Renewable and Sustainable Energy Reviews 192 (2024) 114199

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Original research article

Geospatial and socioeconomic prediction of value-driven clean cooking uptake

Micaela Flores Lanza, Alycia Leonard [*], Stephanie Hirmer

Department of Engineering Science, University of Oxford, Parks Road, OX1 3PJ, UK

ARTICLE INFO

Keywords:
User-perceived values
LMICs
Clean cooking
QGIS
Value perception
Machine learning

ABSTRACT

Understanding the community-specific values and needs of consumers is essential for effective targeting and planning of energy services such as clean cooking. Many clean cooking programmes do not however consider these values and needs in targeting, as they can be difficult and time-consuming to ascertain. This work therefore explores whether community needs and values related to cooking can be predicted, using a novel approach that understands the relationship between socioeconomic, demographic, and geospatial data. Specifically, this study investigates (i) which values are most closely linked to cookstoves in rural Uganda; and (ii) whether it is possible to predict cookstove prioritisation and related values using openly-available data. Using machine-learning approaches, user-perceived value data from 199 rural low-income households in Uganda are mapped against socioeconomic, demographic, and geospatial data to identify correlations and intersections. The values most closely related to cookstoves were found to be food security, time benefit, accessibility to services, fixed costs, and being healthy. The most important parameters in predicting who would hold these values were found to be: the number of people living in a house; age; quintile 2 of the wealth index; annual accumulated precipitation; forest density; night time luminance; and distance to water source, nearest forest within ten kilometers, and nearest road. This study takes a first step towards enabling energy service providers to target areas with a greater likelihood of uptake based on open-source datasets. While cooking in Uganda is analysed herein, the proposed method can be applied for different geographies and energy services.

1. Introduction

This study investigates why people in rural Uganda value cookstoves and explores which geospatial, socioeconomic, and demographic parameters are predictive of cookstove prioritisation and related values. Considering energy services to be “those functions performed using energy which are means to obtain or facilitate desired end services or states”, [1] the aim is to identify where the energy service of cooking is most needed and valued based on open datasets to improve targeting of clean cooking initiatives.

Approximately 2.6 billion people lack access to clean cooking services [2], with 82% of people in sub-Saharan Africa [3] and 94% in rural areas of this region [4] being affected. This research focuses on Uganda, where approximately 45.5 million people rely on traditional cooking technologies and fuels for cooking, of which 76% reside in rural areas [5].

The use of traditional cooking methods in low- and middle-income countries (LMICs) exacerbates societal problems relating to health, safety, and gender equality. The World Health Organisation estimates that household air pollution caused by inefficient kerosene, biomass, and coal stoves was responsible for around 3.2 million deaths in 2020 [6]. The smoke produced from traditional cooking methods causes strokes, respiratory infections, obstructive pulmonary disease, and lung cancer [6]. The most affected groups are women and children, as women usually prepare food for their families and children help with household chores [7,8]. This smoke also releases greenhouse gases: every year, around one gigaton of carbon dioxide is produced by burning firewood, representing around 2.1% of global emissions [9]. More than half of the black carbon (which is 460–1500 times more polluting than carbon dioxide) generated by human activity comes from burning wood fuels [9]. Wood collection degrades forests and accelerates climate change; an estimated 34% of the firewood collected for cooking is environmentally unsustainable [10]. Women and girls spend around ten hours per week collecting firewood, and walk an average of five kilometres per trip [11]. The journey to collection points far from their communities puts them at risk for physical and sexual violence [12].

* Corresponding author.
E-mail address: alycia.leonard@eng.ox.ac.uk (A. Leonard).

https://doi.org/10.1016/j.rser.2023.114199
Received 30 March 2023; Received in revised form 26 October 2023; Accepted 9 December 2023
Available online 4 January 2024
1364-0321/

Nomenclature

Abbreviations

5TopVals	Top 5 Values
CSV	Comma-Separated Values
DT	Decision Tree
ERA	Electricity Regulatory Authority
FO-DS	Final Observations Dataset
ID	Identification
KNN	K-Nearest Neighbours
LMICs	Low and Middle Income Countries
LPG	Liquefied Petroleum Gas
ML	Machine Learning
PR	Participation Rate
QGIS	Quantum Geographical Information System
RF	Random Forest
SVM	Support Vector Machine
UPV-DS	User-Perceived Values Dataset
UPVs	User-Perceived Values

Notations/Symbols

η	Correlation ratio
$\overline{y_x}$	Mean of category 'x'
$\overline{y}$	Mean of the whole observations
n_x	Number of observations in category 'x
x	Value to be normalised
x'	Normalised value
x_{max}	Maximum value
x_{min}	Minimum value
y_{xi}	Observation to be studied

These harms can be alleviated by replacing traditional stoves with efficient "clean cooking" technologies which use cleaner fuels and allow for better ventilation. By reducing smoke exposure, clean cookstoves can improve quality of life and health. They can help women save time and protect the natural environment. Clean cookstoves are seen by the Global Alliance for Clean Cookstoves to favour ten of the Sustainable Development Goals, contributing to the effective achievement of the 2030 Agenda [13].

While the clean cooking sector has made progress – from 2010 to 2021, 400 million people have gained access to clean cooking fuels and technologies [14] – initiatives are seeing low uptake or only short-term effects [15,16]. A "failure to meet consumers needs" has often been cited as "an impediment to cookstove adoption, especially in those cases where cookstoves were subsidised for the end user but not selected to best meet their needs" [7]. To overcome this challenge, it has been proposed to focus on perceptions, needs, values [17,18], and sociocultural [19] influences to maximise cookstove adoption in rural communities. For instance, [20] found that when local artisans participated in cookstove design and sale, dissemination programmes achieved higher uptake rates, and production could be compatible with consumer needs while stimulating local industry.

It is therefore important to look at how technologies match local values when designing and distributing cookstoves. Local values can be elicited through bottom-up processes such as pairwise ranking, matrix ranking [21], the nominal group technique [22], observations during field work [23], and the user-perceived values (UPV) approach [17]. While these approaches are effective, they can be costly and time-consuming to replicate at scale. In addition, they often involve travelling long distances to insecure places with poor infrastructure where villagers might be culturally hesitant to welcome foreign people [24]. It is therefore important to look for ways to accelerate targeting in less labour-intensive and invasive ways. This is not to say that bottom-up data collection is not needed once target areas have been identified. Rather, it hypothesises that initial targeting can be accelerated using scalable data sources at community-level resolution.

To do this initial targeting, parameters collected in large-scale geo-tagged surveys which have been found to influence cookstove adoption can be investigated. These include economic factors (e.g., family income, cookstove price) and demographics (e.g., gender, age, level of education, household composition) [18]. Geographic and contextual parameters including fuel availability, location (rural or urban) [25], deforestation rates, proximity to forests [26], vegetation density, land cover, luminance, proximity to distribution and transmission substations, proximity to roads [27], and precipitation rates have also been found to be related to the adoption of cookstoves. This is in line with findings that geographic features do improve the accuracy of development program targeting [28].

Computational methods are needed to explore the potential for these data to enable predictive targeting at scale. Fortunately, machine learning (ML) algorithms are designed specifically to identify these prediction patterns in large datasets. They are receiving increasing attention in the development sector, particularly in resource targeting applications [29]. Early explorations of ML in development program targeting show great promise, including in the targeting of aid towards the ultra poor in Afghanistan [30] and COVID-19 relief aid targeting in Togo [31]. So long as fairness and bias prevention are kept top of mind in ML design for developing contexts [32], these tools can powerfully accelerate and development program targeting. While early applications primarily explore ML-driven targeting based on poverty, this study probes whether these methods could equally apply to targeting based on value alignment.

Within this context, the main question of this work is: can the needs and values related to the energy service of cooking be predicted from scalable geospatial, socioeconomic, and demographic data to improve technology deployment and uptake? We hypothesise that this can be accomplished using openly available data and ML methods. To test this hypothesis, this work first develops an approach to investigate which values are most strongly related to cooking through analysis of a case study value-perception dataset. Then, it investigates whether any strong correlations exist between geospatial, demographic, or socioeconomic survey data and these values. Finally, it explores whether the importance of cooking to communities can be reliably predicted using these data and ML methods. This is, to our knowledge, the first time a study intends to comprehend the relationships between geospatial, socioeconomic, and demographic parameters, and the values associated with energy services, particularly cooking.

Through this investigation, this study develops a novel open-access predictive model designed to accelerate energy services targeting. It is implemented to be scalable to other geographic contexts and energy service technologies. It could assist policy-makers, organisations, businesses, and donors in better targeting their energy service initiatives towards people that are more likely to value and adopt them, saving time and money while improving intervention effectiveness. This could support LMICs as they craft clean cooking strategies in line with their policy agendas. It is important to note that this work assumes that need of a service and value of a service drive consumers to adopt it and generate uptake. The model is made openly available online here: https://github.com/Nocandidate/clean_cooking_values_Uganda.

