Journal of Infrastructure, Policy and Development 2024, 8(12), 9817.
https://doi.org/10.24294/jipd.v8i12.9817

Article

Machine learning-based clustering analysis of the banking green credit for the renewable energy industry

Quynh Vu-Thi-Nhu

Institute of Financial Management, Vietnam Maritime University, Hai Phong 484, Vietnam; quynhvn.qtc@vimaru.edu.vn

CITATION

Vu-Thi-Nhu Q. (2024). Machine learning-based clustering analysis of the banking green credit for the renewable energy industry. Journal of Infrastructure, Policy and Development. 8(12): 9817. https://doi.org/10.24294/jipd.v8i12.9817

ARTICLE INFO

Received: 20 August 2024
Accepted: 30 October 2024
Available online: 5 November 2024

Abstract: Credit policies for clean and renewable energy businesses play a crucial role in supporting carbon neutrality efforts to combat climate change. Clustering the credit capacity of these companies to prioritize lending is essential given the limited capital available. Support Vector Machine (SVM) and Artificial Neural Network (ANN) are two robust machine learning algorithms for addressing complex clustering problems. Additionally, hyperparameter selection within these models is effectively enhanced through the support of a robust heuristic optimization algorithm, Particle Swarm Optimization (PSO). To leverage the strength of these advanced machine learning techniques, this paper aims to develop SVM and ANN models, optimized with the PSO, for the clustering problem of green credit capacity in the renewable energy industry. The results show low Mean Square Error (MSE) values for both models, indicating high clustering accuracy. The credit capabilities of wind energy, clean fuel, and biomass pellet companies are illustrated in quadrant charts, providing stakeholders with a clear view to adjust their credit strategies. This helps ensure the efficient operation of banking green credit policies.

Keywords: green credit; renewable energy; particle swarm optimization; machine learning; clustering

1. Introduction

1.1. Background and motivation

Renewable energy is one of the critical factors that necessitate increased attention, development, and investment like never before (Bui-Duy et al., 2023). Green fintech is recently influencing the utilization of natural resources (Kai et al., 2024; Leng et al., 2024) and determining the raw materials for the production of renewable energy. Green credit clustering is a crucial tool of green fintech for the renewable energy industry, particularly in developing nations where financial access is often constrained. In regions such as Southeast Asia and Sub-Saharan Africa, the demand for renewable energy is growing rapidly. According to the International Renewable Energy Agency (IRENA), global renewable energy capacity increased by 260 GW in 2021, with developing countries accounting for nearly two-thirds of this expansion. However, the financing gap for renewable energy projects in these nations remains significant, with an estimated annual shortfall of over $200 billion required to meet international climate goals.

Artificial intelligence (AI) techniques, encompassing machine learning and deep learning algorithms, exhibit remarkable superiority in forecasting and classification tasks, achieving impressive levels of accuracy (Samal and Kumar, 2024b). These methods are increasingly playing a pivotal role in addressing complex challenges across various domains (Samal and Kumar, 2024a; Le, 2025). The application of machine learning in green credit forecasting has become increasingly significant, particularly as financial institutions seek efficient and reliable methods to assess credit risk and allocate resources toward sustainable investments (Sun, 2022). Machine learning algorithms, such as SVM, ANN, and decision trees, have demonstrated high accuracy in forecasting credit risk by analyzing historical financial data, environmental, social, and governance (ESG) metrics, and other relevant financial indicators (Ben Jabeur et al., 2021). In green credit forecasting, these models can identify patterns in financial and environmental data, helping lenders predict a firm's creditworthiness based on its adherence to sustainable practices. Additionally, advanced machine learning models like ensemble methods or deep learning architectures allow for the integration of various data sources, which enhances predictive performance and offers a more comprehensive understanding of risk factors associated with green credit. Such applications not only improve the credit evaluation process but also promote green finance by systematically directing funds toward environmentally responsible companies, contributing to broader sustainability goals.

Green credit clustering, driven by machine learning, enables financial institutions to effectively assess the creditworthiness of renewable energy projects by grouping them based on risk factors, financial performance, and environmental impact. For example, clustering algorithms can analyze large datasets on total capital, asset valuations, P/E ratios, profitability, and collateral values, allowing banks to optimize their credit portfolios. In a study by the World Bank, clustering-based credit analysis improved loan approval times by 25%, significantly enhancing the efficiency of capital allocation to renewable energy projects.

Moreover, green credit clustering ensures that resources are directed toward projects with the highest potential for both financial viability and environmental benefits. For instance, solar energy projects, which have seen an average annual growth rate of 20% globally, can be prioritized in regions with high solar irradiance, while biomass energy projects, which are crucial for waste-to-energy initiatives, may be clustered separately based on resource availability and technological maturity. In this way, clustering supports targeted financing, which is essential for accelerating the transition to clean energy in developing countries. This strategic allocation of credit can enhance project success rates, reduce default risks, and contribute to achieving sustainable development goals (SDGs), especially SDG 7, objecting to fostering access to affordable, reliable, and sustainable energy.

1.2. Research objectives and contributions

To address the research gap, this paper implements PSO-SVM and PSO-ANN models to solve the clustering problem of banking green credit for renewable energy enterprises using seven input variables and three output variables. The clustering and validation results indicate which companies have high, medium, or low credit coefficients, approval times, and credit ratios. This assists credit policy managers and credit enterprises in assessing their credit status and capacity, allowing them to develop suitable strategies to leverage credit resources in support of the carbon neutrality transition.

This paper provides significant contributions to literature and financial operators as follows:

(1) We develop machine learning models for optimizing hyperparameters using the heuristic Particle Swarm Optimization (PSO) algorithm to cluster banking green credit.

(2) The clustering model offers a robust perspective for financial strategy operators in energy companies to assess their credit capabilities.

(3) Based on the clustering results, both regulatory authorities and banks, along with energy enterprises, can formulate appropriate credit strategies to support and promote the renewable energy sector.

(4) The rest of this paper is structured as follows. A review of relevant studies is provided in the Section 2; Section 3 describes the collected data and proposes two machine learning models; In Section 4, the clustering results will be presented; Section 5 discusses some implications and the conclusion.

2. Related works

2.1. Review of banking green credit

Banking green credit clustering is a data-driven approach aimed at improving the allocation and management of financial resources for sustainable projects (Wang et al., 2021), particularly those in renewable energy sectors such as wind, solar, and biomass energy. This method involves grouping or clustering green credit portfolios based on various financial, environmental, and project-specific attributes using machine learning algorithms. The primary goal of green credit clustering is to enable banks and financial institutions to better assess the risk (Al-Qudah et al., 2023), profitability, and creditworthiness of renewable energy projects by identifying patterns and similarities within large datasets.

