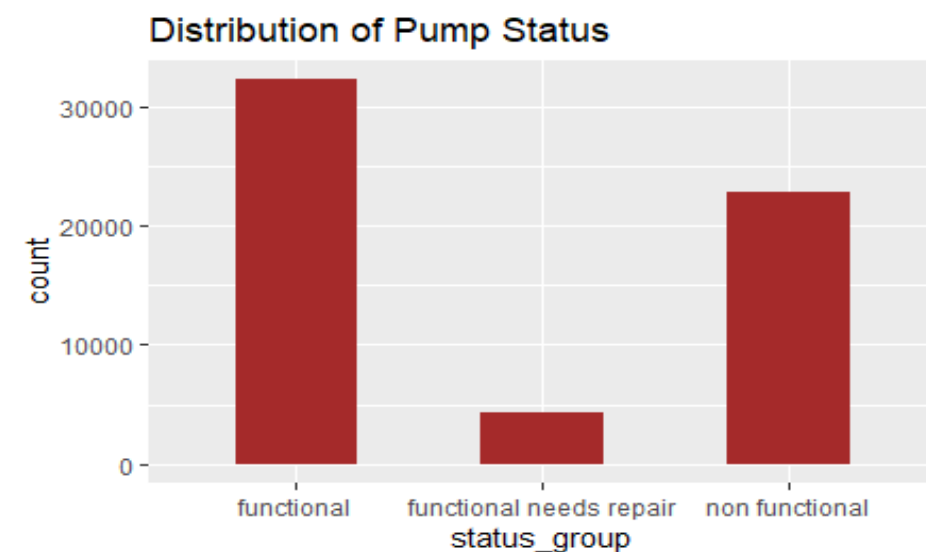


## Problem Statement

This dataset consists of details about the pumps that are available in Tanzania. A smart understanding of which waterpoints will fail can improve maintenance operations and ensure that clean, potable water is available to communities across Tanzania. In order to find the faulty pump, we are going to predict the condition of the pump(functional, non-functional, functional needs repair) based on a number of variables about what kind of pump is operating, when it was installed, and how it is managed.



## Data Description



Our dataset consists of 59400 rows and 41 columns with the details of pump locality, water extraction type, quality, quantity of water, source from where the pump gets the water and many more.

## Models Used

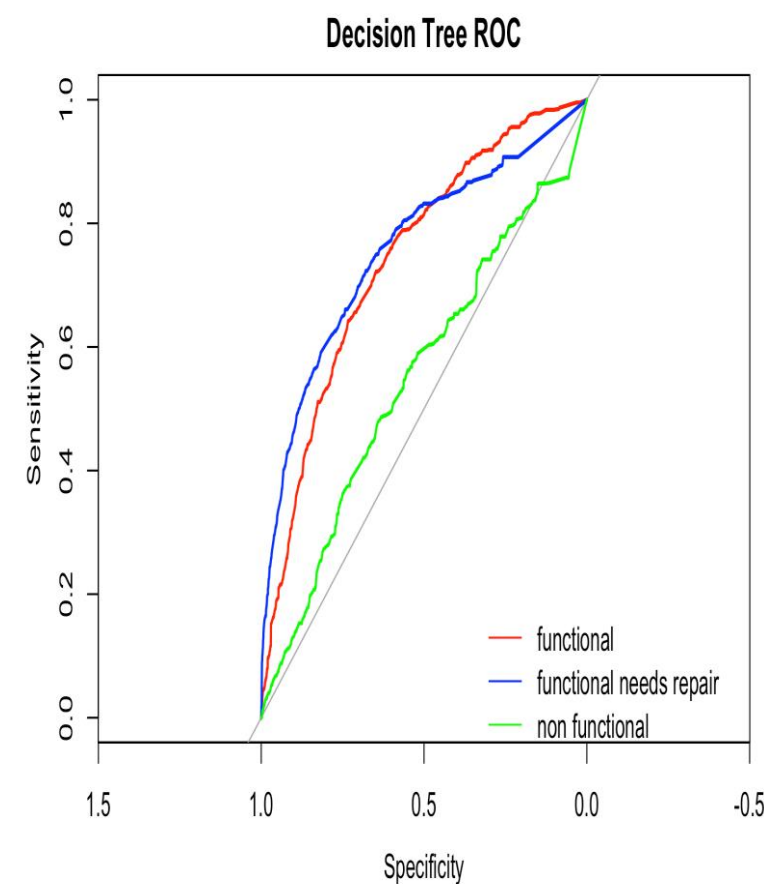
To predict the status of the pump operations, we are using different classification models.

