CHAPTER THREEMATERIAL AND METHODS

3.1 Materials

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).

3.20 Method A python program will be developed to implement the model based on the collected dataset as follows: data collection, the preprocessing stage which involves the handling of missing data, the training, and testing of the machine learning models and lastly, performance analysis and comparison. The data from the dataset will be categorized into LOW, MEDIUM and HIGH respectively and this approach will be used to determine which adsorbent has the higher removal efficiency prioritize on that. The dataset is being categorized based on the %R from the dataset, so for values ranging From 0-30% will be classified as LOW, 31-60% - MEDIUM and 61% and above will be Classified as HIGH.

The first step is to collect or gather relevant data. The data will be collected from Etuk’s library (Etuk 2022).The next step is to import the necessary libraries such as pandas, NumPy, matplotlib, seaborn, and scikit-learn. The necessary libraries for prediction will be imported in order to carryout data analysis. The next step in the process is to prepare the data. It involves cleaning the data to remove unwanted and redundant values, converting it into the right format eg. .csv, and making it ready for analysis. It also requires data wrangling. Data wrangling is the process of transforming and structuring data from one raw form into a desired format with the intent of improving data quality and making it more consumable and useful for analytics. The data collected will be pre-processed, the preprocessing is done so as to handle the missing values in the data and a feature scaling technique is employed to normalize the data. Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. After the data is ready, data exploration is done using various data visualization techniques to find unseen trends from the data. The data is split into training set (80%) and test set (20%). The training is used to train the prediction models, while the test set is used for validation purpose. The next step is to build your predictive models using machine learning algorithms to make future predictions. Model evaluation Metrics, correlation coefficient (R2), mean square error (MSE) will be used to determine the model’s predictive capabilities. After analyzing the data and fitting the models, the next step is to compare different adsorption models and kinetic models to identify the best-fitting model that describes the adsorption behavior of cobalt ions onto the adsorbent material.