

Article

Enhancing Survey Efficiency and Predictive Ability in Energy System Design through Machine Learning: A Workflow-Based Approach for Improved Outcomes

Andrew Chapman 

International Institute for Carbon Neutral Energy Research (WPI-I2CNER), Kyushu University,
Fukuoka 819-0395, Japan; chapman@i2cner.kyushu-u.ac.jp

Abstract: The design of a desirable, sustainable energy system needs to consider a broad range of technologies, the market landscape, and the preferences of the population. In order to elicit these preferences, both toward lifestyle factors and energy system design, stakeholder engagement is critical. One popular method of stakeholder engagement is the deployment and subsequent analysis of a survey. However, significant time and resources are required to design, test, implement and analyze surveys. In the age of high data availability, it is likely that innovative approaches such as machine learning might be applied to datasets to elicit factors which underpin preferences toward energy systems and the energy mix. This research seeks to test this hypothesis, utilizing multiple algorithms and survey datasets to elicit common factors which are influential toward energy system preferences and energy system design factors. Our research has identified that machine learning models can predict response ranges based on preferences, knowledge levels, behaviors, and demographics toward energy system design in terms of technology deployment and important socio-economic factors. By applying these findings to future energy survey research design, it is anticipated that the burdens associated with survey design and implementation, as well as the burdens on respondents, can be significantly reduced.



Citation: Chapman, A. Enhancing Survey Efficiency and Predictive Ability in Energy System Design through Machine Learning: A Workflow-Based Approach for Improved Outcomes. *Energies* **2023**, *16*, 4911. <https://doi.org/10.3390/en16134911>

Academic Editor: Alan Brent

Received: 22 May 2023

Revised: 21 June 2023

Accepted: 22 June 2023

Published: 24 June 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: energy system; sustainability; system preference; machine learning; survey analysis

1. Introduction

As nations around the world struggle with the technical aspects of energy transitions, i.e., the need to deploy new, clean energy technologies to meet their future energy needs, policy makers and governments are wrestling with policy and energy system design which meets our energy, economic and social needs. These challenges can be summarized as the need for equitable and sustainable energy system design, often described as the just transition [1]. The just transition deals with a number of critical issues associated with rapidly changing energy systems, ranging from impacts on employment [2], pressures on energy prices [3], health impacts [4], the allocation of subsidies [5], and a number of potential trade-offs depending on stakeholder preferences [6–8].

In order to better understand these trade-offs, policy makers and researchers utilize a number of tools to undertake stakeholder engagement and, ideally, to propose energy systems which can meet the needs of all stakeholders, and which can achieve common energy transition goals including sustainability, equitability and the ability to meet carbon reduction targets. These stakeholder engagement tools include public focus groups, interviews, surveys, social media and a variety of interactive engagement methodologies, among others [9–11]. Considering research approaches, in addition to literature reviews, survey analysis is often used to understand stakeholder's opinions and to improve outcomes across a range of fields [12]. Focal energy issues have shifted over time, i.e., from environmental justice, concerned with unrestricted stakeholder involvement in environmental policy development [13], toward climate justice, more concerned with sharing the

burdens of climate change [14,15] and to the more recent ‘energy justice’ which is concerned with the distribution, recognition and procedural justice aspects of energy [16,17]. For this reason, surveys, over time, have become a useful tool to elicit the critical issues surrounding energy systems. With a recent focus on the just transition, i.e., a transition which achieves carbon reduction goals, while upholding the tenets of environmental justice, climate justice and energy justice [16], targeted surveys and their analysis can complement efforts toward designing energy systems which are both equitable and sustainable.

One major challenge experienced by researchers or energy policy makers is the successful design of surveys that can meet these needs, followed by their application to energy system design. Although some unique approaches exist, such as responsive survey design to aid in developing surveys that improve response rates and reduce bias [18], there is also a significant body of survey work in the energy and energy transition space that can be drawn upon for retroactive analysis to determine which responses influence energy technology and system preferences the most. This paper leverages statistical analysis and innovative machine learning methods applied to recent, topical surveys regarding energy system preferences, opinions and design priorities to attempt to extract the factors that are most influential in the design of desirable (i.e., equitable and sustainable) energy systems.

The aim of this research is to provide an evidence base for the streamlining of stakeholder engagement which can be utilized to underpin bottom-up, desirable future energy system design. Four energy-related surveys from Japan and the United States (US) are utilized to derive the key factors that should be explored as a priority in order to best inform desirable energy system design stakeholder response prediction.

The study is presented as follows: Section 2 details recent literature regarding stakeholder engagement, survey design challenges and approaches to energy system design; Section 3 details the machine learning methods used in this study to identify influential survey responses which are useful to predict energy factor and system design preferences and their importance regarding survey design; Section 4 describes the results and variation across investigated surveys and nations as well as the efficacy of various machine learning approaches; Section 5 provides the discussion and implications for future energy survey design; and conclusions are provided in Section 6.

2. Literature Review

The challenge of designing appropriate surveys to elicit desired responses while being conscious of cost, timeliness and effective response elicitation is not unique to the energy field. For example, in the design of a national survey in the Netherlands, an increasing need to provide more targeted and detailed information and to improve analysis was identified. In order to meet these needs, researchers undertook a fundamental change in their surveying processes in order to achieve an integrated design and to move away from single-purpose surveys [19]. In the realm of information and communications technology, researchers sought to design a web survey that would identify travel behaviors. In order to design this survey, they undertook a review of previous work to identify gaps and necessary questions (i.e., literature review) and attempted to design a survey that would elicit the desired data points while limiting respondent fatigue and cognitive burden [20]. In the case of a US survey to assess social life, health and aging, the design of the survey instrument and the definition of measurement domains took several months. In addition, a pretest of the questionnaire was undertaken, identifying issues with the time required to administer, thus leading to a prioritization of factors to be extracted and identified non-essential factors being cut from the final survey instrument [21].

Cognizant of the need to reduce both the cost and respondent burden of surveys, a modular design was proposed for survey design using a random search algorithm which attempts to maintain precision requirements in light of other constraints. One shortcoming of such an approach is the identified need to estimate required design effects for the algorithm, or to undertake a pilot sample [22]. Recognizing that innovation in survey design can often yield unexpected results or have a negative impact on outcomes such as

bias or survey costs, the idea of ‘responsive survey design’ has been proposed. Utilized in Germany, researchers proposed a self-administered survey and experimented with incentives for completion, question ordering (sequential or simultaneous mode) and cost, finding that response rates can be improved through prepaid incentives, and that mode choice sequence had only a small impact on cost overall, while response bias was not largely impacted [18]. Another source of response bias was reported to occur in the use of different devices for responding to surveys, particularly for mobile devices (anticipated to account for up to 60% of respondents) and the need to ensure consistent display of response ranges to avoid prioritization of visible elements [23].

In terms of general survey design ideals, the Pew Research Center offers guidance in terms of keeping question numbers to a minimum to ascertain the defined goals of the survey, using closed-ended questions to compare specific traits, and being conscious of the order of questions, placing open-ended questions before closed-ended ones to avoid order effects and leading bias [24]. The key considerations of brevity, consistency, avoiding leading questions and having appropriate categorical response ranges are also extolled by the General Medical Council in their National Training Survey best practice guidelines [25]. Having a clear purpose for the survey is also critical, along with only asking questions that add value and relate specifically to the research goals. Testing a survey is also identified as a critical step prior to large scale deployment [26]. Further, the use of literature review to establish if model questions have been implemented and tested in previous surveys which could meet the needs of new research is also identified as valuable by Harvard University, with the additional benefit of comparative analysis across surveys [27].

