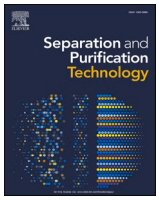
Separation and Purification Technology 315 (2023) 123712

Contents lists available at ScienceDirect

Separation and Purification Technology

journal homepage: www.elsevier.com/locate/seppur



Adsorption of organic dyes onto nanozeolites: A machine learning study

Leandro Rodrigues Oviedo, Vinícius Rodrigues Oviedo, Lissandro Dornelles Dalla Nora, William Leonardo da Silva *

Nanoscience Graduate Program, Universidade Franciscana, Santa Maria - RS, Brazil

ARTICLE INFO

Keywords:
Nanozeolites
Batch adsorption
Random Forest
LGB
XGB
ANN

ABSTRACT

Wastewater pollution with organic dyes has generated great concern in society due to the hazardous effects these contaminants pose to humans and aquatic life. In this view, the application of the adsorption process using nanoadsorbents has been a promising alternative due to the relatively low cost, high efficiency and simple operation. In addition, nanozeolites were highlighted in scientific literature due to their properties (surface area, porosity, ion exchange capacity, chemical and thermal stability), being useful for dye removal from wastewater. However, time and cost in experimental procedures are required to find optimal conditions for the adsorption of dyes onto these nanozeolites. Therefore, machine learning methods have emerged as a suitable tool for the prediction of the adsorption capacity of the nanoadsorbents in an efficient manner, being capable of recognizing patterns in the process and addressing the process feasibility. In this context, the present work aims to develop a machine learning (ML) study of the adsorption of organic dye onto nanozeolites and to identify the main variables that affect the adsorption capacity and removal of organic dye from wastewater. Thus, four ML algorithms (RF, LGB, XBG, and ANN) were tested as a regression model. This study revealed that XBG showed the best performance in comparison to the other models, being suitable in the prediction of adsorption capacities of nanozeolites for cationic dyes. Additionally, an exploratory analysis and hypothesis testing confirmed the great effect of the dye and nanoadsorbent concentrations, contact time and pH in the adsorption process. Therefore, the XGB proved to be capable to address the predicted adsorption capacity of nanozeolite from a relatively small dataset, being characterized as a starting point before experimental procedures and scale-up of wastewater treatment concerned with organic dye removal.

1. Introduction

The contamination of wastewater with organic dyes has been increasing fast mainly due to the expansion of the textile, leather and pulp industries, and due to the inadequate discharge of these dyes [1]. Among these industries, the textile industry is the one that consumes a great amount of water and then, generates wastewater with high content of organic dyes in it [2,3]. In summary, 700,000 tons of organic dyes are produced annually, where 140,00 tons are generated as a constituent of textile wastewater [4].

Organic dyes can change the fauna and flora, obstructing light absorption by the water body and interfering with photosynthetic processes [5]. Moreover, these contaminants can cause severe damage to humans, such as eye and skin irritation and respiratory issues. In addition, organic dyes can show carcinogenic and mutagenic effects depending on the water concentration [6]. For example, measurement informed in some scientific reports for a real sample of textile wastewater has been quantified dye concentrations ranging from 10 to 200 mg L^{-1} [7].

Regarding environmental issues, synthetic organic dyes show low biodegradability, high chemical stability, water solubility and complex structure, which turns their removal extremely difficult through conventional wastewater treatment [8]. In this view, efficient and cost-effective technologies, such as physical–chemical treatments using nanomaterials, can be promising in fixing these environmental problems [9].

Adsorption is characterized as a simple process with a relatively low cost, ease of operation and low generation of secondary hazardous pollutants [10,11]. Adsorption can be either physical (physisorption) or chemical (chemisorption), depending on the nature of the interaction between adsorbate (i.e., an organic dye) and adsorbent [12]. Physisorption involves weak interactions, such as hydrophobic and

* Corresponding author.
E-mail address: williamleonardo_silva@hotmail.com (W.L. da Silva).

https://doi.org/10.1016/j.seppur.2023.123712
Received 17 March 2023; Received in revised form 24 March 2023; Accepted 26 March 2023
Available online 31 March 2023

hydrophilic interactions or Van der Waals forces, being reversible, once requires considerably low heat to desorb the adsorbate and, consequently, to regenerate the adsorbent [13]. Moreover, chemisorption is based on stronger interaction forces than those associated with physisorption, including π- π interaction, and covalent chemical bonds, which in turn required high energy consumption to regenerate the adsorbent [14].

Nanomaterials, such as nanozeolites, have proved to be good nanoadsorbent used for the removal of synthetic organic dyes from water and wastewater, mainly due to their size-dependent (high surface area, meso and microporosity, unique electronic behavior) and intrinsic properties (high chemical and hydrothermal stability, ion exchange capacity, uniform porosity) [15].

Nanozeolites can be synthesized by (i) chemical methods, in which organic solvents and templates are used to tunnel the proprieties of the product; (ii) simple physio-chemical methods without the use of organic solvents and templates (hydrothermal, geothermal or ionothermal synthesis using whatever chemical source of alum and silicon) at closed system; and (iii) alternative methods (mainly hydrothermal synthesis), in which the synthesis is based on the use of residual materials as a source of alum and silicon, i.e., rice husk, coal fly ash, and alum sludge [16,17].

Nanozeolites showed good adsorption capacity for synthetic organic dyes. However, due to the vast diversity of nanozeolite topologies and morphologies, an experimental study to verify the best type of zeolite-based nanoadsorbent is inviable from an economic and temporal viewpoint [18]. Therefore, computational studies, such as machine learning methods, are characterized as a good alternative to fix environmental problems of dye-contaminated wastewater, once it requires data selection from literature and pattern identification through a specific algorithm for further predictions, being relatively cheap and time-effective procedure prior experimental runs [19]. Also, computational studies can help in the identification of the main variables that affect the adsorption capacity or dye removal [20,21].

In this view, the present work aims to apply a machine learning-based regression model for predicting the adsorption capacity of nanozeolites for synthetic organic dyes. In addition, the effect of the type of agitation, dye nature and molecular weight on the adsorption capacity was investigated through exploratory data analysis.

2. Material and methods

2.1. Dataset

The adsorption of the dyes onto the nanozeolites (nZ) was studied by collecting adsorption data published in scientific research papers. For this purpose, the Scopus indexer (https://www.scopus.com) was used to filter and select the articles for the machine learning study. Then, the following descriptors were used: “nanozeolite” AND “adsorption” AND “dye”.

