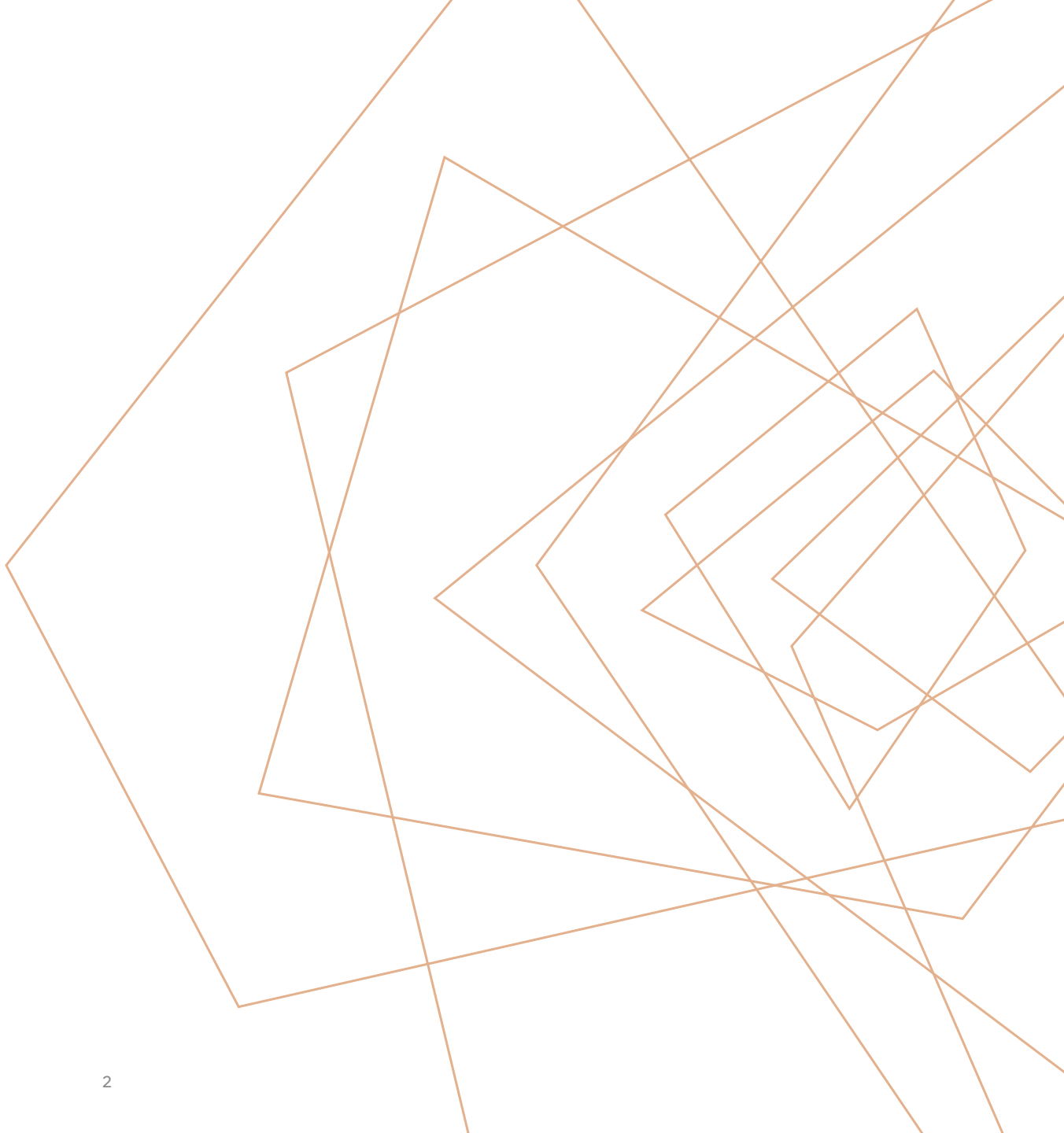
TITLE: TANZANIAN WATER WELLS STATUS PREDICTION

SUBTITLE: DATA ANALYSIS & MACHINE LEARNING APPROACH (OKECH MAX)

INTRODUCTION

Context: Water access issues in Tanzania and the impact of faulty water pumps

- **Objective**: Diagnose pump status, classify into categories, and inform stakeholders
- **Importance**: Improving public health through proactive maintenance



BREAKDOWNS

Frequent malfunctions of water pumps lead to severe water access issues

CLASSIFICATION

Classification challenge:

- Functional (Fully Operational)
- Functional Needs Repair (Partially Functional)
- Non Functional (Faulty)

