sustainability

MDPI

Article

Rough-Set-Based Rule Induction with the Elimination of Outdated Big Data: Case of Renewable Energy Equipment Promotion

Chun-Che Huang [1], Wen-Yau Liang [2,*], Roger R. Gung [3] and Pei-An Wang [1]

[1] Department of Information Management, National Chi Nan University, 1 University Road, Nantou 545, Taiwan
[2] Department of Information Management, National Changhua University of Education, 1, Jin-De Road, Changhua 511, Taiwan
[3] Department of Business Analytics & Operations Research, University of Phoenix, 4025 S Riverpoint Parkway, Phoenix, AZ 85040, USA
* Correspondence: wyliang@cc.ncue.edu.tw

Abstract: As developing economies become more industrialized, the energy problem has become a major challenge in the twenty-first century. Countries around the world have been developing renewable energy to meet the Sustainable Development Goals (SDGs) of the United Nations (UN) and the 26th UN Climate Change Conference of the Parties (COP26). Leaders of enterprises have been made aware of the need to protect the environment and have been practicing environmental marketing strategies and green information systems (GISs) as part of ESG practices. With the rapid growth of the available data from renewable electricity suppliers, the analyses of multi-attribute characteristics across different fields of studies use data mining to obtain viable rule induction and achieve adaptive management. Rough set theory is an appropriate method for multi-attribute classification and rule induction. Nevertheless, past studies for Big Data analytics have tended to focus on incremental algorithms for dynamic databases. This study entails rough set theory from the perspective of the decrement decay alternative rule-extraction algorithm (DAREA) to explore rule induction and present case evidence with managerial implications for the emerging renewable energy industry. This study innovates rough set research to handle data deletion in a Big Data system and promotes renewable energy with valued managerial implications.

Keywords: rough set theory; Big Data; renewable energy; decrement algorithm

check for updates

Citation: Huang, C.-C.; Liang, W.-Y.; Gung, R.R.; Wang, P.-A. Rough-Set-Based Rule Induction with the Elimination of Outdated Big Data: Case of Renewable Energy Equipment Promotion. *Sustainability* **2023**, *15*, 14984. https://doi.org/10.3390/su152014984

Academic Editors: Eduard Petlenkov and Larbi Chrifi-Alaoui

Received: 6 September 2023
Revised: 9 October 2023
Accepted: 10 October 2023
Published: 17 October 2023

1. Introduction

Maintaining sustainable energy supplies using limited natural resources in the world has become one of the major challenges in the twenty-first century. Many countries have been actively seeking energy alternatives [1,2], and green energy engineering has been the key to this exercise [3]. However, many technical, economic, socio-cultural and institutional barriers limit the growth of renewable energy technologies [4], and grand challenges (GCs) affect organizations and institutions [5]. To achieve better energy efficiency, energy management strategies can achieve better outcomes than traditional methods [6]. The promotion of renewable energy is a key success factor that largely relies on government policies and entrepreneur coalitions [7].

The Sustainable Development Goals (SDGs) of the United Nations (UN) consist of 17 goals that were adopted by all UN Member States in 2015 that set out a 15-year plan to achieve them. Goal 7 is to ensure access to affordable, reliable, sustainable and modern energy for all [8]. Meanwhile, at the 26th UN Climate Change Conference of the Parties [9], the most important outcome was directly related to fossil fuels. Leaders of enterprises have been made aware of the need to protect the environment and have been practicing environmental marketing strategies as part of ESG practices (Environment, Social, and Government) [10]. From a political perspective and in terms of the subsequent ramifications of fundamental and critical decisions, digital and information systems technology has to play a role in finding potential solutions [11].

Information systems can help manage and save resources pertaining to operations [12]. To achieve environment sustainability, green information systems (Green ISs) have emerged as a crucial research area in order to reduce carbon footprints [13]. Over the last two decades, past data systems have been insufficient in terms of analyzing rapid environmental changes due to the explosive growth of data. The analysis of renewable energy to generate decision rules has a considerable impact, as it helps to formulate policy and strategy implementation for the promotion of renewable energy [14–16]. Information research through data analytics has therefore become a practical demand in energy and resource optimization [17]. The advent of Big Data analytics has changed existing data application development in information management. It is involved in a wide range of applications, not only in medicine, business, social analysis and public services, but also in solving the energy crisis.

There have been many studies associated with Big Data in renewable energy [18]; nonetheless, most of them have focused on renewable energy production, planning and forecasting [19–21]. This is inadequate to illustrate how to promote these applications to stakeholders in the aforementioned areas. In addition, the repeated use of data should be limited, as constant recalculations are a waste of computing resources. There have been many studies that have focused on the increment algorithm but have also ignored the occurrence of the decrement situation.

The long-term accumulation of data in today's rapidly changing society may result in massively outdated information. Previous studies have indicated that the cumulated data are often deleted. Since decrement case studies are less popular than increments, data decrement operations would be inadequate in using Big Data for renewable energy applications, which would limit the promotion of renewable energy sources. In such a situation, how to achieve operational efficiency and rule induction is not only a consideration of the elimination of outdated data, but also one that needs to result in effective decision making, which is an important topic.

The scope of renewable energy data contains a number of different types of qualitative and quantitative attributes that increase the difficulties in rule induction. When conducting analysis, standardizing and defining attributes can take up a considerable amount of time. Rough set theory is a quantitative method by which to deal with imprecise and vague concepts and to classify knowledge that has been widely applied to many types of decision-making problems and in various industries. It can potentially generate rules and use rule induction to find key rules under the condition of attribute reduction and also use the attribute value model to describe dependencies between attributes, evaluate the significances of attributes and, ultimately, lead to the decision rules.

In addition, while dealing with the current environmental situation, the impact of using Big Data in promoting renewable energy must be considered. In summary, rough set theory is suitable to analyze the problems pertaining to renewable energy. The objective of this study was to effectively analyze the renewable energy data and improve the renewable energy market by using rough set theory with a consideration of how to handle data deletion in a Big Data system, which in turn results in efficient data processing.

This study proceeds with Big Data perspectives and explores the feasibility of using rough set theory in decreasing renewable energy data through generating decision rules that are used in the assessment scheme to promote renewable energy and propose managerial implications. This paper is organized as follows: The related literature is reviewed and summarized in Section 2. The solution approach is proposed in Section 3. A case study that demonstrates the superiority of the proposed approaches is shown in Section 4. Finally, the conclusions of this study are summarized in Section 5. The novel contributions of the paper are summarized as follows:

- Considerations of how to handle data deletion in a Big Data system;
- An effective analysis of renewable energy data to improve the renewable energy market;
- The generation of decision rules used in an assessment scheme to promote renewable energy and proposed managerial implications.

2. Literature Review

2.1. Renewable Energy

From the mid-1970s onwards, many studies of the energy crisis were conducted due to oil shocks. Common solutions might effectively reduce energy consumption, including the development of eligible renewable energy and the construction and use of green energy equipment to reduce carbon emissions [22,23]. The design, development and dissemination of appropriate renewable energy technology are important not only to meet the growing energy requirements for economic growth, but also for improvement in the quality of human life [24,25].

Compared with other industries, renewable energy has the nature of public utility. Therefore, interdisciplinary cooperation with related government agencies is essential to ensure that data sources are extensive [26]. Big Data based on myriad sources, e.g., environmental and social aspects are closely related, and many different aspects of research from data analysis to identify the decision rules can be linked [27–29]. Moreover, renewable energy data compared with normal data must take non-use of the additional value of renewable energy into account for considerable inclusiveness, as the descriptive data can also help in understanding user intentions [30,31]. These data sources cross many different attributes and fields, and such multi-attribute data also have different aspects of information data. Hence, how to effectively and efficiently deal with such data becomes a critical and important step. Rough set theory for dealing with such multi-attribute and different criterion datasets has yielded significant insightful results.

