

Pump it Up: Data Mining the Water Table



Deepali Sharma
January, 2023



Tanzania begins water rationing due to drought



<https://www.africanews.com/2022/10/28/tanzania-begins-water-rationing-due-to-drought//>

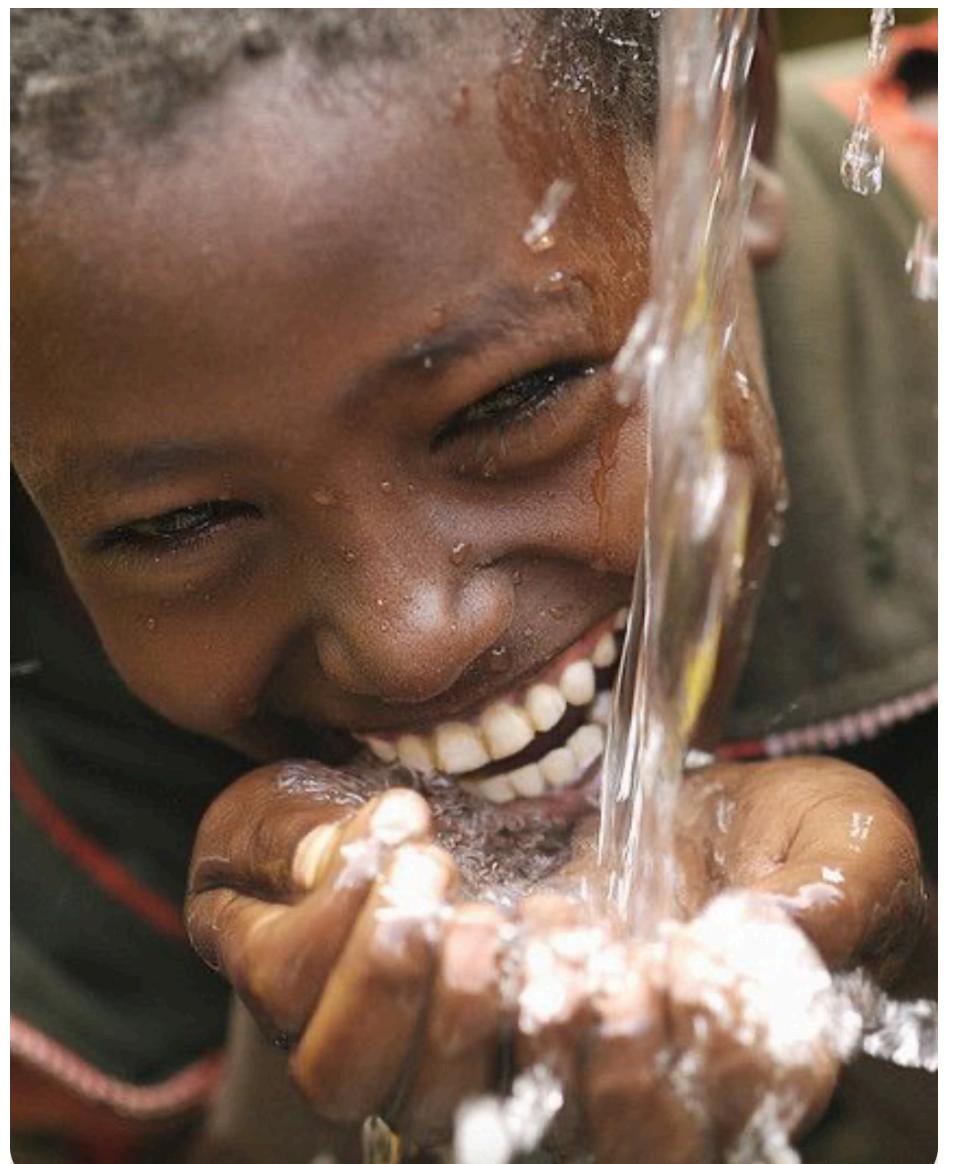
- **Stakeholder:**
 - Ministry of water; Government of Tanzania
 - Predict the pumps that are functional/faulty/need repairs
 - Proper Allocation of resources based on pump status
 - Find other factors that are responsible for non-functioning pumps
- **Business Problem:**

