

## Article

# Analysis and Prediction of the Leaching Process of Ionic Rare Earth: A Data Mining Study with Scarce Data

Zhenyue Zhang <sup>1,2</sup>, Jing Yang <sup>1</sup>, Wenda Guo <sup>1,3,\*</sup>, Ling Jiang <sup>1</sup>, Wendou Chen <sup>1</sup> , Defeng Liu <sup>1</sup> , Hanjun Wu <sup>4</sup>  and Ruan Chi <sup>1,2</sup>

<sup>1</sup> School of Resources and Safety Engineering, Xingfa School of Mining Engineering, Wuhan Institute of Technology, Wuhan 430073, China; zyzxmu@wit.edu.cn (Z.Z.); yangj090301@163.com (J.Y.); jll18711260641@163.com (L.J.); witcwd@163.com (W.C.); liudefengoop@126.com (D.L.); lljhalo777@163.com (R.C.)

<sup>2</sup> Key Laboratory for Green Chemical Process, Ministry of Education, Wuhan 430073, China

<sup>3</sup> Hubei Xingfa Chemicals Group Co., Ltd., Yichang 443000, China

<sup>4</sup> School of Chemistry and Environmental Engineering, Wuhan Institute of Technology, Wuhan 430205, China; wuhj1204@wit.edu.cn

\* Correspondence: gedaya123@163.com; Tel.: +86-13080853386

**Abstract:** To unveil the impact of each condition variable on the leaching efficiency index during the heap leaching process of rare earth ore and establish a prediction model for leaching conditions and efficiency, common parameters in the heap leaching process of rare earth ore were selected. In addition, the pilot-scale test data were collected over 50 days. Based on the collected data, the Ordinary Least Squares (OLS) linear regression method was used for fitting analysis to determine each variable's influence on the change in leaching efficiency. The results indicated a linear relationship between the flow rate of the leaching solution and leaching efficiency. In contrast, no obvious linear relationship was observed between other condition variables and leaching efficiency. Spearman's rank correlation coefficient was calculated to analyze the nonlinear correlation between the abovementioned variables and the leaching efficiency index. The correlation coefficients were found to be  $-0.78$ ,  $0.88$ ,  $-0.93$ ,  $-0.53$ ,  $0.71$ , and  $-0.93$  for ammonium content in the leaching agent, pH of the leaching agent, rare earth content, ammonium content in the leaching solution, pH of the leaching solution, and the flow rate of the leaching solution, respectively. This suggests that the flow rate of the leaching solution, rare earth content, and pH of the leaching agent significantly influence leaching efficiency, thus affecting the rare earth leaching efficiency index. Based on the correlation analysis results of leaching conditions and efficiency, a dataset with limited data trained by the common Ordinary Least Squares model, linear regression model, random forest model, and support vector machine regression model was selected to develop a prediction model for the leaching process data. The results indicated that the random forest model had the lowest mean square error of 7.47 among the four models and the coefficient of determination closest to 1 (0.99). This model can effectively analyze and predict condition variables' data and leaching efficiency index in the heap leaching process of rare earth ore, with a prediction accuracy exceeding 90%, thus providing intelligent guidance for the heap leaching process of rare earth ores.

**Keywords:** rare earth ore; heap leaching; data mining; process prediction



**Citation:** Zhang, Z.; Yang, J.; Guo, W.; Jiang, L.; Chen, W.; Liu, D.; Wu, H.; Chi, R. Analysis and Prediction of the Leaching Process of Ionic Rare Earth: A Data Mining Study with Scarce Data. *Minerals* **2024**, *14*, 596. <https://doi.org/10.3390/min14060596>

Academic Editor: Ilhwan Park

Received: 20 May 2024

Revised: 4 June 2024

Accepted: 5 June 2024

Published: 6 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Rare earth elements have been widely utilized in petroleum, chemical, aerospace, military, new permanent magnet material, and other high-tech fields, particularly the heavy rare earth elements found in ionic rare earth ores, which are categorized as national strategic resources [1–5]. In recent years, with the rapid advancement of science and technology, the global demand for medium and heavy rare earth elements (such as gadolinium, dysprosium, and yttrium) has been on the rise [6], and China's total reserves of medium and heavy

rare earth resources account for more than 80% of the world's total [7]. These resources are primarily distributed across eight southern provinces: Jiangxi, Fujian, Zhejiang, Hunan, Guangdong, Guangxi, Yunnan, and Guizhou. They possess the characteristics of wide distribution and low radioactivity and can be considered a “trump card” in the Sino–US trade war. With the frequent leaching hydrodynamics of weathered elution-deposited rare earth ores with ammonium salt solutions, the exploitation of ionic rare earth ores has garnered increasing attention from the state [8]. Since the discovery of ionic rare earth ores in the Zudong area of Longnan, Jiangxi Province, in 1969, its mining process has evolved through pool leaching, heap leaching, and in situ leaching [9]. Currently, compared to the other two processes, the in situ leaching process offers advantages such as minimal damage to mountains and vegetation, low labor intensity, and low production costs [10]. It is considered the most effective combined mining and beneficiation process and has been widely adopted [11]. However, this process primarily utilizes ammonium sulfate as the leaching agent, leading to the consumption of a significant amount of ammonium sulfate solution during leaching [12]. As a result, more ammonia nitrogen remains in the tailings, leading to excessive ammonia nitrogen levels in mine tailings, surface water near mining areas, and groundwater. This escalation in production costs and ammonia nitrogen pollution severely impacts the healthy development of the medium and heavy rare earth industry [13].

A substantial body of scholarly research has investigated the impact of various conditions on leaching effectiveness within the leaching process. For instance, Moldoveanu et al. [14] compared the leaching efficacy of several monovalent salt solutions under ambient conditions on weathered crust elution-deposited rare earth ores. Their findings revealed, in a descending order,  $\text{Cs}^+ > \text{NH}_4^+ > \text{Na}^+ > \text{Li}^+$ . Meanwhile, Ye et al. [15] employed an orthogonal design to ascertain the aqueous radon solubilities at different temperatures and salinities. They concurrently measured radon solubilities at various pH levels and explored the influence of temperature, salinity, and pH on the leaching agent's solubility in the leaching solution. Other researchers have investigated the dynamics relevant to the heap-leaching process of rare earth ores. For example, Tian et al. [16] analyzed the kinetics of leaching rare earth from weathered crust elution-deposited rare earth ores using an ammonium sulfate solution. They examined the influence of critical leaching parameters, such as leaching velocity and leaching agent concentration, on mass transfer. Similarly, Zhang et al. [17] examined the impact of regulating injection flow rates on the in situ leaching range. They developed a hydrodynamic model under eight different pumping and injection conditions, suggesting that optimizing injection well flow rates at various positions could effectively control the leaching range. Additionally, studies have focused on leach solutions. For example, Wang et al. [18] investigated interactions among clay particles in ion-adsorption-type rare earth ores within aqueous solutions. They elucidated the influence of solution properties on the percolation process of ion-type rare earth ores during leaching. Zeng et al. [19], on the other hand, prepared three surfactant leaching solutions of varying concentrations for agitation leaching experiments. They identified the optimal surfactant leaching solution to study the seepage characteristics of acid leaching solution affected by surfactants in the ore-bearing layer during in situ leaching. Lastly, Hu et al. [9] developed a new composite swelling inhibitor solution in different concentrations to address clay minerals' swelling issue upon contact with water. They determined the optimal solution system and the corresponding concentration. Research has also delved into the microscopic mechanisms underlying the leaching reaction process. For instance, Feng et al. [20] investigated the electrochemical properties generated by ion migration accompanied by charge movement during the in situ leaching of ion-adsorption-type rare earth ores. They explored the effects of chemical reaction rate and pore size on resistance parameters. Wang et al. [21] proposed the new understanding of the engineering property modification of weathered crust elution-deposited rare earth ores, as well as the underlying microscopic mechanisms. They revealed the internal processes and effects of ammonium sulfate leaching on weathered crust elution-deposited rare earth ores. Moreover, Gao

et al. [22] examined the impact of leaching agent concentration and pH on the stability of agglomeration of ion-adsorbed rare earth deposits. They analyzed surface zeta potential, double electric layer thickness, particle gradation, and pore structure to assess the different concentrations and pH levels of leaching solutions' impact on the stability of ore agglomerates. While single-factor condition experiments can identify the influence of individual factors on leaching efficiency, they cannot elucidate the correlation between these experiments and the final production or provide production guidance. Therefore, using continuous experimental data from pilot scale experiments, common parameters in the rare earth ore heap leaching process were selected. Modern intelligent technology analysis methods were adopted to analyze, excavate, and predict the data, guiding the final production effectively.

