

# *Tanzanian Wells*

Predicting non functional water wells



# *Table Of Contents*

1.

## Overview

Objectives and  
Context

2.

## Modeling

Models and  
Tuning  
Techniques

3.

## Evaluation

Best Model's  
Classification  
Metrics

4.

## Recos

How to Achieve  
Better  
Predictions

5.

## Next Steps

Limits and  
Overcoming  
Them

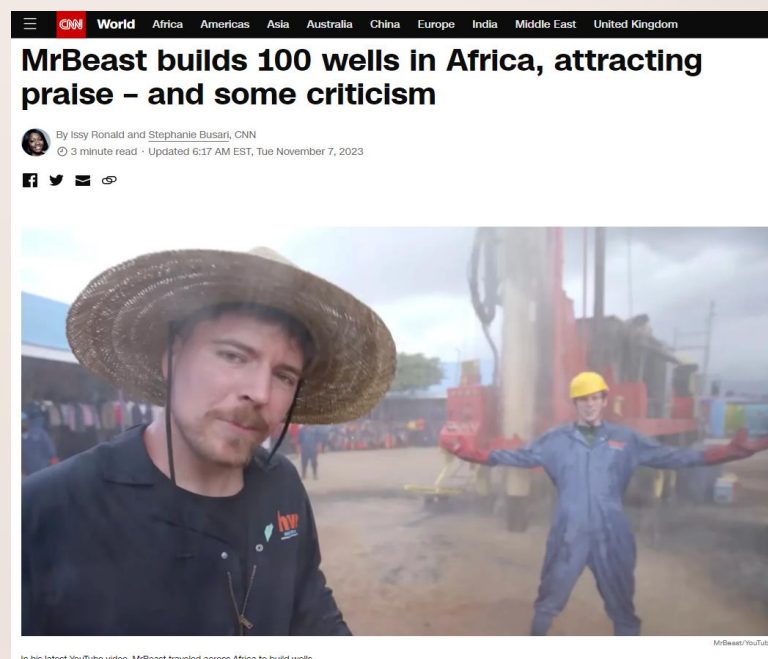




# 1. Overview

Objectives and  
Context

# *A Predictive Tool for Wells in Tanzania*



## 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?





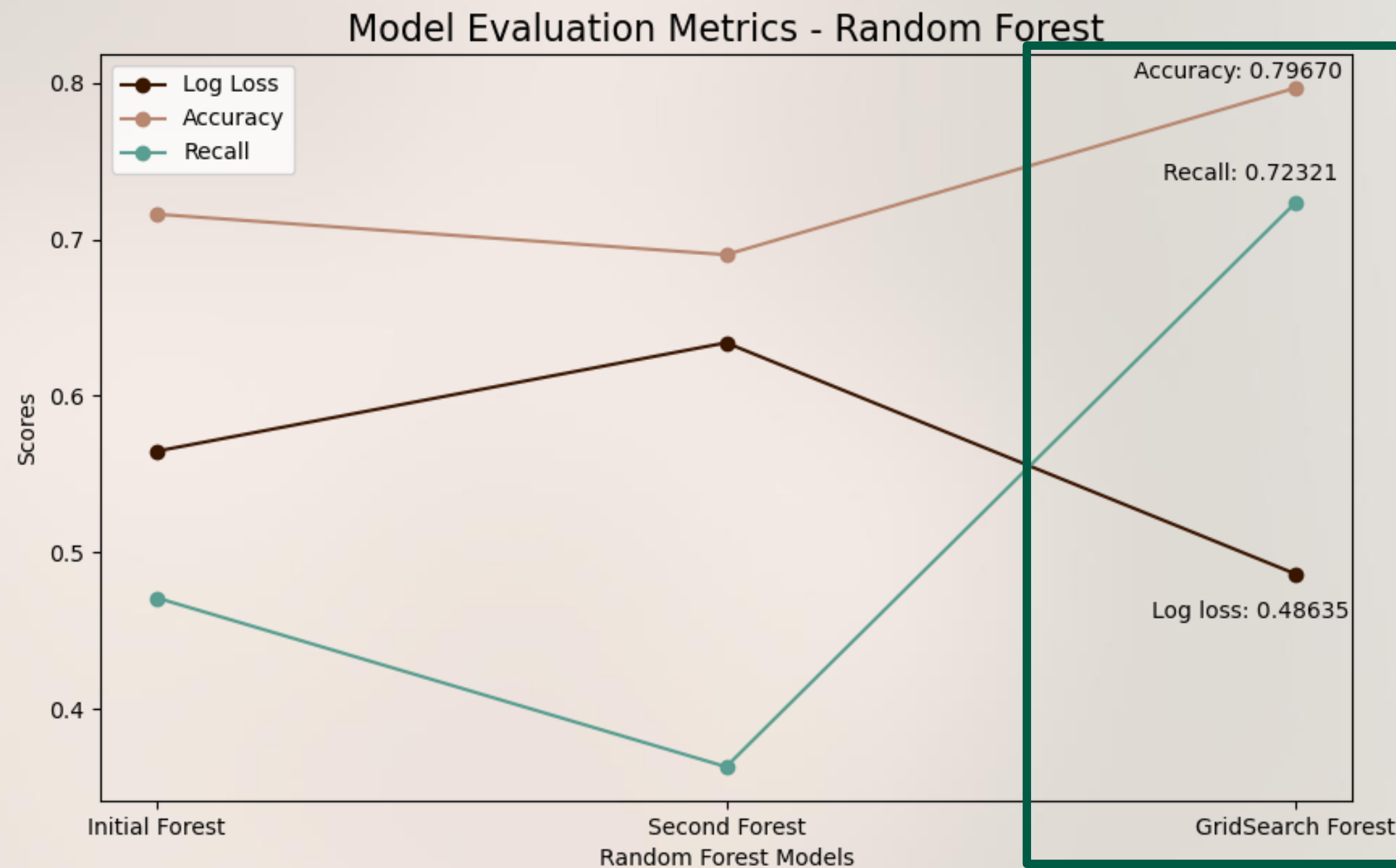
## 2. *Modeling*

---

Models and Tuning  