Data:



- Data collected by **Taarifa** and **Tanzanian Ministry of Water**
 - Information is provided for water pumps that contain geographical locations, quality of water, water extraction method, installed/funded by etc.etc

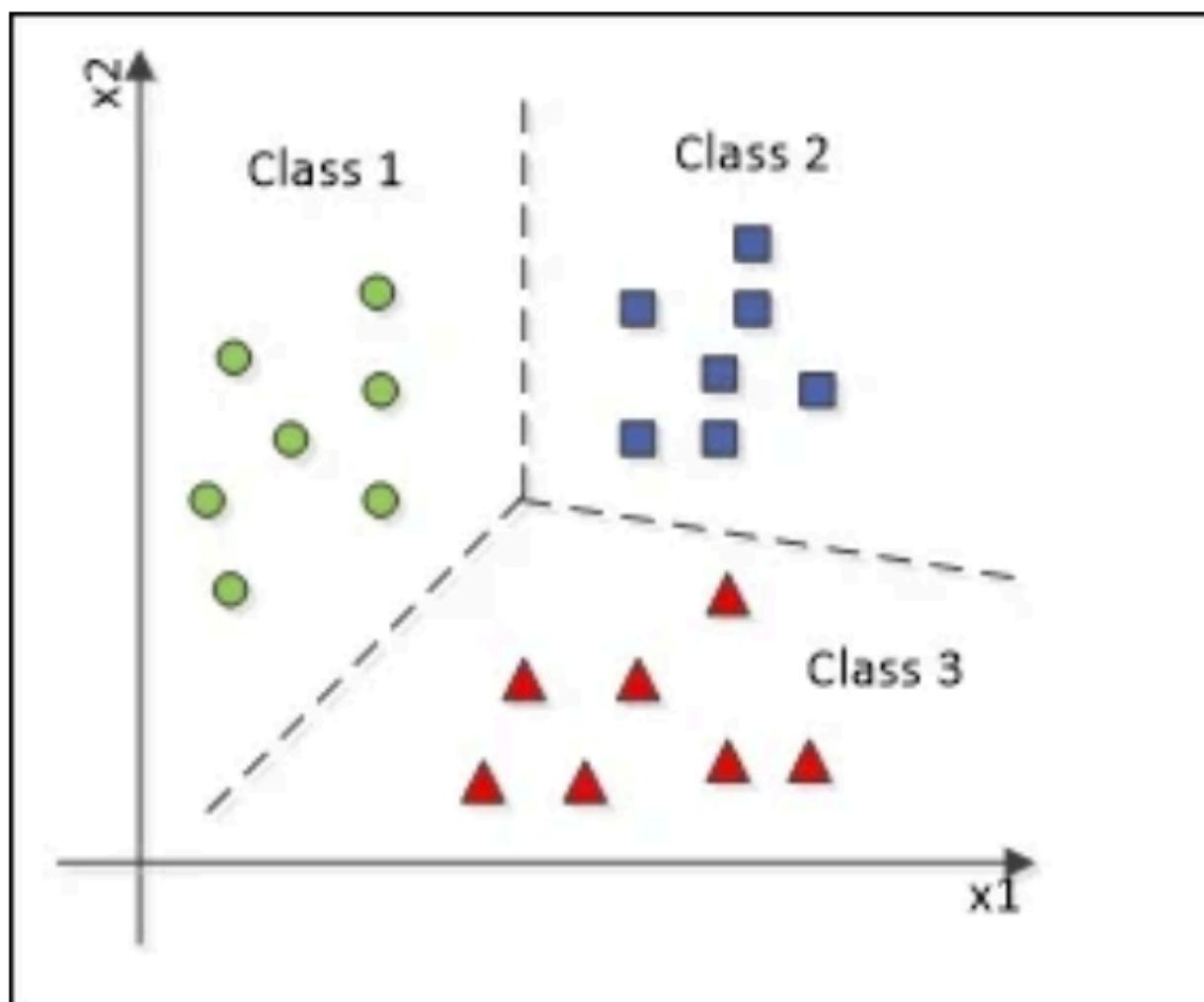
Goal:



