


Research Article

Application of Data Mining in Traditional Benchmark Evaluation Model for Buildings Energy Consumption

Donglei Wu ^{1,2} and Minwei Guo¹

¹School of Design, Jiangnan University, Wuxi 214122, China

²School of Arts and Design, Nanjing University of Forestry, Nanjing, Jiangsu 210037, China

Correspondence should be addressed to Donglei Wu; wudongleinanlin@njfu.edu.cn

Received 3 August 2021; Revised 30 August 2021; Accepted 9 September 2021; Published 1 October 2021

Academic Editor: Mian Ahmad Jan

Copyright © 2021 Donglei Wu and Minwei Guo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since the beginning of data mining technologies, buildings have become not just energy-intensive but also information-centric. Data mining technologies have been widely used to utilize the huge quantities of buildings' operational data to improve their energy systems. Conventional benchmarking of buildings' energy performance reflects a variety of parameters, such as the number of inhabitants, the environment, the energy efficiency of equipment utilized, and the adjustment of internal temperature. These various elements are then assigned weights to generate a single general indicator. This study presents a reasonable benchmark assessment methodology of conventional buildings' energy usage based on a data-mining algorithm for acquiring more specific information, like the energy management efficacy of a building, and aiming at the problem of ineffective use of large amounts of energy consumption in public buildings. A mathematical-statistical approach and a data-mining tool are used to analyse the data. The degree of connection between numerous influencing variables (i.e., characteristic parameters) and building's energy usage is determined using grey correlation analysis. In this work, we have used an enhanced Apriori algorithm to identify the link between the different forms of systems in the same area. In short, the fundamental idea and process of the Apriori algorithm are presented, and preliminary designs of the preprocessing of experimental data as well as the analysis methods are studied to analyse the outcome of the proposed work.

1. Introduction

Every technological improvement produces a series of products and services as part of the social evolution, but it also leads to a rapid increase in resource and energy consumption. Although a variety of technical developments may improve resource usage and energy efficiency, per capita energy consumption continues to rise. Buildings' energy efficiency is improved by having reasonable demand-side energy consumption, whereas supply-side energy demand is improved by having appropriate supply-side energy consumption [1]. To analyse the rationality of buildings' energy consumption on the demand side, we need to determine the energy wastes component and causes. It is important to explain the energy consumption characteristics of the particular interior environment. This achieves a

change from the supply to the demand side in the building's energy-saving mode. The energy consumption of buildings has become increasingly important as the construction industry has risen rapidly in conjunction with urbanization [2–4].

The design and construction industries account for 36% of worldwide final energy consumption and almost 40% of the total direct and indirect carbon dioxide emissions [5]. In the context of future construction management in an urban-based living environment, energy savings may be realized by increasing the dynamical energy efficiency of a building [6]. Moreover, due to the prevalence of smart sensors and the deployment of intelligent building management systems [7], construction activities are data-intensive. A significant quantity of building operational data is gathered to provide a framework for creating performance analysis. As a result,

using big data-driven techniques to construct smart energy management is a potential option for addressing energy conservation.

According to [8], the share of energy usage of buildings in the overall energy consumption is increasing steadily, putting enormous strain on energy consumption. The operational waste of buildings is a particularly critical component of energy consumption, accounting for roughly 80% of total energy consumption [9, 10]. Therefore, the only option to establish a resource-conserving society and economic development is to increase knowledge and cognition of buildings' energy conservation. The major challenge for the building sector is how to extract precious data quickly and efficiently from vast data, uncover the problems with using energy consumptions, and enhance the efficient and rational use of buildings' energy.

Traditional Chinese architecture has a complex and varied framework. First, classic architecture is brilliantly constructed which inspires future generations and provides a solid foundation for modern architectures. Furthermore, the traditional architecture allows for natural building ventilation, such as ventilation with pressure differences and ventilation between levels, which improves people's living conditions [11–13]. Traditional architecture requires a lot of climate adaption design. It often depends on the nature of area features, but people's demand for comfortable shelter is also quite strong. The primary goal of the construction industry is to offer pleasant living quarters. Because technology is continuously improving, the construction sector is looking for ways to reduce energy usage. The kind and number of small and large equipment in metropolitan public buildings, as well as data on energy use, are rapidly expanding. A large number of building energy consumption data have been accumulated, with the establishment of an energy consumption monitoring system and the implementation of air conditioning, lighting, power supply, and other items of measurement. The main characteristics of these data are a large amount of data and a large amount of information. Traditional data analysis methods cannot meet the potential value of these incomplete, irregular, and huge data. Data mining can effectively solve such problems. The particular mining process comprises data processing, preprocessing, data mining energy consumption, assessment of results from data information mining, and application of the model. Many other researchers have utilized algorithms like K-means, Chameleon, and DBSCAN to construct an energy-reduction cluster model [14–16]. When applying these methods to the energy consumption clustering of buildings such as buildings, offices, and malls, the distribution of energy and the average value of each cluster can be calculated. Following that, the data are utilized to create an energy conservation evaluation index of buildings. This gives us a scientific and acceptable basis for making building energy conservation decisions. Several data mining approaches may be used for data processing depending on the application areas, user expectations, and application procedures.

Data mining methods have been widely utilized to unlock the values of enormous volumes of building operation data, as the author in [17] worked to improve the

operational performance of building energy systems. This study attempts to give a comprehensive review of the applicability of data mining technologies in this industry. In general, there are two types of data mining technologies: supervised data mining and unsupervised data mining. For forecasting building energy load and detecting/diagnosing problems, supervised data mining algorithms are frequently utilized in this industry. The importance of energy-efficient building systems cannot be overstated, given the construction sector's continual development as a significant energy user in the modern world [16]. Today, most buildings have an electric dashboard for recording demand forecasts, which offers several study opportunities by employing this data in energy modelling. This paper investigates standard methods for regression in energy estimation and presents three models with data classifications to improve their performance. Regression methods and an artificial neural network model with data categorization for projecting hourly or hourly energy use in four different buildings are among the recommendation strategies. Energy data from a building energy simulation program as well as existing buildings are collected to develop models for a thorough study.

Figure 1 illustrates the fundamental procedure for the application of data mining in the building industry. After identifying issues and targets, we have gathered the appropriate data and a database. Among other techniques, we have utilized a building management system and field measurement and then our desired database prepared on the obtained data. On this foundation, a data warehouse or data mart is built. The data are then analysed, and the most valuable patterns or rules are determined using the proper data mining technologies. Finally, experts in the building sector can extract the information related to these patterns or principles.

