Tackling Energy Poverty with Artificial Intelligence: Challenges and Opportunities

Emergent Research Forum (ERF) Paper

Bastin Tony Roy Savarimuthu
University of Otago, New Zealand
Tony.Savarimuthu@otago.ac.nz

Jacqueline Corbett
Université Laval, Canada
Jacqueline.Corbett@fsa.ulaval.ca

Abstract

Energy poverty occurs when households do not have access to, or cannot afford the energy services necessary to support their daily needs, including heating/cooling, washing, cooking, lighting and other activities. Energy poverty is a hidden, but important social problem because living in conditions without adequate energy access can lead to other health and social problems. Addressing energy poverty in a meaningful way requires first detecting households in precarious situations and then putting in place appropriate strategies to support them. In this research, we explore the potential for using artificial intelligence (AI) to model energy patterns that reflect situations of poverty. Using simulated energy consumption data in the New Zealand context, we show that AI, specifically machine learning models can achieve high predictive accuracy. We discuss challenges associated with using finer-grained approaches and opportunities for better prediction, prevention, and remediation of energy poverty.

Keywords

Artificial intelligence, energy poverty, justice, machine learning, sustainability.

Introduction

The United Nations' Sustainable Development Goals call for universal access to affordable, reliable and modern energy services (United Nations n.d.). This objective may seem trivial in developed countries, however, an estimated 4.5% to 25% of New Zealanders face energy hardship (Stats NZ 2017). Energy hardship occurs when a household cannot meet its energy needs, either because it does not have access to a reliable energy supply or cannot afford the necessary energy services. The extreme case of hardship, called energy poverty, manifests in developed countries when households are unable to pay their energy bills (Dalla Longa et al. 2021), which can result in disconnections. Energy poverty is a complex social problem with cascading effects from individuals to society. Seniors, single-parent families, and those living with long-term illness or disability are at higher risk of energy poverty (Riva et al. 2021). Energy poverty is associated with significantly increased likelihood of poor health (Riva et al. 2021) and negatively impacts educational outcomes in developing countries (Banerjee et al. 2021).

Tackling energy poverty involves two main steps. The first is the detection of households facing energy poverty. Energy poverty is a relatively invisible problem and, while related to financial insecurity, arises from diverse factors within and beyond the households' control (e.g., building characteristics, weather, energy prices). The second step is the development and delivery of appropriate and effective interventions. To date, these efforts have used broad income or cost measures; these are blunt instruments that may fail to reach those in need while benefiting those who are not (Romero et al. 2018). Artificial intelligence (AI) offers great potential for addressing energy poverty by providing insights and tools for more effective detection and intervention (Corbett and Savarimuthu Forthcoming). Despite a growing corpus of energy-related research in information systems (IS) (Zhang et al. 2023), research at the intersection of AI and energy poverty is limited (Lopez-Vargas et al. 2022). To address these gaps, we have developed a multi-method research program to investigate the use of AI for tackling energy poverty. In the first study (reported here), we focus on detection and aim to answer two questions: RQ1) *can machine learning models be developed to accurately predict energy poverty,* and RQ2) *can machine learning be used to identify different clusters of energy poor households with similar energy consumption patterns*.

Background

Energy poverty is a global problem affecting both developing (Abbas et al. 2022; Zhang et al. 2023) and developed nations (Kez et al. 2024). Estimations of energy poverty have moved from simple, economic-based measures to a range of multidimensional energy poverty indexes (MEPI) that focus on physical access to electricity, affordability, and access to equipment, cost, and efficiency (Wang et al. 2021). Energy poverty detection faces two related challenges: lack of data and appropriate metrics (Lopez-Vargas et al. 2022). Data often comes from self-reported survey data that is subjective, time-consuming to collect and difficult to obtain (Kez et al. 2024). On the supply side, measures for energy poverty are often restricted to electricity connection and omit the other aspects of energy poverty, such as affordability, reliability, safety, and availability (Wang et al. 2021). With increasing digitalization, traditional datasets are being supplemented with environmental data gathered through sensors (e.g., outside temperature) (Riva et al. 2021) and satellite imagery (Kez et al. 2024). Despite the diversity of data sources, current approaches to measuring data rely on coarse-grained data (e.g., household income, energy costs, dwelling size), rather than considering fine-grained energy consumption data (e.g., how much energy is used for different activities). The use of fine-grained data could provide deeper and more nuanced understanding of energy poverty. For example, it could help identify households that forego space heating by prioritizing cooking and water heating. It could also provide more accurate profiles of different types of energy poor households with the goal of designing highly targeted interventions.

