# *Tanzanian Wells*

Predicting non functional water wells

# *Table Of Contents*

# 1. Overview

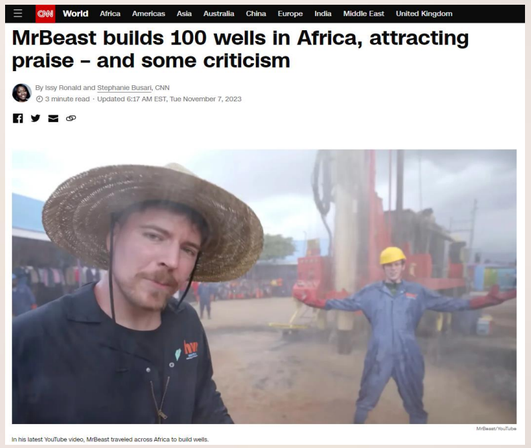
Objectives and Context

# *A Predictive Tool for Wells in Tanzania*

WaterAid

charity: water

the Tanzania Water Project




Current Situation:

In Tanzania, only **61%** have **basic water access**, 32% have basic sanitation, and less than half enjoy basic hygiene.

Source: World Bank

Objectives:

- Develop a machine learning classifier
- To **predict non-functionality** of water wells in Tanzania

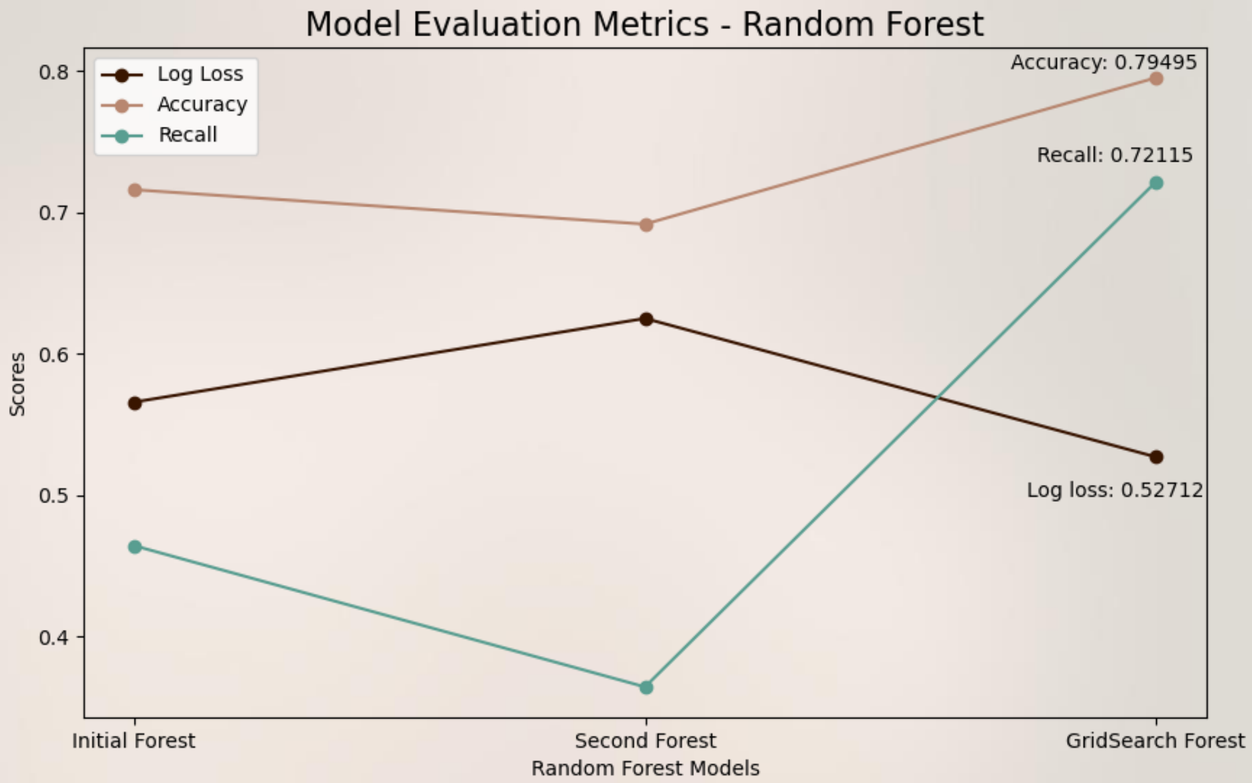
Intended Users:

- **Organizations** involved in water & infrastructure management
  - **Non-Governmental**
  - **Non-Profit**
  - YouTubers?

