Modeling short-term solar energy generation: an integrated approach

Eric K. H. Leung
School of Management,
Univ. of Liverpool
Liverpool, L69 7ZH, United Kingdom
K.H.Leung@liverpool.ac.uk

Mark C. P. Poo
Liverpool Logistics, Offshore
and Marine Research Institute
Liverpool John Moores Univ.
Liverpool, United Kindgom
C.P.Poo@ljmu.ac.uk

***Abstract*—The non-renewable energy generation process emits undesirable CO_2 emissions which has long been regarded as a threat to our environment, and ultimately, human beings. Renewable energy is perceived as a viable solution in response to transitioning to a greener future and tackling climate change. However, the major challenge associated with most renewable energy sources is the intermittency caused by fluctuating weather conditions. This paper proposes an integrated approach in predicting the short-term solar energy generation based on changing weather conditions. The proposed approach is generic and thus can be treated as a systematic framework of predicting the generation of different renewable energy. An illustrative example is provided, demonstrating the practicability of the approach.**

Keywords—Climate change, Solar energy generation, Short-term real-time prediction, Big data prescriptive analytics

I. Introduction

Climate change adaption is unavoidable [1]. This subject has received a growing attention in recent years due to the more frequent occurrence of extreme weather events [2]. Energy consumption and generation both have a strong relation to climate change. The dramatic increase in global energy consumption owing to a higher standard of living and the increasing world population is a major cause of the global warming problem [3]. Today, most of our energy, still, is generated from burning fossil fuels. The energy generation process emits undesirable CO_2 emissions which has long been regarded as a threat to our environment, and ultimately, human beings. Renewable energy is a viable solution to transitioning to a greener future and tackling climate change. Nonetheless, electricity is the basis for most of the modern-day functions and thus the availability of constant sources of electricity is of fundamental importance. The major challenge associated with most renewable energy sources is the intermittency caused by fluctuating weather conditions [4]. Renewable energy sources are most influenced by the geographical location as the energy generation depends on environmental and weather-related factors, such as wind, rainfall, cloud cover, sunshine, etc.

Solar energy is free and clean in most places throughout the year [5]. This energy is the most abundant renewable resource and therefore much of the focus on sustainable energy is targeting optimum solar energy exploitation [6]. Like many other renewable energy sources, solar energy is weather-dependent. In other words, geographical locations with rapidly changing weather conditions heavily affect the efficiency of generating solar energy. This arises the need of predicting solar energy generation. In the literature, many solar energy models have been presented. Khatib et al. [7] review solar energy modeling techniques which are classified based on the nature of the modeling technique, such as linear, nonlinear functions, artificial neural networks (ANN), and fuzzy logic. They summarize five major challenges in modeling solar energy: prediction accuracy, model simplicity, model inputs, availability of data, and architecture of ANN models. For instance, they mention that heuristics have been proven to be capable of predicting solar energy generation as compared to linear and nonlinear optimization method. The simplicity of the models is of prime importance when considering which models to be adopted. There are inputs such as forecasted weather conditions currently not integrated into the modeling of solar energy. Finally, gathering a long-term weather data is also a challenging issue in solar energy prediction. Taking these challenges into account, this study develops an integrated approach, utilizing multiple regression, association rule mining and fuzzy logic techniques, for modeling solar energy. The rationale of developing this integrated model can be explained according to the above specified solar energy modeling challenges:

Model simplicity and inputs – we combine multiple regression, association rule mining and fuzzy logic techniques, three simple, yet important tools that are able to extract essential factors affecting the solar energy production the most and subsequently predict the dynamic changes in solar energy generation solar energy generation.

Availability of data – India and China are the two countries with the highest consumption of energy, and hence, also the biggest polluters [8]. This study, therefore, takes data from India and develops a model for a selected Indian solar plant.

The rest of this paper is organized as follows. Our integrated solar energy prediction model is presented in Section II. An illustrative example is given in Section III. Implications and discussions are provided in Section IV, followed by concluding remarks in Section V.

II. An Integrated Approach of Solar Energy Generation Prediction

The integrated approach of modeling solar energy is comprised of three techniques, multiple regression, association rule mining and fuzzy logic, as depicted Fig. 1. We systematically discuss the use of each technique as follows.