- Goal is to build a **classifier** that would predict the status of water pump utilizing the available information
 - Using the predicted information about water pumps, Tanzanian Govt. can appropriately redistribute their crews and funds to fix non-functioning pumps
 - Find the most important features affecting the pump quality that provide insights into e.g, where not to build pumps, which extraction techniques to use etc.
 - Ultimate goal is to provide clean water access to everyone

Modeling Algorithm

- Clean and pre-process the data
- Find the model that best describes the data
- Check the model against validation and test data
- Find the most important features that affect pump functionality
- Use the knowledge gained to make recommendations that will improve the pump status



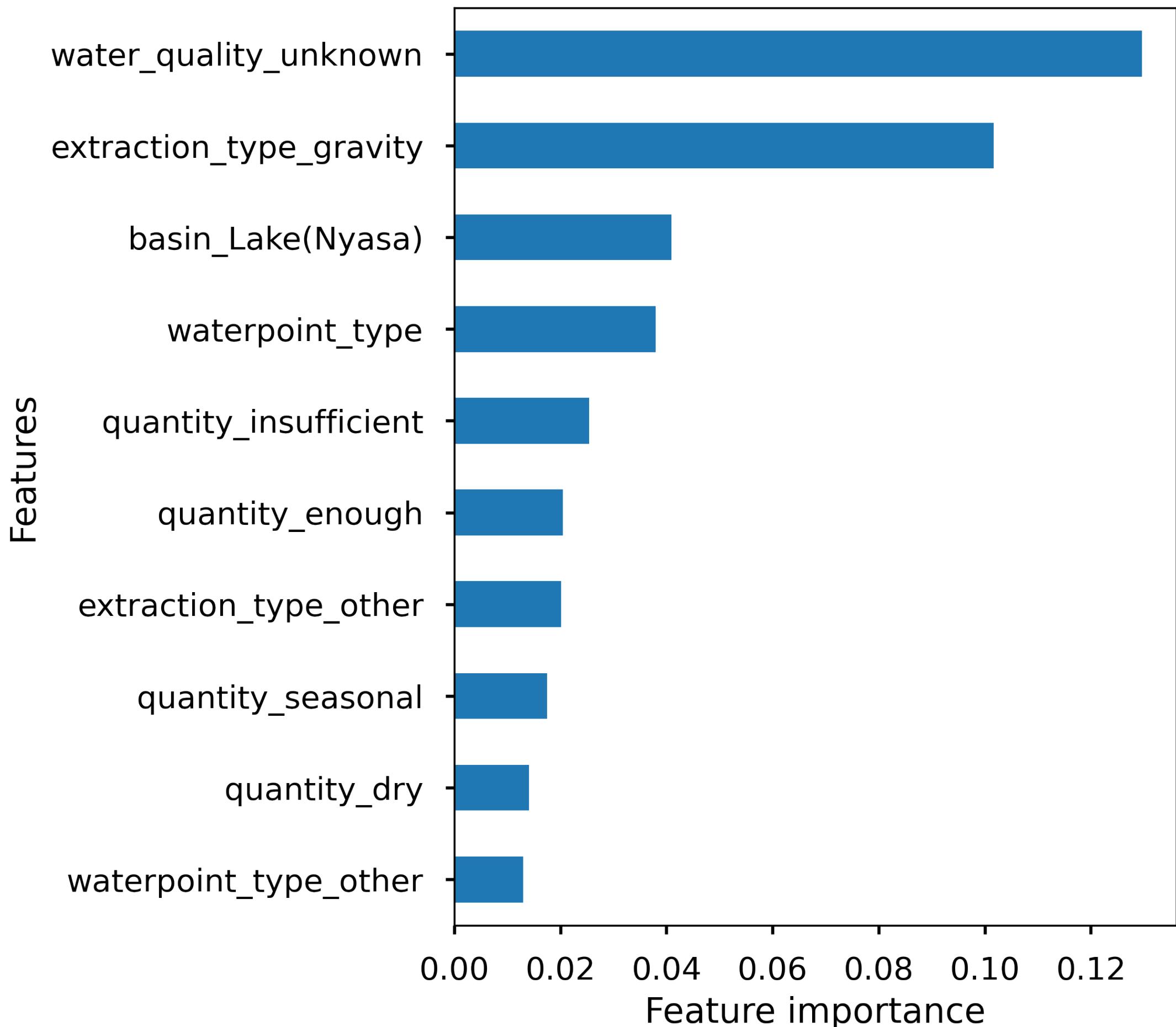
Classification Problem

Results from the Best Model:

Model Performance:

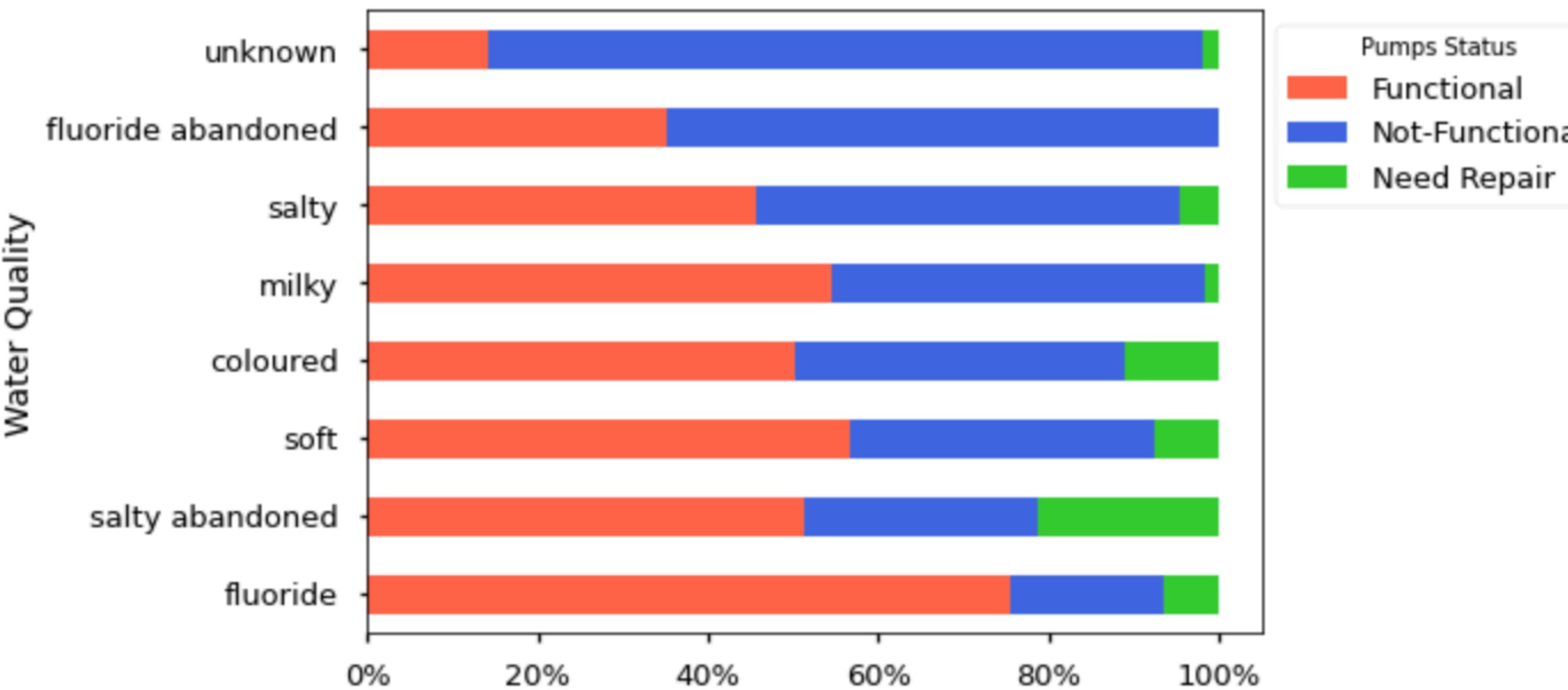
ROC_AUC	91%
ACCURACY	78%
RECALL	80%
PRECISION	78%

