DRUG AUTHENTICITY VERIFICATION SYSTEM

AI/ML Solution for UN SDG 3: Good Health and Well-being

Name: BRIAN MWANGI KAIGAI

Course: AI for Software Engineering

1. SDG PROBLEM ADDRESSED

Counterfeit medicines cause over **1 million deaths annually**, with developing countries facing up to **30% fake drug prevalence**. WHO reports that **1 in 10 medical products** in low- and middle-income countries is substandard or falsified. In Kenya, the Pharmacy and Poisons Board has documented rising counterfeit penetration, threatening public health and trust.

This project supports **SDG 3 (Good Health & Well-being)**, Target 3.8 (access to safe, quality medicines), while also contributing to **SDG 9 (Innovation)** and **SDG 17 (Partnerships)**.

2. MACHINE LEARNING APPROACH

Method: Supervised multi-class classification for pharmaceutical verification.
Target Classes: Authentic (1), Suspicious (0), Counterfeit (-1).

Models Used:

- **Random Forest (Primary):** 200 trees, balanced weights, max depth 10.
- **K-Nearest Neighbors:** 3 neighbors, cosine similarity (comparative baseline).

Feature Engineering: TF-IDF (1–3 n-grams), 150-dimension vector space capturing manufacturer and drug-name patterns.

Multi-Layered Verification System:

1. ML prediction + confidence score
2. PPB registry fuzzy matching (RapidFuzz)
3. OpenFDA real-time API cross-reference
4. Simulated blockchain hash verification (SHA-256)
5. Combined 0–100 drug risk score

Dataset: 75 samples (35 authentic, 20 suspicious, 20 counterfeit), stratified 75/25 split, 5-fold cross-validation.

3. RESULTS

Model	Accuracy	F1 Score	Cross-Validation
Random Forest	**90–95%**	**88–92%**	**90–93%**
KNN	85–90%	80–85%	82–87%

Deployment: Streamlit web app with global access, real-time API checks, color-coded risk levels (0–20 authentic → 81–100 counterfeit) and downloadable verification reports.

4. ETHICAL CONSIDERATIONS

Bias & Limitations:

- Small dataset; synthetic data may miss real-world complexity.
- Optimized for Kenya/East Africa; English-only interface.
- Internet-dependent, may exclude low-connectivity regions.

Risk Impact:

- **False negatives:** Highest danger—may allow harmful drugs through.
- **False positives:** Reduce trust; mitigated via "suspicious" middle class.

Fairness & Access:

- Free, open-source, no login required
  – Requires digital literacy and device access

Transparency: System is a **screening tool**, not a replacement for pharmacists or regulators. Real blockchain, chemical testing, and full PPB integration are not yet implemented.

Responsible AI: Prioritizes safety, clear disclaimers, transparent methods, and open public access.

CONCLUSION

The Drug Authenticity Verification System demonstrates how AI/ML can meaningfully support global health by detecting counterfeit medicines with **90–95% accuracy** and multi-layer verification. Future improvements include dataset expansion (10,000+ samples), barcode scanning, packaging image analysis, full blockchain backend, multilingual support, and offline mobile capability for rural healthcare.

Technologies: Python • Scikit-learn • Streamlit • TF-IDF • OpenFDA API • SHA-256