This study is rooted in the case study context of Uganda. Uganda's government shows enthusiasm towards clean cooking, as it works towards achieving Sustainable Development Goal 7 [33] and universal access to modern cooking solutions by 2030 [34]. In December 2021, Uganda's Electricity Regulatory Authority (ERA), introduced a 'cooking tariff' to encourage the replacement of biomass cooking fuel with electricity by making the "cost of electric cooking lower than cooking using charcoal in homes" [35]. Uganda's third National Development Plan incorporates an Energy Development Programme which aims to reduce the use of biomass for cooking in homes to 50% by replacing it with electricity, liquefied petroleum gas (LPG), and biogas [36]. The ERA has piloted a programme called the 'charcoal to power project,' which aims to incentivise 500 institutions and eventually 50,000 households to transition from biomass to electricity for cooking [37]. The Ugandan government also undertakes awareness raising for greater penetration of cleaner cooking appliances in the country [38]. With an existing drive for clean cooking, Uganda is an ideal case study to determine values linked to the energy service of cooking and identify parameters that can aid in better targeting to increase clean cookstove uptake. It is hoped that the Ugandan case study can generate insights applicable in other LMICs, and that the developed method can translate to other energy technologies and services.

This analysis proceeds as follows. Section 2 locates the definition of value adopted in this work amongst existing definitions, and refines this understanding of value in the context of cooking. Section 3 then explains the methodology used to determine which parameters are most correlated and predictive of cookstove prioritisation and related values. The results of all analyses and their discussion are detailed in Sections 4 and 5 respectively. Finally, Section 6 presents the conclusions of this study and future work.

2. Understanding user value of cookstoves

2.1. Energy service values

There are many definitions of 'value', ranging from the notion of 'utility' in economics [39] to moral connotations in philosophy [40]. However, in sociology and marketing, 'values' are connected to a 'user' as "culturally approved, internalised wishes that motivate our actions [...] inspire our beliefs and attitudes and determine what we strive for" [41], and the "consumer's assessment of the ability of a product or service to meet their needs" [42] respectively. This can be seen as an applied collective version of the psychological definition of value as "internalised cognitive structures that guide choices by evoking a sense of basic principles of right and wrong, a sense of priorities, and a willingness to make meaning and see patterns" [43]; or as [44] defines it, "desirable trans-situational goals, varying in importance, that serve as guiding principles in the life of a person or other social entity".

This work uses Hirmer's concept of User-Perceived Values (UPVs) to understand value in the delivery of energy services: "the benefits, concerns, feelings and underlying drivers that vary in importance and act as the main motivators in the lives of the people — as perceived and defined by the [people] themselves at a given time" [17]. Considering this definition in concert with those previously discussed, and noting that "individual values have a significant impact on consumers' inclinations to adopt new products", [45], it becomes clear that understanding what is important (valued) for people on the ground is necessary to ensure that a product or service can address user needs and values effectively. As [46] put it: "different stakeholders may have conflicting views on the success of a delivery model; in other words, local communities may have very different priorities from businesses, donors or government". Understanding local priorities and values is critical to the delivery of energy services such as cooking.

UPVs can be determined through the 'UPV game', a method proposed by Hirmer which uncovers what is important to people without asking them directly [17]. This game "bypasses the interviewees' predispositions and preconceptions by redirecting the focus of the participants to the game itself". It is based on methods used in product design and market research and can be played by groups of people with different backgrounds and abilities. The game requires participants to select items they value most from a locally-appropriate visual list and then explain why they value each one. Their reasoning is probed through multiple layers of 'why' questioning. Their speech is collected verbatim and annotated against the UPV framework of 56 values (i.e., the values related to the reasoning are annotated against the text snippet) to extract meaning. For more information on this process, refer to [47]. A socioeconomic survey is undertaken alongside the game to understand people's perceptions within the larger context of their lives [48].

2.2. Cooking and values

There are several value-driven reasons why households may hesitate to transition away from traditional cooking. These include high upfront cookstove costs, safety concerns, and impacts on the taste of food. Additionally, households may have reservations about investing time to learn how to use a clean-cooking appliance [38,49] Clean-cooking appliances also do not perform all of the same functions as traditional cookstoves (e.g., heating on cold days [50]). Lastly, improved cookstoves that are currently available on the market may not be suitable for large families: modern stoves tend to be too small too cook the amount of food needed to feed a large family [38].

There is also however evidence of value-driven clean cookstove adoption. Vigolo's [18] thematic analysis of consumer perceptions of improved cookstoves identified four main categories: convenience and uses, aesthetics, health-related impacts, and environmental impacts. Aesthetics are especially valued by cookstove users exposed to mass media and located in urban and peri-urban areas [51]. For instance, in Zimbabwe, women tended to value the 'aesthetics' of improved cookstoves and thus the 'social status' involved in having an improved cookstove [52]. Similarly, another study in Kenya mentioned that the appearance of improved cookstoves pleased them as they were 'stylish,' 'well-designed,' and 'beautiful' [53]. As previously discussed, health-related impacts of clean cookstoves include better health and a reduction in child mortality rates, especially among women and young girls. Finally, environmental impacts, though highlighted by Vigolo [18], are rarely mentioned across studies in the field of user perceptions and values. Reference is made to environmental sustainability as an important factor when deciding to acquire clean cookstoves as "less smoke means less respiratory infections for mothers and children" and "[the] use of different fuels will help the environment in the community" [54]. Others mentioned that, although people knew about the negative environmental effects of traditional cookstoves, this was not a determining factor in their decision-making due to "a poverty-related short term planning horizon" [55].

The cookstove-related values (e.g., 'time-benefit, 'safety', 'aesthetics', 'social status') identified in previous research (from e.g., [18,52]) align with the values annotated in Hirmer's UPV dataset [17]. For instance, the following phrases were recorded in discussions on the item 'stove' (where annotated values are listed in brackets following each quote): "*I promise, if I had a stove my husband will be eating on time thus saving time on cooking*" (convenience), "*it also produces less smoke which is important for our environment*" (environmental impact), and "*even the gas cooker is fast it will help me eat my food in time which is health and even eases my work*" (convenience and health benefits). The consistency between the existing body of research on this topic and the UPV data is encouraging. It suggests a significant cookstove-related value patterns that could be indicative of a higher potential for cooking uptake. Hereafter, this study aims to develop and test methods to quantify this observation and predict values for improved targeting of clean cooking interventions.

2.3. Spatial and demographic influences

The value of a cookstove to a family is also known to be linked to their socioeconomic status and spatial context. With regard to socioeconomic factors, the systematic review on the consumer behaviour elaborated by [18] concluded that economic factors (e.g., income, appliance upfront cost) are the most important elements influencing the adoption of improved cookstoves. Higher-income households tend to acquire cookstoves more frequently than lower-income households. The higher the upfront cost of a cookstove, the lower probability for a household to purchase it, especially in rural areas of LMICs [56].

Demographics such as gender, age, education, and household size have been highly studied as determinants of cookstove adoption. However, there is no general agreement about how they affect consumer choices. Despite women being more likely to use cleaner cookstoves than men, as the primary users of cooking appliances in many developing society households [55], they usually do not have the economic power and authority to decide to purchase one [57]. Education may also play an important role, as it is generally agreed that more educated people are more likely to adopt cleaner forms of cooking. This is supported by a recent study in Pakistan that demonstrates a positive effect of primary or secondary education level on the appropriation of improved cookstoves [58]. Some research has found that younger people are more likely to use sustainable cooking systems. For example, [59] identified a negative effect between a married woman's age and the rate of improved cookstoves adoption: the older the woman, the greater the intention to keep with traditional cookstoves. However, looking at UPV-DS, similar proportions of the oldest people (from 74 to 80 years old) and younger people (from 20 to 34) prioritised cooking [17,60], conflicting with the previous research. The number of people living in a household might also have a negative effect on improved cookstove adoption because, as previously discussed, they could be too small to feed a large family [50].

Some key geospatial features influencing cookstove adoption found throughout the corpus of research in this area are proximity to forests, density of forests, and land cover type. For instance, [61] suggested that people with their own forest resources around them show more interest in obtaining cleaner cookstoves because they want to conserve their natural heritage. However, other studies found accessibility to open forests is negatively correlated with improved cookstove acquisition [62]; that is, the closer a person is to a forest, the less likely they are to acquire one. A household's location (urban or rural) may also affect cookstove adoption. A study among cookstove consumers in Malawi, Mali, South Africa, and Mozambique, found that households in rural areas are less likely to pay for an improved cookstove than urban consumers [51]. It also seems that people from rural areas are less interested in adopting improved cookstoves because they have easy access to firewood [63].

