WaterNet A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes

ABSTRACT

Water is a fundamental requirement for human, animal, and plant survival. Despite its importance, quality water is not always \_t for drinking, domestic and/or industrial use. Numerous factors such as industrialization, mining, pollution, and natural occurrences impact the quality of water, as they introduce or alter various parameters present therein, thus, affecting its suitability for human consumption

or general use. The World Health Organization has guidelines which stipulate the threshold levels of various parameters present in water samples intended for consumption or irrigation. The Water Quality Index (WQI) and Irrigation WQI (IWQI) are metrics used to express the level of these parameters to determine the overall water quality. Collecting water samples from different sources, measuring the various parameters present, and bench-marking these measurements against pre-set standards, while adhering to various guidelines during transportation and measurement can be extremely daunting. To this end this study proposes a network architecture to collect data on water parameters in real-time and use Machine Learning (ML) tools to automatically determine suitability of water samples for drinking and irrigation purposes. The developed monitoring network is based on LoRa and takes the land topology into consideration. Results of simulations done in Radio Mobile revealed a partial mesh network topology as the most adequate.

Due to the absence of large and open datasets on drinking and irrigation water, datasets usable for training ML models were developed. Three ML models - Random Forest (RF), Logistic Regression (LR) and Support Vector Machine (SVM) were considered for the water classification process and results obtained showed that LR performed best for drinking water, while SVM was better suited for irrigation water. Recursive feature elimination was then combined with the three ML models to reveal which of the water parameters had the greatest influence on the classification accuracies of the respective model.

**EXISTING SYSTEM**

In [12], a network for measuring and monitoring water parameters in a metal producing city in Brazil was developed. Twelve water monitoring stations were setup to measure several physico-chemical water parameters, including pH, dissolved solids, Zinc, Lead etc. Finally, obtained results were analysed using principal component analysis. In a similar manner, [13] developed a system to monitor water quality in Limpopo River Basin in Mozambique and set up 23 monitoring stations to measure physico-chemical and microbiological parameters, and ultimately assess the quality of water in the river basin. To address the challenges of optimal placement of gauges and sampling frequencies, which are often faced when developing water monitoring systems, the authors in [14] developed an economically viable model that combined genetic algorithm with 1-D water quality simulation. Though the work was only simulated by using genetic algorithm, the authors were able to solve the NP hard problem of optimally placing monitoring stations.

Monitoring water parameters often entails periodically sampling a body of water to capture relevant metrics. These metrics might include physico-chemical and microbiological measurements, such as potential of hydrogen (pH), temperature, sodium levels etc. In a water monitoring network, measured parameters need to be transferred to a base station where relevant decision(s) would be taken. Due to the sparse nature of transmitted data, light weight communication protocols capable of transmitting relatively small data over long distance are required for water monitoring networks. From literature, Low Power Wide Area Network (LPWAN) technologies have been favoured for such applications. An extensive discussion on LPWAN technologies was done in [19]. The work compared a few sub-GHz solutions including Sig- Fox, LoRa, Ingenu and Telensa, with respect to their range,

transmission rate, and channel count. Ingenu was reported to have the longest range in city settings at 15 km, followed by SigFox at 10 km (in cities) and 50 km (in rural areas); then LoRa at 5 km (in cities), and 15 km in rural settings.

Regarding the assessment of communication technologies, there has been a long-drawn debate over the efficacy of software simulations versus real-world testing. Though this debate still rages, several researchers have shown that simulation results are often at par with real-world tests. For instance, using LoRa, the authors in [20] compared simulation results with real world test for intervehicle communication. They used NS3 as a simulation platform and an Arduino UNO C Dragino LoRa module for the real-world tests, while Propagation loss, coverage Packet Inter-reception (PIR), Packet Delivery Ratio (PDR) and Received Signal Strength Indicator (RSSI) level were used as benchmark metrics. They concluded that the results of the simulator were consistent with those of the real-world tests. In a similarwork,

Hassan [21] also compared the efficacy of simulation results (from Radio Mobile simulator) with real-world tests (using micro controllers C LoRa modules) when using LoRa as a bridge for Wi-Fi. Unlike [20], [21] did not give a side-byside comparison of simulated vs. real-world results for each metric considered but concluded that the simulator performed well. [22] set up seven pairs of XBee modules and compared communication performance using both the 800/900MHz

and 2.4GHz frequencies. They concluded that simulation results from the Radio Mobile simulator corroborated with those of real-world tests.

**Disadvantages**

* An existing methodology doesn’t implement DATA PRE-PROCESSING & LABELLING method.
* The system not implemented Calculating WQI for Irrigation Water for prediction in the datasets.

Proposed System

The water monitoring network proposed in this work is to be deployed in the City of Cape Town in Western Cape, South Africa, with the intention of monitoring water parameters in water storage dams and/or water treatment plants across the city. Data gathered by the monitoring network are then passed through Machine Learning (ML) models to determine their suitability for consumption or irrigation purposes.

1) Build a network for real-time collection and monitoring of water quality across water storage dams in the city of Cape Town. This network takes into consideration the unique geographical features of Cape Town, such as mountains and elevations that might obstruct radio frequency propagation.

2) Curate ample sized datasets on drinking and irrigation water that can be used to train (and test) machine learning models to automatically determine the `fitness for use'' of a sample of water for drinking and/or irrigation purposes.

3) Build models that determine the most critical parameters that influence the accuracy of machine learning models in analyzing water for drinking or irrigation.

**Advantages**

* The purpose of WaterNet is to gather data on water parameters from dams across the city. These parameters are then used to assess the quality of water with regards ``fitness for use'' for drinking and irrigation purposes.
* In this work, rather than relying on instrumental and physico-chemical analysis carried out in laboratories to assess water parameters, we propose the use of machine learning (ML) models, which take the numerous water parameters into consideration and automatically determine if a sample of water is potable or fit for agricultural use.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).