Data mining is a multidisciplinary technology that combines artificial intelligence, machine learning, data visualization, and other modern technology to retrieve different algorithms from a large, irregular, complex database using cluster analysis, preliminary analysis, correlation analysis, and other methods to find information that has potential value. Data mining technology has the characteristics of mass and relativity, which can effectively solve the effective classification and information mining of a large number of data and has given full play to its advantages in many fields. For example, on the Internet, IT industry, data mining technology can help the industry realize efficient processing of large-scale e-commerce data, obtain valuable results for enterprises, help enterprises make scientific and correct marketing decisions, and finally achieve accurate push and marketing. The application of data mining in the field of building HVAC is mainly divided into data mining framework process, preprocessing and specific application, etc. Presently, it mostly entails the analysis of building energy consumption data, problem diagnosis and detection, and system and data operation and control optimization, among other things. In construction management, the existing building energy management and automatic control system, which can store a lot of construction operation data,

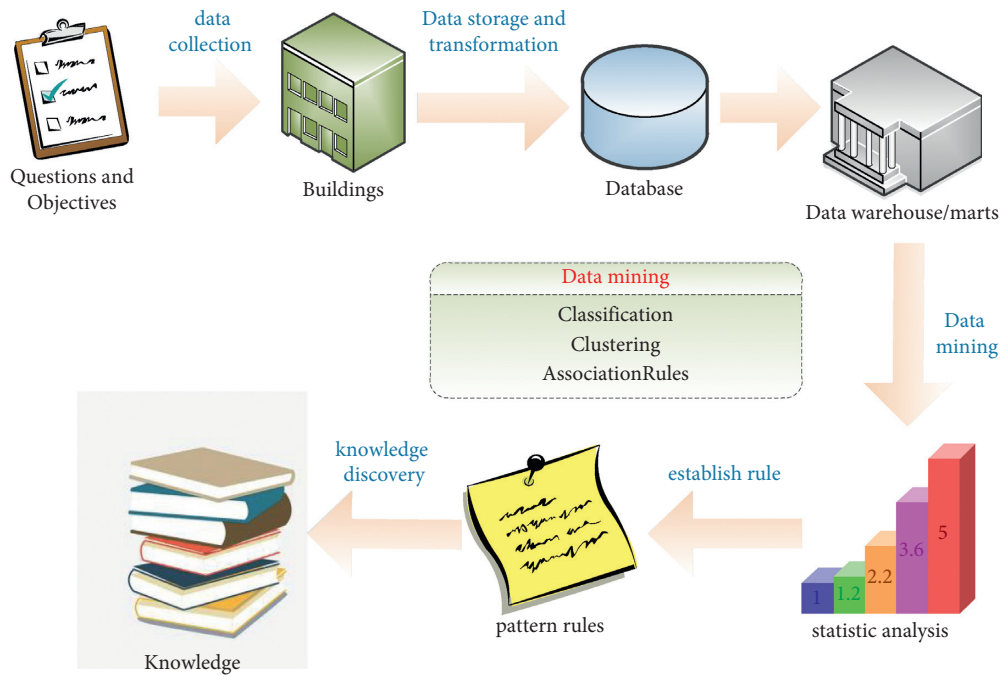


FIGURE 1: The basic process of data mining application in the field of construction.

provide data basis for the research in this field and foundation. The fusion of data mining technology is to provide computational tools and methods for research, by using data mining technology, unearth the operation law of the HVAC system, and promote the progress of research and scientific development. As mentioned in Section 2, various literature evaluations on the uses of data mining techniques in the area of buildings have been published in the last decade. Yet, in the realm of building energy systems, thorough literature evaluations on data mining methods used for load forecast, fault detection/diagnosis, and pattern recognition are still lacking. It is critical to review the findings of past studies and identify potential areas for future study. The main contributions of the proposed work are listed as follows:

- (1) Firstly, this work proposed a strategy for addressing the efficacy of traditional building energy throughout the household building process by using a benchmark evaluation approach. Here we optimized the planning and run a clustering algorithm on the overall data.
- (2) Secondly, we used three-dimensional (3D) rendering technique to develop a dynamical data analysis model for energy control. Besides, we have used the system for the adaptation construction of traditional building energy.
- (3) Thirdly, data mining methods have been widely utilized to identify actual values of huge quantities of building information, in order to improve the overall performance of building energy systems.
- (4) Finally, different evaluation methods are used to confirm that the data are evaluated using a mathematical statistics technique and a data-mining algorithm in order to improve the building energy systems.

The remainder of the paper is organised as follows: Section 2 shows a review of relevant work; Section 3 shows our benchmark evaluation model and an enhanced data mining algorithm; and Section 4 shows results obtained and discussion. Finally, Section 5 of our paper brings us to a conclusion.

2. Related Work

Identifying hidden energy consumption features in energy consumption data is a critical step in achieving building energy savings. The association mining analysis technique is used to analyse a large number of building energy consumption data, using a simple and progressive analysis procedure. Relevant scholars utilized the Apriori algorithm mining method to build strong association rules, from which they discovered the air conditioning system's unreasonable operational problems before and after the noon break. Then, targeted improvement measures were taken to improve the operating efficiency of the air conditioning system, achieve effective energy saving in building air conditioning, and promote the pace of building energy construction. The building energy consumption benchmark can be determined by comparing it with other buildings of the same type or with the energy consumption of its own history [18].

Many authors have performed depth research on it, and the three primary evaluation tools are as follows: constructing a score evaluation method, a simulation analysis method, and a statistical analysis method are all methods that may be used to evaluate a score [19]. For example, literature [20] and literature [21] respectively used cluster analysis to classify commercial buildings and hotel buildings. This method can classify buildings by considering the

shadow response theory of multiple characteristic parameters on energy consumption at the same time. However, its main limitation is that the influence of different characteristic parameters on classification is ignored, and it is simply treated as the same, which will inevitably lead to a large error of classification results. In addition, residential buildings were not involved in the above research. Considering the particularity of the composition and influencing factors of residential building energy consumption, it is necessary to conduct a separate study on the benchmark evaluation method. At present, many scholars have developed a large number of research works on energy consumption anomaly measurement.

The authors in [22] proposed a real-time monitoring method of building energy consumption based on data mining technology. By combining the DBSCAN algorithm with the classification method, the building consumption value was extracted by category and the new generation energy consumption value was identified as the category, to judge whether it was an abnormal value. Bourdeau [23] improves the modified Z-score algorithm based on GESD, which can reflect the dispersion degree of outlier data while detecting outliers and is suitable for the detection of building energy consumption data. Although these methods can detect the abnormal building energy consumption data, when the spatial density distribution of samples is not uniform or the class spacing is very different, the detection results will show deviation, and the energy consumption data cannot be processed quickly [24].

To summarise, when building equipment and energy consumption increase, data on building energy consumption must be quantified. The key development direction that the construction industry has to pay attention to is how to utilize data mining technology to find valuable data information from large-scale data and give data reference for building energy conservation. The application of data mining technologies in building energy saving will become more prevalent as the technology advances. Inspired from the work of above, this research work presents an evaluation model for building energy consumption based on traditional benchmark.

3. Traditional Benchmark Evaluation Model of Building Energy Consumption

3.1. Determination of Subitem Energy Consumption Benchmark. In this section, in certain circumstances, the energy consumption index values for each unit area of the building obtained through calculation and analysis are determined, but the distribution of energy consumption of each building is reasonable given the reasonable definition of the conditions of use in the research process. Building energy consumption levels may be fairly managed using the energy consumption index per unit area to satisfy the purpose of building service. Based on the index of energy consumption per unit area, at the same time, the average level is determined by fractal levels at 25 percent, 50 percent, and 75 percent. As a result, the related median is chosen as the energy consumption base value in addition to the overall

building energy consumption. Figure 2 depicts the energy-saving ratio derived using energy consumption simulation, as well as the associated score.