2.2. Machine learning study

To develop a machine learning-based predictive model four algorithms were carried out: Light Gradient Boosting (LGB), Xtreme Gradient Boosting (XGB), Random Forest (RF) and Artificial Neural Network (ANN). Thus, from this model is possible to make time and cost-effective generalizations, which is extremely useful before experimental runs and scale-up. All ML models were performed in Python 3.8 (open-source version at Google Colaboratory).

2.3. Data processing

The data set was arranged in a matrix of 446 rows and 16 columns, summing up to 7,136 data points. These data were divided into two subsets, the training (80%) and testing dataset (20%). The molecular weight (MW), pH, dye ([Dye]) and nanoadsorbent ([nZ]) concentrations and contact time (t). The MW was used to identify the dye and its nature (cationic and anionic). The Ag_{mode} were sonication and adsorption under magnetic stirring, respectively. Additionally, an exploratory analysis of the dye nature, MW and the type of agitation (Ag_{mode}) was carried out to verify how these variables affect the response variable. Thereafter, the input data were normalized using Eq. (1), in which the coefficients $\beta = 0.5$ and $\gamma = 2$ were used in this work [22].

$$x_{in} = \gamma \frac{x_i - x_{min}}{x_{max} - x_{min}} - \beta\gamma \tag{1}$$

Where x_i is i-th value of the raw input variable x; x_{min} and x_{max} are the minimum and maximum values of the variable x_i of the input data; x_{in} is the i-th value of the normalized input variable x. Thus, Table 1 shows the ML models used and their configurations tested to find the best model performances. Although all ML methods are black-box models, the ANN showed the well-known continuous equations used in the activation step, which are expressed as follow:

$$\phi(x_i) = \sum_i^n w_i \bullet x_i + b_k \tag{2}$$

$$f(\phi) = \frac{1}{1 - e^{-\phi(x_i)}} \tag{3}$$

$$f(\phi) = \tanh(\phi(x_i)) = \frac{2}{1 + e^{-2[\phi(x_i)]}} - 1 \tag{4}$$

Where w_i is the i-th weight associated to the input variable x_i; $\phi(x_i)$ and $f(\phi)$ are the summation/transfer and activation functions. The former is responsible for the codification of the input data, whereas the latter is useful to define if the neuron (node in the hidden layer) is activated or not, depending on the product between w_i and x_i. The parameter b_k is the bias (uncertainty coefficient) associated to the model.

Table 1
ML models and configurations used for the prediction of nZ adsorption capacity.

Model	Configurations	Reference
RF	Number of estimators: 1, 2, 5, 10, 20, 30, 50, 100 and 200 Maximum depth of the decision tress: 1, 5, 10, 15, 20, 25, 50, and 100 Sampling method: bootstrap sampling Model characterization: complex model Loss function: RMSE	[23]
LGB	Number of estimators: 10, 100, and 1000 decision trees Maximum depth of the decision tress: 1, 2, 5, 10, 15, 20, and 25 Model characterization: baseline model Loss function: RMSE	[24]
XBG	Learning rate: 0.5Booster methods: Gradient Boosting Decision Tree (GBDT) and Dropout Regularization in Boosting Ensembles (Dart) Maximum depth of the decision tress: 1, 2, 5, 10, 15, 20, and 25 Sampling method: bootstrap sampling Loss function: RMSE	[25]
ANN	Neural network structure (input-hidden layer-output): 6-K-1Number of neurons tested (K) : 1–15Method: Perceptron Regressor Model (one hidden layer)Activation functions: logistic (Eq. (3)) and hyperbolic tangent function (Eq. (4)) based in the summation function (Eq. (2))Solver: Stochastic Gradient Descend (Adam function) Regression method: non-linear regression bases on Levenberg-Marquardt algorithm Loss function: RMSE Threshold value: tolerance $\varepsilon < 10^{-4}$	[26]

2.4. Model performance

The ML model performances were evaluated in terms of the correlation coefficient (R^2) and the root means squared error (RMSE), according to Eq. (5–6), where higher R^2 and lower RMSE values indicate greater model performance [27].

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_{i,exp} - y_{i,pred}\right)^2}{\sum_{i=1}^{N}\left(y_{i,exp} - \hat{y}_{i,pred}\right)^2} \quad (5)$$

$$RMSE = 1 - \sqrt{\frac{\sum_{i=1}^{N}\left(y_{i,exp} - \hat{y}_{i,pred}\right)^2}{N}} \quad (6)$$

Where: $y_{i,\mathrm{exp}}$ and $\hat{y}_{i,pred}$ are the actual and the predicted value of the response (q_e); N is the data size. The RMSE is expressed in mg g^{-1}.

2.5. Feature importance

Owing to the nature of the ML models tested, which are black-box models, a feature importance study was carried out to explain the contribution of each input variable on the target variable q_e. Thus, a permutation feature score was used to demonstrate which input variables can significantly change the response of the model when these input variables are changed. Higher permutation feature scores higher the contribution of the feature in the response [28].

3. Results and discussion

3.1. Exploratory analysis

The dataset was comprised of 7 dyes randomly distributed in 446 rows. To investigate the effect of the molecular weight on the target variable adsorption capacity, a kernel density estimation plot (kdeplot) was generated, as can be seen in Fig. 1.

According to Fig. 1, the MW of the dyes of the dataset ranged from 240-260 to 1100-1300 g mol^{-1}, in which the main values for q_e concentrated between 0-25 mg g^{-1} and between 0-60 mg g^{-1} (Fig. 2). Moreover, it was noticeable that the MW values in the dataset do not follow normal distribution and the Kolmogorov-Smirnov Normality test was applied to the data, in which p-value ($p<0.005$) confirmed the existence of a non-normal distribution of data. Spearman correlation was performed to address the correlation between these two variables (Fig. 3).

According to Fig. 2, it is noticeable that the MW values in the dataset do not follow normal distribution. Moreover, the Kolmogorov-Smirnov Normality test was applied to the data, in which p-value ($p < 0.005$) confirmed the existence of a non-normal distribution of data. Furthermore, the Spearman correlation was performed to address the correlation between these two variables, MW and q_e. The result is shown in Fig. 3.