**Relative Importance of Top10 Features
for Predicting Water Pump Status**



Water Quality: the top feature

Relationship between Water quality and Pump status



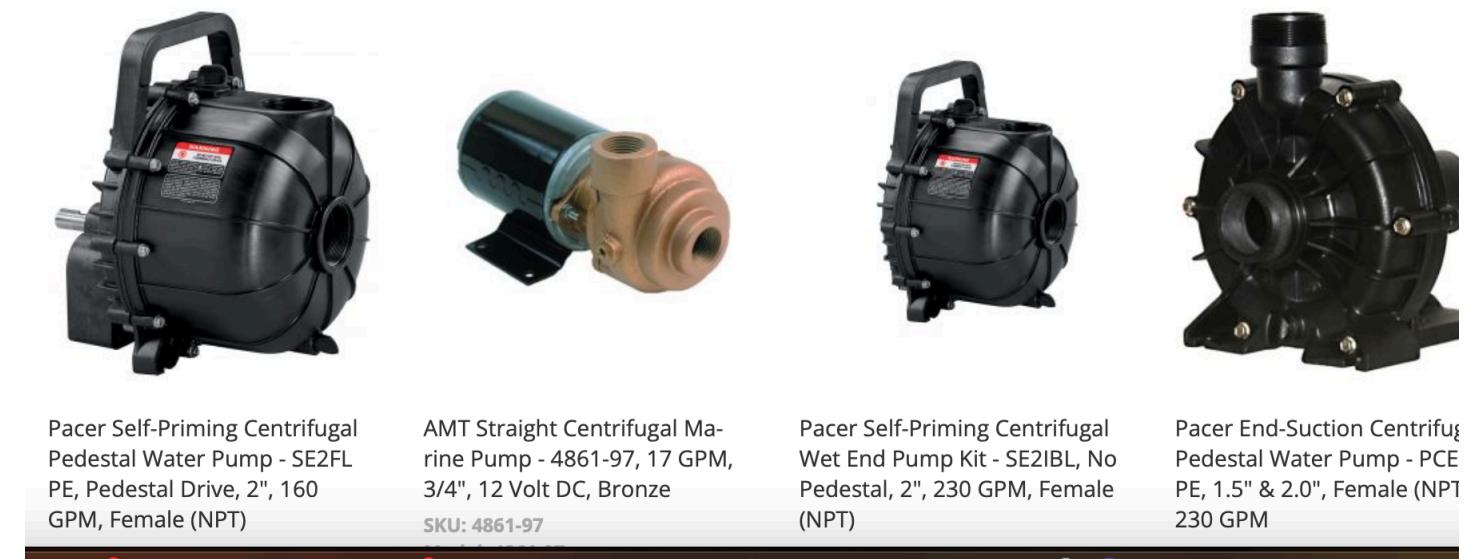
- Most of the **salty** water and **unknown** type water pumps fall in the category of non-functional/need repairs category

Salt Water Pumps

Salt water pumps are commonly equipped with materials such as stainless steel, aluminum, and thermoplastics, that perform better than cast iron in corrosive environments. Absolute Water Pumps suggests thorough flushing of your pump's inside and out with fresh clean non-salty and non-corrosive water after every use in a salt water or other corrosives environment.

