

Budría, Santiago; Fermé, Eduardo; Freitas, Diogo Nuno

**Working Paper**

## Toward Proactive Policy Design: Identifying 'To-Be' Energy-Poor Households Using Shap for Early Intervention

IZA Discussion Papers, No. 17669

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Budría, Santiago; Fermé, Eduardo; Freitas, Diogo Nuno (2025) : Toward Proactive Policy Design: Identifying 'To-Be' Energy-Poor Households Using Shap for Early Intervention, IZA Discussion Papers, No. 17669, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/314566>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

DISCUSSION PAPER SERIES

IZA DP No. 17669

**Toward Proactive Policy Design:  
Identifying 'To-Be' Energy-Poor  
Households Using Shap for Early  
Intervention**

Santiago Budría  
Eduardo Fermé  
Diogo Nuno Freitas

FEBRUARY 2025

## DISCUSSION PAPER SERIES

IZA DP No. 17669

# **Toward Proactive Policy Design: Identifying 'To-Be' Energy-Poor Households Using Shap for Early Intervention**

**Santiago Budría**

*Universidad Antonio de Nebrija, CEEAplA and IZA*

**Eduardo Fermé**

*Universidade da Madeira and NOVA LINCS*

**Diogo Nuno Freitas**

*Universidade da Madeira, NOVA LINCS and Interactive Technologies Institute*

FEBRUARY 2025

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9  
53113 Bonn, Germany

Phone: +49-228-3894-0  
Email: [publications@iza.org](mailto:publications@iza.org)

[www.iza.org](http://www.iza.org)

## ABSTRACT

---

# Toward Proactive Policy Design: Identifying 'To-Be' Energy-Poor Households Using Shap for Early Intervention

Identifying at-risk populations is essential for designing effective energy poverty interventions. Using data from the HILDA Survey, a longitudinal dataset representative of the Australian population, and a multidimensional index of energy poverty, we develop a machine learning model combined with SHAP (SHapley Additive exPlanations) values to document the short- and long-term effects of individual and contextual factors—such as income, energy prices, and regional conditions—on future energy poverty outcomes. The findings emphasize the importance of policies focused on income stability and may be used to shift the policy focus from reactive measures, which address existing poverty, to preventive strategies that target households showing early signs of vulnerability.

**JEL Classification:** I32, D12, C53

**Keywords:** Energy poverty, panel data, explainable AI, time-series analysis, public policy, temporal dynamics, feature importance

**Corresponding author:**

Santiago Budría  
Department of Business Administration  
Universidad Nebrija  
C/ de Sta. Cruz de Marcenado, 27  
28015 Madrid  
Spain  
E-mail: sbudria@nebrija.es

## 1. Introduction

Energy poverty, a condition where households are unable to access or afford adequate, reliable, and clean energy services, has emerged as a critical global issue. Recent estimates suggest that 750 million people still lack access to electricity worldwide, and more than 2 billion people lack access to clean cooking fuels (International Energy Agency (2024)). In economic literature, energy poverty has garnered independent attention and is now studied as a distinct subject. This is because energy poverty is only moderately correlated with income poverty and yet negatively related to relevant economic outcomes, including human capital formation (Phoumin and Kimura (2019)), well-being (Nguyen-Phung and Le (2024)), and health (Pondie et al. (2024)).

In this context, the identification of populations at risk is essential for formulating effective policy interventions. While there has been significant research on the socioeconomic determinants of energy poverty (Fry et al. (2022); Awan et al. (2022); Koomson and Awaworyi Churchill (2022)), most studies look only at contemporaneous relationships, assuming that control variables fully represent the information influencing the observed outcome. This approach may overlook the potential role of long-memory processes and the enduring influence of contextual factors. If energy poverty is indeed a chronic state shaped by an individual’s history, such perspectives could provide an incomplete understanding. Evidence on the long-term effects of specific characteristics on energy poverty remains scarce, highlighting the need for further research in this area.

This paper takes a step forward by analyzing the short- and a long-term effects of individual variables and contextual factors—such as income, energy prices, regional conditions, and other socioeconomic variables—on future energy poverty outcomes. We use the 2007–2021 waves of the Household, Income and Labour Dynamics in Australia (HILDA) Survey<sup>1</sup>, a micropanel survey representative of the Australian population, which allows us to track people over up to 15 consecutive years. Since energy poverty is a multifaceted construct, we utilize a Multidimensional Energy Poverty Index (MEPI) incorporating five items that capture both objective (expenditure-based) and subjective (self-assessed) dimensions. We then use Machine Learning (ML) models to forecast the MEPI and employ SHapley Additive exPlanations (SHAP) (Lundberg and Lee (2017)) to interpret these predictions. SHAP

---

<sup>1</sup><https://melbourneinstitute.unimelb.edu.au/HILDA>.

quantifies the contribution of each variable, highlighting how specific factors at different points in time influence future energy poverty outcomes. This approach provides an understanding of the importance of each variable and the temporal dynamics shaping energy poverty trajectories.

This paper makes three significant contributions to the literature. First, the paper aims to advance the methodological toolkit for studying energy poverty. SHAP has been applied successfully to the study of financial time series data (Mokhtari et al. (2019)), short-term load forecasting (Lee et al. (2023)), and aviation’s predictive maintenance (Alomari and Andó (2024)). To the best of our knowledge, this is the first analysis to combine SHAP techniques with high-quality micropanel data to identify household-level drivers of energy poverty. By using SHAP, the paper highlights how its application in energy poverty research extends beyond the capabilities of traditional analytical methods, offering dynamic and actionable insights for policymakers.

Second, the paper contributes to the growing body of literature employing ML techniques to analyze the determinants of energy poverty. This approach is still limited but increasingly recognized for its potential to guide alleviation strategies<sup>2</sup> (Dalla Longa et al. (2021); van Hove et al. (2022); Spandagos et al. (2023); Gawusu et al. (2024)). However, much of the existing ML-based evidence relies on contemporaneous relationships between explanatory variables and energy poverty, largely due to the prevalence of cross-sectional or short-duration datasets. In contrast, our study leverages a 15-year longitudinal dataset to explore predictive dynamics. While the previous studies utilized Extreme Gradient Boosting (XGBoost),  $k$ -Nearest Neighbors ( $k$ -NN), Random Forest (RF), and Artificial Neural Networks (ANN), this paper focuses on capturing temporal dependencies. By employing SHAP, we quantify and disentangle the contribution of each feature to predictions, both overall and at specific points in time, offering insights into how past conditions influence future energy poverty.

Third, historically, public initiatives that address energy poverty—particularly in developed nations—have primarily focused on providing financial assistance and energy subsidies to individuals currently classified as (energy) poor. This approach operates on the assumption that immediate

---

<sup>2</sup>Unlike traditional regression methods, which require prior assumptions about potential correlations and their functional forms, ML techniques allow these relationships to naturally emerge during model training and excel at capturing complex, non-linear dependencies.

69 interventions can effectively alleviate energy poverty in the short term. How-  
70 ever, this focus overlooks households that are at risk of becoming energy-poor  
71 in the future, leaving a significant portion of the potentially vulnerable popu-  
72 lation unaddressed. Our approach challenges this perspective by emphasizing  
73 the importance of both the timing and the magnitude of key variables, such  
74 as income stability and energy prices, in shaping energy poverty trajectories.  
75 By identifying these “to-be energy-poor” households, our paper paves the  
76 way for more proactive policies that tackle the historical, individual-  
77 level causes of energy poverty, moving beyond temporary relief measures.

78 We consider Australia to be a compelling subject for our research. Es-  
79 calating energy costs have been a major concern over the last decade in  
80 Australia as electricity prices have almost tripled (Proctol (2022)). Forward  
81 electricity prices for 2023 delivery in Australia’s National Electric Market  
82 surged from approximately \$48 in 2021 to \$156/MWh in 2022 (the 52-week  
83 average), peaking around \$247/MWh in October 2022 (Simshauser (2023)).  
84 This substantial surge, relative to household income, has placed a heavier  
85 burden on household budgets and exacerbated issues related to energy ac-  
86 cess and affordability (OECD (2023)). Furthermore, despite its fragmented  
87 system of energy assistance—varied across states and territories—existing  
88 programs are largely focused on mitigating costs through price subsidies and  
89 welfare payments for energy bills (Willand (2022)). These measures heavily  
90 rely on means-testing and target low-income groups (Simshauser and Miller  
91 (2023)), often overlooking individuals who are not currently energy-poor but  
92 are at risk of becoming so. By identifying to-be energy-poor individuals be-  
93 fore they fall into vulnerability, our paper shifts the focus from reactive policy  
94 interventions to preventive, forward-looking interventions.

95 The paper shows that historical household income levels are pivotal in  
96 forecasting energy poverty outcomes, particularly over longer time horizons  
97 and in more severe cases. It also highlights the significant impact of income  
98 variations, independent of static income levels. This effect intensifies when  
99 transitioning from short-term to long-term poverty, suggesting that income  
100 volatility is particularly harmful in the long run. Additionally, energy prices  
101 have a moderate, non-linear effect in the short term but become less relevant  
102 for longer horizons. These findings emphasize the importance of policies  
103 focused on income stability and may be used to shift the policy focus from  
104 reactive measures, which address existing poverty, to preventive strategies  
105 that target households showing early signs of vulnerability.

106 The paper is structured as follows: Section 2 reviews the relevant liter-

107 ature on energy poverty, its determinants, and the application of machine  
108 learning methods in this context. Section 3 describes the data, key variables,  
109 and the construction of the MEPI. Section 4 outlines the methodological ap-  
110 proach, including model development and the use of SHAP for interpretabil-  
111 ity. Section 5 presents the results, highlighting the predictive performance of  
112 the models and the temporal dynamics of key variables. Section 6 discusses  
113 sensitivity analyses and robustness checks. Finally, Section 7 concludes with  
114 key findings, policy implications, and limitations of the study. The paper  
115 includes three appendices with technical details.

## 116 **2. Review of the literature**

117 Energy poverty can be defined as a household’s inability to afford or access  
118 energy services needed to support adequate living conditions and human  
119 development. While translating into practice conceptual definitions of energy  
120 poverty is typically a challenge and has been the object of extensive discussion  
121 in the literature (for an overview see Sy and Mokaddem (2022)) the focus has  
122 generally been put on the inability of households to afford and have access  
123 to adequate energy services.

124 The global interest in energy poverty stems from its far-reaching conse-  
125 quences, which are multifaceted. Research based on international macroe-  
126 conomic data shows that the prevalence of energy poverty negatively affects  
127 development, health outcomes, and average schooling levels (Banerjee et al.  
128 (2021)). Moreover, energy access and affordability are crucial dimensions of  
129 multidimensional poverty, and, as such, they can be negatively related to  
130 economic growth (Bao and Liao (2024)). Studies based on microeconomic  
131 panel data are consistent with this notion, showing that energy poverty sig-  
132 nificantly affects a number of personal-level outcomes, including subjective  
133 well-being (Lin and Okyere (2021)) and health (Zhang et al. (2021b); Pondie  
134 et al. (2024)). Energy poverty is also negatively related to children’s academic  
135 performance (Zhang et al. (2021a)) and human capital formation (Phoumin  
136 and Kimura (2019)).

137 Using international comparable data, research shows that country-level  
138 factors such as education, governance quality, technology advancements,  
139 economic development, and health expenditures are relevant determinants  
140 of household-level energy poverty depending on the country’s GDP (Boğa-  
141 Avram et al. (2024)). Moreover, income inequality and, to a lesser extent,  
142 climate conditions also play a role (Igawa and Managi (2022)). Furthermore,



143 the sources of electricity production also contribute to shaping energy poverty  
144 outcomes, reflecting the importance of a country’s energy mix (Kocak et al.  
145 (2023)). Additionally, high energy costs, accessibility, and the types of en-  
146 ergy sources further shape these outcomes (Primc et al. (2021)). Inefficient  
147 building structures, dwelling size, age, thermal insulation, floor area, and  
148 heating system can be significantly correlated with various forms of energy  
149 deprivation (Karpinska and Śmiech (2020)).