Deforestation rates also seem to have an impact on improved cookstove adoption — higher rates of deforestation might indicate extensive use of firewood [27]. Proximity to roads (paved or unpaved) could also influence the adoption of stoves since accessibility of roads affects supply chains and ergo stove cost. Fuel availability is another geospatial factor, as the location of petrol stations and LPG shops determine how easily a person can get this fuel to keep using a cookstove [64]. It has been found that "access to infrastructure and location are important determinants for the use of electricity for [...] cooking" [65]. As such, the proximity to distribution and transmission substations might also be a geospatial determinant for adopting clean cooking systems. Nightlight data (i.e., luminance at night) might indicate access to electricity and, therefore, the potential for clean cooking systems adoption [66]. Finally, the distance or time between a person's location and the water collection source is likely to influence the decision to adopt cleaner cooking technologies [67].

3. Methods

This section outlines the methods employed to investigate:

(i) which values are most closely linked to cookstoves in a specified study context; and
(ii) whether it is possible to predict cookstove prioritisation and related values based on geospatial, socioeconomic, and demographic parameters.

Statistical and machine learning techniques are used to analyse UPV data, identify key parameters correlating to cookstove prioritisation and related values, and develop a predictive model. Fig. 1 summarises this approach.

3.1. UPV dataset

The baseline UPV dataset used in this analysis (hereafter UPV-DS) was collected by Hirmer in rural Uganda and by Efficiency for Access [68]. UPV-DS was produced using the UPV game described in Section 2. The variables in UPV-DS are 'Interview ID,' 'Text,' 'Item,' and 'Annotations.' Each interviewee (Interview ID) selected 20 out of 40 items based on what was most important to them (Item). Subsequently, they ranked them in order of importance and then gave reasons for choosing that particular item (Text); these reasons were divided into sentences that were tabulated in different cells. Finally, the value data (Annotations) were determined by a large group of experts in the area.

3.2. Selecting parameters for correlation analysis and predictive modelling

UPV-DS is analysed alongside geospatial, socioeconomic, and demographic data to determine value correlations and build a predictive model. As many demographic and geographic parameters will have no relation to cookstoves, the most relevant parameters to investigate were identified based on a literature review. This can be understood as feature engineering for subsequent predictive modelling. From the review conducted in Section 2.3, the following parameters are identified as key to explore in terms of relationship to cookstove adoption:

- **Demographics**: Age, sex, number of people in the household, level of education;
- **Socioeconomic factors**: Main type of fuel used for cooking in the household, source of water used in the household, type of toilet commonly used;
- **Distance to closest**: Location to fetch water, river, lake, fuel station, forest, smallest forest in a fixed land radius, distribution and transmission substations, national and subnational roads, city (travel time);
- **Other geospatial factors**: Density of the forest, grade of deforestation, land cover, luminance at night, and precipitation levels.

These parameters are hereafter referred to as the independent variables. Table 1 lists the sources from which the data on these variables were collected. Note that the survey undertaken alongside the UPV game is used as a data source. However, similar data can be extracted from a number of national survey microdata sources to replicate this process using only open data. It is important to note that the base year of this study is 2016 based on the UPV-DS collection dates.

3.3. Data preparation and pre-processing

Data pre-processing was necessary to obtain a cleaned final observations dataset (hereafter FO-DS), wherein UPV-DS observations (i.e., value-annotated text snippets) are attached to location-specific independent variables. Pre-processing is undertaken using Python and Quantum Geographic Information System (QGIS) software.

Datasets containing independent variables are extracted and appended to the UPV-DS observations considering the geographic coordinates of each observation (i.e., where each UPV game took place). Table B.9 in Appendix B describes the independent variables, their keys, and their types in detail. To handle outliers, the box-plot method (numeric outlier) was used, which is a simple non-parametric outlier detection method using the interquartile range [82]. To handle missing values, the mean imputation method [83] was used, as only the values of two interviewees' ages were missing.

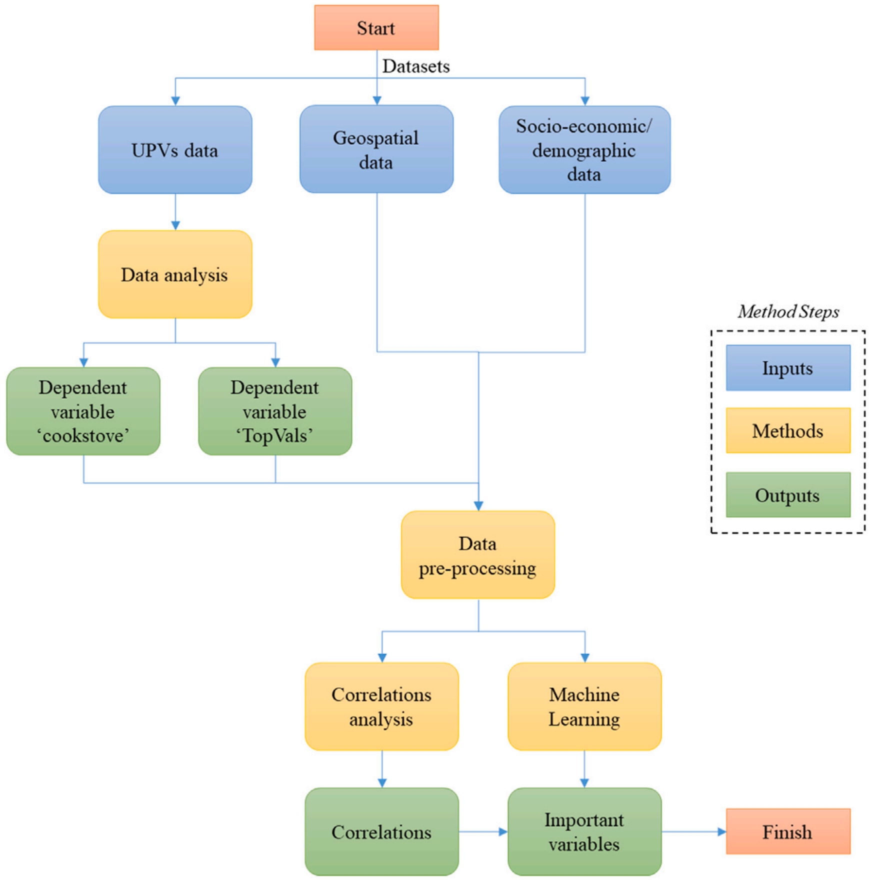

Fig. 1. Method structure for the analysis of UPV data and subsequent predictive modelling.

Numeric data were normalised to avoid biases in the predictive modelling produced by very large or small values. The min–max scaling method is used for this, defined as:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x' is the normalised value, x is the value to be normalised, x_{min} is the minimum value, and x_{max} is the maximum value.

Two additional columns are then added to Fo-DS for analysis. First, FO-DS is filtered by 'Interview ID' and 'Item' to select only observations from individuals who chose a 'Stove' as an important item. These observations are assigned a value of 1 in a new categorical dependent variable called 'cookstove'. All other observations are assigned a value of 0. Second, a new categorical dependent variable called 'TopVals' is added to FO-DS to represent those who hold the top five values related to the item stove. To do this, the values associated with the reasons that people gave when choosing a 'Stove' (i.e., the 'Annotations' for observations where 'Item' = 'Stove') are extracted from FO-DS. Then, a count is made of all the repeated values to obtain a frequency. By ranking these frequencies, the values most related to the selection of 'Stove' as a priority 'Item' are obtained. Values that represent more than 9.5% of the total counted values are noted and collectively named '5TopVals'. A value '1' is assigned to observations in the new column 'TopVals' for both individuals who had chosen 'Stove' as one of their priority items, and individuals who did not choose 'Stove' as a priority item but who hold the '5TopVals' values. All other observations are assigned the value '0'.

3.4. Correlation analysis

Correlation analysis is undertaken to understand any relationships between variables in FO-DS. FO-DS contains both numerical and categorical (i.e., nominal or ordinal) data. Table 2 outlines the variables belonging to each type. For the categorical variables, one-hot encoding was used.

To estimate correlation among pairs of numerical data, Pearson's correlation coefficient is used with a statistical significance level (p-value) of 0.05. This coefficient has a range between −1 and 1, with 1 meaning that the relationship between a pair is positively correlated and −1 that it is negatively correlated. To estimate the strength of correlation among categorical data, the Cramer's V coefficient is used. This coefficient has a range between 0 (no association) and 1 (strong association). Again, a threshold of a p-value of 0.05 is considered. Finally, if a correlation between numerical and categorical data is desired, the correlation ratio (η) is used:

$$\eta = \sqrt{\frac{\sum_x n_x(\overline{y_x} - \overline{y})^2}{\sum_{x,i}(y_{xi} - \overline{y})^2}} \tag{2}$$

where n_x is the number of observations in category 'x', y_{xi} is each observation, $\overline{y_x}$ is the mean of category 'x', and $\overline{y}$ is the mean of the whole population [84]. The value of η can vary between 0 and 1. When a category cannot be determined by a continuous measurement, η is 0 (no association), and if it can, η is 1 (strong association).