By clustering similar projects together, banks can optimize their lending strategies, allocate funds more efficiently, and mitigate potential financial risks (Moradi and Mokhatab Rafiei, 2019). This is especially crucial for supporting renewable energy investments in developing countries, where financial risk is often higher due to unstable economic conditions and less mature regulatory frameworks. Green credit clustering facilitates more informed decision-making, allowing institutions to balance the need for profitability with sustainability goals, while also promoting the wider adoption of clean energy technologies.

Given its increasing importance, numerous studies have been conducted to explore this topic. A typical paper examined how financial clustering influences green development, considering both linear and nonlinear effects (Tao et al., 2023). Empirical findings indicated that a 1% rise in financial clustering led to a 0.1012% improvement in green development at the city level. The spatial Durbin model revealed that financial clustering played a significant role in supporting local green development, although its spillover effects remain relatively limited. Another paper introduced a novel framework for customizing green building assessment tools tailored to regions or countries with diverse climates (Sadeghi et al., 2022). The framework utilized the K-means clustering method to group different climatic conditions, along with the silhouette value (SV) to verify the clustering accuracy.

Additionally, local experts' input was incorporated for further customization of the assessment tools. To fine-tune the regulations specific to each climatic zone, the Fuzzy Analytical Hierarchy Process (AHP) was applied. The methodology was validated through a real-world case study conducted in Iran. An investigation of the effect of the "Green Credit Guidelines" (2012 Guidelines) policy was conducted on the green transformation of heavily polluting industries (HPIs) by employing both the Difference in Differences (DID) and Propensity Score Matched-Difference in Differences (PSM-DID) models (Tian et al., 2022).

The relationship between banks' green lending practices and their credit risk was c examined, focusing on how Chinese green finance regulations influenced both the solvency of individual banks and the stability of the financial system (Zhou et al., 2022). The findings indicated that while China's Green Credit Policy reduced credit risk for large state-controlled banks, it increased credit risk for city and regional commercial banks. This disparity in performance was largely attributed to information and expertise asymmetries, as city and regional banks had more limited access to the knowledge and resources needed to accurately assess the credit risks associated with green lending.

2.2. Review of green credit for the renewable energy industry

As countries aim to meet carbon reduction targets and transition to cleaner energy sources (Hoang Huong et al., 2021), green credit schemes offer a vital mechanism for directing capital toward environmentally responsible investments (D'Orazio and Popoyan, 2019). These financing options often come with favorable terms, such as lower interest rates or extended repayment periods, making them attractive to renewable energy developers (Li et al., 2024). Green credit not only stimulates the growth of the renewable energy industry but also encourages innovation by making funding more accessible to new technologies and projects in developing countries.

Additionally, green credit promotes long-term environmental and economic benefits by aligning financial institutions with sustainability goals. By assessing the environmental impact and risk of projects, banks and lenders can contribute to a more resilient and greener economy. However, challenges remain, particularly in standardizing green credit evaluation metrics and ensuring that smaller financial institutions, which often lack expertise in green finance, can participate effectively. Therefore, green credit is a powerful tool in the push for global energy transformation (Hassan et al., 2024), but its potential must be fully realized through collaborative efforts between governments, banks, and renewable energy stakeholders.

To investigate the effect of green credit on energy efficiency, the research employed the SE-SBM model and the spatial Durbin model (Zhao et al., 2023). The findings revealed that this tool not only considerably improved energy efficiency but also generated a notable positive spatial spillover influence beyond the local area. As a result, implementing green credit might boost efficiency in energy usage within an area while simultaneously promoting improvements in neighboring regions. Su et al. (2023) examined how the policy influences innovative technological advancements enacted by China's financial operator. The quasi-natural experiment revealed that the tool could foster innovative technological advancements in renewable energy companies and encourage incremental growth, balancing innovation categories by alleviating the financial limitations these firms face. The heterogeneity analysis indicated the positive effect of green credit for both revolutionary and progressive technical innovation for renewable energy enterprises that obtained superior efficiency in production while receiving reduced government aid. Lai et al. (2022) examined how the policy influenced the capital of new energy firms, along with the intermediary functions of financial limitations and external scrutiny in the correlation between green credit and the economic benefits of new energy companies. The findings indicated that green credit created a substantial positive effect on the value of new energy firms, and this beneficial impact could be sustained over the long term.

The EBM (epsilon-based measure) super-efficiency model was developed to evaluate the extent of energy-effective consumption in China (Ma et al., 2021). The authors employed the regression discontinuity design (RDD) approach for empirically analyzing the net impact of the policy on energy efficiency levels, and examining the local heterogeneity of the policy. A coupling degree of coordination paradigm for the vibrant finance–clean power grid was developed (Zhao et al., 2023). They further introduced a fuzzy set qualitative comparative analysis (fsQCA) model to investigate various patterns for enhancing coupling coordination. Empirical findings revealed that the coupling coordination degree between green finance and clean energy in China increased from 0.3341 to 0.4718, though it remained near an imbalanced state. Additionally, the regional coupling coordination degree showed uneven development, with a trend toward high-value clustering.

In the domain of clustering green credit banking and renewable energy, machine learning algorithms demonstrate significant advantages over traditional techniques due to their high accuracy, multilayer separation capabilities, suitability for large datasets, and ability to address complex issues (Le et al., 2020; Le and Xuan-Thi-Thu, 2024). Bucur et al. (2021) examined various energy indicators assessed over a 12-year period using statistical methods and machine learning techniques, including an unsupervised clustering algorithm utilizing Self-Organizing Maps (SOM). A back-propagation neural network (BPNN) model was validated using principal component analysis and factor analysis, and its performance was tested with sample data (Feng, 2022). The results indicated that the BPNN-based credit risk assessment model achieves 95% accuracy.

To the best of our knowledge, previous studies have not focused on the implementation of clustering models utilizing advanced machine learning techniques to explore green credit strategies for renewable energy enterprises in developing countries. To address this research gap, this paper develops an advanced machine learning model optimized through hyperparameter tuning using one of the most powerful heuristic algorithms, Particle Swarm Optimization (PSO). The proposed model is designed to evaluate the clustering of credit indicators for renewable energy enterprises, using Vietnam as a case study.

3. Materials and methods

3.1. Data collection and preprocessing

The data was collected from the databases of 13 banks in Vietnam from 2015 to 2020, including the following variables: company name, main business sector, total capital, total assets, total debt, P/E ratio, profitability ratio, collateral value, guarantee coefficient, credit coefficient, credit ratio, and approval time. The dataset contains 1,236,329 observations. We selected companies based on industry groups, consisting of 3 wind power companies, 8 clean energy companies, and 21 biomass pellet companies.

