



# **Machine Learning for Water Crisis: Identifying Non-Functional Wells in Tanzania**


**Jennifer Casias**

# Tanzania




# Data


59,400 entries and 40 features



[Tenders](#) [Contact Us](#) [F.A.Q's](#) [Staff Mail](#) [En](#) [Sw](#)



## The United Republic of Tanzania Ministry of Water




[Home](#) [About Us](#) [Instrument](#) [Projects and Programmes](#) [Document Repository](#) [Media Center](#) [Contacts for Uwssa's](#) [Water Charges Price](#) [PSP](#) [Maji Week](#)

### Wizara

- [The Minister](#)
- [Deputy Minister](#)
- [Management Team](#)
- [Divisions](#)
- [Units](#)
- [Ministry Institutions](#)

### Budget Speeches

- [Budget Speeches](#)
- [Questions and Answers](#)



Rais wa Jamhuri ya Muungano wa Tanzania Mhe. Samia Suluhu Hassan

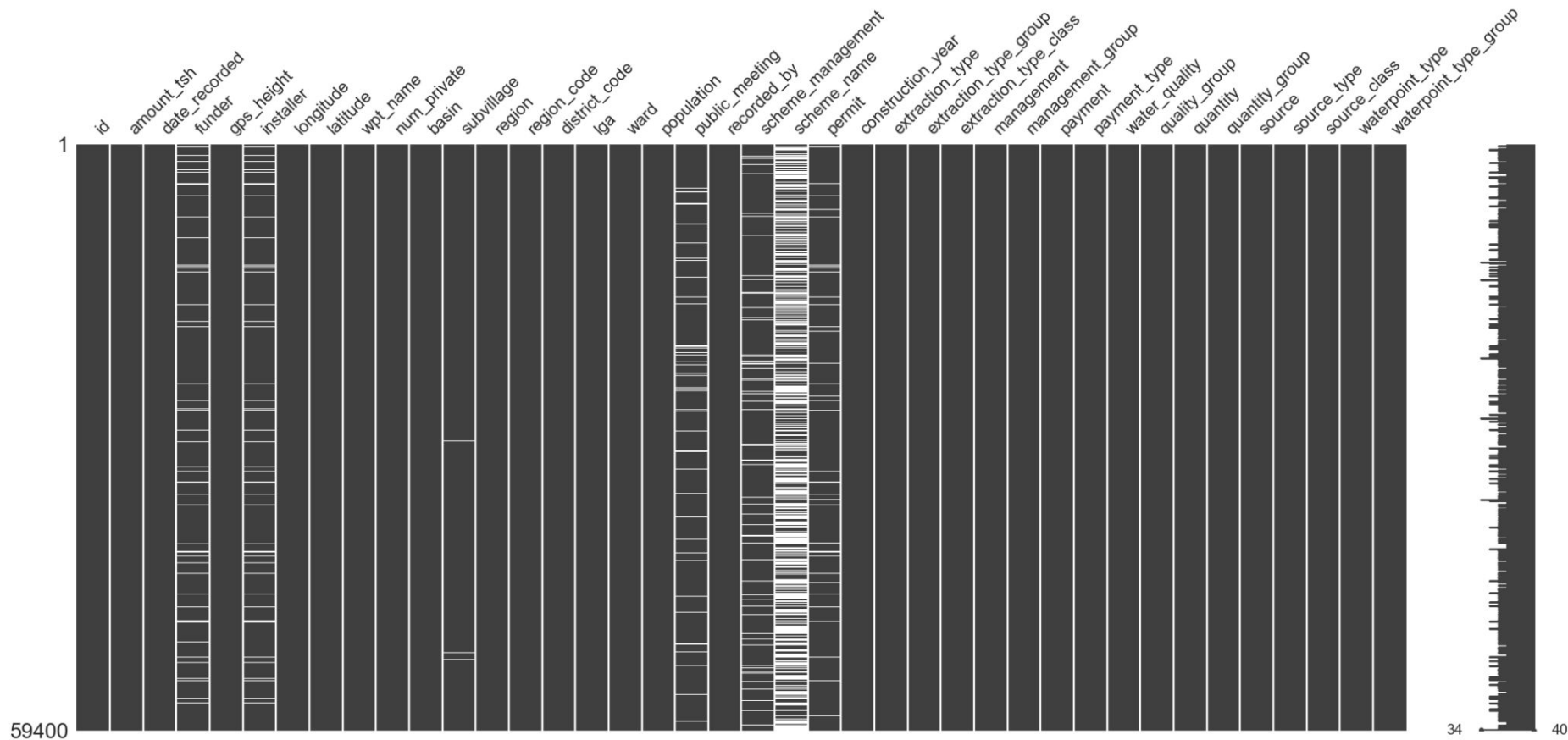
### Welcome Note

You are welcome to the Ministry of Water's website. Water is a basic need and right for all human beings. Please, value and conserve it for the ecosystem. Remember that development comes only when there is plenty of quality water, which is sustainably uti

[...Read More](#)

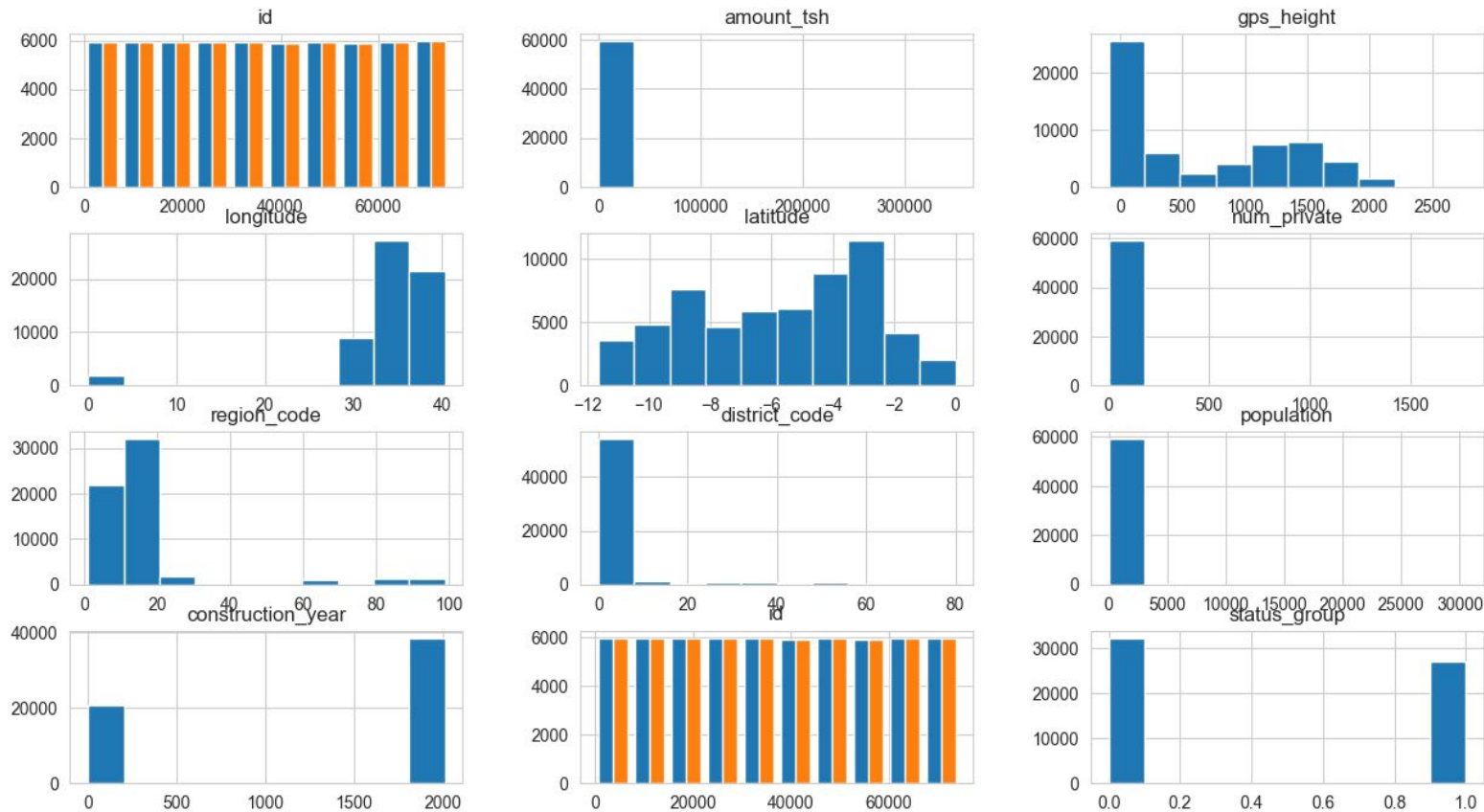
[Client Service Charter](#)