IS represent a promising tool for alleviating energy poverty (Lopez-Vargas et al. 2022), especially when such systems leverage AI. For instance, Corbett & Savarimuthu (2024) propose an eWISe system, an integrated platform that collects, analyzes and diffuses relevant information to address energy poverty at multiple levels: household, community, and regional/national. Operationalizing such a system requires developing machine learning (ML) models that can detect energy poverty. Indeed, energy poverty prediction is one of the most popular topics in this area (Lopez-Vargas et al. 2022). ML offers the advantages of being able to find patterns in large data sets, while explainable AI can provide insights regarding the relationships between inputs and outputs of models (van Hove et al. 2022).

ML models have been used to study energy poverty at multiple levels as mentioned above. Dalla Longa et al. (2021) use ML to categorize energy poverty risk at the national level in the Netherlands. They have identified four categories: those who are not at risk, those who face an expenditure risk, those with an income risk, those that have a double risk relating to both expenditures and income. This framework is also used by van Hove et al. (2022). At the regional level, Wang et al. (2021) explored the potential of satellite remote sensing data in energy poverty prediction combined with socioeconomic survey data, achieving 90 % accuracy for prediction of energy poverty districts. Kez et al. (2024) combined nighttime lighting remote sensing satellite information combined with socio-economic data to identify areas of energy poverty at the community level. Spandagos et al. (2023) applied ML to predict energy poor individual households from a large and heterogeneous pool of data from 28 European countries. However, in the research work reported above, the ML models developed make use of coarse-grained data, while our current work considers fine-grained energy consumption data in a household (i.e., different energy usage streams such as space heating, lighting and cooking).

Methods

This study follows a Design Science Research approach and responds to two specific research objectives (Baskerville et al. 2018): i) the development of design artifacts, specifically ML models to detect energy poverty at the household level and cluster households based on similar energy consumption patterns; and ii) demonstrating the utility of the developed artifacts. We applied the Design Science Research Methodology (DSRM) (Peffers et al. 2007) using three steps. First, we defined the objectives of the artifacts as stated above. In addition, we developed an approach to simulate energy poverty data based on real-world data. Second, we developed ML models using appropriate features of households. Third, we evaluated the goodness of the models.

Simulated Data Creation. Acquiring quality data for data poverty detection can be difficult. Thus, following other research (e.g., Bienvenido-Huertas et al. 2023), we created a simulated data set for training and validating our ML models. We chose two-person senior citizen households as our target population due to their vulnerability to energy poverty (Serrano et al. 2023). We simulated the data across four energy consumption streams: space heating, water heating, lighting, and cooking (e.g., range usage). New Zealand households utilize on average 34% of the total energy consumption for water heating, 12% for space heating, 12% for lighting and 7% for cooking (Isaacs et al. 2006). Based on prior work (CSIRO 2018), we assumed that a two-person household uses 12.5 Kwh of energy per day on average. We modeled four categories of households: i) *energy poor* who consume much (i.e., 50% to 90%) less than 12.5 Kwh of energy per day; ii) *energy hardship* who consume somewhat (i.e., 0% to 49%) less than average; iii) *energy adequate* who use enough power to cover their energy needs, using 1% to 25% more than the average; and iv) *energy surplus* who use 26% to 100% more energy than average. For each category, the range of energy consumption deficit or surplus was fixed. Then, a specific value was drawn from this range for each of the four energy usage streams, which was used as a multiplier to assign a specific energy usage value for a given energy usage stream. For example, if the energy specific value is 0.35 for space heating, out of 12% of total power that is used for space heating, the average space heating for a day was computed as: 0.65 * 0.12 * 12.5 = 0.975 Kwh. Note that 0.65 represents the actual usage (i.e., 1 – 0.35 = 0.65). This way, we modeled the variability of energy usage among households across the four energy consumption streams. Once the base loads for the different energy consumption streams were set, seasonal adjustments were added. In New Zealand, there is a substantial increase (> 200% increase) in space heating during the winter as compared to the summer, and also energy consumption for lighting, water heating and cooking range increases in the winter, by 50% (Isaacs et al. 2006). These values were used for data scaling.