Python syntax is succinct and clear, facilitating ease of learning and usage. With robust data processing libraries such as Pandas, NumPy, and others, Python efficiently performs tasks including data cleaning, conversion, and aggregation [23]. Moreover, sophisticated data visualization tools like Matplotlib, Seaborn, and Plotly enable the creation of high-quality charts for visual representation. Additionally, with support from a diverse array of machine learning libraries like scikit-learn, TensorFlow, and PyTorch, Python facilitates the implementation of various machine learning algorithms and modeling tasks. Therefore, this paper adopts Python (version 3.7.0) as the primary language for data processing to conduct data analysis and prediction. Data preprocessing plays a crucial role in data analysis, encompassing tasks such as cleaning, integration, transformation, and normalization. Effective preprocessing aids in understanding data relationships, reducing errors, and enhancing the accuracy and reliability of prediction analysis results. Scatter plots visually represent data points on a rectangular coordinate system, illustrating the degree of interaction between variables. By examining the distribution of data points on a scatter plot, correlations between variables can be inferred. A random distribution of discrete points suggests no correlation, whereas a dense concentration of data points with a discernible trend indicates a certain correlation [24]. Therefore, following data preprocessing, this paper generates scatter plots and performs linear fitting to preliminarily ascertain the correlation between condition variables and leaching efficiency. Correlation analysis measures the relationship between variables, aiding in understanding the data's internal structure and guiding the selection of appropriate variables when constructing models. Prediction models focus on forecasting unknown data, providing insights into future trends, and aiding in decision making.

In response to the scarce available data on the heap leaching process of rare earth ores, this study utilized Python to address missing data, perform data cleansing, and execute data format conversion, thus completing data preprocessing. The processed data were visualized in the form of scatter plots to preliminarily assess the correlation between each condition variable and the leaching efficiency index. A linear regression model was employed to fit a straight line, verifying the linear relationship between the condition variables and the leaching efficiency index. In cases where linear regression assumptions were not met, the correlation coefficient for nonlinear relationships was calculated to identify key factors influencing leaching efficiency. Building upon correlation analysis among the condition variables, four common models were trained using multivariate datasets. The goodness of fit of each model was evaluated based on mean square error and coefficient of determination. Additionally, a comparison graph of predicted versus actual leaching efficiency was generated to facilitate data mining and the prediction of condition variables in the traditional rare earth ore heap leaching process, employing modern technical methods.

## 2. Data Collection for the Heap Leaching Process of Rare Earth Ore

Continuous industrial testing was conducted at the rare earth ore heap leaching site to collect and process data for the entire process. The heap leaching process diagram is depicted in Figure 1. This process comprises three main steps: liquid infusion, liquid

collection, and ammonium recycling. The process flow is as follows: initially, the ore pile is constructed, and the leaching agent solution is infused. Considering both cost and leaching efficiency, the leaching agent selected for this experiment was a complex ammonium salt composed of ammonium sulfate and ammonium chloride in a 3:1 ratio. Once the rare earth leaching solution is collected, rare earth recovery is achieved through impurity removal precipitation, and the precipitation mother liquor is collected. Subsequently, sulfuric acid is added to the precipitated mother liquor until the pH is adjusted to meet the leaching agent requirements for the late leaching agent solution. Upon the completion of the rare earth ore heap leaching operation, the tailings are washed with clean water, and the leaching liquid and mother liquor are combined to form a new leaching solution. Ammonium salt leaching agent is then added to meet the leaching agent requirements, facilitating the recycling and reuse of ammonium. Among these, the reaction process for leaching rare earth elements with ammonium salt is as Equation (1).

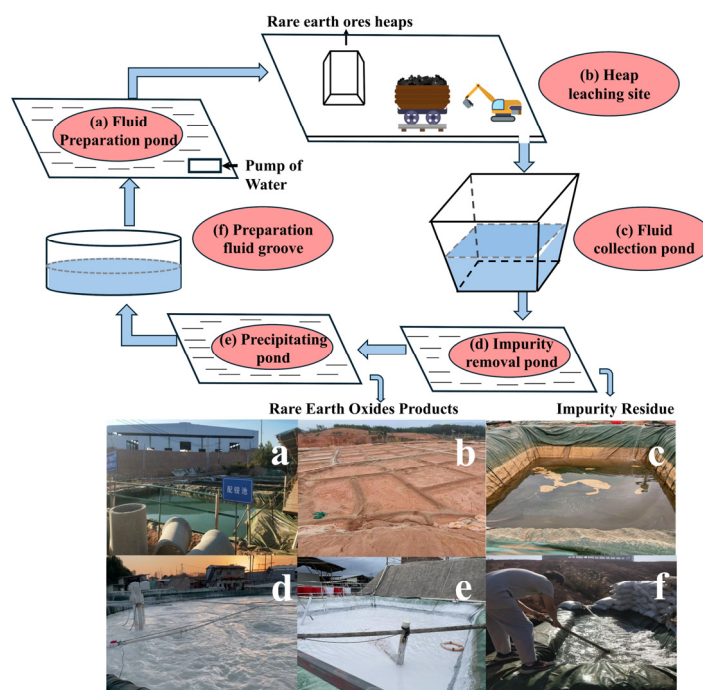
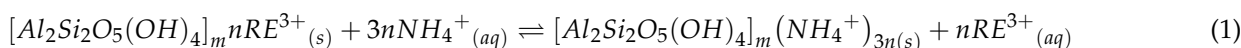
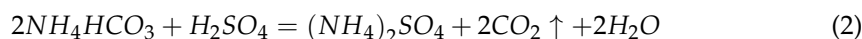


Figure 1. Data acquisition process.

Rare earth elements through this chemical reaction are exchanged in the solution to obtain the rare earth leaching solution. The rare earth elements are then recovered by precipitation using ammonium bicarbonate. The mother liquor, after the precipitation of rare earth elements, is treated with the sulfuric acid solution, and the chemical reaction is as Equation (2).



After collecting and sorting the continuous pilot scale test data obtained from the heap leaching site, point diagrams are drawn based on the leaching start time, as depicted in Figures 2–4.

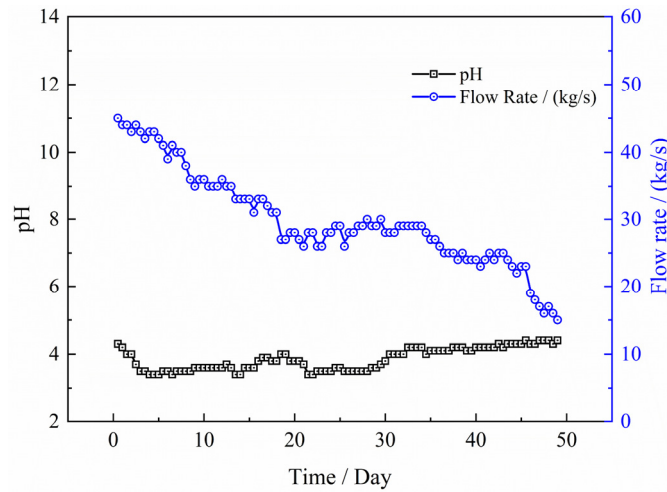


Figure 2. pH of leaching solution and flow rate.