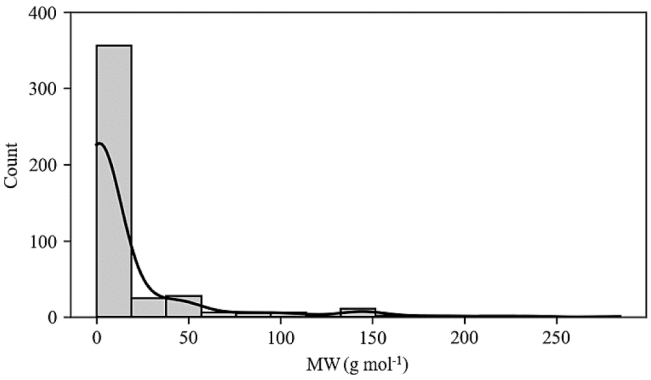


Fig. 2. Data distribution of *MW* with respect to q_e.

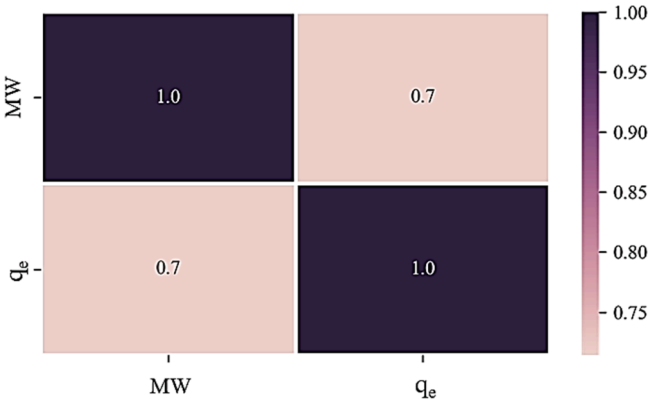


Fig. 3. Spearman correlation between MW and q_e data.

As can be observable in Fig. 3., there is a strong correlation between q_e and MW ($\rho = 0.7$), suggesting that lower values of MW tend to result in higher values for q_e, although this dependency does not follow a linear behavior.

Fig. 4 shows the bar plot used to investigate the effect of the type of agitation on dye adsorption.

According to the Fig. 4, higher qe values were reported when adsorption occurs under magnetic stirring and the Kolmogorov-Smirnov and the Mann-Whitney tests were carried out to investigate the data normality and whether there was a statistical difference of qe for the two

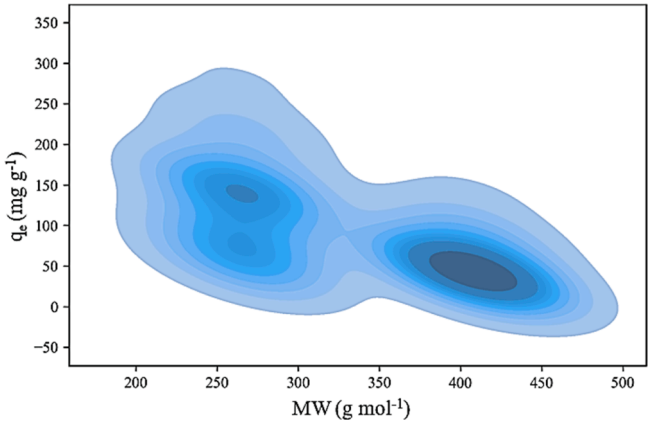


Fig. 1. Effect of the molecular weight (MW) on the adsorption capacity (q_e).

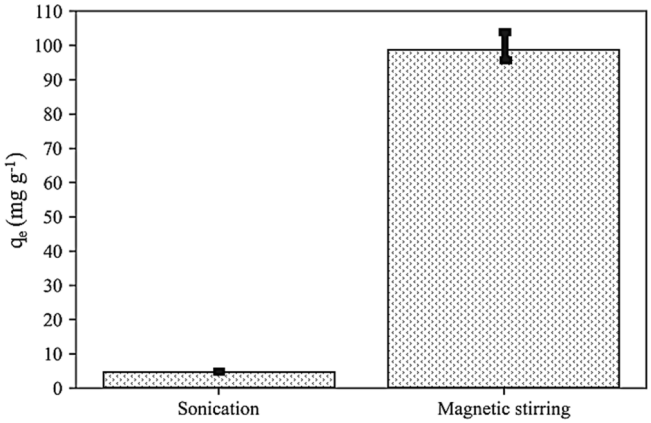


Fig. 4. Effect of the type of agitation on the dye adsorption in nanozeolites.

types of agitation, respectively (Table 2).

As can be seen in Table 2, the data did not follow a normal distribution. Furthermore, the Mann-Whitney hypothesis test showed that there was a significant difference in the q_e value regarding the type of agitation ($p<0.001$). Thus, it suggests that magnetic stirring (71.25 ± 8.640 mg g^{-1}) indicated higher values of adsorption capacity (q_e) in relation to sonication (0.31 ± 0.016 mg g^{-1}). Fig. 5 shows the q_e values for different dye natures where the dataset used in this work was based on 258 datapoints for anionic dyes and 188 for cationic ones.

According to Fig. 4, higher q_e values were reported for cationic dyes than anionic ones, due to the electrostatic interactions between the positive surface of the cationic dye with the negative surface of the nZ promoting greater adsorption capacity (q_e). Additionally, normality (Kolmogorov-Smirnov) and hypothesis (Mann-Whitney) tests were performed to investigate the data distribution and the effect of the dye nature on the target variable q_e (Table 3).

According to Table 3, non-normal distribution was observable for data and the Mann-Whitney test showed a statistical difference for the q_e value regarding the nature of the dye ($p<0.001$), where cationic dyes (12.28 ± 0.895 mg g^{-1}) resulted in higher q_e values in relation to the anionic dyes (0.23 ± 0.014 mg g^{-1}).

3.2. Machine learning models

Type of agitation and the dyes nature were the chosen parameters for the selected dataset in the machine learning study, where only two cation dyes (Crystal Violet and Acridine Orange) were reported, and the dataset was constituted of 1088 data points (68 rows and 16 columns). The ML models tested were RF, LGB, XBG and ANNT. Fig. 5 shows a violin plot generated to investigate the distribution of qe data for the two dyes.

