IBM **Research** | Africa

The world is our lab.

MACHINE LEARNING APPROACHES TO SAFEGUARDING CONTINUOUS WATER SUPPLY IN ARID & SEMI-ARID (ASAL) LANDS OF NORTHERN KENYA

Fred Otieno

Software Engineer

Impact Science – Future of Climate

MACHINE LEARNING APPROACHES TO SAFEGUARDING CONTINUOUS WATER SUPPLY IN ARID & SEMI-ARID (ASAL) LANDS OF NORTHERN KENYA

—

Climate Shocks

UK weather

Heatwave: Paris suffers 42.6C hottest day ever as UK temperatures set record - as it happened

Updated 25 Jul 2019

Belgium,
day ever,

Likelihood of Cape Town water crisis tripled by climate

studied the effect of climate change on the
and found that it made

A climate in crisis

How climate change is making drought and humanitarian disaster worse in East Africa

1.2–2.2°C increase in temperatures by 2050

Increase in severity of dry spells and duration of heat waves

People cool off and sunbathe by the Trocadero Fount 2019 as a new heatwave hits the French capital.
Bertrand Guay/AFP/Getty Images

Climate Change and California's Water

- **Managing water is at the forefront of climate change adaptation in California.**

Source: Oxfam – Briefing Ref: 01/2017, USAID – Climate Risk Profile Kenya 2018, Guardian, ABC News, World Weather Attribution – Likelihood of Cape Town water crisis tripled by climate change, PPIC – Climate change and California's water

Problem

Inconsistent Water Access

Persistent drought emergencies and water inaccessibility have plagued communities in Sub-Saharan Africa, impacting their health, livelihoods, and education. Presently, the number of people in SSA without access to an improved water source is increasing, while the other geos are decreasing.

Source: OCHA, ourworldindata.org/water-access

Case Study: Northern Kenya

Kenya RAPID (Kenya Resilient Arid Lands Partnership for Integrated Development)

ASAL Counties

1. **84%** of total land mass of Kenya
2. Home to:
   - **36%** of the human population,
   - **70%** of the livestock herd
   - **65%** of the wildlife
3. **Natural disasters** have become **more frequent**; with droughts or floods every 3-4 yrs.
4. 2 to 4 million people require food relief each year

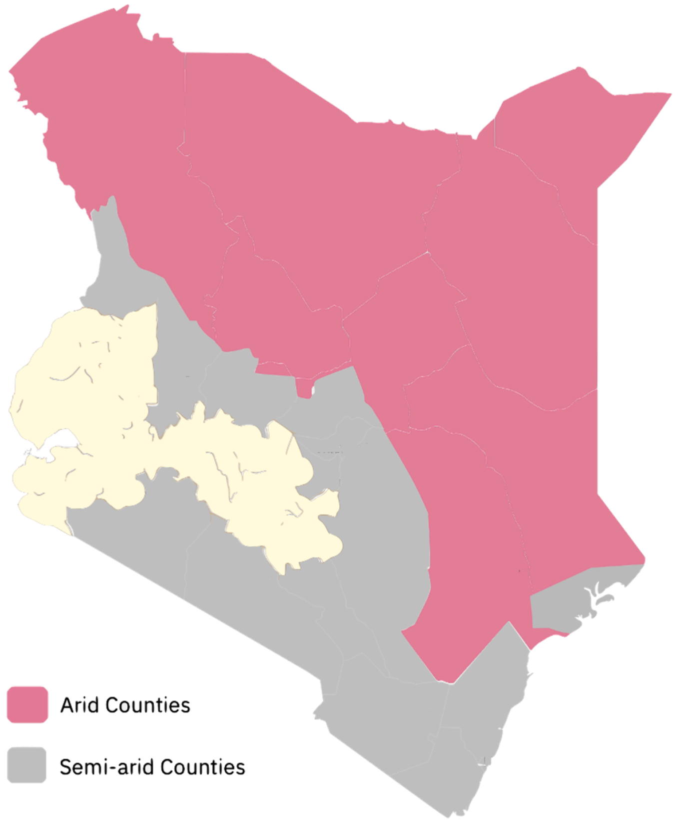
Kenya RAPID

Consortium of private and public partners including county governments, SweetSense among others and led by Millennium Water Alliance (MWA), our role is technology partner

Goal

Increase the average water access coverage in the five counties from 37% to more than 50%.