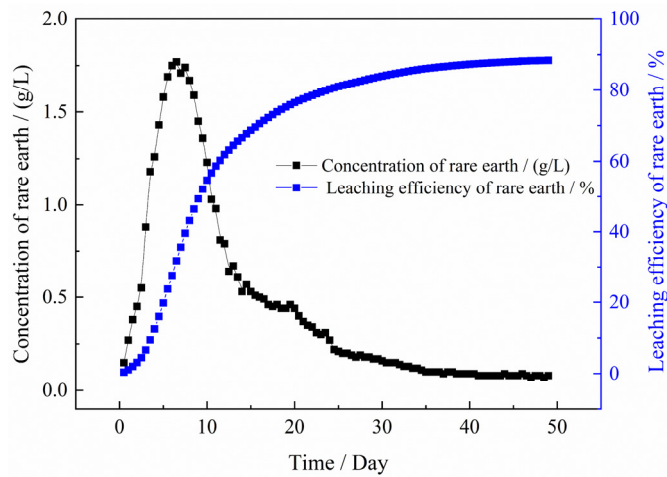


Figure 3. Concentration of rare earth and leaching efficiency.

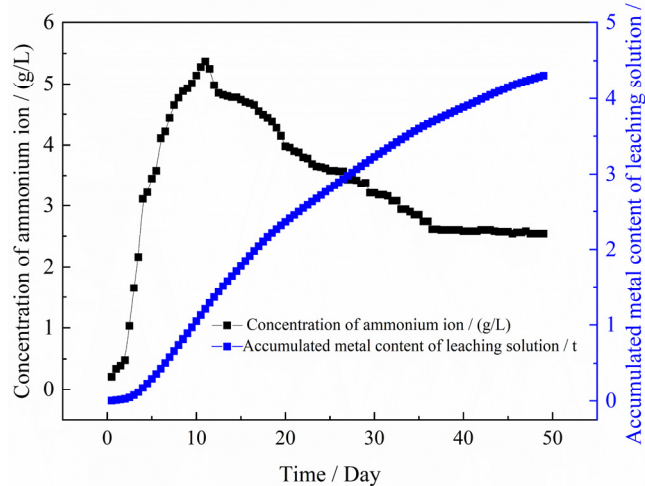


Figure 4. Concentration and content of ammonium in leaching solution.

As the leaching process progressed, the pH of the leaching solution consistently ranged between 3 and 5, indicating its acidic nature, as illustrated in Figure 2. However, it is noteworthy that, at the onset of the leaching process, the pH initially exceeded 5, indicating no rare earth leaching. This initial phase of the data is not represented in the

figure. At the beginning of leaching, the pH of the leaching solution is weakly acidic due to the presence of clay minerals in rare earth minerals, which specifically act as an acid–base buffer [25]. When an excessive amount of acidic leaching agent is added to the ore body, the pH of the leaching solution decreases and is maintained between 3 and 5. During this time, an exchange reaction occurs between the leaching agent cation and rare earth ion, resulting in the collection of a leaching solution containing rare earth. Moreover, the flow rate exhibited a gradual decline throughout the leaching process. From days 20 to 40, the trend was relatively steady with minor fluctuations, while towards the end of the leaching process, the flow rate diminished considerably. The leaching liquid flow rate is related to the amount of leaching agent solution. At the beginning of leaching, leaching agent solution is added according to a certain liquid–solid ratio, resulting in a high leaching liquid flow rate. After the leaching agent solution is added proportionally, the injection stops, leading to a gradual decrease in the flow rate of the leaching solution as the leaching cycle progresses.

As illustrated in Figure 3, as the heap leaching process progressed, there was a gradual enhancement in the leaching efficiency of rare earth. The growth rate was faster in the first 20 days, and the rare earth leaching efficiency reached 90.28% by the end of liquid collection. The concentration of rare earth increased rapidly in the first ten days, gradually decreased after reaching the peak, and stabilized after 25 days. This occurs because, during the initial stage of the ion exchange reaction, a substantial amount of rare earth ions is released from the mineral surface or pores and enter the solution, leading to a rapid increase in the rare earth leaching rate and the concentration of rare earth in the solution. Adjusting the pH value of the solution is necessary to promote the release of rare earth ions. Under acidic conditions, the release of rare earth ions is more efficient [26]; hence, the leaching solution is typically acidic. As the ion exchange reaction begins, rare earth ions undergo migration within the solution, a phenomenon influenced by various factors including solution concentration, temperature, pH value, and other variables.

As depicted in Figure 4, during the 50-day leaching period, the concentration of ammonium ions peaked on the 10th day of the leaching process, followed by a gradual decrease, ultimately stabilizing towards the end. Upon the completion of leaching, the accumulated metal content in the leaching solution reached 4.44 tons. The content of ammonium ions is related to the amount of leaching agent. With the addition of a leaching agent solution, some ammonium ions and rare earth ions are adsorbed by clay minerals in a large amount of ammonium ion leached ore body. Most ammonium ions percolate into the solution and finally reach the outlet. During the injection period, most of the ammonium ions were collected simultaneously with the rare earth leaching solution, and the ammonium ions in the leaching solution increased rapidly. When the liquid injection is stopped, the residual ammonium ions in the ore body are collected after percolating with the solution, and the ammonium ions gradually decrease.

To sum up, exploring the relationship between leaching process parameters and rare earth leaching plays a beneficial guiding role in efficiently extracting rare earth ions and optimizing mine production.

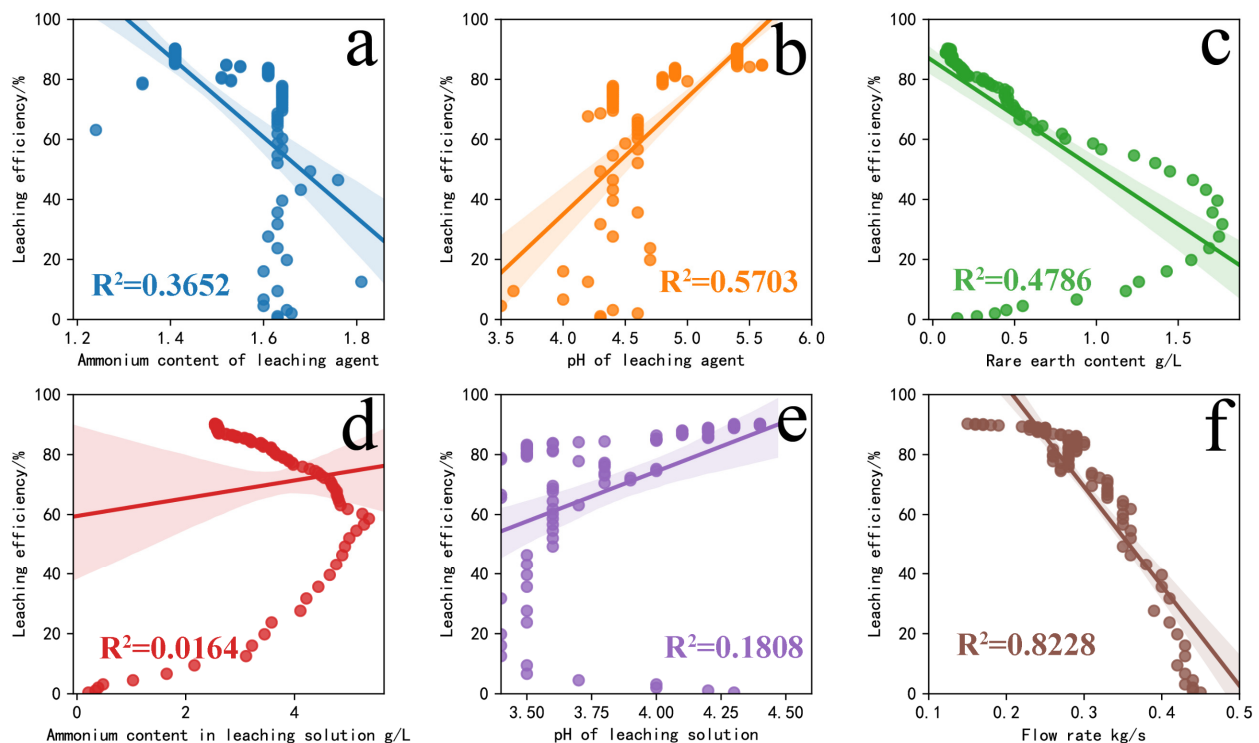
### 3. Analysis of Correlation between Leaching Condition Variables and Leaching Efficiency