According to Fig. 6(a), the MW of the cationic dyes concentrated from 260 to 270 g mol^{-1} and from 380-440 g mol^{-1}, where higher q_e values were observed for dyes with the lowest MW. Furthermore, Fig. 6 (b) confirmed this pattern, in which q_e ranged from 50 to 250 mg g^{-1} (mean ~ 120 mg g^{-1}) was observed for AO dye, whereas q_e values were smaller than 100 mg g^{-1} to CV dye (mean ~ 48 mg g^{-1}). 20% of the dataset (476 datapoints) was used for testing and training each ML model. The RMSE and R^2 metrics for train and test models were compared for further model selection. Table 4 shows the comparison among the models and the best configuration for each of them.

According to Table 4, the LGB showed the lowest model performance due to extremely low R^2 for training and testing data, although low RMSE. It was probably due to the number of the dataset, which might be short enough to generate a good predictive LGB model [29]. Moreover, the fact of the LGB algorithm using discrete values and do not consider the variables with feature importance equal to or close to zero (such as temperature and adsorbate MW, which are important to adsorption), despite showing strong effect on the adsorption capacity (response) as proved in experimental run published in literature [30]. Regarding the algorithm that generated the best model performance, the opposite behavior was observed, which is better described in the following section.

Table 2
Normality and hypothesis test for qe regarding the type of agitation with 95% confidence interval.

Kolmogorov-Smirnov normality test		
	Magnetic stirring	**Sonication**
Statistical value	1.0000	0.5264
p-value	<0.001	<0.001
Mann-Whitney hypothesis testing		
Statistical value	129.0000	
p-value	<0.001	

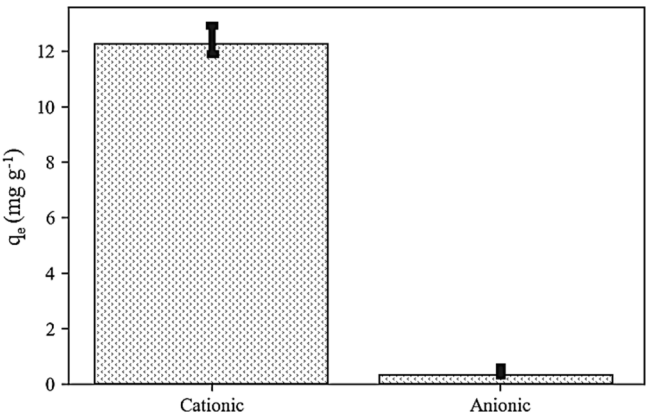


Fig. 5. Effect of the dye nature on the dye adsorption in nanozeolites.

Table 3
Normality and hypothesis test for q_e regarding dye nature

Kolmogorov-Smirnov normality test		
	Anionic dyes	**Sonication**
Statistical value	0.5264	0.5264
p-value	<0.001	<0.001
Mann-Whitney hypothesis test		
Statistical value	8.0000	
p-value	<0.001	

3.3. Model selection

Regarding the results reported in Table 3, it was noticeable that the XGB model showed better performance than the other ML models, once the highest R^2 and lowest RMSE values were reported for this model. Furthermore, the effect of the max depth of the decision trees used in the model in the training step and the booster method on the ML model performance were investigated. Fig. 7 shows the effect of the max tree depths on the RMSE generated from the training step.

According to Fig. 7, the RMSE converged to 0.6628 mg g^{-1} when 10 m was set as the maximum depth of the decision trees used in the XGB model. Thus, this value was further used in the regression model for predicting the adsorption capacity of the nanozeolites for the cationic dyes CV and AO. However, the booster method (Dart and GBtree) showed no significative difference and hence the little effect on the model performance. Thus, Dart function was set in the model selection step.

Fig. 8 shows the correlation between the actual and the predicted q_e values for (a) train and (b) test dataset.

According to Fig. 8, the training model showed a higher correlation ($R^2 = 0.9994$) between the predicted and observed q_e values. In addition, the training RMSE value (0.6628 mg g^{-1}) lower than test RMSE (4.1048 mg g^{-1}) confirmed this observation. Owing to the relatively high RSME of the test model, it was expected to find a R^2 ranging from 0.80 to 0.90. The test R^2 reported was 0.8030. Moreover, this behavior can be explained due to the nature of the test dataset, which is based on data new for the XGB model, and hence, the generation of disparities should be normal. However, this difference algo suggest no overfitting associated to the literature [23]. Nevertheless, the XGB model developed for the test dataset was characterized as a good generalizer model, once it showed a good ability to predict untrained data (R^2 greater than 0.75 and low RMSE, which is an indicates that good ability of the model to predict data with low absolute standard deviation) [31].

3.4. Data correlation

The correlation among the variables selected for the ML study was

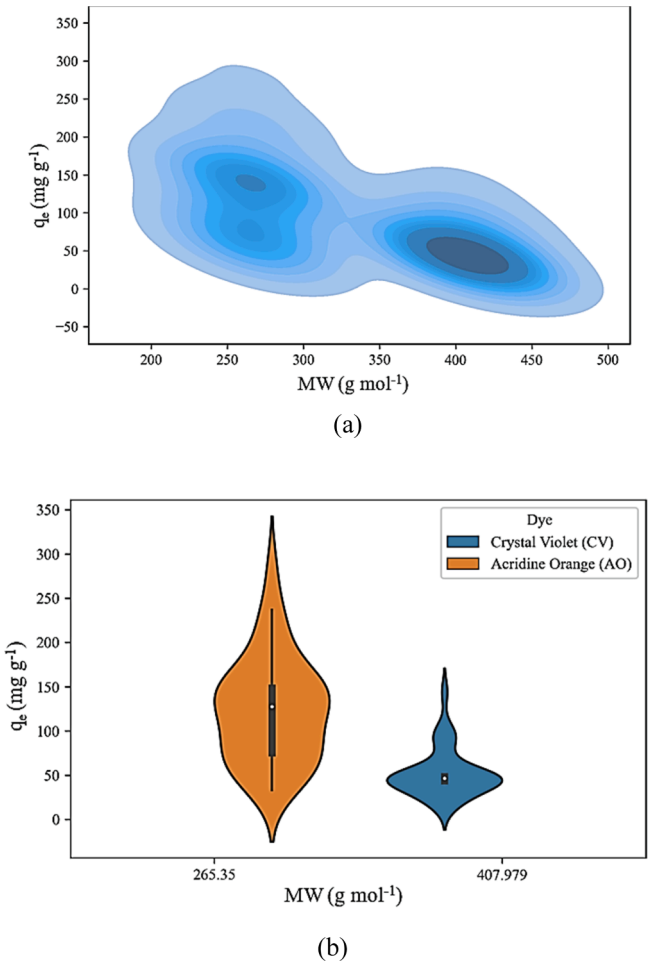


Fig. 6. Data distribution addressed by (a) kdeplot and (b) violin plot.
Operational parameters: [nZ] = 0.5-4.5 g L^{-1} | [Dye] = 15-900 mg L^{-1} | pH 3-11 | contact time = 5-150 min.