Focus counties: Turkana, Marsabit, Isiolo, Wajir & Garissa



Source: IUCN-Kenya's Drylands, IDMP- Drought Conditions and Management Strategies in Kenya, MWA - Kenya RAPID Program Background, UNDP – Kenya National Disaster Profile

Key Intervention Points

Utilizing groundwater

Monitoring and management of water supply system

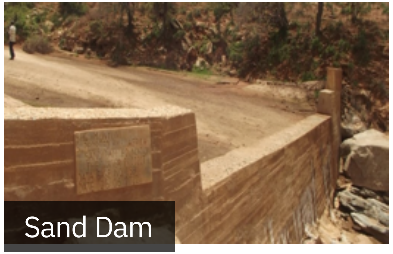
Decision support for water supply systems management

Predict waterpoint failure, detect recurrent behaviours or unknown trends; feed early warning

Types of Waterpoints


Water Pan


Shallow Well


Sand Dam


Well


Spring

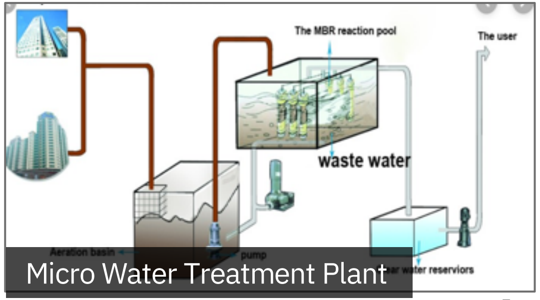

Motorized Borehole


Water Tank


River

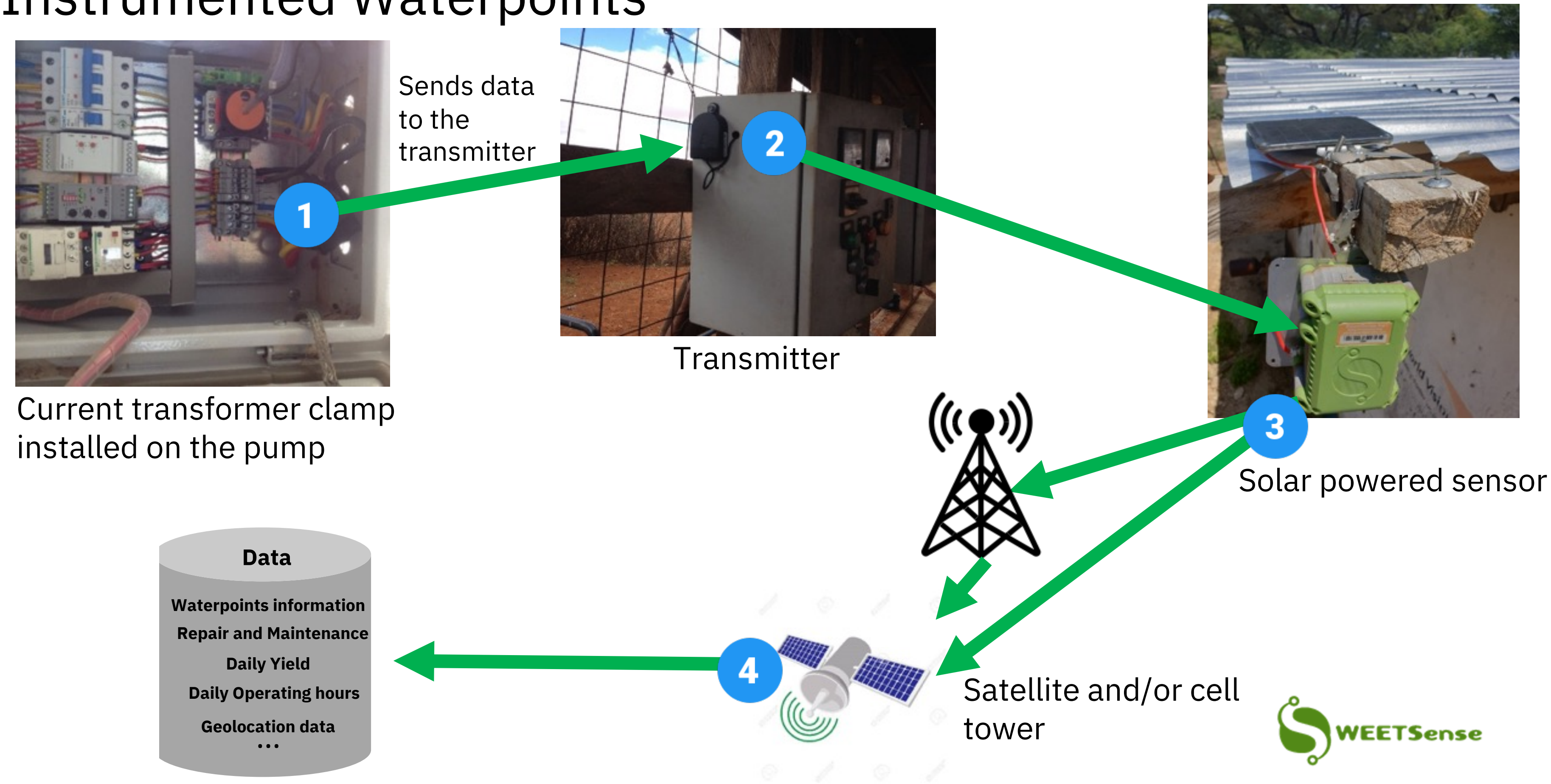
Water Truck


Handpump Borehole



Micro Water Treatment Plant

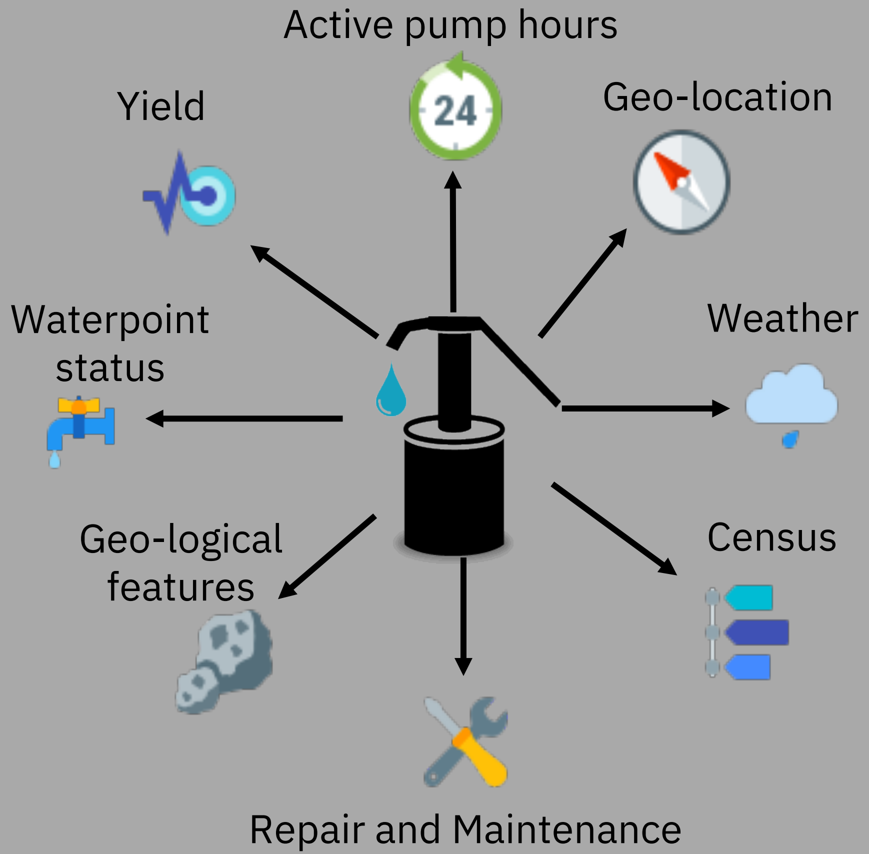
Instrumented Waterpoints



Data Collection

- 5 Counties in Northern Kenya
- 122 instrumented waterpoints
  - Up from 59 on Jan 2018
- 400 inventoried waterpoints