In addition, timely and cost-effective analytics on Big Data has emerged as one of the key success factors for most success stories in businesses, scientific and engineering disciplines and government endeavors [32,33]. To add or remove a single data point, a conventional single increment–decrement algorithm can be used to update the model efficiently. However, to add and or remove multiple data points, the computational cost of the current update algorithm becomes inhibitive because we need to repeatedly apply it for each data point [34]. With the development of science and technology and the growing internet, the amount of information is also increasing. People often need to remove or add lots of data when companies use databases or data warehouses. Using traditional data mining methods is neither suitable nor cost-effective [35]. Especially when considering rapidly changing data, finding methods of using data mining to reduce time expenditure is the goal of many research endeavors. Many dynamic algorithms have been developed not only to maintain the accuracy of the calculation, but also to enhance the computation speed [36].

Considering the data changes in dynamic Big Data systems, the majority of studies based on data incrementing might cause serious errors and irreparable mistakes if data reduction is not taken into consideration. For this reason, this study focuses on the research of Big Data systems in the renewable energy industry based on the perspective of data reduction. Rough set theory is applied to facilitate multi-attribute decision analysis. Under the data reduction principle, this research could provide a critical supplement to the renewable energy industry where key success factors have not been identified.

2.2. Rough Set Theory

Rough set theory was proposed by Pawlak to classify imprecise, uncertain and incomplete information or knowledge expressed by data acquired from experience [37,38]. It is an improvement of the set theory used in the study of systems that are characterized by insufficient, incomplete and uncertain information [39]. Rough set theory not only possesses rule induction but also the ability to figure out critical rules under the attribute reduction condition. Additionally, it is able to use attributes in specific models to describe dependencies among attributes and assess the significance of the properties concluded by the decision rules. Moreover, rough set theory allows for an information reduction procedure based on the consequence of particular attribute subsets [40].

Attribute reduction is an important concept in rough set theory, as it can find the core problem via attribute reduction. The core idea of attribute reduction is to obtain the sub-attribute set (or feature set) that can maintain the same discriminability as the full attribute set [41]. There are many studies that indicate that rough set theory is an effective tool for data analysis [42–44]. It has been widely applied in the improvement of decision making.

In addition, market statuses may often change from time to time, leading to changes in Big Data since data are "dynamic" in the sense that users may periodically or occasionally insert or delete data in the data storage method, e.g., a database. Such real data may increase/decrease dynamically in size and this phenomenon is particularly notable in the energy domain. Morley showed that many empirical studies suffer from a lack of dynamic structure and ignore the fact that data are added [45]. Such updates to the data may cause not only the generation of new rules but also the invalidation of some existing rules [46]. As an effective and efficient mechanism to deal with such data, an in/decremental technique has been proposed in the literature and has attracted much attention, stimulating results [47]. The in/decremental technique allows new data to be added or updated data to be removed without re-implementing the algorithm in a dynamic database [48]. In the RST, all updates should lead to changes in a decision table.

To specify the domain of rough set theory, in/decremental techniques essentially reuse previously inducted information and combine this information with the freshly updated data to efficiently generate new decision rules or new reductions. For example, Liang et al. introduced incremental mechanisms and then developed a group incremental rough feature selection algorithm based on information entropy [47]. Zhang et al. proposed composite rough sets and combined them with the incremental learning technique for quickly updating approximations [49]. Luo et al. analyzed the updating mechanisms for computing approximations with the variation of the object set and proposed two incremental algorithms for updating the approximations [50]. Li et al. proposed incremental approaches and algorithms for updating approximations [51]. Zhang developed incremental approaches for updating rough approximations in IvIS under attribute generalization, which refers to the dynamic changing of attributes [52]. Shu et al. proposed a neighborhood entropy-based incremental feature selection framework via the neighborhood rough set model [53]. Yang et al. first combined incremental technology and the accelerated strategy in attribute reduction and proposed matrix-based incremental mechanisms for dynamic attribute reduction when the objects are evolved over time [54]. Huang et al. proposed an incremental feature selection approach in hierarchical classification by employing a fuzzy rough set technique with the consideration that data in the real world may arrive dynamically [55]. However, most of the previous incremental algorithms require the calculation of the decision matrix for each decision attribute, while the number of decision attributes is usually very large (e.g., in mass data information systems). This is time- and memory-consuming [56]. Therefore, Ref. [57] proposed an incremental rule-extraction algorithm based on the work of [58]. Ref. [59] proposed an incremental attribute extraction based on the work of Ref. [44]. However, both studies only discussed the object incremental issue and posited that one object is added at a time. As a result, these studies can only resolve each additional object and implement the solution approach n times. They can not handle issues when a set of n attributes is removed.

2.3. Research Gap

With the development of science, technology and the growing internet, the volume of information is increasing, driving industry to often remove and add lots of data when companies are using databases or data warehouses. Traditional rough set approaches may consider the data changes in a dynamic world based on data increment. This might cause serious errors and irreparable mistakes if data reduction is not taken into consideration. In addition, current dynamic algorithms mainly focus on the accuracy of the calculation. Therefore, algorithms to enhance agility in industry are constantly in demand.

3. Solution Approach

In this study, a data deletion method, i.e., the decrement decay alternative rule-extraction algorithm (DAREA), was applied. It is a decrement alternative rule-extraction algorithm that modifies some rules in the original rule set based on the results of renewable energy decrement rule induction analysis. The advantage of using the DAREA is to inherit the nature of the alternative rule extraction algorithm (AREA), adding the concept of weight and finding more meaningful rules in accordance with the weight. The concept of weight can be used to improve the credibility of rules and resolve the problems due to the repeatability of rules. In addition, the DAREA can also solve the problem of rule induction when data are deleted and directly concluded, as it is unnecessary to reload all databases to conduct rule induction and compare with old rules to obtain the results. This improves the overall efficiency of data management as it is difficult to accurately reflect the current situation otherwise, especially when rough sets are previously changed at a large scale.

In accordance with the current state of renewable energy social environment, a user satisfaction index from a customer relation management (CRM) perspective is considered as an optimistic direction in renewable energy promotion. It is also integrated with government data to achieve authenticity and effectiveness for renewable energy promotion. The differences in rules and potential impact of the models with and without outdated data are summarized in this study.

3.1. Customer Analysis Rules

In this study, the following three categories of attributes were identified from the data:

1. Customer characteristic attributes are customer data that enterprises apply to customer composition analysis, such as customer type, gender and annual income.
2. Government data attributes are data acquired from government agencies and integrated with customer data, such as residential electricity consumption, electricity change, regional tariff and regional payroll.
3. Customer behavior and experience attributes are customer data based on customer feedback and the evaluations of their prior purchase factors.
4. For the DAREA methodology, every conditional attribute is given a weight to calculate the strength index (SI), and the third-category attributes are given higher weights. Adjusting the weights is a key process of the DAREA developed in this study. The summary of all attributes is shown in Table 1.

Customer characteristic attributes are used to classify customers. Public sector objectives are integrated with customer behavior and experience attributes and used as the feedback to improve customer satisfaction. Ultimately, customer satisfaction serves as the decision attribute for the use of rough sets to develop marketing rules and assess corporate strategies.

3.2. Rule Change

Databases often change by adding, deleting and updating data. Practical applications must take these changes into consideration. DAREA methodology is a decrement algorithm that can effectively deal with changes in data reduction. There are five different cases of object increment shown in Table 2.

Table 1. Customer value analysis framework.

	Customer Characteristic Attributes				Government Data Attributes			Customer Behavior and Experience Attributes		Output	
Object	**Gender**	**Age**	**Revenue**	…	**Kilowatt**	**Electricity Growth**	…	**Purchase Factors**	**Purchase Time**	**Satisfaction**	**Cardinality**
1	M	20	23 k	…	200	1%	…	Cost	1 year	H	20
2	F	40	50 k	…	300	3%	…	Environmental	2 years	M	15
3	M	35	100 k	…	500	−1%	…	Environmental	Over 5 years	L	5
…	…	…	…	…	…	…	…	…	…	…	…
Weight	0.1	0.2	0.5	…	0.4	0.6	…	0.7	0.8		

Table 2. Structure of case.