Table 4
ML model performance.
Parameters: k-fold cross validations as validation method (cv = 3, 5, 7, 10) | Loss functions: RMSE (mg g^{-1}) | Metrics: R^2 for training and testing

Model	Training R^2	Training RMSE	Testing R^2	Testing RMSE	Best model configuration
RF	0.9558	2.5114	0.6261	4.4754	Number of decision trees: 50, max. tree depth: 15 *m*
LGB	−0.1211	5.1125	−1.5071	5.6757	Number of decision trees: 200, max. tree depth:15 *m*
XGB	0.9994	0.6628	0.8030	4.1048	Booster method: dart; max. tree depth: 15 *m*
ANN	0.9991	1.4996	0.8416	32.4899	Number of epochs: 80,000; activation function: hyperbolic tangent function; neural network: 6–12-1 (1 hidden layer with 12 neurons)

presented in Fig. 9. Due to the non-normality of the dataset, the Spearman correlation was used to correlate data of adsorption.

According to Fig. 9, the organic dye removal (R_{dye}) is negatively affected by nanozeolite and dye concentration and positively affected by

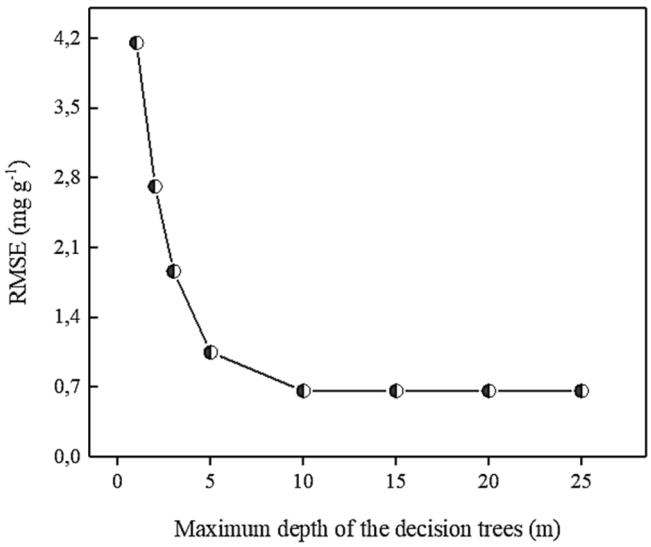


Fig. 7. RMSE as a function of the maximum depth of the decision trees used in the XGB model.
Model parameters: Boosting method: dart, k-fold cross validation = 10.

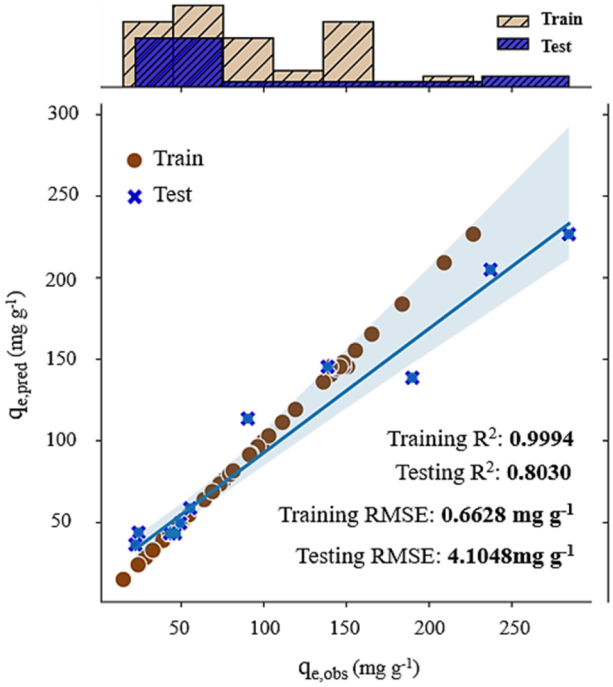


Fig. 8. Predicted versus observed (actual) values for training and testing dataset.
Operational parameters: [nZ] = 0.5-4.5 g L^{-1} | [Dye] = 15-900 mg L^{-1} | pH 3-11 | contact time = 5-150 min.

temperature, contact time, molecular weight and pH. On the contrary, the dye molecular weight showed to negatively influence the adsorption capacity q_e. It suggests that lower MW should result in higher q_e values, in a non-linear manner [32]. However, the variables pH, time and T showed a negative influence on the qe, suggesting that the high q_e range reported in the dataset was due to the high uptake of the adsorbate in the adsorbent in a short period of contact time. Also, the positive effect of [Dye] and [nZ] on qe is probably due to the concentration range found in the dataset, which was not too high to result in low dye adsorption capacities.

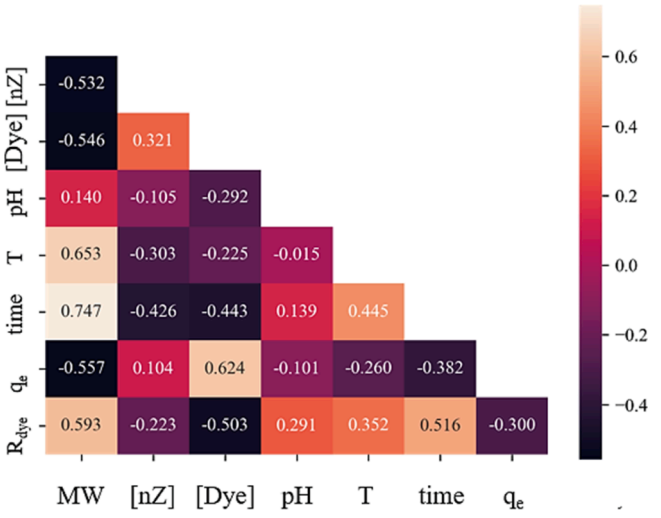


Fig. 9. Spearman correlation among the input and the response variables.