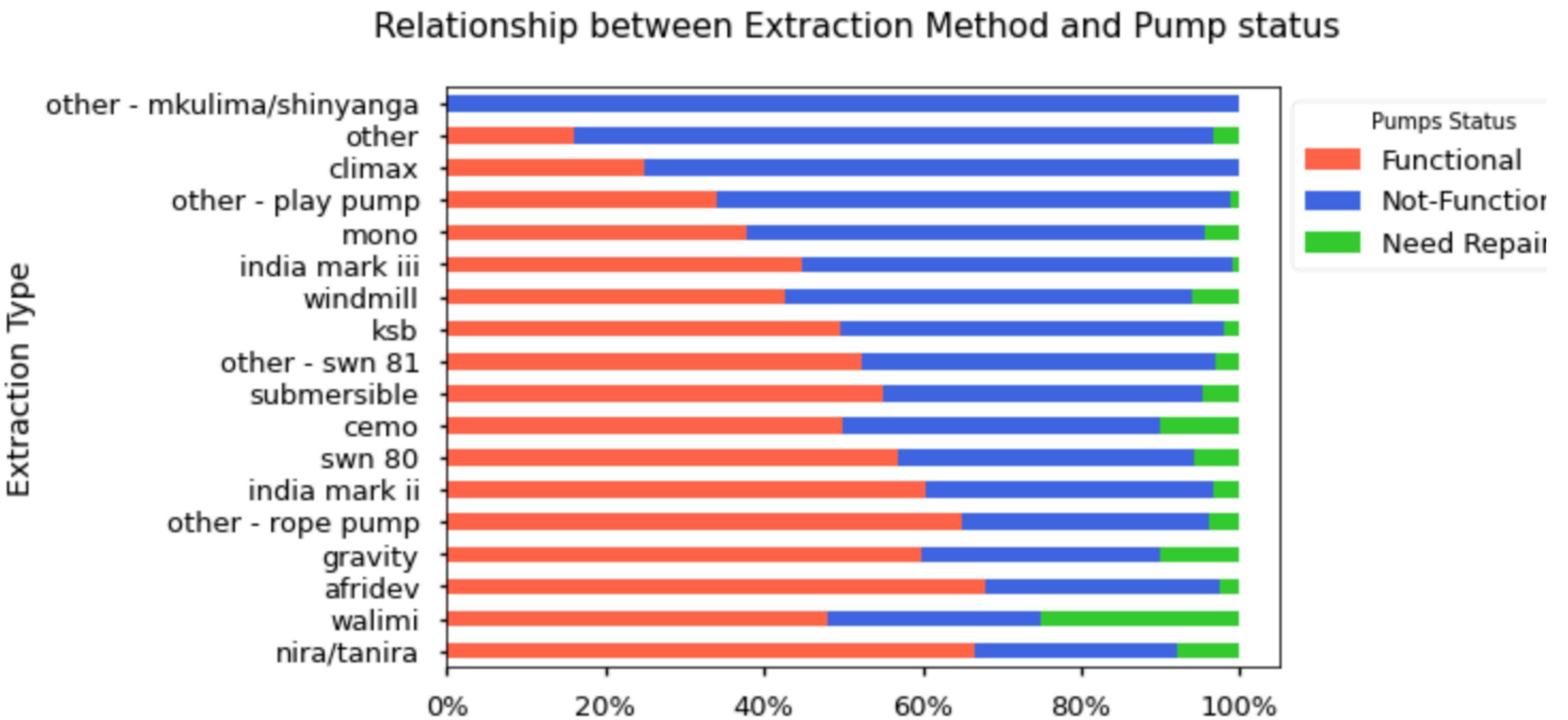
Items 1-12 of 427

Sort By Special Price ▾



- The government should prioritize fixing those water pumps with appropriate types such as Aluminum pumps for salty locations!