# 2. *Modeling*

Models and Tuning Techniques

# *Predictive Models to Assess Well Functionality*




<u>4 models</u>:

1. Logistic Regression
2. K-Nearest Neighbor
3. Decision Tree
4. **Random Forest**

<u>Preprocessing:</u>

- Categorization
- Missing values
- Encoding categorical features
- Scaling

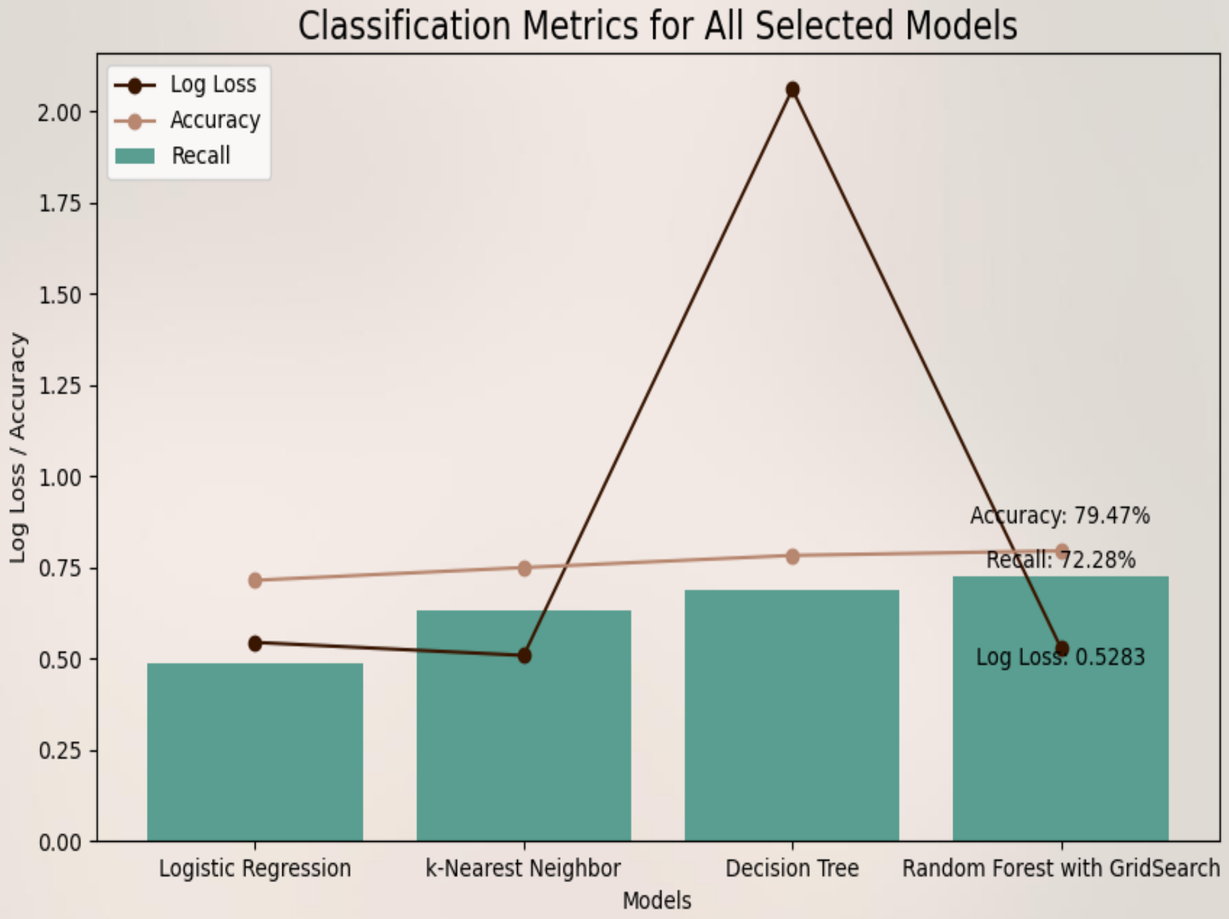
<u>Modeling Techniques</u>

- Synthetic Minority Over-sampling Techniques (SMOTE)
- Recursive Features Elimination
- **Combinatoric Grid Searching**


# 3. *Evaluation*

Best Model's
Classification
Metrics

# *Best Model Selection: Classification Metrics*




Classification Metrics:

1. **Recall**
2. Accuracy
3. Log loss

Risk for populations' health & lives if a well is predicted functional when it was not (false negative).