Multiple regression – It explains the relationship between various independent or multiple predictor variables and one

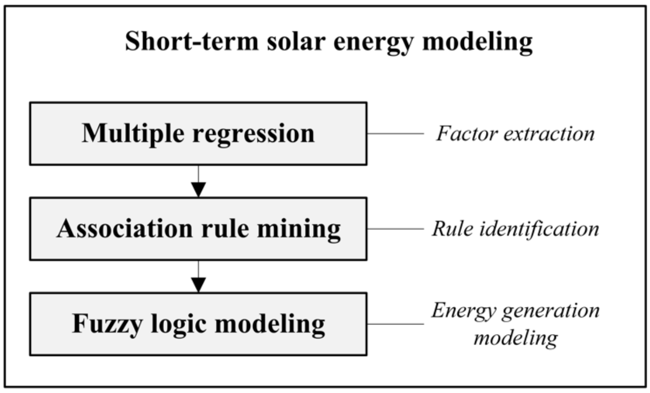

Fig. 1. An intrgrated approach of short-term solar energy modeling.

dependent or criterion variable [9]. In our model, it is used to discover the factors that most affect the solar energy generation.

Association rule mining – It is a commonly used data mining technique to explore and discover patterns and rules in a large dataset. It uses if-then statements to show the chances of relationships between items in one data record [10]. Once the dominant factors are identified by applying multiple regression, association rules can mine the underlying conditions for these factors that are important for modeling solar energy generation.

Fuzzy logic – It conceptualizes the ambiguity and fuzziness in the weather data into proper quantifiable boundaries. Therefore, a fuzzy logic model can be implemented for efficient renewable energy planning to reach realistic solutions. It deals with the concept of truth, and values that lie between absolute truth and absolute false [11]. It's a way of explaining and expressing uncertainty [12]. Weather has high degrees of truth. Therefore, it forms the perfect basis for application of fuzzy logic. The rules discovered in applying association rule mining technique are used to build an accurate and effective fuzzy logic model further to predict the effect of these variables on solar energy production.

III. An Illutrative Example

In this section, we present an illustrative example to demonstrate the modeling process using our integrated approach. The process consists of four stages: (a) Data pre-processing, (b) factor extraction, (c) rule identification, (d) fuzzy model construction.

A. Data pre-processing

Historical data for solar energy generation and various weather conditions has been taken from an Indian solar plant. It contains over 4,000 data records for different days and covers a range of different weather conditions. This dataset has been refined of the extreme or inaccurate values. The columns thought to be irrelevant (i.e. including mostly NIL or 0 values) or repetitive (high/mid/low cloud cover) have been removed to make the model simpler.

B. Factor extraction

To reduce the number of factors, multiple regression is used to identify the key independent variables. For the purpose of association rule mining, the variables have been

TABLE I. List of Variables Used in Multiple Regression

Notation	Variable
a	Temperature 2m above ground
b	Relative humidity 2m above ground
c	Mean sea level pressure
d	Total cloud cover
e	Angle of incidence
f	Wind speed 10m above ground
g	Wind direction 10 m above ground
h	Wind speed
i	Wind direction
j	Wind gust 10m above ground
k	Generated solar energy power **(Dependent variable)**

discretized into 3 crisp categories, namely, low, med, high. We normalize the variables using (1):

$$x' = \frac{x - average(x)}{\max(x) - \min(x)} \quad (1)$$

Regression helps discover the variables that most affect the output, i.e. generated solar energy power. Table 2 depicts the regression result. $P < 0.001$ is considered to be statistically highly significant. From Table 2, the variables with sig. <0.001 are (a) **temperature**, (c) **mean sea level pressure**, (d) **cloud cover**, and (e) **angle of incidence**. Therefore, these four variables have the most impact on the power generated. Data of these four variables will be extracted for rule identification in the next stage.