150 At the household level, income constraints, coupled with high energy  
151 prices, can culminate in the difficulty of paying bills, energy debt, and  
152 even the disconnection of energy supplies (Awan et al. (2022); Manasi and  
153 Mukhopadhyay (2024)). Educational attainment is inversely correlated with  
154 energy poverty, primarily due to energy-saving practices and an improved  
155 economic situation. Education enhances knowledge and the capacity to make  
156 choices that benefit household welfare, leading to better living conditions  
157 through improved decision-making and the adoption of more efficient energy  
158 sources (Crentsil et al. (2019)). Place of residence, gender, and household  
159 size also exhibit a statistically significant relationship with multidimensional  
160 poverty due to increased energy consumption needs (Abbas et al. (2020)).  
161 Additionally, age effects may arise from life cycle patterns, household ar-  
162 rangements, and risk-taking behavior, while poor health conditions may hin-  
163 der access to energy services and goods by altering spending priorities and  
164 consumption patterns (Fry et al. (2022)). Labor market status, as well as  
165 marital status, are frequently found to be significantly associated with en-  
166 ergy deprivation, with the effect being particularly pronounced in developing  
167 economies (Abbas et al. (2020); Awan et al. (2022); Manasi and Mukhopad-  
168 hyay (2024)). Cultural characteristics and parental behavior (Prakash et al.  
169 (2022)), and energy subsidies also contribute to shaping energy deprivation  
170 outcomes (Hosan et al. (2023)).

171 Despite these advances, a significant gap in the literature persists: un-  
172 derstanding how current circumstances shape energy poverty outcomes later  
173 in life. The studies discussed above primarily emphasize contemporaneous  
174 relationships between explanatory factors and energy poverty, regardless of  
175 whether the findings stem from cross-sectional or panel data analyses. Stud-  
176 ies on energy poverty dynamics are scarce, with only a few papers addressing  
177 this issue through dynamic panel models in which energy poverty is allowed  
178 to depend on past energy poverty (Alem and Demeke (2020); Drescher and  
179 Janzen (2021); Halkos and Kostakis (2023)).

## 2.1. Machine learning models in energy poverty research

A recent body of literature has introduced ML techniques to predict energy poverty outcomes. Evidence based on an XGBoost framework to predict the risk of experiencing energy poverty in the Netherlands identifies income, house value, and house ownership as the main drivers of energy poverty (Dalla Longa et al. (2021)). In a similar setting, and based on 11 European countries, income, household size, and floor area were consistent predictors (van Hove et al. (2022)). Evidence based on an RF classifier across the European Union uncovers household- and country-level predictors like dwelling conditions, energy efficiency, and gas supplier switching rates (Spandagos et al. (2023)).

While the previous studies are based on a single energy poverty indicator, other studies define a multidimensional energy poverty index similar to ours. These studies showed that in Asian and African countries, wealth, marital status, and residence attributes are significant predictors of poverty (Abbas et al. (2020)). Recent research has further advanced these methodologies by employing ensemble models, such as XGBoost, combined with RF and ANN, revealing the critical importance of education and food security indicators in determining energy poverty (Gawusu et al. (2024)).

## 2.2. Measurement

The literature typically distinguishes between objective (expenditure-based) and subjective (self-assessed) approaches. Because poorer households often spend higher proportions of their budget on energy-related expenses relative to higher-income households (Sy and Mokaddem (2022)), expenditure-based measures label a household as energy-poor when the income that households spend on energy is above a specific threshold. For instance, a household may be classified as energy poor if i) its share of income spent on energy is greater than twice the national median (the 2M indicator); ii) its share of income spent on energy exceeds 10% (the Ten Percent Rule, TPR); or iii) its actual energy expenditures are above the national median and, at the same time, their income net of energy costs is below the official national income poverty line (the Low Income High Costs indicator, LIHC). These measures have been used extensively in the literature (Fry et al. (2022); Awan et al. (2022); Manasi and Mukhopadhyay (2024)).

However, while expenditure-based measures are objective and transparent, they may overlook intentional reduction in energy consumption by low-income households. If vulnerable households limit their energy consumption

217 to prioritize other services and goods, measures based on the actual energy  
218 costs may underestimate the true prevalence of energy poverty. Moreover,  
219 low-income families can resort to energy credits and repayments to smooth  
220 their monthly energy costs over time. To overcome these limitations, applied  
221 research has relied on individuals' self-evaluations of their ability to afford  
222 and access specific energy services (Prakash et al. (2022); Spandagos et al.  
223 (2023)). Following this criterion, several multidimensional energy poverty  
224 indexes have been proposed, gathering information related to basic energy  
225 services, including cooking, lighting, and household appliances in developing  
226 countries (Abbas et al. (2020); Gawusu et al. (2024)).

### 227 3. Data and key variables

228 We use the HILDA Survey, a comprehensive, nationally representative  
229 longitudinal study that examines the economic, social, and demographic dy-  
230 namics of Australian households. Initiated in 2001 and conducted annually,  
231 it tracks individuals and households over time, providing important infor-  
232 mation about income, labor market activities, health, education, and family  
233 relationships, among other factors. The original 2001 sample included ap-  
234 proximately 7,600 households and 13,000 individuals, with periodic updates  
235 to account for attrition. While panel data is subject to selection and attri-  
236 tion bias, potentially limiting the generalizability of findings, HILDA has a  
237 high average retention rate of over 90% across waves. Nonetheless, to ad-  
238 dress concerns about attrition bias, several sensitivity checks are presented  
239 in Section 6.

240 We utilize a balanced panel, allowing for varying durations. Our bench-  
241 mark analysis relies on data spanning up to  $T = 8$  consecutive years, enabling  
242 us to conceptualize energy poverty at time  $T$  as a function of characteristics  
243 from the previous  $T - 1$  periods. This approach yields 106,475 observations  
244 from a cohort of 7,977 individuals with complete records. To enhance the ro-  
245 bustness of our findings, in Section 6 we present additional results for panels  
246 spanning  $T = 2$ ,  $T = 4$ ,  $T = 12$ , and  $T = 14$  years.

247 We model energy poverty as a function of socioeconomic factors that are  
248 standard in the literature. These include household income, employment  
249 status, schooling, age, marital status, parenthood, health status, and house-  
250 hold size. We also include controls for remoteness, region of residence (the  
251 six states and two territories of Australia, reference: New South Wales), and  
252 wave-specific effects. Due to their potential impact on energy poverty, we use

253 annual electricity and gas prices at the state level drawn from the Australian  
254 Bureau of Statistics (Australian Bureau of Statistics (2024)). All income and  
255 price variables used in the paper are transformed using the OECD equiva-  
256 lence scale and normalized into real terms using the yearly consumer price  
257 index. We also include variables to control for macroeconomic conditions at  
258 the regional level. The economic cycle affects the chance to find and keep  
259 jobs, and it also impacts the likelihood of having a stable income source.  
260 We include controls for the regional unemployment rate, per capita GDP,  
261 and GDP growth. We also include the regional participation rate to capture  
262 competition effects in the labor market and the labor force share of part-time  
263 workers to account for the fact that areas with a higher proportion of tem-  
264 porary and/or part-time contracts typically experience greater uncertainty  
265 in work hours and income stability. In Appendix A we provide a detailed  
266 summary of the variables used in the analysis.

### 267 3.1. Energy poverty

268 Energy poverty is a multifaceted construct; therefore, we rely on five  
269 items that capture both expenditure-based and subjective dimensions. The  
270 expenditure-based measures include the 2M, TPR, and LIHC indicators,  
271 which are widely recognized in the energy poverty literature and detailed  
272 in Section 2.2. We also consider two self-assessed indicators based on the  
273 household’s inability to pay to heat their home because of a shortage of  
274 money (Heat) and pay electricity, gas, or telephone bills on time (Arrears).

275 The MEPI index is calculated as follows: Let  $J = 5$  represent the set of  
276 poverty indicators, with element  $j$ ,  $j \in J$  and  $m = \text{card}(J)$ . Let  $I$  be a set  
277 of individuals, with element  $i$ ,  $i \in I$ , and  $T$  be a set of time periods,  $t \in T$ ,  
278 representing a specific moment when the survey was conducted. Let  $EP_{ijt}$   
279 denote the status of the  $i$ th individual in the  $j$ -th indicator during period  
280  $t$ . If an individual  $i$  is poor under indicator  $j$  in the period  $t$ , then  $EP_{ijt}$   
281 takes the value of one, and zero otherwise. Following the family of indexes  
282 typically described in the literature on material deprivation (Dhongde et al.  
283 (2019)), individual  $i$ ’s weighted poverty score is given by:

$$\text{MEPI}_{it} = \left( \sum_{j \in J} w_j EP_{ijt} \right), \quad \forall i \in I, t \in T_i; T_i \subseteq T, \quad (1)$$

284 where  $w_j$  denotes the weight assigned to the poverty indicator  $j$ , with  
285  $\sum_{j \in J} w_j = 1$ . Hence, the  $\text{MEPI}_{it}$  ranges from 0 to 1 and captures the per-  
286 centage of dimensions in which the individual is deprived. An individual  $i$  is

287 regarded as energy poor if  $\text{MEPI}_{it} > \bar{m}$ , where  $\bar{m}$  is a cut-off point. Thus, our  
 288 dependent variable is a binary variable that takes value one if the individual  
 289 is energy-poor, and zero otherwise. For the baseline parametrization, we set  
 290  $\bar{m} = 0$ . In Section 6, we provide robustness checks with alternative cut-off  
 291 points, namely  $\bar{m} = 0.2$  and  $\bar{m} = 0.4$ .

292 While it is common to assign equal weights to the indicators, we empha-  
 293 size the indicators where deprivation is less common, the so-called frequency-  
 294 based weighting approach (Decancq and Lugo (2013)). The weight given to  
 295 an indicator is proportional to the percentage of individuals *not* classified as  
 296 poor under that specific indicator within a particular state. In other words,

$$w_j = \frac{(1 - n_j)}{\sum_{j \in J} (1 - n_j)}, \quad (2)$$

297 where  $n_j$  is the proportion of poor individuals in dimension  $j$ . This choice  
 298 is motivated by the idea that not having access to common items should be  
 299 a more relevant determinant of deprivation than less common items. Addi-  
 300 tionally, the weights are based on the distribution of achievements in society  
 301 without considering any value judgment about what the trade-offs between  
 302 items should be. For greater granularity and accuracy, the weights are cal-  
 303 culated separately for each wave. There are two advantages to using that  
 304 approach. Firstly, it allows the poverty of a given individual to increase if  
 305 their conditions do not change and the conditions of all others improve. Sec-  
 306 ondly, it adapts automatically over time, considering economic conditions  
 307 and social and cultural preferences when accessing items.

