Project 3 – Tanzanian Waterpoint Classification

—

By Htoo Aung Latt

The Project – Goal

- Predicting the conditions of water points to allow repairs

- Insights into the reason why some water points are more easily broken
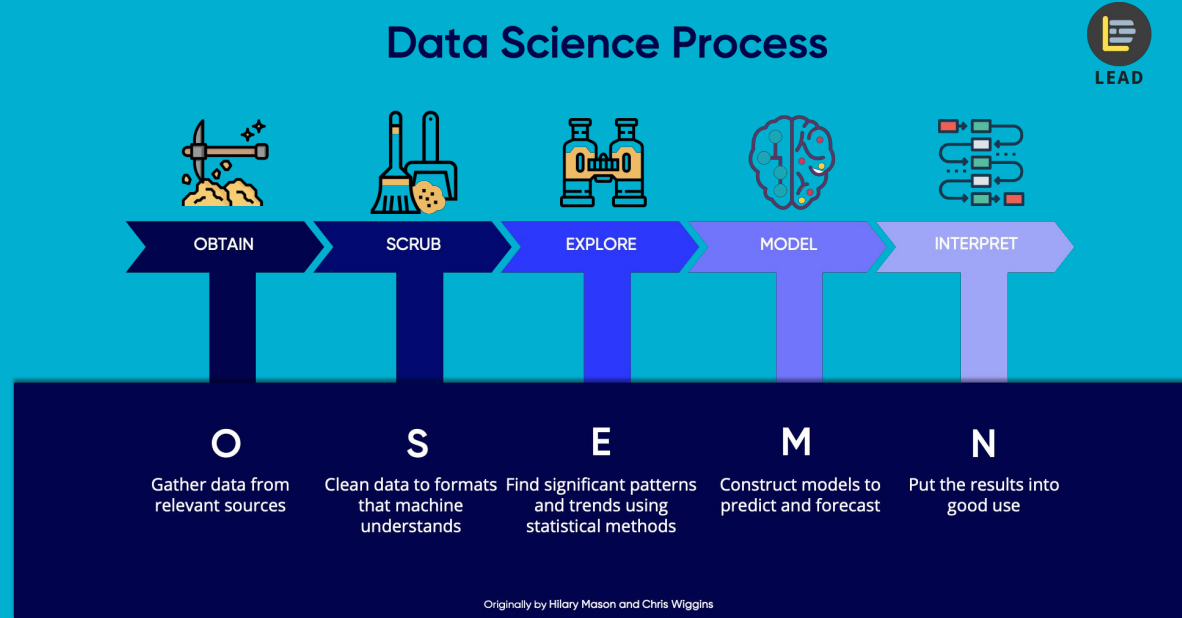
The Process – OSEMN
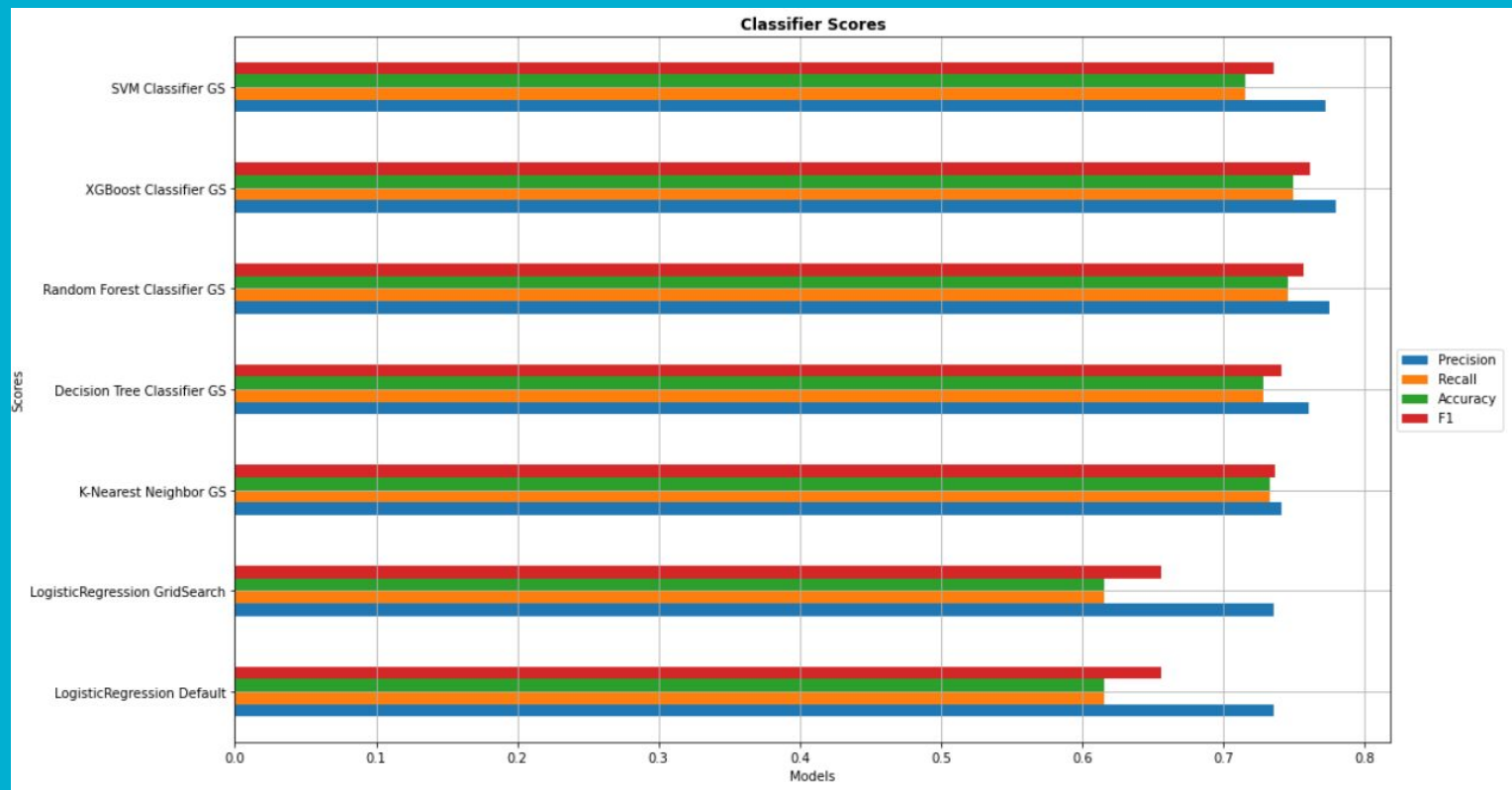
O - **O**btaining the Data