C. Rule identification

The four essential input variables extracted in the previous stage are used to generate association rules important for different levels of power generation. In association rule mining, support and confidence are the main rule evaluation metrics. A minimum support threshold and confidence threshold are to be defined so that every association rule generated fulfills the support and confidence criteria. A rule generated by association rule mining has the If-Then form, i.e. 'LHS (left hand side) => RHS (right hand side)', where LHS (antecedent) and RHS (consequent) are the disjoint sets of items. This rule expresses that the consequent set is likely to occur whenever the antecedent set occurs. For a rule, $X \geq Y$,

$$Support\ (X \leq Y) = P(X \cap Y) \quad (2)$$

$$Confidence\ (X \leq Y) = P(X|Y) = \frac{P(X \cap Y)}{P(X)} \quad (3)$$

Support is a fraction of data records that contain both X and Y, which represents the probability of the presence of both X and Y within the whole dataset, whereas confidence measures how often items in Y appear in data records that contain X, which in other words reflect how sure we can be that a result with a certain combination of items will occur under a particular condition [13]. In this modeling example, taking the characteristics of the dataset into consideration, the minimum support and confidence threshold are set to be 0.005 and 0.65. To mine association rules from the dataset,

TABLE II. INFLUENCE OF VARIABLES ON SOLAR ENERGY GENERATION

	Unstandardized B	Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
(constant)	-7802.350	1849.186		-4.219	<0.001
a	11.583	2.106	0.109	5.500	<0.001
b	2.037	0.739	0.051	2.757	0.006
c	10.088	1.775	0.076	5.684	<0.001
d	-6.264	0.258	-0.286	-24.26	<0.001
e	-23.202	0.411	-0.659	-56.45	<0.001
f	5.550	9.176	0.058	0.605	0.545
g	-0.158	0.248	-0.018	-0.638	0.523
h	-15.187	8.937	-0.160	-1.699	0.089
i	-0.043	0.247	-0.005	-0.174	0.862
j	-1.725	1.803	-0.023	-0.957	0.339

amongst the four input variables we have chosen, we observe their data characteristics, including their minimum, maximum and average values, and subsequently assign three data categories for each input and the output. For example, "temperature" is categorized into "cold", "normal" and "hot", whereas the rest of the input variables, "mean sea level pressure", "cloud cover", and "angle of incidence", as well as the output, "generated solar energy power", each of them is categorized into "low", "medium" and "high" with specific and specified ranges. We then use Apriori algorithm, the first and most influential algorithm for efficient association rule discovery [14], to obtain association rules that generates "high", "medium" and "low" solar energy power. By so doing, we understand at what circumstances will a "high", "medium" and "low" levels of solar energy power be generated. A total of 15 rules are generated, as shown in Table 3.

D. Fuzzy model construction

After mining the important rules, a model based on fuzzy logic is build. Wider ranges were used in the parameters for membership function to ensure that when the data changes, the values still fall within the categories. Membership type "trimf" and "trapmf" are used for the simplicity that they offer. In fuzzy logic model construction, three crucial steps are fuzzification, fuzzy inference process and defuzzification.

In fuzzification phase, we make use of the three data categories of each variable (such as "temperature" = high, medium, low) assigned in the previous rule identification stage. Data of the four variables and one output variable are studied once again to define the fuzzy set ranges of each membership functions. The membership function configuration of the output variable, i.e. Solar energy power generation, is shown in Fig. 2 as an example of undergoing a fuzzification procedure in fuzzy modeling.

With the types and ranges of the membership functions for each variable defined, a fuzzy inference engine, which consists of four inputs and one output, as depicted in Fig. 3, is constructed. As shown in Fig. 3, a rule block in the fuzzy inference engine is one that processes the values of each input to obtain a corresponding output values. The set of rules as listed in Table 3 are stored in the rule block to govern the

TABLE III. 15 RULES MINED FROM DATASET

	RHS is "Generated solar energy power = high", when:			
Rule	***LHS***	***S***	***C***	***Lift***
1	a = high, e = low	0.038	0.661	3.858
2	a = high, c = low, e = low	0.023	0.688	4.014
3	a = high, d = med, e = low	0.038	0.661	3.858
4	a = high, c = low, d = med, e = low	0.023	0.688	4.014
	RHS is "Generated solar energy power = medium", when:			
5	c = low, e = high	0.007	0.969	1.541
6	a = high, e = high	0.005	0.917	1.458
7	c = low, d = med, e = high	0.007	0.969	1.541
8	a = high, d = med, e = high	0.005	0.917	1.458
	RHS is "Generated solar energy power = low", when:			
9	a = low, c = high, e = med	0.006	1	4.998
10	a = low, c = high, e = low	0.005	1	4.998
11	a = low, c = high, d = med, e = med	0.006	1	4.998
12	c = high, d = med, e = low	0.022	0.939	4.692
13	c = high, e = low	0.023	0.932	4.658
14	a = med, c = high, d = med, e = low	0.018	0.925	4.623
15	a = med, c = high, e = low	0.018	0.914	4.566

S = Support; C = Confidence; Lift is a performance measure specifically for association rule; all values in 3 decimal places.

processing of the output, that is, the predicted solar energy power that will be generated.