Table 1
The geospatial, socioeconomic, and demographic datasets containing independent variables analysed for correlations with cookstove prioritisation and related values as drawn from UPV data.

Dataset	Description	Availability	Source
Demographic and Health Surveys	Socioeconomic/demographic data from more than 90 countries.	Available upon request, with institutional email addresses.	United States Agency for International Development [69]
Socioeconomic & demographic data	CSV file containing personal information, water-related data, and energy-related data, among other data.	Restricted for the use of this study.	Hirmer [17], [60]
Precipitation	Raster map of annual accumulated precipitation measured in (mm) with spatial resolution of 0.1°	Publicly available, worldwide scope.	National Aeronautics and Space Administration [70]
Luminance	Raster map of Night lights in colour scale (3 bands) with a resolution of 500 m × 500 m and a value range from 3 to 255.	Publicly available, worldwide scope.	National Aeronautics and Space Administration [71]
Deforestation	Raster map of forest loss from 2000 to 2021 with resolution of 30 m × 30 m and values range from 0 to 20.	Publicly available, worldwide scope.	Hansen et al. [72]
Land Cover	Raster map of 23 types of land fractions with a resolution of 100 m × 100 m.	Publicly available, worldwide scope.	Buchhorn et al. [73]
Forest Density	Raster map of the 'forest' type with a resolution of 100 m × 100 m and with a range from 0 to 100.	Publicly available, worldwide scope.	Buchhorn et al. [73]
Travel Time	Raster map of time in minutes to access cities of more than 50,000 inhabitants with a resolution of 1 km × 1 km.	Publicly available, worldwide scope.	Malaria Atlas Project [74]
Protected Areas	Vector map depicting the aerial extension of protected areas.	Publicly available, worldwide scope.	United Nations Environment Program - World Conservation Monitoring Centre [75]
Fuel Station	Vector map showing the location of fuel stations.	Publicly available, worldwide scope.	OpenStreetMap [76]
Substations	Vector map describing the location of distribution and transmission substations.	Publicly available.	ENERGYDATA.INFO [77], [78]
Roads	Vector map describing de-sealed and un-sealed (national and subnational) roads.	Publicly available, Uganda only.	Data.ug [79]
Rivers	Vector map describing the rivers or streams in the country.	Publicly available.	ENERGYDATA.INFO [80]
Lakes	Vector map describing the lakes present in the country.	Publicly available.	ENERGYDATA.INFO [81]

Table 2
Variable types considered in this analysis.

Data type	Variables
Numerical (continuous)	Age; Household size; Distance to the closest: river, lake, fuel station, forest, smallest forest within a determined radius, distribution substation, transmission substation, road (national or subnational), road (national); Density of forests (including and without reserves); Grade of deforestation; Luminance; Travel Time; Accumulated Precipitation; Wealth Indices: Quintile 1, 2, 3, 4, 5.
Categorical (nominal or ordinal)	Sex; Education level; Distance to fetch water; Main type of fuel used for cooking; Source of water; Type of toilet used; Land Cover Type

3.4.1. Predictive modelling

Various ML algorithms are available which can perform prediction of 'cookstove' or 'TopVals' based on independent variables. Four ML algorithms – namely, Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbours (KNN) – are therefore trained and tested on FO-DS to identify the highest performing algorithm for this task. The best performing algorithm is taken forward in subsequent analysis.

For each algorithm, three rounds of ten-fold cross-validation are applied with FO-DS randomly split into 80% training and 20% testing data. Each algorithm is used with its default settings in the sci-kit learn implementation. At the end of each cross-validation round, the performance indicators of accuracy, balanced accuracy, and F1 score are calculated. These indicators are based on the elements of the confusion matrix and the foundational indicators of precision, recall, sensitivity, and specificity [85]. The performance indicators are averaged across the three rounds of cross-validation. These averages are compared, and the best-performing algorithm is identified.

After selecting the algorithm with the best performance, several analyses are carried out to see which independent variables are most important in making predictions. Twelve basic analyses are employed. These use six different subsets of the independent variables for prediction of each of the two elements to predict (i.e., 'cookstove' and 'TopVals'). The variable subsets are based on data type (i.e., geospatial, socioeconomic/demogarphic, or all) and correlation (i.e., most correlated to the variable being predicted or all). The analyses are described in Table 3.

Ideally, when using all of the independent variables in the predictive model, the most important ones should match those that are most highly correlated with the predictor variables ('cookstove' or 'TopVals') in the correlation matrices. Careful analysis is necessary in case of the contrary. Ultimately, in this predictive model stage, the independent variables that could have the greatest impact on an individual's decision to prioritise the adoption of a cookstove or the acquisition of certain UPVs related to the prioritisation of cookstoves are obtained.

4. Results

Through the initial analysis of FO-DS, it was discovered that only 26 out of 199 people interviewed (13%) considered a 'stove' in their

Table 3
Definitions of the 12 core predictive analyses conducted.

Analysis	Predicting	Variables considered
1	TopVals	All independent variables
2	TopVals	Independent variables most correlated to Topvals
3	TopVals	Geospatial independent variables only
4	TopVals	Geospatial independent variables most correlated to Topvals
5	TopVals	Socioeconomic and demographic independent variables only
6	TopVals	Socioeconomic and demographic independent variables most correlated to Topvals
7	cookstove	All independent variables
8	cookstove	Independent variables most correlated to cookstove
9	cookstove	Geospatial independent variables only
10	cookstove	Geospatial independent variables most correlated to cookstove
11	cookstove	Socioeconomic and demographic independent variables only
12	cookstove	Socioeconomic and demographic independent variables most correlated to cookstove

Table 4
Top five values related to prioritisation of cookstoves in descending frequency. Definitions from [17].

Value	Definition	Count	Percent
Food security	A reliable and continuous supply of a diverse variety of foods.	16	15.2%
Time benefit	Being able to accomplish a task with the least waste of time or minimum expenditure of time.	15	14.3%
Accessibility to services	The ability to access certain services and regions as well as the ability to be easily reached. It holds five value categories: access to area, banking access, mobile phone access, mobility and transportation, water, energy, and health.	14	13.3%
Fixed costs	Fixed one-time expenditure incurred through the purchase of an item or service.	11	10.5%
Being healthy	Practices performed to prevent illness or injury.	10	9.5%

five most important items. Further analysis was conducted to understand the reasons behind their selection. The top five values that most commonly influenced their decision to select a 'Stove' were 'food security,' 'time benefit,' 'accessibility to services,' 'fixed costs,' and 'being healthy.' These values are defined as '5TopVals' and Table 4 lists them and contains their definitions. As mentioned in Section 3, these five values were selected as they represent more than 9.5% of the total counted values.

Individuals who expressed all '5TopVals' values but did not select a 'Stove' as important were next identified. There were 116 individuals in this group. These observations were combined with observations from those individuals who had chosen a stove. This resulted in 142 individuals (out of the 199 interviewed) who were identified as likely to desire a cookstove based on their priority items or values. Their observations were assigned a value of '1' for variable 'TopVals' as explained in Section 3.

4.1. Correlations

An analysis was undertaken to determine the independent variables which correlate with 'cookstove' and 'TopVals'. Table C.10 in Appendix C shows the full results set using the correlation ratio (η). The Pearson's coefficients (between numeric variables), Cramer's V coefficients (between categorical variables), and η matrices can be seen in Figs. D.9–D.11 in Appendix D. As η can vary between 0 and 1, the correlations are generally low, with the highest at 0.27. To identify the most important correlations, here, values with $\eta > 0.1$ are considered. Tables 5 and 6 present these correlations separately for 'TopVals' and 'cookstove'. Common variables in these tables are the density of forest around a point ('forest'), density around a point without considering protected areas ('forest_no_res'), accumulated annual precipitation ('precipitation'), distance to collect water between 500 and 1000 meters ('Dist_500-1000m'), and river or stream as a water collection source ('Water_river_stream').

Table 5
Most correlated independent variables with 'TopVals' and their participation rates (PR) in the predictive analyses. Variables shaded in green are also correlated with 'cookstove' (see Table 6). PR values shaded in darker blue show the variables that participate the most in the analyses (PR ≥ 75%). PR values shaded in lighter blue show the variables with a medium-high participation rate in the analyses (50% ≤ PR < 75%).