The original data from the rare earth ore heap leaching process were manually collected from pilot scale tests. However, several issues such as omissions, losses, and incomplete record preservation arose. To mitigate potential data loss, anomalies, redundancies, and other factors that could compromise the accuracy of the analysis, the collected data underwent preprocessing [27–33]. In the collected dataset, less than 3% of the data groups exhibited a loss of more than 80% of the values for each condition variable on the same day, representing a negligible proportion of the overall raw data. Consequently, these data groups were directly discarded. Regarding duplicate data, owing to the nature of heap leaching tests, instances occurred where the values of conditional variables were identical.

Therefore, this portion of the data was retained to uphold the integrity of the dataset and ensure the accuracy of model construction. Subsequently, variables were selected and determined based on the rare earth ore heap leaching data, and feature extraction of conditional variables was completed. The data value range was established, the data type was modified, and the data were formatted to facilitate easy readability and analysis using code. Additionally, a multidimensional dataset was created to furnish a more comprehensive and accurate foundation for modeling and visualization. Through the steps of data cleaning, conversion, and integration, we effectively preprocessed the data, guaranteeing its quality and reliability, and strengthening the accuracy and stability of our data mining and prediction models [34,35].

### 3.1. Linear Correlation Analysis between Condition Variables and Leaching Efficiency

The Python language is used for generating a scatter plot, attempting to illustrate the linear correlation between each condition variable and the leaching efficiency, as shown in Figure 5. Figure 5a–f, respectively, depict scatter diagrams representing the relationship between leaching agent ammonium content and leaching efficiency, leaching agent pH and leaching efficiency, rare earth content and leaching efficiency, leaching solution ammonium content and leaching efficiency, leaching solution pH and leaching efficiency, and flow rate and leaching efficiency. In Figure 5a–c, noticeable patterns emerge in the scatter plots, suggesting a correlation between the variables, although not all scattered points are directly related. However, different variable values correspond to the same leaching efficiency, indicating a one-to-many variable relationship in the two-dimensional scatter plot. Figure 5f demonstrates that the flow rate and leaching efficiency exhibit an approximate monotone linear relationship. Conversely, there is no obvious monotone linear relationship between the other condition variables and leaching efficiency. To further investigate the linear relationship, the linear fit between each condition variable and leaching efficiency was visualized.



**Figure 5.** Linear regression fitting diagram between leaching efficiency and (a) ammonium content of the leaching agent, (b) pH of the leaching agent, (c) rare earth content, (d) ammonium content in the leaching solution, (e) pH of the leaching solution, and (f) flow rate.

The linear regression fitting line between each condition variable and the leaching efficiency index was generated using the OLS algorithm, as displayed in Figure 5 alongside the scatter plot. OLS was utilized to fit linear regression models. The least squares method is a widely used parameter estimation technique. Given independent and dependent variables, OLS identifies an optimal line that minimizes the sum of squares of the residuals (the differences between actual and predicted values) for all data points. This method is commonly employed in fitting linear regression models and other parametric models [36–38].

As depicted in the linear regression fitting line diagram in Figure 5, the visual scatter plot of linear regression provides a more intuitive means to assess whether a monotonic linear correlation exists between each condition variable and leaching efficiency. The shaded area in the figure is a 95% confidence interval, that is, the interval constructed according to the 95% accuracy and accuracy of the samples extracted from the population, which is used as an estimate of the range of the truth value of the distribution parameters of the population. Also, the  $R^2$  of each line is added to each line. Consistent with the initial observation from the scatter plot, only the flow rate and leaching efficiency exhibited a well-fitted straight line, with data points primarily aligning along the straight line and within the error range, indicating a strong negative linear correlation. However, the other variables did not demonstrate a clear linear correlation with leaching efficiency, resulting in a small  $R^2$  value. To further evaluate the strength of correlation among the condition variables, excluding the flow rate of the leaching solution, and leaching efficiency, we selected the correlation coefficient as a numerical indicator to assess the extent of influence of each variable on the leaching efficiency outcomes during the heap leaching process of rare earth ore, thereby quantifying the degree of correlation.

### 3.2. Nonlinear Correlation Analysis between Condition Variables and Leaching Efficiency

The correlation coefficient is utilized to assess the strength of the correlation between variables expressed numerically. Given the various correlation coefficient types available, they can accommodate both linear and nonlinear correlation relationships [39,40]. In the previous analysis, we did not observe a satisfactory linear fitting effect between the condition variables of the heap leaching process and the leaching efficiency index. Therefore, we opted to employ the correlation coefficient to more intuitively demonstrate the correlation strength between independent and dependent variables. Subsequently, we computed the correlation coefficient for each condition variable and the leaching efficiency index, identifying the main control variable with the strongest correlation to leaching efficiency. This analysis provided valuable guidance for data analysis and prediction within the heap leaching process.

The commonly used correlation coefficients include the Pearson correlation coefficient, Spearman's rank correlation coefficient, and Kendall's rank correlation coefficient. The value of the correlation coefficient ranges from  $-1$  to  $1$ , where  $1$  signifies a perfect positive correlation, and  $-1$  signifies a perfect negative correlation.

Pearson's correlation coefficient, which measures the strength and direction of the linear relationship between two continuous variables, is sensitive to outliers and is often utilized to gauge the linear correlation between two continuous variables. The calculation formula is given by Equation (3), where  $r$  represents the Pearson's correlation coefficient, and  $X$  and  $Y$  denote the two variable values to be calculated.

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \quad (3)$$

Spearman's rank correlation coefficient is a non-parametric method that measures a monotonic relationship between two variables, rather than a strictly linear relationship. It calculates correlations by comparing the ranks of variables and applies to any monotonic relationship, not just linear relationships. The calculation formula is given by Equation (4).

In this formula,  $r_s$  is Spearman's rank correlation coefficient,  $d_i$  represents the difference of the variables after ordering, and  $n$  represents the number of samples.