With this integrated model, organizations who gather weather information can input the current temperature, mean sea level pressure, cloud cover, and angle of incidence in exact values to obtain a predicted solar energy power. For example, as shown in Fig. 4, when the temperature is 7.29 °C, sea level pressure is 1070 millibars, cloud-cover is 50% and angle of incidence is 28°, the predicted solar energy power to be generated is 694 kw. The input values can also be forecasted values so that our model can predict the solar energy power generation based on the forecasted input values.

IV. IMPLICATIONS AND DISCUSSIONS

The integrated model with an illustrative example demonstrate the potentials of utilizing multiple data analytical tools for decision-making and prescriptive analytics. The research and practical implications are discussed below.

A. Research implications

Regression is simple yet an effective mean of identifying the most important weather factors and how they impact the efficiency of generating solar energy. Once this has been identified, association rules mined and fuzzy model built, it can give an apt indication of power generation for the next few hours, depending on the data. In this application, for the association rules, crisp discretization was used to categorize numerical data. Fuzzification reduces the influence of endpoints in a particular category and distributes it over to another class. This is a better way of mining association rules using fuzzification rather than crisp discretized classes [15].

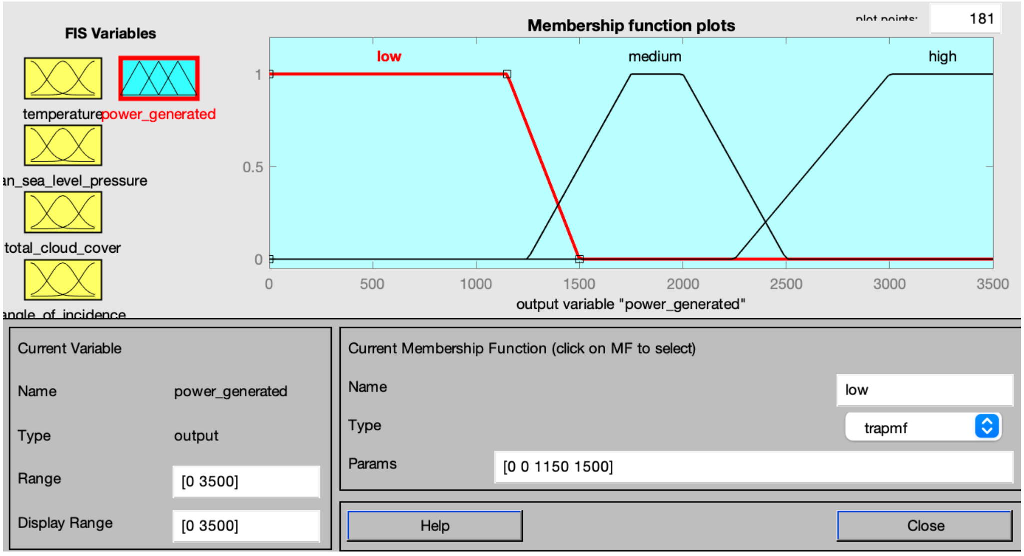

Fig. 2. Fuzzy inference engine for predicting solar energy power

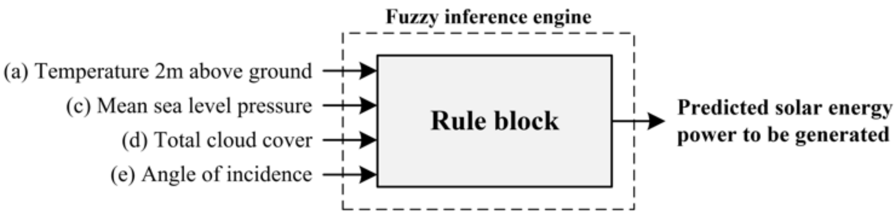

Fig. 3. Fuzzy inference engine for predicting solar energy power

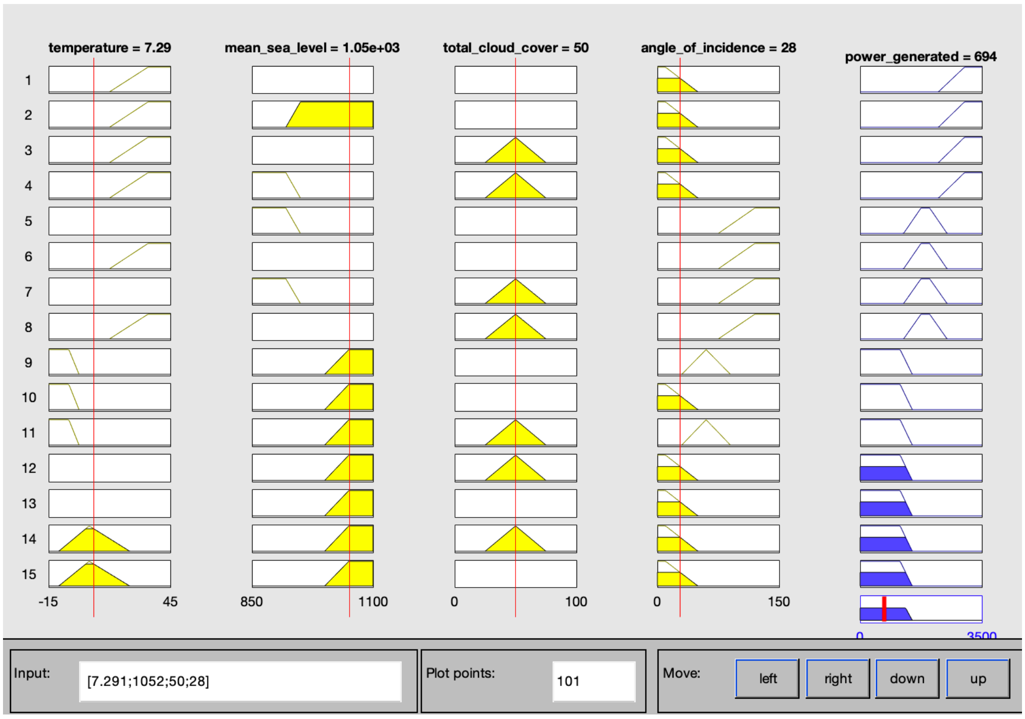

Fig. 4. An example of generating solar energy prediction based on the integrated model

