ROBERT GORDON UNIVERSITY ABERDEEN

SCHOOL OF COMPUTING SCIENCE AND DIGITAL MEDIA

MSc Big Data Analytics

Student Name: B. A. Malith Abeywardena	Matriculation Number: 2017671 IIT Index Number: 20200753
Supervisor: Dr.Alroy Mascrenghe	Second Marker:

| Project Title: **Using Machine Learning to develop a predictive model for pipe-born water production during drought season in Sri Lanka:** Focusing on Kalatuwawa Treatment Plant ||

Word Count - 18500	Start Date:17/01/2022
	Submission Date: 10/01/2023

CONSENT

I agree ☑
I do not agree ☐

That the University shall be entitled to use any results, materials or other outcomes arising from my project work for the purposes of non-commercial teaching and research, including collaboration.

DECLARATION

I confirm:

- **That the work contained in this document has been composed solely by myself and that I have not made use of any unauthorised assistance.**

- **That the work has not been accepted in any previous application for a degree.**

- **All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.**

Student Signature: **SIGNED**	Date Signed: 10/01/2023

RGU SYMBOL OF SUCCESS

ABSTRACT

Scarcity of water is a problem that is being commonly discussed in modern days, where water is being a basic human need out of which a single human being can exist only up to three days without water. Among many other uses water plays an important role in consumption. To live longer free of diseases it is important to consume clean water where many countries suffer in finding clean water resources because of lack of water resources due to seasonal effect, pollution of water and other factors. Every country allocates a lot of resources yearly to keep the water resources clean as well as to convert the raw water to consumable potable water where Sri Lanka spends 20,000 million per year in the purification process. National Water Supply and Drainage Board (NWSDB) is the national service provider responsible for providing potable water to the nation. Even though NWSDB provides potable water to the population, during the drought season usually a shortage of water causing water supply interruptions occur, so NWSDB is unable to produce and supply the required demand for water. Every year Sri Lanka usually faces drought causing deaths to living beings and failure in agriculture occurs due to a deficiency in the rainfall.

Since this is the first time this research is being carried out in the worldwide context as well as Sri Lankan context, the aim of this research is to develop a predictive model which predicts the outflow or the optimum capacity that can be purified within a purification plant by considering intake quantities, quality parameters of water (Biological Oxygen Demand (BOD), Biological Oxygen Demand (COD), Nitrogen, Ammonia), rainfall, humidity, temperature, visibility, and windspeed. Where the author follows data pre-processing techniques, feature selection techniques (Backward selection, Feature importance), Feature engineering techniques (Principal component analysis). Since this is a regression type scenario the author utilizes the commonly used regression models namely, Random Forest, Decision Tree (DT), Linear Regression, Lasso Regression, XGBoost Regression, Bayesian Ridge Regression, Support Vector Regression (SVR) and Ridge Regression. The author also applies hyper parameter tuning per each model. Each model is being evaluated by the model evaluators R^2, RMSE, MAE in order to select the most applicable parameter combination along with the best model.

Key Words: Regression, Machine Learning, Feature Selection, Feature Engineering, Hydrological predictive modelling.