308 The MEPI shows two desirable characteristics, as it can be used to mea-  
 309 sure the prevalence and average intensity of energy poverty in a population.  
 310 Prevalence is given by:

$$p = \frac{q}{\text{card}(I)}, \quad (3)$$

311 where  $q$  is the number of deprived individuals,  $q = \sum_{i \in I, t \in T_i} \mathbb{I}(\text{MEPI}_{it} > \bar{m})$ ,  
 312 where the indicator function  $\mathbb{I}(\cdot)$  equals one if its argument holds, and zero  
 313 otherwise. The intensity of energy poverty, i.e., the average poverty score of  
 314 individuals identified as energy poor, is:

$$a = \frac{\sum_{i \in I, t \in T_i} \text{MEPI}_{it} \times \mathbb{I}(\text{MEPI}_{it} > \bar{m})}{q}. \quad (4)$$

315 The average population MEPI is then:

$$\text{MEPI} = a \times p. \quad (5)$$

316 The advantages of these axiomatic properties have been highlighted in  
317 previous work of Crentsil et al. (2019).

## 318 4. Methodological approach

319 We employed ML techniques to model energy poverty at time  $T$  as a  
320 function of historical socioeconomic and demographic variables from the pre-  
321 ceding years. Importantly, no data from year  $T$  were used in the predictions,  
322 ensuring that our forecasting is based entirely on prior historical data. Al-  
323 though the model’s accuracy can be improved by including contemporaneous  
324 characteristics, we refrain from doing so for two main reasons. First, our fo-  
325 cus is on the role of historical factors. Introducing contemporaneous variables  
326 could potentially mask the contribution of lagged effects, especially if auto-  
327 correlation exists in the data. Second, and more relevant, including contem-  
328 poraneous variables may introduce reverse causality between energy poverty  
329 and socio-demographic variables, such as health and schooling (Phoumin and  
330 Kimura (2019); Pondie et al. (2024)). By only considering past variables, we  
331 eliminate the risk of current energy poverty influencing these characteristics.

332 We then integrate the ML techniques with an interpretability framework.  
333 This integration allows us not only to predict energy poverty outcomes but  
334 also to understand the contribution of each historical factor to these predic-  
335 tions. This involves a systematic process of data preparation, model devel-  
336 opment, and the application of feature importance and explainability tech-  
337 niques.

### 338 4.1. Data preparation

339 To capture the temporal dynamics of the variables, we created lagged  
340 features, which serve as the input to the predictive models. Generically,  
341 for each original feature, we obtained new features representing their values  
342 from each of the previous years. This transformation ensures that the model  
343 has access to the full temporal history of each variable, enabling it to learn  
344 patterns and relationships that may influence the energy poverty indicator in  
345 the  $T$ -th year. We split the dataset into training, validation, and test subsets  
346 to facilitate model development and evaluation. For our baseline estimates  
347 ( $T = 8$ ), out of the 7,977 participants in our dataset, 6,382 (80%) were  
348 randomly selected for training and validating the predictive models, while  
349 the remaining 1,595 participants (20%) were included in the test set. The

test set was held out and used exclusively to evaluate the final performance of the models, providing an unbiased estimate of their forecasting accuracy.

To avoid data leakage across the splits, each individual was assigned exclusively to one subset, ensuring that no participant’s data appeared in more than one split. Additionally, we removed any user identifiers from the data to prevent the models from learning user-specific patterns, which could limit their generalizability. The year variable (“wave” variable) was also excluded from the input features to ensure that the models focus on patterns within the socioeconomic and demographic variables rather than relying on specific temporal markers.

Before training the models, we standardized the data to ensure consistency and reliability in our modeling process. This involved removing the median and scaling the data according to the interquartile range, a method particularly effective at managing outliers and recommended as a best practice in machine learning (Sullivan et al. (2021)). Such standardization is crucial in predictive modeling; it normalizes all input features to a similar scale, thereby enhancing the model’s generalization capabilities and preventing variables with larger magnitudes from disproportionately influencing the learning process (Mahmud Sujon et al. (2024)). To prevent data leakage, the scaling parameters were calculated using only the training set and then applied to the test sets.

Feature engineering was explored in this study to improve the forecasting power of our models. Specifically, we expanded the set of socioeconomic, geographical, and contextual factors by including a range of interaction terms and decomposing variables into levels and yearly variations. While this approach increased the model’s ability to identify energy-poor households to 78.04% compared to the final model (cf. Table 1), it reduced the overall accuracy, with the ability to correctly classify non-energy-poor households dropping to 59.31%. Moreover, the increased complexity introduced by additional variables would pose practical challenges for policymakers, making the results harder to interpret and apply. Consequently, we retained the model configuration that provided a better balance between performance and practical usability for policy design. The results of the models with the expanded feature set can be provided by the authors upon request.

#### *4.2. Model development*

We treat the energy poverty forecasting task as a classification problem. Specifically, households are classified as energy-poor depending on whether

387 their MEPI is greater than  $\bar{m}$  (cut-off point), where  $\bar{m} = 0$  for the baseline  
388 model.

389 The dataset used in this study exhibited a significant class imbalance,  
390 with most participants (73.55%) being classified as not energy-poor and a  
391 smaller proportion (26.45%) classified as energy-poor. This imbalance poses  
392 challenges for predictive modeling, as standard machine learning methods  
393 tend to favor the majority class, potentially leading to poor performance in  
394 identifying the minority class (Provost (2000)). To address this issue, we  
395 tested three ensemble classifiers, namely random under-sampling boost clas-  
396 sifiers (Seiffert et al. (2010)), balanced bagging classifier (for a review on  
397 bagging classifiers, see, e.g., Galar et al. (2012)), and easy ensemble classifier  
398 (Liu et al. (2009)). Due to space constraints, we describe here only the bal-  
399 anced bagging classifier. The descriptions of the other classifiers are provided  
400 in Appendix B.

401 A balanced bagging classifier is an ensemble technique that combines the  
402 predictions of multiple base models, e.g., decision trees, in order to improve  
403 the robustness and accuracy of the outcomes. This method specifically ad-  
404 dresses class imbalance by ensuring that each decision tree in the ensemble  
405 is trained on a balanced subset of the dataset. These subsets are created by  
406 resampling the original training data, wherein each subset contains a rep-  
407 resentative distribution of both minority (energy-poor) and majority (not  
408 energy-poor) classes. In order to further refine the modeling approach, we  
409 implemented the classifiers in an One-vs-the-Rest (OvR) binary classification  
410 framework (Murphy (2012)). OvR decomposes the problem into multiple  
411 binary classification tasks, where each class is treated as a separate binary  
412 problem against all other classes. Although OvR is commonly used for multi-  
413 class classification tasks, this methodology fits one classifier per class, which  
414 enables the models to focus on the distinctions between the two groups.

415 We optimized the hyperparameters of our classifiers using a grid search,  
416 which tested various configurations to identify settings that maximize model  
417 performance. For details on the specific hyperparameters and grid configura-  
418 tions, see Appendix B. We employed 5-fold cross-validation on the training  
419 dataset to ensure the robustness of the hyperparameters across different data  
420 splits, selecting the best set based on the highest Receiver Operating Char-  
421 acteristic - Area Under Curve (ROC AUC) score. This metric is crucial for  
422 datasets with class imbalances, like the HILDA Survey, as it fairly assesses  
423 the model’s discriminatory power between energy-poor and non-energy-poor  
424 households.



425 The final model was trained on the complete training set using the identi-  
 426 fied optimal hyperparameters and subsequently evaluated on a held-out test  
 427 set of 1,595 participants. This approach ensured an unbiased assessment of  
 428 the model’s forecasting accuracy. All experiments were conducted with a  
 429 fixed seed to guarantee reproducibility.

### 430 4.3. Feature importance and explainability

431 To interpret the forecast of our model and understand the contributions of  
 432 individual features, we employ the SHAP. SHAP is a well-known method for  
 433 explainability in the literature due to its theoretical consistency and ability  
 434 to provide both local and global explanations of model behavior (Lundberg  
 435 and Lee (2017)). It is rooted in cooperative game theory and assigns each  
 436 feature a contribution value toward the model’s prediction, being thus model  
 437 agnostic. The SHAP value ( $\phi$ ) for a given feature  $k$  is given by,

$$\phi_k = \sum_{S \subseteq N \setminus \{k\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{k\}) - f(S)], \quad (6)$$

438 where  $N$  is the set of all features,  $S$  is a subset of features excluding feature  
 439  $k$ , and  $f(S)$  is the model’s prediction based only on the features in the subset  
 440  $S$ . This equation ensures that each feature’s contribution is fairly allocated  
 441 by accounting for all possible combinations of features.

442 We use SHAP values to evaluate the importance of each input variable  
 443 in the model and to identify when a particular variable had the most sig-  
 444 nificant influence on the predictions. We did not perform feature selection  
 445 before training the models, despite its potential to improve the overall model  
 446 performance. This decision ensured that no variables were excluded prema-  
 447 turely, allowing the model to consider all socioeconomic, geographical, and  
 448 contextual factors and their interactions. With this approach, we can iden-  
 449 tify not only which variables to target for interventions but also the optimal  
 450 timing for these interventions.

## 451 5. Results

### 452 5.1. Models evaluation

453 Among the models evaluated, the balanced bagging classifier achieved the  
 454 highest average ROC AUC of  $73.22\% \pm 4.48\%$ , outperforming the random  
 455 under-sampling boost classifier ( $69.50\% \pm 6.22\%$ ) and the easy ensemble

Table 1: Predictive performance of the balanced bagging classifier model across varying time windows.

Window size ( $T$ )	Sensitivity (%)	Specificity (%)	ROC AUC (%)
2	70.39	69.23	69.81
4	70.82	73.65	72.24
<b>8</b> (baseline)	<b>73.25</b>	<b>66.77</b>	<b>70.01</b>
12	71.60	65.79	68.69
14	78.74	66.31	72.52

**Notes:** This table highlights the trade-offs between sensitivity, specificity, and ROC AUC. Sensitivity reflects the model’s ability to correctly identify energy-poor households, while specificity measures its ability to correctly identify non-energy-poor households. ROC AUC evaluates the model’s overall capacity to discriminate between energy-poor and non-energy-poor households across varying decision thresholds. These results were obtained from the evaluation of unseen data (i.e., unseen participants).

456 classifier ( $70.59\% \pm 4.82\%$ ). Based on these findings, the balanced bagging  
 457 classifier was selected for detailed analysis. Performance metrics for the other  
 458 models are provided in Appendix B Table C.3. A grid search was conducted  
 459 to optimize the balanced bagging classifier’s configuration for the baseline  
 460 window. The best setup included 100 estimators with bootstrapping of fea-  
 461 tures but not samples. Each estimator sampled 50% of the data, and the  
 462 sampling strategy ensured an equal representation of energy-poor and non-  
 463 energy-poor instances. Replacement was used in the resampling process.

464 Table 1 presents the performance metrics. Sensitivity reflects the model’s  
 465 ability to identify energy-poor households, specificity measures its ability to  
 466 identify non-energy-poor households, and ROC AUC assesses overall discrim-  
 467 ination performance. The baseline window ( $T = 8$ ) achieved a ROC AUC of  
 468 70.01%, with a sensitivity of 73.25% and a specificity of 66.77%.

469 The results in Table 1 reveal how window size influences performance.  
 470 Shorter windows ( $T = 2$ ) yield balanced sensitivity (70.39%) and specificity  
 471 (69.23%), with a ROC AUC of 69.81%. At  $T = 4$ , specificity improves sig-  
 472 nificantly to 73.65%, leading to a higher ROC AUC of 72.24%. The baseline  
 473 window ( $T = 8$ ) prioritizes sensitivity, achieving the highest value 73.25%,  
 474 but with a slightly lower specificity. For longer windows ( $T = 12$  and  $T = 14$ ),  
 475 performance varies:  $T = 14$  achieves the highest ROC AUC of 72.52% by

476 increasing sensitivity to 78.74%, though specificity stabilizes at 66.31%. On  
 477 the other hand,  $T = 12$  obtains the lowest ROC AUC.

