Short 1-2 Page document

1. What is the problem you want to solve?

My goal is to improve the maintenance operations in Tanzania to ensure that clean, and portable water is available to the Tanzanian communities. I would analyze past records of water points that were functional, non-functional, and functional but need repair to construct a predictive model. The model will be used to predict current operating conditions of waterpoints in Tanzania with a recent data, to be used as a test. My goal is to maximize the classification rate, so that the Tanzanian Ministry of Water can rly on the model in making quick and necessary actions to provide water to their country.

1. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

My client is the Tanzanian Ministry of Water and Taarifa. They care about this problem, because they want to provide water throughout the Tanzanian community. Based on the analysis and the recommendation proposed, they would have a better idea of why the non-functional water points are not working and which waterpoints needs repairs. This can speed up the process of errors with the water point and satisfy the community.

1. What data are you going to do for this? How will you acquire this data?

The data I would use for this problem is provided by Taarifa, with the help of the Tanzanian Ministry of Water. The data consists over 40+ features to help understand the bigger picture of the problem to create a predictive model. The model will then be used in a test dataset provided to measure the effectiveness of the model.

1. In brief, outline your approach to solving this problem (knowing that this might change later).

I intend to approach this problem by first organizing the data. I will need to merge two datasets, the labels and the values, to figure out which water point is functional, non- functional or functional needs repair. I would the visualize various variables such as quantity, quality group, water point type, etc. to have a better picture of the cause of non- functional water points. Furthermore, I can identify any missing plots that may cause the graphs to skew. I will also research various R packages, for example googleVis, and identify ways to visually represent the data as well. I will explore various techniques such as Random Forest and creating new features to create the best predictive model.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

My deliverables would be to provide the code to execute the predictive model I have found and the presentation slide deck of my approach to solve the problem.

Appendix

Provided data features

* amount\_tsh - Total static head (amount water available to waterpoint)
* date\_recorded - The date the row was entered
* funder - Who funded the well
* gps\_height - Altitude of the well
* installer - Organization that installed the well
* longitude - GPS coordinate
* latitude - GPS coordinate
* wpt\_name - Name of the waterpoint if there is one
* num\_private -
* basin - Geographic water basin
* subvillage - Geographic location
* region - Geographic location
* region\_code - Geographic location (coded)
* district\_code - Geographic location (coded)
* lga - Geographic location
* ward - Geographic location
* population - Population around the well
* public\_meeting - True/False
* recorded\_by - Group entering this row of data
* scheme\_management - Who operates the waterpoint
* scheme\_name - Who operates the waterpoint
* permit - If the waterpoint is permitted
* construction\_year - Year the waterpoint was constructed
* extraction\_type - The kind of extraction the waterpoint uses
* extraction\_type\_group - The kind of extraction the waterpoint uses
* extraction\_type\_class - The kind of extraction the waterpoint uses
* management - How the waterpoint is managed
* management\_group - How the waterpoint is managed
* payment - What the water costs
* payment\_type - What the water costs
* water\_quality - The quality of the water
* quality\_group - The quality of the water
* quantity - The quantity of water
* quantity\_group - The quantity of water
* source - The source of the water
* source\_type - The source of the water
* source\_class - The source of the water
* waterpoint\_type - The kind of waterpoint
* waterpoint\_type\_group - The kind of waterpoint