Tanzanian Water Well Classification

Jeremy Pagirsky
April 2nd, 2021


Bunia
Uganda
Kisangani
Kampala
Eldoret
Kenya
Butembo
Entebbe
Kisumu
Nakuru
Garissa
Lake Victoria
Nairobi
Rwanda
Democratic Republic of the Congo
Kindu
Mwanza
Burundi
Arusha
Malindi
Mombasa
Kasango
Kigoma
Tanga
Kongolo
Kananga
Mbuji-Mayi
Kabinda
Tanzania
Dar es Salaam
Gandajika
Morogoro
Mwene-Ditu
Manono
Iringa
Sumbawanga
Kamina
Upemba National Park
Mbeya
Kasama
Lindi
Songea
Masasi
Mtwara
Likasi
Mzuzu
Lake Malawi
Mpika
Comoros
Mutsamudu
Fomboni
Mayotte
Solwezi
Kitwe
Ndola
Pemba
Kasungu
Zambia
Chipata
Malawi
Namapa
Kabwe
Lilongwe
Nacala
Kafue National Park
Ribaue
Nampula
Lusaka
Zomba
Gurue
Mogadishu
Baraawe
Antsohihy
Imagery ©2021 TerraMetrics, Map data ©2021
United States
Terms
Privacy
Send feedback
100

Background and Focus

- Tanzania abundant in natural resources.
- German and British colonization until 1961 independence.
  - Maintained presence.
  - Political destabilization, hunger, poverty, resource scarcity.
- **Decreased ability to access improved water sources.**

Objective

- Determine factors that predict water well functionality.
  - Functional
  - Functional Needs Repair
  - Non-Functional
- Examine the extent to which funding predicts functionality.
  - Tanzanian vs Non-Tanzanian

Data

- Data from Taarifa and Tanzanian Ministry of Water on over 59,000 water wells.
  - Provided by Flatiron School and DrivenData.
- Data as recent as 2013.

Methods

- Selected 20 most prevalent funders.
- Classified as **government** or **non-government**.

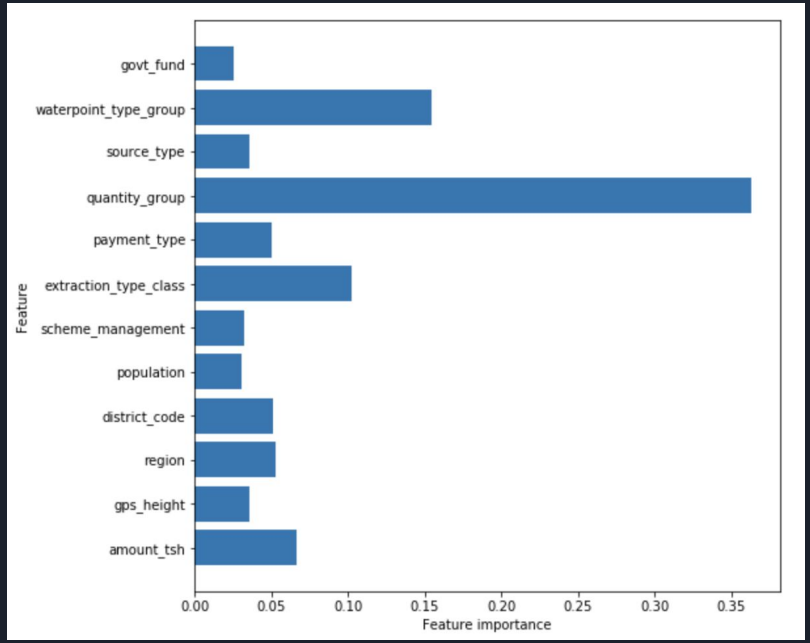
Government Of Tanzania	15.29
unknown	6.12
Danida	5.24
Hesawa	3.71
Rwssp	2.31
World Bank	2.27
Kkkt	2.17
World Vision	2.10
Unicef	1.78
Tasaf	1.48
District Council	1.42
Dhv	1.40
Private Individual	1.39
Dwsp	1.37
0	1.31
Norad	1.29
Germany Republi	1.03
Tcrs	1.01
Ministry Of Water	0.99
Water	0.98

Methods

- Trained, tested, and compared four different model types for highest accuracy.
- Selected model based on accuracy.
- Tuned model for optimization.

Results

- **Government / Non-government:** 22.5% / 77.5%
- Selected Random Forest Classifier with **73.4%** accuracy.
- **Main predictors:**
  - Quantity in well
  - The type of well
  - Extraction method
- Type of funding was **not** a significant predictor in well functionality.



Random Forest Classifier

Implications

- This model is able to predict functionality based on important features.
- Funding source is not important to this model.
- Model type is teachable and can continue to learn.

Limitations

- First iteration; model is not yet optimally tuned.
  - Classified unknown funders into categories.
- Colonization is multi-dimensional.
  - Funding source is one aspect.
- Approach has not thoroughly explored geographic factors.
  - Urban-rural disparities.

Future Work

- Technical aspects of model tuning and different approaches to the data.
- Examine other aspects of infrastructure as result of colonization.
- Utilize more robust geographical data methodology for urban-rural disparities.

References

- Tanzania High Commission: https://ke.tzembassy.go.tz/tanzania/natural-resources-and-mining-in-tanzania#:~:text=The%20main%20natural%20resources%20in,%2C%20fishing%20and%20minerals'%20mining
- TASAF: https://www.tasaf.go.tz/index.php/about-us/organization/tasaf-i
- Water.org: https://water.org/our-impact/where-we-work/tanza nia/
- Fall (2016): https://researchrepository.wvu.edu/etd/5569/
- National Bureau of Statistics: https://web.archive.org/web/20161123044258/http://www.nbs.go.tz/nbs/Stastical%20Abstract/Statistical%20Abstract%20Report%202013.pdf

Thank you!

E-mail: jeremy.pagirsky@gmail.com

GitHub: https://github.com/jeremypagirsky/Tanzanian_Water_Wells