Although no specific machine learning approaches are proposed in the literature to retrospectively design surveys, the idea itself is consistent with literature review processes, utilizing existing survey data. On the other hand, the use of machine learning techniques to analyze survey results is becoming increasingly common, with multiple benefits in addition to the computing power offered expected as a result [28]. For example, machine learning algorithms are expected to offer advantages in terms of quality assurance and real-time data monitoring and trend identification [29]. Furthermore, neural networks have been identified as more capable of modeling non-linear relationships than traditional analysis approaches such as statistical regression [30]. Other machine learning approaches such as decision tree classification algorithms have also been shown to perform better than other classifiers for dependent variable classification related to stakeholder intent [31]. Research has also been undertaken regarding the broad range of computational methods applicable to response prediction, including data hierarchy, collaborative filtering, supervised, and semi- and unsupervised learning-based approaches, with a focus on online advertising responses [32]. Considering the energy-related fields of study which employ machine learning and artificial intelligence, energy and eco-efficiency have been investigated in some depth recently. For example, machine learning approaches were combined with optimization methods to evaluate cement companies’ eco-efficiency, detailing model efficiencies [33]. Further, an evaluation of machine learning employing recommender systems was undertaken to show how these could be used to improve building energy efficiency, as well as the possibility of combining these approaches with smart meters and Internet of Things sensors [34]. Additionally, a review-based study on enhancing energy efficiency among other aspects via occupancy prediction was undertaken, demonstrating the value of both machine learning and neural network approaches in combination with sensors [35]. An additional review was undertaken, considering machine learning toward achieving thermal comfort, finding that machine learning-based controls can improve indoor air quality while reducing carbon dioxide levels and energy consumption [36].

Toward energy system design, neural networks are often utilized in the fields of evaluation, control, and operation or in the application of artificial intelligence for the realization of a modern smart grid or renewable energy systems, with applications increasing rapidly over time [37]. To support energy system optimization, a modeling approach was developed using neural networks and a hybrid optimization algorithm (HOA) which

can reach Pareto optimal solutions approximately 17 times faster than by using an actual engineering model (AEM) alone [38]. Further, it was identified that a data-driven approach based on reinforcement learning to design distributed energy systems by altering the number of hidden layers in the neural network can effectively predict system needs, and reinforcement learning with deep neural networks can help energy systems operate more efficiently with minimal impacts on the grid itself [39]. Research utilizing metamodels which can significantly reduce implementation effort and runtime per scenario for artificial intelligence-based modeling when compared to traditional white box modeling approaches has been explored in Germany, achieving a mitigation of complexity while maintaining a high level of accuracy [40].

Recognizing the challenges toward survey design, the existing approaches toward determining ideal questions and response ranges, and the recent utilization of machine learning for survey response analysis and prediction, this study proposes a methodology for the streamlining of survey design, using energy as an example. Specifically, the aim of this research is to leverage existing survey data (with similar applicability to consumption or behavioral data) and machine learning to reduce survey burden. Burden here refers to three key aspects: (1) the burden imposed by the high cost of the deployment of surveys; (2) the burden imposed on researchers in designing and testing survey instruments; and (3) the burden on respondents, in terms of fatigue in having to respond to an excessive number of questions. By reducing these burdens, it is anticipated that the deployment costs can be reduced, enabling a larger sample to be obtained for the same budget expenditure. In addition, the time and effort required to develop surveys can be reduced, allowing for the allocation of limited resources toward analyses and energy system design. Finally, it is anticipated that respondent fatigue can also be reduced, improving the quality and accuracy of responses while reducing response bias. By reducing these burdens and improving research outcomes, a flow-on benefit will be the usefulness of obtained data to inform future energy system design and analysis activities. Overall, we are aiming to apply machine learning to streamline survey design, improve deployment and data analysis, and to provide prediction-based insights toward energy system design.

3. Methodology

The methodology employed in this study utilizes machine learning evaluation of previously administered energy system-related surveys to establish the demographic factors or response types which are most influential toward our variables of interest, i.e., those related to energy factor preference and system design. Further, machine learning algorithms are compared to identify those which exhibit optimal prediction (classification) accuracy in eliciting critical demographics and response types.

3.1. Data Sources

The surveys utilized in this study include three conducted in Japan and one conducted in the US. The deployment date, original use case, number of questions asked, samples gathered, and common factors analyzed across surveys are detailed in Table 1.

In addition to the common factor questions identified in Table 1, each survey captures demographic data and, in some cases, additional specific enquiries about educational achievement, region, race, etc. A list of all factors analyzed is presented in Appendix A. All surveys are used to assess energy system design factors, while only surveys 3 and 4 are used to analyze key future energy system preferences due to the different design specifications of surveys utilized in this study.

Table 1. Survey specifics and common factors investigated.

Survey Specifics					Common Factors	
#	Nation	Date	Use Case	#Q/Samples	Preferences	System Design
1	Japan	March 2017	Energy Market Liberalization [41]	30/4148	➤ Energy Policy ➤ Energy Use ➤ Energy Cost ➤ Behavior	➤ Energy tech. knowledge ➤ Energy policy knowledge ➤ Energy mix ➤ Participation
2	Japan	March 2019	Energy Market Liberalization Impacts and Future Energy Systems [42]	51/4247	➤ Energy Policy ➤ Environmental Issues ➤ Energy Cost ➤ Behavior	➤ Energy tech. knowledge ➤ Energy policy knowledge ➤ Energy Mix ➤ Participation
3	US	August 2020	Cultural dynamics of Energy Systems [43,44]	17/3000	➤ Energy Policy ➤ Environmental Issues ➤ Energy Cost	➤ Energy tech. knowledge ➤ Energy policy knowledge ➤ Energy Mix ➤ Energy Location
4	Japan	January 2022	Knowledge of the Just Transition in Japanese Regions [45]	24/6000	➤ Energy Policy ➤ Environmental Issues ➤ Energy Cost	➤ Energy tech. knowledge ➤ Energy policy knowledge ➤ Energy Mix

3.2. Analysis Methods

Survey data are prepared by coding responses to numerical ranges (for levels of knowledge, preferences, etc.) and categorical (importance, increase or decrease in specific energy sources in the future energy mix, etc.) inputs. Prediction machine learning models including naïve Bayes, generalized linear model (GLM), logistic regression, large margin, deep learning, decision tree, random forest, gradient boosted trees and support vector machine (SVM) are run for each of the target factors of interest for this study, as detailed in Table 2.

Table 2. Prediction targets selected for this study.

Energy System Preferences	Energy System Design Factors
Stakeholder self-reported importance level of:	Stakeholder desire for an increase or decrease in generation from the following energy sources:
➤ Environmental Protection • (i.e., clean water, clean air, reduced waste, etc.) ➤ Climate Change Mitigation • (i.e., reducing greenhouse gases, restricting temperature rises, etc.) ➤ An Equitable Society • (i.e., reducing the gap between rich and poor, shared environmental cost and burden allocation)	➤ Renewable Energy • Solar, wind, biomass, hydro, geothermal, etc. ➤ Fossil-based Energy • Coal, LNG, oil, etc. ➤ Nuclear Energy

For energy system preferences, the survey respondents were asked to indicate how important they felt each factor was using a Likert scale response ranging from 1 to 5. These responses were then summarized into the three categories of unimportant (responses in the range of 1–2), neutral (a response of 3), and important (responses in the range of 4–5). Similarly to the energy system design factors, where respondents were asked to indicate their preference toward increasing or decreasing specific energy sources, we used categories

such as decrease, neutral and increase, using a composite score for fossil fuels (oil, coal, natural gas, etc.) and renewables (solar, wind, hydro, etc.).

Machine learning modelling and algorithm performance comparisons are undertaken using RapidMiner Studio v9.10.010 utilizing a 64-bit architecture. The data preparation, modeling, comparison, and influential factor extraction process flow is summarized in Figure 1.