The energy consumption data was simulated on a monthly basis for each of the 100 households for one year for the four energy consumption streams. Thus, the simulated dataset contains 1200 rows and five columns: one for the household category and four for the energy consumption streams (independent variables). Then, we added a sixth column for the dependent variable called *energy poverty* and assigned the value 1 for all households in the *energy poor* category and 0 otherwise.

Development of Machine Learning (ML) models. We developed two sets of ML models. First, the goal of the *classification* models was to predict whether a particular household is energy poor. We developed three models using the programming language R: Logistic Regression (LR), Support Vector Machines (SVM), and Random Forest (RF) (James et al. 2013). These models were selected based on prior works in energy poverty classification (Kez et al. 2024; Spandagos et al. 2023). In developing the models, we split the data into a training dataset (80%) and a test dataset (20%). For each model, we repeated the model building activity 10 times, called experiments, using different combinations of training data and then testing the model using the test dataset. In other words, each experiment was conducted with a different 80% of the data and evaluated on the remaining 20%. We then computed the average of these 10 experiments and report these results.

Second, the goal of the *clustering* model was to determine whether distinct patterns of energy consumption could be identified, by grouping energy poor households by similarities in the four energy consumption streams. We employed K-means clustering algorithm for this purpose (James et al. 2013) and Principal Component Analysis (PCA) (Labrín and Urdinez 2020) to visualize and present the clusters. The clusters were made up of two dimensions of PCA derived from the four energy consumption streams.

Preliminary Results

In response to RQ1, our results show that the three models developed (LR, SVM and RF) were able to detect energy poor households with high accuracy (97.96% 98.00% and 98.46% respectively). The average Mathew's Correlation Coefficient scores, the commonly used metric for imbalanced datasets, for the three models were: 0.947, 0.951 and 0.960 respectively. These results demonstrate that ML models can be useful in detecting energy poverty patterns based on fine-grained household-level, energy consumption data. Our results are also consistent with prior energy poverty studies where RF have yielded superior results (Kez et al. ; Spandagos et al. 2023).

Figures 1 and 2 answer RQ2. Figure 1 shows two clusters: households experiencing energy poverty in red, and energy sufficient households in blue. Figure 2 shows sub-clusters within energy poverty households (i.e., sub-categories of the red cluster in Figure 1). Dimension 1, shown on the x-axis, is dominated by the lighting, water heating, and space heating streams (as shown by arrows), while dimension 2 on the y-axis is dominated by the cooking consumption stream. More concretely, the cluster in blue contains mostly positive values for dimension 1 and mostly negative values for dimension 2. This means the households in this cluster have values similar for lighting, water heating and space heating, but not for cooking. On the other hand, the cluster in yellow is based on (mostly) positive values for dimension 1 and positive values for dimension 2, meaning that these households had similar energy consumption for cooking. The grey cluster comprises households with the most dissimilarity on both dimensions. This analysis thus helps to identify clusters that have similar energy consumption behaviours across the four main areas of use.

