARY OBJECTIVES

* Flag anomalies: price drops, sales spikes, rare brands, near-expiry drugs.
* Score and rank pharmacies, suppliers, and brands by risk.
* Deliver an interactive dashboard for stakeholders.

1. DATA OVERVIEW

**Dataset:** 10,000 pharmacy transactions, 12 columns.

KEY FEATURES

* Date, Expiry Date: Transaction and expiry tracking.
* Pharmacy, Location: Urban vs rural differences.
* Drug, Brand, Price, Quantity: Core sales information.
* Supplier Name, Sales Channel: Supplier behaviours and distribution methods.
* Drug Form: Delivery format (tablet, injection, syrup).

CHALLENGES

* Inconsistent date formats → cleaned with parsing rules.
* Brand spelling variations → standardized using fuzzy matching.
* Missing values in pharmacy, price, supplier → imputed or dropped.

1. FEATURE ENGINEERING

Features designed to reveal **risk patterns**:

* **Pharmacy-Level Features**

Total sales volume per pharmacy.

Average price per pharmacy.

Unusual discounts (Price < 70% of brand average).

* **Supplier-Level Features**

Rare brand ratio per supplier.

Suspicious distribution channel flag (Mobile Van, Market Stall).

* **Drug-Level Features**

Sales spike flag (Quantity > 2× average for drug).

* **Expiry Features**

Days to expiry.

Near-expiry flag (< 90 days).

* Composite Risk Score

Unusual discount + Rare brand + Suspicious channel + Near expiry + Sales spikes

1. EXPLORATORY DATA ANALYSIS

* **Price distribution:** genuine brands (BioCare, MediPlus, PharmaTrust) clustered around ₦2600–₦2700; anomalies as low as ₦100 suggesting deep discounts or fakes.
* **Supplier distribution:** Jones Inc handled ~650 sales, but several suppliers had <350
* **Rare brands:** “Unknown” brand (<10 sales) raised suspicion.
* **Urban vs rural:** Transactions were evenly split, but rural locations carried higher average risk scores.
* **High-risk pharmacies:** Gruzman, Rodriguez-Graham had average risk scores >1.3.

1. ANOMALY DETECTION MODELS

Unsupervised models applied with contamination rate ~2%:

* **Isolation Forest (IF):** good for tabular anomalies, flagged ~2%.
* **Local Outlier Factor (LOF):** captured local deviations, flagged ~2%.
* **One-Class SVM:** Captures nonlinear patterns, slightly more sensitive, flagged ~2.21%.
* **Ensemble Majority Voting:** flagged if ≥2 models agreed; conservative, ~1.3% flagged.

**Interpretation:**

* IF and LOF aligned well for global/local anomalies.
* OCSVM captured additional patterns but risked more false positives.
* Ensemble reduced false positives.

1. DASHBOARD AND VISUALIZATION

We built a **Streamlit dashboard** for stakeholders. Key views:

* **Suspicious Pharmacies:** Gracia-James, Rodriguez-Graham, Unknown Pharmacy.
* **Suspicious Suppliers:** Novak PLC, Hansley, Powell, David.
* **Price Monitoring:** Low-price outliers at Rodriguez-Graham & Harrell LLC.
* **High-risk brands:** MediPlus, PharmaTrust (high sales, require monitoring).
* **Expiry monitoring:** 31% of flagged sales were near expiry.
* **Urban vs Rural:** anomalies higher in urban absolute terms, but rural pharmacies riskier due to weak oversight.

1. RESULTS AGAINTS RESEARCH QUESTIONS
2. **Which pharmacies sell at significantly lower prices?**
   * Rodriguez-Graham & Harrell LLC underpriced by >30% compared to brand average.
3. **Are there brands with disproportionately high sales but unknown origins?**
   * Yes, “Unknown” brand and a small supplier-linked antibiotic brand.
4. **Are certain suppliers repeatedly linked to anomalies?**
   * Yes, Novak PLC, Hansley.
5. **How do urban vs rural sales differ?**
   * Similar volumes, but rural transactions show higher anomaly density.
6. **Are expiry dates suspicious?**
   * 31% of anomalies had near-expiry drugs (<90 days).
7. RECOMMENDATIONS

SHORT-TERM

* Audit pharmacies and suppliers flagged by the system.
* Investigate near-expiry sales and verify expiry labels.
* Increase awareness campaigns in Lagos and rural regions.

LONG-TERM

* Integrate anomaly detection pipeline into **NAFDAC systems**.
* Mandate supplier traceability (batch-level tracking).
* Deploy nationwide monitoring with continuous retraining.

1. LIMITATIONS

* Dataset covers only one month → no long-term trend analysis.
* No ground-truth fraud labels → risk of false positives.
* Location data coarse (Urban vs Rural only).
* Supplier names inconsistently recorded.

1. CONCLUSION

This project shows how **AI-driven anomaly detection of pharmacy sales** can complement regulatory inspections. By flagging suspicious entities, it allows better targeting of scarce inspection resources.

While not replacing lab verification or field audits, the system provides a **scalable early-warning tool** that can save lives and rebuild trust in Nigeria’s healthcare system.