Variable	η with TopVals	PR
precipitation	0.25	80%
Dist_500-1000m	0.17	50%
g2	0.15	100%
members_house	0.14	100%
Water_river-stream	0.14	33%
small_forest	0.14	60%
nat_sub_road_dist	0.14	40%
forest_no_res	0.13	0%
luminance	0.13	20%
forest	0.12	20%
close_forest	0.1	20%
age	0.1	100%

4.2. Prediction

The RF, DT, SVM, and KNN ML methods were evaluated for their performance on this prediction problem. Using all the variables in FO-DS (i.e., analyses 1 and 7 in Table 3), the best algorithm for predicting both 'TopVals' and 'cookstove' is RF. Figs. 2 and 3 compare the performance indicators of the different algorithms for predicting 'TopVals' and 'cookstove' respectively, while Tables 7 and 8 show the numerical results. Note that Fig. 3 and Table 8 do not include the F1 score since the values obtained were very low, which did not allow a perceptible differentiation in magnitude.

4.3. Variable importance

Using the RF algorithm, the twelve analyses outlined in Table 3 are undertaken to determine which variables are most important in predicting 'TopVals' and 'cookstove'. First, all independent variables

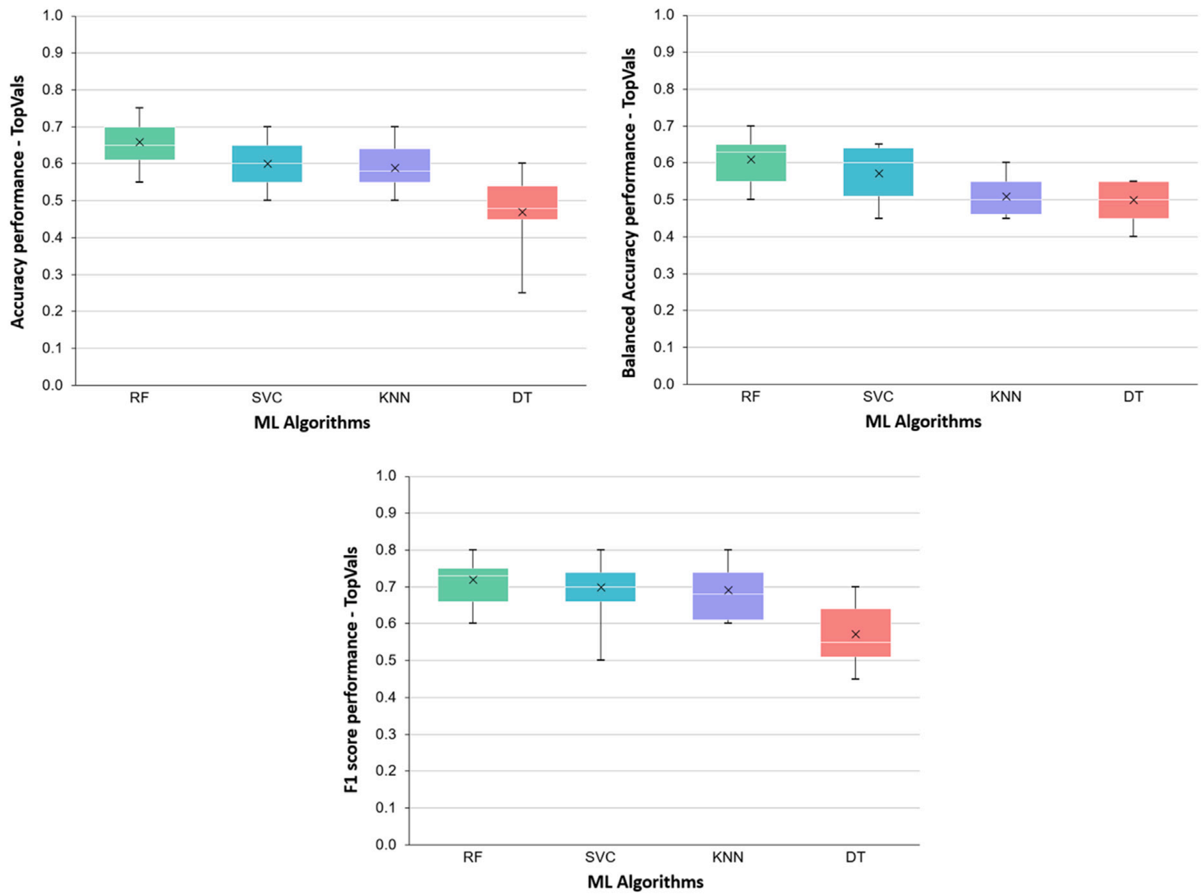

Fig. 2. Performance of each ML algorithm in predicting if a person has the five top values related to cookstoves ('TopVals'). The location of 'x' represents the mean and the white line represents the median.

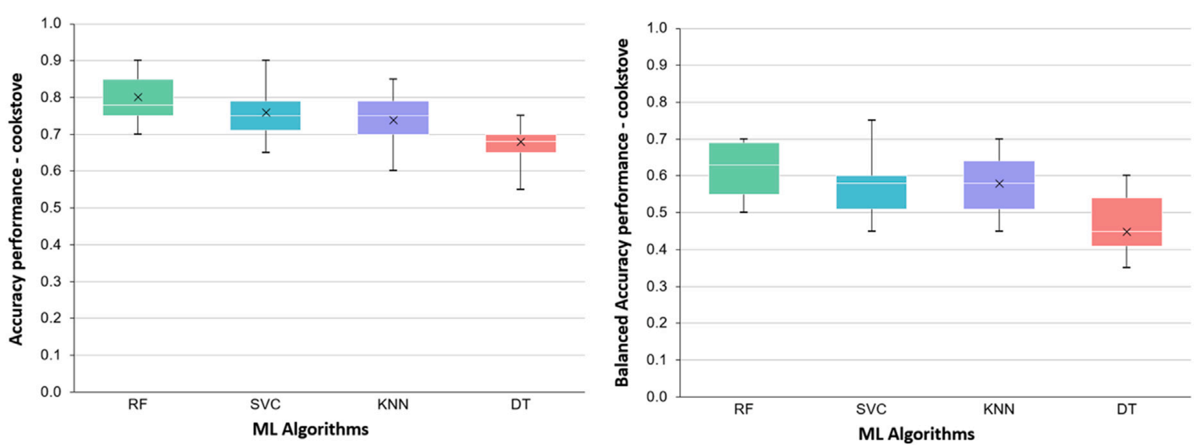

Fig. 3. Performance of each ML algorithm in predicting if a person prioritises a cookstove in their top five items ('cookstove'). The location of 'x' represents the mean and the white line represents the median.

were considered in prediction (i.e., analyses 1 and 7). After three rounds of 10-fold cross-validation, the average results for predicting 'TopVals' in analysis 1 show an accuracy of 0.66, a balanced accuracy of 0.55, and an F1 score of 0.77. The most relevant variable for predicting 'TopVals' for analysis 1 is 'age', with a participation rate of 12%, as shown in Fig. 4. The results of analysis 7 for predicting 'cookstove' show an accuracy of 0.84, a balanced accuracy of 0.51, and an F1 score of 0.07 with, again, 'age' as the most relevant variable in the prediction, with a 13% in importance.

The relevance of 'age' in the predictions is surprising. In Table 5, it can be seen that the correlation that 'age' has with 'TopVals' is 0.1; meanwhile, 'age' is not included in the group of variables most correlated with the variable 'cookstove' as its η is less than 0.1 (Table 6). As such, a deeper analysis was undertaken.

