Fuel Poverty Prediction Using Socio-Economic Factors and Clustering Analysis

BY RAHIL DEJKAM[a] AND REINHARD MADLENER[a,b]*

Abstract

Our study predicts fuel poverty risk by grouping households based on data from a survey in England. The analysis reveals important differences between household groups, helping policymakers to better understand which factors contribute most to fuel poverty and suggesting targeted interventions to address the issue.

1. Introduction

Fuel poverty has garnered significant attention from both academics and policymakers in the EU (Castaño-Rosa et al. 2019). Despite numerous government-proposed solutions, such as the Winter Payment and Warm Home Discount, the fuel poverty rate continues to rise, with current solutions only reaching 10% of affected households (Charlier and Legendre 2021). Enhancing energy efficiency in housing requires substantial funding (Rzetelska and Combrinck 2022). Access to basic household energy services—heating, cooling, lighting, and such from appliances—is critical to welfare in the EU and UK. The EU Energy Poverty Observatory (EPOV) focuses on ensuring equitable energy access without imposing financial strain. Despite the UK being one of the world's leading economies, fuel poverty remains widespread due to socioeconomic factors, substandard housing, rising energy costs, and inefficient energy use (Boardman 2013). Vulnerability studies highlight that low-income households and disabled individuals are disproportionately affected by fuel poverty in the UK (Snell, Bevan, and Thomson 2015).

Current research attempts to address energy vulnerability by integrating social, political, and techno-economic perspectives [6]. However, these approaches often fail to account for unobserved heterogeneity within household characteristics and energy consumption patterns. Traditional regression-based models and spatial interpolation techniques lack the ability to capture the complex dynamics of fuel poverty (Abbas et al. 2020; Liu et al. 2021; Qurat-ul-Ann and Mirza 2021). Recent studies have utilized machine learning algorithms for more accurate predictions, but many still overlook important household features (Wong et al. 2018we can use spatial interpolation (SI; Robinson 2019; Puttanapong et al. 2022).

Unlike previous methodologies, our study introduces a novel cluster-based method that groups households based on socioeconomic and energy-related characteristics, allowing for a more nuanced analysis and targeted policy interventions (Dejkam and Madlener 2023). Using England as a case study provides an opportunity to apply this cluster-based method, given its large and diverse population that reflects many of the fuel poverty challenges seen across the UK. By focusing on England, where more comprehensive data is available, our approach captures more detailed patterns of fuel poverty, enabling policymakers to develop strategies to mitigate fuel poverty more effectively (Xu et al. 2021; Wang, Maruejols, and Yu 2021).

Our study addresses key gaps in the literature by offering a novel machine learning-based approach to fuel poverty prediction, helping to identify the most vulnerable households and the factors that contribute most to their energy struggles.

2. Methodology

This study employs a multi-step methodology to analyze fuel poverty in England using data from the English Housing Survey (EHS). Data was collected from April 2018 to March 2020, with April 2019 serving as the midpoint. The dataset includes 11,974 households and covers variables such as energy costs, household income, dwelling type, and heating characteristics. The methodology begins with data preprocessing, where missing values are removed, and categorical data that is converted to numerical form using one-hot encoding. The features for analysis were chosen based on a combination of literature review and a Pearson Correlation Coefficient analysis to remove irrelevant features. Households were grouped using a k-prototypes clustering algorithm, which combines both categorical and numerical data, making it ideal for mixed datasets. The optimal number of clusters was identified using the "elbow method", ensuring that households with similar characteristics were grouped together (see below). Microsoft Power BI was employed to visualize the clusters, helping to identify patterns within the data. In a next step, the fuel poverty risk within each cluster was predicted using a modeling algorithm. Finally, the contribution of each feature to the model's predictions was determined, providing insights into the factors that most influence fuel poverty.