The data was preprocessed to remove N/A values and outliers using z-score, Cook's distance, and Mahalanobis distance methods. Subsequently, the data was normalized using min-max normalization to eliminate large disparities in measurement units and ensure more accurate model performance. The data is split into training and testing sets with a ratio of 80:20.

Table 1. Descriptive statistics of selected input variables (Source: author).

Variables	Min	Max	Mean	Standard deviation	VIF
Total capital (TC) (10^3 USD)	10,387	52,837	19,746	5,877	1.736
Total assets (TA) (10^3 USD)	12,987	65,837	29,835	18,972	0.987
Total debt (TB) (10^3 USD)	7862	34,927	7,927	5,110	1.973
P/E ratio (PE)	3.23	28.76	16.67	3.45	2.654
Profitability ratio (PR) After-tax–(%)	16.76	65.26	27.65	21.02	1.997
Collateral value (CV) (10^3 USD)	987.23	32,465.14	7046.28	2198.17	2.926
Guarantee ratio (GR) (%)	0	50.35	10.26	5.37	1.765
Credit coefficient (CC)	456.37	621.14	499.77	123.05	-
Credit ratio (CR) (%)	30.56	86.72	52.47	12.99	-
Approval time (AT) (day)	27.63	134.16	78.72	33.35	-

The descriptive statistics of the selected input variables provide key insights into the distribution and variability of financial and credit-related indicators for the dataset. Total capital (TC) and total assets (TA) have mean values of 19,746 and 29,835 (in thousand USD), respectively, with standard deviations of 5877 and 18,972, indicating moderate variation in capital and significant variation in assets across firms. Total debt (TB) displays a lower mean (7927) but a higher variability relative to its mean, as evidenced by a standard deviation of 5110 in **Table 1**.

The P/E ratio (PE) has an average of 16.67 with a standard deviation of 3.45, demonstrating moderate fluctuation, while the profitability ratio (PR) shows more variability, with a mean of 27.65% and a high standard deviation of 21.02%, indicating wide-ranging profitability among firms. Collateral value (CV), with a mean of 7046.28 (in thousand USD) and a relatively large standard deviation (2198.17), highlights considerable variation in the assets firms use to secure financing.

The guarantee ratio (GR) has a mean of 10.26%, with a standard deviation of 5.37%, showing that most firms have a relatively low guarantee level. Both credit coefficient (CC) and credit ratio (CR) exhibit moderate averages of 499.77 and

52.47%, respectively, though CR has a wider range (min: 30.56, max: 86.72), suggesting diverse credit allocation practices. Finally, approval time (AT) ranges widely from 27.63 to 134.16 days, with an average of 78.72 days, reflecting variability in the time required for credit approvals.

Regarding multicollinearity, the variance inflation factor (VIF) values are all below 10, with most being below 2, indicating that multicollinearity is not a significant concern in the dataset.

3.2. Methodologies

3.2.1. PSO-SVM

We optimize the hyperparameters of a Support Vector Machine (SVM) model using Particle Swarm Optimization (PSO) in order to predict credit-related outputs like credit coefficient, credit ratio, and approval time based on financial input features. The goal is to find optimal parameters that improve clustering performance across these outputs. The objective function to be optimized is based on clustering-specific metrics using the Silhouette score. **Figure 1** presents the flowchart of PSO-SVM and PSO-ANN.

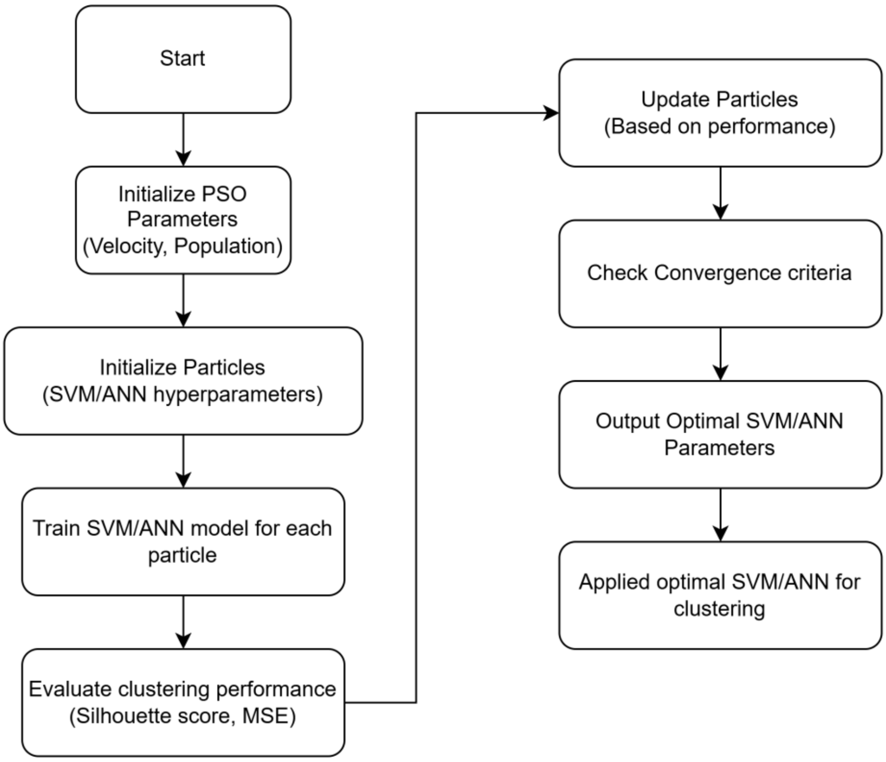


Figure 1. Algorithms flowchart of PSO-SVM and PSO-ANN.

The SVM decision function with an RBF kernel can be expressed as:

$$f(x) = sign\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right) \tag{1}$$

where the kernel function $K(x_i, x)$ for the RBF kernel is:

$$K(x_i, x) = exp(-\gamma \|x_i - x\|^2) \tag{2}$$

The parameters to optimize include the regularization parameter C and the kernel parameter γ. We define the loss function $L(C, \gamma)$ as the performance measure for clustering outputs, such as minimizing the difference between predicted and actual values of credit coefficient, credit ratio, and approval time:

$$L(C, \gamma) = \frac{1}{N} \sum_{i=1}^{N} (y_i - f(x_i))^2 \tag{3}$$

Next, we initialize a swarm of particles, where each particle represents a potential set of SVM hyperparameters $x_i = \{C_i, \gamma_i\}$ and initialize each particle's position and velocity within the search space defined by feasible ranges for C and γ.