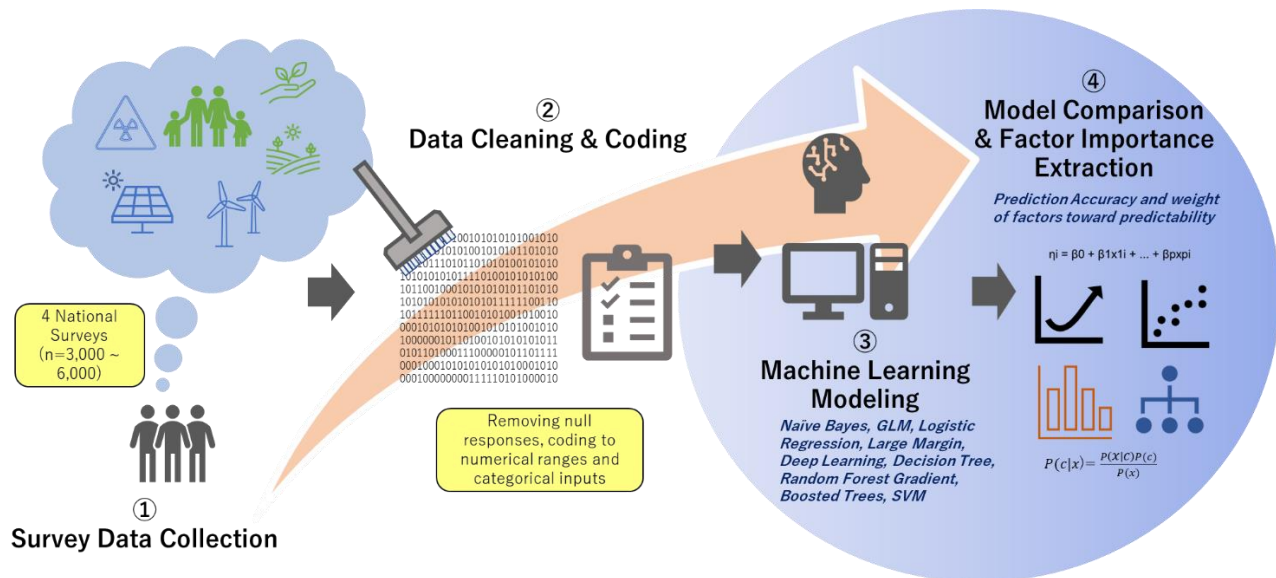


Figure 1. Workflow for data collection, cleaning, coding, model testing, and influential factor extraction.

The results are summarized to detail influential factors toward predicting desired targets for each survey, optimal machine learning models and their predictive ability (i.e., accuracy). Finally, sensitivity analysis is conducted to test whether combined survey samples improve prediction ability where survey variables and ranges allow.

4. Results

The results include machine learning algorithm predictive accuracy results, followed by the weight of the most influential inputs toward target variables. Hereafter, surveys are referred to by number, as identified in Table 1.

4.1. Energy System Design

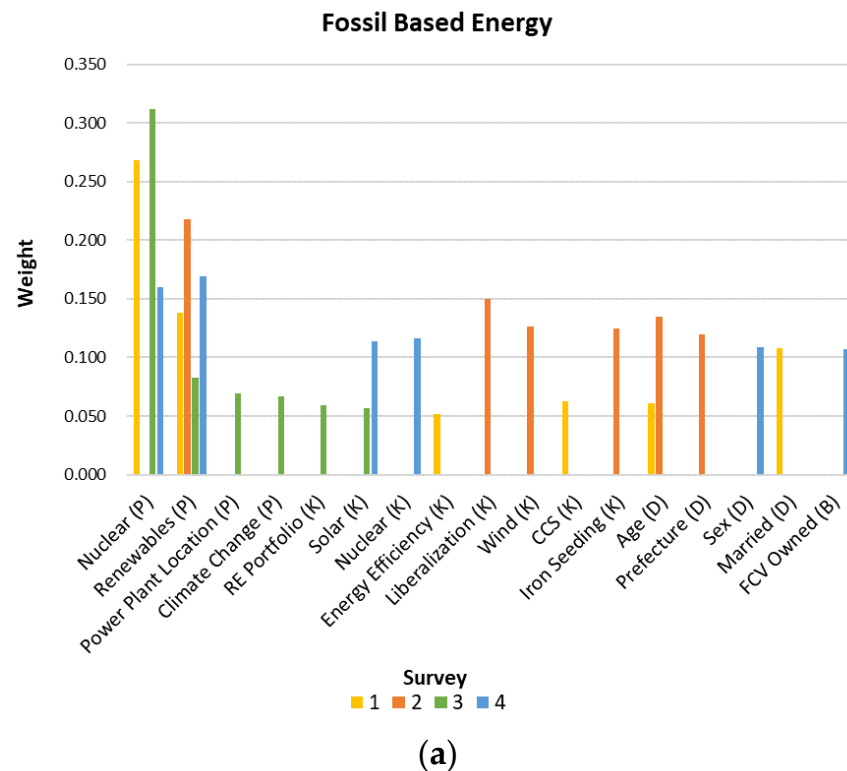
Energy system design factors are evaluated based on whether the respondents indicated a preference to increase or decrease fossil fuels (encoded FF), nuclear energy (NE) and renewable energy (RE) within the energy mix. Machine learning algorithm predictive accuracy for energy system design target factors is summarized in Table 3, with the best performing algorithms being identified in bold text.

The predictive model accuracy ranges from a low of 54.9% for predicting nuclear energy preferences in survey 3 to a high of 78.9% for predicting preferences toward renewable energy in survey 1. In all surveys, the prediction accuracy for renewable energy preferences was the highest. The sample size and the number of questions posed did not appear to heavily affect prediction accuracy for single survey analysis. No one machine learning approach was consistently superior, with GLM, large margin and gradient boosted trees models each having the best predictive ability three times (25% of the time, respectively), decision tree twice, and random forest once. Naïve Bayes, logistic regression, deep learning and SVM models did not demonstrate superior predictive ability for any of the survey targets.

Table 3. Energy system design machine learning algorithm predictive accuracy by target (best prediction percentage shaded and bold).

Survey		1			2			3			4		
Model	Target	FF	NE	RE	FF	NE	RE	FF	NE	RE	FF	NE	RE
		Accuracy (%)											
Naive Bayes		63.5	62.3	72.2	58.1	59.2	74.7	57.2	48.6	70.8	61.5	50.4	73.3
GLM		63.5	67.1	78.2	57.2	64.6	78.1	62.0	51.7	73.7	62.1	57.1	76.0
Logistic Regression		63.5	66.1	77.0	57.5	64.2	77.9	62.0	54.0	71.3	58.8	56.3	75.0
Large Margin		63.5	67.4	78.9	58.3	63.8	78.0	60.5	53.8	75.1	61.3	51.4	74.1
Deep Learning		63.6	66.4	75.2	58.0	63.6	76.9	58.2	54.1	72.1	59.4	53.5	75.0
Decision Tree		63.5	67.4	73.6	56.9	65.5	76.7	52.5	51.4	70.6	57.4	50.4	72.9
Random Forest		63.9	63.8	78.1	57.1	65.0	77.1	59.1	53.7	73.8	59.3	56.9	75.9
Gradient Boosted Trees		63.5	67.3	78.0	57.0	65.0	77.2	62.5	54.9	73.0	60.3	57.6	75.3
SVM		63.5	67.0	77.8	57.6	65.2	77.7	56.4	49.0	72.5	60.3	50.4	74.1

Utilizing the outcomes of the best performing models for each survey and factor, the most influential factors underpinning predictions can be extracted, as shown in Figure 2, according to their comparative weights. Response variables are categorized as preferences (P), knowledge levels (K), demographics (D), and behavior (B).

