Predictive Modeling: Tanzanian Water Wells

Snake Solutions LLC: Ely Lin, Titilayo Amuwo, Samantha Nasti

Agenda

- Problem
- Data
- Modeling Process
- Results
- Recommendation
- Next Steps

Problem

Context:

- Limited Clean Water Access
- 57 million people struggling
- Trouble maintaining large number of wells

Mission:

- Detect faulty wells
- Minimize impacted population
- Consider maintenance cost

Data

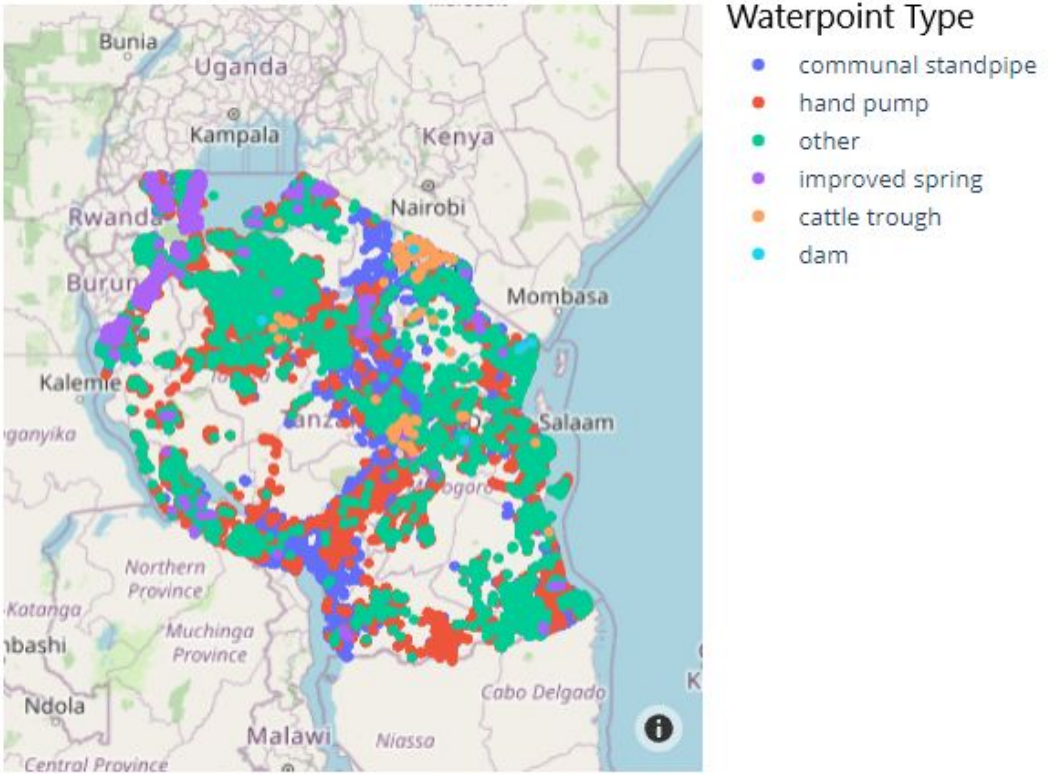
Public Records of existing Water Wells in Tanzania

- Taarifa
- Tanzanian Ministry of Water
- 50,000 + data points
- Dropped redundant information

Taarifa

Modeling Process: How It Works

- Predicts the condition of existing water wells in Tanzania
- Data contains **location coordinates**
  - Allows us to identify where repairs are needed



Modeling Process: Our Approach

263 Models Built

- Logistic Regression
- KNN Nearest Neighbors
- Decision Tree Classifier
- Random Forest Classifier
- CatBoost Classifier