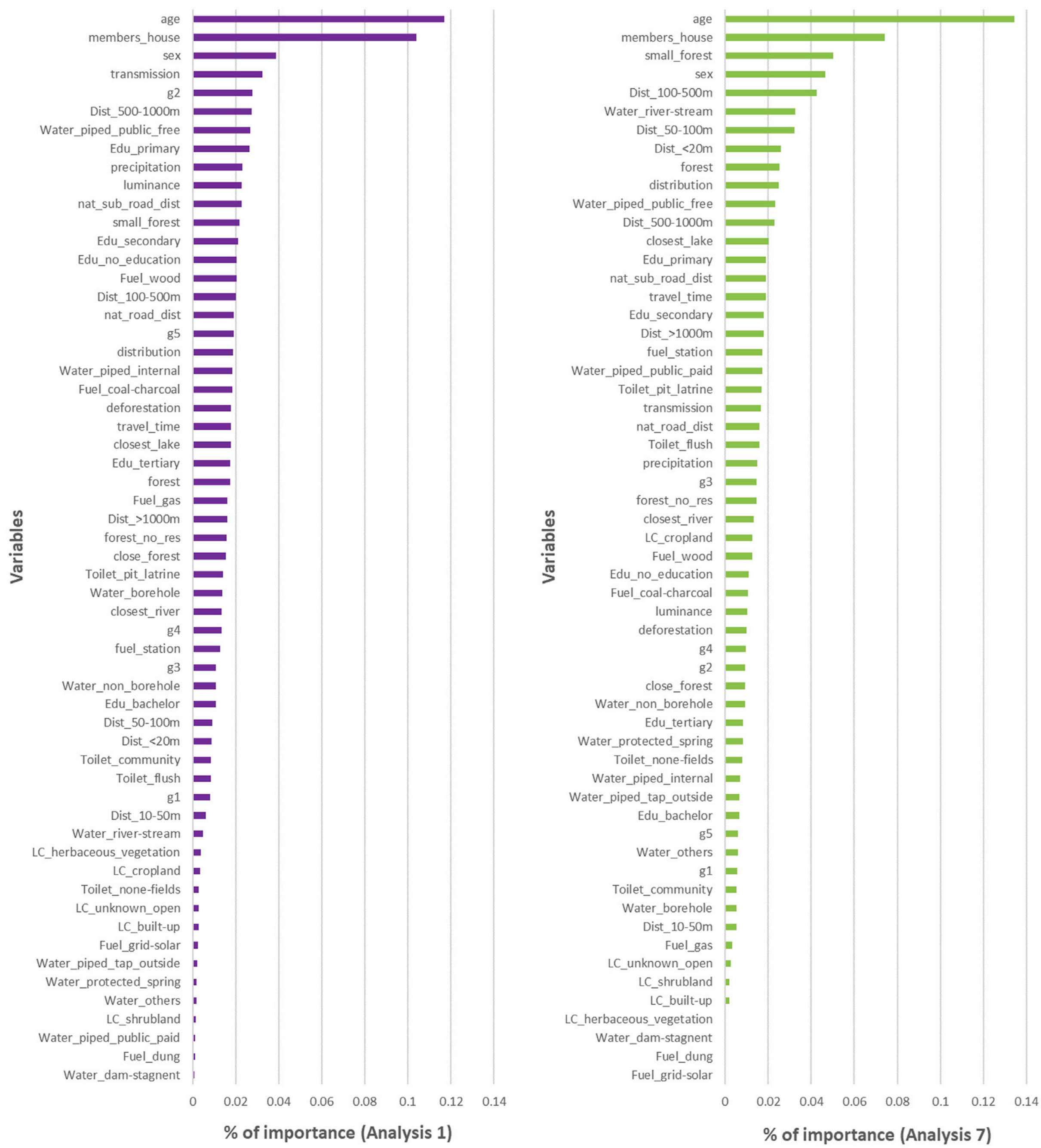

Fig. 4. Variable importance for the prediction of 'TopVals' (analysis 1) and 'cookstove' (analysis 7).

To determine the nature of the relationship between 'age' and 'cookstove', the value of the overlap coefficient between normalised bell distributions of these variables was calculated. Figs. 5 and 6 show histograms of normalised distributions were created considering 10-year age bins for the entire FO-DS, for those in FO-DS with 'cookstove' = 1, and for those in FO-DS with 'cookstove' = 0. The resulting overlap coefficient is 62%. According to [86], an overlap coefficient of more than 90% would indicate that the normalised distributions are strongly similar, meaning an insignificant relationship. Therefore, an overlap coefficient of 62% indeed suggests that there is a significant relationship. However, looking at Fig. 6, the most noticeable difference is between the 70–80 age range, which may be the reason that 'age' correlates somewhat with 'cookstove' and 'TopVals.' This result is supported by other studies in this field [67].

Two analyses were therefore added to avoid biased results due to the apparent influence of 'age' in the predictions, and to create a model that fits the rest of the variables. Analysis 2.2 considers all independent variables that seem to be more correlated with 'TopVals' (not including age), and analysis 6.2 considers only the socioeconomic/demographic independent variables that seem to be more correlated with 'TopVals' (not including age). Details on the results of all analyses are provided in Appendix A.

Table 6
Most correlated independent variables with 'cookstove' and their participation rates (PR) in the predictive analyses. Variables shaded in green are also correlated with 'TopVals' (see Table 5). PR values shaded in darker blue show the variables that participate the most in the analyses (PR ≥ 75%). PR values shaded in lighter blue show the variables with a medium-high participation rate in the analyses (50% ≤ PR < 75%).

Variable	η with cookstove	PR
Dist_100-500m	0.27	75%
LC_cropland	0.23	0%
forest	0.18	50%
Dist_>1000m	0.18	25%
Dist_50-100m	0.15	25%
Dist_500-1000m	0.15	50%
travel_time	0.15	25%
precipitation	0.14	25%
distribution	0.14	25%
Water_river-stream	0.13	50%
LC_built-up	0.13	0%
Water_piped_internal	0.11	0%
Water_piped_public_paid	0.11	25%
transmission	0.11	0%
forest_no_res	0.11	50%

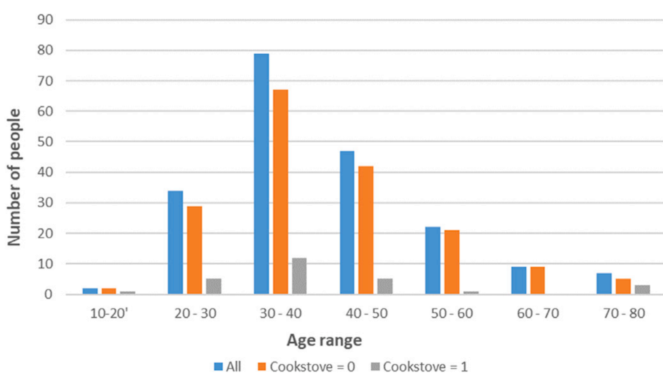

Fig. 5. Distribution of ages for people that chose and did not choose a cookstove.

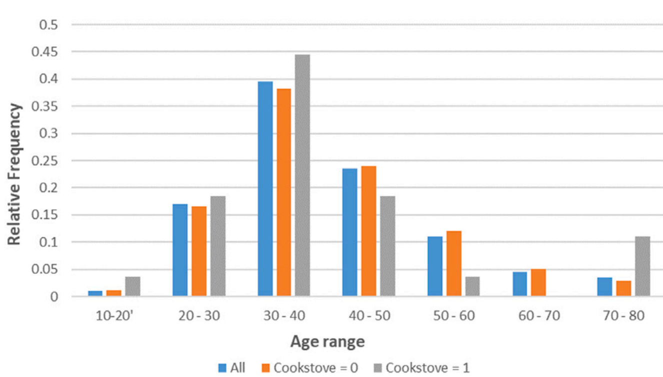

Fig. 6. Normalised distribution for people that chose and did not choose a cookstove.

The participation of each of the most correlated variables in the predictive analyses was determined, considering the top five most important variables. To illustrate, if 'small_forest' was included in five analyses but was only present in the top five most important variables in three analyses, its participation was 60%. Table 5 shows that the variables with the highest participation in predicting 'TopVals' were 'g2', 'age', 'members_house', 'precipitation', 'small_forest', and 'Dist_500-1000m'. Thus, these variables could be the most influential in the predictions. Definitions of these variables can be seen in Appendix B.9.

5. Discussion

These results allow a rich discussion on how geospatial, demographic, and socioeconomic factors may influence the prioritisation of cookstoves at household level, the values related to this, and how this can influence targeting of clean cooking programmes.

5.1. Values associated to cookstove prioritisation

Several of the values found to be related to cookstove prioritisation in this analysis agree with previous studies. Obtaining 'food security' as a top value confirms the conclusions of [87,88], which agree that villagers find it easier to reheat prepared foods or boil liquids with a cookstove, making them appreciate food security. Interviewees who chose 'Stove' as an important item in the Hirmer UPV dataset [17] highlighted the importance of food in phrases such as *"this is used to cook food to overcome hunger"* or *"it helps us to cook our food in homes so as not to starve"*. The value 'time benefit' being related to cookstoves also corroborates previous findings, since having a cookstove can save time for other activities. For example, in surveys conducted in national parks in western Uganda, it was found that the most significant perceived advantages of improved cookstoves were cooking more than one food simultaneously (perhaps implying fuel stacking) and cooking quickly [89]. Obtaining 'fixed costs' as a value related to cookstoves supports several findings from previous studies [56,67]. Finally, 'being healthy' is not a surprising top value. Numerous studies mention that the smoke emitted by burning firewood has harmful effects on health [6]. This is reflected in observations in Hirmer's UPV dataset [17]: e.g., *"This helps us to get rid of smoke caused by firewood because charcoal does not emit smoke hence keeping health"*, *"When I have a gas cooker in my house [...] I will not suffer from smoke like that from the firewood"*.