**Figure 2.** Cont.

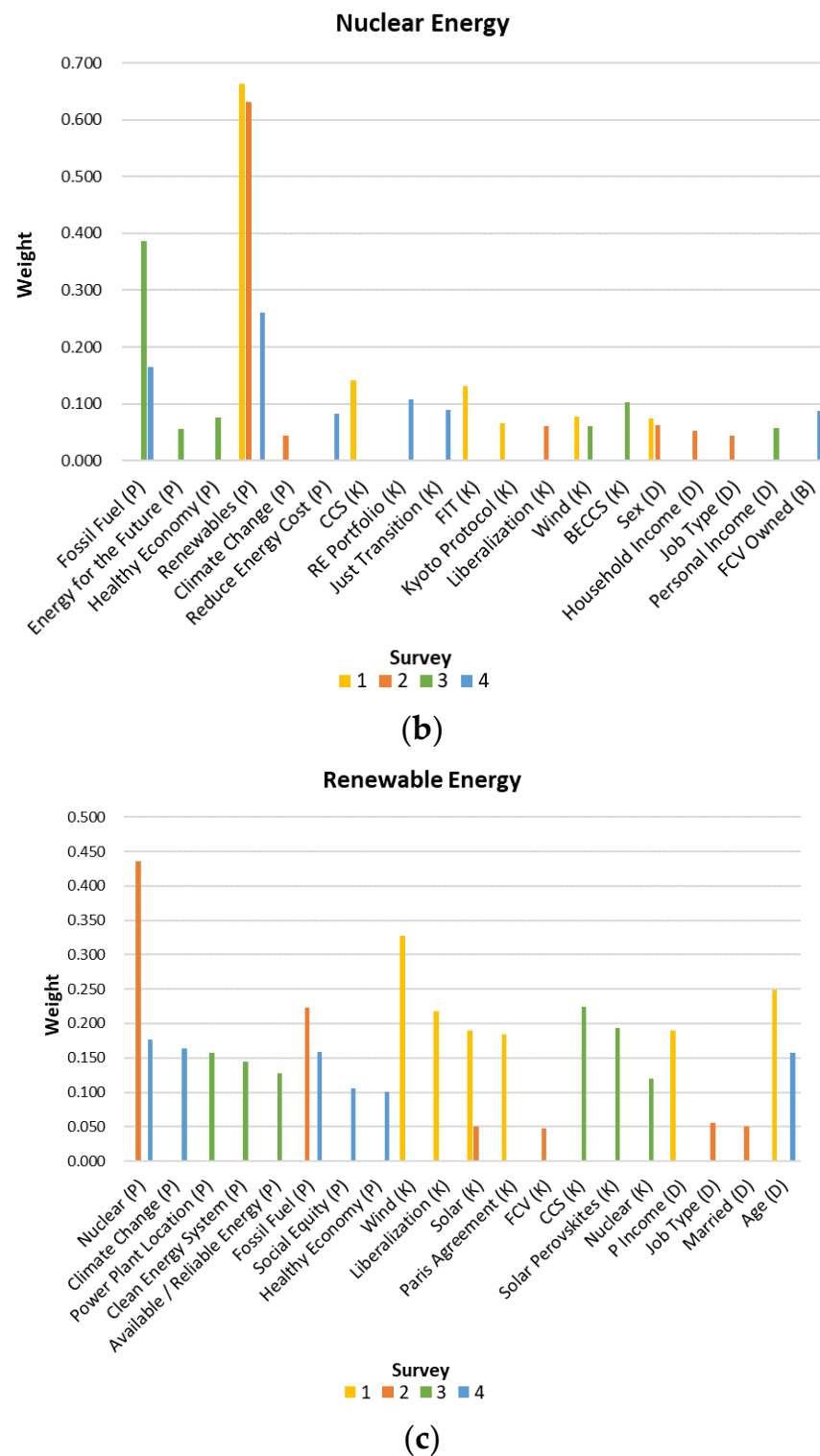


Figure 2. Influential factors toward energy system design preference prediction for (a) fossil-based, (b) nuclear and (c) renewable energy.

Energy system design preferences toward fossil fuels appear to be highly influenced by people's preferences toward nuclear and renewable energy types. Other commonly influential factors identified were knowledge of solar energy and the respondents' age. In terms of types of predictors, knowledge levels were the most commonly influential responses, followed by preferences and demographics, and behavior. For nuclear energy, renewable preferences were highly influential, followed, though to a lower degree, by fossil

fuel preferences. Commonly influential factors included knowledge of wind and the demographic factor of sex. Again, knowledge-based responses were most commonly influential on the predictability of energy system design preferences. For renewable energy preference prediction, the most accurate among energy types, fossil and nuclear energy preferences were influential but at a relatively lower level than for other energy types. Solar knowledge levels and the age demographic factor were commonly influential across surveys.

4.2. Energy System Preferences

The energy system preferences tested include environmental protection (coded as EP), climate change response (CC), and social equity (SE); these acted as proxies for equitable and sustainable energy systems (i.e., desirable in accordance with a just transition). Machine learning algorithm predictive accuracy is detailed for each energy system preference target variable for surveys 3 and 4 (the two surveys exploring these factors) in Table 4. The best performing algorithms are identified in bold text.

Table 4. Energy system preference machine learning algorithm predictive accuracy by target (best prediction percentage shaded and bold).

Survey		3			4		
Model	Target	EP	CC	SE	EP	CC	SE
	Accuracy (%)						
Naive Bayes		68.3	58.9	56.5	65.7	65.8	60.4
GLM		71.6	62.1	58.0	70.8	68.9	69.3
Logistic Regression		71.9	62.1	58.2	70.0	66.9	68.7
Large Margin		70.9	64.1	58.8	71.0	67.5	68.6
Deep Learning		70.4	63.0	56.8	68.6	69.7	67.3
Decision Tree		69.9	57.5	54.7	66.0	69.1	71.5
Random Forest		73.0	60.4	59.3	71.3	67.9	70.4
Gradient Boosted Trees		72.9	63.0	61.1	72.2	71.9	72.2
SVM		71.4	62.7	60.1	70.0	69.1	68.2

The results show that the predictability of people's preference toward environmental protection is higher than that of climate change response and social equity. The gradient boosted trees algorithm was the most accurate in four out of six predictions, having the second highest performance in the remainder of cases.

The six most influential response variables toward the prediction of energy system preferences from each analyzed survey are detailed in Figure 3. As was the case for energy system design, the response variables are categorized as either preferences (P), knowledge levels (K), demographics (D), and behavior (B).

As shown in Figure 3, preferences regarding the energy system and energy sources are the most influential toward predicting respondent's overall views toward the importance of environmental protection, climate change response and social equity. Some preferences are common across surveys, notably respondent's preferences toward solar power in the future energy mix, and the desire for energy availability. Knowledge of energy types was influential toward environmental protection and climate change response; however, this was not the case for social equity importance. In terms of influential demographics, age was influential toward environmental protection importance, sex was influential on climate change response and social equity importance, and in the case of the US survey (Survey 3), ethnicity was also influential.