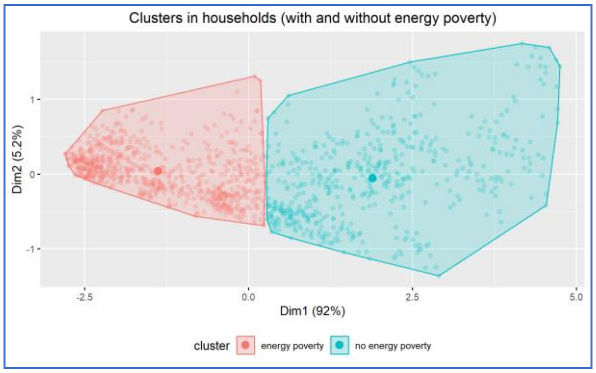

Figure 1. Clusters in households with and without energy poverty

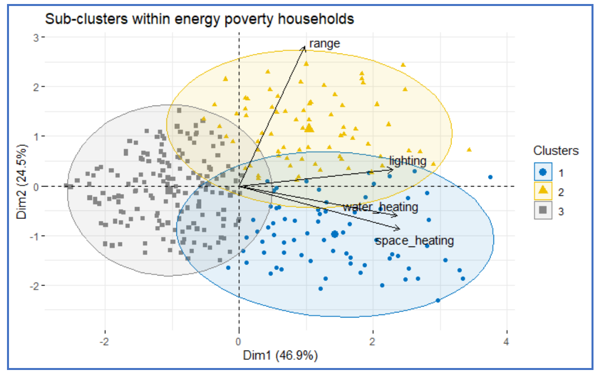

Figure 2. Sub-clusters within energy poverty households

Discussion and Conclusion

The overarching goal of this research program is to find ways of using AI to alleviate energy poverty. To do this, we must first be able to identify those facing energy poverty so that more effective interventions can be developed and implemented. This paper presents preliminary results related to the first part of this research program. There are a number of novel aspects and implications of this work.

First, a major challenge associated with energy poverty is the lack of household-level data. While energy providers may have total consumption information, sharing of such data may not be possible due to privacy concerns. Alternatively, sensors could be installed to capture real-time fine-grained household data. However, this is costly and intrusive. To overcome this challenge, we created a simulated dataset with realistic assumptions based on empirical studies and energy reports. Our research will continue to expand the size of the dataset by adding additional predictors of energy poverty (e.g., building characteristics, outdoor temperatures, household types and sizes, and seasonal prices) to the simulated dataset. *Second*, we use a design science research approach to develop and validate ML models to detect patterns of energy poverty. The results provide proof-of-concept, showing that it is possible to use ML with a high level of accuracy to detect energy poverty and cluster groups of households with similar consumption patterns at the community level. This provides preliminary empirical support for the concept of a multi-level, integrated energy well-being information system (Corbett and Savarimuthu Forthcoming). Future research could extend the context to the regional/national level to complement other work (e.g., Abbas et al. 2022). *Third*, this research extends the discourse on energy poverty in the IS literature. While other studies consider IS interventions in developing nations (Abbas et al. 2022; Zhang et al. 2023), this research is situated in a developed nation. As energy poverty threatens societies at all levels of development, being able to accurately predict energy poverty is a necessary step to achieve SDG 7 (Kez et al. 2024).

The next steps of the research program will involve refining and expanding the ML models for energy detection by considering more variables and validating the designed artifacts through focus groups involving domain experts. Then we will focus on developing intervention strategies. Subsequent research will investigate AI-based approaches to reduce or eliminate energy poverty. We intend to use agent-based simulation and modelling (Besse et al. 2019) to investigate the impacts of various interventions. For example, based on time-series data of energy usage across different energy consumption streams, we will define and assess the effectiveness of different mitigation strategies (e.g., suggesting alternative time-of-use for certain devices, joining microgrid-based energy coalitions, and applying for energy subsidies). These simulations will contribute to explanatory theory building as well as inform design and action.

REFERENCES

Abbas, K., Butt, K.M., Xu, D., Ali, M., Baz, K., Kharl, S.H., and Ahmed, M. 2022. "Measurements and Determinants of Extreme Multidimensional Energy Poverty Using Machine Learning," *Energy* (251), pp. 1-15.

Banerjee, R., Mishra, V., and Maruta, A.A. 2021. "Energy Poverty, Health and Education Outcomes: Evidence from the Developing World," *Energy economics* (101), p. 105447.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., and Rossi, M. 2018. "Design Science Research Contributions: Finding a Balance between Artifact and Theory," *Journal of the Association for Information Systems* (19:5), pp. 358-376.

Besse, J., Haki, M.K., Aier, S., and Winter, R. 2019. "Simulation-Based Research in Information Systems," *Business Information Systems Engineering* (61:4), pp. 503-521.

