Data

Files and Description

Submission format: The format for submitting your predictions

Test set values: The independent variables that need predictions

Training set labels: The dependent variable (status\_group) for each of the rows in

Training set values: The independent variables for the training set

The features in this dataset

Your goal is to predict the operating condition of a waterpoint for each record in the dataset. You are provided the following set of information about the waterpoints:

* 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

## The labels in this dataset

The labels in this dataset are simple. There are three possible values:

* functional - the waterpoint is operational and there are no repairs needed
* functional needs repair - the waterpoint is operational, but needs repairs
* non functional - the waterpoint is not operational

## Submission format

The format for the submission file is simply the row id and the predicted label (for an example, see SubmissionFormat.csv on the data download page).