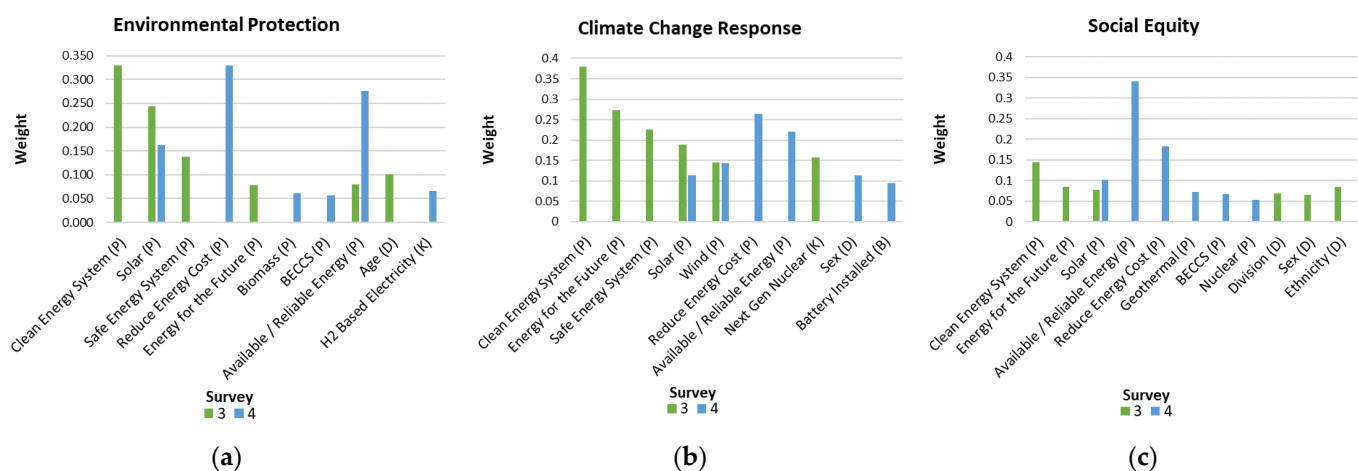


Figure 3. Influential factors toward energy system preferences for (a) environmental protection, (b) climate change response, (c) social equity.

4.3. Sensitivity Analysis

As the addition of data points has been empirically shown to improve machine learning performance, albeit with some caveats such as introducing samples from different time periods [46], our sensitivity analysis combines compatible surveys to develop larger datasets, each within 2 years of each other.

Following the combination of compatible surveys, i.e., surveys 1 and 2 for energy system design factors and surveys 3 and 4 for energy system preferences, we repeat the machine learning multi-model prediction analysis and contrast the accuracy results. As the number of questions (input variables) is limited to the survey with the smallest number of usable questions, predictive accuracy is contrasted with survey 2 for energy system design factors, and survey 3 for energy system preferences, with the results detailed in Figure 4.

For fossil fuel and nuclear energy preference predictability, the combined survey samples offered consistently higher accuracy across all models ranging between 1.7 and 3.7% for fossil fuels and 0.9 and 2.6% for nuclear energy, respectively. In the case of renewable energy preference predictability, which was predicted at a much higher level of accuracy than that for fossil fuels or nuclear energy, the combined sample did not improve predictability in all cases, with changes ranging in a tight band from -0.9% to 0.6% . The only models which improved predictive accuracy were the deep learning (0.3%) and gradient boosted trees (0.6%) algorithms, suggesting an upper limit to the predictive accuracy for this dataset. For fossil fuel preference predictability, the top influential factors of preferences toward nuclear and renewables remain relevant, along with age and technology knowledge for fossil fuel preferences. For nuclear, preferences toward renewable energy deployment remained overwhelmingly the top response. Finally, for renewable energy, attitudes toward nuclear and fossil fuel deployment remain influential, along with technology knowledge.

For energy system preference predictability, an increase in accuracy was again observed for all factors, ranging from 5.1 to 7.5% for environmental protection, 11.2 to 21% for climate change response and 8.4 to 12.7% for social equity. Within energy system preferences, the combined survey samples identified that for environmental protection, preferences toward the energy system (resource preservation, energy cost) and technology knowledge and energy mix preferences toward solar were important, consistent with single sample findings. For climate change response, preserving limited resources and thus ensuring energy for the future was most influential toward predictability, along with technology and policy knowledge. For social equity, preferences toward the future energy system were the most important, particularly regarding energy availability and cost.

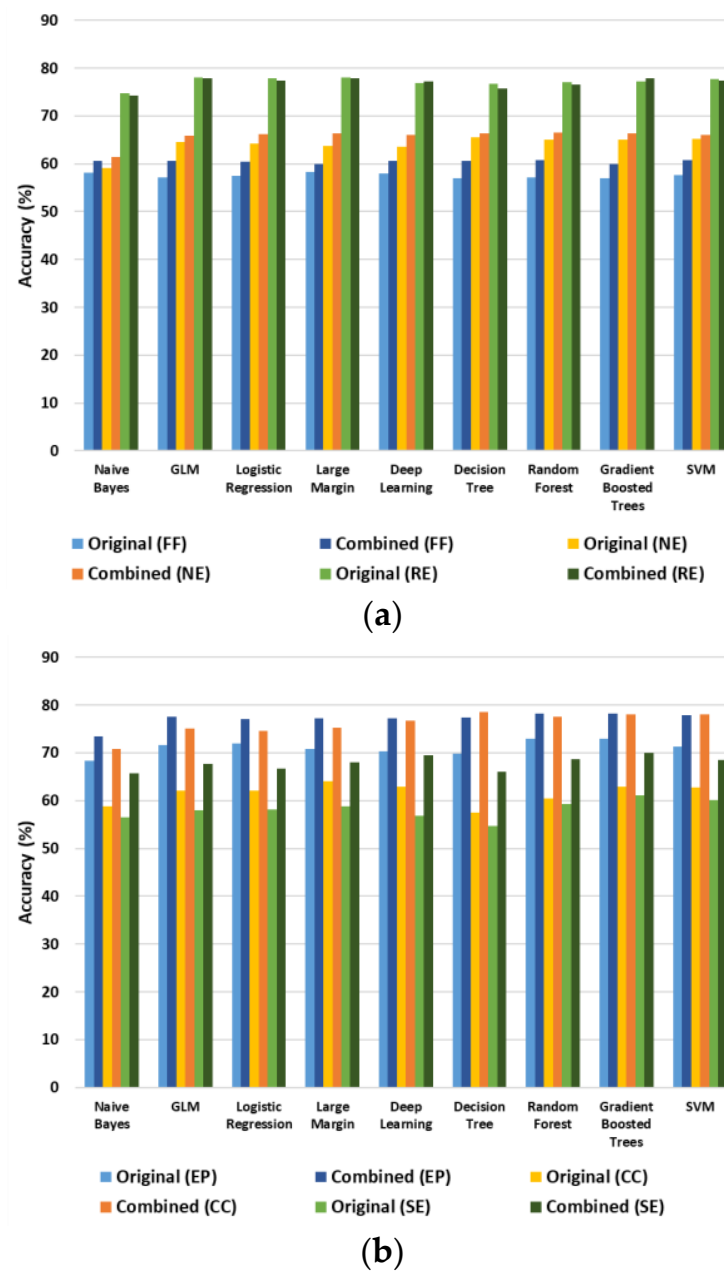


Figure 4. Combined survey predictive performance change compared to single survey analysis for (a) energy system design and (b) energy system preferences.

Generally speaking, a higher number of samples improves the model's overall prediction accuracy, with a maximum accuracy of 78.2% being achieved for environmental protection preferences; however, it should be noted that before combining samples, a maximum predictive accuracy of 78.1% had been achieved for social equity, suggesting an upper limit on predictability for energy system preferences being reached. There is also a suggestion that the ratio of sample increase may also be linked to model predictive accuracy increase.

