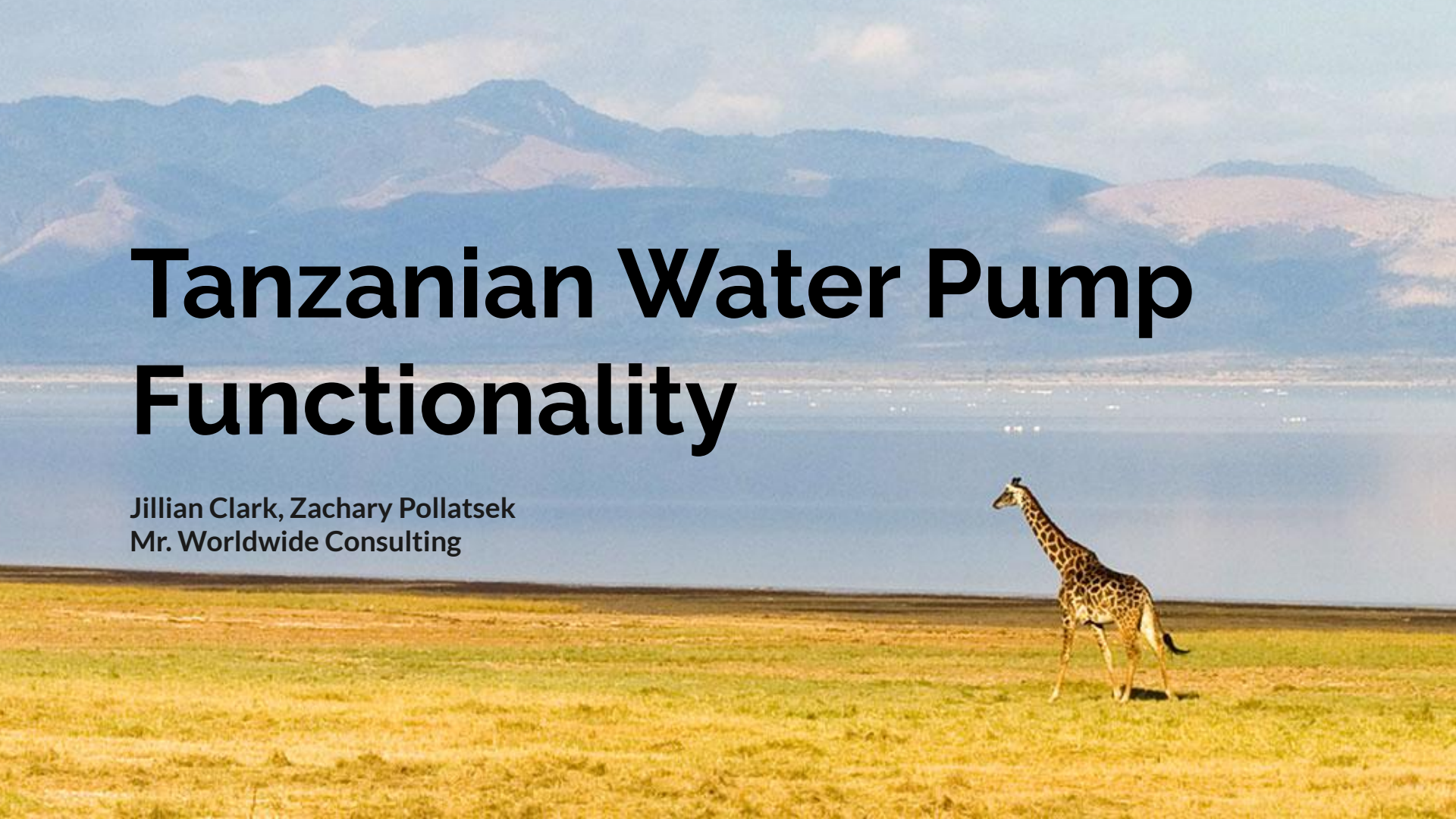
Tanzanian Water Pump Functionality

Jillian Clark, Zachary Pollatsek
Mr. Worldwide Consulting



Agenda

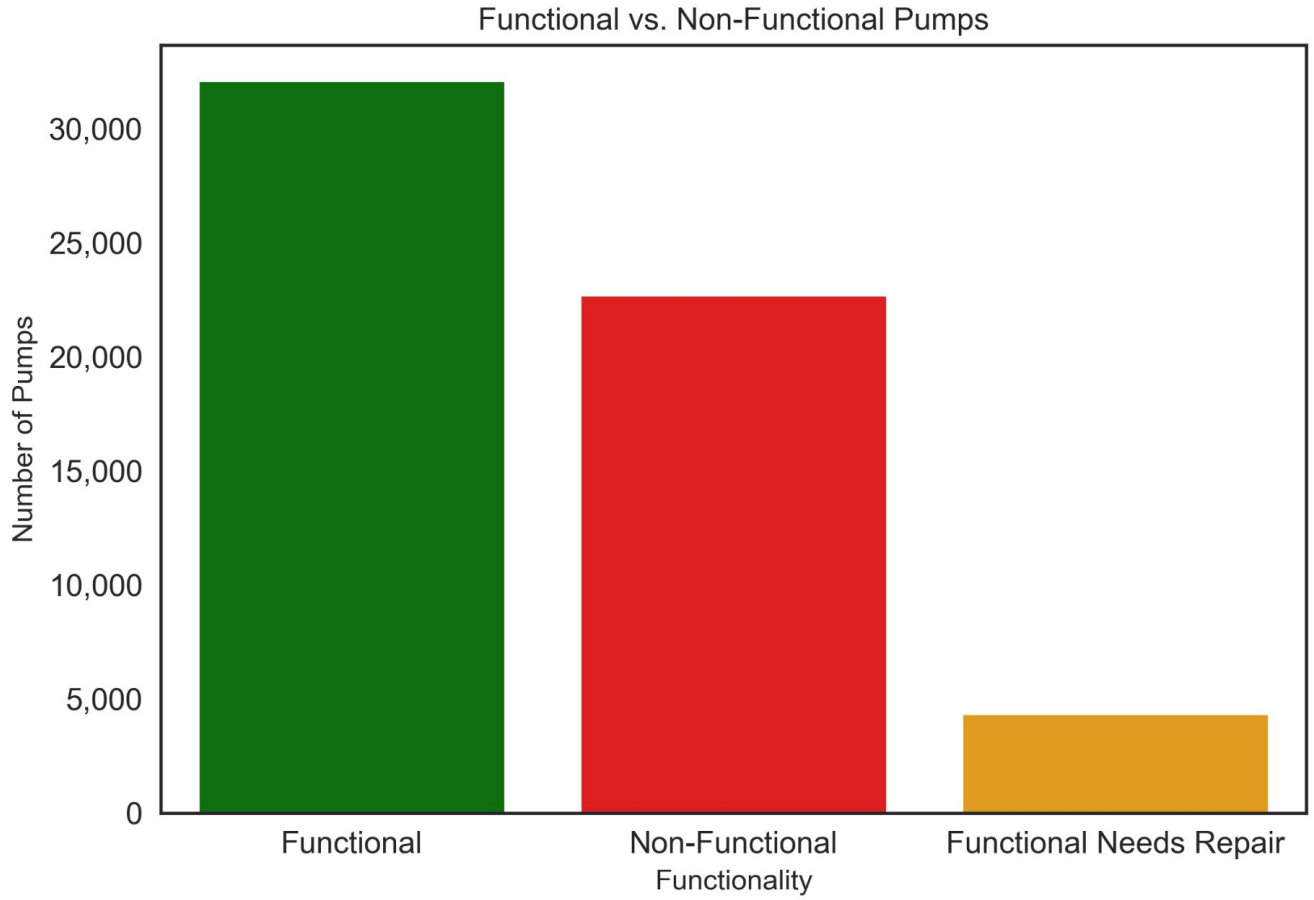
Problem

Modeling

Next Steps

Problem to Solve

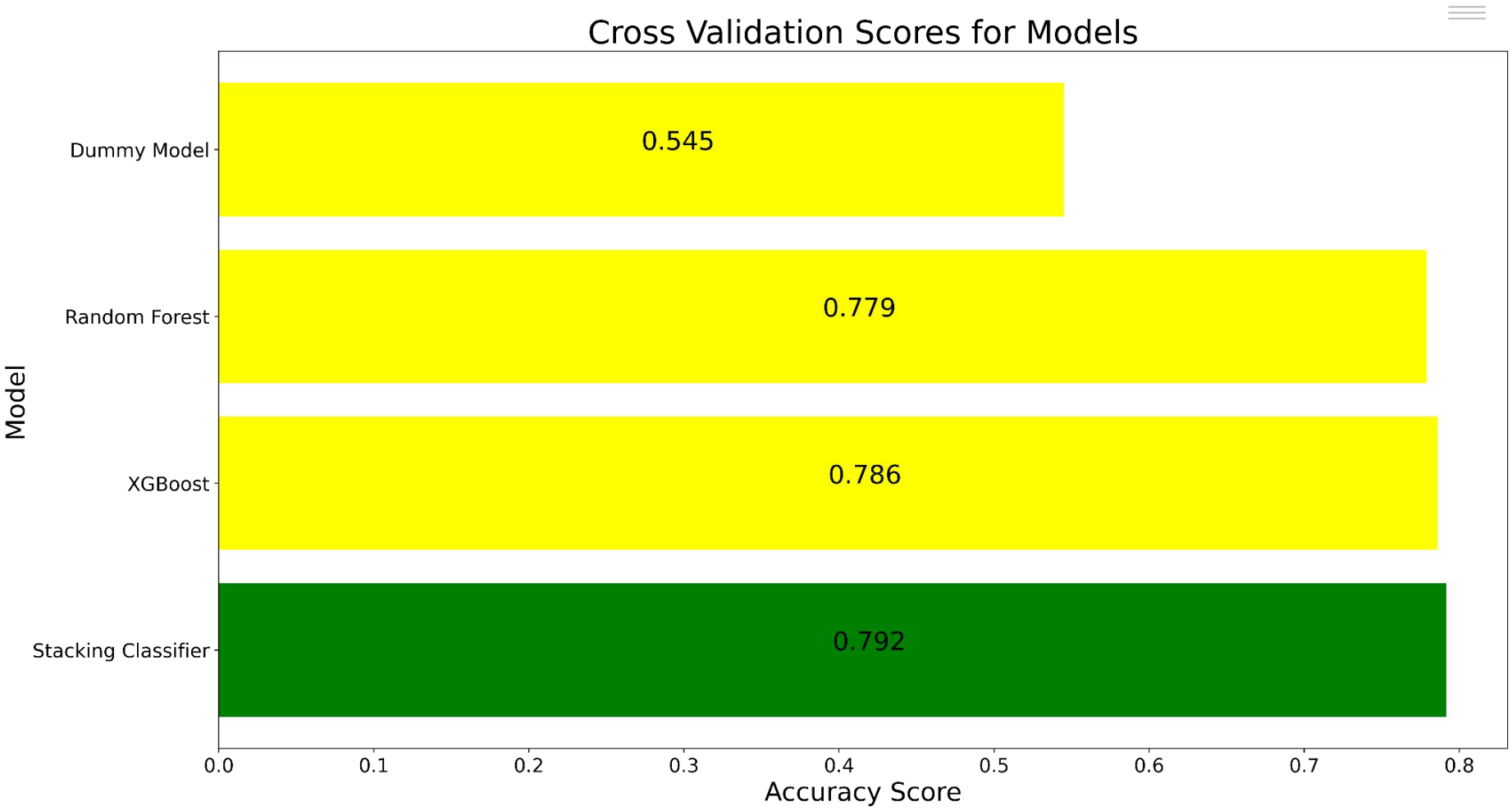
- Which pumps are functional, which need repairs, and which don’t work at all?
- Ensure no one is lacking access to clean, potable water
- Maximize time and resources to repair pumps


Functional vs. Non-Functional Pumps
Number of Pumps
30,000
25,000
20,000
15,000
10,000
5,000
0
Functional
Non-Functional
Functional Needs Repair
Functionality

Data Overview

- Data from over 59,000 water pumps provided by Taarifa and the Tanzanian Ministry of Water
- Variables include information about what kind of pump is operating, when it was installed, location, and how it is managed