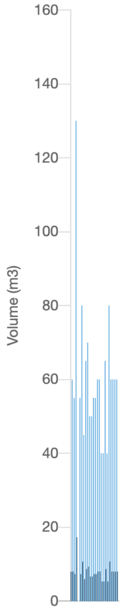
Water Point Data



eMaji Manager

Marsa Ticke

General



Assi

Ticket I

155948-MOFQ1

453020-I

993385-\

559657-\

Sub-county Water Officer

Number of Open Tickets

9

Non-functional Waterpoints

0 Strategic ⋮ **12** Non-strategic

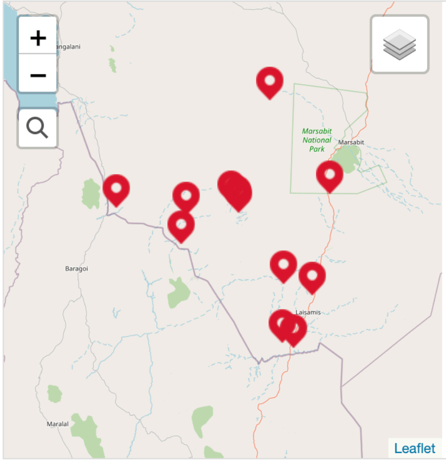
Livestock Affected

7,845

People Affected

1,365

Non-Functional Waterpoints



Name	Classification	Reported Issue	Distance away	Livestock served	People served	View
Marsabit - Kamboe	Non-strategic					
Marsabit- Kubi Qallo	Non-strategic		21 kms	45	65	
sakaldala borehole	Non-strategic			0		
kamatonyi	Non-strategic					
merille borehole 1	Non-strategic			0		

Items per page: 5 1 - 5 of 12

Pattern Mining and Failure Prediction For Waterpoints

Goal

To predict events such as waterpoint failure and high yield using waterpoint data and repair logs to trigger targeted interventions that facilitate continuous access to water

Related Works

i. Early failure detection for predictive maintenance of sensor parts(Kuzin 2016)

ii. Predictive model of failure based on event sequences observed at wire bonding process(Lim 2017)

iii. LSTM to find sessions that are prone to code failure in apps that rely on telemetry data for system health monitoring (Hajiaghayi 2019)

Our Proposed Methodology

i. Subsequence discriminatory models to mine for recurrent behaviors or unknown trends

ii. Waterpoint failure prediction using LSTM

Tomás Kuzin and Tomás Borovicka. Early failure detection for predictive maintenance of sensor parts. In ITAT, pp. 123–130, 2016.
Hwa Kyung Lim, Yongdai Kim, and Min-Kyoon Kim. Failure prediction using sequential pattern mining in the wire bonding process. IEEE Transactions on Semiconductor Manufacturing, 30(3): 285–292, 2017

Hajiaghayi, M. and Vahedi, E., 2019, April. Code Failure Prediction and Pattern Extraction using LSTM Networks. In 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService) (pp. 55-62). IEEE.

Discriminatory Subsequence Mining(Waterpoint Pattern Discovery)

1. Extended PrefixSpan to mine for patterns that are different between the two classes of interest. There are two parameters that need to be set, σ_{min} and *max_len*.

2. We report each pattern found is tagged with various metrics.
   - Coverage left and right, the coverage $\sigma_T = \frac{1}{n}|K_T(I)|$ of a pattern T, for left and right datasets.
   - Lift(T, left, right), this is the ratio of $\frac{\sigma_T^{left}}{\sigma_T^{right}}$ defining how much more often does a pattern occur in the left dataset compared to the right.

Algorithm 1: A pseudo-code for the proposed discriminatory sequence mining.

input : Collections $S_n = [S_1, S_2]$, σ_{min}, max_len, I
output: S_{out}, where $(|S_{out}| << |S_1|, |S_2|)$ such that $\forall\ s \in S$, $d_s > threshold$

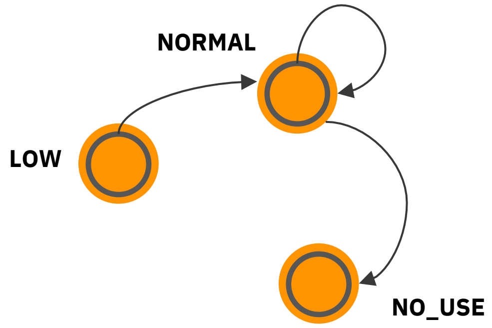
for $c \to 1$ **to** $|S_n|$ **do**
    $freq_{S_c} = \Phi_{S_c}(\sigma_{min})$;
    $\sigma_c = \frac{1}{n}|K_{freq_{S_c}}(I)|$;
end
for $c \to 1$ **to** $|freq_n|$ **do**
    if $S_1.isPrefix(S_2)$ *and* $Lift(freq_{S_c}, S_1, S_2) > Lift(freq_{S_c}, S_2, S_1)$ **then**
        S_{out} = pruneByDominance($freq_{S_c}, S_1$);
    end
    if $\frac{match(S_1,S_2)}{match(S_1)} > threshold$ **then**
        S_{out} = pruneByShadows(S_1);
    end
end

Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U. and Hsu, M., 2001, April. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In proceedings of the 17th international conference on data engineering (pp. 215-224). IEEE Washington, DC, USA.