5. Discussion and Implications

Previous research has explored how people's preferences can influence energy system design; for example, in Vietnam, a site selection tool for the future deployment of solar and other renewables was developed based on experts' opinions, knowledge and judgements [47]. In the US, factors such as ethnicity, region and lived experience were

found to be influential on the bottom-up-based development of a desirable energy mix [43]. People's aversion to risk-taking, curiosity, environmental awareness and overall active-ness toward energy system participation was found to be influential on future energy provider selection in Japan [42]. In Spain, householder's propensity toward improving their homes' insulation was found to be linked to demographic factors such as age and income level, as well as the heating technologies employed in their home [48]. Likewise, in India, socio-demographic factors such as age, gender, educational achievement, job type and vehicle ownership influence people's propensity toward investing in solar PV or car-charging infrastructure [49]. In a multi-nation study across Romania, Hungary and Serbia, it was identified that peoples' preference toward participation in demand response initiatives is largely contingent on a perceived increase in renewable energy deployment as a result, reducing carbon emissions and global warming, and with the aim of reducing energy bills [50]. In light of this, recent research has suggested that both end-users and experts can influence the future energy system through their preferences, behavior and perceived benefits; this research offers a methodology for the rapid acquisition of these preferences through previously conducted national surveys (the method used in the majority of these studies), utilizing machine learning to uncover critical preferences, knowledge and demographics toward the prediction of target variables.

Examining the influence weightings of energy system design preferences toward fossil, nuclear and renewables in the future energy mix, we found that preferences related to generation technologies, followed by policy factors, were the most important. Demographics were influential at a much lower level, and only consistently for age, toward renewables and fossil fuels. The only influential behavior was the ownership of an FCV toward fossil and nuclear energy at a similar level to the influence of demographic factors. These findings have some commonality with those of previous research, specifically for the socio-demographic factors of vehicle ownership and age. For energy system preferences toward environmental protection, climate change response and social equity, preferences toward technologies and lifestyle factors were most influential, significantly more so than demographics or knowledge levels. Although demographics appear to play a small role toward people's preferences with regard to energy systems, they were not as influential as hoped, as these kinds of data can be elicited from other sources, including census records or consumer survey data.

In terms of machine learning algorithms utilized, gradient boosted trees showed excellent performance overall for training and scoring times, in addition to being consistently the most accurate. This algorithm has proven to be popular in social science research and has multiple applications toward classification, including human behavior-related factors [51,52]. On the other hand, the SVM algorithm, although intensive in terms of computing resources and solution times, was not superior in our classification models; however, as has been noted in previous sentiment-based research, it generally offers accuracy advantages over the less intensive, simpler Naïve Bayes classification algorithm [53].

The availability of additional data appears to improve overall accuracy, as shown by the combination of survey samples, and, although not conclusive with the limited investigation offered in this research, it appears to scale somewhat according to overall sample sizes in line with expectations [54]. Overall, irrespective of the size of the samples processed, ranging from 4148 to 9000 in our study, the prediction accuracy never exceeds 79.8%. Although this accuracy is sufficient for our purposes, further investigation is required to uncover related factors that may increase accuracy in the future. Machine learning has proven to be accessible and useful for the prediction of influential factors of energy system preferences and design; however, some factors have proven to be more difficult to predict than others. In this study, for energy system design factors, opinions toward renewable energy proved easier to predict than for fossil fuels and nuclear power. This may be due to an overwhelming preference toward the increased deployment of these types of energy. For energy system preferences, environmental protection opinions were the most accurately predicted. In this case, the ease of understanding of climate change and

social equity concepts may play a role, and the relationship between other demographic factors (education, age, income, etc.) and these concepts may require further investigation.

Although some work has been undertaken regarding energy and eco-efficiency utilizing machine learning, this study is unique in that it focuses specifically on energy system preferences and underpinning demographics to aid in energy system design. Through the provision of a methodology for the rapid acquisition of preferences that can be utilized for energy system design, as well as appropriate weightings of energy system technology deployment preferences (i.e., stakeholder desires for future deployment of fossil fuels, renewable and nuclear), survey design can be improved toward deriving sustainable, desirable energy systems.

The findings of this research provide a framework for the investigation of the linkage between people's preferences, knowledge, demographics and behaviors and desirable energy system design based on survey data over a period of time and in two discrete jurisdictions. This approach could be generalized to other research areas, and could be utilized in future survey design, streamlining survey design processes and allowing for a targeted selection of questions, reducing survey-related burdens.

6. Conclusions

In the age of big data and increased data availability, the operations described in this research are likely useful toward reducing a significant portion of menial work involved in survey preparation, analysis and application. In addition, it is hoped that cost-prohibitive portions of survey design and implementation, including workshops and question-testing, can be streamlined, and that the predictive ability of surveys can be improved by utilizing machine learning in research workflows.

A lower number of questions ultimately result in less demanding surveys, increased response rates and sample sizes achievable from available budgets while reducing respondent fatigue. Considering that machine learning is relatively accessible and, due to our finding that resource-intensive algorithms such as SVM do not necessarily offer predictability advantages, less resource-intensive algorithms that can be comfortably deployed on mid-range hardware have been identified as sufficient and suitable to the task. Among these algorithms, gradient boosted trees showed excellent performance in predicting influential factors of energy system preferences and design, and offering accuracy advantages over simpler algorithms such as naïve Bayes classification. Within predictions, people's preferences related to generation technologies and policy factors were identified as the most important factors influencing energy system design preferences, followed by demographic factors.

Furthermore, by employing a human-guided artificial intelligence-based workflow, we do not seek to automate our research processes; thus, we ensured human oversight and the provision of expertise, in line with the concept of making artificial intelligence engagement trustworthy [55]. This research identifies insights toward the rapid, fit-for-purpose deployment of surveys and their utilization toward energy system design.

Our proposed methodology has identified key factors to be investigated in future surveys, and thus meets our stated goals of reducing survey-related burdens for both researchers and respondents alike. With regard to future work, the investigation of the leveragability of publicly available data, including big data and open access databases, to assess their utilization toward predictive ability would be a useful endeavor. Ideally, the design of desirable energy systems can be a positive outcome based on this work, leading to the meeting of energy system goals at a national level while deriving desirable and equitable outcomes for individuals.

The findings of this research provide a framework for investigating the linkages between people's preferences, knowledge, demographics, behaviors, and desirable energy system design, with potential applications in survey design and targeted question selection to streamline processes and reduce survey-related burdens, which is broadly applicable to a number of research fields.

Funding: This research was funded by the Japanese Society for the Promotion of Science under Kaken Grant 22K18039, Social Energy System Design Incorporating AI and Lived Experience.

Data Availability Statement: Data will be made available upon reasonable request.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Demographics.

#	Sex	Age	Race	State	Residence Length	Education	Married	Children	Household Income	Personal Income	Job Type	Household Members	Dwelling Type	Drivers License
1	☒	☒		☒			☒	☒	☒	☒	☒	☒	☒	
2	☒	☒		☒			☒	☒	☒	☒	☒	☒	☒	☒
3	☒	☒	☒	☒		☒				☒				
4	☒	☒		☒	☒	☒								

Table A2. Technology knowledge.

#	Solar PV	Wind	Biomass	Geother	Hydro	Hydroge	CCS	CCU	BECCS	FCV	Nuclear	Fossil Fuel	Iron Seeding
1	☒	☒	☒				☒	☒		☒			
2	☒	☒	☒				☒	☒		☒	☒		☒
3	☒	☒	☒	☒	☒		☒	☒	☒		☒	☒	
4	☒	☒	☒	☒	☒	☒	☒		☒		☒	☒	

Table A3. Policy knowledge.

#	Liberalization	Energy Efficiency	Paris Agreement	INDC	FIT	COP	RE Portfolio	Kyoto Protocol	IPCC	SDGs	Energy Mix	Just Transition
1	☒	☒	☒	☒	☒	☒		☒	☒			
2	☒											
3			☒	☒	☒	☒	☒	☒	☒	☒	☒	
4			☒	☒	☒	☒	☒	☒	☒	☒	☒	☒

Table A4. Technology preferences.

#	Coal	LNG	Oil	Nuclear	Solar	Wind	Biomass	Hydro	Geothermal
1	☒	☒	☒	☒	☒	☒	☒	☒	☒
2	☒	☒	☒	☒	☒	☒	☒	☒	☒
3	☒	☒	☒	☒	☒	☒	☒	☒	☒
4	☒	☒		☒	☒	☒	☒	☒	☒

Table A5. Lifestyle preferences.