		Generate New Rule	Replace Old Rule	Delete Old Rule
Case 1		----	----	----
Case 2		----	----	✓
Case 3		----	✓	----
Case 4		✓	----	----
Case 5	Case 5.1	✓	✓	----
	Case 5.2	✓	----	✓
	Case 5.3	----	✓	✓
	Case 5.4	✓	✓	✓

Case 1: The increment of a new object does not cause any conflict with the original rules and the original rules can dominate this new object.

Case 2: The increment of a new object does not cause any conflict with the original rules, but the original rules cannot dominate this new object.

Case 3: The increment of a new object causes conflict with the original rules, but at least one of the original rules can dominate this new object.

Case 4: The increment of a new object causes conflict with the original rules and none of the original rules can dominate this new object.

Case 5: The increment of a new object causes conflict with the original rules and new object features are identical to one of the original objects. However, these two objects have different outputs.

3.3. Algorithms

The notation and variables used in this data reduction algorithm are presented as follows:

Algorithm procedure:

Our algorithm (Algorithm 1) is based on the theory proposed by Pawlak [60]. Three lemmas are described as follows:

Lemma 1. *If* $D(X_i)_{Aj} - n_d = \varnothing$*, then a new reduct is generated.*

Lemma 2. $Reduct_{new} \notin Reduct_{old}$*; the new reduct must be different from the previous reduct.*

Lemma 3. *If Temp = 0, then old rule set $\subseteq$ new rule set. If Temp was not changed, this means that only one rule is reduced or the rules are not affected.*

Algorithm 1: DAREA

Input: The number of a removed object, n_d.
Output: The set of decision rules and alternative rules, F_{new}.
Step 0. Initialization
(i). When an object is removed, set the object number to be n_d.
(ii). Set $S = 1$, $l = 1$, F_{old} = original rules set, $F_{new} = \varnothing$, $F_{gen} = \varnothing$, $F_{re} = \varnothing$, $F_{del} = \varnothing$.
Step 1.
For $i = 1$ to q
    For $j = 1$ to r
If $n_d \in D(X_i)_{Aj}$ && $D(X_i)_{Aj} - n_d$ == empty
        Based on *Lemma 1*, go to step 1.1.
Else go to step 2
      End If
    Endfor
Endfor
Step 1.1. Apply the reduct generation procedure of Pawlak to generate a new reduct of $D(X_i)_{Aj}$.
Step 1.2. Check whether the new reduct exists in the R.
For $i = 1$ to k
    If the new reduct $\in R_i$
      Based on *Lemma 2*, go to step 1.2.1.
    Else go to step 1.2.2.
    End If
Endfor
Step 1.2.1. Add the new reduct to R.
Step 1.2.2. Merge the new reduct with the identified original reduct into R. The new reduct of object number joins R_{MO} and the cardinality is also added to R_{OC}.
Step 2. Check whether the new reduct is better than the original.
For $l = 1$ to L
    If a new reduct is generated,
      Go to step 2.1.
    Else go to step 3.
    End If
Endfor
Step 2.l. Check whether the intersection of A_j and $D(X_i)$ is empty.
If it is empty
Go to Step 1.1.
Else go to step 2.
  End If
Step 3. Find any reduct that is possibly affected by n_d in R.
For $n = 1$ to k
If $n_d \in R_{MO}(n)$
      Go to step 3.1.
    End If
Endfor
Go to step 3.2.
Step 3.1. When the n_d is removed from R_{MO}, subtract $I(n_d)_{OC}$ from R_{OC}.
Step 3.2. Re-compute the strength index, SI. Sort SI according to case number, S, and T_{new} is stored, with the order, for each reduct.
Step 4. Check whether the new order has been changed in R.
Temp = 0
For $n = 1$ to k
    If $T_{new}(n)$!= $T_{old}(n)$
      Temp = 1
    End If
Endfor

Algorithm 1: *Cont.*
Step 5. Decide whether to re-extract the rules, according to the value of *Temp*. Based on *Lemma 3*, if *Temp* = 1, the rules must be re-extracted, so go To Step 5.1, else go To Step 6. Step 5.1. Re-extract the rules, according to AREA Set F_{old} = the original rules, F_{gen} = the new rules, F_{re} = the replaced rules, F_{del} = the deleted rules. Step 6. Print $F_{new} = \{F_{old} + F_{gen} + F_{re} - F_{del}\}$.

The flowchart of the algorithm is presented in Figure 1.

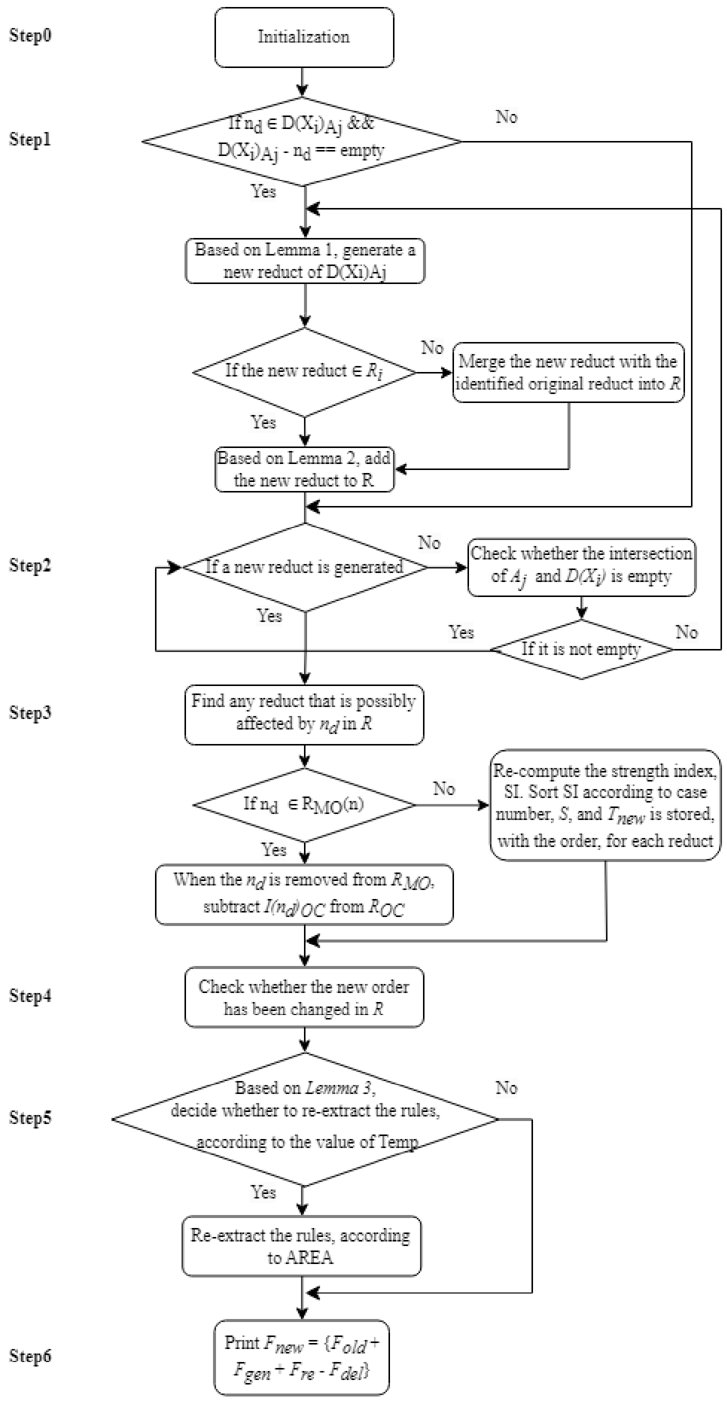

Figure 1. The flowchart of the algorithm.