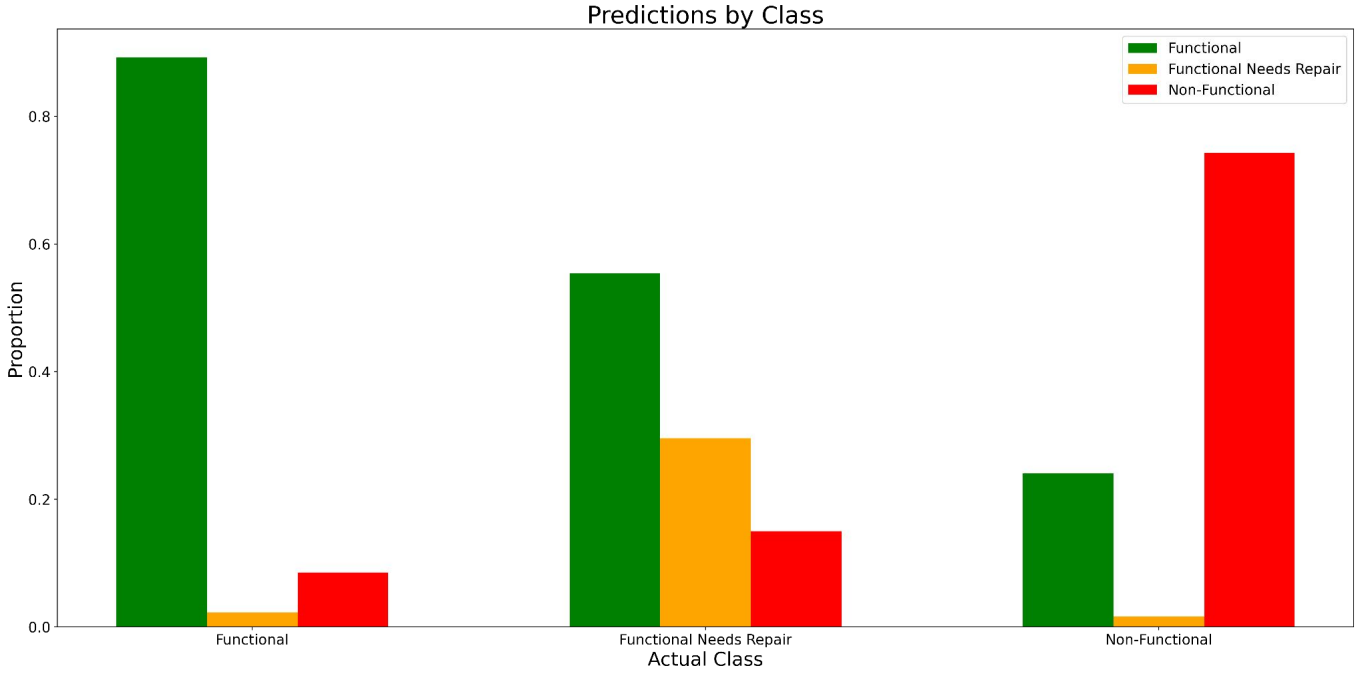

Prediction Model

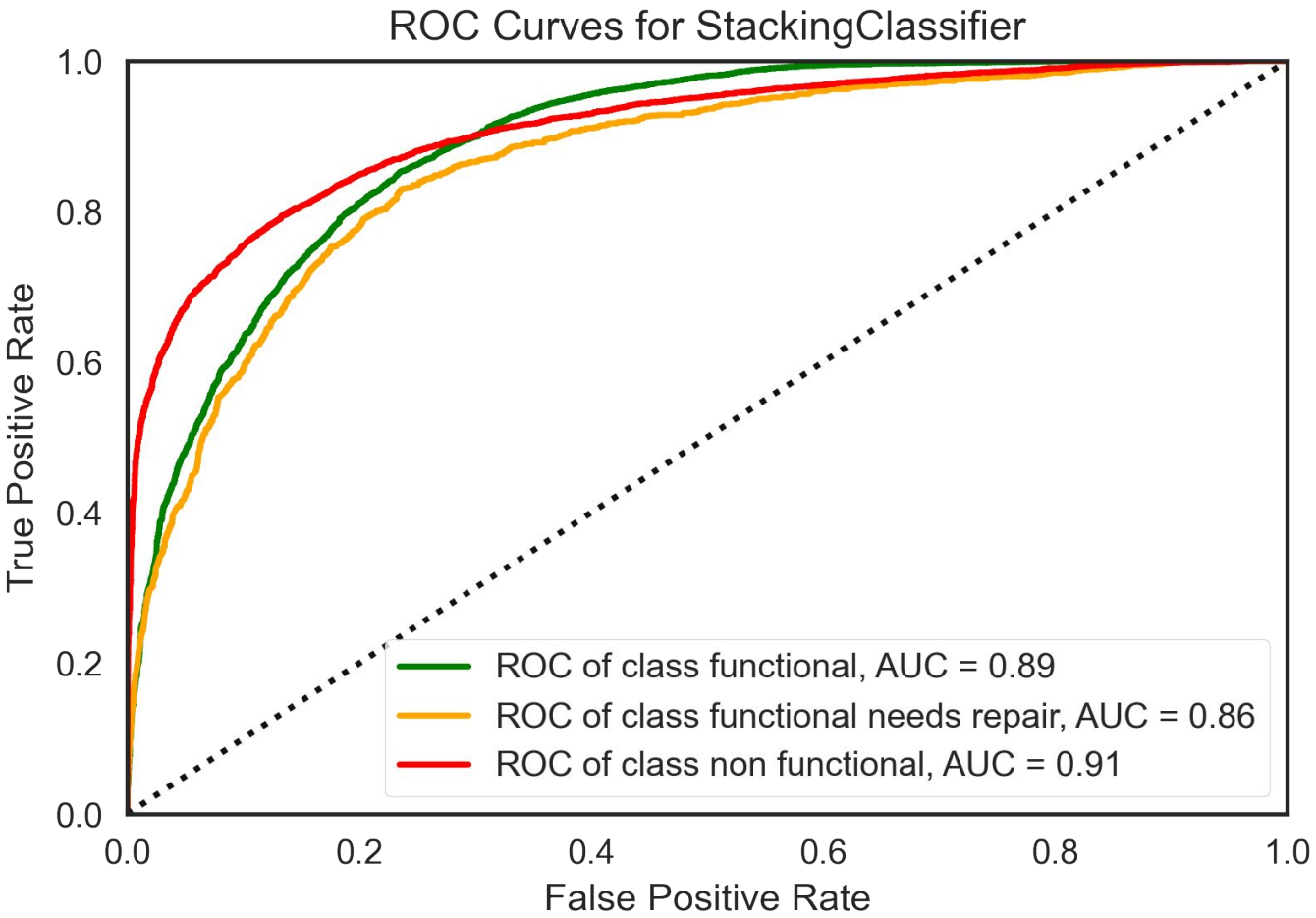

Cross Validation Scores for Models
Dummy Model
Random Forest
XGBoost
Stacking Classifier
0.545
0.779
0.786
0.792
Model
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Accuracy Score

Final Model Results

- Unseen Data Accuracy: 79.0%
- Model appears to be slightly overfit
- Best at classifying functional and non-functional

Final Model Results




ROC Curves for StackingClassifier
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC of class functional, AUC = 0.89
ROC of class functional needs repair, AUC = 0.86