3. Results

The study identified three distinct household groups in England that are most at risk of fuel poverty, using

a Institute for Future Energy Consumer Needs and Behavior (FCN), School of Business and Economics / E.ON Energy Research Center, RWTH Aachen University, Mathieustraße 10, 52074 Aachen, Germany
b Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway
* Corresponding author. Tel. +49 241 80 49 820, E-mail: RMadlener@eonerc.rwth-aachen.de (R. Madlener).

a method that helps to categorize homes based on shared characteristics. By applying the elbow method, which helps determine the optimal number of groups, it was found that splitting the households into three clusters offered the best balance between complexity and insight (cf. Figure 1). Figure 2 illustrates the three clusters, each representing different types of households and facing unique challenges when it comes to energy costs and affordability.

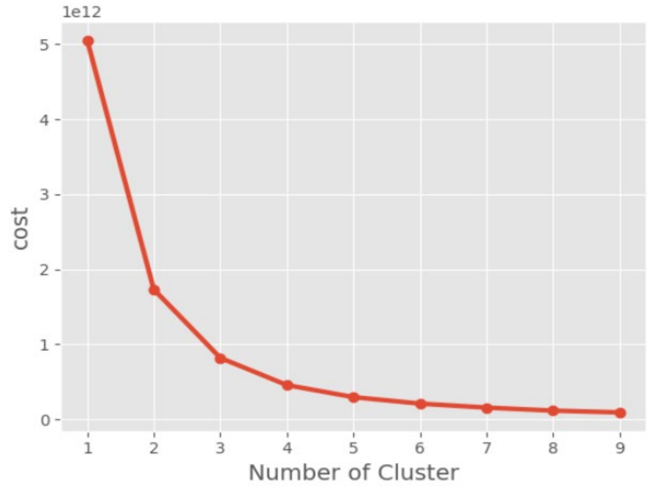

Figure 1: Optimal number of clusters determined by the elbow method

The first group (Cluster 0) consists of households with moderate energy costs, mostly living in detached homes with relatively good health and manageable expenses. The second group (Cluster 1) includes younger households that tend to have higher lighting and appliance costs, often due to more active household members and larger homes. The final group (Cluster 2) represents the most vulnerable households—older, low-income individuals struggling to meet their energy needs, especially for essential services like heating and lighting.

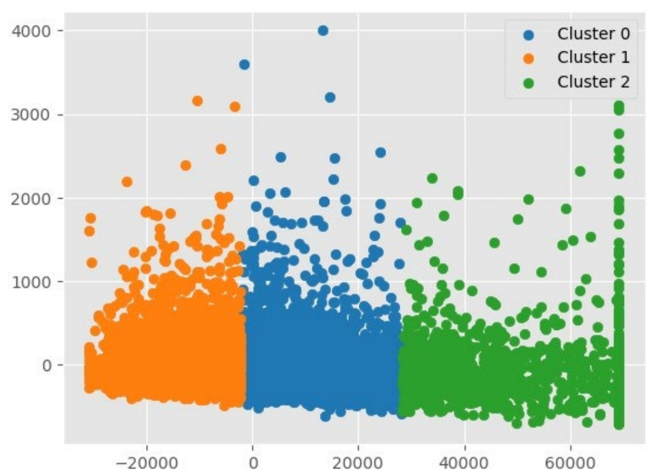

Figure 2: K-prototypes clustering of English households into three clusters

4. Discussion

This study shows the importance of tailored interventions for different household groups. Cluster 2, in particular, would benefit from direct financial aid and energy-saving measures, such as better insulation or energy-efficient appliances. In contrast, Cluster 1 would benefit from programs that help to reduce energy consumption for lighting and appliances, such as energy efficiency grants or appliance replacement programs.

Additionally, this study highlights the critical features contributing to fuel poverty, offering a clearer pathway for policymakers to design targeted interventions. Table 1 shows the key predictive features in each cluster; for instance, water heating costs and household income were significant predictors in Cluster 0, while age and lighting costs were more important in Cluster 2.

In conclusion, this study underscores the importance of tailored, data-driven interventions to effectively address fuel poverty. The combination of clustering analysis and machine learning provides a powerful tool for identifying at-risk households and guiding policymakers in designing targeted solutions. The insights gained from this research offer a clear path forward for combating fuel poverty, ensuring that the most vulnerable populations receive the support they need to improve their quality of life.

