

An aerial photograph of a vast, calm body of water, likely a lake or bay, with a green, hilly shoreline on the left. In the distance, a range of mountains is visible under a clear blue sky. A small blue dot is located in the top right corner of the image. A large white oval is superimposed on the left side of the image, containing the title and presenter information.

# **TANZANIA WATER POINT PREDICTION**

Presented by Todd Strain

# AGENDA

1. Overview
2. Business and Data Understanding
3. Modeling
4. Evaluation
5. Next Steps
6. Questions
7. Thank You

# OVERVIEW

This project analyzed data from the Tanzania Ministry of Water to determine the state of water points in the country.

We tried several Machine Learning models to get the best predictions.

We are able to predict the state of water points with 78% accuracy.



# BUSINESS AND DATA UNDERSTANDING

## **Make predictions**

The Tanzanian Ministry of Water has engaged us to predict which water pumps are functional, which need repairs and which don't work at all.

## **Our Data**

We've been provided with a database holding information on 15,000 water points and over 30 attributes for each pump.

## **Good Quality Data**

Information is almost complete for the 15,000 pumps with very few missing values.





# MACHINE LEARNING MODELING

We used two types of models, Logistic Regression, and Decision Tree, with several iterations of each.

## **Logistic Regression**

A very simple model to see how well Logistic Regression will fit to the data.

## **Logistic Regression with Tuned Parameters**

We used the simple model as a based and performed model tuning in an attempt to get a more accurate model.

## **Decision Tree Classifier**

This is a different type of model that produced slightly better results than the Logistic Regression model. Untuned and tuned versions were tried.

## **Random Forrest Classifier**

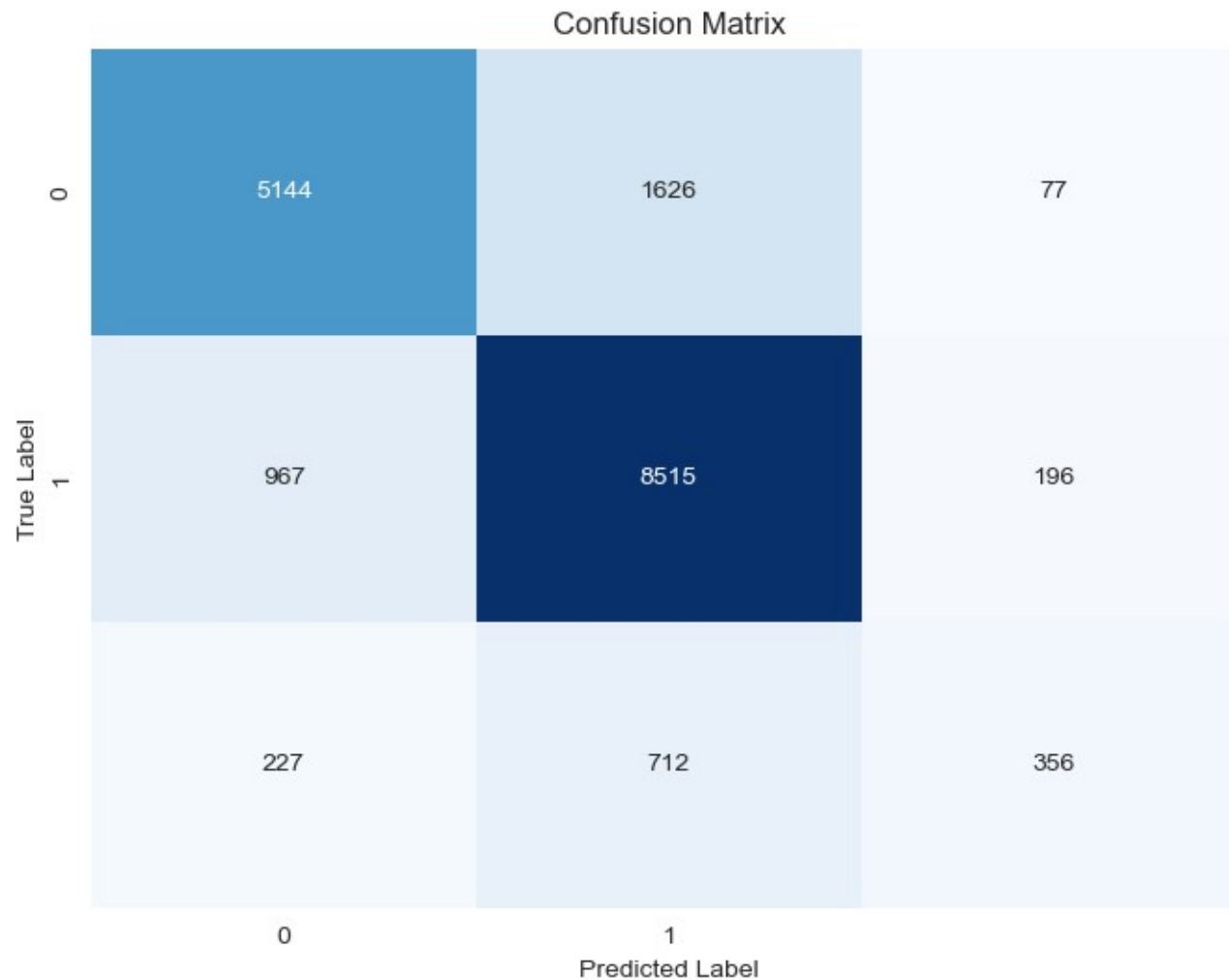
Another type of Decision Tree which produced the best results. Our final version had tuned parameters.

# EVALUATION

We were able to correctly predict a working point 8515 times out of 9678. That is a 88% success rate.

Non working points we correctly predicted 5144 out of 6847, a 75% average.

Pumps that were working and need repair were 356 correct predictions out of 1295. That is 27%



# 78%

Overall correct prediction rate

# NEXT STEPS

## **More sophisticated models**

More Machine Learning models are available than we tried in this project.

## **Beter Data**

Each model we tried had similar results with small variations. This could be due to the limits of the available data.

## **Focus on points that need repair**

Our prediction rate was 27% for points that worked but needed repair. There is room for a hugh improvement here.



# QUESTIONS?

---

Please ask us anything.