3.4. Time Complexity and Comparison

DAREA time complexity in different cases is shown in Table 3.

Table 3. Time complexity.

Case No.	Time Complexity
Case 1	$O(nm(N_{cor}))$
Case 2	$O(nm(N_{cor}) + r(N_r))$
Case 3	$O(nm(N_{cor}) + q(N_{nr}) + r(N_r))$
Case 4	$O(nm(N_{cor}) + q(N_{nr}))$
Case 5	$O(nm(N_{cor}) + q(N_{nr}) + r(N_r))$
AREA	$O(m^2\ k\ (N_{cor}) + qk(N_{nr}) + r(N_r))$

Figure 2 is the time of the original data processed via rough set theory with outdated data. Figure 3 is CPU time when measured with the deleted ratio. Figures 2 and 3 show that, with a growth in data size, the original processing time increases substantially. This is a shortcoming of the rough set theory algorithm, as it wastes a lot of resources and time by reloading all databases in every update. Although the cost of time increases with data size, compared with the growth in original data, the time increased could be moderated by the DAREA, which uses the principle of difference in sets to make database management more efficient. When an object is deleted, it can effectively respond to the changes in rules without reloading the entire database. With the characteristics of Big Data, the percentage of data being changed is relatively low. The new dataset is almost the same size as the original dataset, which means that the calculation time would be close to the original processing time of the entire database. We should avoid reading the database repeatedly to update the rules. As the DAREA can effectively modify the rule directly for the decrement dynamic database, it can handle rule changes, as well as resource consumption, under a decrement algorithm.

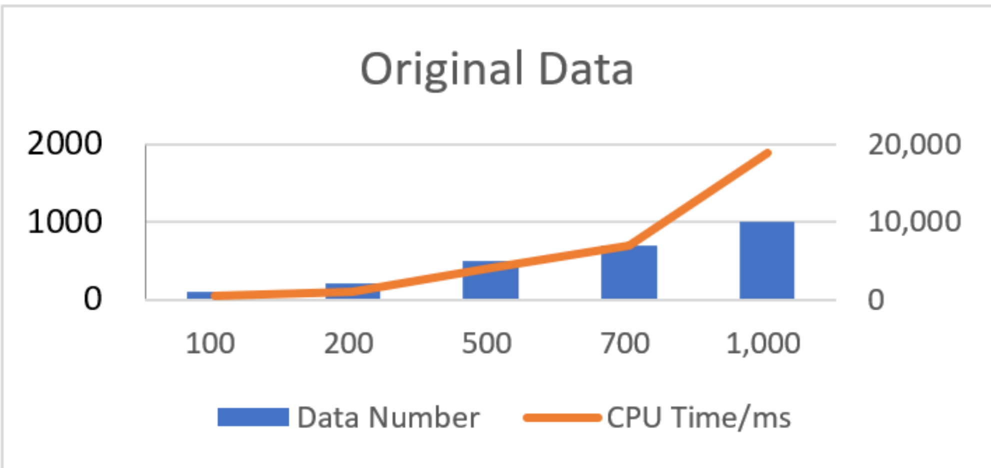

Figure 2. Original data time.

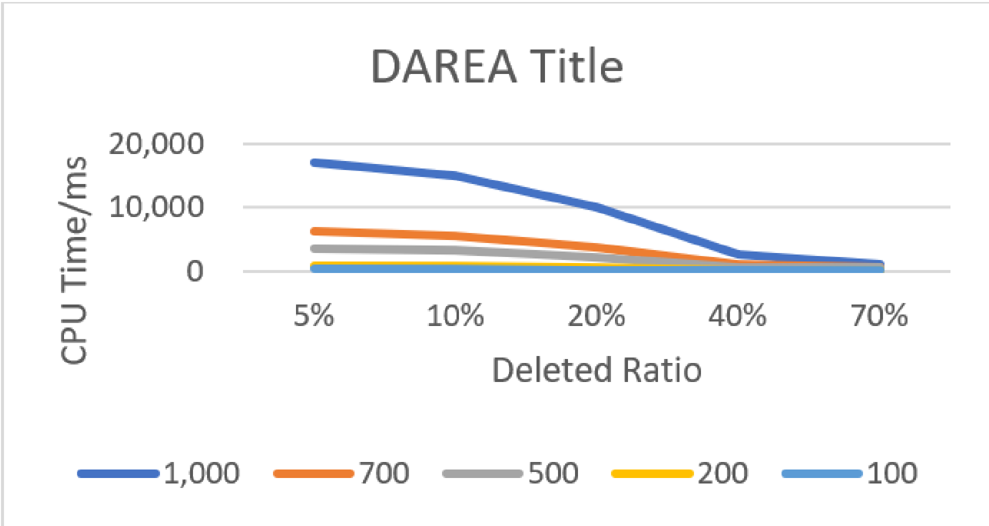

Figure 3. Deleted rate time.

4. Case Study

A case in the renewable energy company ABC was studied to illustrate the effectiveness of using DAREA methodology under commercial conditions and to explore its management implications.

4.1. Background

ABC Company produces and sells a wide range of renewable energy products, not only in household energy-saving products but also in building energy systems. In addition to production and marketing, ABC Company has invested substantial resources into research and development of renewable energy technologies. The company also has joint research with other renewable energy laboratories. Moreover, with ABC's experience and excellent efficiency in the construction of renewable energy, government agencies often seek their cooperation to assist government-owned projects. ABC Company provides a full range of green energy products and construction services for residents, businesses and government agencies. Their products and services are committed to achieving a sustainable future.

With a huge customer database and a variety of products, ABC Company provides customers with various analyses that consider a number of factors to develop different marketing strategies for different target groups, aiming to achieve a good balance in the operation and expansion of renewable energy. ABC Company analyzes customers' buying behavior and recommends appropriate merchandise to customers. This does not only help target customers to make accurate purchases, but also allows them access to a tailored product portfolio. The company investigates customer purchase factors and takes surveys to understand user satisfaction for the products after a month or two. Although new technology and products often replace the existing ones, some green energy equipment lasts a long period of time and users often keep using old products for several years or even decades, despite technology upgrades or the emergence of new products. This historical data might affect the analysis and cause difficulty in the analysis due to the vast accumulation of data. To solve this situation, ABC Company uses dynamic data analysis to delete outdated data when performing analysis so as to avoid an impact on current analysis from the outdated data.