# Missing Values

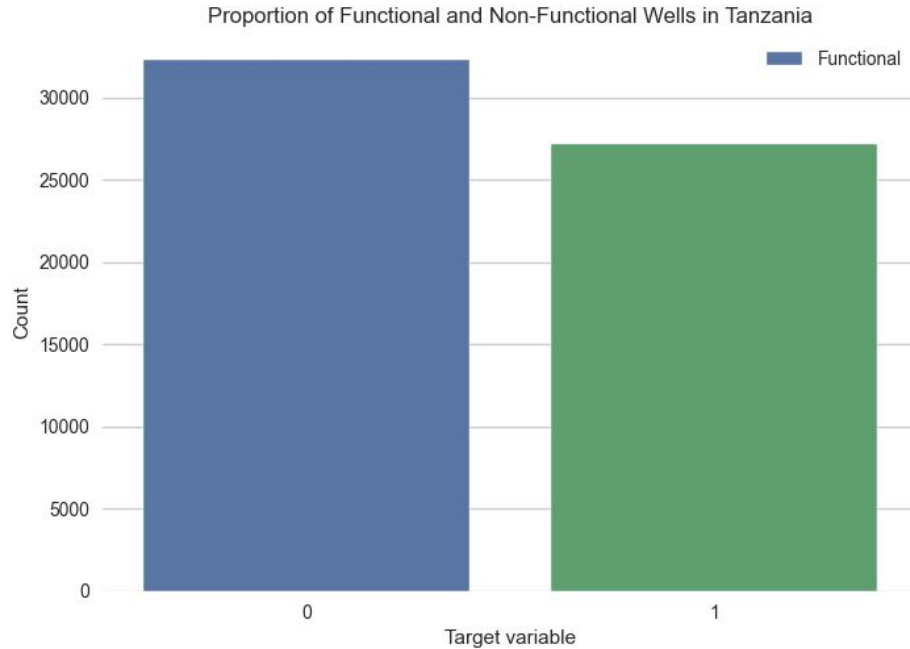




# Numerical Data



# Target



0 - Functional

Imbalance

1 - Non-Functional

0 = 32,259

1 - Needs Repair

1 = 27,141



# Metrics

Accuracy

Measures the proportion of correct predictions out of all predictions.

Precision

Measures the proportion of true positives out of all positive predictions.

Recall

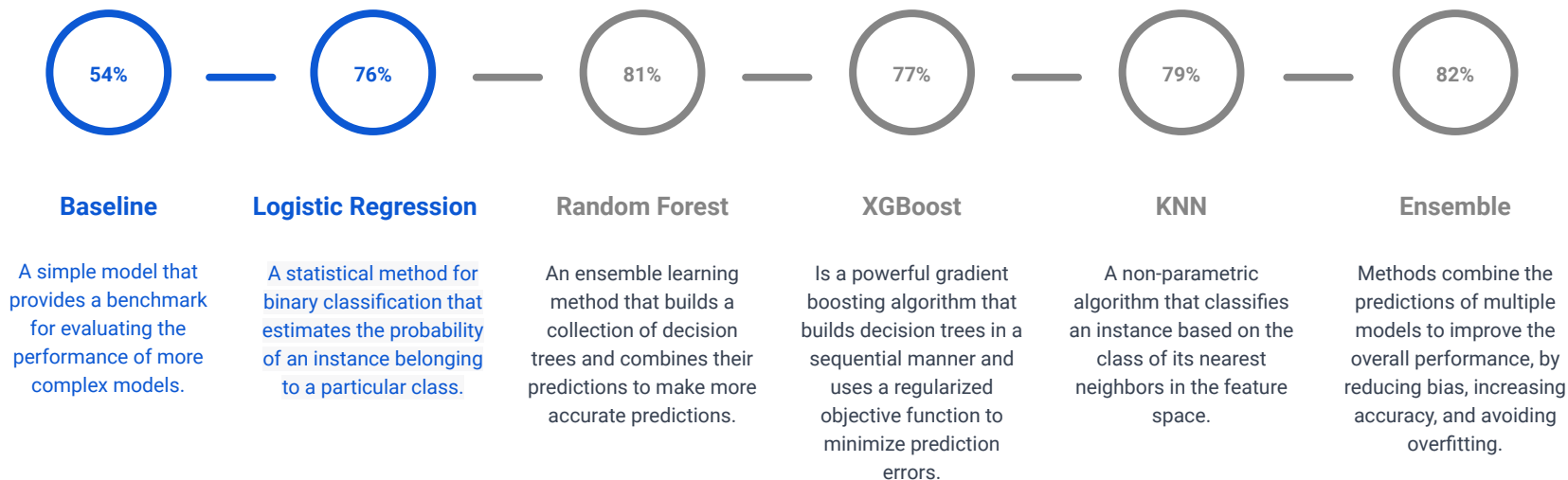
Measures the proportion of true positives out of all actual positives.