This can help mine better and more useful association rules. In fact, fuzzy logic and association rule mining has been used in various disciplines. For example, Ho et al. [16] use fuzzy association rule mining approach to identify financial data association. Leung et al. [17] apply fuzzy association rule mining for managing a quotation process in supply chains. Leung and Mo [18] developed a fuzzy-AHP approach for evaluation and selection of digital marketing tools. In summary, this approach shows potentials to facilitate companies or the government in understanding the necessary conditions for generating different types of energy. Researchers are suggested to consider integrating this tool to make use of the available big data to generate insights or inform decision-making.

B. Practical implications

Solar power is one of the fastest growing sources of renewable energy. The investments in solar plants are astronomical. This deters the companies from foraying into this business. Predicting solar power output is pivotal for grid operators to optimize their functioning. Accurate forecasting of the power generation in the short run is of immense importance for delivery, storage, efficient power management of the grid, and for decision making for auctioning the energy. Having a good understanding of the output can help organizations in the industry optimize the pricing and maximize the returns on their investment, and hence increase the participation of companies in this field. An extended version and combined energy systems can be build, having vast implications for planning energy systems, auctioning in energy markets, etc.

C. Future opportunities

We discuss three major areas of extending our work, they include: Wider applications of solar energy efficiency management, integration of IoT devices for data capturing, and Extension of fuzzy logic modeling.