The SVM model is trained using the parameters C_i and γ_i from each particle. For each particle, the architecture evaluates its performance based on the loss function $L(C, \gamma)$ calculated on the predicted outputs (credit coefficient, credit ratio, approval time). The velocity and position of each particle are updated using the following equations:

$$v_i^{(t+1)} = \omega v_i^{(t)} + c_1 r_1 (p_i^{best} - x_i^{(t)}) + c_2 r_2 (g^{best} - x_i^{(t)}) \tag{4}$$

where ω is the inertia weight, controlling the impact of previous velocities; c_1 and c_2 are cognitive and social coefficients guiding particles towards their best positions and the global best position; r_1 and r_2 are random numbers between 0 and 1. The position is then updated by $x_i^{(t+1)} = x_i^{(t)} + v_i^{(t+1)}$ The swarm's positions and velocities are continuously updated and velocities until a stopping criterion is met (e.g., the loss function converges).

3.2.2. PSO-ANN

For PSO-ANN, we define the loss function based on the Mean Squared Error (MSE) as Equation (5). MSE is a widely used metric in machine learning for regression tasks, as it provides a clear measure of prediction accuracy by penalizing larger errors more heavily due to its squared term (Samal and Kumar, 2024b; Samal and Kumar, 2024c). In clustering tasks, MSE can be effective in evaluating the compactness of clusters by measuring the average squared distance between each point and its assigned cluster center, helping to assess clustering precision (Samal and Kumar, 2024a).

$$L(W, b) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{5}$$

where y_i is the actual output and $\hat{y}_i$ is the predicted output based on the network's weights W and biases b. The forward pass through the ANN is given by:

$$a^{(l+1)} = \sigma(W^{(l)} a^{(l)} + b^{(l)}) \tag{6}$$

where: $a^{(l+1)}$ is the activation of the $l + 1 - th$ layer; $W^{(l)}$ is the weight matrix for the $l - th$ layer; $b^{(l)}$ is the bias vector for the $l - th$ layer; σ is the activation function. Each particle in the swarm represents a potential set of ANN parameters, including weights W and biases b. Additionally, the learning rate can be included as a particle dimension. Then we randomly initialize the positions and velocities of the particles within predefined ranges. For each particle, we construct an ANN using the current weights and biases, then perform forward propagation through the network. The MSE is calculated for each particle. The velocity and position of each particle are updated using the standard PSO formulas:

$$v_i^{(t+1)} = \omega v_i^{(t)} + c_1 r_1 (pi^{best} - x_i^{(t)}) + c_2 r_2 (g^{best} - x_i^{(t)}) \tag{7}$$

$$x_i^{(t+1)} = x_i^{(t)} + v_i^{(t+1)} \tag{8}$$

The updated positions represent new sets of ANN weights, biases, and learning rates. The architecture continues evaluating and updating the particles until the predefined convergence criterion is satisfied. Once PSO has converged, the optimal set of parameters is used to retrain the ANN on the entire dataset.

4. Analysis results

The results from the hyperparameter tuning of the PSO-SVM and PSO-ANN models demonstrate a well-optimized configuration that enhances both accuracy and generalization (see **Tables 2** and **3**). For the PSO-SVM, the optimal penalty parameter (C) was set at 150, indicating a focus on reducing misclassification while maintaining an appropriate margin. The kernel coefficient (γ) of 0.01 suggests a moderate influence of individual training points in the RBF kernel, which was selected for its ability to capture non-linear relationships in the data. An epsilon value of 0.05 provides flexibility in regression performance by allowing small deviations without penalty, and the stopping tolerance of 0.001 ensures that the model halts training once performance plateaus, avoiding unnecessary computation. The model was allowed up to 1500 iterations, ensuring thorough training without overfitting.

Table 2. Optimal hyperparameters of PSO-SVM (Source: author).

Hyperparameter	Description	Search Range	Optimal Value
C (Penalty parameter)	Controls trade-off between maximizing margin and minimizing classification error	[0.1, 1000]	150
γ (Kernel coefficient)	Determines the influence range of each training sample in the RBF kernel	[0.0001, 1]	0.01
Kernel	Kernel type to be used in the SVM. RBF is used for non-linear classification	Linear, RBF, Sigmoid	RBF
Epsilon	Epsilon specifies the margin of tolerance where no penalty is given in the SVM's loss function	[0.01, 0.1]	0.05
Stopping tolerance	Stopping criterion when model performance improvement is minimal	[0.0001, 0.01]	0.001
Max iterations	The maximum number of iterations for the SVM training process	1000, 2000	1500

Table 3. Optimal hyperparameters of PSO-ANN (Source: author).

Hyperparameter	Description	Search Range	Optimal Value
Number of hidden layers	Number of hidden layers in the neural network	1–5	3
Neurons per layer	Number of neurons in each hidden layer	[16, 128] (per layer)	64 (Layer 1), 32 (Layer 2), 16 (Layer 3)
Activation function	Activation function for the hidden layers. Common choices include ReLU, sigmoid, tanh	ReLU, Sigmoid, Tanh	ReLU
Output activation function	Activation function for the output layer. Linear for regression tasks	Linear, Sigmoid	Linear
Learning rate	Controls the step size during gradient descent	[0.0001, 0.1]	0.001
Batch size	Number of samples per gradient update	[16, 128]	64
Momentum	Used in gradient descent optimization to accelerate convergence	[0.5, 0.99]	0.9
Number of epochs	Number of times the learning algorithm will work through the entire training dataset.	[50, 500]	300
Optimizer	Optimization algorithm for updating the weights (e.g., SGD, Adam, RMSprop)	SGD, Adam, RMSprop	Adam
Regularization	Regularization method to prevent overfitting (L1, L2, or dropout)	L1, L2, Dropout	L2 (lambda = 0.001)
Dropout rate	Probability of setting a neuron to zero during training (if dropout is used)	[0, 0.5]	0.3
Cross-validation folds	Number of folds for cross-validation to assess model performance	3, 5, 10	5

For the PSO-ANN model, the architecture with 3 hidden layers was found optimal, with 64, 32, and 16 neurons in successive layers, indicating a decreasing complexity as the model processes more abstract features of the data. The ReLU activation function was chosen for the hidden layers to handle non-linearities effectively, while a linear activation was used in the output layer for regression tasks. A learning rate of 0.001 and batch size of 64 were selected to ensure smooth and efficient gradient descent. The momentum value of 0.9 helped accelerate convergence, and training over 300 epochs provided sufficient iterations for the model to learn without overfitting. The Adam optimizer was chosen for its adaptability in complex datasets, and L2 regularization (lambda = 0.001) along with a dropout rate of 0.3 helped prevent overfitting. Additionally, 5-fold cross-validation ensured the model’s generalizability and robustness.