F1 Score

Is a weighted harmonic mean of precision and recall.

I will prioritize **recall** over precision on the basis that the cost of a false negative(predicting a waterpoint as functional when it's actually non-functional) is much higher than the cost of a false positive (predicting a waterpoint as non-functional when it's actually functional).

# Modeling





# Scores

## Logistic Regression

	0	1	accuracy	macro avg	weighted avg
precision	0.750498	0.795569	0.767311	0.773033	0.770924
recall	0.860525	0.654852	0.767311	0.757688	0.767311
f1-score	0.801754	0.718384	0.767311	0.760069	0.763970
support	7012.000000	5812.000000	0.767311	12824.000000	12824.000000

## Random Forest

	0	1	accuracy	macro avg	weighted avg
precision	0.821070	0.812714	0.817452	0.816892	0.817282
recall	0.851641	0.776230	0.817452	0.813935	0.817452
f1-score	0.836076	0.794053	0.817452	0.815064	0.817024
support	7010.000000	5814.000000	0.817452	12824.000000	12824.000000

## XGBoost

	0	1	accuracy	macro avg	weighted avg
precision	0.740032	0.858177	0.778228	0.799104	0.793595
recall	0.916120	0.611971	0.778228	0.764045	0.778228
f1-score	0.818715	0.714458	0.778228	0.766586	0.771448
support	7010.000000	5814.000000	0.778228	12824.000000	12824.000000