- *Wider applications of solar energy efficiency management*

The inspiration and methods proposed in this paper can further be extended for scouting locations for solar panel installations and building solar plants, be it specific narrow locations within the city, or as wide as the whole city itself. Gunderson et al. [19] suggest a model for locating potential sites in the Black Sea region. This can be combined with IoT sensors installed at specific localities and locations to understand the patterns and collect data for that particular area and compare with other locations. Fuzzy logic has already been heavily used in non-renewable electricity pricing. Arciniegas and Rueda [20] use fuzzy logic to predict the prices of electricity in Ontario, based on weather (to understand demand) and power generation variables such as outages, excess-capacity, excess-energy, temperature etc. All of these instances, pricing, and optimal locations can help maximize the returns of the investment.

- *Integration of IoT devices for data capturing*

This study only discusses the environmental factors affecting the solar energy generation. We suggest that our integrated approach can be combined with IoT devices on the panels to understand the device optimality for electricity generation at a particular location, as proposed by Wedashwara et al. [21].

- *Extension of fuzzy logic modeling*

Fuzzy logic can be used in conjunction for demand and supply fluctuations, predict the same for future. These predictions can be combined with some other variables, and finally fuzzy logic can be applied on these predictions to create a dynamic model for pricing solar energy. The data can be used to understand co-relation between different environmental factors and predict weather patterns hours and, in the future, days in advance to make the model even more dynamic. Fuzzy logic can be used to build entire interconnected energy systems and help optimize the renewable energy sources.

There is a growing need for more flexible energy and grid-systems. In 2021, in Texas, a major power-cut, that happened due to rapid fall in temperatures wreaked havoc on the gas, electricity and water network. This was caused to due drastic changes in weather conditions, and maintenance scheduling. This rendered loss of 45GW capacity of electricity generation [22]. This is possible when using just one source of renewable energy. Proactive co-ordination and management of different energy sources, with weather conditions, can help build an efficient and effective energy ecosystem. When one source is predicted (using fuzzy) to have lesser generation, another source can be used to better cater to the energy needs, depending on the environmental conditions. In this way, fuzzy logic can be extended to other renewable energy sources. However, these are complex models, which require immense experimentation and calculations to build dynamic, accurate models

V. Concluding Remarks

Big data and data analytics are emerging fields that have drawn much attention from many research disciplines. In the past decade, researchers have been paying great efforts in utilizing data analytical tools to extract essential information from datasets. This paper makes use of multiple regression to identify the key weather-related factors affecting the efficiency of solar energy power generation. Four factors, namely temperature, mean sea level pressure, cloud cover, and angle of incidence, are found to be the statistically significant independent variables affecting solar energy power generation. Association rule mining is then used to extract hidden association rules among these four variables. Using this method, rules are in "If-Then" structure, denoting the antecedent (the "If" part, a.k.a. LHS) item sets and the consequent (the "Then" part, a.k.a. RHS). By restricting the consequent part to only include "Generated solar energy power = low, medium or high", we identify 15 rules from the four variables using Apriori algorithm. These set of rules are then inputted into a fuzzy inference engine for the construction of a fuzzy model that is capable of predicting the next period, short-term solar energy power generation. Only by inputting the current or forecasted temperature, mean sea level pressure, cloud cover, and angle of incidence, organizations can predict the solar energy power generation. This is the first known study that develops an integrated approach, utilising multiple regression, association rule mining and fuzzy logic to build a prediction model with simplicity.

References

[1] L. Berrang-Ford, J. D. Ford, and J. Paterson, "Are we adapting to climate change?" Global environmental change, vol. 21(1), pp. 25-33, 2011.

[2] V. Weilnhammer, J. Schmid, I. Mittermeier, F. Schreiber, L. Jiang, V. Pastuhovic, ..., and S. Heinze, "Extreme weather events in europe and their health consequences–A systematic review" International Journal of Hygiene and Environmental Health, vol. 233, 113688, 2021.

[3] B. K. Bose, “Global warming: Energy, environmental pollution, and the impact of power electronics,” IEEE Industrial Electronics Magazine, vol. 4, no. 1, pp. 6-17, 2010.

[4] S. Asiaban, N. Kayedpour, A. E. Samani, D. Bozalakov, J. D. De Kooning, G. Crevecoeur, and L. Vandevelde, “Wind and Solar Intermittency and the Associated Integration Challenges: A Comprehensive Review Including the Status in the Belgian Power System,” Energies, vol. 14, no. 9, pp. 2630, 2021.