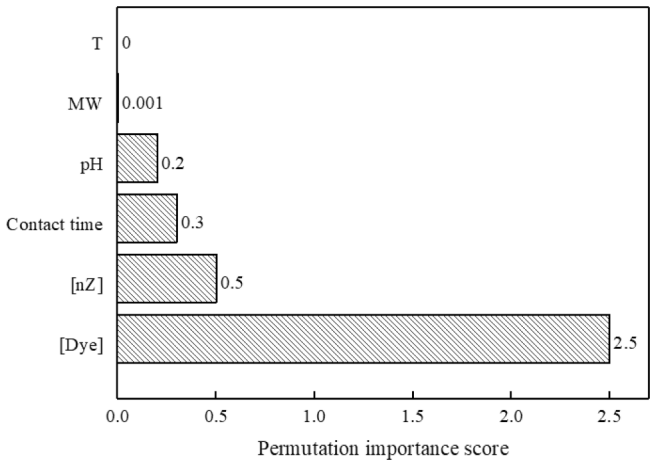


Fig. 10. Importance of each input variable on the responses q_e and R_{dye}.

3.5. *Feature importance*

The effect of each feature (input variable) on the q_e is studied by permutation importance. The permutation importance informs us how a variable affects the value of the response variable, that is, the greater the feature importance for a feature greater the change in the output value, which can result in errors and hence reduce the regression model performance. Fig. 10 shows the permutation importance of the variables selected in the ML model development, in which variables with score greater than zero was considered to have great contribution to the response. However, higher scores suggested higher influence of the input variable on the response.

According to Fig. 10, dye and nanozeolite concentrations ([Dye] and [nZ]) showed a higher effect on the adsorption capacity. Thus, higher modification in [Dye] and [nZ] can significantly change the value of q_e, which is generally observed in experimental procedures. Furthermore, the contact time and pH also showed to affect the response variable, to some extent.

4. Conclusion

In this work, four black-box ML regression models were used to predict the adsorption capacity of low-silica nanozeolites for organic dyes. Thus, the effect of dye nature (cationic and anionic), type of agitation (sonication and magnetic stirring) and molecular weight (low and high MW) were investigated on the response variable q_e. According to the results, the XGB and ANN models showed the higher performance to predict the adsorption capacities of nanozeolites. The XGB was selected as the best model due to the highest determination coefficient (training $R^2 = 0.9994$, testing $R^2 = 0.8030$) and RSME (0.6628 mg g^{-1} for train and 4.1048 mg g^{-1} for test) values. The exploratory analysis revealed that the adsorption capacity of nZ was higher for MW cationic dye (~120 mg g^{-1}) than high MW cationic dye (~48 mg g^{-1}). The feature importance revealed that dye (PIS = 2.5) and nanozeolite concentrations (PIS = 0.5) showed a higher effect on the adsorption capacity, and to a lesser extent, the contact time (PIS = 0.3) and pH (PIS = 0.2). In summary, the XGB model developed was characterized as a good generalizer prediction model, once resulting from R^2 greater than 0.80 and RSME less than 5.00. Therefore, the ML regression models proved to be a useful tool for the prediction of experimental data of adsorption, reducing costs and time associated with experimental runs.

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author.

CRediT authorship contribution statement

Leandro Rodrigues Oviedo: Conceptualization, Data curation, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **Vinícius Rodrigues Oviedo:** Conceptualization, Writing – review & editing. **Lissandro Dornelles Dalla Nora:** Conceptualization, Validation, Writing – review & editing. **William Leonardo da Silva:** Conceptualization, Data curation, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was financed by the CAPES (Coordination of Superior Level Staff Improvement) - Finance Code 001.

References

[1] D.A. Yaseen, M. Scholz, Textile dye wastewater characteristics and constituents of synthetic effluents: a critical review, Int. J. Environ. Sci. Technol. 16 (2019) 1193–1226, https://doi.org/10.1007/s13762-018-2130-z.
[2] B.L. Alderete, J. da Silva, R. Godoi, F.R. da Silva, S.R. Taffarel, L.P. da Silva, A.L. H. Garcia, H.M. Júnior, H.L.N. de Amorim, J.N. Picada, Evaluation of toxicity and mutagenicity of a synthetic effluent containing azo dye after Advanced Oxidation Process treatment, Chemosphere. 263 (2021), 128291, https://doi.org/10.1016/J.CHEMOSPHERE.2020.128291.
[3] K. Chaieb, B. Kouidhi, L. Ayed, S. Bakr Hosawi, J. Abdulbaqi Abdulhakim, A. Hajri, H.N. Altayb, Enhanced textile dye removal from wastewater using natural biosorbent and Shewanella algae B29: Application of Box Behnken design and genomic approach, Bioresour. Technol. 374 (2023), 128755, https://doi.org/10.1016/J.BIORTECH.2023.128755.
[4] L. Hossain, S.K. Sarker, M.S. Khan, Evaluation of present and future wastewater impacts of textile dyeing industries in Bangladesh, Environ. Dev. 26 (2018) 23–33, https://doi.org/10.1016/j.envdev.2018.03.005.
[5] D. Li, S. Ge, Y. Xiang, J. Gong, C. Liu, G. Sun, J. Xu, W. Fa, J. Ma, A simple and facile bioinspired catalytic strategy to decolorize dye wastewater by using metal