Table 4 presents the clustering metric results of the performance of the PSO-SVM and PSO-ANN models. Both models demonstrated strong predictive accuracy and clustering ability, as reflected in the metrics. For the PSO-SVM, the training mean squared error (MSE) was 0.0073, and the validation MSE was slightly higher at 0.0118. These low error values indicate that the model was able to fit the training data well and generalize effectively to the validation data, though some slight overfitting may have occurred. The Silhouette Score of 0.78 suggests that the PSO-SVM achieved a high degree of cohesion within clusters and separation between different clusters, indicating solid clustering performance. The PSO-ANN model, on the other hand, demonstrated even better accuracy with a training MSE of 0.004 and a validation MSE of 0.006, showing minimal discrepancy between training and validation, which indicates good generalization. The Silhouette Score of 0.81 further suggests that the PSO-ANN outperformed the PSO-SVM in terms of clustering quality, with even tighter clusters and clearer separation between groups. Both models exhibited strong performance, but the PSO-ANN achieved slightly better results, particularly in terms of clustering quality and prediction accuracy, making it the superior model in this context.

Table 4. Clustering metric results (Source: author).

Model	Evaluation Metric	Value
PSO-SVM	Training MSE	0.0073
	Validation MSE	0.0118
	Silhouette Score	0.78
PSO-ANN	Training MSE	0.004
	Validation MSE	0.006
	Silhouette Score	0.81

Figure 2 depicts the Mean Squared Error (MSE) for both PSO-SVM and PSO-ANN models during the training and testing phases across 100 epochs. The MSE serves as an indicator of the models' performance, with lower values indicating better predictive accuracy. At the start (epoch 0), both models have relatively high MSE values, approximately 0.35–0.40, which is expected as the models begin with initial random weights. As training progresses, the MSE values for both the PSO-SVM and PSO-ANN models gradually decline, indicating that both models are learning from the data and improving their performance. Around epoch 40, the MSE values for both models begin to converge, with only slight variations between the training and testing phases. PSO-ANN shows slightly better performance, with its training MSE (green line) and testing MSE (red dashed line) consistently lower than those of PSO-SVM (blue and orange lines). This suggests that PSO-ANN is more efficient in capturing the underlying patterns of the data, leading to a more accurate model. By epoch 100, the MSE for both models plateaus at a low value of approximately 0.01 for PSO-SVM and 0.004 for PSO-ANN, suggesting that further training provides diminishing returns in terms of error reduction. This plateau signifies that both models have effectively converged, and no significant overfitting is observed as the testing MSE remains very close to the training MSE for both models. The PSO-SVM and PSO-ANN models demonstrate strong performance in minimizing the MSE over time, with PSO-ANN showing a slight edge in terms of lower overall error, making it the more effective model for this particular task. The small gap between training and testing errors in both models further confirms the robustness and generalization capabilities of the approaches.

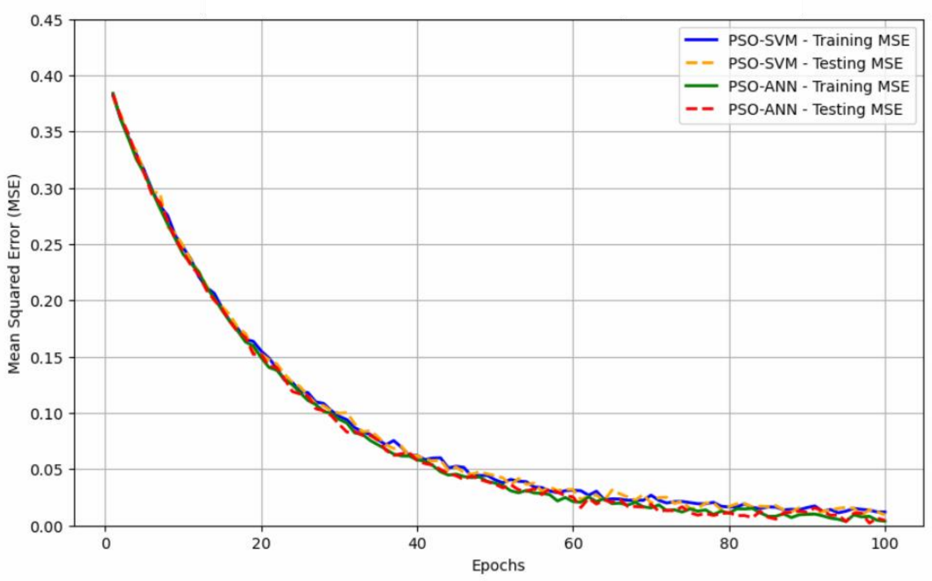


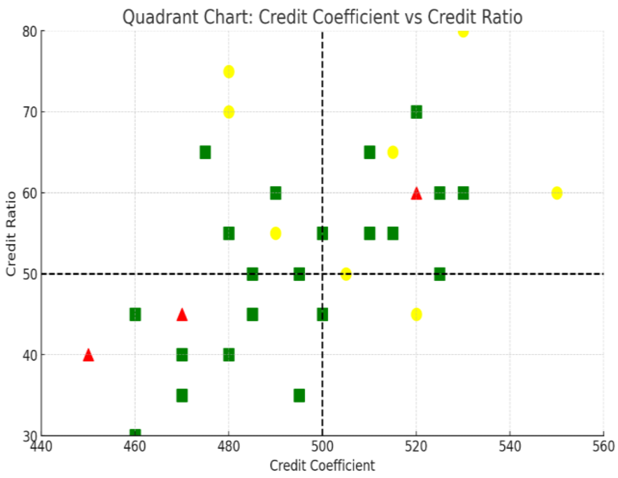
Figure 2. Training and testing results of PSO-SVM and PSO-ANN.

Figure 3 illustrates the clustering results of green credit for energy firms, specifically focusing on three energy sources: Biomass Pellet, Clean Fuel, and Wind Energy. These visualizations provide valuable insights into how these energy types perform in terms of credit coefficient, approval time, and credit ratio, revealing trends and disparities in green credit distribution. **Figure 3a**, which explores Credit Coefficient versus Credit Ratio, further highlights differences among the energy sources. The top-right quadrant (high credit coefficient and high credit ratio) features a mix of Clean Fuel and Biomass Pellet firms, indicating that these projects have both high credit exposure and strong creditworthiness. In contrast, Wind Energy firms are scattered in the bottom-left quadrant (low credit coefficient and low credit ratio), suggesting they may face challenges in securing large amounts of green credit despite their potential for lower financial risk. Biomass Pellet and Clean Fuel firms also appear in the lower-right quadrant, where credit coefficients are high but credit ratios are moderate or low, indicating a stable yet cautious investment approach from lenders.