4.2. Data Analysis

In this study, we used the DAREA to analyze ABC's customer data with the 18 attributes shown in Table 4. Attribute 1 is customer type {1—general public, 2—company, 3—government, 4—other}, which corresponds to ABC's main product sales targets. Attribute 2 is gender {1—male, 2—female, 3—other}, which leads to a significant difference in buying behavior. If the customer type is not the general public, this attribute is empty. Attribute 3 is age, divided into six age groups {1—under 24, 2—25 to 34, 3—35 to 44, 4—45 to 54, 5—54 to 64, 6—over 65}, which is used to analyze the buying behavior of different ages. Attribute 4 is revenue per year, divided into five levels in thousands of NTD {1—under 250, 2—260 to 500, 3—510 to 750, 4—760 to 1000, 5—over 1010}. Attribute 5 is education level {1—junior high, 2—senior high, 3—universities and colleges, 4—above}. Education level affects consumer buying behavior and purchase intention [61]. Attribute 6 is marital status {1—married, 2—unmarried}. Marital status affects the type of purchased goods. If customer type is not the general public, this attribute is empty. Attribute 7 is willingness to pay the cost of green energy {1—yes, 2—no}. Compared to traditional energy, customers who are willing to pay extra for green energy costs have higher purchase intention. Attribute 8 is the age of the house in years {1—under 5, 2—6 to 30, 3—over 3}. Attribute 9 is house area space in squared meters {1—under 75, 2—75 to 200, 3—over 201}. Attribute 10 is the floor level {1—under 5, 2—6 to 15, 3—over 16}. Attribute 11 is the first three digits of the zip code. Attribute 12 indicates if a customer is using green power {1—yes, 2—no}. Attribute 13 is monthly electricity consumption in kilowatts {1—under 300, 2—301 to 600, 3—601 to 1000, 4—over 1001}. Attribute 14 is the average monthly electricity bill in NTD {1—under 1000, 2—1001 to 2000, 3—2001 to 5000, 4—over 5001}. When the use of electricity is higher, the price per kilowatt is higher. This affects willingness to save electricity. Attribute 15 is electricity growth rate per year {1—under 1% including negatives, 2—1% to 3%, 3—over 3%}. Attribute 16 is the time from the last purchase in years {1—under 1, 2—1 to 2, 3—2 to 5, 4—over 5}. Attribute 17 is the purchased product ID. Attribute 18 is the result of the user purchase intention survey {1—cost, 2—social and environmental, 3—brand, 4—policy, 5—other}. Outcome is customer satisfaction based on an after-sales survey {1—very bad, 2—bad, 3—normal, 4—good, 5—very good}. The ABC Company customer decision table is partially presented in Table 5. The calculated customer information based on the decision rules is partially presented in Table 6.

Table 4. Customer data attribute table.

	Attribute Name	1	2	3	4	5	6
A1	Customer Type	General public	Company	Government	Other	-	-
A2	Gender	Male	Female	another	-	-	-
A3	Age	<24	25~34	35~44	45~54	55~64	>65
A4	Revenue/Year	<25 K	26~50	51~75	76~100	>101	-
A5	Education level	Junior high	Senior high	Universities and colleges	Above	-	-
A6	Marital	Married	Unmarried	-	-	-	-
A7	Willingness	Yes	No	-	-	-	-
A8	Age of house	<5	6~30	>30	-	-	-
A9	House area (m^2)	<75	75~200	>201	-	-	-
A10	Floors	<5	6~15	>16	-	-	-
A11	Area	The first three digits of the zip code					
A12	Using green power	Yes	No	-	-	-	-
A13	Monthly electricity consumption	<300	301~600	601~1000	>1001	-	-

Table 4. *Cont.*

	Attribute Name	1	2	3	4	5	6
A14	Average monthly electricity bill	<1000	1001~2000	2001~5000	>5001	-	-
A15	Electricity growth rate/year	<1%	1~3%	>3%	-	-	-
A16	Time from the last purchase products (year)	<1	1~2	2~5	>5	-	-
A17	Product ID	The purchased product ID					
A18	Purchase intentions	Cost	Social and environmental	Brand	Policy	Other	-
Outcome	Customer satisfaction	Very bad	Bad	Normal	Good	Very good	-

Table 5. Company customer data decision table.

Object	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	O	Cardinal Number
1	1	2	2	2	2	3	2	1	2	1	3	2	2	2	2	1	7	2	3	3
2	1	1	2	3	3	2	1	2	2	2	2	1	1	1	1	1	4	1	3	9
3	2	2	3	2	2	1	2	1	1	1	2	2	2	2	2	4	5	2	1	15
4	2	2	1	2	2	1	3	1	3	1	2	2	3	3	3	3	3	3	2	11
5	2	1	4	2	3	1	3	1	2	2	2	1	1	1	1	4	1	4	4	17
6	2	1	3	2	3	2	2	2	1	1	2	2	3	3	3	3	3	3	4	2
7	2	1	3	3	3	1	2	1	3	2	2	1	3	3	3	3	10	2	5	53
8	1	3	1	2	2	1	2	1	1	2	3	2	1	1	3	1	1	2	1	13
9	3	1	4	3	3	2	2	2	2	1	2	2	2	2	2	2	4	2	3	12
10	3	1	3	2	2	2	2	1	2	2	3	1	3	3	3	4	3	3	2	7
11	2	2	3	1	3	2	1	1	2	2	2	1	3	3	3	3	8	3	3	12
12	1	2	2	4	2	1	2	2	1	1	3	2	2	2	2	2	2	1	2	27
13	2	1	3	2	3	1	2	1	3	2	3	2	3	3	3	4	3	3	5	22
14	4	2	1	2	3	1	2	1	3	2	3	1	3	3	3	3	9	4	4	8
15	2	3	2	1	3	2	1	1	2	1	3	2	3	3	3	3	6	3	2	11
……																				
500	1	2	3	5	2	1	4	2	3	3	2	1	3	4	2	2	7	2	4	6
wj	0.75	0.28	0.75	0.92	0.61	0.25	0.67	0.65	0.6	0.4	0.65	0.67	0.82	0.81	0.79	0.32	1.0	0.91		

ABC Company product sales strategies are driven by these analysis results. For example, according to Rule 1, a male married customer with "Revenue per Month" between 51 and 75 (NTD 10,000), if the purchased product is "SWHs System", will have a customer satisfaction of "very good". Hence, in the marketing for SWHs Systems, this segment of customers is recommended as the primary sales target.

For another example, according to Rule 4, for a customer with monthly electricity consumption over 1001 kilowatts and a house over 30 years old, if the customer purchases "PV System", customer satisfaction is "very bad". Hence, ABC Company's marketing should avoid recommending products of PV systems to this segment of customers. ABC Company developed a new kind of solar photovoltaic panels which, compared with PV systems, have lower prices and better conversion efficiency. Through a decision process, the company decided to fully replace the old products with the new products. Because the old PV systems can be used for a long time, customers might continue using the old systems instead of using new systems. This outdated customer data could cause difficulty in analysis, and we were unable to determine a significant value. Therefore, ABC Company decided to delete customer data that were more than five years old or the data of customers who purchased old products (Table 7). This action can facilitate better analysis with current market conditions.

Table 6. Rule table.

Rule	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	O
1	1	1		2		1											2		5
2				3			1					1			1			2	5
3	2			3						3	106						9	1	3
4	1							3		1			4				8		1
5		2		4									2				4		1
6	2			5							110		4			4	8		5
7		1	2			1								2				1	4
8			3	3	4		1											2	4
9	2							2	3	3							12		5
10	3								2	2					3			3	2
11			4		2												1	4	3
12		2			3	1							2				4		4
13	1							1	1								8	1	5
14	2											1				2	9		3
15	1			3		2											1	3	4
….																			

Table 7. Deleting outdated objects.

Object	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	O	Cardinal Number
1	1	2	2	2	2	3	2	1	2	1	3	2	2	2	2	1	7	2	3	3
2	1	1	2	3	3	2	1	2	2	2	2	1	1	1	1	1	4	1	3	9
3	2	2	3	2	2	1	2	1	1	1	2	2	2	2	2	4	8	2	1	15
4	2	2	1	2	2	1	3	1	3	1	2	2	3	3	3	3	3	3	2	11
5	2	1	4	2	3	1	3	1	2	2	2	1	1	1	1	4	1	4	4	17
6	2	1	3	2	3	2	2	2	1	1	2	2	3	3	3	3	3	3	4	2
7	2	1	3	3	3	1	2	1	3	2	2	1	3	3	3	3	10	2	5	53
8	1	3	1	2	2	1	2	1	1	2	3	2	1	1	3	1	1	2	1	13
9	3	1	4	3	3	2	2	2	2	1	2	2	2	2	2	2	4	2	3	12
10	3	1	3	2	2	2	2	1	2	2	3	1	3	3	3	4	3	3	2	7
11	2	2	3	1	3	2	1	1	2	2	2	1	3	3	3	3	8	3	3	12
12	1	2	2	4	2	1	2	2	1	1	3	2	2	2	2	2	2	1	2	27
13	2	1	3	2	3	1	2	1	3	2	3	2	3	3	3	4	3	3	5	22
14	4	2	1	2	3	1	2	1	3	2	3	1	3	3	3	3	9	4	4	8
15	2	3	2	1	3	2	1	1	2	1	3	2	3	3	3	3	6	3	2	11
……																				
500	1	2	3	5	2	1	4	2	3	3	2	1	3	4	2	2	7	2	4	6
wj	0.33	0.28	0.75	0.88	0.61	0.25	0.67	0.65	0.6	0.4	0.85	0.67	0.92	0.81	0.79	0.32	1.0	0.62		

Because the original Object3 purchase was more than five years ago and the purchased product was an old version of a PV system, the analysis result using this data would not fit today's market trend. For this reason, Object3 and other objects over five years were deleted from the data for this analysis. Table 8 presents the partial 430 objects for analysis.