Because a building's total energy consumption comprises total electricity, total heat, total gas, and different subenergy consumption, the overall energy consumption of a structure is rather high. There are several contributing elements, and mining characteristic data and constructing a model from a large number of building energy consumption data is simple. The grading end performance assessment technique of calculating a building's energy consumption uses the grading evaluation method and examines the data provided by each grade in a step-by-step manner. Building's energy consumption assessment approach is used extensively in green building assessment systems [25].

The energy benchmark of office buildings requires a benchmark model and data processing to be determined. The following equation is the office building model of multiple linear regressions:

$$EUI = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_2 x_2 + \dots + \beta_k x_k^k, \quad (1)$$

where β_0 is regression constant and R_i denotes the regression coefficient.

The data will be initially screened and the viability of the regression model may be confirmed by ANOVA, multiple correlation coefficients (R), etc., depending on the actual circumstances of the sample data. Among these, variance analysis is also known as variance analysis and F-test analysis. The aim is to determine if the total average of two or more data groups is equal or not and to assess whether there is statistically significant or nonstatistical differentiation of two or more sample media. Thus, R is referred to as a multivariate coefficient of correlation:

$$R^2 = \frac{\sum (\hat{y} - y_0)^2 - \bar{y}}{\sum (y - y_0)^2 - \bar{y}}, \quad (2)$$

where R^2 is the efficiency of the model and the summarised regression is shown in Figure 3. Through a comparative analysis of benchmark energy usage, the building owner or management will understand how the building functions and analyse the energy consumption difference between the building and other comparable structures. If the energy consumption of similar buildings is determined to be greater, appropriate actions can be made to minimize the energy consumption.

The appropriate variables were defined using the same technique. As indicated in Table 1, X_1 represents the building area (numerical variable), X_2 represents the end form of air conditioning, X_3 represents the kind of heat source, X_4 represents the type of cold source, X_5 represents the type of glass, and X_6 represents the building structure.

3.2. Traditional Buildings Energy Consumption System. The energy consumption system consists largely of many linked energy consumption devices. During the inquiry process, data from the direct and indirect components of the energy consumption system are gathered. General status of

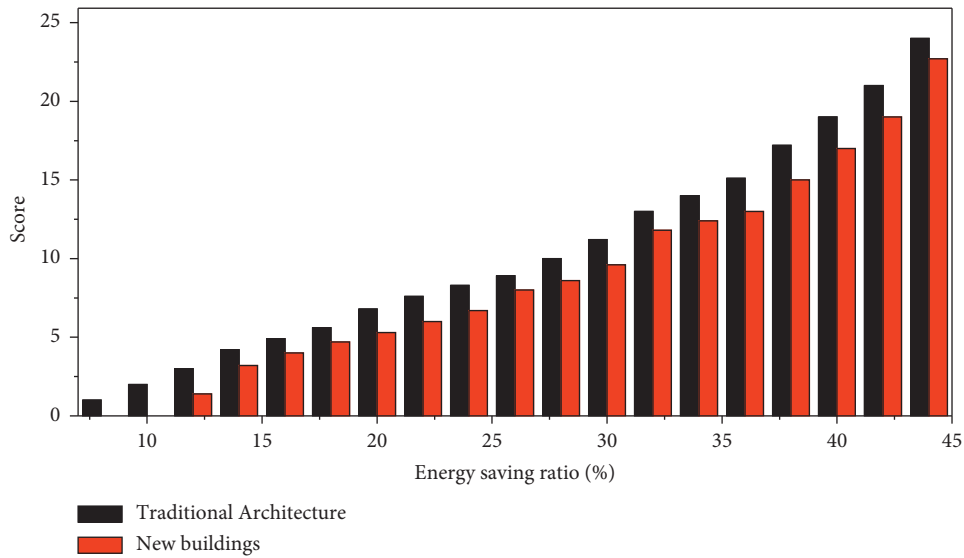


FIGURE 2: Energy-saving ratio and corresponding score.

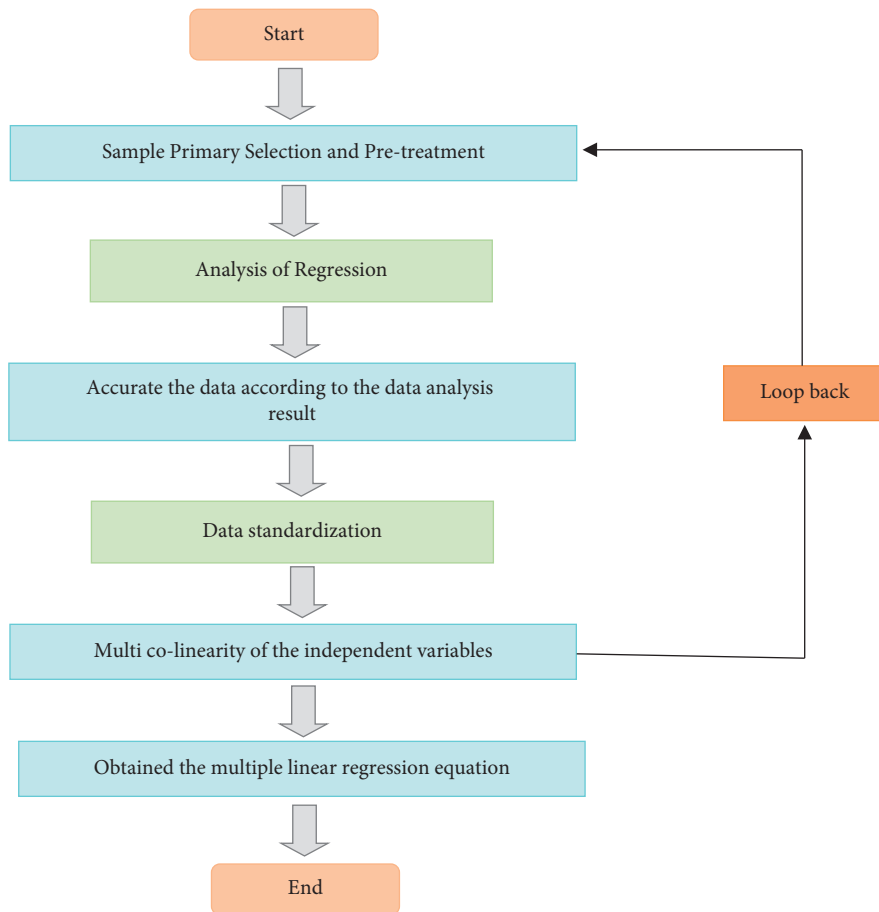


FIGURE 3: Flowchart of regression calculation.

the general details, structure of buildings of palisades, composition of the retaining structures, information on the energy use of the equipment (including air conditioning, lights, and other power equipment) and operating status of

the buildings, total annual energy, electricity, energy from month to month, and management data are carefully examined. The acquired data were examined using the mathematical statistics approach. Figure 4 depicts an office

TABLE 1: Variable number definition.

