# CHAPTER 3

# METHODOLOGY

## 3.1 INTRODUCTION

In this chapter, we describe the methodology used for the predictive modelling of adsorption efficiency of nanocellulosic composite for the removal of cobalt II ion from waste water using Python. The methodology consists of four main steps: data collection, data preprocessing, model development, and model evaluation.

## 3.2 DATA COLLECTION

Dataset: Dataset with relevant parameters will be used for the prediction (Etuk, 2022). The independent parameters include initial adsorbate concentration (Co), contact time (t), influent flow rate (Q) and bed height (H) or adsorbent mass (m). While the dependent parameter is the Removal efficiency. (%R). The experiment was carried out in batches, at different bed height of 4cm, 8cm, and 12cm at constant flow rates, concentration and absorbent, another batch with different flow rates of 4ml/min, 8ml/min, and 12ml/min with constant bed height, concentration and absorbent, another batch with different concentration (50mL, 100mL, and 150mL) at constant bed height value, flow rate and concentration, The last batch consist of different absorbent(C-NFC, MF-NFC, C-MNFC, MF-NFC) at constant bed height, concentration, and flow rate.

## 3.3 DATA PREPROCESSING

The data were preprocessed using Python to prepare them for the model development. The preprocessing steps included the following:

* Checking for missing values and outliers, and handling them appropriately.
* Splitting the data into training and testing sets, using a 70:30 ratio and a random seed of 45, to ensure reproducibility of the results.
* Scaling the data using the standard scaler from the scikit-learn library, to transform the data into a standard normal distribution with zero mean and unit variance.
* Performing feature selection using the mutual information criterion from the scikit-learn library, to identify the most relevant features for predicting the adsorption efficiency.

## 3.4 MODEL DEVELOPMENT

The model development involved the following steps:

* Choosing the machine learning algorithms to be used for the predictive modelling. Based on the literature review and the nature of the dataset, three algorithms were selected: Ridge Regression, support vector regression (SVR), and Random Forest Regressor.
* Fitting the models on the training data using the scikit-learn library.

## 3.5 MODEL EVALUATION

The model evaluation involved the following steps:

* Applying the models on the testing data and obtaining the predicted values of the adsorption efficiency.
* Calculating the performance metrics of the models, such as the coefficient of determination (R2), the root mean square error (RMSE), and the mean absolute error (MAE), using the scikit-learn library.
* Comparing the performance metrics of the models and selecting the best model based on the highest R2 and the lowest RMSE and MAE values.
* Plotting the actual vs. predicted values of the adsorption efficiency for the best model using the matplotlib library, and assessing the goodness of fit and the residual distribution.

## 3.6 KINETIC MODELS AND THEIR ERRORS

The kinetic models are mathematical equations that describe the rate of adsorption of the adsorbate on the adsorbent as a function of time and other parameters. The kinetic models can provide information about the mechanism and the controlling factors of the adsorption process, as well as the adsorption capacity and the equilibrium time of the adsorbent. The kinetic models can also be used to compare the performance of different adsorbents and to optimize the adsorption conditions.

In this study, six kinetic models, namely Adams-Bohart, Thomas, Weibull, Wolborska, Yan, and Yoon-Nelson, were applied to the experimental data of the adsorption of cobalt II ion on the nanocellulosic composite. The equations and the parameters of the kinetic models are shown in Table 3.1.

|  |  |  |
| --- | --- | --- |
| **Model** | **Equation** | **Parameters** |
| Adams-Bohart | Ln = kBA C0t - kBAzN0  V0 | kBA corresponds to the kinetic rate constant for the Bohart-Adams’ model (Lmg-1min-1)  N0 represent the adsorbent saturation concentration (mgL-1) |
| Thomas | Ln = kTHq0 m -  kTHC0 t  Q | kTH - Thomas model constant (L min-1mg-1)  q0 - maximum adsorption capacity (mg g-1) |
| Weibull | = 1 – exp [ -] | > 0 is a rate parameter, and b > 0 is a shape parameter |
| Wolborska | Ln = βα  - βα  t  V0  N0 | βα represents the kinetic coefficient of the external mass transfer (min-1)  N0 is equivalent to saturation concentration or adsorption capacity (mgL-1) |
| Yan |  | is the maximum uptake capacity (mg/g),  a is an empirical parameter that decides the slope of the regression function |
| Yoon-Nelson | Ln= kYN t - kYN Ʈ, | kYN stands for the Yoon-Nelson rate constant (min-1)  Ʈ refers to the length of time (min) necessary for the attainment of 50% initial adsorbent concentration in the effluent stream |

Table 3. 1

## 3.7 ERROR FUNCTIONS

The parameters of the kinetic models were estimated by fitting the models to the experimental data using the nonlinear least squares method from the scipy library in Python. The errors of the kinetic models were calculated using the following metrics:

**Error functions Definition/expressions**

The coefficient of determination

The sum of squares of the errors (ERRSQ/SSE)

The hybrid error functions (HYBRID)

Marquardt’s percent standard deviation (MPSD) 2

The average relative error (ARE) i

The sum of absolute errors (SAE) i

Mean square error (MSE)

Nonlinear chi square (ꭓ2)

Source: Oboh (2011)

Where, qe, cal is the theoretical concentration of adsorbate on the adsorbent, which have been

calculated from one of the isotherm models, qe, exp is the experimentally measured adsorbed

solid phase concentration of the adsorbate adsorbed on the adsorbent. (n) is the number of data

points, (p) is the number of parameters in the equation.