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (4)$$

Kendall's rank correlation coefficient is also a non-parametric method, similar to Spearman's rank correlation coefficient, as it is calculated based on the rank of the variable. Kendall's rank correlation coefficient is more robust for cases of outliers and small sample sizes in the data. Unlike Spearman's rank correlation coefficient, Kendall's rank correlation coefficient is less sensitive to rank correlations, so in some cases, it may yield different correlation results. The calculation formula is given by Equation (5), where  $\tau$  is Kendall's rank correlation coefficient,  $C$  represents the number of consistent pairs,  $D$  signifies the number of inconsistent pairs, and  $n$  denotes the number of samples.

$$\tau = \frac{C - D}{\frac{1}{2}n(n - 1)} \quad (5)$$

When calculating the correlation coefficient, we can determine whether it is linear or based on the distribution and trend of the scatter plot diagram and select the appropriate correlation coefficient accordingly. Due to the one-to-many relationship and the small amount of data in the heap leaching dataset after preprocessing, there was no linear fitting relationship between the other conditional variables except for the flow rate of the leaching solution and the leaching efficiency. Therefore, Spearman's rank correlation coefficient was used to assess the strength of the nonlinear relationship between the conditional variables and the leaching efficiency, aiming to improve the accuracy of correlation assessment. Spearman's rank correlation coefficient was employed to calculate the nonlinear correlation between the conditional variables and the leaching efficiency. The final results of the nonlinear correlation between the two-dimensional variables in the rare earth heap leaching process were presented as a heat map.

The nonlinear correlation coefficient thermal map provides the correlation coefficient for each pair of variables. This study focuses mainly on identifying the main control variable with the greatest influence on the leaching efficiency index, concentrating on data in the last row or column. In Figure 6, the red series denotes positive correlations, while the blue series represents negative ones. The stronger the correlation, the closer the absolute value of the correlation coefficient is to 1. From Figure 6, it is evident that the flow rate and rare earth content of the leaching solution exhibit a strong negative correlation with leaching efficiency, with a correlation coefficient of  $-0.93$ . Moreover, the thermal diagram shows a strong correlation coefficient of  $0.85$  between flow rate and rare earth content. This suggests a significant relationship between these variables. It is inferred that the correlation between flow rate and rare earth content may indicate the two primary controlling variables of leaching efficiency. The correlation coefficient of ammonium content in the leaching solution is only  $-0.53$ , a weaker correlation than other conditional variables. Consequently, it is believed that the correlation between ammonium content in the leaching solution and leaching efficiency is relatively weak. Specifically, the correlation between each conditional variable and the index of leaching efficiency was ranked as follows: flow rate = rare earth content ( $-0.93$ ) > leaching agent pH ( $0.88$ ) > leaching agent ammonium content ( $-0.78$ ) > leaching solution pH ( $0.71$ ) > leaching solution ammonium content ( $-0.53$ ). In other words, the conditional variables guiding the leaching efficiency index are ranked from strongest to weakest correlation as follows: flow rate, rare earth content, pH of leaching agent, ammonium content of leaching agent, pH of leaching solution, and ammonium content in leaching solution. The data prediction model for the heap leaching process of rare earth ores can be constructed using as few condition variables as possible, focusing on control variables with strong correlations.

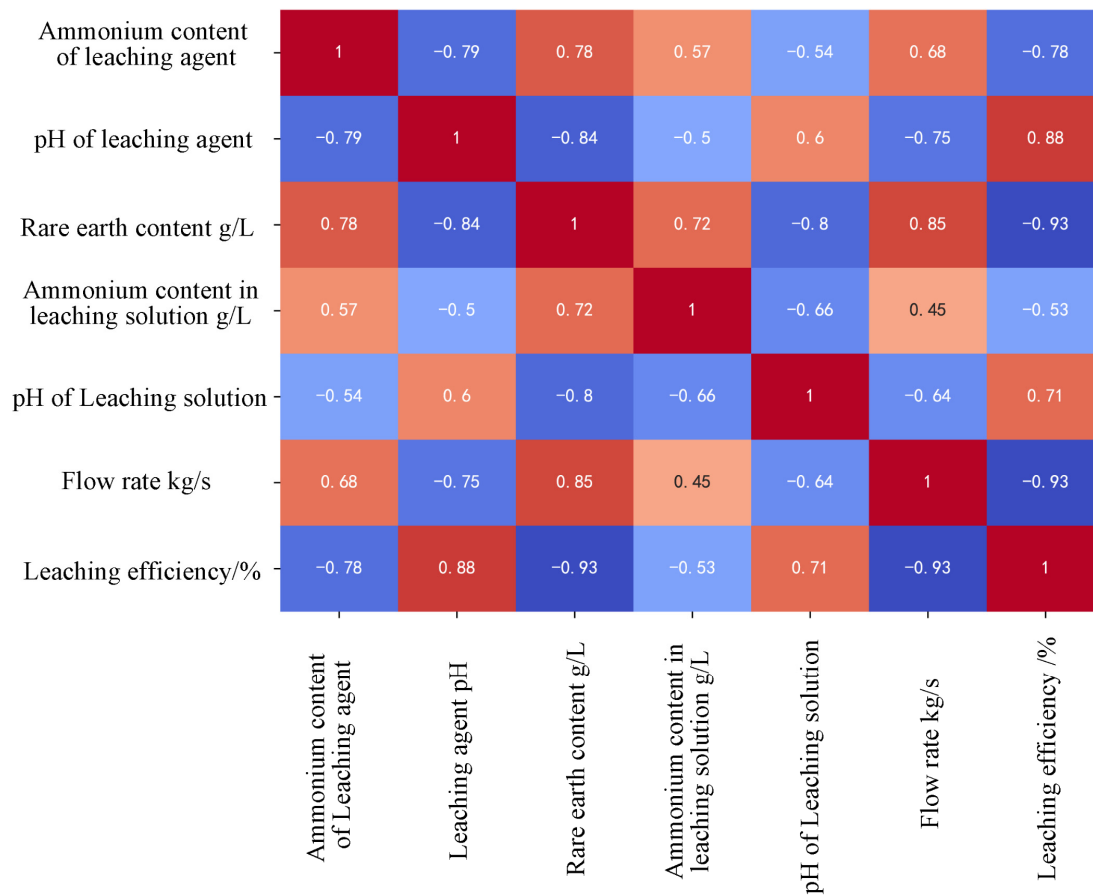


Figure 6. Correlation coefficient thermal map.

#### 4. Study on Data Prediction Model of Heap Leaching Process of Rare Earth Ores

After examining the influence of each conditional variable on the leaching efficiency index, it was discovered that the correlation coefficient between each conditional variable and the leaching efficiency in the heap leaching process exceeded 0.5. Considering that the leaching efficiency index is affected by multiple factors and results from synergistic action, multi-variable co-fitting training was chosen for predicting the leaching efficiency data in the heap leaching process of rare earth ore to achieve optimal results. Regarding model selection, four common models were chosen: linear regression model, OLS model, random forest model, and support vector machine regression (SVR) model.

Linear regression is a statistical model for establishing a linear relationship between dependent and independent variables. This model is often employed to predict the value of the dependent variable or to analyze the influence of the independent variables on the dependent variable [41]. Its general form is represented by Formula (6).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \tag{6}$$

where Y is the dependent variable (also called the response variable or the target variable),  $X_1, X_2, \dots, X_n$  are the independent variables (also called explanatory variables or predictor variables),  $\beta_0$  is the intercept (constant term),  $\beta_1, \beta_2, \dots, \beta_n$  are the coefficients of the independent variables, indicating the influence of each independent variable on the dependent variable, and  $\epsilon$  is the error term (residual), which represents the factors that the model fails to account for.

The OLS model is also a kind of linear regression model. In the OLS model, the optimal coefficients are estimated by minimizing the residuals sum of squares  $\beta_1, \beta_2, \dots$ ,

$\beta_n$  [42]. Specifically, the OLS model is implemented by minimizing the loss functions as represented by Formula (7).

$$\text{minimize} \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}))^2 \quad (7)$$

where  $Y_i$  represents the  $i$ -th observation, and  $X_{i1}, X_{i2}, \dots, X_{in}$  are the corresponding argument values. Once an optimal coefficient estimate is obtained, the model can be used for predictive and inferential analyses. The OLS model typically assumes that the residuals satisfy some basic assumptions, such as a mean of 0, homogeneity of variance, and independent homoscedasticity.

