1 Introduction

1.1 Background

1.2 Objectives

1.2.1 To evaluate baseline model performance of forecasting NH_3N by developing models with traditional and deep learning algorithms.

1.2.2 To develop raw data cleaning methodologies for improved machine learning model performance of forecasting NH_3N.

1.2.3 To explore feature engineering with the use of domain knowledge in wastewater treatment to creating new variables for optimizing model performance of forecasting NH_3N.

1.2.4 To construct an optimized procedure of training a machine learning model for NH_3N forecasting.

1.3 Organization of the thesis

2 Literature Review

2.1 Wastewater reclamation and reuse

2.1.1 Reclaimed water standard in wastewater treatment plant

2.1.2 Chlorination for reclaimed water disinfection

2.2 Control strategies for reclaimed water quality

2.2.1 Feed-forward control

2.2.2 Feed-back control

2.3 Water quality forecasting in wastewater treatment plant

2.3.1 Tools and technologies for parameter forecasting in wastewater treatment plant

2.3.2 Case study of machine learning models for water quality forecasting

2.4 Introduction to machine learning

2.4.1 Machine learning algorithms

2.4.2 Introduction to time-series data

2.4.3 Data pre-processing methodologies

2.4.3.1 Data cleaning

2.4.3.2 Outlier removal

2.4.3.3 Feature engineering

2.4.4 Model evaluation

2.4.4.1 Common metrics to evaluate model performance

2.4.4.2 Select baseline model performance

3 Materials and methods

3.1 Wastewater treatment plant description

3.1.1 Treatment processes in SHWEPP

3.1.2 Historical water quality data in SHWEPP

3.2 Raw data collection and preparation

3.2.1 NH_3N data monitoring and collection

3.2.2 Data cleaning and pre-processing

3.2.2.1 Data smoothing with mathematical solutions

3.2.2.1.1 Exponentially Weighted Moving Average filter

3.2.2.1.2 Savitzky–Golay filter

3.2.2.2 Outlier detection and removal

3.2.2.2.1 Peak detection analysis

3.2.3 Feature Engineering

3.2.3.1 Positional encoding

3.2.4 Data transformation

3.2.4.1 Split train/test dataset fot model input

3.3 Architectures of the machine learning algorithms

3.3.1 Random Forest

3.3.2 Multilayer perceptron

3.3.3 Long Short-Term Memory

3.3.4 Transformer

3.4 Proposed time series forecasting workflow

4 Results and Discussion

4.1 Comparisons of forecasting accuracy with traditional and deep learning models

4.1.1 Selection of baseline model performance

4.2 The influence of data pre-processing on model performance

4.2.1 Data cleaning

4.2.2 Outlier removal

4.3 The influence of feature engineering on model performance

4.3.1 Positional encoding as new input features

5 Conclusions and recommendations

Title:
Forecasting the Ammonia Concentration in the Reclaimed Water using Machine Learning
Study 1: Algorithm Selections and Data Pre-processing Methodologies in Developing Machine Learning Models for NH_3N Forecasting.
Study 2: To incoporate the use of domain knowledge in water/wastewater treatment and established methodology to construct a optimized model for NH_3N forecasting.