S no.	Air conditioning end from X2	Heat source from X3	Cool source from X3	Type of glass
1	Fan coil	Municipal water	Water chilling unit	Low-e glass
2	Radiator	Ground source heat pump	Ground source heat pump	General coated glass
3	Cold runner	Direct-fired machine	Direct-fired machine	Common glass
4	Automatic wind	—	—	—

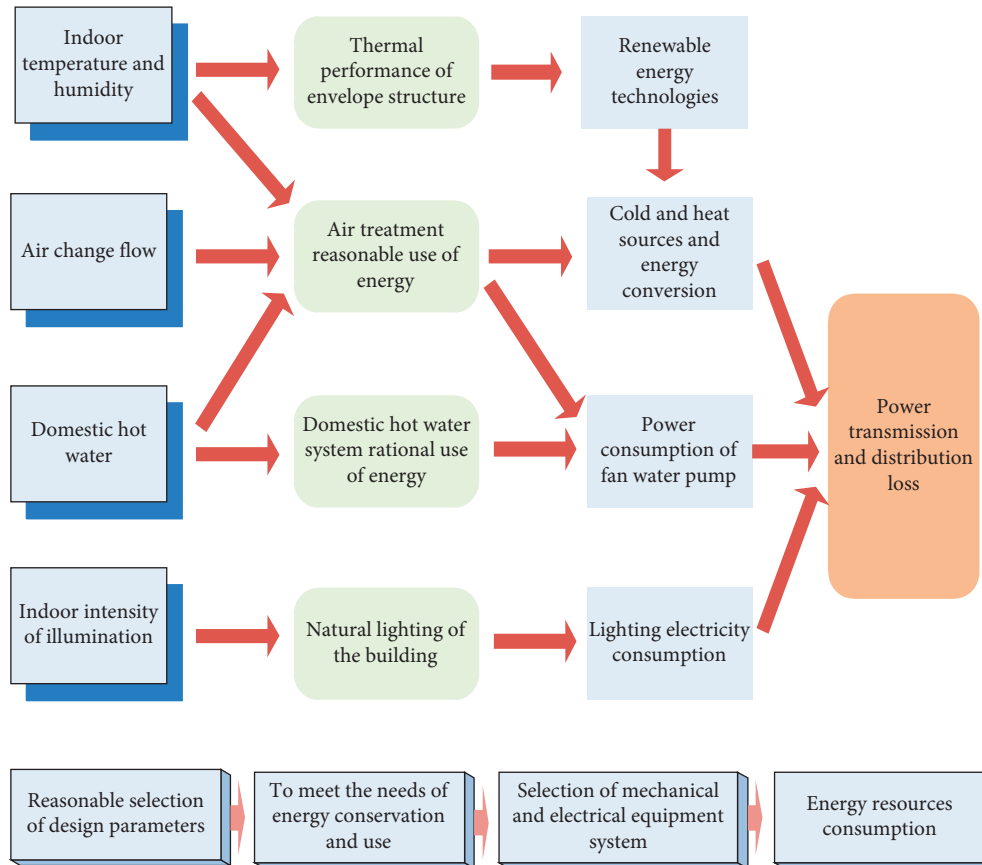


FIGURE 4: The main energy flow of an office building.

building’s primary energy consumption process. Electric electricity, natural gas, heat, and steam, among other energy sources, are utilized by the structure.

Solar photovoltaic systems and photo-thermal systems are the most common forms of renewable energy. Solar energy systems are used in 8 buildings, with 32 percent of those installed after 2005. Only three of the eight buildings employ both photovoltaic and photo-thermal systems at the same time, with the remainder relying solely on photo-thermal. It can be observed that the overall solar energy usage ratio is low, and the photovoltaic system utilization ratio is much worse. Whether the building energy consumption level can be effectively analysed mainly depends on the availability of basic building information and the accuracy and completeness of continuous energy consumption data. The operational data of the system and equipment of the selected study samples are given by property management and energy firms or collected by on-site real-time measurement and transcription. In addition, the effective energy consumption in the data only comprises the

energy consumed by the structure to maintain its function, omitting the power, gas, and water consumed by the kitchen or special room.

3.3. *Data Mining Algorithm.* Cluster analysis rules mining associations and decision tree techniques are the most used approaches for data mining. The two first are used mostly for descriptive data mining, the last being used for predictive data mining purposes. In this article, an algorithm for association regulation mining is mostly utilized. The transaction database is the topic of the common association rules for mining, whereas the training dataset is the subject of a benchmark assessment of building energy use. Because the parameters in the database have varying generalities, it is difficult to relate the row comparison and analysis to them. It is tough to connect the two if the air conditioning operational status is a classification property (on or off) and the building surface product is a numerical value attribute.

The multiple states of multivariate variables, such as the four LEED certification grades for green construction in the United States (certification, silver, gold, and platinum), must be sorted and transformed to the interval $[0, 1]$ according to

$$x_i = \frac{\text{rank}_i - x_0}{\text{rank}_{\max} - x_0}, \quad (3)$$

where x_i is the transition value of a certain state; rank_i is the sorted value of the state; rank_{\max} is the maximum sorted value for all states. Calculate the grey correlation degree γ of y_0 and y_i , and the calculation equation is

$$\gamma(y_0, y_i) = \frac{1}{n} \sum_{k=1}^n [\xi_i(k) - \gamma_0]^2, \quad (4)$$

$$\xi_i(k) = \frac{\min_k \min_i |y_i(k) - 1| - \alpha \min_k \min_i |y_i(k) - 1|}{\alpha \min_k \min_i |y_i(k) - y_0(k)| + y_0(k)}.$$

The grey correlation order of each variable is obtained by sorting according to the grey correlation degree.

3.4. Decision Tree. A decision tree is a supervised learning method that can handle both discrete and continuous data [26]. It divides the dataset into subgroups based on the dataset's most important attribute. The algorithms determine how the decision tree recognizes and divides this characteristic. The most important predictor is the root node, which is divided into decision nodes and terminal or leaf nodes that do not divide further. In the decision tree, the dataset is divided into homogenous and nonoverlapping areas. It employs a top-down method, with the top area showing all of the observations in one location before separating into two or more branches, each of which then splits further. This strategy is also regarded as a greedy approach since it just analyses the current node between the worked on without focusing on the future nodes:

- (1) Produce a decision tree from the training tuples of data divider D
- (2) Data divider, D , which is a set of training tuples and their related class labels
- (3) *attribute_list*, the set of candidate attributes

Our proposed attribute selection approach determines the splitting criterion that best divides the data tuples into separate classes. This criterion includes attribute splitting as well as the possibility of a split point. Algorithm 1 shows the phases of our recommended Decision Tree algorithm after splitting the subset.

3.5. Association Rule Mining Algorithm. An association rule is a form of inference statement: $X \longrightarrow Y$, with being disjoint item-sets: $X \cap Y = \emptyset$. The strength of an association may be determined by the amount of support and faith it has. Support determines the frequency with which a rule applies to a certain data collection, but confidence determines the

frequency with which items in Y appear in transactions containing X [27]. These metrics can be seen in equation (5) and in equation (6):

$$\text{Support}, s(X \longrightarrow Y) = \frac{\alpha(X \cup Y)}{N}, \quad (5)$$

$$\text{Confidence}, c(X \longrightarrow Y) = \frac{\alpha(X \cup Y)}{\alpha(X)}. \quad (6)$$

3.6. Apriori Algorithm. Association rules are the most frequent data mining technique. The Apriori algorithm is a well-known method for mining association rules. Many strategies have been developed concerning the rules of mining associations and associated mutations, which depend on the Apriori algorithm. The huge quantity of candidate 2 item-sets and the inefficiency with which they are tallied are the two obstacles of frequent item-sets mining. One superfluous C2 pruning operation is eliminated using the suggested method. If the number of frequently litem-sets is n and pruning operations are C_n , the number of linked candidate 2 item-sets is C_n . Candidate 2 item-set pruning procedures are reduced using the recommended approach, saving time, and increasing efficiency. The suggested technique leverages the transaction tag to improve subset operations and speed up support computations, which addresses the bottleneck of inefficient support counts. Algorithm 2 shows the phases of our recommended Decision Tree algorithm.