Best model results:

1. The model **correctly identified** over 72% of the actual **non-functional** water wells
2. Model's predictive power: 79% of **all** water **wells** were **correctly predicted**
3. High **probability** that a well **predicted** as non-functional **is actually** non functional

# 4.

# *Recommendations*

How to Achieve
Better Predictions

# *Recommendations*

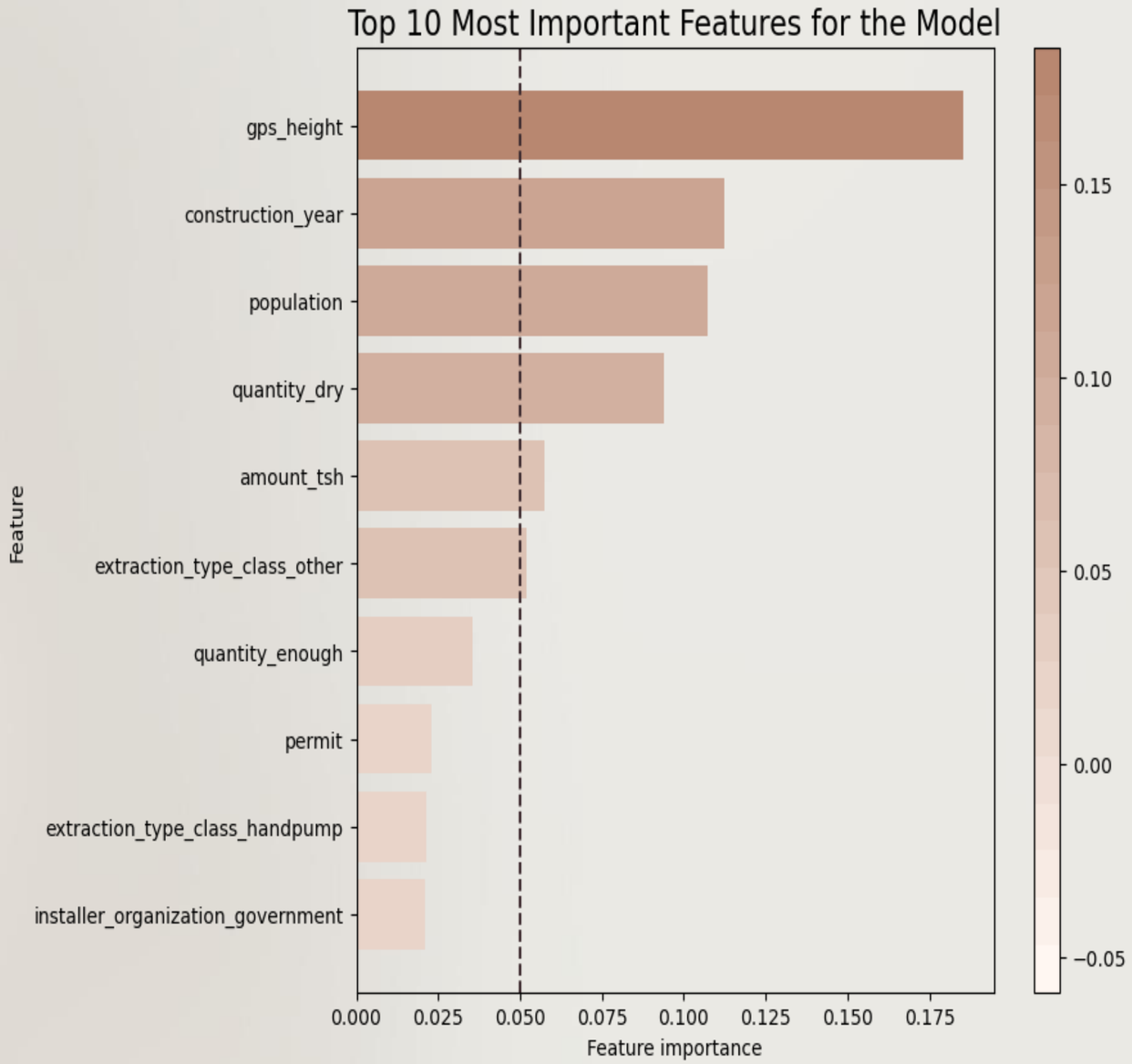
Purposes of Model's Predictions

- Useful
  - Prioritize maintenance efforts
  - Allocate resources and funding

- Not Useful
  - Data is outdated or inaccurate
  - Most important features are not reported

How to Achieve Better Predictions

1. Enhance **quality**: collect recent & accurate data
2. Include **local knowledge**
3. Share knowledge with other NGOs: **open data**
4. Implement a **feedback system**



# 5.
# *Next Steps*

Limits and
Overcoming Them

# 5.

# *Next Steps*

1. **Collect More Recent Data:** Data recorded is outdated: it is from 2011 to 2013. More recent information must be collected for more accurate predictions.

2. **Verify Actuals Before Relying on Predictions :** Accuracy is good, over 79%. Nevertheless, room for error still exist.

3. **Scalability Concerns:** Consider optimizing the model with more efficient algorithms or modifying processing techniques to predict larger datasets.

# *Contact Information*

For more details, contact:
albane.colmenares@gmail.com

*Thank you!*