478 Overall, shorter windows ( $T = 2$ ) favor specificity, while longer windows  
 479 enhance sensitivity. This can be related to the nature of shorter windows  
 480 capturing more immediate and recent information, which tends to reduce  
 481 false positives and improve specificity. In contrast, longer windows incor-  
 482 porate cumulative historical data, allowing the model to better detect pat-  
 483 terns associated with energy poverty over time, which enhances sensitivity  
 484 by reducing false negatives. The choice of window size thus depends on the  
 485 specific policy objective, whether it prioritizes minimizing false positives or  
 486 false negatives.

## 487 5.2. Main explanatory factors and initial policy recommendations

488 In this section, we analyze the predictive power of the model’s features,  
 489 while the discussion of their directional effects is addressed in the next sec-  
 490 tion. Figure 1 shows those factors that contribute at least 1 % to the observed  
 491 outcome over the entire 8-year time window, ranked by order of importance.  
 492 Household income emerges as the most critical determinant, contributing  
 493 38.84% to the total predictive importance. Notably, changes in household  
 494 income rank as the second most influential predictor, accounting for 11.29%.  
 495 Variables with medium explanatory power include the part-time employ-  
 496 ment rate (7.31%), which underscores the role of labor market dynamics  
 497 in shaping energy poverty, and household size (6.82%), likely due to the  
 498 balance between higher energy consumption needs and economies of scale.  
 499 Energy prices (5.80%) emerge as the fifth predictive factor, and years of ed-  
 500 ucation (5.67%) emphasize the interplay between human capital and energy  
 501 poverty. Lower-contribution factors include poor health (3.09%), employ-  
 502 ment status (2.78%), macroeconomic indicators such as the unemployment  
 503 rate (2.53%), Gross State Product (GSP) per capita (2.38%), GSP per capita  
 504 growth (2.26%), and the total labor force participation rate (2.11%). Finally,  
 505 demographic and family characteristics such as the number of children at  
 506 home, age groups, and marital status round out the analysis.

507 Figure 2 breaks down the results from Figure 1 across the different time  
 508 lags. Household income consistently stands out as the most critical predic-  
 509 tor, with its impact peaking at  $T - 1$  (10.85%) and gradually diminishing  
 510 over longer lag periods ( $T - 2$  : 8.55%,  $T - 3$  : 4.50%,  $T - 4$  : 4.41%,  $T - 5$  :  
 511 4.17%,  $T - 7$  : 3.99%). Household income changes are also among the top

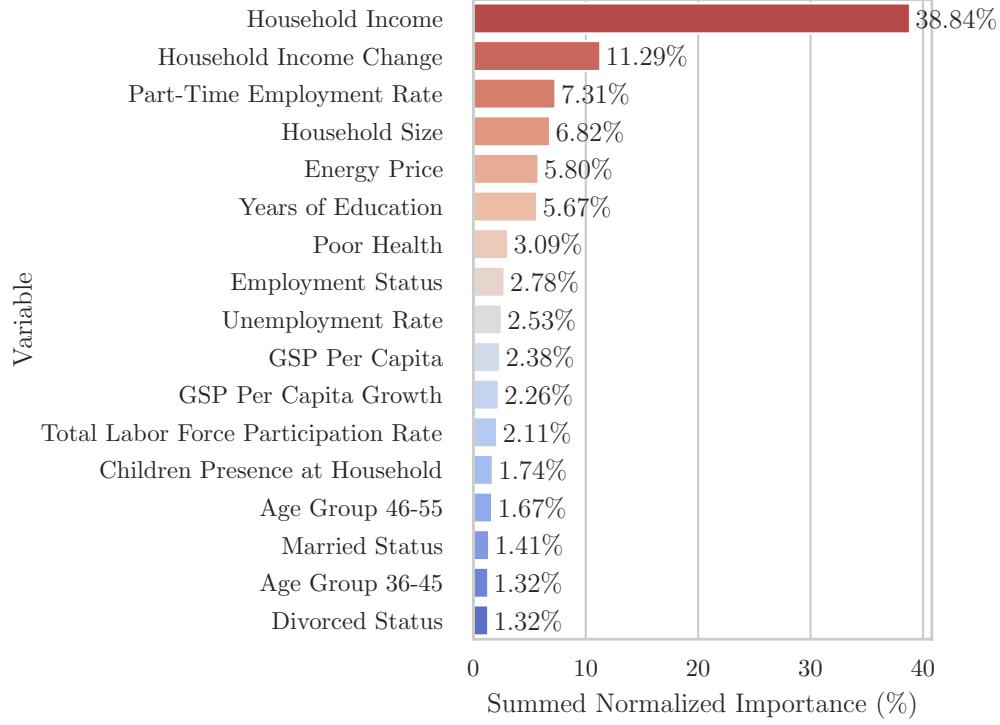


Figure 1: Relative contribution (%) of predictive variables for energy poverty outcomes across a 8-year time window.

**Notes:** i) The figure presents the top predictors with a summed normalized importance of at least 1% for energy poverty outcomes; ii) Source: HILDA 2007–2021 waves.

512 predictors, particularly at  $T-1$  (2.55%) and  $T-2$  (2.21%). Additional contri-  
513 butions come from household income changes at  $T-3$  and  $T-6$ , suggesting  
514 that historical fluctuations in income continue to influence household energy  
515 vulnerability years later. Energy prices operate mainly through a one-year  
516 lag, highlighting the effects of short-term fluctuations. Household size at  
517  $T-1$  (1.80%) and  $T-2$  (1.22%) reflect the immediate impact of family com-  
518 position on energy poverty. The part-time employment rate also emerges  
519 as an important variable, particularly at  $T-2$  (1.55%) and  $T-6$  (1.50%),  
520 pointing to the relevance of the regional employment structure for household  
521 energy vulnerability.

522 Overall, from a policy perspective, the results offer a set of initial insights.

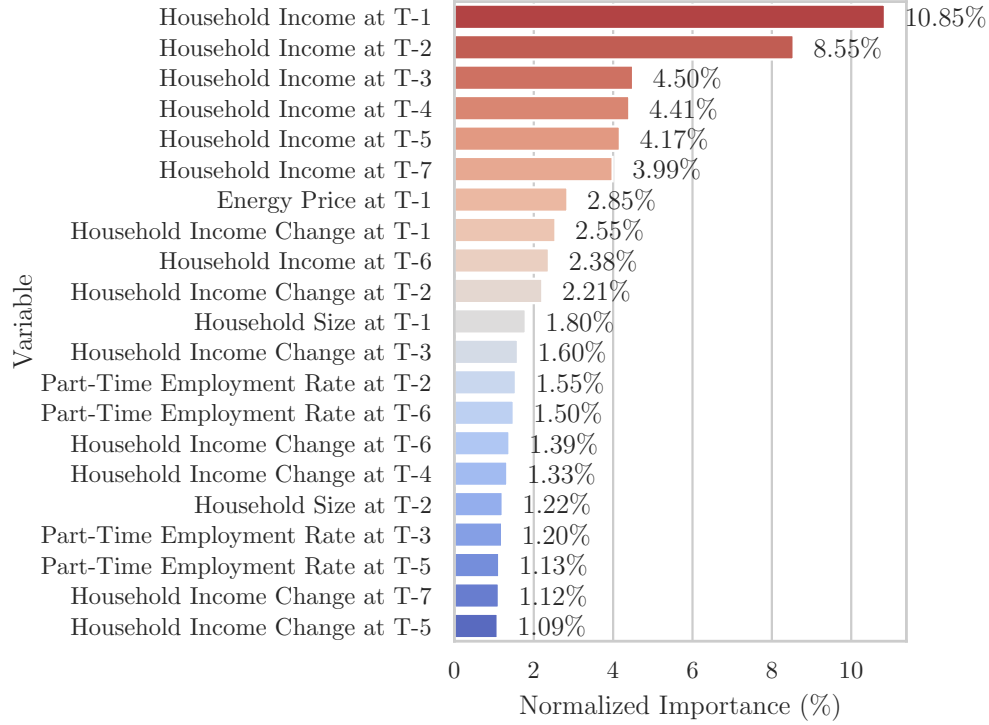


Figure 2: Relative contribution (%) of predictive variables for energy poverty outcomes across an 8-year time window—discriminated by period.

**Notes:** (i) The figure presents the normalized importance of predictive variables for energy poverty outcomes across individual time lags. Only predictors with a normalized importance of at least 1% at any lag are shown; (ii) The suffix “ $T - j$ ” indicates the time lag of the feature relative to the prediction for year  $T$ ; (iii) Source: HILDA 2007–2021 waves.

523 First, the strong association between income across all lags and current en-  
 524 ergy poverty suggests that income can serve as an indicator to identify in-  
 525 dividuals at risk of energy poverty, even in the long-term. Second, Figure 3  
 526 focuses on the top 5 contributing variables and their relative contribution  
 527 over time. The growing importance of household income and income changes  
 528 toward  $T - 1$  suggests that policies aimed at stabilizing income in the short  
 529 term can have a great impact on mitigating immediate energy poverty risks.  
 530 According to our results, such policies may benefit not only those with low  
 531 incomes but also individuals with moderate incomes who experience above-

average income volatility. Third, the contribution of energy prices to energy poverty rises from  $T - 3$  onwards, reflecting the fact that the energy burden over the last 3 years is partly responsible for current energy poverty outcomes. Therefore, price stabilization strategies that extend beyond just one year or rely on occasional interventions could be beneficial for policy. Finally, the contribution of household size also grows steadily over the time window, suggesting that energy poverty is critically influenced by recent adjustments in household arrangements and the changes in energy needs and economies of scale associated with them. In the next section, we identify key household sizes.

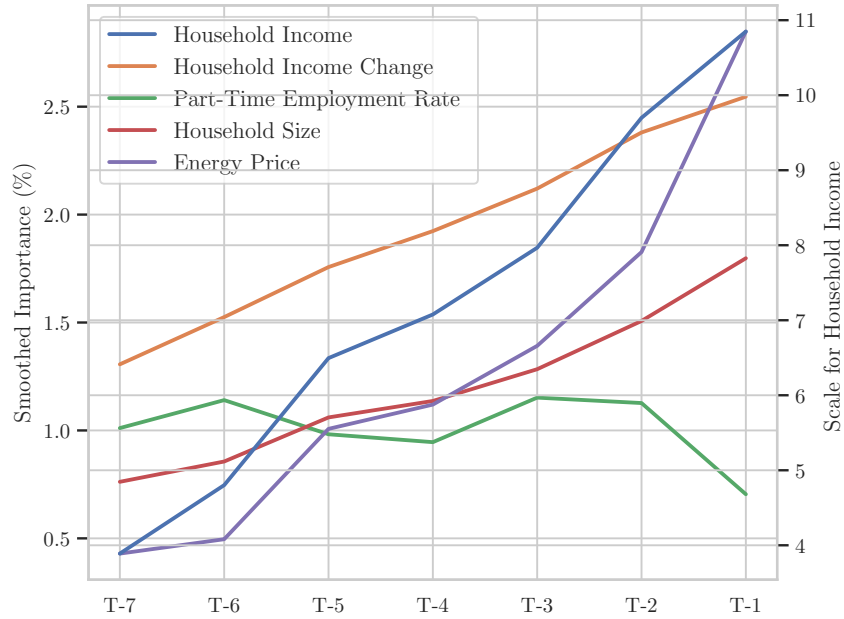


Figure 3: Evolution of the relative contribution (%) of the top five predictive variables for energy poverty outcomes across a 8-year time window.

**Notes:** i) the figure highlights the temporal trends, persistence, and shifts in the influence of different variables across the individual time lags; ii) Source: HILDA 2007–2021 waves.

