

**BUSITEMA
UNIVERSITY**
Pursuing Excellence

**FACULTY OF ENGINEERING AND TECHNOLOGY DEPARTMENT
OF WATER RESOURCES ENGINEERING
ASSESSMENT OF THE APPLICATION OF MACHINE LEARNING IN
GROUNDWATER PROSPECTING FOR SITING OF BOREHOLES
CASE STUDY: AWOJA CATCHMENT**

BY

SSEKYANZI CHRISTOPHER

BU/UG/2021/0029

SUPERVISOR: Ms. Nabunya Victor

*This final year project report submitted to the department of Water Resources Engineering
as a partial fulfilment of the requirements for the award of Bachelors degree of Science in
Water Resources Engineering.*

MAY 2025

ABSTRACT

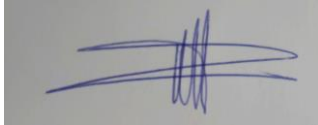
This report presents an evaluation of the role of machine learning in predictive studies focusing on groundwater. Machine learning algorithms such as Artificial Neural Networks (ANN) and Extreme Gradient Boost(XGB) to offer solutions in groundwater prospecting through data-driven modeling, requiring less labor, cost, and time. The main objective of this study was to evaluate the applicability of two machine learning algorithms, ANN and XGB in the prediction of static water level and total yield for new borehole sites within the Awoja catchment. The study utilized borehole data for the existing boreholes, including all the borehole information together with groundwater explanatory variables, soil, lithology, landuse and precipitation, and remote sensed data to train a multivariate ANN and XGB models for predicting Static Water Level and total yield. Both models were evaluated using the coefficient of correlation (R^2), MAE, and MSE and the models performed well with the XGB model better than the ANN for the prediction of both the static water level and total yield having scores of 0.901, and 0.019 for R^2 and MAE. It has been revealed that ML techniques cannot replace the existing methods of prospecting but are very important when used alongside each other. The results of this study aim to inform the practicability of data driven modeling in groundwater prospecting for the siting of boreholes as a way of improving the efficiency of the process.

Key words: Machine-Learning, Artificial Neural Networks, Extreme Gradient Boost, Ground-water

DECLARATION

I **SSEKYANZI CHRISTOPHER** declare that this research report was originally written and organized by me and has not been previously published, therefore presented as a requirement for the award of Bachelor of science in Water Resources Engineering.

Date ; 15/05/202025

A handwritten signature in blue ink, consisting of several horizontal strokes and a central vertical flourish.

Signature

Email: ssekyanzichristopher19@gmail.com

Contact: 0700118849/ 0761417053

ACKNOWLEDGEMENT

I take this opportunity to thank the almighty God for the good health and the ability to research and gather information in this report. I thank my supervisor Ms. Nabunya Victo and Mr. Maseruka Benedicto Ssajjabi, a mentor for the time, support and advice provided to me for the period I have carried out this research.

DEDICATION

I dedicate this report to Mr. & Mrs. Mukaaya Christopher, my beloved parents, as an appreciation for the love and care provided to me up to this point in my academics.

APPROVAL

I SSEKYANZI CHRISTOPHER, submit this report of my final year project report to the faculty of engineering in the department of Water resources engineering for the examinations with approval of my supervisor.

Ms. NABUNYA VICTO



Signature

Date: 14/05/2025

Table of Contents

ABSTRACT	i
DECLARATION	ii
ACKNOWLEDGEMENT	iii
APPROVAL	v
List of figures.....	ix
List of tables	x
List of Acronyms	xi
CHAPTER ONE: INTRODUCTION.....	12
1.1 BACKGROUND	12
1.2 PROBLEM STATEMENT.....	14
1.3 JUSTIFICATION	14
1.4 OBJECTIVES.....	14
1.4.1 Main objective	14
1.4.2 Specific Objectives	15
1.5 SIGNIFICANCE OF THE STUDY	15
1.6 RESEARCH QUESTIONS	15
1.7 SCOPE OF THE STUDY.....	16
1.7.1 Conceptual scope	16
1.7.2 Geographical scope.....	16
1.7.3 Time scope	16
CHAPTER TWO: LITERATURE REVIEW.....	17

