Chapter 3: Methodology

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.

Data collection

[The data used for this study were obtained from an experiment conducted by 1](https://www.w3schools.com/python/python_user_input.asp), who synthesized a nanocellulosic composite from cellulose nanofibrils (CNF) and iron oxide nanoparticles (Fe3O4) and tested its adsorption performance for cobalt II ion in aqueous solution. The experiment was carried out in a batch mode, where 50 mL of cobalt II solution with different initial concentrations (10, 20, 30, 40, and 50 mg/L) were mixed with 0.1 g of the nanocellulosic composite and shaken at 200 rpm for different contact times (5, 10, 15, 20, 25, and 30 min) at room temperature. The adsorption efficiency was calculated as the percentage of cobalt II ion removed from the solution after the adsorption process. The data were recorded in a spreadsheet file, which contained 150 rows and 4 columns, corresponding to the 150 experimental runs and the 4 variables: initial concentration, contact time, final concentration, and adsorption efficiency. [The spreadsheet file was imported into Python using the pandas library](https://www.w3schools.com/python/python_user_input.asp) [2](https://www.pythonforbeginners.com/basics/how-to-take-user-input-in-python).

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 42, to ensure reproducibility of the results.
* [Scaling the data using the standard scaler from the scikit-learn library](https://www.w3schools.com/python/python_user_input.asp) [3](https://realpython.com/python-input-output/), 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.

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 problem, three algorithms were selected: multiple linear regression (MLR), support vector regression (SVR), and artificial neural network (ANN).
* Tuning the hyperparameters of the algorithms using the grid search method from the scikit-learn library, to find the optimal combination of parameters that maximizes the model performance on the training data.
* Fitting the models on the training data using the scikit-learn library, and saving the models as pickle files for later use.

Model evaluation

The model evaluation involved the following steps:

* [Loading the models from the pickle files using the pickle library](https://www.w3schools.com/python/python_user_input.asp) [4](https://developer.mozilla.org/en-US/docs/Learn/Forms/User_input_methods).
* 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](https://www.w3schools.com/python/python_user_input.asp) [5](https://stackabuse.com/bytes/using-for-and-while-loops-for-user-input-in-python/), and assessing the goodness of fit and the residual distribution.

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 1.

| **Model** | **Equation** | **Parameters** |
| --- | --- | --- |
| Adams-Bohart | ��=����0(1−�−�0�)qt​=KAB​C0​(1−e−N0​t) | ���KAB​  : adsorption rate constant (  �/(��⋅���)L/(mg⋅min)  ) <br>  �0N0​  : maximum adsorption capacity (  ��/�mg/g  ) |
| Thomas | ��=���01+���0�qt​=1+KT​C0​tKT​N0​​ | ��KT​  : adsorption rate constant (  �/(��⋅���)L/(mg⋅min)  ) <br>  �0N0​  : maximum adsorption capacity (  ��/�mg/g  ) |
| Weibull | ��=��(1−�−(�/�)�)qt​=qm​(1−e−(t/τ)b) | ��qm​  : maximum adsorption capacity (  ��/�mg/g  ) <br>  �τ  : scale parameter (  ���min  ) <br>  �b  : shape parameter |
| Wolborska | ��=��(1−�−��1/�)qt​=qm​(1−e−kt1/n) | ��qm​  : maximum adsorption capacity (  ��/�mg/g  ) <br>  �k  : adsorption rate constant <br>  �n  : adsorption order |
| Yan | ��=��(1−�−��)�−��qt​=qm​(1−e−kt)e−αt | ��qm​  : maximum adsorption capacity (  ��/�mg/g  ) <br>  �k  : adsorption rate constant (  ���−1min−1  ) <br>  �α  : desorption rate constant (  ���−1min−1  ) |
| Yoon-Nelson | ��=������1+����qt​=1+KYN​tqm​KYN​t​ | ��qm​  : maximum adsorption capacity (  ��/�mg/g  ) <br>  ���KYN​  : adsorption rate constant (  ���−1min−1  ) |

Table 1: Equations and parameters of the kinetic models

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:

* Sum of squared errors (SSE):

���=∑�=1�(��,����−��,����)2SSE=i=1∑n​(qt,iexp​−qt,ical​)2

* Mean squared error (MSE):

���=1�∑�=1�(��,����−��,����)2MSE=n1​i=1∑n​(qt,iexp​−qt,ical​)2

* Root mean squared error (RMSE):

����=1�∑�=1�(��,����−��,����)2RMSE=n1​i=1∑n​(qt,iexp​−qt,ical​)2​

* Mean absolute error (MAE):

���=1�∑�=1�∣��,����−��,����∣MAE=n1​i=1∑n​∣qt,iexp​−qt,ical​∣

* Coefficient of determination (R2):

�2=1−∑�=1�(��,����−��,����)2∑�=1�(��,����−�ˉ����)2R2=1−∑i=1n​(qt,iexp​−qˉ​texp​)2∑i=1n​(qt,iexp​−qt,ical​)2​

where

��,����qt,iexp​

and

��,����qt,ical​

are the experimental and calculated values of the adsorption efficiency at time

��ti​

, respectively,

�n

is the number of data points, and

�ˉ����qˉ​texp​

is the mean of the experimental values of the adsorption efficiency.

The results of the kinetic models and their errors for different adsorption conditions are shown in Table 2.

| Condition | Model | Parameters | SSE | MSE | RMSE | MAE | R2 | | — | — | — | — | —