### 5.3. How key predictive variables shape energy poverty outcomes

This section explores how key predictive variables influence their SHAP contributions. A positive SHAP value indicates a higher probability of energy

545 poverty, while a negative value reflects a reduced risk. The results, shown  
546 in Figure 4, are suggestive of some non-linear relationships. To facilitate  
547 interpretation, a fourth-degree least squares polynomial fit was applied to  
548 highlight the main trends. However, caution is advised at the plot extremes,  
549 where sparse data points may undermine the reliability of interpolations. As  
550 household income increases, the SHAP value decreases sharply. However,  
551 this effect is more intense at low and moderate income levels than at high  
552 incomes. Similarly, the scatter plot for yearly income variations is suggestive  
553 of a somewhat asymmetric effect, with income losses being relatively more  
554 relevant for energy poverty than income gains. This pattern reinforces earlier  
555 insights that interventions like income insurance, unemployment benefits, or  
556 programs aimed at shielding households from income shocks are essential for  
557 mitigating these risks.

558 The part-time employment rate contributes to the energy poverty risk,  
559 particularly in areas where the part-time employment rate exceeds 30%. One  
560 possible explanation is that part-time jobs reflect labor market and income  
561 instability. These positions often lack critical benefits, such as health insur-  
562 ance or retirement plans, which heightens financial vulnerability. Addition-  
563 ally, fluctuating hours and earnings further amplify economic uncertainty.  
564 At lower part-time employment rates (below 25%), SHAP values remain rel-  
565 atively stable, indicating a minimal influence. These findings indicate that  
566 policies promoting income stability, benefits for part-time workers, and ac-  
567 cess to full-time employment opportunities are crucial for tackling energy  
568 poverty in regions with high part-time employment rates. Additionally, the  
569 results in Figure 3 reveal that regional labor market dynamics can have de-  
570 layed impacts on energy poverty, suggesting that such policies could produce  
571 lasting effects.

572 The relationship between household size and SHAP values highlights a  
573 clear risk group: people living alone or in two-person households. This is  
574 likely because fixed energy costs pose a disproportionately heavy burden on  
575 them. As household size increases to 3–4 members, the likelihood of energy  
576 poverty decreases, likely reflecting economies of scale in energy consumption,  
577 which reduce the per-capita cost burden.

578 Lastly, energy prices display a notable pattern, suggesting that below  
579 a certain threshold, they are not relevant for energy poverty. At low price  
580 levels, SHAP values remain relatively stable, but they increase steadily above  
581 \$0.228 and rise significantly beyond \$0.266. This is a relevant finding, as  
582 most representations in the literature describe the effect of energy prices on

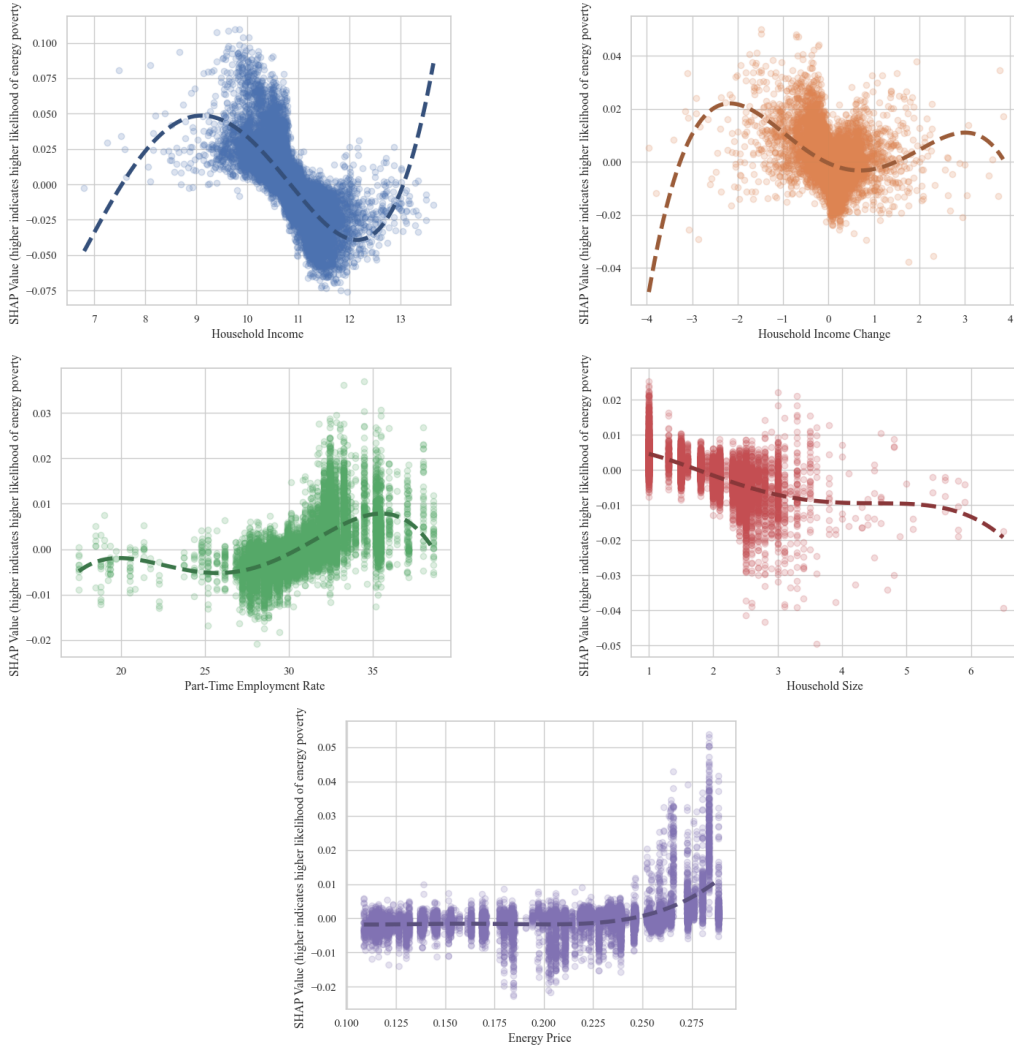


Figure 4: Relationship between key predictive variables and SHAP values for energy poverty outcomes.

**Notes:** i) Each point represents a household, with the x-axis indicating the feature value and the y-axis showing the SHAP value, which reflects the feature's contribution to the forecasting. Positive SHAP values indicate a higher likelihood of energy poverty, while negative values suggest a reduced risk. The plots highlight how changes in the variables influence the model's forecasts; ii) The dashed lines summarize the underlying trends and were calculated using a least squares polynomial fit of degree 4. The interpolations at the extremes of the plots may lack reliability due to the sparse data points in these regions, potentially leading to less accurate representations of the trend; iii) Source: HILDA 2007–2021 waves.



energy poverty in a linear, average manner. However, the results indicate that the relationship between energy prices and energy poverty is non-linear and nearly flat within certain ranges. In this context, energy subsidies and price controls may be ineffective within these ranges, whereas informational campaigns and targeted support for individuals exposed to high prices could play a crucial role.

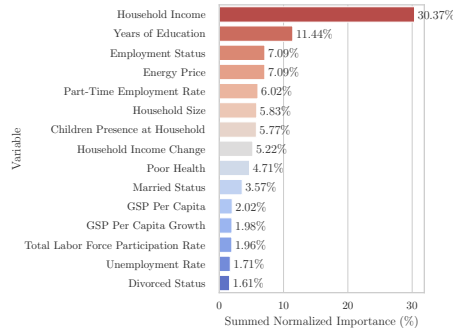
## 6. Sensitivity checks

In this section, we conduct a series of supplementary analyses. Specifically, we explore the sensitivity of the results to variations in i) the time length considered for the analysis and ii) the chosen cut-off point for defining the energy poverty line. We also examine to what extent our findings might be affected by selection and attrition bias.

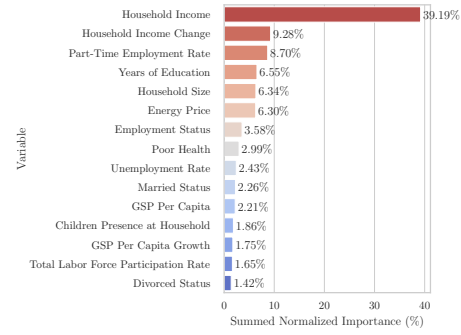
In Figure 5 we depict the relative contribution of the predictive variables for alternative time spans. The results show robust consistency across scenarios, with household income emerging as the most significant determinant of energy poverty, irrespective of the time span. Notably, the predictive power of income changes significantly, increasing more than threefold from about 5% when  $T = 2$  to over 17% when  $T = 12$  or  $T = 14$ . This suggests that income volatility and the uncertainty it creates are crucial factors influencing long-term energy poverty outcomes.

Additionally, energy prices are relatively important in the short term (6–7%), but their relevance decreases over the long-term (< 4%). Similarly, improvements in education levels are more strongly associated with short-term energy poverty outcomes than with long-term ones. Household size maintains a consistent level of importance across both short- and long-term periods, reinforcing its stable role as a determinant of energy poverty.

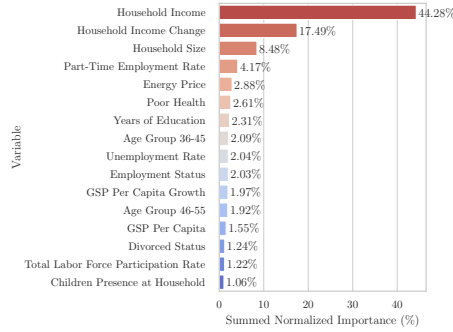
In Figure 6 we discriminate across the different time lags. Perhaps the most relevant finding is that household income in the immediate past ( $T - 1, T - 2, T - 3$ ) holds less accumulated relevance for long-term energy poverty outcomes compared to short-term outcomes. This underscores the notion that energy poverty is influenced by a “long memory” process, where the individual’s entire history—albeit with diminishing weight—plays a critical role. Lagged energy prices are in the list of top contributors for energy poverty outcomes at  $T = 2$  and  $T = 4$ . However, they disappear for  $T = 14$ , suggesting that in the long-term, the structural aspects of the individual are relatively more relevant than energy prices.



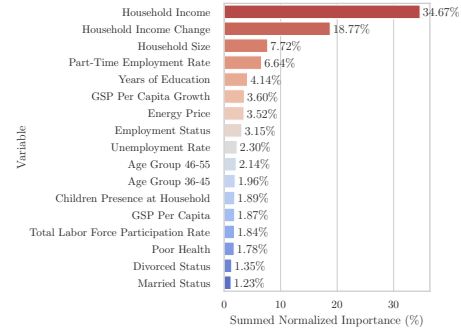
(a) ( $T = 2$ )



(b) ( $T = 4$ )



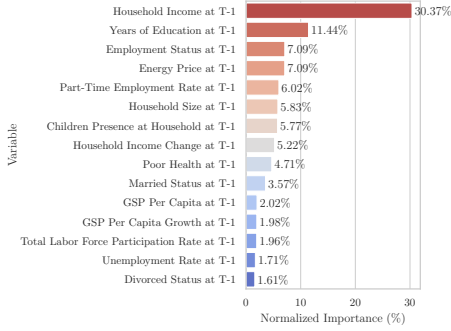
(c) ( $T = 12$ )



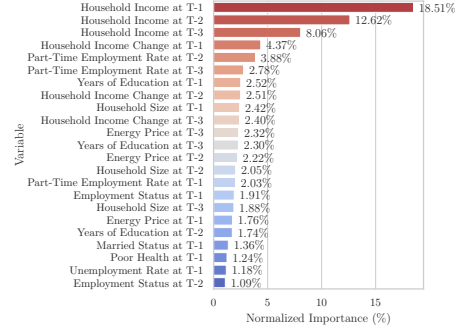
(d) ( $T = 14$ )

Figure 5: Relative contribution (%) of predictive variables for energy poverty outcomes across alternative time windows.