ACTION POINTS

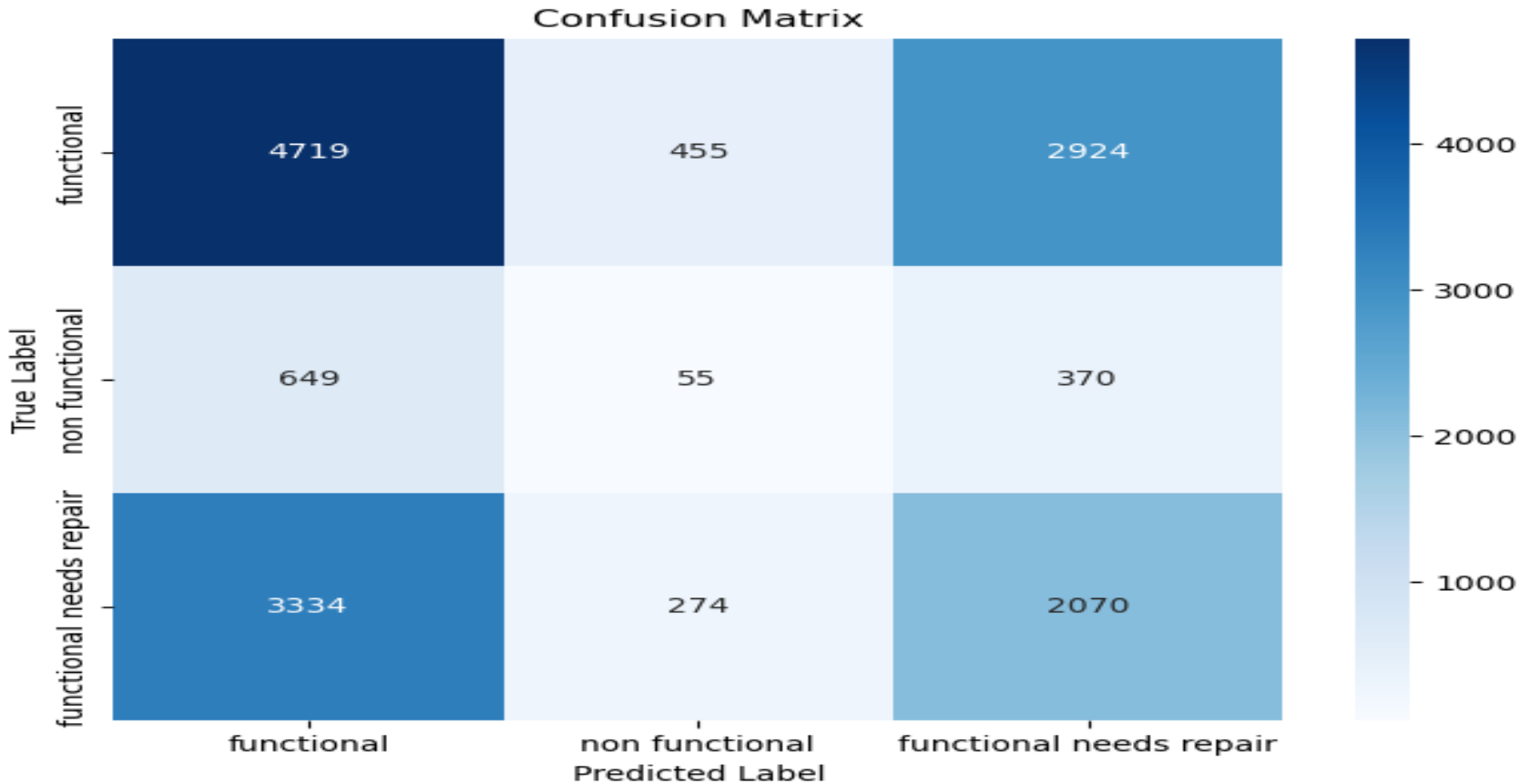
Transforming data into actionable insights

PROBLEM

Challenges

- **Class Imbalance:** Fewer malfunctioning pumps relative to fully functional ones
  - **Data Quality:** Noisy sensor data and incomplete maintenance logs
  - **Changing Conditions:** Environmental factors affecting pump performance over time

```
Confusion Matrix:
[[4719  455 2924]
 [ 649   55  370]
 [3334  274 2070]]
```



	precision	recall	f1-score	support
functional	0.54	0.58	0.56	8098
non functional	0.07	0.05	0.06	1074
functional needs repair	0.39	0.36	0.37	5678
accuracy			0.46	14850
macro avg	0.33	0.33	0.33	14850
weighted avg	0.45	0.46	0.45	14850

SUMMARY

Our model could predict 54% of the occurences in the functional class, and 39% of the functional needs repair class. It, however, could only predict 7% of the non-functional class.For the functional and functional-needs-repair classes f1-scores of 0.58 and 0.37 reflects a poor performance by the model. For the non-functional class ,moreover, a f1-score of 0.06 indicates that both precision and recall are low for this class. The actual number of occurences "support" for the non-functional class is very small compared to the other classes. The model generally had a poor score only being able to predict less than 10% of the non-functional class and barely 60% for the functional class which was the best performing class. A low number of samples is most likely the reason for the poor precision, recall and f1 score