4. Results and Discussion

This portion of the article covers the experiments conducted and the simulation findings produced by the research. Many simulation tests on traditional building energy usage were conducted using an improved Apriori algorithm and a combined decision tree procedure. For household use, we have designed a model to save energy. The classification of traditional building energy consumption has a set of requirements (laptop computer) that can be seen in Table 2.

4.1. Characteristics of Total and Subitem Energy Consumption of Traditional Buildings. To determine the energy consumption value of building reference, information may be provided by analysing and comparing the features of overall energy consumption and the subenergy consumption of homes in traditional structures. It may also better comprehend the energy use of buildings and their energy-saving potential and clarify the emphasis of home saving energy. The average total yearly energy usage of four major household types and the proportions of each kind of home are shown in Figure 5.

The above figure shows that homes in Category 1 have the greatest overall average energy usage, whilst those in category 3 have the least. Of the eight categories of electricity consumed heating, air conditioning, and hot household water, the total energy consumptions of the four types of households account for over 20 percent and the total of two

```

(1) Make a node  $N$ ;
(2) If all of the tuples in  $D$  belong to the same class,  $C$  then;
(3)  $N$  will be returned as a leaf node with the class  $C$ ;
(4) if  $attribut\_list$  is unfilled then;
(5) return  $N$  by way of a leaf node categorized with the majority class in  $D$ ; //majority elective
(6) apply  $attribute\_selection\_method(D, attribut\_list)$  to discovery the finest  $splitting\_criterion$ ;
(7) tag node  $N$  with  $splitting\_criterion$ ;
(8) if  $splitting\_attribute$  is discrete-valued then; //not limited to binary-trees
(9)  $attribut\_list \leftarrow attribut\_list - splitting\_attribute$ ; //eliminate  $splitting\_attribute$ 
(10) for each outcomes  $j$  of  $splitting\_criterion$ 
    //divider the tuples and produce sub-trees for each divider
(11) let  $D_j$  be the set of data tuples in  $D$  filling the results  $j$ ; //a divider
(12) if  $D_j$  is unfilled then;
(13) assign a leaf labeled with the majority class in  $D$  to node  $N$ ;
(14) else assign the node returned by  $Generate\_Decision\_tree(D_j, attribut\_list)$  to node  $N$ ;
    end for
(15) return  $N$ ;

```

ALGORITHM 1: Decision Tree.

```

(1)  $K = 1$ 
(2)  $F_k = \{i \mid i \in I \wedge \sigma(\{i\}) \geq N * \min\ sup\}$ 
    {Find all frequent 1 item-sets}
(3) Repeat
(4)  $k = k + 1$ 
(5)  $C_k = \text{Apriori-gen}(F_{k-1})$ 
    {Make candidate items}
(6) for each transaction  $t \in T$  do
(7)  $C_t = \text{subset}(C_k, t)$ 
(8) {Identify all candidates that fit to  $t$ }
(9) for each candidate item-set
(10)  $c \in C_t$  do
(11)  $\sigma(c) = \sigma(c) + 1$ 
    {Increase support count}
(12) end for
(13) end for

```

ALGORITHM 2: An improved Apriori algorithm for mining association rules.

for over 60 percent. Consequently, they are “large customers”, and they should be the emphasis for the conservation of home energy. Among the remaining 6 categories of energy consumption, lighting, kitchen, and refrigeration accounted for a relatively large proportion (the third category of household lighting energy consumption is the smallest and lower than other terminal energy consumption, which may be due to their strong awareness of energy saving and the use of energy-saving lamps in lighting equipment). The energy consumption of lighting and the kitchen, in addition to refrigerators, varies substantially amongst houses. This is because refrigeration equipment runs constantly for long periods and is less affected by user behavior.

Figure 5 further shows that despite having the lowest outside annual average temperature and the highest wind speed, Category 1 households consume substantially less heating energy than Category 4 households. One reason might be that their enclosure construction provides superior heat insulation. The number of homes has a significant impact on the amount of

hot water energy used in the household. It is difficult to assume that the greatest number of homes will always result in the greatest residential hot water energy usage (for example, Category 4 has the largest number of households, but not the highest domestic hot water energy consumption).

Figure 6 shows the cumulative frequency distribution curve of the total energy consumption of building, heating, conditioning, and domestic hot water of Category 1 households. The energy consumption values corresponding to the cumulative frequency of 50% and 25% are taken as the reference value and target value. As can be seen from Figure 6, the base value of Category 1 household TEC is 391 MJ/m²/year and the nominal value is 305 MJ/m²/year. Subitem energy consumption index values for households may be calculated by using the cumulative frequency distribution curves of HC and HWS. If the baseline for the same sample is 50 percent of the water level, therefore the results are 111 MJ/m²/year and 127 MJ/m²/year. In the same way, the TEC reference and target values of the other three categories of households can be obtained.

TABLE 2: Hardware requirements for the experimental work.

S No	Requirements	Specification
1	System core	Core-i7
2	Generation	3 rd
3	CPU	2.9 GHz
4	RAM	16 GB
5	Hard drive	1TB
6	Operating system	Windows 10, 64 bits