MODEL	ACCURACY (IN %)
Decision Tree	75.52
Random Forest Classifier	77.69
Gradient Boosting Machines	73.72
Support Vector Machine	74.37
Extreme Gradient Boosting	75.16
Association Rule Mining	

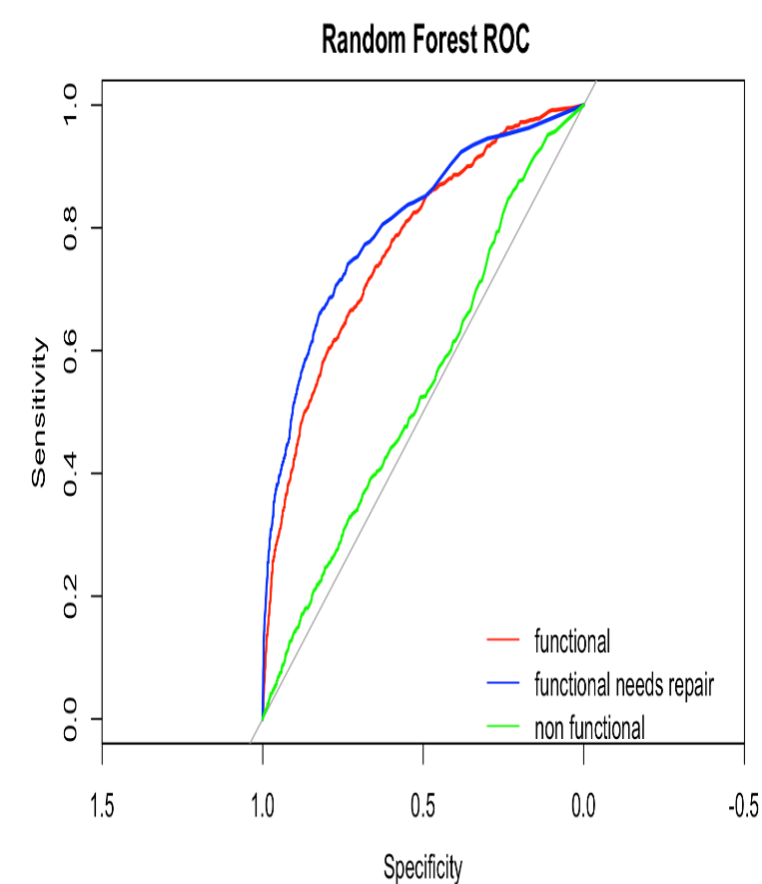
## Model Comparison:

We tested Decision Tree, Random Forest Classifier, Gradient Boosting Machine, Support Vector Machine and Extreme Gradient Boosting to predict the faulty pumps. Here, we have compared the best two models which has higher accuracy.

### Decision Tree



### Random Forest



## Performance Evaluation

Training and testing data were split into 80-20 split and below are the results of two models in form of confusion matrix.

		Decision Tree				Random Forest			
		Predicted				Predicted			
Actual		F	FNR	NF		F	FNR	NF	
	F	5527	510	1247		5631	486	1154	
	FNR	178	213	87		183	259	72	
	NF	746	140	3230		637	118	3338	

F – Functional

FNR – Functional needs repair

NF – Non functional

## Conclusion

Based on various evaluation metrics for different tasks, Random forest model was time consuming. Decision Tree model is 20 times faster than Random Forest. Therefore, it's a trade-off between better accuracy and time complexity.

ALGORITHM	ACCURACY	TIME TAKEN
Decision Tree	75.52%	~61 sec
Random Forest	77.69%	~1142 sec

## Shiny App

We have created a web application that will allow the user to modify various parameters in order to predict the condition of the pump.