Random forest is an ensemble learning model based on decision trees. Its core algorithms consist of decision trees and ensemble learning methods, commonly employed for classification and regression tasks. Random forest comprises multiple decision trees, each functioning as a weak learner that contributes to overall predictions through voting or averaging. The random forest model features decision tree composition, random feature selection, voting method selection, determination of average values, and strong robustness [43]. It is worth noting that the random forest model has parameters that require tuning, such as the number of trees, maximum depth, and method of random feature selection. Hence, for optimal model performance when employing the random forest model, it is essential to fine-tune its parameters.

SVM is commonly utilized for classification problems but can also be applied to regression problems; in such cases, it is referred to as support vector machine regression (SVR) [44]. The objective of SVR is to find a function that fits the data with minimal error while maintaining conciseness. Unlike traditional linear regression models, SVR employs an approach called the “kernel trick,” enabling data to be mapped into a high-dimensional feature space to handle nonlinear relationships more effectively. SVR also requires the adjustment of hyperparameters, such as the selection of the kernel function (e.g., linear, polynomial, or Gaussian kernels) and the regularization parameter  $C$ , to achieve optimal model performance.

Mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R-squared) are pivotal evaluation metrics for assessing prediction model fitting. These metrics are calculated using the predicted values from the trained model and the actual dependent variable values to evaluate the goodness of fit of the model. To validate the effectiveness of the prediction model for rare earth ore heap leaching process data, this study divides the preprocessed dataset into two parts: one for training the model and the other for validation. Utilizing MSE and R-squared, models with high degrees of fitting are selected with both intuition and accuracy. The original leaching efficiency index and predicted leaching efficiency data are concurrently output to represent the predicted results visually.

Mean square error (MSE) and coefficient of determination ( $R^2$ ) are commonly used to evaluate regression models, measuring the goodness of fit. MSE quantifies the average error between the predicted and actual values of the model, while  $R^2$  indicates the model's explanatory power for variable changes. The formula for calculating MSE is as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

where  $n$  represents the number of samples,  $y_i$  denotes the actual value, and  $\hat{y}_i$  signifies the model's predicted value. The smaller the value of MSE, the better the model fits [45,46].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

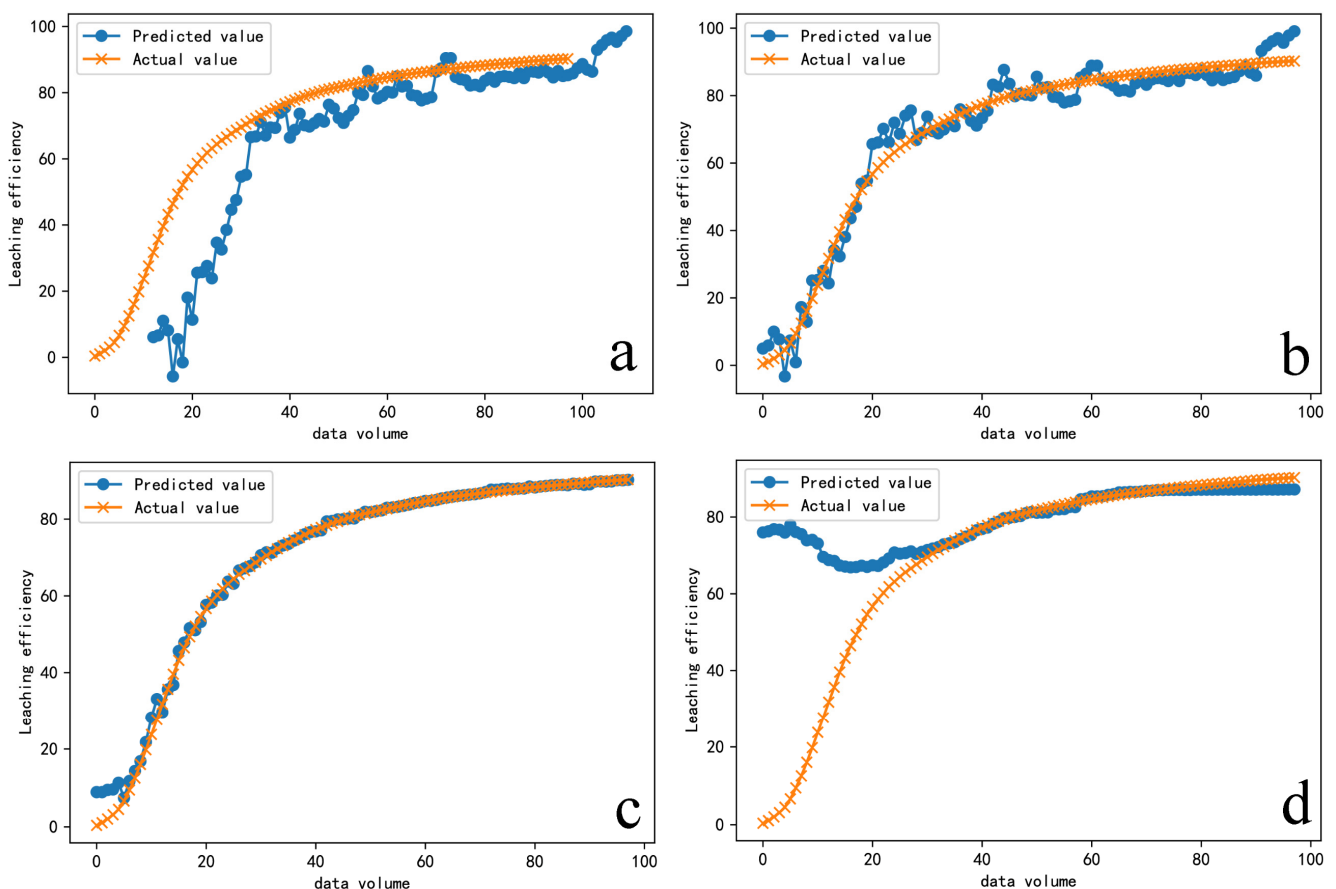
where  $y_i$  represents the real value,  $\hat{y}_i$  represents the predicted value of the model, and  $\bar{y}$  stands for the mean of the dependent variable.  $R^2$  ranges from 0 to 1, and the closer the value is to 1, the better the model fits the data. When  $R^2$  is 1, the model perfectly fits the data [47,48].

The MSE and  $R^2$  values of the OLS model, linear regression model, random forest model, and SVR model are presented in Table 1.

**Table 1.** Four common models’ mean square error and coefficient of determination.

Model	MSE	$R^2$
Ordinary Least Squares	27.512530	0.965140
Linear regression	25.441179	0.967764
Random forest	7.469664	0.990535
Support vector machine regression	748.356873	0.051775

When selecting a model, opt for the one with the lowest MSE and the  $R^2$  closest to 1. Upon examining the evaluation indicators of model predictions, the MSE values, from smallest to largest, were as follows: random forest model, linear regression model, OLS model, and SVR model. Similarly, the  $R^2$  values followed the same order: random forest model, linear regression model, OLS model, and SVR model. The MSE of the random forest model is 7.47, and the  $R^2$  is 0.99, indicating that the random forest model provides the best prediction. Figure 7 illustrates the actual leaching efficiency values and the predicted leaching efficiency index of the four common models.



**Figure 7.** Comparison between actual leaching efficiency data and predicted leaching efficiency curve of (a) OLS model, (b) linear regression model, (c) random forest model, and (d) support vector machine regression model.