ROC of class non functional, AUC = 0.91

Limitations

- Class imbalance
- Time to run models
- Lack of domain expertise

Next Steps

- Treat as binary classification
- Take more time to fine tune model
- Explore other models
- Consult with domain expert

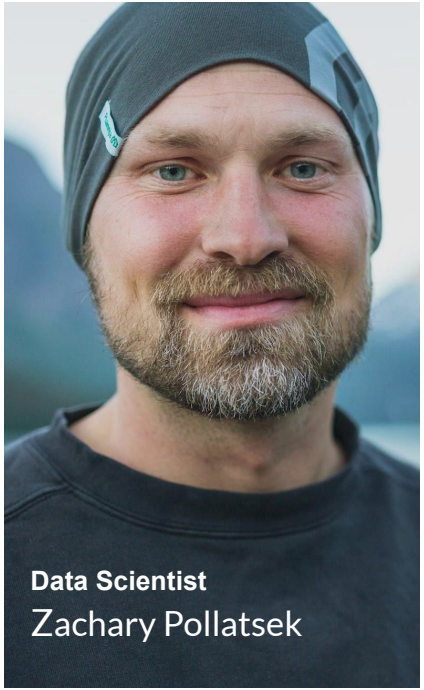
Conclusion

- Predict well functionality with 80% accuracy
- Maximize time and resources to repair pumps
- Minimize number of non-functional pumps to ensure clean and potable water for all

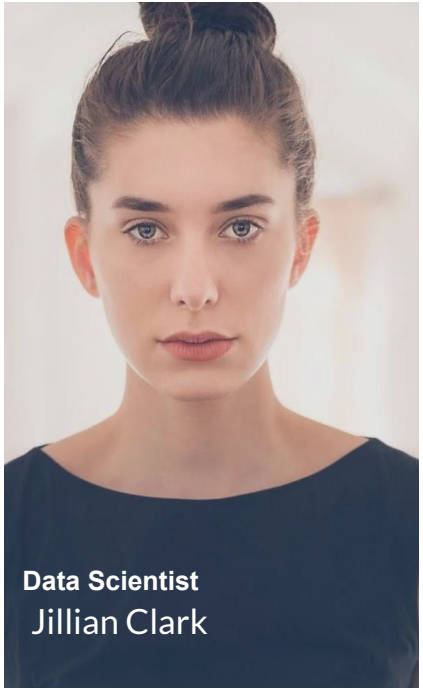
Thank you!

Any questions?

Website



LinkedIn
GitHub



LinkedIn
GitHub