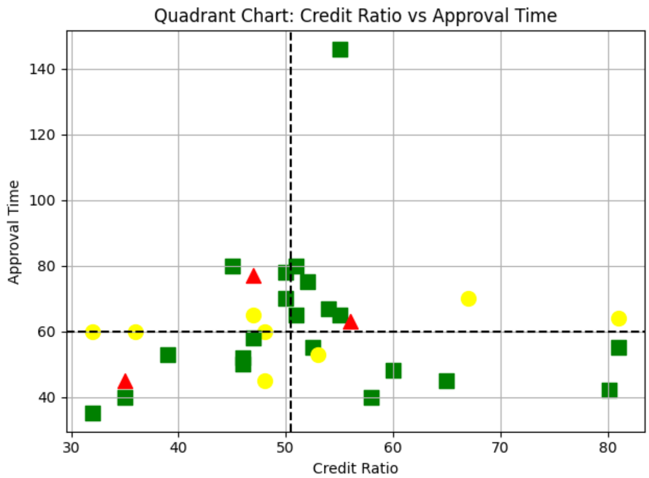
In **Figure 3b**, which plots the Credit Coefficient against Approval Time, distinct patterns emerge. The top-right quadrant, representing firms with high credit coefficients and longer approval times, is primarily occupied by Biomass Pellet and Clean Fuel firms. This suggests that, while these projects are perceived as financially sound, they require a lengthier assessment process before approval. On the other hand, the bottom-left quadrant includes firms with lower credit coefficients and shorter approval times, dominated by Biomass Pellet firms and a few Wind Energy firms, indicating that these projects are generally viewed as lower-risk and thus undergo quicker approval processes. Notably, Wind Energy firms are concentrated in the bottom-right quadrant, reflecting high credit coefficients paired with relatively short approval times. This suggests that despite their financial complexity, wind energy projects are considered attractive by green credit schemes, receiving faster approvals.

Figure 3c represents the relationship between Credit Ratio and Approval Time. This visualization offers insights into how these factors interact across different energy sources, revealing trends in credit allocations and the speed at which credit applications are approved. In the top-right quadrant, representing both high credit ratios and long approval times, the Biomass Pellet firms are notably present. This suggests that while these projects are granted a significant portion of the credit, they also experience delays in approval, possibly due to their complexity or the thorough assessment required. Projects in this quadrant are likely seen as high-return but high-risk, necessitating more extensive due diligence before approval. On the other hand, Clean Fuel firms appear primarily in the bottom-left quadrant, where both the credit ratio and approval time are relatively low. This indicates that these projects receive a smaller share of credit but are approved more quickly. The clustering of these firms in this area may imply that clean fuel projects are considered lower-risk but also less capital-intensive, requiring less scrutiny and faster processing times. Wind Energy firms appear scattered across the quadrants, with a few located in the bottom-right quadrant (high credit ratio and short approval time). This positioning reflects a favorable scenario for wind energy projects, where they receive substantial credit while enjoying faster approvals, possibly due to the maturity of wind energy technology and its recognized potential for contributing to green development.

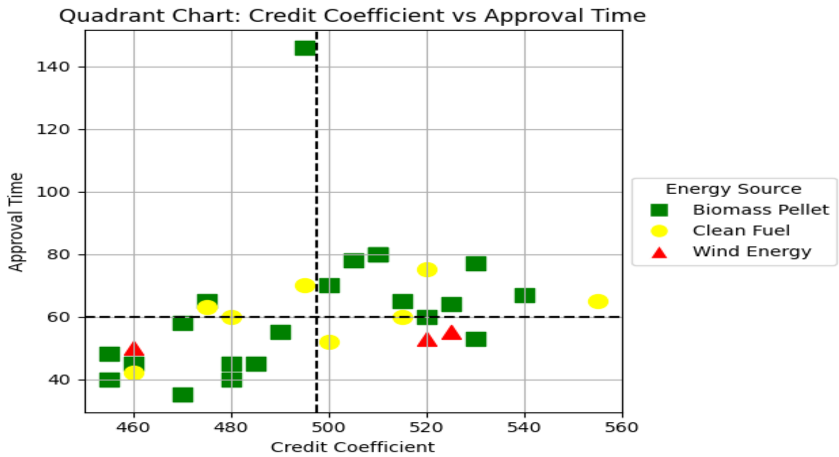
Biomass Pellet firms are distributed more uniformly across the quadrants, reflecting a balanced approach to their credit allocation and risk profile. Wind Energy firms, while securing favorable credit terms with high coefficients and low approval times, struggle to receive high credit ratios, suggesting that lenders may be cautious about extending large credit amounts to these projects. Clean Fuel firms, on the other hand, show greater variability in both credit ratios and coefficients, implying that these projects face more uncertainty in credit evaluations, possibly due to the complexity or novelty of the technology involved. Biomass Pellet projects benefit from more stable and consistent credit support, while Clean Fuel projects face longer approval times and higher variance in credit conditions, indicating a need for more specialized financial assessments. The clustering patterns across these charts suggest that different renewable energy types require tailored financial mechanisms to address their unique risks and opportunities in the green credit landscape. These patterns underscore the need for tailored financial strategies in green credit policies, ensuring that various renewable energy types receive the right amount of credit support while minimizing delays in the approval process. The chart suggests that Wind Energy projects are particularly well-positioned to capitalize on green credit, while Biomass Pellet and Clean Fuel projects may require further adjustments in credit mechanisms to enhance their financial viability and reduce delays.



(a)



(b)



(c)

Figure 3. Clustering results of green credit for energy firms. **(a)** Quadrant chart of credit coefficient and credit ratio; **(b)** quadrant chart of credit ratio and approval time; **(c)** quadrant chart of credit coefficient and approval time.

5. Discussion and conclusion

This study successfully addressed the initial research problem. First, the authors developed a robust machine learning model optimized through the PSO algorithm to identify effective hyperparameters. The clustering model achieved high accuracy with an MSE of approximately 0.01 for the PSO-SVM model and 0.004 for the PSO-ANN model. Second, the clustering model effectively segmented energy enterprises using quadrant charts. Based on this, companies can forecast their credit coefficient, credit ratio, and approval time. Lastly, the results provide recommendations for firms to strategically manage their credit capacity, enabling them to efficiently and promptly access credit resources by leveraging the input variables of the model.