[5] T. Khatib, A. Mohamed, K. Sopian, and M. Mahmoud, “Solar energy prediction for Malaysia using artificial neural networks,” International Journal of Photoenergy, 2012.

[6] P. Szuromi, B. Jasny, D. Clery, J. Austin, and B. Hanson, “Energy for the long haul,” Science, vol. 315, no. 5813, pp. 781-781, 2007.

[7] T. Khatib, A. Mohamed, and K. Sopian, “A review of solar energy modeling techniques,” Renewable and Sustainable Energy Reviews, vol. 16, no. 5, pp. 2864-2869, 2012.

[8] C. Li, C. McLinden, V. Fioletov, N. Krotkov, S. Carn, J. Joiner, ..., and R. R. Dickerson, “India is overtaking China as the world’s largest emitter of anthropogenic sulfur dioxide,” Scientific reports, vol. 7, no. 1, pp. 1-7, 2017.

[9] C. H. Mason, and W. D. Perreault Jr, “Collinearity, power, and interpretation of multiple regression analysis,” Journal of marketing research, vol. 28, no. 3, pp. 268-280, 1991.

[10] J. Hipp, U. Güntzer, and G. Nakhaeizadeh, “Algorithms for association rule mining—a general survey and comparison,” ACM sigkdd explorations newsletter, vol. 2, no. 1, pp. 58-64, 2000.

[11] K. H. Leung, H. C. Lau, D. Nakandala, X. T. Kong, and G. T. S. Ho, “Standardising fresh produce selection and grading process for improving quality assurance in perishable food supply chains: an integrated Fuzzy AHP-TOPSIS framework,” Enterprise Information Systems, vol. 15, no. 5, pp. 651-675, 2021.

[12] L. Suganthi, S. Iniyan, and A. A. Samuel, A. A. “Applications of fuzzy logic in renewable energy systems–a review,” Renewable and sustainable energy reviews, vol. 48, pp. 585-607, 2015.

[13] C. H. Lee, Y. H., Kim, and P. K. Rhee, “Web personalization expert with combining collaborative filtering and association rule mining technique,” Expert Systems with Applications, vol. 21, no. 3, pp. 131-137, 2001.

[14] M. Hegland, “The apriori algorithm–a tutorial,” Mathematics and computation in imaging science and information processing, pp. 209-262, 2007

[15] A. Kaushik, D. K. Tayal, and K. Yadav, “A fuzzy approach for cost and time optimization in agile software development,” In Advanced Computing and Intelligent Engineering, pp. 629-639. Springer, Singapore, 2020.

[16] G. T. S. Ho, W. H. Ip, C. H. Wu, and Y. K. Tse, “Using a fuzzy association rule mining approach to identify the financial data association,” Expert Systems with Applications, vol. 39, no. 10, pp. 9054-9063, 2012.

[17] K. H. Leung, C. C. Luk, K. L. Choy, H. Y. Lam, and C. K. M. Lee, “A B2B flexible pricing decision support system for managing the request for quotation process under e-commerce business environment,” International Journal of Production Research, vol. 57, no. 20, pp. 6528-6551, 2019.

[18] K. H. Leung, and Daniel Y. Mo, “A fuzzy-AHP approach for strategic evaluation and selection of digital marketing tools,” 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE, pp. 1422-1426, 2019.

[19] I. Gunderson, S. Goyette, A. Gago-Silva, L. Quiquerez, and A. Lehmann, “Climate and land-use change impacts on potential solar photovoltaic power generation in the Black Sea region,” Environmental Science & Policy, vol. 46, pp. 70-81, 2015.

[20] A. I. Arciniegas, and I. E. A. Rueda, “Forecasting short-term power prices in the Ontario Electricity Market (OEM) with a fuzzy logic based inference system,” Utilities Policy, vol. 16, no. 1, pp. 39-48, 2008.

[21] W. Wedashwara, I. W. A. Arimbawa, A. H., Jatmika, A. Zubaidi, and T. Mulyana, “IoT based Smart Small Scale Solar Energy Planning using Evolutionary Fuzzy Association Rule Mining,” In *2020 International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, pp. 1-6. IEEE, 2020.

[22] Casey, J. (2022). Big data and modelling data: Encoord on data in energy. Retrieved 19 April 2022, from https://www.power-technology.com/analysis/encoord-big-data-processing-renewable-energy/