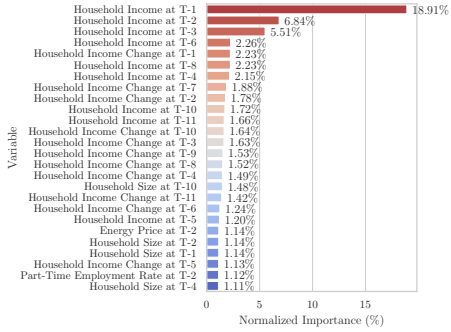
**Notes:** i) This figure presents the top predictors with a summed normalized importance of at least 1% for energy poverty outcomes. ii) Source: HILDA 2007–2021 waves.



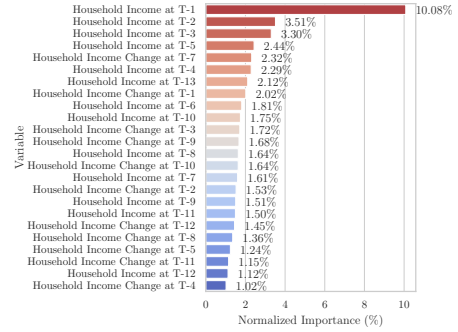
(a) ( $T = 2$ )



(b) ( $T = 4$ )



(c) ( $T = 12$ )



(d) ( $T = 14$ )

Figure 6: Relative contribution (%) of predictive variables for energy poverty outcomes across a T-year time window- discriminated by period.

**Notes:** i) This figure presents the normalized importance of predictive variables for energy poverty outcomes across individual time lags. Only predictors with a normalized importance of at least 1% at any lag are shown; ii) The suffix “ $T - j$ ” indicates the time lag of the feature relative to the prediction for year  $T$ ; iii) Source: HILDA 2007–2021 waves.

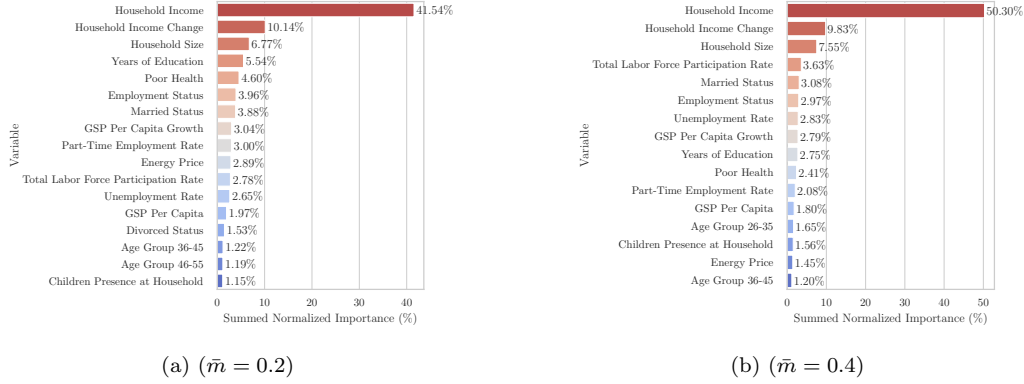


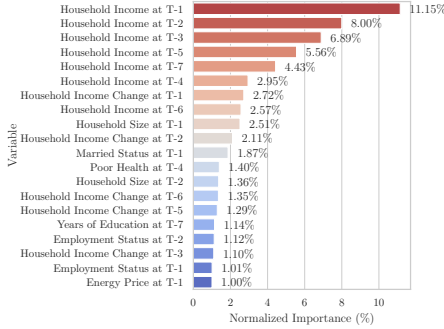
Figure 7: Relative contribution (%) of predictive variables for energy poverty outcomes across a 8-year time window for different cut-off values.

**Notes:** This figure presents the normalized contribution of predictive variables for energy poverty outcomes across individual time lags. Only predictors with a normalized importance of at least 1% at any lag are shown. The results are for a  $T = 8$  year time window.

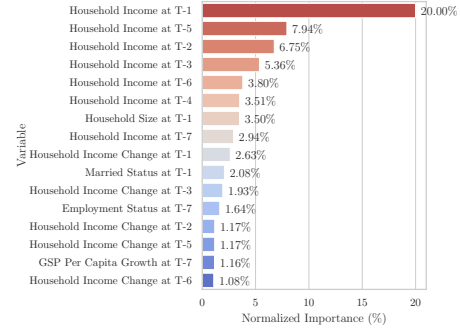
In Figure 7 we conduct additional sensitivity checks and present results using more stringent criteria for energy poverty ( $\bar{m} = 0.2$  and  $\bar{m} = 0.4$ ). The estimates are based on  $T = 8$ , as in the baseline estimates. The contribution of income rises from approximately 39% in the baseline estimates ( $\bar{m} = 0$ ) to 50.3% when  $\bar{m} = 0.4$ . Reversely, the contribution of energy prices falls from 5.8% in the baseline model to 1.5% when  $\bar{m} = 0.4$ , suggesting that energy prices are not a primary driver of severe energy poverty. Additionally, marriage emerges as a protective factor against stricter definitions of poverty ( $>2.5\%$ ), highlighting its buffering effect in more vulnerable contexts. Finally, Figure 8 documents the timing effects, with income in the previous year gaining importance when accounting for the most stringent definition of energy poverty.

### 6.1. Is attrition endogenous?

Although the average entry rate (individuals not in the sample in the previous period who are in the current period) and exit rate (individuals who leave the sample) are very moderate in our sample (8.9% and 7.4%, respectively), the nonrandom exit and entry of individuals for reasons related to energy poverty is a potential concern. To address this issue, we conducted a regression using a dummy variable that equals one if the individual exits the sample in the following year and zero otherwise, against energy poverty



(a) ( $\bar{m} = 0.2$ )



(b) ( $\bar{m} = 0.4$ )

Figure 8: Relative contribution (%) of predictive variables for energy poverty outcomes across a 8-year time window for different cut-off values - discriminated by period.

**Notes:** Notes: This figure presents the normalized contribution of predictive variables for energy poverty outcomes across individual time lags. Only predictors with a normalized importance of at least 1% at any lag are shown. The results are for a  $T = 8$  year time window.

and all controls and obtained a coefficient equal to -0.003 ( $p - value = 0.454$ ). In other words, leaving the sample is not significantly related to energy poverty. We proceeded likewise with individuals entering the sample, and energy poverty showed a significant negative effect -0.005 ( $p - value = 0.082$ ). . This suggests that the incorporation of new panelists in the sample over the years is not completely random, with a slight tendency to incorporate people who are less likely to suffer energy poverty. These individuals may be either less difficult to contact or more ready to join the panel, although once they decide to participate, their attrition is mostly random.

## 7. Discussion and conclusions

This study highlights the potential of AI-based methodologies, particularly SHAP, for analyzing the dynamics of energy poverty. It examines the short- and long-term effects of key variables and contextual factors—such as income, energy prices, and regional conditions—on future energy poverty outcomes. By capturing both the timing and magnitude of past events, the study offers a perspective on how these factors shape energy poverty over time. This approach sets our research apart from previous studies, which predominantly rely on static models or contemporaneous relationships between energy poverty and explanatory variables.

658 The paper shows that current energy poverty is the outcome of historical  
659 trajectories. The results are robust to a battery of sensitivity checks, includ-  
660 ing alternative definitions of multidimensional energy poverty and varying  
661 time spans. Income levels emerge as the most critical factor, particularly for  
662 long-term outcomes and under strict definitions of poverty. While the con-  
663 temporaneous relation between income and energy poverty has been high-  
664 lighted in previous work (Dalla Longa et al. (2021); van Hove et al. (2022)),  
665 our results uncover the association between income across all lags and cur-  
666 rent energy poverty. From a policy design perspective, we provide evidence  
667 that income can serve as an effective screening tool for identifying 'future'  
668 energy-poor individuals—those at risk of becoming energy-poor in the years  
669 ahead. Moreover, our results emphasize the critical role of income *changes*.  
670 Historical income fluctuations have lasting effects on household energy vul-  
671 nerability, persisting over time. This insight introduces a new dimension to  
672 combating energy poverty, showing that beyond income levels, individuals  
673 experiencing income volatility and uncertainty constitute a high-risk group.

674 Consistent with numerous studies, we find a positive association between  
675 energy prices and energy poverty (Primc et al. (2021); Spandagos et al.  
676 (2023)). However, our study adds that energy prices have a significant im-  
677 pact in the short term and under less stringent definitions of poverty. In con-  
678 trast, their influence diminishes when addressing long-term energy poverty  
679 or more severe cases. Furthermore, our findings suggest the existence of a  
680 price threshold beyond which energy prices become particularly detrimen-  
681 tal. In this context, measures such as energy subsidies, price controls, and  
682 informational campaigns specifically aimed at individuals facing high energy  
683 prices could play a critical role.

684 The AI-approach used in the paper provides insights that may be used to  
685 shift the policy focus from reactive measures, which address existing poverty,  
686 to preventive strategies that target households showing early signs of vul-  
687 nerability. Specifically, our findings suggest that policymakers can enhance  
688 resilience and reduce long-term socioeconomic disparities by balancing im-  
689 mediate relief measures—such as energy price support, energy benefits, and  
690 income transfers—with structural policies addressing systemic vulnerabilities  
691 identified in our study, particularly income volatility, labour market condi-  
692 tions and small households.

693 This exploratory study has several limitations that warrant further in-  
694 vestigation. A key shortcoming is the failure to account for the endogeneity  
695 of life events, such as income shocks, which may be driven by unobserved

behaviors, situational factors, or omitted variables. Addressing these issues in future research is essential and could involve incorporating more advanced econometric and AI techniques to ensure a better understanding of the mechanisms at work. Another limitation is the uniform treatment of households, which overlooks heterogeneity in responses to energy poverty predictors. Factors such as income, age, education, and personal traits likely influence how individuals experience and respond to energy challenges (Cong et al. (2022)). Future research could improve the granularity and relevance of conclusions by conducting separate analyses based on these dimensions. Finally, this study is focused on a single dataset, the HILDA Survey. While this dataset provides rich, longitudinal information about Australian households, testing the methodology on additional datasets from other regions and contexts with varying degrees of energy poverty would help assess the generalizability of the findings and derive equally meaningful insights for other jurisdictions.

#### **CRedit authorship contribution statement**

**Santiago Budría:** Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review and editing. **Eduardo Fermé:** Investigation, Methodology, Supervision, Writing – original draft. **Diogo Nuno Freitas:** Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review and editing.

#### **Declaration of competing interest**

The authors declare no potential conflict of interest.

#### **Acknowledgments**

We want to thank seminar participants at the I Workshop on “The ecological transition challenge: Energy poverty, Well-being and Sustainability,” held in Madrid in November 2024 for their comments and suggestions.

This work was partially supported by FCT-Fundação para a Ciência e a Tecnologia, Portugal through project PRODY: PTDC/CCI-COM/4464/2020; by NOVA LINC ref. UIDB/04516/2020 (<https://doi.org/10.54499/UIDB/04516/2020>) and ref. UIDP/04516/2020 (<https://doi.org/10.54499/UIDP/04516/2020>) with the financial support of FCT.IP; by LARSyS ref. LA/P/0083/2020

729 (<https://doi.org/10.54499/LA/P/0083/2020>) and ref. UIDP/50009/2020  
730 (<https://doi.org/10.54499/UIDP/50009/2020>) the financial support of  
731 FCT.IP; and with the financial support provided by the 2022 R&D&I  
732 National Projects and 2021 Strategic Projects Oriented to the Ecological  
733 and Digital Transition by the Spanish Ministry of Sciences and Innovation  
734 (Refs: PID2022-143254OB-I00 and TED2021-132824B-I00).