Bienvenido-Huertas, D., Sanchez-Garcia, D., Marin-Carcia, D., and Rubio-Bellido, C. 2023. "Analysing Energy Poverty in Warm Climate Zones in Spain through Artificial Intelligence," *Journal of Building Engineering* (68), pp. 1-19.

Corbett, J., and Savarimuthu, B.T.R. Forthcoming. "Energy Poor No More: Intelligent Approaches to Realizing Energy Well-Being," *IEEE Power & Energy Magazine*), pp. 1-13.

CSIRO. 2018. "Household Types and Energy Use." *CSIRO Energise Insight*, from https://www.studocu.com/en-gb/document/university-of-central-lancashire/development-and-change-in-the-asia-pacific-region/insight-28-household-types-and-energy-use/62346694

Dalla Longa, F., Sweerts, B., and van der Zwaan, B. 2021. "Exploring the Complex Origins of Energy Poverty in the Netherlands with Machine Learning," *Energy Policy*), pp. 1-12.

Isaacs, N., Camilleri, M., French, L., Pollard, A., Saville-Smith, K., Fraser, R., Rossouw, P., and Jowett, J. 2006. "Energy Use in New Zealand Households–Heep Year 10 Report," *Building Research Association of New Zealand*).

James, G., Witten, D., Hastie, T., and Tibshirani, R. 2013. *An Introduction to Statistical Learning*. Springer.

Kez, D.A., Foley, A., Abdul, Z.K., and Del Rio, D.F. 2024. "Energy Poverty Prediction in the United Kingdom: A Machine Learning Approach," *Energy Policy* (184), pp. 1-17.

Labrín, C., and Urdinez, F. 2020. "Principal Component Analysis," in *R for Political Data Science*. Chapman and Hall/CRC, pp. 375-393.

Lopez-Vargas, A., Ledezma-Espino, A., and Sanchis-de-Miguel, A. 2022. "Methods, Data Sources and Applications of the Artificial Intelligence in the Energy Poverty Context: A Review," *Energy & Buildings* (268), pp. 1-13.

Peffers, K., Tuunanen, T., Rothenberger, M.A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45-77.

Riva, M., Makasi, S.K., Dufresne, P., O'Sullivan, K., and Toth, M. 2021. "Energy Poverty in Canada: Prevalence, Social and Spatial Distribution, and Implications for Research and Policy," *Energy Research & Social Science* (81), pp. 1-12.

Romero, J.C., Linares, P., and López, X. 2018. "The Policy Implications of Energy Poverty Indicators," *Energy policy* (115), pp. 98-108.

Serrano, P.Y., Gómez, D.T., and Bieńkowska, Z. 2023. "Energy Vulnerability and Self-Imposed Austerity: An Ethnographic Approach to Adaptation Strategies to Extreme Heat among Older Adults in Madrid," *Energy Research & Social Science* (103), p. 103207.

Spandagos, C., Reanos, M.A.T., and Lynch, M.A. 2023. "Energy Poverty Prediction and Effective Targeting for Just Transitions with Machine Learning," *Energy Economics* (128), pp. 1-19.

Stats NZ. 2017. "Investigating Different Measures of Energy Hardship in New Zealand." from https://www.stats.govt.nz/reports/investigating-different-measures-of-energy-hardship-in-new-zealand

United Nations. n.d. "Goal 7: Targets and Indicators." from https://sdgs.un.org/goals/goal7#targets_and_indicators

van Hove, W., Dalla Longa, F., and van der Zwaan, B. 2022. "Identifying Predictors for Energy Poverty in Europe Using Machine Learning," *Energy & Buildings* (264), pp. 1-12.

Wang, H., Maruejols, L., and Yu, X. 2021. "Predicting Energy Poverty with Combinations of Remote-Sensing and Socioeconomic Survey Data in India: Evidence from Machine Learning," *Energy Economics*), pp. 1-13.

Zhang, X., Pan, S., Tim, Y., and Wu, L. 2023. "Leveraging Is-Based Energy Systems for Energy Poverty Alleviation in Zambia: An Interpretive Case Study," in: *Pacific Asia Conference on Information Systems*. pp. 1-10.