Discriminatory Subsequence Mining (Cont.)

What the domain expert can do with this information?

Given the following sequence:



One month of low use followed by three months of normal use and a final month of no use

With

$coverage_{left}$ = 0.4321; 43% of sensors with this sequence end in failure

$coverage_{right}$ = 0.0532; low frequency on the rest of the population; hence the pattern is discriminatory regarding failure class in sensors

lift = 4.095; pattern is 4 times more likely to appear on sensors with failure events *vis*-à-*vis* functional ones

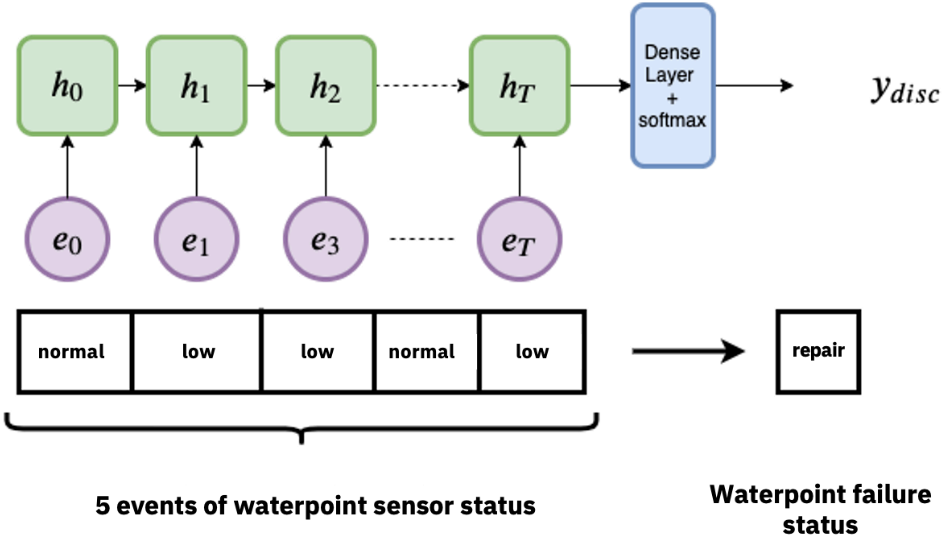
This insight would be valuable for developing an automated repair regime