Techniques

# *Predictive Models to Assess Well Functionality*

Several versions of each model was run multiple times to find the one with the highest recall. Below is the example of **Random Forest**' iterations.



4 models:

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

Preprocessing:

- Missing values handling
- Categorization
- Encoding categorical features
- Scaling



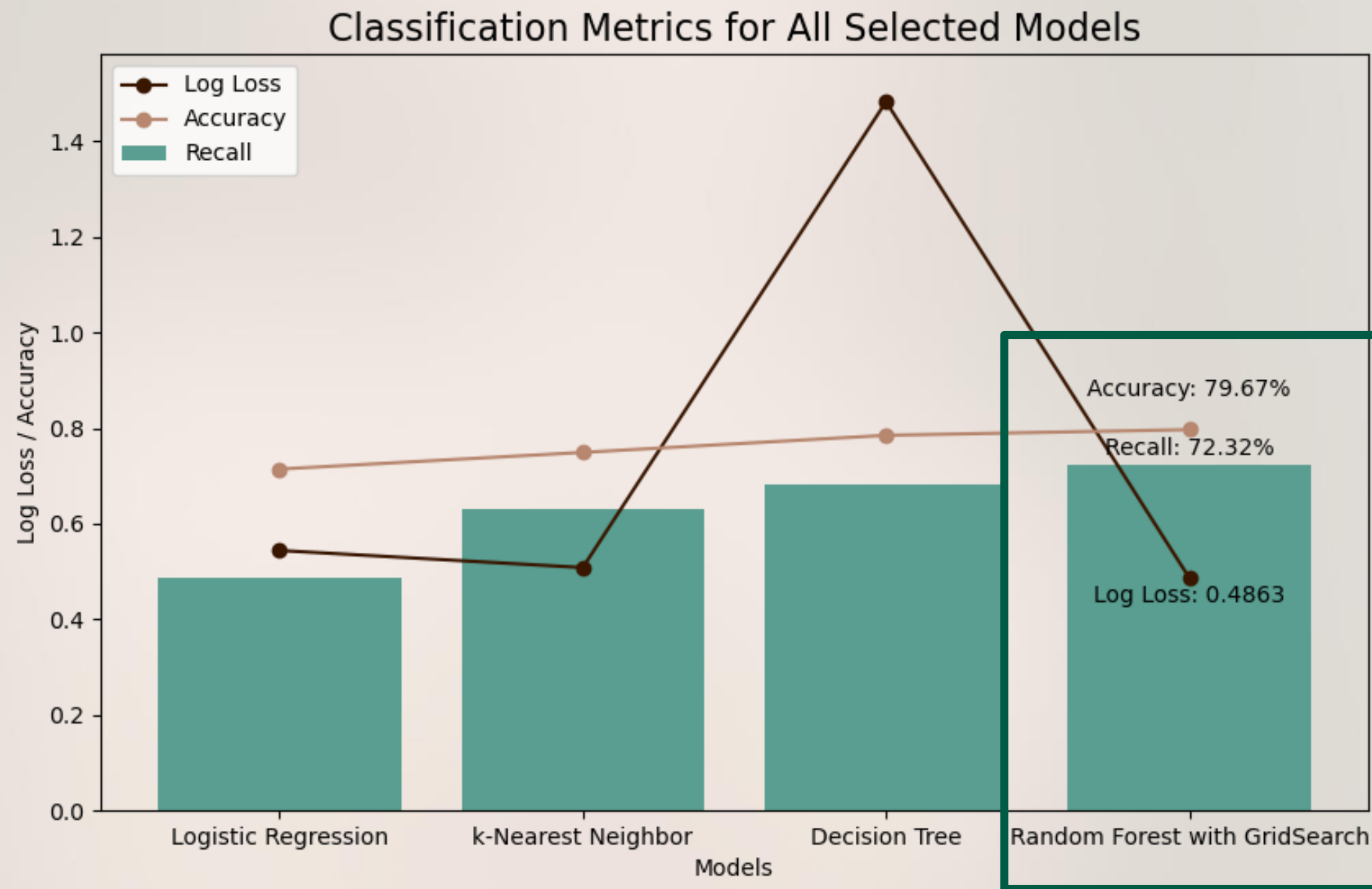


# 3. *Evaluation*

---

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*

The **higher** these features are on this graph, the **more likely** a well will be **non-functional**.