This study presents significant theoretical and practical implications. Theoretically, the machine learning model optimized by PSO represents an advancement in the implementation techniques for credit clustering. Compared to traditional segmentation methods, the accuracy achieved by this model is notably impressive (Mirza et al., 2023; Machado and Karray, 2022). By leveraging these sophisticated machine learning techniques, the research introduces a novel approach to green credit analysis for wind energy, clean fuel, and biomass pellet sectors in Vietnam. The integration of Particle Swarm Optimization (PSO) with Support Vector Machines (SVM) and Artificial Neural Networks (ANN) not only enhances the accuracy of the clustering process but also demonstrates the potential for optimizing hyperparameters in complex financial scenarios. This methodological advancement moves beyond traditional clustering approaches by providing a more precise and tailored solution for segmenting enterprises based on credit criteria. Furthermore, this study extends the theoretical framework of green finance by incorporating machine learning and heuristic optimization, opening new avenues for research in credit risk assessment and green credit policy development, especially in emerging markets with diverse energy sectors. Compared to some machine learning models applied to clustering tasks, the results of this study have not shown a significantly superior accuracy, with an MSE of approximately 0.004. Other studies, such as those conducted by Yang et al. (2019) and Hassan et al. (2021), report considerably lower MSE values. However, within the scope of this topic, our results demonstrate marked improvement over traditional classification methods, particularly in the context of banks in developing countries. This outcome will facilitate more efficient and streamlined credit assessment and approval procedures for energy-sector enterprises, enabling higher credit limits and thereby accelerating the carbon-neutral transition.

The practical implications of this study are significant for the banking sector and renewable energy enterprises in Vietnam, particularly those involved in wind energy, clean fuel, and biomass pellet production. By implementing PSO-SVM and PSO-ANN clustering models for green credit analysis, banks can enhance their credit risk assessment and lending strategies for renewable energy projects. The models developed in this study offer improved accuracy in predicting credit coefficients, approval times, and credit ratios for renewable energy enterprises, enabling more informed lending decisions.

For financial institutions, the results of this study provide a practical tool to better segment enterprises based on creditworthiness, allowing for more efficient allocation of green credit resources. Banks can optimize their credit approval processes by identifying which firms are more likely to meet credit requirements and benefit from green financing, ultimately reducing loan processing time and associated risks. Additionally, this methodology helps align financial practices with Vietnam’s sustainable development goals, encouraging investment in renewable energy sectors.

For renewable energy companies, the study offers actionable insights into how they can improve their credit standing and access to green financing. By understanding the key input factors that influence credit decisions—such as total capital, asset value, and debt levels—enterprises can strategically manage their financials to increase their chances of securing green loans. This fosters more effective financial planning and strengthens their capacity to scale up renewable energy projects. In turn, this supports the broader goal of transitioning to a low-carbon economy, helping Vietnam meet its renewable energy targets.

This paper opens up future research directions, including the potential to further compare the performance of various hyperparameter optimization algorithms—such as Bayesian optimization, Genetic Algorithm (GA), and Whale Optimization Algorithm (WOA)—to validate clustering results. Additionally, as machine learning techniques excel in high accuracy but are less focused on interpretability, future studies could integrate explanatory methods, such as the method of moment quantile regression (Kai et al., 2024; Leng et al., 2024), to examine influential factors in green credit policies across detailed quantiles. This approach could facilitate deeper application in managing credit projects for energy enterprises. Moreover, the findings of this research have the potential for broad application not only within the energy sector but could also extend to other fields like transportation and logistics, which utilize green fuel solutions.

Acknowledgments: This research was financially supported by Vietnam Maritime University.

Conflict of interest: The authors declare no conflict of interest.

References

Al-Qudah, A. A., Hamdan, A., Al-Okaily, M., Alhaddad, L. (2023). The impact of green lending on credit risk: evidence from UAE's banks. Environmental Science and Pollution Research, 30(22), 61381–61393. https://doi.org/10.1007/s11356-021-18224-5

Ben Jabeur, S., Khalfaoui, R., Ben Arfi, W. (2021). The effect of green energy, global environmental indexes, and stock markets in predicting oil price crashes: Evidence from explainable machine learning. Journal of Environmental Management, 298. https://doi.org/10.1016/j.jenvman.2021.113511

Bucur, C., Tudorica, B. G., Oprea, S. V., Nancu, D., & Dusmanescu, D. M. (2021). Insights into Energy Indicators Analytics towards European Green Energy Transition Using Statistics and Self-Organizing Maps. IEEE Access, 9, 64427–64444. https://doi.org/10.1109/ACCESS.2021.3075175

Bui-Duy, L., Le, L. T., Vu-Thi-Minh, N., Hoang-Huong, G., Bui-Thi-Thanh, N., & Nguyen, P. N. (2023). Economic and environmental analysis of biomass pellet supply chain using simulation-based approach. Asia Pacific Management Review, *28*(4), 470–486. https://doi.org/10.1016/j.apmrv.2023.02.002

D'Orazio, P., & Popoyan, L. (2019). Fostering green investments and tackling climate-related financial risks: Which role for macroprudential policies? Ecological Economics, 160, 25–37. https://doi.org/10.1016/j.ecolecon.2019.01.029

Feng, Y. (2022). Bank Green Credit Risk Assessment and Management by Mobile Computing and Machine Learning Neural Network under the Efficient Wireless Communication. Wireless Communications and Mobile Computing, 2022. https://doi.org/10.1155/2022/3444317

Hassan, B. A., Rashid, T. A., Mirjalili, S. (2021). Performance evaluation results of evolutionary clustering algorithm star for clustering heterogeneous datasets. Data in Brief, 36. https://doi.org/10.1016/j.dib.2021.107044

Hassan, Q., Viktor, P., J. Al-Musawi, T., Mahmood Ali, B., Algburi, S., Alzoubi, H. M., Khudhair Al-Jiboory, A., Zuhair Sameen, A., Salman, H. M., & Jaszczur, M. (2024). The renewable energy role in the global energy Transformations. Renewable Energy Focus, 48. https://doi.org/10.1016/j.ref.2024.100545

Hoang Huong, G., Ta Anh, S., Le, L. T., Linh, B. D., & Vu Thi Minh, N. (2021). Supply chain disruption for crude oil due to the effect of Covid-19 pandemic: Evidence from Vietnam. International Journal of Logistics Systems and Management, 1(1), 1. https://doi.org/10.1504/ijlsm.2021.10044296

Kai, Z., Sharaf, M., Wei, S. Y., Shraah, A. Al, Le, L. T., Arvind Bedekar, D. A., Bani Ahmad, A. Y. A. (2024). Exploring the asymmetric relationship between natural resources, fintech, remittance and environmental pollution for BRICS nations: New insights from MMQR approach. Resources Policy, 90. https://doi.org/10.1016/j.resourpol.2024.104693