Other values found here to be related to cookstove prioritisation are not directly supported by existing research. For instance, there is minimal explicit support 'accessibility to services' as a value related to cookstoves. Still, [57] concluded that "access to electricity and to modern and clean cooking facilities [. . .] may contribute to improvements in several domains of development". In other words, if people have access to modern clean cooking facilities, it might mean that they could live in more economically adequate conditions, since owning a cookstove would generate greater economic productivity (possibly due to maybe time-saving). Additionally, the reasons associated with the value 'fixed costs' of people who chose 'Stove' in Hirmer UPV dataset [17] do not explicitly describe the cost of the cookstove as a barrier. They mention that firewood is *"scarce and expensive"*, so it could be implied that the cost of maintaining a traditional cooking technology would be high. Therefore, this research also relates the value 'fixed costs' with another value proposed by [17]: 'continuous costs.' Indeed, the cost of firewood and charcoal has been increasing in recent years in Sub-Saharan Africa due to declining forests and high urban demand [90], so it could be evident that people who prefer a cookstove value a cheaper maintenance cost.

5.2. Variables most predictive of cookstove prioritisation and related values

The most important parameters in predicting 'cookstove' and 'TopVals' were found to be: age; household size; quintile 2 of the wealth index; precipitation; forest density; night time luminance; and distance to water source, nearest forest within ten kilometres, and nearest road. These are discussed hereafter with an emphasis on the results of prediction for 'TopVals'.

- **Age**: 'Age' is the most important variable in all the analyses in which it was included, even though its η values with 'TopVals' and 'cookstove' were low. The normalisation distributions in Fig. 6 suggest that people between 70 and 80 years old tend to prioritise

Table 7
Numeric results used to create the box-plots for the comparison of the models' performance to predict if a person can have the 5 top values related to cookstoves ('TopVals'). It can be seen that Random Forest (RF) is the model best suited for the data considered in this work.

	Accuracy				Balanced Accuracy				F1 score			
	RF	SVC	KNN	DT	RF	SVC	KNN	DT	RF	SVC	KNN	DT
Mean	0.66	0.60	0.59	0.47	0.61	0.57	0.51	0.50	0.72	0.70	0.69	0.57
Min	0.55	0.50	0.50	0.25	0.50	0.45	0.45	0.40	0.60	0.50	0.60	0.45
Q1	0.61	0.55	0.55	0.45	0.55	0.51	0.46	0.45	0.66	0.66	0.61	0.51
Med	0.65	0.60	0.58	0.48	0.63	0.60	0.50	0.50	0.73	0.70	0.68	0.55
Q3	0.70	0.65	0.64	0.54	0.65	0.64	0.55	0.55	0.75	0.74	0.74	0.64
Max	0.75	0.70	0.70	0.60	0.70	0.65	0.60	0.55	0.80	0.80	0.80	0.70

Table 8
Numeric results used to create the box-plots for the comparison of the models' performance to predict if a person can appreciate a cookstove more than other appliances ('cookstove'). It can be seen that Random Forest (RF) is the model best suited for the data considered in this work.

	Accuracy				Balanced Accuracy			
	RF	SVC	KNN	DT	RF	SVC	KNN	DT
Mean	0.80	0.76	0.74	0.68	0.61	0.58	0.58	0.45
Min	0.70	0.65	0.60	0.55	0.50	0.45	0.45	0.35
Q1	0.75	0.71	0.70	0.65	0.55	0.51	0.51	0.41
Med	0.78	0.75	0.75	0.68	0.63	0.58	0.58	0.45
Q3	0.85	0.79	0.79	0.70	0.69	0.60	0.64	0.54
Max	0.90	0.90	0.85	0.75	0.70	0.75	0.70	0.60

a cookstove more than the other age groups, which contradicts findings of previous studies: namely that the older women are, the more they tend to prefer traditional cooking methods due to the difficulty of changing behaviours [59] and the less they accept improved cookstoves [67]. This discrepancy might be because the sample data was too small to make representative predictions.

- **Household size**: The number of people living in a certain house ('members_house') was in the top five most important variables for all prediction analyses in which it was included (see Appendix A). The participation of 'members_house' in the predictions ranged from 10% to 43%. The influence of this variable on the predictions is high because 'members_house' has one of the highest η with 'TopVals' (0.14). These results suggest that the greater the number of people in the household, the greater the possibility (even if it is weak) that they hold the top five values related to cookstoves. A study carried out in Ouagadougou, Burkina Faso supports this point: it found that the largest families are usually the poorest ones and tend to use firewood more continuously, while richer families are less numerous and tend to use more LPG [91]. Therefore, a positive correlation between the number of members of a household and the acquisition of values related to the prioritisation of a cookstove ('TopVals') could be driven by the needs and values ('5TopVals') resulting from having a large family.
- **Quintile 2 of wealth index**: The proportion of people belonging to quintile 2 of the wealth index ('g2') was in the top five variables in all the prediction analyses in which it was considered. This result aligns with previous findings that, in general, a person's wealth determines the choice of one fuel over another for cooking [67].
- **Precipitation**: Our findings suggest that people are more likely to hold the top cookstove-related values where there are higher precipitation rates. This suggests that the greater the annual rainfall experienced, the more a person may prioritise a cookstove due to their values. One could theorise that rain may preclude continuous use of firewood for cooking due to humidity [92]. This could make people tend to appreciate a cookstove to meet their needs, especially that of 'food security,' as less availability of dry wood might mean less food to eat.
- **Distance to forests**: The importance rates of the predictive analyses also suggest that people that live the longest distances from the smallest forest within a ten kilometre radius are more likely to hold the top cookstove-related values. 'Food security' and 'time benefit' could be the most relevant values here, as it might be difficult to provide a household with food because more time is spent collecting firewood in these forested areas. These concerns might cause a person to consider purchasing a cookstove. However, the acquisition of this appliance also relates to the value of 'fixed costs' since there may be the possibility that these people cannot afford the initial cost. Therefore, the generation of one value could be determined by the presence of two others.
- **Forest density**: The findings suggest that that the higher the forest density around an observation, the greater the likelihood of holding cookstove-related values. However, this relationship may be biased, since few people chose a cookstove when the forest density in their location was high, implying that this same set of '5TopVals' values could be addressed by acquiring a different item. In any case, the forest density around a point could influence the value of 'food security' based on availability (or lack thereof) of natural and imported food resources. Moreover, 'time benefit' could be related to forest density as highly forested areas might prevent people from accessing regions where they could effectively access firewood, making them appreciate a cookstove.
- **Distance to water source**: The distance to the water source also appears to influence the prediction of 'TopVals.' The time spent on collecting water has been previously found to have a large impact on the selection of the type of fuel used for cooking [67], in the sense that if less time were spent in collecting water (i.e., shorter distances to travel for collection), the wealth from performing other tasks would make people consider using cleaner fuels for cooking. This would theoretically then lead them to adopt cleaner technologies.

Other variables and results are also worth considering. For example, the association value of 0.13 between 'TopVals' and luminance could indicate that when people have more access to electricity, they are in a better position to prioritise a cookstove. Furthermore, despite the belief that women value a cookstove more than men, gender does not show any association ($\eta = 0$) with either 'TopVals' or 'cookstove.' It only appears in the top five variables in two prediction analyses with participation percentages of around 4.5%. This might confirm some conclusions made by [48] when they mention that "men and women in rural Uganda held largely the same high-priority underlying values focused on basic human needs such as income, healthcare, information services, food security, and water security".

5.3. Implications

The statistical and predictive ML method developed here to understand the relationship between energy services and values using open spatial data can be used to improve the cooking and energy sectors in a number of ways.

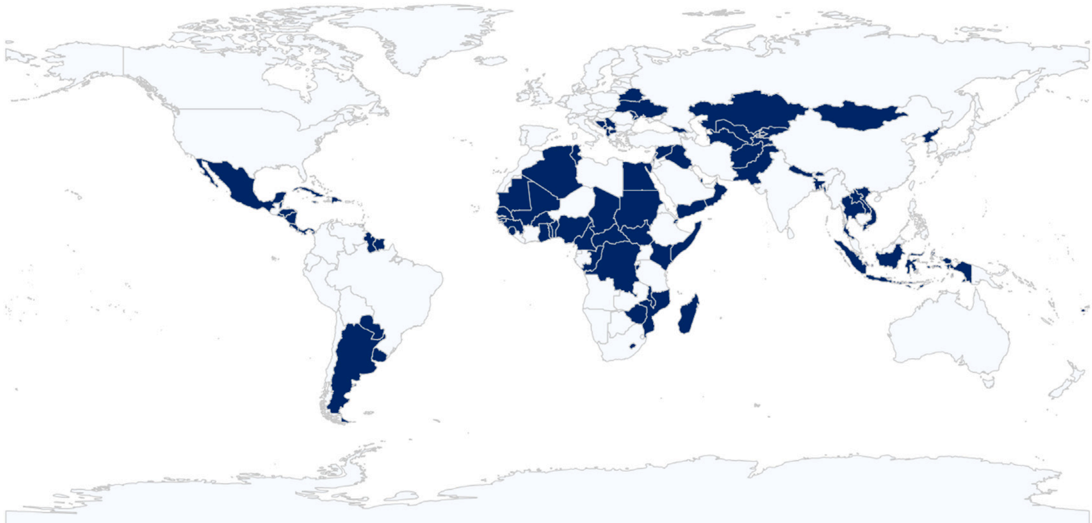

Fig. 7. Availability of Multiple Indicator Cluster Survey data including energy appliances. Countries where data are available are shown in dark blue. This includes 89 low- and middle-income countries. This data source can be used to extend the method applied in this study. Data from [93].