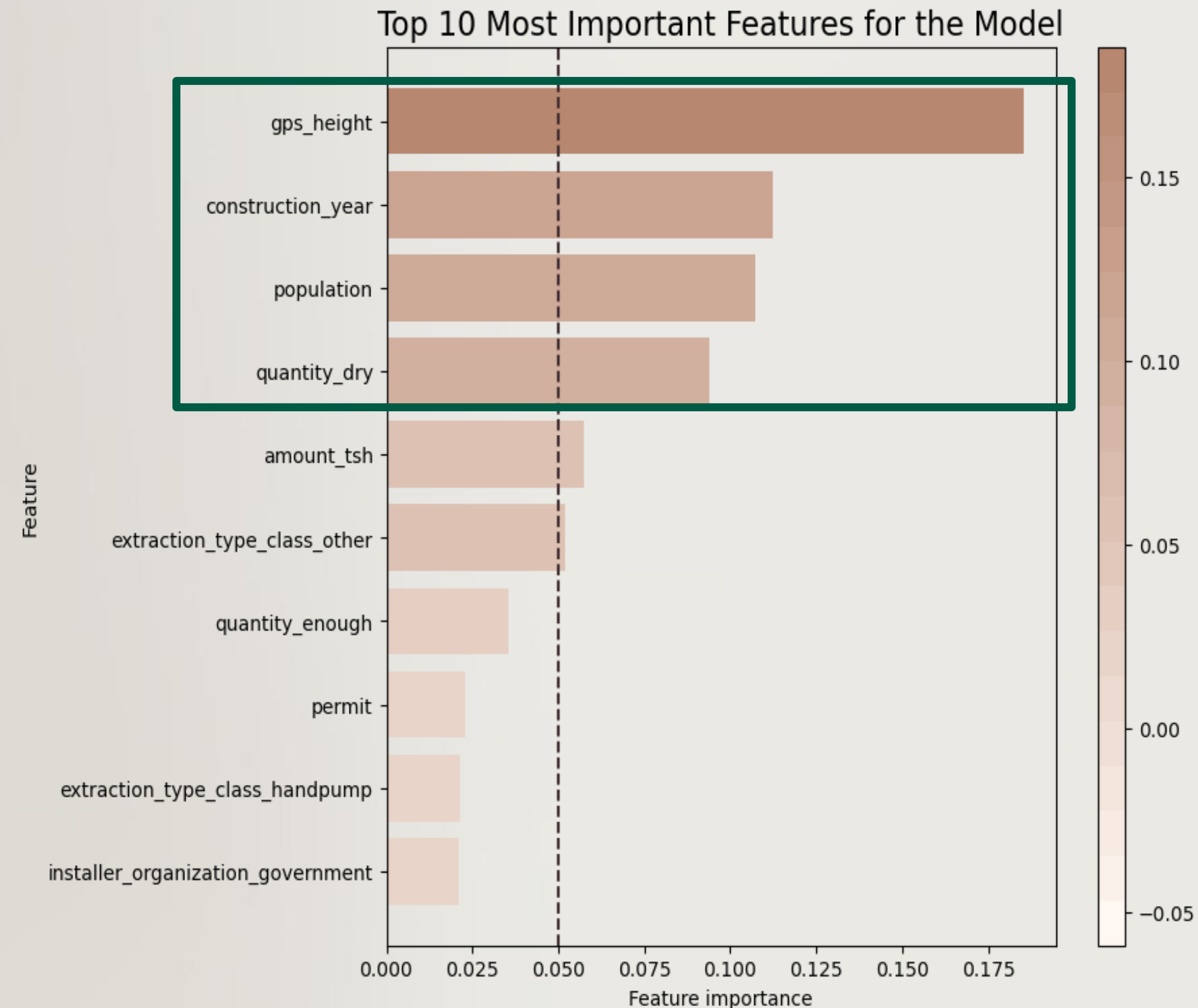
Start by:

1. Ensuring the well is indeed non-functional
2. Then, allocate resources and plan to repair ASAP

The wells with:

- a) A higher altitude
- b) An older construction date
- c) A larger population depending on it
- d) A quantity level identified as dry

Once these wells are repaired, focus on the next section of features indicated.







# 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](mailto:albane.colmenares@gmail.com)



The background of the image is a solid brown color, densely populated with numerous water droplets of varying sizes. Each droplet is rendered with a 3D effect, featuring a dark brown outline and a lighter brown highlight on its upper left side, giving them a realistic, glistening appearance. The droplets are scattered across the entire frame, creating a textured, organic pattern.

*Thank you!*