PREDICTING WATER WELL FAILURES IN TANZANIA

SUBTITLE

A DATA-DRIVEN STRATEGY FOR SMARTER MAINTENANCE

Problem Statement

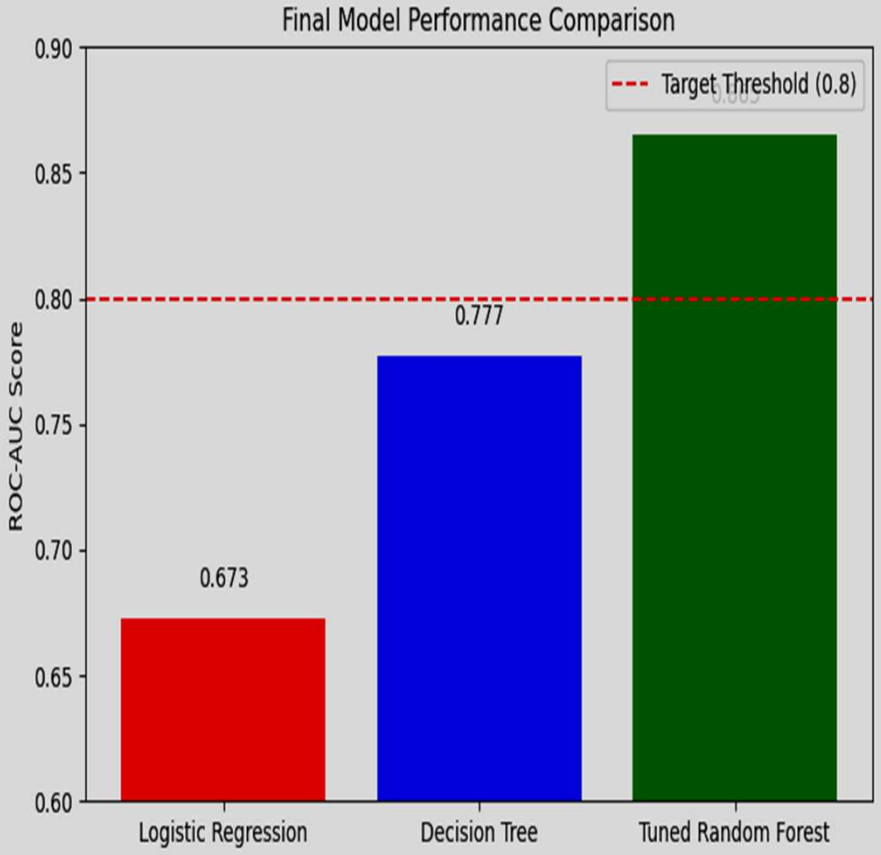
- **Current Situation**: Tanzania has 50,000+ water points, fixed only after they break
- **Impact on Communities:** People wait weeks for water; emergencies cause frustration
- **Cost Issues:** Emergency repairs are expensive and drain resources
- **Our Goal:** Predict which wells are likely to fail, enabling proactive maintenance.
- **Success:** Our model identifies 2 out of 3 at-risk wells, preventing water shortages.

What the Data Tells Us

- **Dataset:** 59,400 wells with info on pump type, age, location among others
- **Data Cleaning**: Fixed placeholders like “0” for construction year and elevation
- **Realistic Adjustments:** Replaced missing years with plausible values plus randomness
- **New Features:** Calculated well age, since older wells fail more often.
- **Simplified Metrics:** Converted water amount and population into **low / medium / high** categories for easier modeling

Our Modeling Journey

- **Logistic Regression:** Simple yes/no checklist; missed most at-risk wells
- **Decision Tree:** Flow-chart style rules; easier to interpret but still inaccurate
- **Random Forest:** Hundreds of trees working together like expert panel
- **Tuning:** Focused on catching wells that need repair, reducing missed cases
- **Outcome:** Final recommended model The Tuned Random Forest – effective at predicting at-risk wells