Extraction type: other important feature



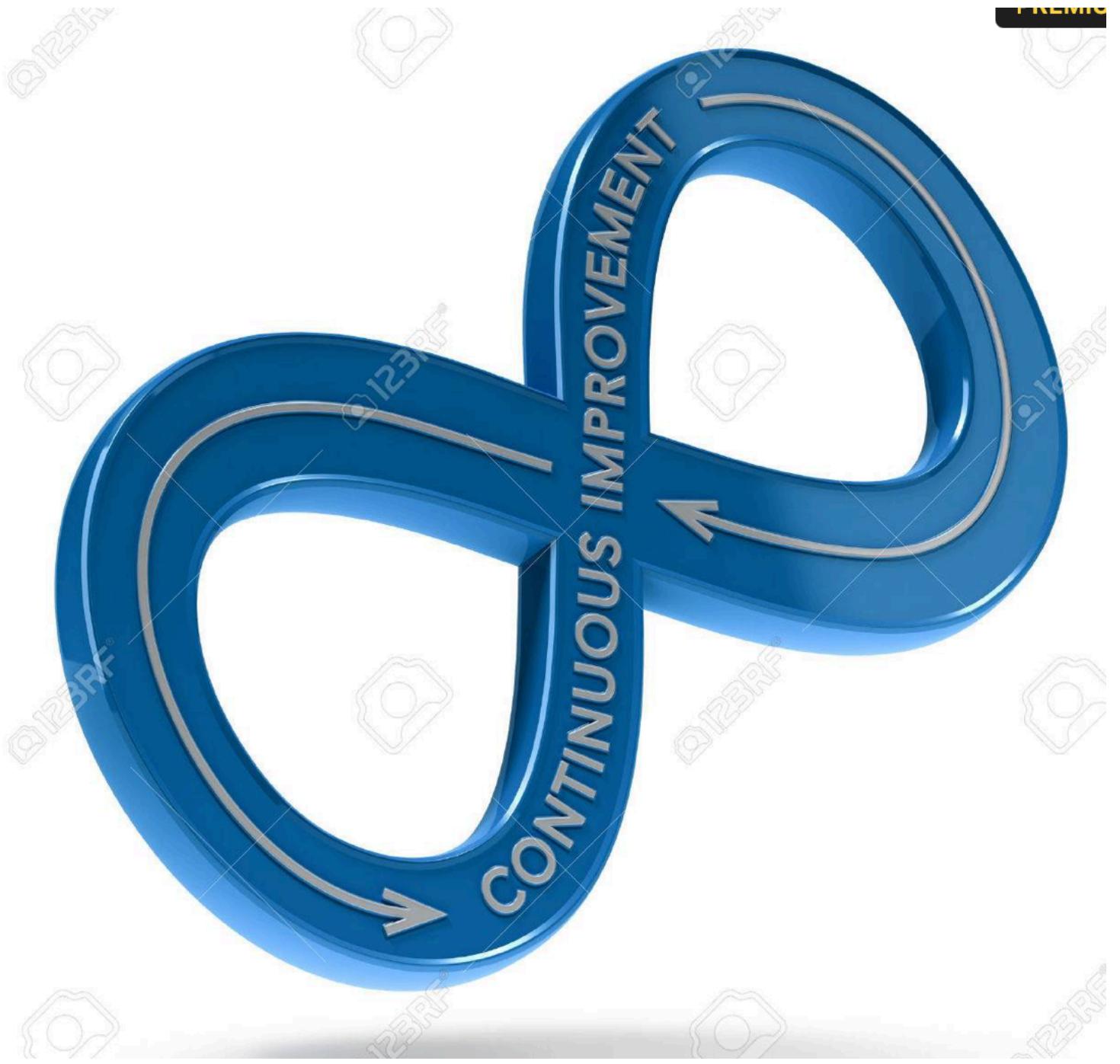
- Looks like the pumps with certain extraction types are performing worse!

- The government should only invest in the extraction techniques that are robust!

Some Recommendations:

- The Water ministry should reallocate funds to replacing pumps at salty locations with the correct types.
- Allocate some budgets for R&D to find out the right extraction type for a given location.

Limitations and Improvements



- Come up with better strategy to clean up some of the columns (installer, funder and some more)
- Find out if I can run GridSearchCV in reasonable amount of time
- Try some more classification models like CNN to see if they perform any better
- Explore if there is any dependence on the scoring metric used in theGridSearchCV
- Eliminate or quantify if the parameters such as installer, funder, scheme_name have effect on improving model performance

THANK YOU!



- **Deepali Sharma:** email:(deeps.sharma@gmail.com, deepali@rcf.rhic.bnl.gov)
- **LinkedIn:** <https://www.linkedin.com/in/deepali-sharma-a83a126/>
- **GitHub:** <https://github.com/deepssharma>