735 Diogo Nuno Freitas was supported by the Portuguese Foundation  
736 for Science and Technology with the grant number 2021.07966.BD  
737 (<https://doi.org/10.54499/2021.07966.BD>).



## Appendix A. Variables of the HILDA survey

Table A.2: Summary of variables used in the study.

Variable	Description	Type	Observations
Participant ID	Identification of the participant.	—	This variable was not included in the prediction model
Year	Year the participant data was collected.	Discrete	This variable was not included in the prediction model
Part-time employment rate	Percentage of the workforce employed part-time.	Continuous	
Unemployment rate	Total unemployment rate.	Continuous	
Unemployment rate change	Year-over-year change in the total unemployment rate.	Continuous	Reflects shifts in labor market conditions and employment opportunities.
Total labor force participation rate	Labor force participation rate.	Continuous	
Total labor force participation rate change	Year-over-year change in the total labor force participation rate.	Continuous	Indicates changes in the proportion of individuals actively participating in the labor market.
GSP per capita	Gross state product per capita.	Continuous	
GSP per capita growth	Growth rate of gross state product per capita.	Continuous	Captures the economic growth fluctuations at the state level.
GSP per capita growth change	Year-over-year change in the growth rate of gross state product per capita.	Continuous	
Energy price	Energy price.	Continuous	Adjusted for inflation.
Energy price change	Year-over-year change in the energy price.	Continuous	Measures fluctuations in energy costs that may impact household energy affordability.
Years of education	Logarithm of the number of years the participant has been educated.	Continuous	
Age	Age	Discrete	
Married	If married (=1) or not (=0)	Binary	
Divorced	If divorced (=1) or not (=0)	Binary	
Widowed	If widowed (=1) or not (=0)	Binary	
Children presence at household	Number of minors at home.	Discrete	
Unemployment status	If unemployed (=1) or not (=0)	Binary	
Poor health	If the individual perceives their health status as bad or very bad (=1) or not (=0).	Binary	
Employment status	If employed (=1) or not (=0).	Binary	
Household size	Household size.	Discrete	
Household income	Logarithm of the household income.	Continuous	
Household income change	Year-over-year change in the household income.	Continuous	Reflects income volatility and its potential effect on household vulnerability to energy poverty.
Household region 2 to 8	Various one-hot encoded regional and household state indicators.	Binary (one-hot encoded)	
MEPI	Poverty indicator.	Continuous	Includes the 2M, TPR, LIHC indicators, and measures of inability to pay for adequate heating (Heat) or utility bills on time (Arrears).

## 738 Appendix B. Model selection and grid search parameters

739 To identify the optimal models and hyperparameter configurations for  
740 predicting energy poverty, a grid search approach was implemented. This  
741 process systematically tested combinations of model parameters and evalu-  
742 ated their performance using cross-validation.

743 We used 5-fold cross-validation on the training dataset to ensure that  
744 the models were tested on various data splits and that the hyperparameters  
745 chosen were robust across different subsets of data. The best set of hyperpa-  
746 rameters was then chosen based on the highest average Receiver Operating  
747 Characteristic - Area Under Curve (ROC AUC) score from the validation  
748 folds. The ROC AUC score measures the model’s ability to discriminate  
749 between energy-poor and non-energy-poor households.

750 For all models, the grid search incorporated time window sizes and cut-off  
751 points to capture the temporal dynamics of energy poverty predictors. Ad-  
752 ditionally, the grid search utilized the one-vs-rest framework, which creates  
753 a binary classifier for each class.

754 The grid search was applied to three machine learning models commonly  
755 used for imbalanced classification tasks. All models were optimized and  
756 trained using Python, and the scikit-learn library for model implementation  
757 and evaluation. The models and their corresponding parameter grids are  
758 described below.

### 759 *Appendix B.1. Model A: Random under-sampling boost*

760 The random under-sampling boost classifier is an ensemble method that  
761 combines boosting with random under-sampling to address class imbalance  
762 efficiently. Based on the AdaBoost algorithm, the classifier under-samples the  
763 training data during each boosting iteration in order to guarantee an equal  
764 representation of the minority (energy-poor) and majority (not energy-poor)  
765 classes. Weak classifiers, such as decision trees, are iteratively trained, with  
766 misclassified instances receiving higher weights. The final model aggregates  
767 predictions from all weak classifiers through weighted voting.

768 The grid search included the following parameters:

- 769 • Number of estimators: 50, 100, 200
- 770 • Learning rate: 0.01, 0.1, 1.0
- 771 • Algorithm type: SAMME, SAMME.R

- 772 • Sampling strategy: Auto, 0.5, 1.0
- 773 • Replacement: True, False

#### 774 *Appendix B.2. Model B: Easy ensemble*

775 The easy ensemble classifier (Liu et al. (2009)) works by creating multiple  
776 balanced subsets of the training data through under-sampling of the majority  
777 class (in the case of this work, the not energy-poor class). For each balanced  
778 subset, an AdaBoost learner (which, in this context, uses a decision tree  
779 as the base estimator) is trained. The outcomes from all subsets are then  
780 combined to form a robust ensemble, and thus, prediction. As a result, the  
781 classifier ensures that the model remains sensitive to the energy-poor class  
782 (minority) while maintaining robust overall performance.

783 The grid search explored the following parameters:

- 784 • Number of estimators: 10, 50, 100
- 785 • Sampling strategy: Auto, 0.5, 1.0
- 786 • Replacement: True, False

#### 787 *Appendix B.3. Model C: Balanced bagging*

788 A balanced bagging classifier (Galar et al. (2012)) is an ensemble tech-  
789 nique that combines the predictions of multiple base models, e.g., decision  
790 trees, in order to improve the robustness and accuracy of the outcomes. This  
791 method specifically addresses class imbalance by ensuring that each decision  
792 tree in the ensemble is trained on a balanced subset of the dataset. These  
793 subsets are created by resampling the original training data, wherein each  
794 subset contains a representative distribution of both minority (energy-poor)  
795 and majority (not energy-poor) classes.

796 The parameter grid included the following parameters:

- 797 • Number of estimators: 10, 50, 100
- 798 • Maximum samples: 0.5, 1.0
- 799 • Maximum features: 0.5, 1.0
- 800 • Bootstrap sampling: True, False
- 801 • Bootstrap feature selection: True, False

- 802     • Sampling strategy: Auto, 0.5, 1.0
- 803     • Replacement: True, False

## 804   **Appendix C. Model performance across time windows and cut-off** 805                                   **value**

806       This appendix presents the performance results of all models tested in  
807 the study, evaluated across various cut-off values ( $\bar{m}$ ) and time windows ( $T$ ).  
808 The results are summarized in terms of sensitivity, specificity and ROC AUC.  
809 The best model was selected based on the highest average ROC AUC value,  
810 calculated independently of specific cut-off values and time windows. This  
811 approach ensures that the chosen model has consistently strong performance  
812 across all configurations tested.

813       Those performance metrics used in this evaluation are briefly described  
814 below:

815   **Sensitivity:** The percentage of correctly identified energy-poor households  
816                   among all actual energy-poor households. Higher sensitivity indicates  
817                   better identification of the minority (energy-poor) class.

818   **Specificity:** The percentage of correctly identified non-energy-poor house-  
819                   holds among all actual non-energy-poor households. Higher specificity  
820                   reflects fewer false positives.

821   **ROC AUC:** (Receiver Operating Characteristic - Area Under Curve)  
822                   A measure of the model’s overall ability to discriminate between  
823                   energy-poor and non-energy-poor households across varying thresholds.  
824                   Higher values indicate better discrimination.

Table C.3: Performance results of models across cutoff values ( $\bar{m}$ ) and time windows. This table summarizes the sensitivity, specificity, ROC AUC for each model evaluated in the study. The results highlight the impact of varying cutoff values and time windows on model performance.

Model	Cut-off ( $\bar{m}$ )	Window size ( $T$ )	Sensitivity (%)	Specificity (%)	ROC AUC (%)	Average ROC AUC (%)
Random under-sampling boost	0	2	67.35	67.66	67.50	69.50 $\pm$ 6.22
		4	65.64	68.83	67.24	
		8	68.26	66.77	67.52	
		(baseline)				
		12	57.40	65.41	61.40	
		14	65.35	70.08	67.72	
	0.2	2	69.08	75.09	72.09	
		4	67.97	76.25	72.11	
		8	68.31	75.59	71.95	
		(baseline)				
		12	55.81	78.57	67.19	
		14	66.67	79.04	72.85	
		2	73.49	78.49	75.99	
		4	74.66	82.32	78.49	
Balanced bagging	0.4	8	79.55	79.80	79.67	
		(baseline)				
		12	40.00	88.71	64.36	
		14	25.00	87.85	56.43	
		2	70.39	69.23	69.81	
		4	70.82	73.65	72.24	
	0	8	73.25	66.77	70.01	73.22 $\pm$ 4.48
		(baseline)				
		12	71.60	65.79	68.69	
		14	78.74	66.31	72.52	
		2	74.32	72.43	73.38	
		4	77.71	72.81	75.26	
		8	73.94	72.21	73.08	
		(baseline)				
Easy ensemble	0.2	12	55.81	85.41	70.61	
		14	85.71	77.66	77.66	
		2	74.46	76.50	75.48	
		4	76.71	78.36	77.54	
		8	79.55	74.27	76.91	
		(baseline)				
	0.4	12	80.00	83.36	81.68	
		14	50.00	76.72	63.36	
		2	67.40	67.47	67.43	
		4	66.39	67.37	66.88	
		8	68.86	65.85	67.35	
		(baseline)				
		12	63.31	62.22	62.77	
		14	69.29	68.19	68.74	
	0	2	69.63	74.47	72.05	70.59 $\pm$ 4.82
		4	69.70	76.15	72.92	
		8	71.13	72.09	71.61	
		(baseline)				
		12	60.47	75.53	68.00	
		14	57.14	84.49	70.81	
	0.2	2	74.22	77.85	76.03	
		4	76.03	79.93	77.98	
		8	77.27	77.63	77.45	
		(baseline)				
		12	40.00	86.40	63.20	
		14	75.00	76.32	75.66	

## 825 References

- 826 Abbas, K., Li, S., Xu, D., Baz, K., Rakhmetova, A., 2020. Do socioeconomic  
827 factors determine household multidimensional energy poverty? Empirical  
828 evidence from South Asia. *Energy Policy* 146, 111754. doi:10.1016/j.  
829 enpol.2020.111754.
- 830 Alem, Y., Demeke, E., 2020. The persistence of energy poverty: A dynamic  
831 probit analysis. *Energy Economics* 90, 104789. doi:10.1016/j.eneco.  
832 2020.104789.
- 833 Alomari, Y., Andó, M., 2024. SHAP-based insights for aerospace PHM:  
834 Temporal feature importance, dependencies, robustness, and interaction  
835 analysis. *Results in Engineering* 21, 101834. doi:10.1016/j.rineng.2024.  
836 101834.
- 837 Australian Bureau of Statistics, 2024. Consumer Price Index, Aus-  
838 tralia. URL: [https://www.abs.gov.au/statistics/economy/  
839 price-indexes-and-inflation/consumer-price-index-australia/  
840 mar-quarter-2024#data-downloads](https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/consumer-price-index-australia/mar-quarter-2024#data-downloads). (Accessed on 29 March 2024).
- 841 Awan, A., Bilgili, F., Rahut, D.B., 2022. Energy poverty trends and de-  
842 terminants in Pakistan: Empirical evidence from eight waves of HIES  
843 1998–2019. *Renewable and Sustainable Energy Reviews* 158, 112157.  
844 doi:10.1016/j.rser.2022.112157.
- 845 Banerjee, R., Mishra, V., Maruta, A.A., 2021. Energy poverty, health and  
846 education outcomes: Evidence from the developing world. *Energy Eco-  
847 nomics* 101, 105447. doi:10.1016/j.eneco.2021.105447.
- 848 Bao, Y., Liao, T., 2024. Multidimensional poverty and growth: Evidence  
849 from India 1998–2021. *Economic Modelling* 130, 106586. doi:10.1016/j.  
850 econmod.2023.106586.
- 851 Boța-Avram, C., Apostu, S.A., Ivan, R., Achim, M.V., 2024. Exploring the  
852 impact of macro-determinant factors on energy resource depletion: Evi-  
853 dence from a worldwide cross-country panel data analysis. *Energy Eco-  
854 nomics* 130, 107341. doi:10.1016/j.eneco.2024.107341.
- 855 Cong, S., Nock, D., Qiu, Y.L., Xing, B., 2022. Unveiling hidden energy  
856 poverty using the energy equity gap. *Nature communications* 13, 2456.