Table 8. New decision table after deletion.

Object	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	O	Cardinal Number
1	1	2	2	2	2	3	2	1	2	1	3	2	2	2	2	1	7	2	3	3
2	1	1	2	3	3	2	1	2	2	2	2	1	1	1	1	1	4	1	3	9
3	2	2	1	2	2	1	3	1	3	1	2	2	3	3	3	3	3	3	2	11
4	2	1	3	2	3	2	2	2	1	1	2	2	3	3	3	3	3	3	4	2
5	2	1	3	3	3	1	2	1	3	2	2	1	3	3	3	3	10	2	5	53
6	1	3	1	2	2	1	2	1	1	2	3	2	1	1	3	1	1	2	1	13
7	3	1	4	3	3	2	2	2	2	1	2	2	2	2	2	2	4	2	3	12
8	2	2	3	1	3	2	1	1	2	2	2	1	3	3	3	3	8	3	3	12
9	1	2	2	4	2	1	2	2	1	1	3	2	2	2	2	2	2	1	2	27
10	4	2	1	2	3	1	2	1	3	2	3	1	3	3	3	3	9	4	4	8
11	2	3	2	1	3	2	1	1	2	1	3	2	3	3	3	3	6	3	2	11
……																				
430	1	2	3	5	2	1	4	2	3	3	2	1	3	4	2	2	7	2	4	6
wj	0.75	0.28	0.75	0.92	0.61	0.25	0.67	0.65	0.6	0.4	0.85	0.67	0.92	0.81	0.79	0.32	1.0	0.62		

After recalculating and comparing with the old rule table, the new rule table is presented in Table 9. The original Rule 6 no longer supports the new decision table after comparing with the new rule table. Because of the deleted objects, Rule 6 is unable to exist and therefore disappears. ABC Company would know that the original customer sales program corresponding to rule 6 does not have enough evidence to support its effectiveness and expectations. This sales program should be abolished so as to save the related resources for other purposes.

Table 9. New rule table after deletion.

Rule	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	O
1	1	1		2		1											2		5
2				3			1					1			1			2	5
3	2			3						3	106						9	1	3
4	1							3		1			4				8		1
5		2		4									2				4		1
6	2			5							110		4			4	8		5
7		1	2			1								2				1	4
8			3	3	4		1											2	4
9	2							2	3	3							12		5
10	3								2	2					3			3	2
11			4		2												1	4	3
12		2			3	1							2				4		4
13	1							1	1								8	1	5

Table 9. *Cont.*

Rule	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	O
14	2											1				2	9		3
15	1			3		2											1	3	4
….																			

4.3. Managerial Implication

The managerial implications of rules developed using the DAREA methodology for analysis are critical. As described in Section 4.1, ABC Company can leverage a customized product portfolio to recommend suitable products to its customers through CRM channels. The following two directions in rules, validated and interpreted by senior managers, provide general business guidance:

Satisfaction is very high (Customer satisfaction = 5)

In order to enhance overall customer satisfaction, knowing the most suitable service and products for customers is the key to increasing revenue and profit. Energy should be focused on identifying rules that lead to very high satisfaction. According to one study, if customer satisfaction is high for a product service, customer loyalty is generally five times the average. On the other hand, if one company's customer satisfaction continues to increase, its annual profit will increase more than 10%, whereas if the company declines in satisfaction, annual profit will drop at least 14% [62].

Some studies point out that the cost of retaining an old customer is one-fifth of the cost of attracting a new customer. If customer satisfaction is high, the cost of retaining old customers becomes very low [63]. Customer retention is one of the key fields that most of the CRM strategies focus on [64]. Because the promotion of renewable energy needs government and community support, a good corporate reputation and a high degree of social acceptance in terms of implementation can obtain greater success.

Referring to these rules can provide suitable product suggestions for an incoming customer. Instead of passively selling products, one should proactively recommend products to achieve customer satisfaction and enhance the effect of promoting renewable energy. Integrating with other information systems, such as CRM and business intelligence (BI) modules of ERP (enterprise resource planning), can further increase the level of competitiveness.

Satisfaction is very low (Customer satisfaction = 1)

If the final output of the rules results in very low customer satisfaction, the company should be alerted and respond accordingly. For example, Rule 4's association properties show IF (Customer Type = general public) AND (revenue/year = over 101) AND (Monthly electricity consumption = over 1001) AND (Age of house = over 30 years) AND (Product ID = 8) THEN (Customer satisfaction = 1). Customer satisfaction is very low. Most likely, the age of the house causes energy inefficiency. Customers meeting these factors can be suggested to purchase goods other than PV systems. In general, the possible causes of ineffective energy saving can be identified. The company's R&D department can improve the effectiveness of PV systems for older buildings, as energy conservation was not previously an apparent problem. Understanding the crux of the problem and investing resources in improving the greatest efficiency of products can enhance customer satisfaction.

The other levels of satisfaction and associated rules can provide the direction to work with and improve business performance. Because DAREA methodology delivers weighted indicators, compared with the original rough set theory, the generated rules do not only have more substantial significance, but also provide more managerial implications.

5. Conclusions

This case study shows that the DAREA method is more efficient than the traditional rough set algorithm, especially in refining the tremendous data cleaning of a database.

When the original rules change, the changes are reflected simultaneously. When processing Big Data, the execution time grows only moderately even if the data size is greatly increased. The DAREA provides a significant improvement to the traditional issues of deleting outdated data in a Big Data system. Meanwhile, when processing a diversity of data types in the renewable energy industry, one can achieve effective decision rules quickly through DAREA rule induction to aid users in making the right choices. Specifically, the long-term accumulation of data in today's rapidly changing society may result in massive amounts of outdated information. This decrement case study may lead to data decrement operations being adequate in Big Data for renewable energy applications, which would help the promotion of renewable energy sources. In such a situation, achieving operational efficiency and rule induction requires the elimination of outdated data as well as effective decision making.

The methodology in this study could be one solution. However, due to the characteristics of Big Data and the variabilities of renewable energy data, there are still a number of challenges to overcome going forward. Some of these challenges are related to how to obtain the right information for rule induction, how to adapt corporate culture and management, if the quality and quantity of data are less than expected, etc. [65]. In addition, the methodology applied in this study still needs further research to verify whether DAREA is the most efficient algorithm in deleting outdated data in a Big Data system. To obtain social and popular support, the shaping of the renewable energy market depends not only on government policies, but also on business collaboration [7]. For enterprise applications, a quantitative test should be performed and analyzed after two quarters when additional outdated data are generated and eliminated. In addition, companies should trace Big Data over a long period and check whether the data needed for the applications have significantly increased. Hence, how to timely and efficiently add customer data could be another future research topic.