Lai, X., Yue, S., Chen, H. (2022). Can green credit increase firm value? Evidence from Chinese listed new energy companies. Environmental Science and Pollution Research, 29(13), 18702–18720. https://doi.org/10.1007/s11356-021-17038-9

Le, L. T., Lee, G., Park, K. S., Kim, H. (2020). Neural network-based fuel consumption estimation for container ships in Korea. Maritime Policy and Management, 47(5), 615–632. https://doi.org/10.1080/03088839.2020.1729437

Le, L. T., Xuan-Thi-Thu, T. (2024). Discovering supply chain operation towards sustainability using machine learning and DES techniques: a case study in Vietnam seafood. Maritime Business Review, EarlyCite. https://doi.org/https://doi.org/10.1108/MABR-10-2023-0074

Le, T. L. (2025). Uncovering Import Document Fraud: Leveraging the Deep Learning Approach. Global Trade and Customs Journal, 20(1), 1–8. https://kluwerlawonline.com/journalarticle/Global+Trade+and+Customs+Journal/20.1 [pre-publication]/GTCJ2025002

Leng, C., Wei, S. Y., Al-Abyadh, M. H. A., Halteh, K., Bauetdinov, M., Le, L. T., & Alzoubi, H. M. (2024). An empirical assessment of the effect of natural resources and financial technologies on sustainable development in resource abundant developing countries: Evidence using MMQR estimation. Resources Policy, 89. https://doi.org/10.1016/j.resourpol.2023.104555

Li, X., Abed, A. M., Shaban, M., Le, L. T., Zhou, X., Abdullaev, S., Alhomayani, F. M., Yasser Elmasry, Mahariq, I., & Rahman, A. A. (2024). Artificial Intelligence Application for Assessment/Optimization of a Double-Flash Geothermal Scheme Tailored Combined Heat/Power Plant. Energy, 133594. https://doi.org/https://doi.org/10.1016/j.energy.2024.133594

Ma, X., Ma, W., Zhang, L., Shi, Y., Shang, Y., & Chen, H. (2021). The impact of green credit policy on energy efficient utilization in China. Environmental Science and Pollution Research, 28(37), 52514–52528. https://doi.org/10.1007/s11356-021-14405-4

Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. Expert Systems with Applications, 200. https://doi.org/10.1016/j.eswa.2022.116889

Mirza, N., Afzal, A., Umar, M., & Skare, M. (2023). The impact of green lending on banking performance: Evidence from SME credit portfolios in the BRIC. Economic Analysis and Policy, 77, 843–850. https://doi.org/10.1016/j.eap.2022.12.024

Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. Financial Innovation, 5(1). https://doi.org/10.1186/s40854-019-0121-9

Sadeghi, M., Naghedi, R., Behzadian, K., Shamshirgaran, A., Tabrizi, M. R., & Maknoon, R. (2022). Customisation of green buildings assessment tools based on climatic zoning and experts judgement using K-means clustering and fuzzy AHP. Building and Environment, 223. https://doi.org/10.1016/j.buildenv.2022.109473

Samal, U., & Kumar, A. (2024a). Empowering software reliability: Leveraging efficient fault detection and removal efficiency. Quality Engineering, 1–12. https://doi.org/https://doi.org/10.1080/08982112.2024.2358889

Samal, U., & Kumar, A. (2024b). Fault Removal Efficiency: A Key Driver in Software Reliability Growth Modeling. In Springer Series in Reliability Engineering. Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-031-55048-5_7

Samal, U., & Kumar, A. (2024c). Incorporating human dynamics into software reliability analysis: learning, fatigue, and efficiency considerations. International Journal of System Assurance Engineering and Management. https://doi.org/https://doi.org/10.1007/s13198-024-02368-x

Samal, Umashankar, & Kumar, A. (2024a). A Neural Network Approach for Software Reliability Prediction. International Journal of Reliability, Quality and Safety Engineering, 31(3). https://doi.org/https://doi.org/10.1142/S0218539324500098

Samal, Umashankar, & Kumar, A. (2024b). Enhancing Software Reliability Forecasting Through a Hybrid ARIMA-ANN Model. Arabian Journal for Science and Engineering, 49(5), 7571–7584. https://doi.org/10.1007/s13369-023-08486-1

Su, T., Meng, L., Wang, K., & Wu, J. (2023). The role of green credit in carbon neutrality: Evidence from the breakthrough technological innovation of renewable energy firms. Environmental Impact Assessment Review, 101. https://doi.org/10.1016/j.eiar.2023.107135

Sun, C. (2022). The correlation between green finance and carbon emissions based on improved neural network. Neural Computing and Applications, 34(15), 12399–12413. https://doi.org/10.1007/s00521-021-06514-5

Tao, M., Silva, E., Sheng, M. S., Wen, L., & Qi, L. (2023). How financial clustering influences China's green development: Mechanism investigation and empirical discussion. Journal of Environmental Management, 347. https://doi.org/10.1016/j.jenvman.2023.119081

Tian, C., Li, X., Xiao, L., & Zhu, B. (2022). Exploring the impact of green credit policy on green transformation of heavy polluting industries. Journal of Cleaner Production, 335. https://doi.org/10.1016/j.jclepro.2021.130257

Wang, M., Li, X., & Wang, S. (2021). Discovering research trends and opportunities of green finance and energy policy: A data-driven scientometric analysis. Energy Policy, 154. https://doi.org/10.1016/j.enpol.2021.112295

Yang, M. S., Chang-Chien, S. J., & Nataliani, Y. (2019). Unsupervised fuzzy model-based Gaussian clustering. Information Sciences, 481, 1–23. https://doi.org/10.1016/j.ins.2018.12.059

Zhao, L. T., Li, F. R., & Wang, D. S. (2023). How to achieve the common developments of green finance and clean energy in China? Evidence from coupling coordination evaluation. Ecological Indicators, 155. https://doi.org/10.1016/j.ecolind.2023.111011

Zhao, X., Zeng, S., Ke, X., & Jiang, S. (2023). The impact of green credit on energy efficiency from a green innovation perspective: Empirical evidence from China based on a spatial Durbin model. Energy Strategy Reviews, 50. https://doi.org/10.1016/j.esr.2023.101211

Zhou, X. Y., Caldecott, B., Hoepner, A. G. F., & Wang, Y. (2022). Bank green lending and credit risk: an empirical analysis of China's Green Credit Policy. Business Strategy and the Environment, 31(4), 1623–1640. https://doi.org/10.1002/bse.2973