OUR MODELING JOURNEY – 3 MODELS, 1 WINNER

❖ **CHECKLIST MODEL(LR)**

Too rigid. Missed nearly every well needing repair

❖ **FLOWCHART MODEL(DT)**

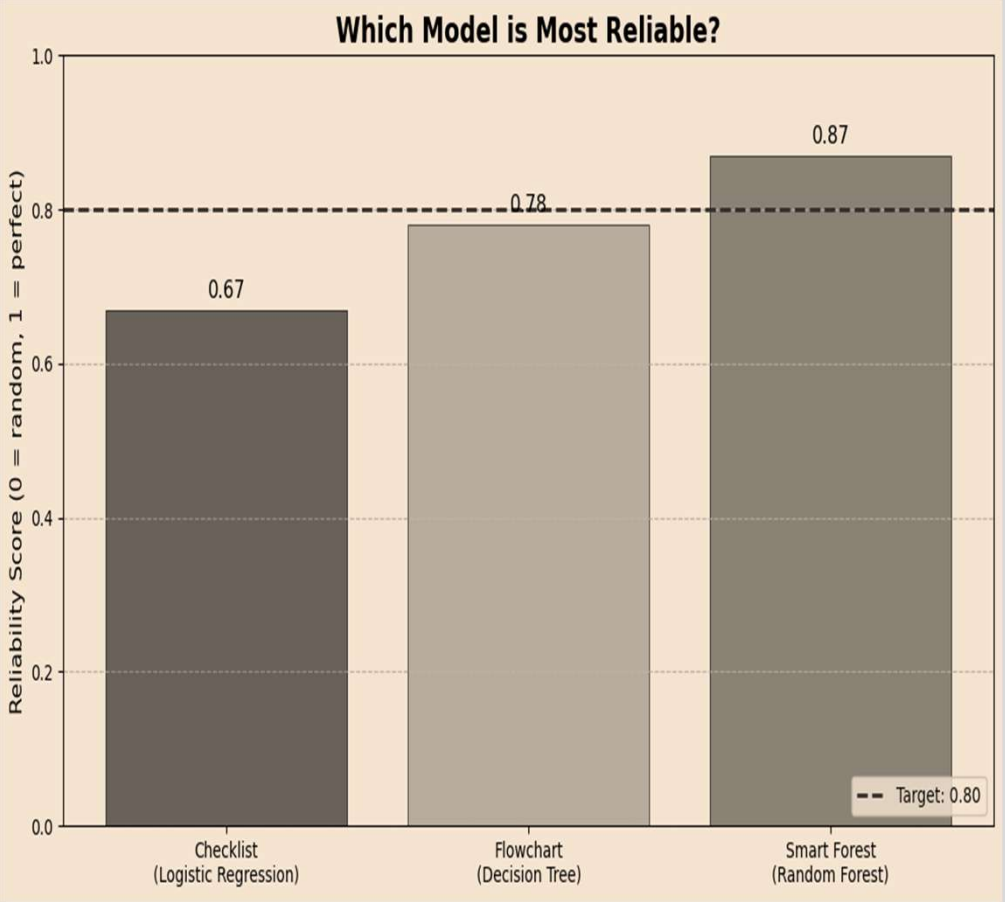
Better logic, but still missed 9 out of 10 at-risk wells