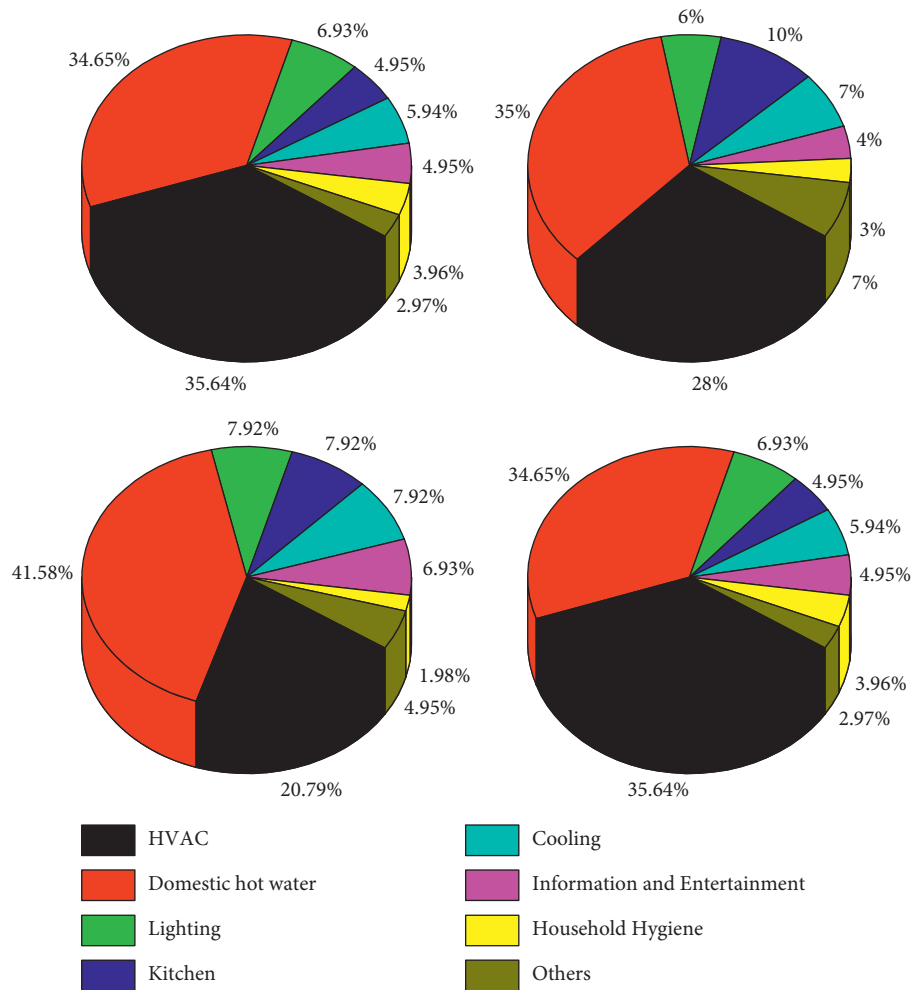


FIGURE 5: Energy consumption characteristics of 4 types of households.

4.2. Overall Evaluation of Building Energy Consumption. The household with the lowest energy usage and the most similar features to the household can be contrasted with similar homes (the similarity of the characteristics of the two households indicates that the two households have the most comparable reference value based on all characteristic parameters). The distance between the two may be used to symbolize their resemblance; the smaller the gap, the higher the similarity. For instance, the energy use evaluation of A in the first category of households is carried out. By calculating the similarity between other households and A, it can be seen that B is the household with the most similar characteristics. The detailed characteristic parameters (T (Total energy

expenditure), WS (Watt Second), RH (Relative Humidity), RA (Resource Adequacy), HLC (Heat Loss Coefficient), ELA (Electrical Load Analysis), and HT (High Tension)) of the two households are shown in Table 3.

In increasing order, Figure 7 depicts the total energy use of all buildings in Category 1 households. As seen in this figure, House A consumes more energy than the baseline, and a structure that consumes a lot of energy is classified as a “nonenergy-saving building”. House B consumes less energy than the goal value; therefore, a structure with low energy consumption qualifies as an outstanding “energy-saving building”. A household’s annual energy-saving potential is $548-391 = 157 \text{ MJ/m}^2$ compared to the base value of energy

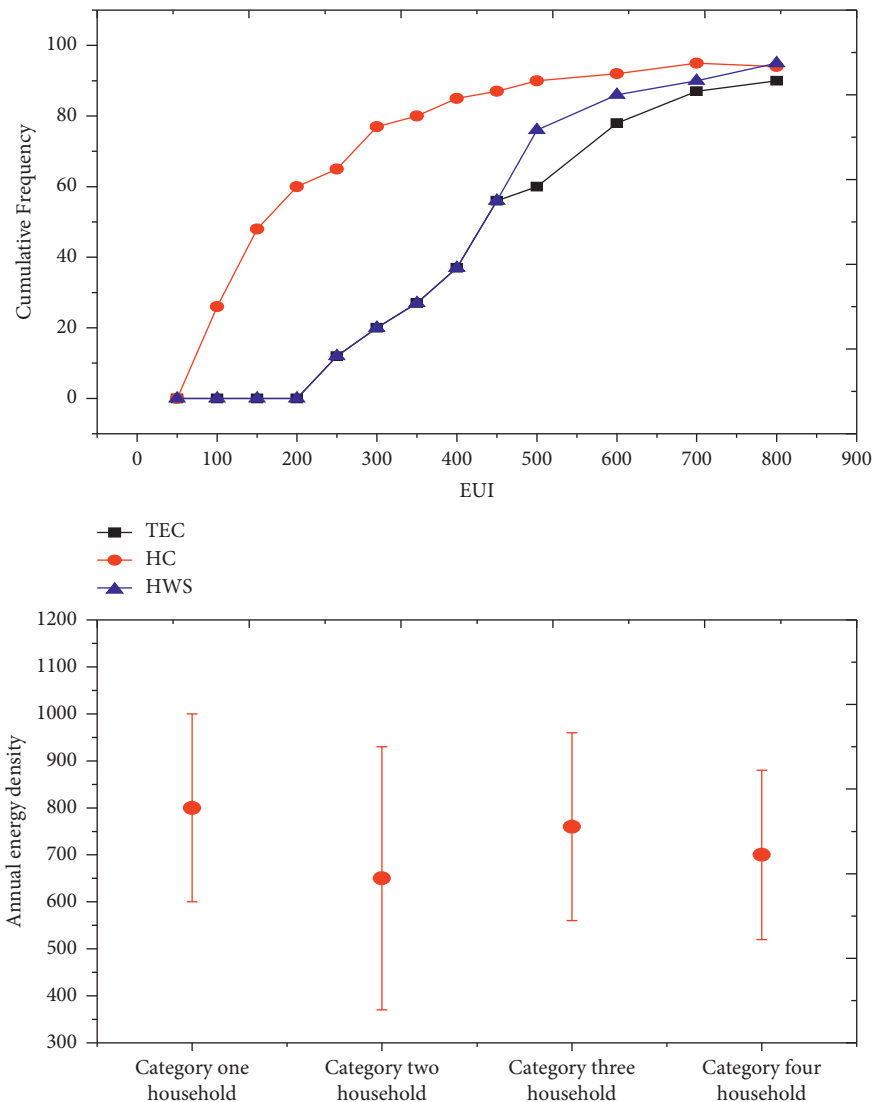


FIGURE 6: The cumulative frequency distribution curve of total energy consumption.

consumption; household B's annual energy-saving potential is $548 - 304 = 244$ MJ/m². When the building features of the two households are compared, it can be seen that the heat loss coefficient and equivalent leakage area of the envelope structure are significantly different, which might be one of the causes for the first household's high energy consumption. Consequently, through energy-saving transformation, a home may reduce building energy consumption by enhancing the thermal insulation performance of the outer structure and the airtightness of doors and windows.

4.3. Abnormal Detections of Energy Consumption of Traditional Buildings. This paper uses daily itemized energy consumption data (air conditioning and lighting) from a shopping center in the second quarter of 2020 (i.e., June to August, a total of 92 days) to perform irregular and practical testing on the data. Figure 8 shows the energy usage of air conditioning and lights. Before the actual test, we have analysed the energy consumption data used in our

experimental work. This is because data from the direct connection source frequently are incomplete and inconsistent and may have a significant impact on the effect of data excavation. To ensure the accuracy and reliability of data, the preanalysis of energy data consumption is necessary for the experiment.