As depicted in Figure 7, panels a, b, c, and d compare the actual leaching efficiency values with the predicted leaching efficiency indices of the OLS model, linear regression model, random forest model, and SVR model, respectively. In each panel, the horizontal axis represents the number of data points involved in verification, while the vertical axis represents the leaching efficiency value. The blue line represents the index predicted by the model after training, and the orange line represents the actual value. It is evident from the figure that the random forest model exhibits the best-fitting degree, consistent with the analysis of evaluation indicators such as MSE and  $R^2$ . The OLS model is a type of linear regression model. It is speculated that the fitting degree of the two models is poor because most of the conditional variables do not satisfy a linear relationship with the leaching efficiency. The latter half of the SVM model fits well, and it is speculated that a certain amount of data is needed to correct the parameter values in the parameter adjustment. The trained prediction model can not only fill in the unrecorded variable values in actual production but also predict the final leaching efficiency index based on the input conditional variable values to guide the actual production process.

## 5. Conclusions

Test data collected from a continuous pilot scale heap leaching process of rare earth ore underwent a series of preprocessing steps, including cleaning, filling, and conversion. Scatter plots were generated to illustrate the relationship between each condition variable and the leaching efficiency index. Linear regression models using OLS were fitted to the data, revealing non-monotonic linear correlations between the condition variables and leaching efficiency, except for the flow rate of the leaching solution. A nonlinear correlation analysis based on Spearman's rank correlation coefficient thermal map indicated significant differences in the nonlinear relationships between each condition variable and leaching efficiency. Notably, the correlation coefficients of rare earth content and flow rate reached  $-0.93$ , indicating a strong negative correlation with leaching efficiency. The absolute values of the correlation coefficients between each condition variable and the leaching efficiency index were above 0.5, suggesting a certain correlation between each conditional variable and leaching efficiency. Comparative analysis of Ordinary Least Squares models, linear regression models, random forest models, and support vector machine regression models for data fitting and prediction of rare earth ore heap leaching processes revealed that the random forest model had the smallest mean square error among the four common models, at 7.47, and a coefficient of determination closest to 1, at 0.99, making it the most suitable for predicting rare earth ore heap leaching processes. Visualization comparisons of the prediction models for heap leaching processes indicated that the prediction accuracy of the random forest model could exceed 90%. These research findings offer fundamental insights into data analysis and exploration within the rare earth ore leaching industry, as well as the advancement of predictive models for production process indices.

**Author Contributions:** Conceptualization, Z.Z. and W.G.; methodology, L.J.; software, J.Y.; validation, J.Y., L.J. and W.C.; investigation, W.C.; resources, Z.Z. and R.C.; data curation, H.W.; writing—original draft preparation, J.Y.; writing—review and editing, W.G. and W.C.; visualization, J.Y., L.J. and W.C.; supervision, Z.Z. and W.G.; project administration, D.L. and R.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (2021YFC2902604).

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** Wenda Guo is an employee of Hubei Xingfa Chemicals Group Co., Ltd. The authors declare no conflicts of interest.

## References

1. Judge, W.D.; Azimi, G. Recent progress in impurity removal during rare earth element processing: A review. *Hydrometallurgy* **2020**, *196*, 105436. [[CrossRef](#)]
2. Krishna, R.; Dhass, A.D.; Arya, A.; Prasad, R.; Colak, I. An assessment of the strategies for the energy critical elements necessary for the development of sustainable energy sources. *Environ. Sci. Pollut. Res.* **2023**, *30*, 90276–90297. [[CrossRef](#)]
3. Coey, J.M.D. Perspective and prospects for rare earth permanent magnets. *Engineering* **2020**, *6*, 119–131. [[CrossRef](#)]
4. Liu, J.Y.; Kong, X.; Zheng, L.R.; Guo, X.; Liu, X.F.; Shuai, J.L. Rare earth single-atom catalysts for nitrogen and carbon dioxide reduction. *ACS Nano* **2020**, *14*, 1093–1101. [[CrossRef](#)]
5. Saraci, F.; Quezada-Novoa, V.; Donnarumma, P.R.; Howarth, A.J. Rare-earth metal-organic frameworks: From structure to applications. *Chem. Soc. Rev.* **2020**, *49*, 7974–7977. [[CrossRef](#)]
6. Hou, X.Z.; Yang, Z.F.; Wang, Z.J. The occurrence characteristics and recovery potential of middle-heavy rare earth elements in the Bayan Obo deposit, Northern China. *Ore. Geol. Rev.* **2020**, *126*, 103737. [[CrossRef](#)]
7. He, Z.; Zhang, Z.; Chi, R.; Xu, Z.; Yu, J.; Wu, M.; Bai, R. Leaching hydrodynamics of weathered elution-deposited rare earth ore with ammonium salts solution. *J. Rare Earths* **2017**, *35*, 824–830. [[CrossRef](#)]
8. Luo, X.P.; Zhang, Y.B.; Zhou, H.P.; He, K.Z.; Luo, C.G.; Liu, Z.S.; Tang, X.K. Review on the development and utilization of ionic rare earth ore. *Minerals* **2022**, *12*, 554. [[CrossRef](#)]
9. Hu, Q.; Xu, Y.L.; Deng, X.Y.; Hu, S.M.; Xu, J.Y.; Zhou, F.; Chi, R.A. Effect of Potassium Salt on Swelling of Halloysite Clay Mineral during Leaching Process of Ionic Rare Earth Ore. *Minerals* **2023**, *13*, 906. [[CrossRef](#)]
10. Xu, Z.G.; Li, G.; Yang, H.F.; Sha, A.Y.; He, Z.Y.; Tang, Y.C.; Wu, M.; Qu, J. Development Review on Leaching Technology and Leaching Agents of Weathered Crust Elution-Deposited Rare Earth Ores. *Minerals* **2023**, *13*, 1223. [[CrossRef](#)]
11. Liu, D.F.; Yan, W.X.; Zhang, Z.Y.; Chi, R.A. Effect of particle gradation on pore structure and seepage law of solution in weathered crust elution-deposited rare earth ores. *Int. J. Min. Sci. Technol.* **2023**, *33*, 1261–1272. [[CrossRef](#)]
12. Yan, H.S.; Liang, T.M.; Liu, Q.S.; Qiu, T.S.; Ai, G.H. Compound leaching behavior and regularity of ionic rare earth ore. *Powder Technol.* **2018**, *333*, 106–114. [[CrossRef](#)]
13. Ou, X.L.; Chen, Z.B.; Chen, X.L.; Li, X.F.; Wang, J.; Ren, T.J.; Chen, H.B.; Feng, L.J.; Wang, Y.K.; Chen, Z.Q.; et al. Redistribution and chemical speciation of rare earth elements in an ion-adsorption rare earth tailing, Southern China. *Sci. Total Environ.* **2022**, *821*, 153369. [[CrossRef](#)]
14. Georgiana, A.M.; Vladimirov, G.P. Recovery of rare earth elements adsorbed on clay minerals: I. Desorption mechanism. *Hydrometallurgy* **2012**, *117*, 71–78. [[CrossRef](#)]
15. Ye, Y.J.; Xia, X.Q.; Dai, X.T.; Huang, C.H.; Guo, Q. Effects of temperature, salinity, and pH on <sup>222</sup>Rn solubility in water. *J. Radioanal. Nucl. Chem.* **2019**, *320*, 369–375. [[CrossRef](#)]
16. Tian, J.; Yin, J.Q.; Chi, R.A.; Rao, G.H.; Jiang, M.T.; Ouyang, K.X. Kinetics on leaching rare earth from the weathered crust elution-deposited rare earth ores with ammonium sulfate solution. *Hydrometallurgy* **2010**, *101*, 166–170. [[CrossRef](#)]
17. Zhang, C.; Xie, T.T.; Tan, K.X.; Yao, Y.X.; Wang, Y.A.; Li, C.G.; Li, Y.M.; Zhang, Y.; Wang, H. Hydrodynamic Simulation of the Influence of Injection Flowrate Regulation on In-Situ Leaching Range. *Minerals* **2022**, *12*, 787. [[CrossRef](#)]
18. Wang, L.; Li, L.; Li, K.Z.; Huang, L. Influence of clay particle interactions on leaching percolation in Ion-adsorption-type rare earth ores. *Minerals* **2024**, *14*, 140. [[CrossRef](#)]
19. Zeng, S.; Sun, J.X.; Sun, B.; Chen, S.T.; Zhao, J.T.; Yang, L.Q. Experimental investigation on the influence of surfactant to the seepage characteristics of acid leaching solution during in-situ leaching of uranium. *J. Radioanal. Nucl. Chem.* **2023**, *332*, 3651–3660. [[CrossRef](#)]
20. Feng, X.J.; Wang, X.Q. Characteristics of electrical resistance alteration during in situ leaching of Ion-adsorption-type rare earth ore. *Minerals* **2024**, *14*, 92. [[CrossRef](#)]
21. Wang, H.; Wang, X.J.; Wang, Y.; Wang, D.H.; Hu, K.J.; Zhong, W.; Guo, Z.Q. Influence of ammonium sulfate leaching agent on engineering properties of weathered crust elution-deposited rare earth ore. *Acta Geotech.* **2023**, *19*, 2041–2062. [[CrossRef](#)]
22. Gao, Z.Q.; Rao, Y.Z.; Shi, L.; Zhang, X.M.; Xiang, R. Investigation of the Impact of Leaching Agent Concentration and pH on the Stability of Agglomeration of Ion-Absorbed Rare Earth Deposits. *Geofluids* **2023**, *2023*, 6333122. [[CrossRef](#)]
23. Thielen, K.; Tien, V. Python and Physical Modeling. *Comput. Sci. Eng.* **2016**, *18*, 8–10. [[CrossRef](#)]
24. Hadzi-Pavlovic, D. Correlations I: Variation in r across subsets of the data. *Acta Neuropsychiatr.* **2007**, *19*, 66–67. [[CrossRef](#)]
25. Wu, X.Y.; Feng, J.; Zhou, F.; Liu, C.F.; Chi, R.A. Optimization of a Rare Earth and Aluminum Leaching Process from Weathered Crust Elution-Deposited Rare Earth Ore with Surfactant CTAB. *Minerals* **2024**, *14*, 321. [[CrossRef](#)]
26. Chen, W.D.; Zhang, Z.Y.; Long, F.; Chen, Z.; Chi, R.A. Rare Earth Occurrence States of Weathered Crust Elution-Deposited Rare Earth Ores in Southern Yunnan. *Minerals* **2023**, *13*, 554. [[CrossRef](#)]
27. Kristin, L.S. The Value of Scatter Plots. *PMR* **2016**, *8*, 1213–1217. [[CrossRef](#)]
28. Joshi, M.A.P. Data Preprocessing The techniques for preparing clean and quality data for data analytics process. *Orient. J. Comput. Sci. Technol.* **2020**, *13*, 78–81. [[CrossRef](#)]
29. Martine, C.; Joseph, D.M. A Data Cleaning Solution by Perl Scripts for the KDD Cup 2003 Task 2. *Acm Sigkdd Explor. Newsl.* **2003**, *5*, 158–159. [[CrossRef](#)]
30. Mellado, M.; Cisternas, L.; Lucay, F.; Gálvez, E.; Sepúlveda, F. A Posteriori Analysis of Analytical Models for Heap Leaching Using Uncertainty and Global Sensitivity Analyses. *Minerals* **2018**, *8*, 44. [[CrossRef](#)]