#	Electricity Cost	Gas Cost	Preserving Resources	Healthy Economy	Convenient Lifestyle	Energy Availability	Energy Affordability	Non-Polluting Energy	Energy for the Future	Safe Energy	No Energy Plants Near Home
1	☒	☒									
2											
3	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
4			☒	☒	☒	☒					

Table A6. Behaviors.

#	Installed Solar PV	Installed Battery	FCV Owned	EV Owned	Changed Energy Provider
1	☒				
2	☒				☒
3					
4	☒	☒	☒	☒	

References

1. Bouzarovski, S. Just Transitions: A Political Ecology Critique. *Antipode* **2022**, *54*, 1003–1020. [\[CrossRef\]](#)
2. Hanbury, H.; Bader, C.; Moser, S. Reducing Working Hours as a Means to Foster Low(Er)-Carbon Lifestyles? An Exploratory Study on Swiss Employees. *Sustainability* **2019**, *11*, 2024. [\[CrossRef\]](#)
3. Goddard, G.; Farrelly, M.A. Just Transition Management: Balancing Just Outcomes with Just Processes in Australian Renewable Energy Transitions. *Appl. Energy* **2018**, *225*, 110–123. [\[CrossRef\]](#)
4. Reddy, B.S. India's Energy System Transition-Survival of the Greenest. *Renew. Energy* **2016**, *92*, 293–302. [\[CrossRef\]](#)
5. Fragkos, P.; Fragkiadakis, K.; Sovacool, B.; Paroussos, L.; Vrontisi, Z.; Charalampidis, I. Equity Implications of Climate Policy: Assessing the Social and Distributional Impacts of Emission Reduction Targets in the European Union. *Energy* **2021**, *237*, 121591. [\[CrossRef\]](#)
6. Chapman, A.; Shigetomi, Y.; Ohno, H.; McLellan, B.; Shinozaki, A. Evaluating the Global Impact of Low-Carbon Energy Transitions on Social Equity. *Environ. Innov. Soc. Transit.* **2021**, *40*, 332–347. [\[CrossRef\]](#)
7. Fell, M.J.; Pye, S.; Hamilton, I. Capturing the Distributional Impacts of Long-Term Low-Carbon Transitions. *Environ. Innov. Soc. Transit.* **2020**, *35*, 346–356. [\[CrossRef\]](#)
8. Frondel, M.; Sommer, S.; Vance, C. The Burden of Germany's Energy Transition: An Empirical Analysis of Distributional Effects. *Econ. Anal. Policy* **2015**, *45*, 89–99. [\[CrossRef\]](#)
9. Cuppen, E.; Bosch-Rekveltdt, M.G.C.; Pikaar, E.; Mehos, D.C. Stakeholder Engagement in Large-Scale Energy Infrastructure Projects: Revealing Perspectives Using Q Methodology. *Int. J. Proj. Manag.* **2016**, *34*, 1347–1359. [\[CrossRef\]](#)
10. Sharpe, L.M.; Harwell, M.C.; Jackson, C.A. Integrated Stakeholder Prioritization Criteria for Environmental Management. *J. Environ. Manag.* **2021**, *282*, 111719. [\[CrossRef\]](#)
11. Dvarioniene, J.; Gurauskienė, I.; Gecevičius, G.; Trummer, D.R.; Selada, C.; Marques, I.; Cosmi, C. Stakeholders Involvement for Energy Conscious Communities: The Energy Labs Experience in 10 European Communities. *Renew. Energy* **2015**, *75*, 512–518. [\[CrossRef\]](#)
12. Boaz, A.; Hanney, S.; Borst, R.; O'Shea, A.; Kok, M. How to Engage Stakeholders in Research: Design Principles to Support Improvement. *Health Res. Policy Syst.* **2018**, *16*, 60. [\[CrossRef\]](#)
13. Walker, G.; Bulkeley, H. Geographies of Environmental Justice. *Geoforum* **2006**, *37*, 655–659. [\[CrossRef\]](#)
14. Bulkeley, H.; Carmin, J.A.; Castán Broto, V.; Edwards, G.A.S.; Fuller, S. Climate Justice and Global Cities: Mapping the Emerging Discourses. *Glob. Environ. Chang.* **2013**, *23*, 914–925. [\[CrossRef\]](#)
15. Pettit, J. Climate Justice: A New Social Movement for Atmospheric Rights. *IDS Bull.* **2004**, *35*, 102–106. [\[CrossRef\]](#)
16. McCauley, D.; Ramasar, V.; Heffron, R.J.; Sovacool, B.K.; Mebratu, D.; Mundaca, L. Energy Justice in the Transition to Low Carbon Energy Systems: Exploring Key Themes in Interdisciplinary Research. *Appl. Energy* **2019**, *233–234*, 916–921. [\[CrossRef\]](#)
17. Heffron, R.J.; McCauley, D. The Concept of Energy Justice across the Disciplines. *Energy Policy* **2017**, *105*, 658–667. [\[CrossRef\]](#)
18. Gummer, T.; Christmann, P.; Verhoeven, S.; Wolf, C. Using a Responsive Survey Design to Innovate Self-Administered Mixed-Mode Surveys. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2022**, *185*, 916–932. [\[CrossRef\]](#)
19. Cuppen, M.; Van Der Laan, P.; Van Nunspeet, W. Re-Engineering Dutch Social Surveys: From Single-Purpose Surveys to an Integrated Design. *Stat. J. IAOS* **2013**, *29*, 21–29. [\[CrossRef\]](#)
20. De Abreu, E.; Silva, J.; De Oña, J.; Gasparovic, S. The Relation between Travel Behaviour, ICT Usage and Social Networks. The Design of a Web Based Survey. *Transp. Res. Procedia* **2017**, *24*, 515–522. [\[CrossRef\]](#)
21. Smith, S.; Jaszczak, A.; Graber, J.; Lundeen, K.; Leitsch, S.; Wargo, E.; O'Muircheartaigh, C. Instrument Development, Study Design Implementation, and Survey Conduct for the National Social Life, Health, and Aging Project. *J. Gerontol.-Ser. B Psychol. Sci. Soc. Sci.* **2009**, *64*, 20–29. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Ioannidis, E.; Merkouris, T.; Zhang, L.C.; Karlberg, M.; Petrakos, M.; Reis, F.; Stavropoulos, P. On a Modular Approach to the Design of Integrated Social Surveys. *J. Off. Stat.* **2016**, *32*, 259–286. [\[CrossRef\]](#)
23. Wardropper, C.B.; Dayer, A.A.; Goebel, M.S.; Martin, V.Y. Conducting Conservation Social Science Surveys Online. *Conserv. Biol.* **2021**, *35*, 1650–1658. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Sue, V.; Ritter, L. Writing Survey Questions. In *Conducting Online Surveys*; SAGE: Thousand Oaks, CA, USA, 2015; pp. 51–75. [\[CrossRef\]](#)
25. General Medical Council. *General Medical Council Survey Design-Best Practice Guidelines*; General Medical Council: London, UK, 2016; pp. 1–8.
26. Fisher, S. How to Create an Effective Survey. 2020. Available online: <https://www.qualtrics.com/blog/how-to-create-a-survey/> (accessed on 6 July 2022).
27. Harrison, C. Questionnaire Design Tip Sheet. *Harvard Univ. Progr. Surv. Res.* 2007. Available online: <https://psr.iq.harvard.edu/files/psr/files/PSRQuest> (accessed on 5 July 2022).
28. Piroddi, L. *Special Topics in Information Technology*; SpringerBriefs in Applied Sciences and Technology; Piroddi, L., Ed.; Springer International Publishing: Cham, Switzerland, 2022; ISBN 978-3-030-85917-6.