Author Contributions: Conceptualization, R.R.G.; Formal analysis, W.-Y.L.; Methodology, C.-C.H.; Writing—review and editing, P.-A.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science and Technology Council of Taiwan, grant number (NSTC 112-2410-H-260-010-MY3, MOST 111-2410-H-018-017-MY2, MOST 110-2410-H-260-016-MY2).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are unavailable due to privacy or ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

U	Set of objects;
A	Set of attributes;
d	Set of decision attributes;
i	Object index;
j	Attribute index;
n	Reduct index;
l	Value of new level;
L	Number of original levels;
q	Number of object data;
r	Number of attributes;
k	Number of reducts in "reduct set of table";
S	Case number;
T_{new}	Number of sort SI according to case number S in the new reduct set of tables;
T_{old}	Number of sort SI according to case number S in the original reduct set of tables;

Temp	The order number of SI that has changed and is different from the original;
X_i	The *i*-th object;
a_{ij}	The *j*-th value set of attributes for object X_i;
n_d	Number of removed objects;
I()	The original information table;
Aj	The *j*-th attribute;
$I(X_i)_{Aj}$	The *j*-th attribute for object X_i column in *I*, the original information table;
D()	The difference set of each attribute A_j (the equivalent class of each object) and attribute *d* (the equivalent class of each object corresponding to decisive) of the table;
$D(X_i)_{Aj}$	The *j*-th attribute for object X_i column in the difference set;
E	Extended table;
R	Reduct set;
$Reduct_{new}$	The new reduct set;
$Reduct_{old}$	The old reduct set;
F	Final rule table;
F_{old}	The set of rules (reducts) selected from old rules;
F_{new}	The set of rules (reducts) selected from new rules;
F_{gen}	The set of rules (reducts) selected from generating rules;
F_{re}	The set of rules (reducts) selected from replacing rules;
F_{del}	The set of rules (reducts) selected from new deleting rules;
OC	Column of object cardinality;
MO	Column of merged object;
R_{OC}	Index set of columns of object cardinality in the reduct set of tables;
R_{MO}	Index set of columns of merged object no in the reduct set of tables;
SO	Column of support object.
n	Total number of original objects;
m	Total number of attributes;
r	Total number of rules;
q	Total number of reducts after being removed;
k	Total number of objects after being removed;
$\mathrm{N_{cor}}$	Total number of original reducts from the object that generates the rule (or reduct);
$\mathrm{N_{nr}}$	Total number of reducts from new data set;
$\mathrm{N_r}$	Total number of rules from rule set.

References

1. Czúcz, B.; Gathman, J.P.; McPherson, G.R. The impending peak and decline of petroleum production: An underestimated challenge for conservation of ecological integrity. *Conserv. Biol.* **2010**, *24*, 948–956. [CrossRef] [PubMed]
2. Fantazzini, D.; Höök, M.; Angelantoni, A. Global oil risks in the early 21st century. *Energy Policy* **2011**, *39*, 7865–7873. [CrossRef]
3. Oncel, S.S. Green energy engineering: Opening a green way for the future. *J. Clean. Prod.* **2017**, *142*, 3095–3100. [CrossRef]
4. Dincer, I. Renewable energy and sustainable development: A crucial review. *Renew. Sustain. Energy Rev.* **2000**, *4*, 157–175. [CrossRef]
5. George, G.; Howard-Grenville, J.; Joshi, A.; Tihanyi, L. Understanding and tackling societal grand challenges through management research. *Acad. Manag. J.* **2016**, *59*, 1880–1895. [CrossRef]
6. Bhowmik, C.; Bhowmik, S.; Ray, A.; Pandey, K.M. Optimal green energy planning for sustainable development: A review. *Renew. Sustain. Energy Rev.* **2017**, *71*, 796–813. [CrossRef]
7. Wüstenhagen, R.; Bilharz, M. Green energy market development in Germany: Effective public policy and emerging customer demand. *Energy Policy* **2006**, *34*, 1681–1696. [CrossRef]
8. SGDs. United Nations. 2015. Available online: https://sdgs.un.org/goals (accessed on 28 April 2023).
9. COP26. 2021. Available online: https://ukcop26.org/cop26-goals/ (accessed on 28 April 2023).
10. Awan, U. Green marketing: Marketing strategies for the Swedish energy companies. *Int. J. Ind. Mark.* **2011**, *1*, 1. [CrossRef]
11. Dwivedi, Y.K.; Hughes, L.; Kar, A.K.; Baabdullah, A.M.; Grover, P.; Abbas, R.; Andreini, D.; Abumoghli, I.; Barlette, Y.; Bunker, D. Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial reflection and call to action. *Int. J. Inf. Manag.* **2022**, *63*, 102456. [CrossRef]
12. Papagiannidis, S.; Marikyan, D. Environmental Sustainability: A technology acceptance perspective. *Int. J. Inf. Manag.* **2021**, *63*, 102445. [CrossRef]
13. Singh, M.; Sahu, G.P. Towards adoption of Green IS: A literature review using classification methodology. *Int. J. Inf. Manag.* **2020**, *54*, 102147. [CrossRef]

