



Developing an Enhanced Soft Sensor for Wastewater Treatment Plants: A Comparative Study of Multiple Machine Learning Approaches

CHARUKA DILSHAN KALUARACHCHI
(Reg. No.: MS24053396)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY
(ENTERPRISE APPLICATION DEVELOPMENT)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Mr Jeewaka Perera

Approved for MSc. Research Project:

MSc in IT Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign: 

Charuka Dilshan Kaluarachchi

Date:27/10/2025.....

ABSTRACT

Developing an Enhanced Soft Sensor for Wastewater Treatment Plants: A Comparative Study of Multiple Machine Learning Approaches

Charuka Kaluarachchi

MSc. in Information Technology

Enterprise Application Development

Supervisor: Mr Jeewaka Perera

October 2025

Wastewater treatment plants (WWTPs) require continuous monitoring of critical water quality parameters to ensure operational efficiency and regulatory compliance. Traditional physical sensors are accurate but expensive and maintenance-intensive, creating a need for cost-effective alternatives. This research investigates the development of enhanced soft sensors using advanced machine learning techniques to estimate key wastewater parameters including Chemical Oxygen Demand (COD) and Total Phosphorus (TP) concentrations at both influent and effluent points. The study addresses fundamental limitations of existing soft sensor implementations particularly their inability to capture complex non-linear relationships which is suspected to have sensor drift and degradation due to seasonal variations and equipment aging. Through comprehensive evaluation of multiple machine learning approaches including Neural Networks and Decision Tree-based methods with the aim to develop robust, adaptable soft sensor models that maintain accuracy over extended periods with reduced recalibration requirements. The methodology involves systematic data collection from a Norwegian WWTP, comprehensive preprocessing to handle data quality issues, feature engineering and rigorous comparative evaluation based on prediction accuracy, computational efficiency and adaptability. Expected outcomes include deployable soft sensor models offering reliable real-time monitoring capabilities, significant cost savings, and improved operational efficiency for WWTPs. The research contributes both theoretical insights into soft sensor design and practical solutions for the wastewater treatment industry.

soft sensors, wastewater treatment, machine learning, neural networks, decision trees, sensor drift, water quality monitoring

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all who contributed to the completion of this work.

First and foremost, I extend my deepest appreciation to Mr. Jeewaka Perera, who served as an exceptional mentor throughout his Journey. His invaluable advice, constant constructive criticism of my ideas and writing, and unwavering support were instrumental in shaping both the direction and quality of this research. His guidance challenged me to think more critically and write more clearly, ultimately making this a far stronger piece of work.

My sincere thanks go to DOSCON AS for providing the resources necessary for this research.

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT.....	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS.....	v
List of Figures	vii
List of Tables	vii
Chapter 1 Introduction	1
Chapter 2 Literature Review	3
2.1 Soft Sensor Fundamentals and Evolution	3
2.2 Machine Learning Approaches in Soft Sensor Development	4
2.2.1 Artificial Neural Networks	5
2.2.2 Decision Tree-Based Methods.....	6
2.3 Advanced Neural Architectures	8
2.4 Soft Sensors in Wastewater Treatment Applications	9
2.5 Data Quality and Preprocessing Challenges	10
2.6 Adaptability and Model Maintenance	12
2.7 Identified Research Gaps	14
Chapter 3 Problem Statement and Current	16
Challenges.....	16
3.1 The Challenge of Sensor Drift in Wastewater Treatment	16
3.2 Quantifying Recalibration Requirements.....	18
3.3 Root Causes of Performance Degradation	19
3.4 Limitations of Current Approaches.....	20
3.5 Economic and Operational Impact.....	20
3.6 Research Motivation	21
3.6.1 The Economic and Operational Imperative.....	21
3.6.2 Root Causes of Performance Degradation.....	22
3.6.3 The Promise of Adaptive Approaches.....	24
3.6.4 Autonomous Calibration and Validation.....	25
3.6.5 Advancing Process Control and Optimization	25
3.6.6 Addressing Root Causes Through Advanced Machine Learning	26
3.6.7 Broader Impacts and Future Vision.....	26
3.6.8 Research Objectives and Scope.....	27
Chapter 4 Methodology and Model Development.....	28

4.1 System Overview	28
4.1.1 Tier 1: Data Acquisition Layer	28
4.1.2 Tier 2: Data Visualization and Storage Server	29
4.1.3 Tier 3: Predictive Analytics Engine.....	29
4.2 Data Collection and Preprocessing	30
4.2.1 Missing Data Treatment	30
4.2.2 Outlier Detection and Treatment	31
4.2.3 Sensor Drift Correction	32
4.3 Decision Tree Model Development	33
4.4 Neural Network Architecture Design.....	34
4.5 Model Training and Validation	35
4.6 Performance Evaluation Metrics	36
Chapter 5 Results and Analysis	37
5.1 Decision Tree V1 Performance Results	37
5.2 Decision Tree V2 Performance Results	39
5.3 Neural Network V1 Performance Analysis.....	41
5.4 Neural Network V2 Performance Analysis.....	43
5.5 Comparative Performance Analysis	45
5.5.1 Comparative Analysis and Conclusions	45
5.5.2 Hyperparameter Optimization Results and Analysis.....	48
5.5.3 Possible Reasons for Performance Stagnant	48
5.6 Prediction Accuracy Comparison.....	49
5.7 Operational Deployment Considerations	49
5.8 Interpretability Evaluation.....	50
5.9 Temporal Performance Analysis	51
Chapter 6 Discussion	52
6.1 Implications for Wastewater Treatment Operations	52
6.2 Practical Implementation Strategies.....	52
6.3 Limitations and Future Directions.....	53
6.4 Economic Considerations.....	53
6.5 Regulatory Compliance and Environmental Impact	54
Chapter 7 Conclusion and Future Work	55
7.1 Conclusion.....	55
7.2 Future Work	56

List of Figures

Figure 1 Overview of TP	16
Figure 2 TP In Lab vs Pred	17
Figure 3 Brief System Overview	28
Figure 4 DT V1 Result.....	37
Figure 5 DT V2 Result.....	39
Figure 6 NN V1 Result	41
Figure 7 NN V2 Result	43

List of Tables

Table 1 Decision Tree Hyperparameters	33
Table 2 Neural Network Hyperparameters	35
Table 3 DT V1 KOF Outlet Lab vs Prediction	37
Table 4 DT V1 Total Phosphorus outlet Lab vs Prediction.....	38
Table 5 DT V2 KOF Outlet Lab vs Prediction	39
Table 6 DT V2 Total Phosphorous Outlet Lab vs Prediction	40
Table 7 NN V1 KOF Outlet Lab vs Prediction.....	41
Table 8 NN V1 Total Phosphorus Outlet Lab vs Prediction	42
Table 9 NN V2 KOF Outlet Lab vs Prediction.....	43
Table 10 NN V2 Total Phosphorous Outlet Lab vs Prediction	44
Table 11 DT V1 vs NN V1 Lab vs Predicted Total Phosphorous Outlet.....	47
Table 12 DT V1 vs NN V1 Lab vs Predicted KOF Outlet	47