29. Shah, N.; Mohan, D.; Bashingwa, J.J.H.; Ummer, O.; Chakraborty, A.; LeFevre, A.E. Using Machine Learning to Optimize the Quality of Survey Data: Protocol for a Use Case in India. *JMIR Res. Protoc.* **2020**, *9*, e17619. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Chan, K.Y.; Kwong, C.K.; Wongthongtham, P.; Jiang, H.; Fung, C.K.Y.; Abu-Salih, B.; Liu, Z.; Wong, T.C.; Jain, P. Affective Design Using Machine Learning: A Survey and Its Prospect of Conjoining Big Data. *Int. J. Comput. Integr. Manuf.* **2020**, *33*, 645–669. [\[CrossRef\]](#)
31. Akour, I.; Alshurideh, M.; Al Kurdi, B.; Al Ali, A.; Salloum, S. Using Machine Learning Algorithms to Predict People's Intention to Use Mobile Learning Platforms during the COVID-19 Pandemic: Machine Learning Approach. *JMIR Med. Educ.* **2021**, *7*, e24032. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Gharibshah, Z.; Zhu, X. User Response Prediction in Online Advertising. *ACM Comput. Surv.* **2021**, *54*, 1–43. [\[CrossRef\]](#)
33. Mirmozaffari, M.; Yazdani, M.; Boskabadi, A.; Dolatsara, H.A.; Kabirifar, K.; Golilarz, N.A. A Novel Machine Learning Approach Combined with Optimization Models for Eco-Efficiency Evaluation. *Appl. Sci.* **2020**, *10*, 5210. [\[CrossRef\]](#)
34. Himeur, Y.; Alsalemi, A.; Al-Kababji, A.; Bensaali, F.; Amira, A.; Sardianos, C.; Dimitrakopoulos, G.; Varlamis, I. A Survey of Recommender Systems for Energy Efficiency in Buildings: Principles, Challenges and Prospects. *Inf. Fusion* **2021**, *72*, 1–21. [\[CrossRef\]](#)
35. Zhang, W.; Wu, Y.; Calautit, J.K. A Review on Occupancy Prediction through Machine Learning for Enhancing Energy Efficiency, Air Quality and Thermal Comfort in the Built Environment. *Renew. Sustain. Energy Rev.* **2022**, *167*, 112704. [\[CrossRef\]](#)
36. Qavidel Fard, Z.; Zomorodian, Z.S.; Korsavi, S.S. Application of Machine Learning in Thermal Comfort Studies: A Review of Methods, Performance and Challenges. *Energy Build.* **2022**, *256*, 111771. [\[CrossRef\]](#)
37. Bose, B.K. Artificial Intelligence Techniques in Smart Grid and Renewable Energy Systems-Some Example Applications. *Proc. IEEE* **2017**, *105*, 2262–2273. [\[CrossRef\]](#)
38. Perera, A.T.D.; Wickramasinghe, P.U.; Nik, V.M.; Scartezzini, J.-L. Machine Learning Methods to Assist Energy System Optimization. *Appl. Energy* **2019**, *243*, 191–205. [\[CrossRef\]](#)
39. Perera, A.T.D.; Wickramasinghe, P.U.; Nik, V.M.; Scartezzini, J.-L. Introducing Reinforcement Learning to the Energy System Design Process. *Appl. Energy* **2020**, *262*, 114580. [\[CrossRef\]](#)
40. Köhnen, C.S.; Priesmann, J.; Nolting, L.; Kotzur, L.; Robinius, M.; Praktiknjo, A. The Potential of Deep Learning to Reduce Complexity in Energy System Modeling. *Int. J. Energy Res.* **2022**, *46*, 4550–4571. [\[CrossRef\]](#)
41. Chapman, A.; Itaoka, K. Curiosity, Economic and Environmental Reasoning: Public Perceptions of Liberalization and Renewable Energy Transition in Japan. *Energy Res. Soc. Sci.* **2018**, *37*, 102–110. [\[CrossRef\]](#)
42. Itaoka, K.; Chapman, A.; Farabi-Asl, H. Underpinnings of Consumer Preferences and Participation in Japan's Liberalized Energy Market. *Util. Policy* **2022**, *76*, 101379. [\[CrossRef\]](#)
43. Chapman, A.; Shigetomi, Y.; Chandra Karmaker, S.; Baran Saha, B.; Huff, K.; Brooks, C.; Stubbins, J. The Cultural Dynamics of Energy: The Impact of Lived Experience, Preference and Demographics on Future Energy Policy in the United States. *Energy Res. Soc. Sci.* **2021**, *80*, 102231. [\[CrossRef\]](#)
44. Chapman, A.; Shigetomi, Y.; Chandra, S.; Saha, B.; Brooks, C. Cultural and Demographic Energy System Awareness and Preference: Implications for Future Energy System Design in the United States. *Energy Econ.* **2022**, *112*, 106141. [\[CrossRef\]](#)
45. Mabon, L.; Chapman, A.; Mclellan, B.; Huang, Y. *Just Transitions in Japan*; The British Academy: London, UK, 2022.
46. Gobeill, J.; Ruch, P.; Meyer, R. Machine Learning for Automatic Encoding of French Electronic Medical Records: Is More Data Better? *Stud. Health Technol. Inform.* **2020**, *270*, 312–316. [\[CrossRef\]](#)
47. Wang, C.N.; Dang, T.T.; Nguyen, N.A.T.; Wang, J.W. A Combined Data Envelopment Analysis (DEA) and Grey Based Multiple Criteria Decision Making (G-MCDM) for Solar PV Power Plants Site Selection: A Case Study in Vietnam. *Energy Rep.* **2022**, *8*, 1124–1142. [\[CrossRef\]](#)
48. Fernandez-Luzuriaga, J.; Flores-Abascal, I.; del Portillo-Valdes, L.; Mariel, P.; Hoyos, D. Accounting for Homeowners' Decisions to Insulate: A Discrete Choice Model Approach in Spain. *Energy Build.* **2022**, *273*, 112417. [\[CrossRef\]](#)
49. Murugan, M.; Marisamynathan, S. Investigating the Individual House Holders' Preference to Adopt Home-Based Charging and Solar Rooftop Facility for Electric Vehicle Charging. *Transp. Lett.* **2022**, 1–4, *In Press, Corrected Proof*. [\[CrossRef\]](#)
50. Tantau, A.; Puskás-Tompos, A.; Stanciu, C.; Fratila, L.; Curmei, C. Key Factors Which Contribute to the Participation of Consumers in Demand Response Programs and Enable the Proliferation of Renewable Energy Sources. *Energies* **2021**, *14*, 8273. [\[CrossRef\]](#)
51. Lou, Y.; Obukhov, M. BDT: Gradient Boosted Decision Tables for High Accuracy and Scoring Efficiency. In Proceedings of the KDD'17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13–17 August 2017; Part F1296. pp. 1893–1901. [\[CrossRef\]](#)
52. Priyadarshini, R.K.; Bazila Banu, A.; Nagamani, T. Gradient Boosted Decision Tree Based Classification for Recognizing Human Behavior. In Proceedings of the 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE), Sathyamangalam, India, 4–6 April 2019; pp. 1–5. [\[CrossRef\]](#)
53. Kusumawati, R.; D'Arofah, A.; Pramana, P.A. Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter's Classification of Tokopedia Services. *J. Phys. Conf. Ser.* **2019**, *1320*, 012016. [\[CrossRef\]](#)
54. Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
55. Kaur, D.; Uslu, S.; Rittichier, K.J.; Duresi, A. Trustworthy Artificial Intelligence: A Review. *ACM Comput. Surv.* **2023**, *55*, 1–38. [\[CrossRef\]](#)

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