❖**SMART FOREST MODEL**

*Learns from 59,400 wells. Catches **2 out of 3 wells** that need repair.*

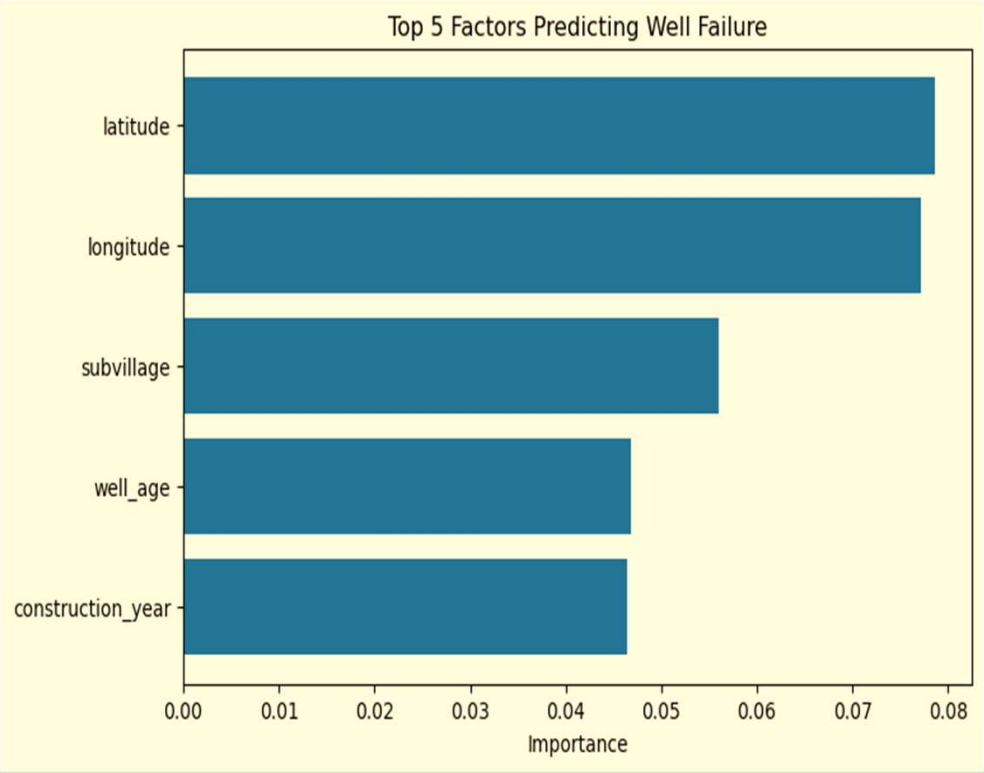
❖ **RESULT: From 0% → 66% catch rate.**

***A real warning system – finally ready for the Ministry**.*



Why Wells Fail – Top 5 Factors

- **Location** – Some regions fail more due to soil, rainfall, or groundwater.
  Action: **Target high-risk zones**
- **Sub village** – Certain communities see repeated failures.
- ***Action: Focus training and maintenance support.***



Why Wells Fail – Top 5 Factors CONT..

- **Well Age** – Older wells are more likely to fail

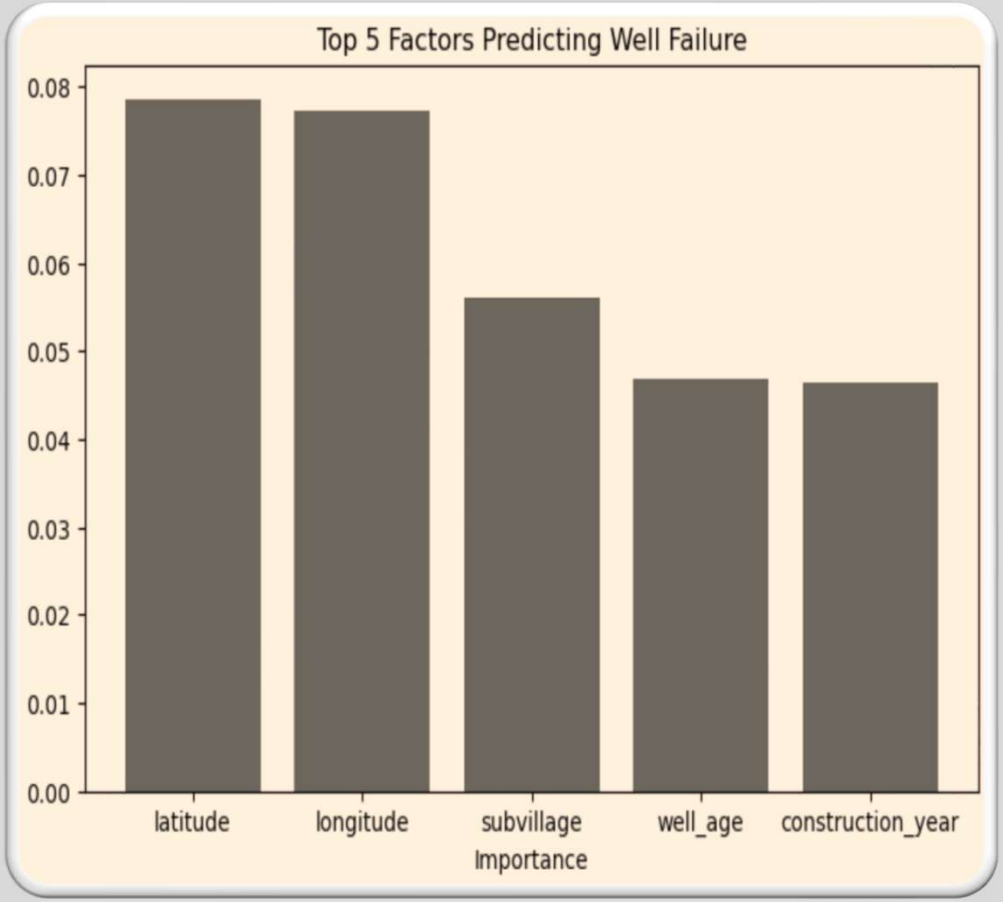
Action: Predict danger zones early

- **Construction Year** – Older builds fail sooner.

Action: Schedule proactive replacements.

- **Combined Insight** – Location & community matter most.

Action: Guide targeted interventions and planning.



RECOMMEDATIONS

❖ Target Dry & Insufficient Wells 79% / 52% non-functional

Action: Rapid response within 7 days

❖ Phase Out Old Wells Pre-1995 wells 2.3x more likely to fail

Action: 5-year replacement plan

❖ Convert “Never Pay” Wells 2x higher failure rate

Action: Introduce low-cost community contributions

GUIDANCE