Table 1: Key predictive features in fuel poverty models

Cluster	Predictive Features
Cluster 0	Water Heating Cost, Floor Area, Income
Cluster 1	Lighting Costs, Household Composition
Cluster 2	Age, Income, Energy Costs

References

Abbas, Khizar, Shixiang Li, Deyi Xu, Khan Baz, and Aigerim Rakhmetova. 2020. “Do Socioeconomic Factors Determine Household Multidimensional Energy Poverty? Empirical Evidence from South Asia.” *Energy Policy* 146 (November):111754. https://doi.org/10.1016/j.enpol.2020.111754.

Boardman, Brenda. 2013. *Fixing Fuel Poverty: Challenges and Solutions* (1st ed.) Routledge. https://doi.org/10.4324/9781849774482.

Castaño-Rosa, Raúl, Jaime Solís-Guzmán, Carlos Rubio-Bellido, and Madelyn Marrero. 2019. “Towards a Multiple-Indicator Approach to Energy Poverty in the European Union: A Review.” *Energy and Buildings* 193 (June):36–48. https://doi.org/10.1016/j.enbuild.2019.03.039.

Charlier, Dorothée, and Bérangère Legendre. 2021. “Fuel Poverty in Industrialized Countries: Definition, Measures and Policy Implications a Review.” *Energy* 236 (December):121557. https://doi.org/10.1016/j.energy.2021.121557.

Dejkam, Rahil, and Reinhard Madlener. (2023). Algorithm-Enhanced Fuel Poverty Prediction Using Household Characteristics, Socio-Economic Factors, and Clustering Analysis, FCN Working Paper No. 12/2023, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, September.

Liu, Mengxiao, Shan Hu, Yong Ge, Gerard B.M. Heuvelink, Zhoupeng Ren, and Xiaoran Huang. 2021. “Using Multiple Linear Regression and Random Forests to Identify Spatial Poverty Determinants in Rural China.” *Spatial Statistics* 42 (April):100461. https://doi.org/10.1016/j.spasta.2020.100461.

Puttanapong, Nattapong, Arturo Martinez, Joseph Albert Nino Bulan, Mildred Addawe, Ron Lester Durante, and Marymell Martillan. 2022.

"Predicting Poverty Using Geospatial Data in Thailand." *ISPRS International Journal of Geo-Information* 11 (5): 293. https://doi.org/10.3390/ijgi11050293.

Qurat-ul-Ann, Abre-Rehmat, and Faisal Mehmood Mirza. 2021. "Determinants of Multidimensional Energy Poverty in Pakistan: A Household Level Analysis." *Environment, Development and Sustainability* 23 (8): 12366–410. https://doi.org/10.1007/s10668-020-01174-2.

Robinson, Caitlin. 2019. "Energy Poverty and Gender in England: A Spatial Perspective." *Geoforum* 104 (August):222–33. https://doi.org/10.1016/j.geoforum.2019.05.001.

Rzetelska, Dorota, and Madeleine Combrinck. 2022. "Fuel Poverty and Low Carbon Emissions: A Comparative Study of the Feasibility of the Hybrid Renewable Energy Systems Incorporating Combined Heat and Power Technology." *Frontiers in Energy* 16 (2): 336–56. https://doi.org/10.1007/s11708-021-0748-x.

Wang, Hanjie, Lucie Maruejols, and Xiaohua Yu. 2021. "Predicting Energy Poverty with Combinations of Remote-Sensing and Socioeconomic Survey Data in India: Evidence from Machine Learning." *Energy Economics* 102 (October):105510. https://doi.org/10.1016/j.eneco.2021.105510.

Wong, Kerry L. M., Oliver J. Brady, Oona M. R. Campbell, and Lenka Benova. 2018. "Comparison of Spatial Interpolation Methods to Create High-Resolution Poverty Maps for Low- and Middle-Income Countries." *Journal of The Royal Society Interface* 15 (147): 20180252. https://doi.org/10.1098/rsif.2018.0252.

Xu, Jianbin, Jie Song, Baochao Li, Dan Liu, and Xiaoshu Cao. 2021. "Combining Night Time Lights in Prediction of Poverty Incidence at the County Level." *Applied Geography* 135 (October):102552. https://doi.org/10.1016/j.apgeog.2021.102552.