The line chart of electricity consumption data obtained after marking the abnormal energy consumption is based on the abnormal detection results (in which the triangular and rhomboid marks represent the detected abnormal energy consumption data and uncertain data of air conditioning electricity, respectively). The round and square marks, respectively, reflect anomalous energy usage data and questionable lighting data that have been discovered. Through the abnormal energy consumption detection model of the MP algorithm, the abnormal value in the data of building energy consumption can effectively be detected and obtained. This can provide necessary help for the management and operation of the construction energy consumption monitoring system.

TABLE 3: Characteristic parameters of households A and B.

Household	T	WS	RH	RA	NO	HLC	ELA	HT
A	9.0	4.2	70	13.1	3	3.08	0.99	0
B	8.9	4.0	62	11.1	3	2.30	0.87	0

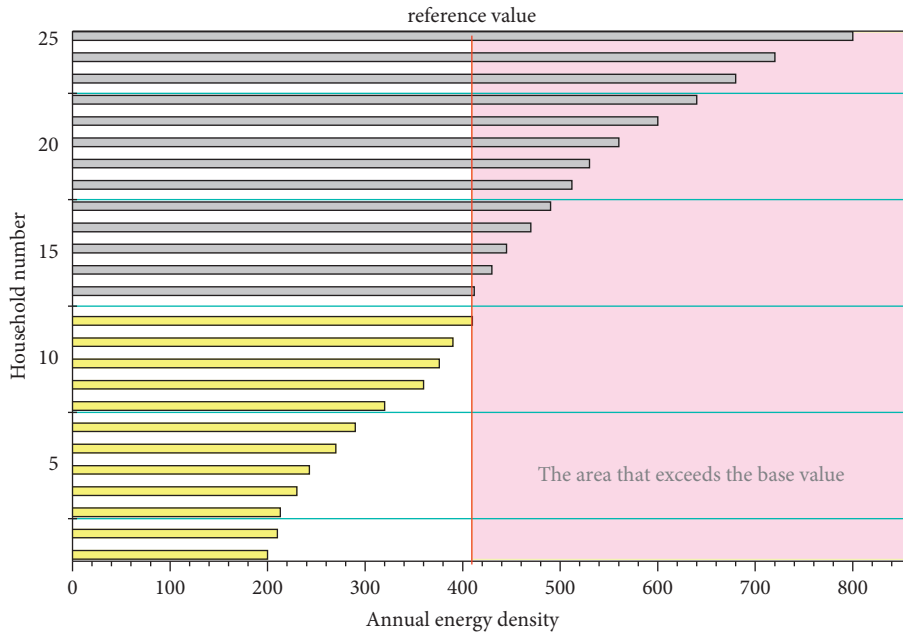


FIGURE 7: Total building energy consumption of Category 1 households.

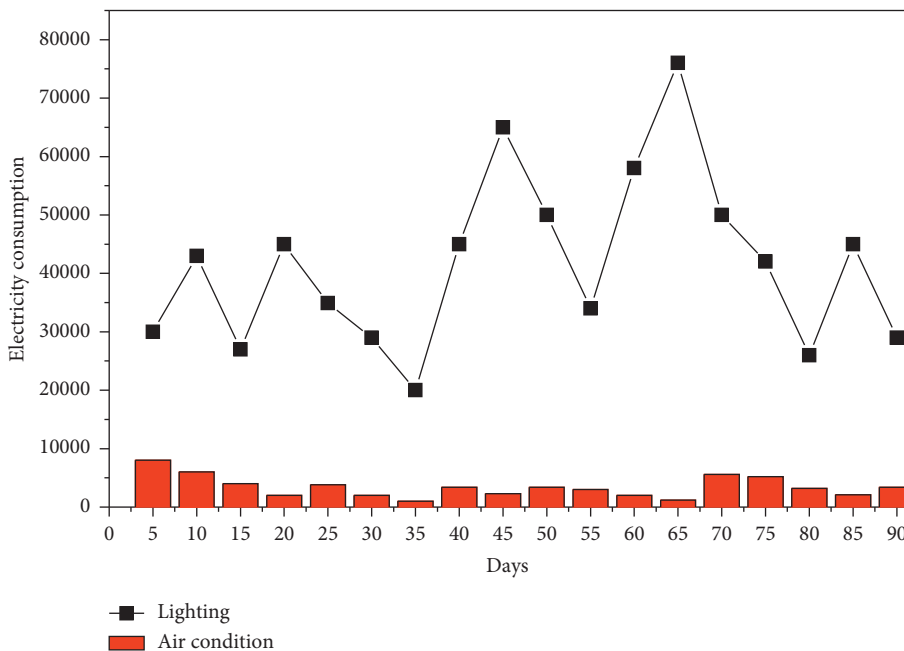


FIGURE 8: Data of itemized energy consumption (electricity consumption for air conditioning and lighting) of shopping malls in the second quarter of 2020.

Figure 9 illustrates the comparisons of parameters used for household A and household B. The total energy expenditure (T) and High Tension (HT) of both the households are the same. However, the Watt Second (WS),

Relative Humidity (RH), Resource Adequacy (RA), Heat Loss Coefficient (HLC), and Electrical Load Analysis (ELA) of household A are 0.2, 8, 2, 0.78, and 0.12, respectively, greater than those of household B.

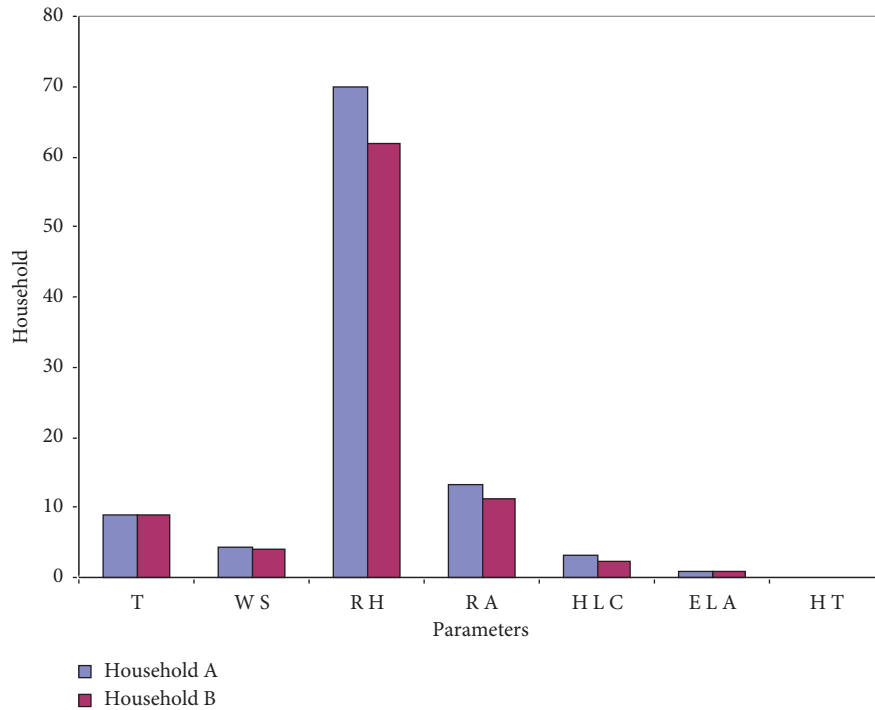


FIGURE 9: Comparison of household A and household B.