2.1 GROUNDWATER AND GROUNDWATER AVAILABILITY	17
2.1.1 Key variables in groundwater availability	19
2.1.2 BOREHOLE SITING	21
2.2 GROUNDWATER MODELLING	24
2.2.1 Machine Learning models.....	25
2.2.2 Commonly used models.....	27
2.2.3 Machine Learning for Prediction studies	29
2.3 EVALUATION OF MACHINE LEARNING MODELS.....	31
2.3.1 Accuracy	31
2.3.2 Mean Absolute Error (MAE).....	32
2.3.3 Root Mean Squared Error (RMSE).....	32
2.3.4 R-squared (R^2).....	32
2.3.5 Precision and Recall.....	33
2.3.6 The F1 score.....	33
2.3.7 Nash–Sutcliffe Efficiency (NSE):.....	33
CHAPTER THREE: METHODOLOGY	34
3.1 DESCRIPTION OF THE STUDY AREA	34
3.2 DATA COLLECTION	35
3.2.1 DATA PREPARATION AND CLEANING.....	38
3.2.2 DATA ANALYSIS.....	41
3.2.3 DATA PROCESSING	42
3.3 MODEL DEVELOPMENT.....	47

3.3.1 Data splitting.....	48
3.3.2 Model Training	48
3.3.3 MODEL EVALUATION	50
3.4 SENSITIVITY ANALYSIS	52
CHAPTER FOUR: RESULTS AND DISCUSION	55
4.1 DESCRIPTIVE STATISTICS FOR INPUTS.....	55
4.2 Model Development	55
4.2.1 Artificial Neural Network (ANN).....	56
4.2.2 Extreme Gradient Boosting (XGB)	56
4.2.3 Analysis of Loss Curves (ANN).....	57
4.2.4 Feature Importance	58
4.2.5 Model Performance.....	59
4.3 Spatial Prediction Results.....	61
4.3.1 Geospatial function for inputs.....	61
CHAPTER FIVE: Conclusions and Recommendations.....	64
5.1 Conclusions	64
5.2 Recommendations.....	65
REFERENCES	66
APPENDIX A.....	77

List of figures

Fig. 2.1: Geological triangulation method	10
Fig. 2.2: Classification of Machine learning models (Aderemi et al., 2023)	14
Figure 3-1; Map of Awoja catchment	38
Figure 3-2; NDVI map for Uganda (2023) and GEE script for extraction	41
Figure 3-3; Correlation matrix for Numerical features	46
Figure 3-4; ANN Architecture	49
Figure 4-1; Code snippet for ANN development.....	56
Figure 4-2; Loss curve for ANN	57
Figure 4-3;XGB feature importance for Total yield prediction	58
Figure 4-4; XGB feature importance for SWL	59
Figure 4-5; Summary of model performance.....	60
Figure 4-6; Mesh points filled with landuse	61
Figure 4-7; Operation of the geospatial function	63

List of tables

Table 3-1: Necessary input variables	22
Table 3-1; Districts making up Awoja Catchment.....	35
Table 3-2; Input data	37
Table 3-3; Table showing the explanatory variables	39
Table 3-4; Imputation techniques used	42
Table 3-5; Preprocessing techniques used	46
Table 3-6; ANN Architecture summary.....	49
Table 3-7; Model development activity summary	50
Table 3-8;Evaluation metrics used.....	52

List of Acronyms

GWPZ	Groundwater potential Zone
ML	Machine Learning
SVM	Support Vector Machine
ANN	Artificial Neural Network
MWE	Ministry of Water and Environment
GBDT	Gradient Boosting Decision Tree
xGB	Extreme Gradient Boosting
RF	Random Forest
LSTM	Long Short-Term Memory
MLP	Multi-Layer Perceptron
VES	Vertical Electrical Sounding
NSE	Nash-Sutcliffe Efficiency
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Squared Error
LULC	Land Use Land Cover
WQI	Water Quality Index