

Tanzania Water Problem

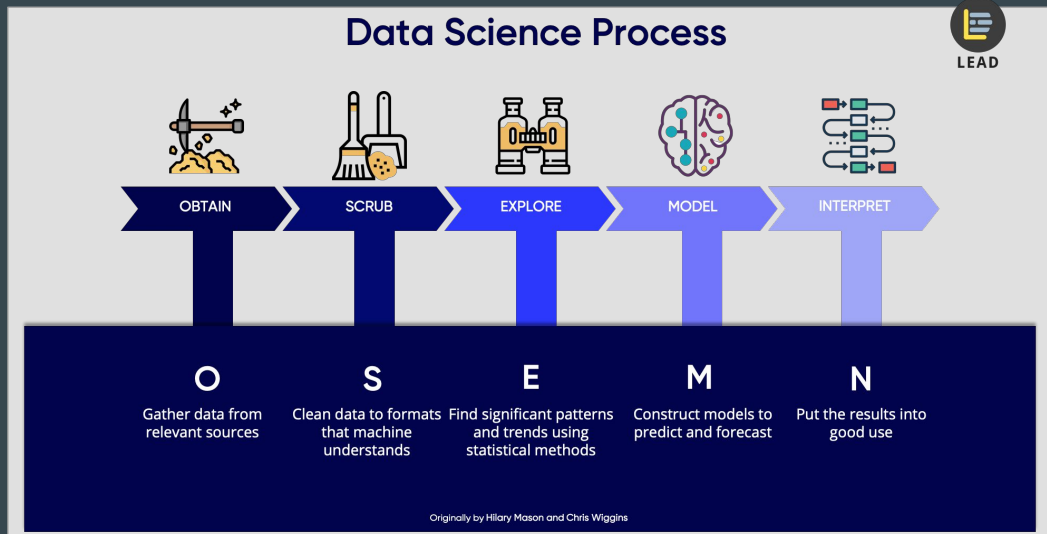
...

By: Cody Hoffman

Project Goals

Accurately predict which wells in Tanzania are broken down or need repair.

The Process

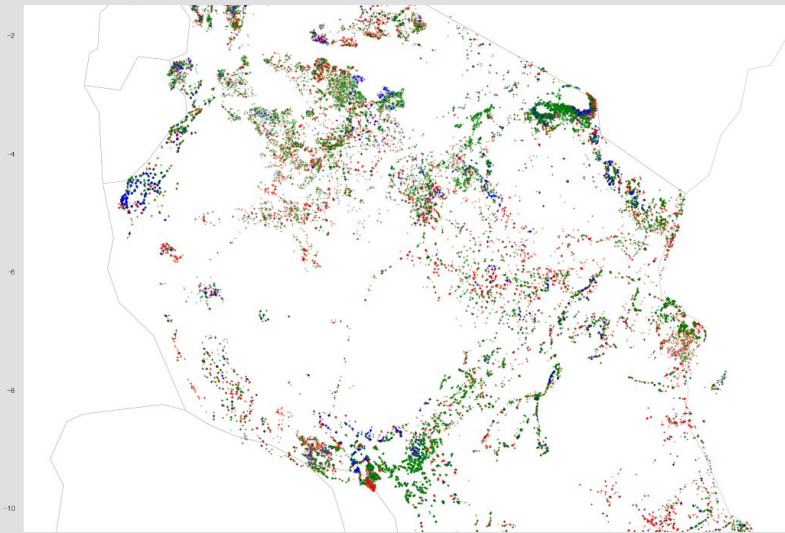


Obtained the data from a taarida for a challenge hosted by drivendata.

Modified data to be used in machine learning applications.

Construct models to predict status of wells.

Project value



From there we can use the information learned to quickly identify and process damaged wells for the betterment of the Tanzanian people.

With accurate enough predictions the Tanzanian Ministry of Water can spend less time manually inspecting wells and dispatch crews to already predicted problem areas.

This means that the system can potentially have fewer downtimes and be maintained at lower cost due to fewer wasted man hours.

Model Findings

Decision Tree
Classification

- 75% Accuracy

AdaBoost
Classification

- 72% Accuracy

K-Nearest Neighbor

- 70% Accuracy

After running 5 different machine learning models
3 stood out as viable.

With this information we have a 75% of knowing
before ever going out into the field whether a
pump needs serviced or not.



Future Work

1. Modify the feature selection.
2. Gather more data such as maintenance records.
3. Apply deep learning techniques.

1. Selecting different features to run through our models could potentially improve our models overall accuracy.
2. Maintenance records would allow us to see if a well has properly maintained which in turn would potentially improve prediction accuracy.
3. More sophisticated models may also provide better results with the same data.

Thank You