14. de la Rue du Can, S.; Leventis, G.; Phadke, A.; Gopal, A. Design of incentive programs for accelerating penetration of energy-efficient appliances. *Energy Policy* **2014**, *72*, 56–66. [CrossRef]
15. Lee, C.W.; Zhong, J. Top down strategy for renewable energy investment: Conceptual framework and implementation. *Renew. Energy* **2014**, *68*, 761–773. [CrossRef]
16. Lee, S.-C.; Shih, L.-H. Renewable energy policy evaluation using real option model—The case of Taiwan. *Energy Econ.* **2010**, *32*, S67–S78. [CrossRef]
17. Pee, L.; Pan, S.L. Climate-intelligent cities and resilient urbanisation: Challenges and opportunities for information research. *Int. J. Inf. Manag.* **2021**, *63*, 102446. [CrossRef]
18. Mohamad Radzi, P.N.L.; Akhter, M.N.; Mekhilef, S.; Mohamed Shah, N. Review on the Application of Photovoltaic Forecasting Using Machine Learning for Very Short-to Long-Term Forecasting. *Sustainability* **2023**, *15*, 2942. [CrossRef]
19. Ceci, M.; Corizzo, R.; Fumarola, F.; Ianni, M.; Malerba, D.; Maria, G.; Masciari, E.; Oliverio, M.; Rashkovska, A. Big data techniques for supporting accurate predictions of energy production from renewable sources. In Proceedings of the 19th International Database Engineering Applications Symposium, Yokohama, Japan, 13–15 July 2015.
20. Yao, Q.; Wu, Z.; Zhu, C.; Song, S.; Qiu, X.; Liu, S. Architecture Design and Application of Global Renewable Energy Reserves Evaluation and planning Platform. In Proceedings of the 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 21–23 November 2019.
21. Devaraj, J.; Madurai Elavarasan, R.; Shafiullah, G.; Jamal, T.; Khan, I. A holistic review on energy forecasting using big data and deep learning models. *Int. J. Energy Res.* **2021**, *45*, 13489–13530. [CrossRef]
22. Li, R.; Wen, H.; Huang, X.; Liu, Y. Renewable Energy: A Curse or Blessing—International Evidence. *Sustainability* **2023**, *15*, 11103. [CrossRef]
23. Duić, N.; da Graça Carvalho, M. Increasing renewable energy sources in island energy supply: Case study Porto Santo. *Renew. Sustain. Energy Rev.* **2004**, *8*, 383–399. [CrossRef]
24. Alharthi, M.; Hanif, I.; Alamoudi, H. Impact of environmental pollution on human health and financial status of households in MENA countries: Future of using renewable energy to eliminate the environmental pollution. *Renew. Energy* **2022**, *190*, 338–346. [CrossRef]
25. Kandpal, T.C.; Broman, L. Renewable energy education: A global status review. *Renew. Sustain. Energy Rev.* **2014**, *34*, 300–324. [CrossRef]
26. Yang, J.; Zheng, C.; Liu, H. Digital Transformation and Rule of Law Based on Peak CO_2 Emissions and Carbon Neutrality. *Sustainability* **2022**, *14*, 7487. [CrossRef]
27. Welch, T.F.; Widita, A. Big data in public transportation: A review of sources and methods. *Transp. Rev.* **2019**, *39*, 795–818. [CrossRef]
28. Fazelpour, F.; Vafaeipour, M.; Rahbari, O.; Rosen, M.A. Intelligent optimization to integrate a plug-in hybrid electric vehicle smart parking lot with renewable energy resources and enhance grid characteristics. *Energy Convers. Manag.* **2014**, *77*, 250–261. [CrossRef]
29. Venetsanos, K.; Angelopoulou, P.; Tsoutsos, T. Renewable energy sources project appraisal under uncertainty: The case of wind energy exploitation within a changing energy market environment. *Energy Policy* **2002**, *30*, 293–307. [CrossRef]
30. Lee, I.; Mangalaraj, G. Big data analytics in supply chain management: A systematic literature review and research directions. *Big Data Cogn. Comput.* **2022**, *6*, 17. [CrossRef]
31. Menegaki, A. Valuation for renewable energy: A comparative review. *Renew. Sustain. Energy Rev.* **2008**, *12*, 2422–2437. [CrossRef]
32. Duan, Y.; Edwards, J.S.; Dwivedi, Y.K. Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *Int. J. Inf. Manag.* **2019**, *48*, 63–71. [CrossRef]
33. Cohen, J.; Dolan, B.; Dunlap, M.; Hellerstein, J.M.; Welton, C. MAD skills: New analysis practices for big data. *Proc. VLDB Endow.* **2009**, *2*, 1481–1492. [CrossRef]
34. Karasuyama, M.; Takeuchi, I. Multiple incremental decremental learning of support vector machines. *IEEE Trans. Neural Netw.* **2010**, *21*, 1048–1059. [CrossRef]
35. Brewer, E.A. Combining systems and databases: A search engine retrospective. *Read. Database Syst.* **2005**, *4*, 247–261.
36. Al Ayub Ahmed, A.; Rajesh, S.; Lohana, S.; Ray, S.; Maroor, J.P.; Naved, M. Using Machine Learning and Data Mining to Evaluate Modern Financial Management Techniques. In Proceedings of the Second International Conference in Mechanical and Energy Technology: ICMET 2021, Greater Noida, India, 28–29 October 2021.
37. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 341–356. [CrossRef]
38. Pawlak, Z.; Skowron, A. Rudiments of rough sets. *Inf. Sci.* **2007**, *177*, 3–27. [CrossRef]
39. Huang, B.; Wei, D.-K.; Li, H.-X.; Zhuang, Y.-L. Using a rough set model to extract rules in dominance-based interval-valued intuitionistic fuzzy information systems. *Inf. Sci.* **2013**, *221*, 215–229. [CrossRef]
40. Abed-Elmdoust, A.; Kerachian, R. Wave height prediction using the rough set theory. *Ocean Eng.* **2012**, *54*, 244–250. [CrossRef]
41. Liu, Y.; Huang, W.; Jiang, Y.; Zeng, Z. Quick attribute reduct algorithm for neighborhood rough set model. *Inf. Sci.* **2014**, *271*, 65–81. [CrossRef]
42. Lingras, P.; Yao, Y. Data mining using extensions of the rough set model. *J. Am. Soc. Inf. Sci.* **1998**, *49*, 415–422. [CrossRef]
43. Swiniarski, R.W.; Skowron, A. Rough set methods in feature selection and recognition. *Pattern Recognit. Lett.* **2003**, *24*, 833–849. [CrossRef]

44. Tseng, T.-L.B.; Huang, C.-C. Rough set-based approach to feature selection in customer relationship management. *Omega* **2007**, *35*, 365–383. [CrossRef]
45. Morley, C.L. A dynamic international demand model. *Ann. Tour. Res.* **1998**, *25*, 70–84. [CrossRef]
46. Huang, C.C.; Tseng, T.L.B.; Jiang, F.; Fan, Y.N.; Hsu, C.H. Rough set theory: A novel approach for extraction of robust decision rules based on incremental attributes. *Ann. Oper. Res.* **2014**, *216*, 163–189. [CrossRef]
47. Liang, J.; Wang, F.; Dang, C.; Qian, Y. A group incremental approach to feature selection applying rough set technique. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 294–308. [CrossRef]
48. Yong, L.; Congfu, X.; Yunhe, P. An incremental rule extracting algorithm based on Pawlak reduction. In Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics, The Hague, The Netherlands, 10–13 October 2004; Volume 6, pp. 5964–5968.
49. Zhang, J.; Li, T.; Chen, H. Composite rough sets for dynamic data mining. *Inf. Sci.* **2014**, *257*, 81–100. [CrossRef]
50. Luo, C.; Li, T.; Chen, H.; Liu, D. Incremental approaches for updating approximations in set-valued ordered information systems. *Knowl. Based Syst.* **2013**, *50*, 218–233. [CrossRef]
51. Li, S.; Li, T.; Liu, D. Incremental updating approximations in dominance-based rough sets approach under the variation of the attribute set. *Knowl. Based Syst.* **2013**, *40*, 17–26. [CrossRef]
52. Zhang, Z. Model building strategy for logistic regression: Purposeful selection. *Ann. Transl. Med.* **2016**, *4*, 111. [CrossRef] [PubMed]
53. Shu, W.; Qian, W.; Xie, Y. Incremental feature selection for dynamic hybrid data using neighborhood rough set. *Knowl. Based Syst.* **2020**, *194*, 105516. [CrossRef]
54. Yang, X.; Li, M.; Fujita, H.; Liu, D.; Li, T. Incremental rough reduction with stable attribute group. *Inf. Sci.* **2022**, *589*, 283–299. [CrossRef]
55. Huang, W.; She, Y.; He, X.; Ding, W. Fuzzy Rough Sets-Based Incremental Feature Selection for Hierarchical Classification. *IEEE Trans. Fuzzy Syst.* **2023**, *31*, 3721–3733. [CrossRef]
56. Liu, D.; Li, T.; Zhang, J. A rough set-based incremental approach for learning knowledge in dynamic incomplete information systems. *Int. J. Approx. Reason.* **2014**, *55*, 1764–1786. [CrossRef]
57. Fan, Y.N.; Tseng, T.L.B.; Chern, C.C.; Huang, C.C. Rule induction based on an incremental rough set. *Expert Syst. Appl.* **2009**, *36*, 11439–11450. [CrossRef]
58. Tseng, T.L.B. *Quantitative Approaches for Information Modeling*; The University of Iowa: Iowa City, IA, USA, 1999.
59. Huang, C.C.; Tseng, T.L.B.; Fan, Y.N.; Hsu, C.H. Alternative rule induction methods based on incremental object using rough set theory. *Appl. Soft Comput.* **2013**, *13*, 372–389. [CrossRef]
60. Pawlak, Z. *Rough Sets*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1991.
61. Sun, B.; Morwitz, V.G. Stated intentions and purchase behavior: A unified model. *Int. J. Res. Mark.* **2010**, *27*, 356–366. [CrossRef]
62. Hallowell, R. The relationships of customer satisfaction, customer loyalty, and profitability: An empirical study. *Int. J. Serv. Ind. Manag.* **1996**, *7*, 27–42. [CrossRef]
63. Bowen, J.T.; Chen, S.L. The relationship between customer loyalty and customer satisfaction. *Int. J. Contemp. Hosp. Manag.* **2001**, *13*, 213–217. [CrossRef]
64. Shirazi, F.; Mohammadi, M. A big data analytics model for customer churn prediction in the retiree segment. *Int. J. Inf. Manag.* **2019**, *48*, 238–253. [CrossRef]
65. LaValle, S.; Lesser, E.; Shockley, R.; Hopkins, M.S.; Kruschwitz, N. Big data, analytics and the path from insights to value. *MIT Sloan Manag. Rev.* **2011**, *52*, 21–32.