5. Conclusion

In this paper, the traditional buildings are classified, and the building energy consumption reference value is determined and evaluated using data mining techniques. The classification of buildings may be fine-tuned with this technique, and the efficiency of building energy consumption reference values can be enhanced. At the same time, a realistic assessment of a building's energy consumption level requires that the structures being compared have a high degree of similarity. The benchmark energy consumption evaluation method proposed in this paper uses grey correlation analysis to determine the degree of correlation between different influencing factors and building energy consumption then uses the correlation degree as the weight of the factors to classify buildings reasonably using cluster analysis. Our proposed approach can evaluate a typical household energy consumption characteristics and energy-saving potential and also make energy-saving recommendations. The reliability of the energy-saving potential obtained and the feasibility of energy-saving recommendations are both high when comparing and evaluating houses whose energy consumption is lower than the base value and whose characteristics are most comparable to those of households in the same category. While evaluating the benchmark energy consumption, this approach may give a lot of information on building energy efficiency.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This study was supported by 2018 Jiangsu Postgraduate Research and Innovation Project, "Research on the Spatial Evolution and Social Changes of Residential Buildings in Central Jiangsu Since Ming and Qing Dynasty," Project Number: KYCX18_1877.

References

- [1] Y. Chen, H. Tan, and U. Berardi, "A data-driven approach for building energy benchmarking using the Lorenz curve," *Energy and Buildings*, vol. 169, pp. 319–331, 2018.
- [2] C. Xu and H. Chen, "A hybrid data mining approach for anomaly detection and evaluation in residential buildings energy data," *Energy and Buildings*, vol. 215, Article ID 109864, 2020.
- [3] J. Liu, J. Wang, G. Li, H. Chen, L. Shen, and L. Xing, "Evaluation of the energy performance of variable refrigerant flow systems using dynamic energy benchmarks based on data mining techniques," *Applied Energy*, vol. 208, pp. 522–539, 2017.
- [4] C.-F. Jeffrey Kuo, C.-H. Lin, and M.-H. Lee, "Analyze the energy consumption characteristics and affecting factors of Taiwan," *Energy and Buildings*, vol. 168, pp. 120–136, 2018.
- [5] Iea, "Energy efficiency: buildings," 2018, <https://www.iea.org/topics/energyefficiency/buildings/>.
- [6] T. Hong, C. Koo, J. Kim, M. Lee, and K. Jeong, "A review on sustainable construction management strategies for

- monitoring, diagnosing, and retrofitting the building's," *Applied Energy*, vol. 155, pp. 671–707, 2015.
- [7] E. O'Dwyer, I. Pan, S. Acha, and N. Shah, "Smart energy systems for sustainable smart cities: current developments, trends and future directions," *Applied Energy*, vol. 237, pp. 581–597, 2019.
 - [8] T. Ahmad, H. Chen, J. Shair, and C. Xu, "Deployment of data-mining short and medium-term horizon cooling load forecasting models for building energy optimization and management," *International Journal of Refrigeration*, vol. 98, pp. 399–409, 2019.
 - [9] X. Yu, Q. Peng, L. Xu, F. Jiang, J. Du, and D. Gong, "A selective ensemble learning based two-sided cross-domain collaborative filtering algorithm," *Information Processing & Management*, vol. 58, Article ID 102691, 2021.
 - [10] Y. Chen and J. Wu, "Distribution patterns of energy consumed in classified public buildings through the data mining process," *Applied Energy*, vol. 226, pp. 240–251, 2018.
 - [11] R. Wang, S. Lu, and W. Feng, "A novel improved model for building energy consumption prediction based on model integration[J]," *Applied Energy*, vol. 262, Article ID 114561, 2020.
 - [12] J. Leprince, C. Miller, and W. Zeiler, "Data mining cubes for buildings, a generic framework for multidimensional analytics of building performance data," *Energy and Buildings*, vol. 248, Article ID 111195, 2021.
 - [13] X. Yu, Y. Chu, F. Jiang, Y. Guo, and D. Gong, "SVMs classification based two-side cross domain collaborative filtering by inferring intrinsic user and item features," *Knowledge-Based Systems*, vol. 141, pp. 80–91, 2018.
 - [14] S. Seyedzadeh, F. P. Rahimian, I. Glesk, and R. Marc, "Machine learning for estimation of building energy consumption and performance: a review," *Visualization in Engineering*, vol. 6, no. 1, pp. 1–20, 2018.
 - [15] E. Sharma, "Energy forecasting based on predictive data mining techniques in smart energy grids," *Energy Informatics*, vol. 1, no. 1, pp. 367–373, 2018.
 - [16] I. Ridwana, N. Nassif, and W. Choi, "Modeling of building energy consumption by integrating regression analysis and artificial neural network with data classification," *Buildings*, vol. 10, no. 11, p. 198, 2020.
 - [17] Y. Zhao, C. Zhang, Y. Zhang, Z. Wang, and J. Li, "A review of data mining technologies in building energy systems: load prediction, pattern identification, fault detection and diagnosis," *Energy and Built Environment*, vol. 1, no. 2, pp. 149–164, 2020.
 - [18] Y. S. Park, N. M. A. Mohamed Abdul Ghani, F. Gebremikael, and G. Egilmez, "Benchmarking environmental efficiency of ports using data mining and RDEA: the case of a U.S. container ports," *International Journal of Logistics Research and Applications*, vol. 22, no. 2, pp. 172–187, 2019.
 - [19] M. Yu, J. Jin, X. Wang, X. Yu, D. Zhan, and J. Gao, "Development and design of flexible sensors used in pressure-monitoring sports pants for human knee joints," *IEEE Sensors Journal*, vol. 12, p. 1, 2021.
 - [20] Y. Wei, X. Zhang, Y. Shi et al., "A review of data-driven approaches for prediction and classification of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1027–1047, 2018.
 - [21] G. Liu, J. Yang, Y. Hao, and Y. Zhang, "Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering," *Journal of Cleaner Production*, vol. 183, pp. 304–314, 2018.
 - [22] R. Hafezi, A. N. Akhavan, M. Zamani, S. Pakseresht, and S. Shamshirband, "Developing a data mining based model to extract predictor factors in energy systems: application of global natural gas demand," *Energies*, vol. 12, no. 21, p. 4124, 2019.
 - [23] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, G. Xiaofeng, and C. Patrice, "Modeling and forecasting building energy consumption: a review of data-driven techniques," *Sustainable Cities and Society*, vol. 48, Article ID 101533, 2019.
 - [24] E. García-Martín, C. F. Rodrigues, G. Riley, and H. Grahm, "Estimation of energy consumption in machine learning," *Journal of Parallel and Distributed Computing*, vol. 134, pp. 75–88, 2019.
 - [25] G. Li, Y. Hu, H. Chen et al., "Data partitioning and association mining for identifying VRF energy consumption patterns under various part loads and refrigerant charge conditions," *Applied Energy*, vol. 185, pp. 846–861, 2017.
 - [26] L. Xuemin Zhang and X. Zhang, "Study of data mining algorithm based on decision tree," in *Proceedings of the 2010 International Conference On Computer Design and Applications*, June 2010.
 - [27] D. M. Tank, "Improved Apriori algorithm for mining association rules," *International Journal of Information Technology and Computer Science*, vol. 6, no. 7, pp. 15–23, 2014.