CatBoost

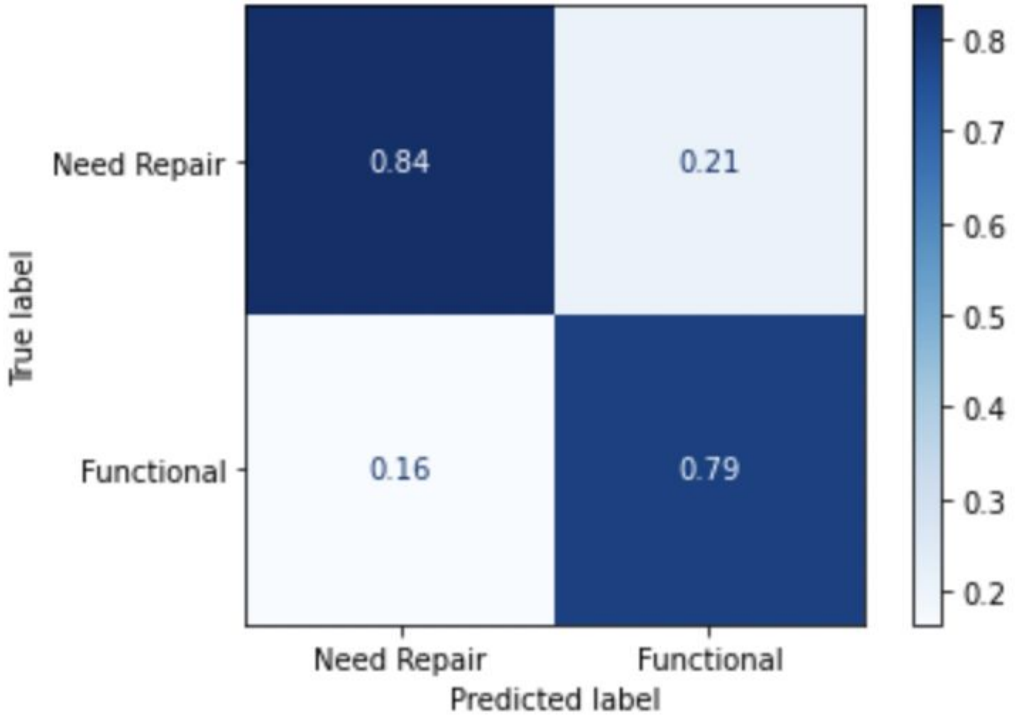
- Newest technology
- High accuracy
- Good with categorical data
- Fast

Results

Overall Accuracy: 80%

Missed repairs: 21%

False alarms: 16%



Recommendations

Use our model to:

- Locate water wells in need of repair
- Determine high risk area

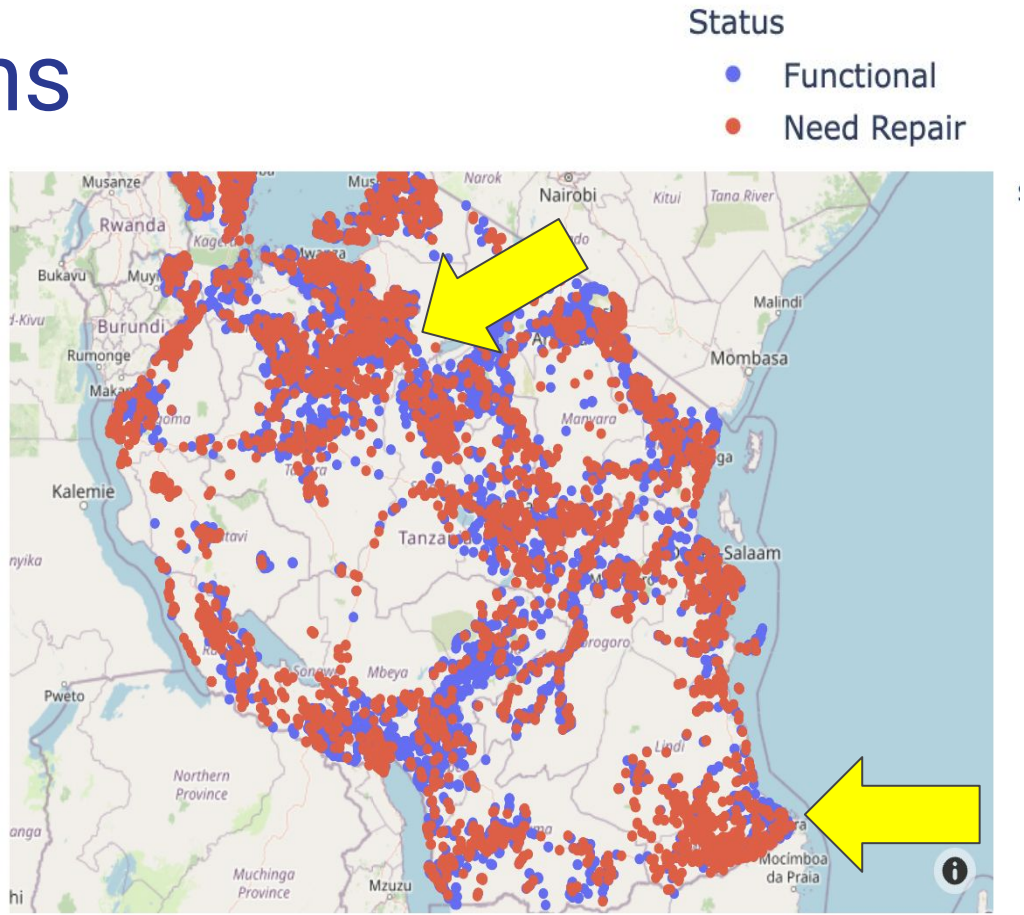
Input:

- Location data
- Water source data
- Wells technical specification

Recommendations

High risk:

- **Mtwara**
- **Shinyanga**



Next Steps

Future analysis

- Optimize for maintenance cost by clustering
- Explore features and trends
- Identify improvement opportunity

Thank You!

Any Questions?

Github:
@ElyLin
@Teetee-lab
@samanthanas