Failure Prediction with LSTM Networks

1. Map the input sequence of events to a fixed-sized vector representation using an RNN.

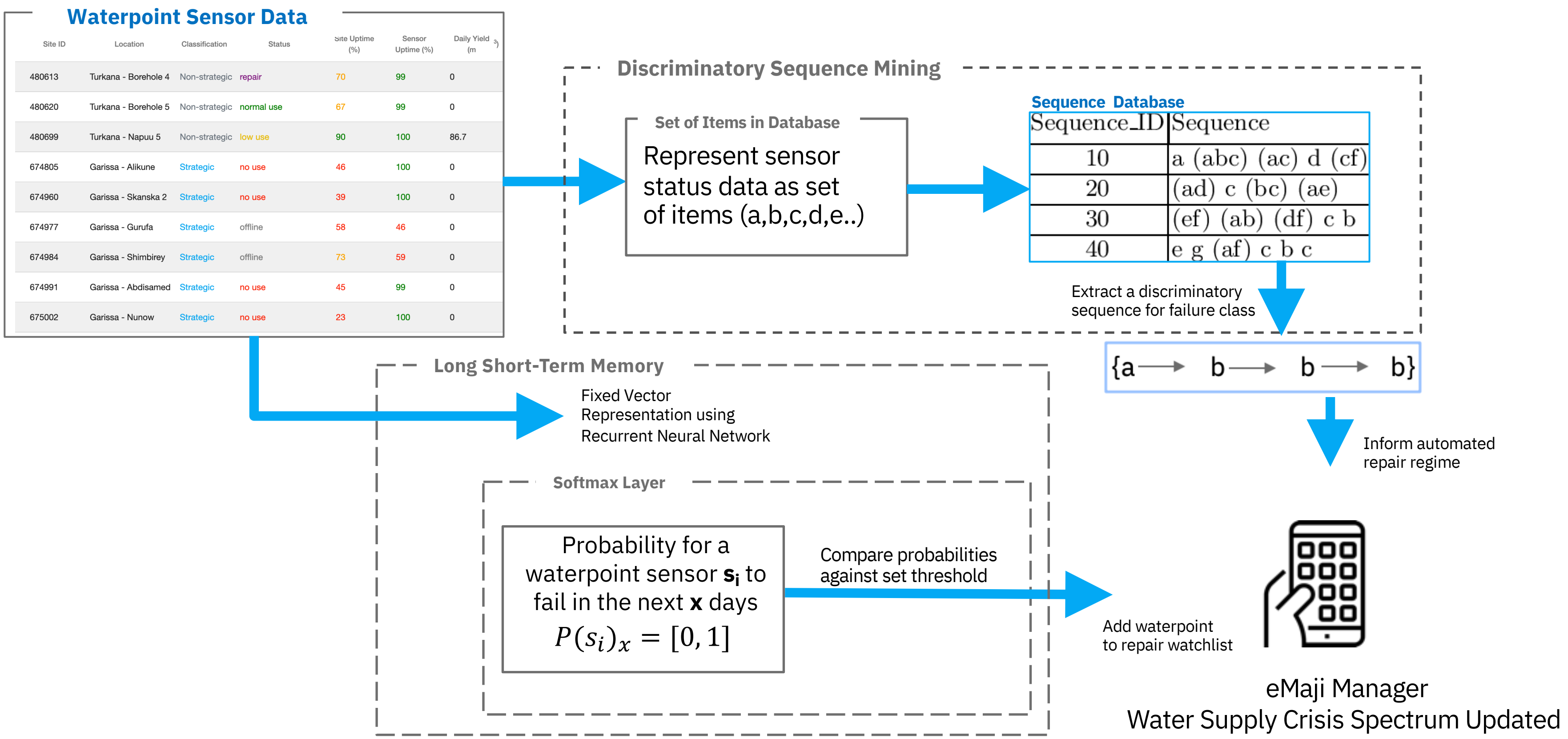
2. Feed h_T vector to a softmax layer for classification on the type of failure status.

3. The softmax non-linear layer predicts the probability of a particular type of failure for that time window of events.



Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.
Hajiaghayi, M. and Vahedi, E., 2019, April. Code Failure Prediction and Pattern Extraction using LSTM Networks. In 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService) (pp. 55-62). IEEE.

Proposed Water Supply Crisis Spectrum



Waterpoint Sensor Data

Site ID	Location	Classification	Status	Site Uptime (%)	Sensor Uptime (%)	Daily Yield (m³)
480613	Turkana - Borehole 4	Non-strategic	repair	70	99	0
480620	Turkana - Borehole 5	Non-strategic	normal use	67	99	0
480699	Turkana - Napuu 5	Non-strategic	low use	90	100	86.7
674805	Garissa - Alikune	Strategic	no use	46	100	0
674960	Garissa - Skanska 2	Strategic	no use	39	100	0
674977	Garissa - Gurufa	Strategic	offline	58	46	0
674984	Garissa - Shimbirey	Strategic	offline	73	59	0
674991	Garissa - Abdisamed	Strategic	no use	45	99	0
675002	Garissa - Nunow	Strategic	no use	23	100	0

Sequence Database

Sequence_ID	Sequence
10	a (abc) (ac) d (cf)
20	(ad) c (bc) (ae)
30	(ef) (ab) (df) c b
40	e g (af) c b c