31. Zhang, X.C.; Wang, M.Q.; Ordieres-Meré, J. Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0. *Sensors* **2018**, *18*, 2146. [[CrossRef](#)]
32. Chazard, E.; Ficheur, G.; Caron, A.; Lamer, A.; Labreuche, J.; Cuggia, M.; Genin, M.; Bouzille, G.; Duhamel, A. Secondary Use of Healthcare Structured Data: The Challenge of Domain-Knowledge Based Extraction of Features. *Stud. Health Technol. Inf.* **2018**, *255*, 15–19. [[CrossRef](#)]
33. Tan, C.; Ji, G.L. Semisupervised local preserving embedding algorithm based on maximummargin criterion for large-scale data streams. *Concurr. Comput. Pract. Exper.* **2017**, *29*, e4246.1–e4246.13. [[CrossRef](#)]
34. Tee-Ann, T. The extraction of urban road inventory from mobile lidar system. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *169*, 012022. [[CrossRef](#)]
35. Zhou, K.; Oh, S.K.; Pedrycz, W.; Qiu, J.L. Data preprocessing strategy in constructing convolutional neural network classifier based on constrained particle swarm optimization with fuzzy penalty function. *Eng. Appl. Artif. Intell.* **2023**, *117*, 105580. [[CrossRef](#)]
36. Ducarme, B.; Barriot, J.P.; Zhang, F.Z. Combination of Tsoft and ET34-ANA-V80 software for the preprocessing and analysis of tide gauge data in French Polynesia. *Geod. Geodyn.* **2023**, *14*, 26–34. [[CrossRef](#)]
37. Shin, K.; You, S.; Kim, M. A Comparison of Two-Stage Least Squares (TSLS) and Ordinary Least Squares (OLS) in Estimating the Structural Relationship between After-School Exercise and Academic Performance. *Mathematics* **2021**, *9*, 3105. [[CrossRef](#)]
38. Lee, J.; Lee, W.S.; Jung, H.; Lee, S.G. Comparison between total least squares and ordinary least squares in obtaining the linear relationship between stable water isotopes. *Geosci. Lett.* **2022**, *9*, 11. [[CrossRef](#)]
39. Chatterjee, S. A New Coefficient of Correlation. *J. Am. Stat. Assoc.* **2021**, *116*, 2009–2022. [[CrossRef](#)]
40. Patrick, S.; Christa, B.; Schwarte, L. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [[CrossRef](#)]
41. Li, T.; Song, X.Y.; Zhang, Y.Y.; Zhu, H.T.; Zhu, Z.Y. Clusterwise functional linear regression models. *Comput. Stat. Data Anal.* **2021**, *158*, 107192. [[CrossRef](#)]
42. Bun, M.J.G.; Harrison, T.D. OLS and IV estimation of regression models including endogenous interaction terms. *Econom. Rev.* **2019**, *38*, 814–827. [[CrossRef](#)]
43. Campozano, L.; Robaina, L.; Gualco, L.F.; Maisincho, L.; Villacís, M.; Condom, T.; Ballari, D.; Páez, C. Parsimonious Models of Precipitation Phase Derived from Random Forest Knowledge: Intercomparing Logistic Models, Neural Networks, and Random Forest Models. *Water* **2021**, *13*, 3022. [[CrossRef](#)]
44. Peng, X.J.; Chen, D. PTSVRs: Regression models via projection twin support vector machine. *Inf. Sci.* **2018**, *435*, 1–14. [[CrossRef](#)]
45. Hodson, T.O.; Over, T.M.; Foks, S.S. Mean Squared Error, Deconstructed. *J. Adv. Model. Earth Syst.* **2021**, *13*, e2021MS002681. [[CrossRef](#)]
46. Rougier, J. Ensemble Averaging and Mean Squared Error. *J. Clim.* **2016**, *29*, 8865–8870. [[CrossRef](#)]
47. Zhang, D.B. A Coefficient of Determination for Generalized Linear Models. *Am. Stat.* **2017**, *71*, 310–316. [[CrossRef](#)]
48. Piepho, H.P. A coefficient of determination (R<sup>2</sup>) for generalized linear-mixed models. *Biom. J.* **2019**, *61*, 860–872. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.