octacarboxyphthalocyanine particles, J. Hazard. Mater. 380 (2019), 120842, https://doi.org/10.1016/j.jhazmat.2019.120842.
[6] K. He, G. Chen, G. Zeng, A. Chen, Z. Huang, J. Shi, T. Huang, M. Peng, L. Hu, Three-dimensional graphene supported catalysts for organic dyes degradation, Appl. Catal. B Environ. 228 (2018) 19–28, https://doi.org/10.1016/j.apcatb.2018.01.061.
[7] M.A. Ahmed, Z.M. Abou-Gamra, H.A.A. Medien, M.A. Hamza, Effect of porphyrin on photocatalytic activity of TiO_2 nanoparticles toward Rhodamine B photodegradation, J. Photochem. Photobiol. B Biol. 176 (2017) 25–35, https://doi.org/10.1016/j.jphotobiol.2017.09.016.
[8] F. Mashkoor, A. Nasar, Inamuddin, A.M. Asiri, Exploring the reusability of synthetically contaminated wastewater containing crystal violet dye using tectona grandis sawdust as a very low-cost adsorbent, Sci. Rep. 8 (2018) 1–16, https://doi.org/10.1038/s41598-018-26655-3.
[9] E.P. Kuncoro, D.R.M. Isnadina, H. Darmokoesoemo, O.R. Fauziah, H.S. Kusuma, Characterization, kinetic, and isotherm data for adsorption of Pb^{2+} from aqueous solution by adsorbent from mixture of bagasse-bentonite, Data Br 16 (2018) 622–629, https://doi.org/10.1016/J.DIB.2017.11.098.
[10] B. Ekka, G. Dhar, S. Sahu, M. Mishra, P. Dash, R.K. Patel, Removal of Cr(VI) by silica-titania core-shell nanocomposites: In vivo toxicity assessment of the adsorbent by Drosophila melanogaster, Ceram. Int. 47 (2021) 19079–19089, https://doi.org/10.1016/j.ceramint.2021.03.254.
[11] E.P. Kuncoro, T. Soedarti, T.W.C. Putranto, H. Darmokoesoemo, N.R. Abadi, H. S. Kusuma, Characterization of a mixture of algae waste-bentonite used as adsorbent for the removal of Pb^{2+} from aqueous solution, Data Br 16 (2018) 908–913, https://doi.org/10.1016/J.DIB.2017.12.030.
[12] H. Darmokoesoemo, Magdhalena, T.W.L.C. Putranto, H.S. Kusuma, Telescope snail (Telescopium sp) and mangrove crab (Scylla sp) as adsorbent for the removal of Pb^{2+} from aqueous solutions, Rasayan J. Chem. 9 (2016) 680–685.
[13] J.N. Naat, Y.A.B. Neolaka, T. Lapailaka, T. Rachmat Triandi, A. Sabarudin, H. Darmokoesoemo, H.S. Kusuma, Adsorption of Cu(II) and Pb(II) using silica@mercapto (hs@m) hybrid adsorbent synthesized from silica of Takari sand: Optimization of parameters and kinetics, Rasayan J. Chem. 14 (2021) 550–560, https://doi.org/10.31788/RJC.2021.1415803.
[14] Y.A.B. Neolaka, A.A.P. Riwu, U.O. Aigbe, K.E. Ukhurebor, R.B. Onyancha, H. Darmokoesoemo, H.S. Kusuma, Potential of activated carbon from various sources as a low-cost adsorbent to remove heavy metals and synthetic dyes, Results Chem 5 (2023), 100711, https://doi.org/10.1016/J.RECHEM.2022.100711.
[15] Y.A.B. Neolaka, Y. Lawa, J. Naat, A.C. Lalang, B.A. Widyaningrum, G.F. Ngasu, K. A. Niga, H. Darmokoesoemo, M. Iqbal, H.S. Kusuma, Adsorption of methyl red from aqueous solution using Bali cow bones (Bos javanicus domesticus) hydrochar powder, Results Eng 17 (2023), 100824, https://doi.org/10.1016/j.rineng.2022.100824.
[16] S.F. Anis, B.S. Lalia, R. Hashaikeh, N. Hilal, Breaking through the selectivity-permeability tradeoff using nano zeolite-Y for micellar enhanced ultrafiltration dye rejection application, Sep. Purif. Technol. 242 (2020), 116824, https://doi.org/10.1016/J.SEPPUR.2020.116824.
[17] S. Sahu, N. Bishoyi, M.K. Sahu, R.K. Patel, Investigating the selectivity and interference behavior for detoxification of Cr(VI) using lanthanum phosphate polyaniline nanocomposite via adsorption-reduction mechanism, Chemosphere 278 (2021), 130507, https://doi.org/10.1016/J.CHEMOSPHERE.2021.130507.
[18] D. Wang, S. Dong, H. Hu, Z. He, F. Dong, J. Tang, X. Lu, L. Wang, S. Song, J. Ma, Catalytic ozonation of atrazine with stable boron-doped graphene nanoparticles derived from waste polyvinyl alcohol film: Performance and mechanism, Chem. Eng. J. 455 (2023), 140316, https://doi.org/10.1016/J.CEJ.2022.140316.
[19] M. Usman, A. Ahmed, B. Yu, Q. Peng, Y. Shen, H. Cong, Photocatalytic potential of bio-engineered copper nanoparticles synthesized from Ficus carica extract for the degradation of toxic organic dye from wastewater: Growth mechanism and study of parameter affecting the degradation performance, Mater. Res. Bull. 120 (2019), 110583, https://doi.org/10.1016/J.MATERRESBULL.2019.110583.
[20] E.P. Kuncoro, D.R.M. Isnadina, H. Darmokoesoemo, F. Dzembarahmatiny, Characterization and isotherm data for adsorption of Cd^{2+} from aqueous solution by adsorbent from mixture of bagasse-bentonite, Data Br 16 (2018) 354–360, https://doi.org/10.1016/j.dib.2017.11.060.
[21] H. Darmokoesoemo, F.R. Setianingsih, T.W.L.C. Putranto, Horn snail (Telescopium sp) and mud crab (Scylla sp) shells powder as low cost adsorbents for removal of Cu^{2+} from synthetic wastewater, ASAYAN J. Chem. 9 (2016) 550–555.
[22] M. Jahandar Lashaki, Z. Hashisho, J.H. Phillips, D. Crompton, J.E. Anderson, M. Nichols, Mechanisms of heel buildup during cyclic adsorption-desorption of volatile organic compounds in a full-scale adsorber-desorber, Chem. Eng. J. 400 (2020), 124937, https://doi.org/10.1016/j.cej.2020.124937.
[23] T. Yin, X. Meng, S. Wang, X. Yao, N. Liu, L. Shi, Study on the adsorption of low-concentration VOCs on zeolite composites based on chemisorption of metal-oxides under dry and wet conditions, Sep. Purif. Technol. 280 (2022), 119634, https://doi.org/10.1016/j.seppur.2021.119634.
[24] S. Shojaei, A. Nouri, L. Baharinikoo, M. Davoodabadi Farahani, S. Shojaei, Removal of the hazardous dyes through adsorption over nanozeolite-X: Simultaneous model, design and analysis of experiments, Polyhedron 196 (2021), 114995, https://doi.org/10.1016/j.poly.2020.114995.
[25] S. Sivalingam, S. Sen, Rapid ultrasound assisted hydrothermal synthesis of highly pure nanozeolite X from fly ash for efficient treatment of industrial effluent, Chemosphere 210 (2018) 816–823, https://doi.org/10.1016/j.chemosphere.2018.07.091.
[26] J. Coronas, Present and future synthesis challenges for zeolites, Chem. Eng. J. 156 (2010) 236–242, https://doi.org/10.1016/j.cej.2009.11.006.
[27] M. Yoldi, E.G. Fuentes-Ordoñez, S.A. Korili, A. Gil, Zeolite synthesis from industrial wastes, Microporous Mesoporous Mater 287 (2019) 183–191, https://doi.org/10.1016/j.micromeso.2019.06.009.
[28] P. Peng, D. Stosic, X.-M. Liu, Z.-F. Yan, S. Mintova, Strategy towards enhanced performance of zeolite catalysts: Raising effective diffusion coefficient versus reducing diffusion length, Chem. Eng. J. 385 (2020), 123800, https://doi.org/10.1016/j.cej.2019.123800.
[29] L. Sellaoui, D. Franco, H. Ghalla, J. Georgin, M.S. Netto, G. Luiz Dotto, A. Bonilla-Petriciolet, H. Belmabrouk, A. Bajahzar, Insights of the adsorption mechanism of methylene blue on brazilian berries seeds: Experiments, phenomenological modelling and DFT calculations, Chem. Eng. J. 394 (2020), 125011, https://doi.org/10.1016/j.cej.2020.125011.
[30] M.S. Alivand, M. Najmi, N.H.M.H. Tehrani, A. Kamali, O. Tavakoli, A. Rashidi, M. D. Esrafili, E. Ghasemy, O. Mazaheri, Tuning the surface chemistry and porosity of waste-derived nanoporous materials toward exceptional performance in antibiotic adsorption: Experimental and DFT studies, Chem. Eng. J. 374 (2019) 274–291, https://doi.org/10.1016/j.cej.2019.05.188.
[31] L.R. Oviedo, P.C.L. Muraro, G. Pavoski, D.C.R. Espinosa, Y.P.M. Ruiz, A. Galembeck, C.R.B. Rhoden, W.L. da Silva, Synthesis and characterization of nanozeolite from (agro)industrial waste for application in heterogeneous photocatalysis, Environ. Sci. Pollut. Res. 29 (2022) 3794–3807, https://doi.org/10.1007/s11356-021-15815-0.
[32] R.A. Khera, M. Iqbal, A. Ahmad, S.M. Hassan, A. Nazir, A. Kausar, H.S. Kusuma, J. Niasr, N. Masood, U. Younas, R. Nawaz, M.I. Khan, Kinetics and equilibrium studies of copper, zinc, and nickel ions adsorptive removal on to Archontophoenix alexandrae: conditions optimization by RSM, Desalin. Water Treat. 201 (2020) 289–300, https://doi.org/10.5004/DWT.2020.25937.