- 857 Crentsil, A.O., Asuman, D., Fenny, A.P., 2019. Assessing the determinants  
858 and drivers of multidimensional energy poverty in Ghana. *Energy Policy*  
859 133, 110884. doi:10.1016/j.enpol.2019.110884.
- 860 Dalla Longa, F., Sweerts, B., van der Zwaan, B., 2021. Exploring the com-  
861 plex origins of energy poverty in The Netherlands with machine learning.  
862 *Energy Policy* 156, 112373. doi:10.1016/j.enpol.2021.112373.
- 863 Decancq, K., Lugo, M.A., 2013. Weights in multidimensional indices of  
864 wellbeing: An overview. *Econometric Reviews* 32, 7–34. doi:10.1080/  
865 07474938.2012.690641.
- 866 Dhongde, S., Pattanaik, P.K., Xu, Y., 2019. Well-being, deprivation, and the  
867 Great Recession in the U.S.: A study in a multidimensional framework.  
868 *Review of Income and Wealth* 65, S281–S306. doi:10.1111/roiw.12411.
- 869 Drescher, K., Janzen, B., 2021. Determinants, persistence, and dynamics  
870 of energy poverty: An empirical assessment using German household sur-  
871 vey data. *Energy Economics* 102, 105433. doi:10.1016/j.eneco.2021.  
872 105433.
- 873 Fry, J.M., Farrell, L., Temple, J.B., 2022. Energy poverty and retirement  
874 income sources in Australia. *Energy Economics* 106, 105793. doi:10.1016/  
875 j.eneco.2021.105793.
- 876 Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F., 2012. A  
877 review on ensembles for the class imbalance problem: Bagging-, boosting-,  
878 and hybrid-based approaches. *IEEE Transactions on Systems, Man, and*  
879 *Cybernetics, Part C (Applications and Reviews)* 42, 463–484. doi:10.  
880 1109/TSMCC.2011.2161285.
- 881 Gawusu, S., Jamatutu, S.A., Ahmed, A., 2024. Predictive modeling of energy  
882 poverty with machine learning ensembles: Strategic insights from socioe-  
883 conomic determinants for effective policy implementation. *International*  
884 *Journal of Energy Research* 2024, 9411326. doi:10.1155/2024/9411326.
- 885 Halkos, G., Kostakis, I., 2023. Exploring the persistence and transience of en-  
886 ergy poverty: Evidence from a Greek household survey. *Energy Efficiency*  
887 16, 50. doi:10.1007/s12053-023-10137-1.

- 888 Hosan, S., Sen, K.K., Rahman, M.M., Karmaker, S.C., Chapman, A.J., Saha,  
889 B.B., 2023. Evaluating the mediating role of energy subsidies on social well-  
890 being and energy poverty alleviation in Bangladesh. *Energy Research &*  
891 *Social Science* 100, 103088. doi:10.1016/j.erss.2023.103088.
- 892 van Hove, W., Dalla Longa, F., van der Zwaan, B., 2022. Identifying pre-  
893 dictors for energy poverty in Europe using machine learning. *Energy and*  
894 *Buildings* 264, 112064. doi:10.1016/j.enbuild.2022.112064.
- 895 Igawa, M., Managi, S., 2022. Energy poverty and income inequality: An  
896 economic analysis of 37 countries. *Applied Energy* 306, Part B, 118076.  
897 doi:10.1016/j.apenergy.2021.118076.
- 898 International Energy Agency, 2024. *World Energy Outlook 2024*. URL:  
899 <https://www.iea.org/reports/world-energy-outlook-2024>. (Ac-  
900 cessed on 10 November 2024).
- 901 Karpinska, L., Śmiech, S., 2020. Invisible energy poverty? Analysing housing  
902 costs in Central and Eastern Europe. *Energy Research & Social Science*  
903 70, 101670. doi:10.1016/j.erss.2020.101670.
- 904 Kocak, E., Ulug, E.E., Oralhan, B., 2023. The impact of electricity from  
905 renewable and non-renewable sources on energy poverty and greenhouse  
906 gas emissions (GHGs): Empirical evidence and policy implications. *Energy*  
907 272, 127125. doi:10.1016/j.energy.2023.127125.
- 908 Koomson, I., Awaworyi Churchill, S., 2022. Employment precarity and en-  
909 ergy poverty in post-apartheid South Africa: Exploring the racial and eth-  
910 nic dimensions. *Energy Economics* 110, 106026. doi:10.1016/j.eneco.  
911 2022.106026.
- 912 Lee, Y.G., Oh, J.Y., Kim, D., Kim, G., 2023. SHAP value-based feature im-  
913 portance analysis for short-term load forecasting. *Journal of Electrical En-*  
914 *gineering & Technology* 18, 579–588. doi:10.1007/s42835-022-01161-9.
- 915 Lin, B., Okyere, M.A., 2021. Does energy poverty affect the well-being of  
916 people: Evidence from Ghana. *Sustainable Production and Consumption*  
917 28, 675–685. doi:10.1016/j.spc.2021.06.031.



- 918 Liu, X.Y., Wu, J., Zhou, Z.H., 2009. Exploratory undersampling for class-  
919 imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics,*  
920 *Part B (Cybernetics)* 39, 539–550. doi:10.1109/TSMCB.2008.2007853.
- 921 Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model  
922 predictions, in: *Proceedings of the 31st International Conference on Neural*  
923 *Information Processing Systems (NeurIPS)*, Curran Associates Inc., Red  
924 Hook, NY, USA. pp. 4768–4777. doi:10.48550/arXiv.1705.07874.
- 925 Mahmud Sujon, K., Binti Hassan, R., Tusnia Towshi, Z., Othman, M.A.,  
926 Abdus Samad, M., Choi, K., 2024. When to use standardization and  
927 normalization: Empirical evidence from machine learning models and XAI.  
928 *IEEE Access* 12, 135300–135314. doi:10.1109/ACCESS.2024.3462434.
- 929 Manasi, B., Mukhopadhyay, J.P., 2024. Definition, measurement and deter-  
930 minants of energy poverty: Empirical evidence from Indian households.  
931 *Energy for Sustainable Development* 79, 101383. doi:10.1016/j.esd.  
932 2024.101383.
- 933 Mokhtari, K.E., Higdon, B.P., Başar, A., 2019. Interpreting financial time  
934 series with SHAP values, in: *Proceedings of the 29th Annual International*  
935 *Conference on Computer Science and Software Engineering (CASCON)*,  
936 IBM Corp., Toronto, Canada. pp. 166–172.
- 937 Murphy, K., 2012. *Machine Learning: A Probabilistic Perspective*. Adap-  
938 *tive Computation and Machine Learning* series, MIT Press, Cambridge,  
939 Massachusetts.
- 940 Nguyen-Phung, H.T., Le, H., 2024. Elderly well-being amidst energy poverty:  
941 Exploring the health, social, and economic impacts in Vietnam. *Energy Re-*  
942 *search & Social Science* 118, 103762. doi:10.1016/j.erss.2024.103762.
- 943 OECD, 2023. *OECD Economic Outlook, Volume 2023 Issue 2*. URL: <https://doi.org/10.1787/7a5f73ce-en>. (Accessed on 1 December 2024).
- 945 Phoumin, H., Kimura, F., 2019. Cambodia’s energy poverty and its effects  
946 on social wellbeing: Empirical evidence and policy implications. *Energy*  
947 *Policy* 132, 283–289. doi:10.1016/j.enpol.2019.05.032.
- 948 Pondie, T.M., Engwali, F.D., Nkoa, B.E.O., Domguia, E.N., 2024. En-  
949 ergy poverty and respiratory health in Sub-Saharan Africa: Effects and

- transmission channels. *Energy* 297, 131158. doi:10.1016/j.energy.2024.131158.
- Prakash, K., Awaworyi Churchill, S., Smyth, R., 2022. Are you puffing your children’s future away? Energy poverty and childhood exposure to passive smoking. *Economic Modelling* 114, 105937. doi:10.1016/j.econmod.2022.105937.
- Primc, K., Dominko, M., Slabe-Erker, R., 2021. 30 years of energy and fuel poverty research: A retrospective analysis and future trends. *Journal of Cleaner Production* 301, 127003. doi:10.1016/j.jclepro.2021.127003.
- Proctol, D., 2022. Coal Use Rises, Prices Soar as War Impacts Energy Markets. URL: <https://www.powermag.com/coal-use-rises-prices-soar-as-war-impacts-energy-markets/>. (Accessed on 2 October 2024).
- Provost, F., 2000. Machine learning from imbalanced data sets 101, in: *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)’2000 Workshop on Imbalanced Data Sets*, AAAI Press, Austin, TX, USA. pp. 1–3.
- Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2010. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics, Part A (Systems and Humans)* 40, 185–197. doi:10.1109/TSMCA.2009.2029559.
- Simshauser, P., 2023. The 2022 energy crisis: Fuel poverty and the impact of policy interventions in Australia’s National Electricity Market. *Energy Economics* 121, 106660. doi:10.1016/j.eneco.2023.106660.
- Simshauser, P., Miller, W., 2023. On the impact of targeted and universal electricity concessions policy on fuel poverty in the NEM’s Queensland region. *Economic Analysis and Policy* 80, 1848–1857. doi:10.1016/j.eap.2023.11.015.
- Spandagos, C., Reaños, M.A.T., Lynch, M.Á., 2023. Energy poverty prediction and effective targeting for just transitions with machine learning. *Energy Economics* 128, 107131. doi:10.1016/j.eneco.2023.107131.

- 981 Sullivan, J.H., Warkentin, M., Wallace, L., 2021. So many ways for assessing  
982 outliers: What really works and does it matter? *Journal of Business*  
983 *Research* 132, 530–543. doi:10.1016/j.jbusres.2021.03.066.
- 984 Sy, S.A., Mokaddem, L., 2022. Energy poverty in developing countries: A  
985 review of the concept and its measurements. *Energy Research & Social*  
986 *Science* 89, 102562. doi:10.1016/j.erss.2022.102562.
- 987 Willand, N., 2022. Opportunity, ideal or distraction? Exploring stakeholder  
988 perceptions of tackling energy poverty and vulnerability among older Aus-  
989 tralians. *Energy Research & Social Science* 94, 102852. doi:10.1016/j.  
990 *erss*.2022.102852.
- 991 Zhang, Q., Appau, S., Kodom, P.L., 2021a. Energy poverty, children’s wellbe-  
992 ing and the mediating role of academic performance: Evidence from China.  
993 *Energy Economics* 97, 105206. doi:10.1016/j.eneco.2021.105206.
- 994 Zhang, Z., Shu, H., Yi, H., Wang, X., 2021b. Household multidimensional  
995 energy poverty and its impacts on physical and mental health. *Energy*  
996 *Policy* 156, 112381. doi:10.1016/j.enpol.2021.112381.