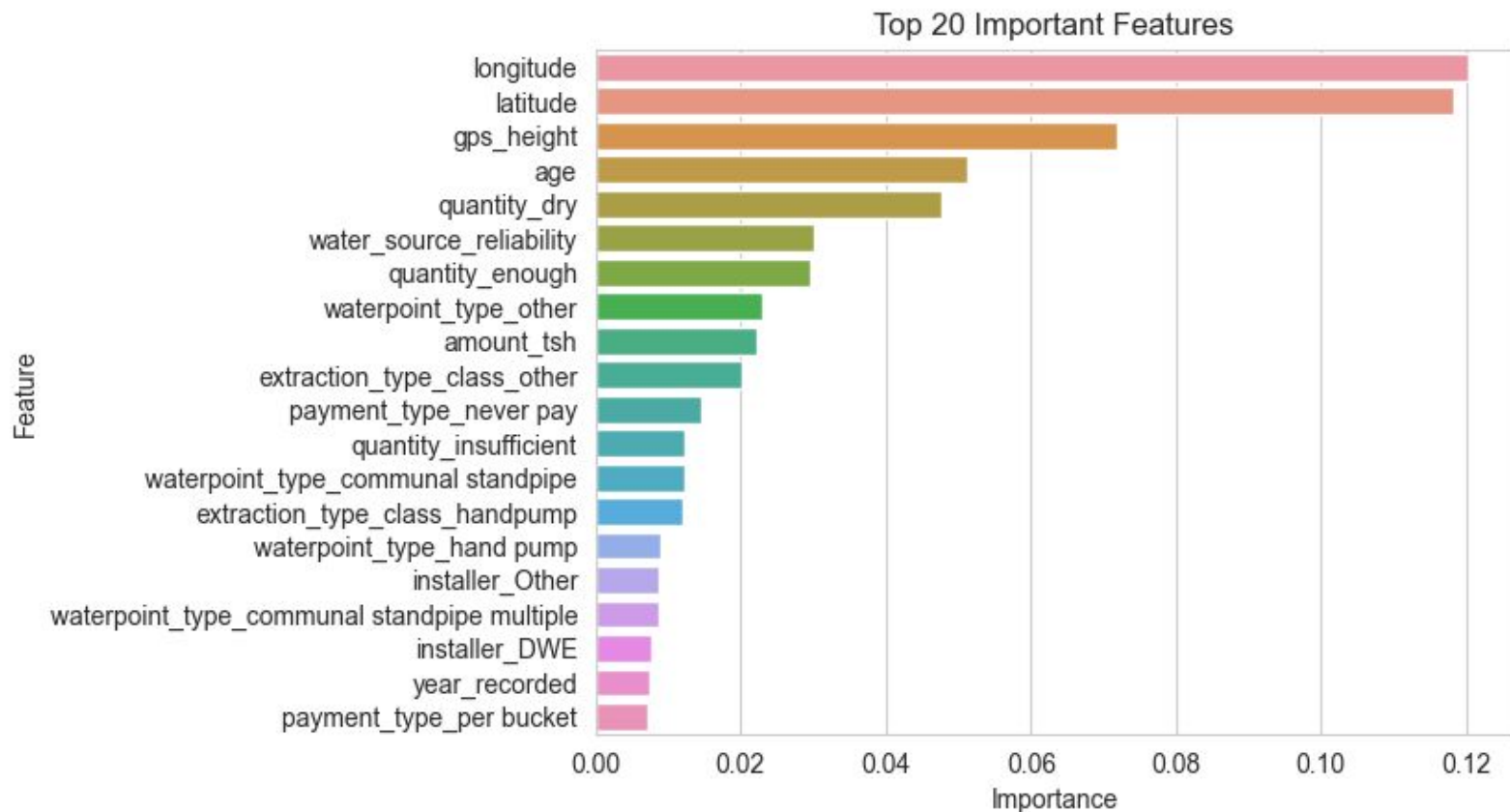
## KNN

	0	1	accuracy	macro avg	weighted avg
precision	0.802123	0.796202	0.799595	0.799162	0.799438
recall	0.840799	0.749914	0.799595	0.795356	0.799595
f1-score	0.821006	0.772365	0.799595	0.796685	0.798954
support	7010.000000	5814.000000	0.799595	12824.000000	12824.000000

## Ensemble

	0	1	accuracy	macro avg	weighted avg
precision	0.754723	0.897580	0.800296	0.826152	0.819490
recall	0.940228	0.631579	0.800296	0.785904	0.800296
f1-score	0.837325	0.741444	0.800296	0.789384	0.793855
support	7010.000000	5814.000000	0.800296	12824.000000	12824.000000

# Top Predictors



# Conclusion

## Final Model

RF Scores:

Accuracy: 0.819557080474111

Precision: 0.8194553198059485

Recall: 0.819557080474111

F1 Score: 0.8190698558570841

ROC AUC Score: 0.8158169542061638

## Improvements:

- Collect data on maintenance history
- Conduct surveys on the condition of water points
-

# Thank You

Jennifer Casias

E-mail: [casiasjc@gmail.com](mailto:casiasjc@gmail.com)



# Scores on Test

## LR Scores:

Accuracy: 76%

Precision: 77%

Recall: 76%

F1 Score: 76%

ROC AUC Score: 75%

## RF Scores:

Accuracy: 82%

Precision: 82%

Recall: 82%

F1 Score: 82%

ROC AUC Score: 81%

## KNN Scores:

Accuracy: 80%

Precision: 80%

Recall: 80%

F1 Score: 80%

ROC AUC Score: 80%

## XGB Scores:

Accuracy: 78%

Precision: 79%

Recall: 78%

F1 Score: 77%

ROC AUC Score: 76%

## ENSEMBLE Scores:

Accuracy: 82%

Precision: 82%

Recall: 82%

F1 Score: 82%

ROC AUC Score: 82%