Recap

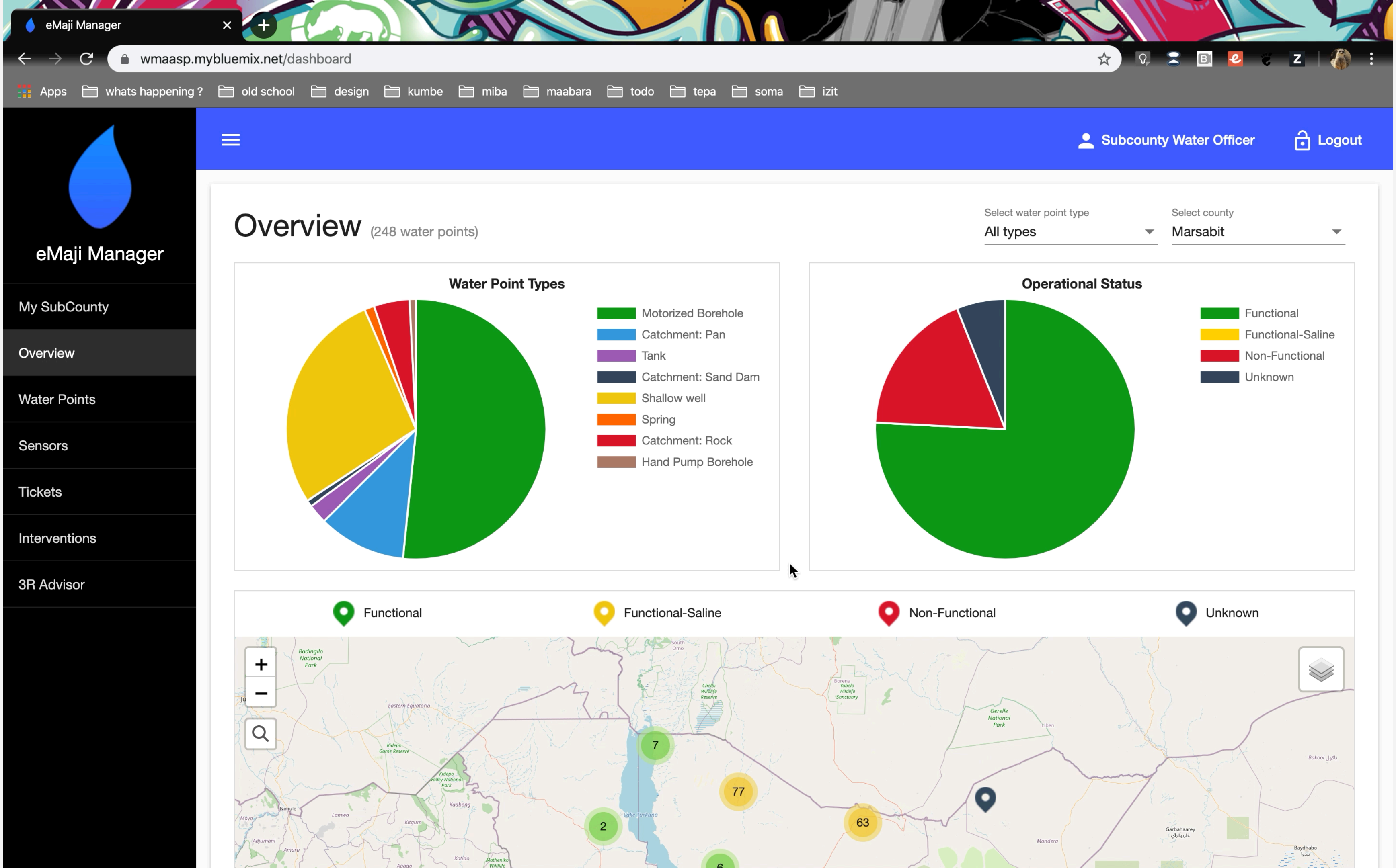
- Climate change impacting livelihoods in Northern Kenya due to increased uncertainty of drought seasons.
- Necessary adaptation to meet acute water shortages
- eMaji Manager an integrated waterpoint inventory, issues tracking, and decision support system for county officials.
- Propose to enhance the quality of data in eMaji Manager by exploiting the waterpoint and sensor data to extract temporal patterns and predict non-functionality to shorten repair response time and provide an emergency alert system.

Next Steps

- Evaluate proposed methodology
- Use alternative data to handle issue of differentiating waterpoint failure between sensor failure and borehole failure

Thank you

Fred Otieno
Software Engineer
—
fred.otieno@ibm.com
ibm.com


eMaji Manager
wmaasp.mybluemix.net/dashboard
Apps
whats happening ?
old school
design
kumbe
miba
maabara
todo
tepa
soma
izit
eMaji Manager
My SubCounty
Overview
Water Points
Sensors
Tickets
Interventions
3R Advisor
Subcounty Water Officer
Logout
Overview (248 water points)
Select water point type
All types
Select county
Marsabit
Water Point Types
Motorized Borehole
Catchment: Pan
Tank
Catchment: Sand Dam
Shallow well
Spring
Catchment: Rock
Hand Pump Borehole
Operational Status
Functional
Functional-Saline
Non-Functional
Unknown
Functional
Functional-Saline
Non-Functional
Unknown
7
77
63
2