Further reading

[33] A. El Bey, M. Laidi, A. Yettou, S. Hanini, A. Ibrir, M. Hentabli, H. Ouldkhaoua, Practical artificial neural network tool for predicting the competitive adsorption of dyes on gemini polymeric nanoarchitecture, Kem. u Ind. 70 (2021) 481–488, https://doi.org/10.15255/KUI.2020.069.
[34] C.E. Golden, M.J. Rothrock, A. Mishra, Comparison between random forest and gradient boosting machine methods for predicting Listeria spp. prevalence in the environment of pastured poultry farms, Food Res. Int. 122 (2019) 47–55, https://doi.org/10.1016/j.foodres.2019.03.062.
[35] J. Li, L. Pan, M. Suvarna, Y.W. Tong, X. Wang, Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning, Appl. Energy. 269 (2020), 115166, https://doi.org/10.1016/J.APENERGY.2020.115166.
[36] J. Li, X. Zhu, Y. Li, Y.W. Tong, Y.S. Ok, X. Wang, Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: Application of machine learning on waste-to-resource, J. Clean. Prod. 278 (2021), 123928, https://doi.org/10.1016/J.JCLEPRO.2020.123928.
[37] M.R. Gadekar, M.M. Ahammed, Modelling dye removal by adsorption onto water treatment residuals using combined response surface methodology-artificial neural network approach, J. Environ. Manage. 231 (2019) 241–248, https://doi.org/10.1016/j.jenvman.2018.10.017.
[38] L. Leng, L. Yang, X. Lei, W. Zhang, Z. Ai, Z. Yang, H. Zhan, J. Yang, X. Yuan, H. Peng, H. Li, Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass, Biochar 4 (2022) 63, https://doi.org/10.1007/s42773-022-00183-w.
[39] S.R. Islam, W. Eberle, S. Bundy, S.K. Ghafoor, Infusing domain knowledge in AI-based "black box" models for better explainability with application in bankruptcy prediction (2019). http://arxiv.org/abs/1905.11474.
[40] W. Cai, R. Wei, L. Xu, X. Ding, A method for modelling greenhouse temperature using gradient boost decision tree, Inf. Process. Agric. 9 (2022) 343–354, https://doi.org/10.1016/j.inpa.2021.08.004.
[41] M. El-Kammah, E. Elkhatib, S. Gouveia, C. Cameselle, E. Aboukila, Enhanced removal of Indigo Carmine dye from textile effluent using green cost-efficient nanomaterial: Adsorption, kinetics, thermodynamics and mechanisms, Sustain. Chem. Pharm. 29 (2022), 100753, https://doi.org/10.1016/j.scp.2022.100753.
[42] X. Yuan, M. Suvarna, S. Low, P.D. Dissanayake, K.B. Lee, J. Li, X. Wang, Y.S. Ok, Applied machine learning for prediction of CO_2 adsorption on biomass waste-derived porous carbons, Environ. Sci. Technol. 55 (2021) 11925–11936, https://doi.org/10.1021/acs.est.1c01849.
[43] M. Despotovic, V. Nedic, D. Despotovic, S. Cvetanovic, Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation, Renew. Sustain. Energy Rev. 56 (2016) 246–260, https://doi.org/10.1016/j.rser.2015.11.058.
[44] Z. Wang, J. Lv, Y. Tan, M. Guo, Y. Gu, S. Xu, Y. Zhou, Temporospatial variations and Spearman correlation analysis of ozone concentrations to nitrogen dioxide, sulfur dioxide, particulate matters and carbon monoxide in ambient air, China, Atmos. Pollut. Res. 10 (2019) 1203–1210, https://doi.org/10.1016/j.apr.2019.02.003.