S - **S**crubbing the Data

E - **E**xploring the Data

M - **M**odeling the Data

N - i**N**terpreting the Data

Data Science Process

OBTAIN	SCRUB	EXPLORE	MODEL	INTERPRET
O	**S**	**E**	**M**	**N**
Gather data from relevant sources	Clean data to formats that machine understands	Find significant patterns and trends using statistical methods	Construct models to predict and forecast	Put the results into good use

Originally by Hilary Mason and Chris Wiggins

LEAD

Selecting The Model

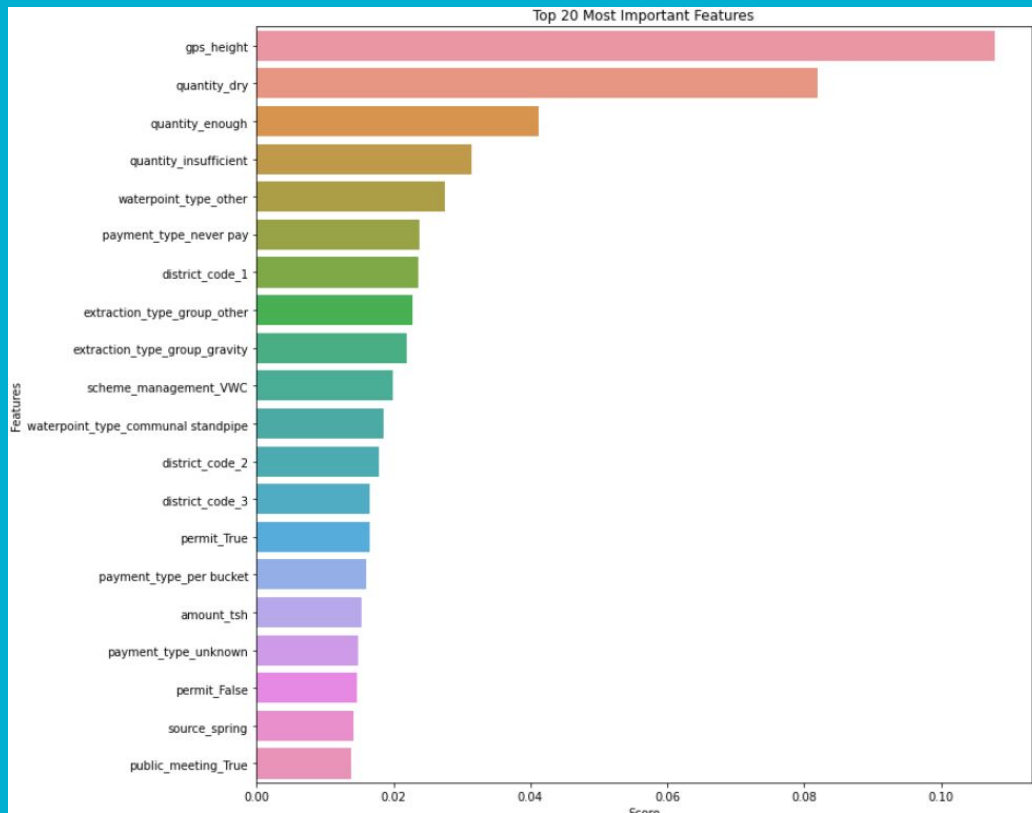
Classifier Scores

Random Forest vs. XGBoost vs. SVM

Balance between Recall and Accuracy

High recall leads to detecting more functional needs repair category but also raise the false positive. Which would lead to wasted manpower and budget.

Interpreting the Model

- Altitude of well
- Quantity
- Waterpoint type
- Extraction Type
- Payment Type
- District Code
- Public Meeting

Top 20 Most Important Features

Thank You