This method can allow faster and cheaper prediction of local values using existing data. By leveraging geospatial, socioeconomic, and demographic data available openly online via existing national surveys to map values, money and time can be saved in preliminary project scoping. Sources such as national census data, Multiple Indicator Cluster Surveys conducted by the United Nations, and Demographic and Health Surveys conducted by the United States Agency for International Development can be used to run this analysis. These data can be found for LMICs throughout sub-Saharan Africa and globally. To illustrate, Fig. 7 shows the availability of Multiple Indicator Cluster Survey data where energy appliances are captured — this includes 89 LMICs [93]. These data could be used within the methodology proposed here to predict the uptake of the energy appliances studied in the survey.

These results can be processed in less time than on-the-ground survey transcription and post-processing. That said, it is still necessary to undertake on-the-ground value data collection and socioeconomic surveys to keep predictive model training data up-to-date. On-the-ground data collection is also still needed prior to implementing energy service programmes; this model just reduces the data collection burden to perform initial value-informed targeting of energy services.

In practice, this method can be used to improve distribution programs for cookstoves or other energy services across LMICs. Understanding which parameters influence the prioritisation of cleaner cooking technologies could help governments and organisations in LMICs improve marketing for their cookstove distribution programs. This aligns with on-the-ground experience, for instance in the Room to Breathe programme on cookstove marketing in India [94], where a key lesson learned is listed to be: "Once you have identified the triggers and barriers you need to tightly match your communications messaging to them" [p. vi].

Understanding values related to cooking can help programme managers to effectively communicate cookstove benefits in a way that speaks to the users' daily life. This can help to drive greater acceptance and uptake of clean cooking. It could translate into greater lasting demand based on continued desire to adopt larger ('food security,' 'time benefit'), modern ('aesthetics'), cleaner ('environmental sustainability,' 'being healthy') cookstoves, where these values are found to be present. By understanding why and where cookstoves are prioritised, last mile cookstove distribution can be improved to create maximum uptake with minimum investment. The resources saved through effective targeting could allow increased focus on those people living in small, remote, under-serviced communities.

Third, the eventual consequences of improved targeting and distribution are more funding and higher rates of clean cooking access. Higher adoption rates over time could attract more funding to support more cookstove distribution programs in a virtuous cycle. With more successful cookstove distribution programs that reach the last mile, receive increasing financial support for their continuation, and constantly innovate, current low rates of access to clean cooking in sub-Saharan Africa can increase. This could help to not only reach Sustainable Development Goal 7 faster, but other goals closely tied to clean cooking. These include 3 (good health and well-being), 4 (quality education), and 13 (climate action), among others [13].

5.4. Study limitations

This study has some limitations. First, this study is limited by the size of the UPV-DS dataset. There are only 199 individuals represented, which is insufficient to generate a robust and precise predictive model. In addition, these observations are geotagged by village — as there are multiple observations per village, the geospatial data attached to multiple interviewees is often the same. That is why not only more observations but also greater geographic diversity of observations is required to build a more robust model. Second, this study is limited to considering the importance of people towards cookstoves at a given time. A certain society can experience changes over time, and values can change due to any event. That is why a continuous collection of values over time, including moments that can generate a sudden change, such as COVID, would allow for improved dynamic modelling. Third, this study considers geospatial and socio-economic/demographic datasets from sources that are often not up-to-date or do not correspond to the year in which values data may be collected. Complete and up-to-date data are required to obtain more representative results. Therefore, this study should be seen as a first step towards developing a methodology for the prediction of value-driven need via socioeconomic and geospatial modelling. The specific results should not be extrapolated beyond the case study context.

6. Conclusions

This work has proposed a method to predict the value-driven prioritisation of the energy service of clean cooking using geospatial, socioeconomic, and demographic parameters. Correlations between these parameters and the values associated with cookstove prioritisation were analysed using a User-Perceived Value dataset collected in rural Uganda (labelled UPV-DS) [17,60]. Then, through machine learning (ML) methods, it was determined which parameters are most predictive of cookstove prioritisation and related values.

Of the 199 individuals in the case study dataset, only 26 chose a cookstove as an important item. The values most frequently related to the prioritisation of a cookstove in the dataset were found to be 'food security,' 'time benefit,' 'accessibility to services,' 'fixed costs,' and 'being healthy.' Another 142 individuals in the dataset who did not choose a cookstove still held this group of values in their records. Thus, there is the possibility that about 71% of the people in this dataset would be willing to adopt a cookstove to satisfy their value-driven needs.

In the correlation analysis, it was found that correlation ratios between each of potential predictor parameters and either (a) prioritisation of a cookstove or (b) holding values related to cookstoves were in the range 0 to 0.27. The overall weakness of correlations was attributed to the low number of observations. Variables with a correlation ratio greater than or equal to 0.1 were considered as possible predictors in subsequent analyses.

Moving to predictive modelling, it was determined that Random Forest (RF) was the best performing algorithm for predicting both cookstove prioritisation and holding cookstove-related values. Using RF across 12 different analyses, age was found to be the most determining variable in prediction whenever it was included. Other important variables for predicting the cookstove-related values were the number of people living in a house, quintile 2 of the wealth index, annual accumulated precipitation, distance to the nearest forest within a radius of ten kilometres, forest density, distance to the water source, luminance, and distance to the nearest road. These parameters were consistent with what had been found in previous studies as determinants for cookstove adoption.

Through this investigation, this study develops a novel open-access predictive model designed to accelerate energy services targeting. It is implemented to be scalable to other geographic contexts and energy service technologies. This method can be used to improve value-driven targeting programs for cookstoves or other energy services across LMICs. It demonstrates how the combination of ML and openly available data can allow improved development program targeting aligned not only with economic factors such as poverty levels but with deeper values and needs.

6.1. Future work

This research aims to ignite interest in mapping deep value-driven needs against geospatial, socioeconomic, and demographic parameters for better energy service targeting. There are a number of opportunities for future work in this area.

- **Expand sample size and geography.** As discussed in Section 5.4, this study is limited by the small sample size of the case study dataset. As such, future work could collect additional values data across diverse geographies. This would allow for the development of a more robust, precise, and generlisable predictive model.
- **Study a broader array of energy services and technologies.** Future work should analyse which geospatial, socioeconomic, and demographic variables influence the acquisition of values related to the prioritisation of other items which provide energy services (e.g., television, radio, computer, fridge, computers, generators, water pumps). This could allow the design of initiatives or programmes that address several problems or needs simultaneously with sets of technologies.
- **Move beyond analysis which is instantaneous in time.** As discussed in Section 5.4, this study uses a dataset which represents a particular snapshot in time. This work therefore cannot account for changes in values across a society over time. Future work should therefore collect panel data on values to consider changes in the values of societies.
- **Improve data availability.** As discussed in Section 5.4, this type of analysis is dependent on complete and up-to-date open data. Therefore, as more up-to-date data is made available, this analysis can be repeated to generate more accurate results.

CRediT authorship contribution statement

Micaela Flores Lanza: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Visualization. **Alycia Leonard:** Writing – original draft, Writing – review & editing, Supervision, Visualization. **Stephanie Hirmer:** Conceptualization, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Code is linked on GitHub in-article. Data sources listed in table in article.

Acknowledgements

This work has been undertaken as part of the Climate Compatible Growth programme, which is funded by Climate Compatible Growth research programme, which is supported by the FCDO (Foreign Commonwealth and Development Office) of the United Kingdom.

Appendix A. Results of all analyses

Fig. A.8 shows the top five variables that most influence the predictive model of each analysis and their performance indicators.

Appendix B. Variables definitions, keys, and types

Table B.9 lists the independent variables that were considered in this study, their keys, and their types.

Appendix C. Correlation ratios with 'TopVals' and 'cookstove'

Table C.10 lists the correlation ratios for each independent variable with 'TopVals' and 'cookstove'.

Appendix D. Correlation matrices

Figs. D.9, D.10, and D.11 show the Pearson's coefficients (between numerical variables), Cramer's V coefficients (between categorical variables), and correlation ratios (between categorical and numerical variables) as matrices.

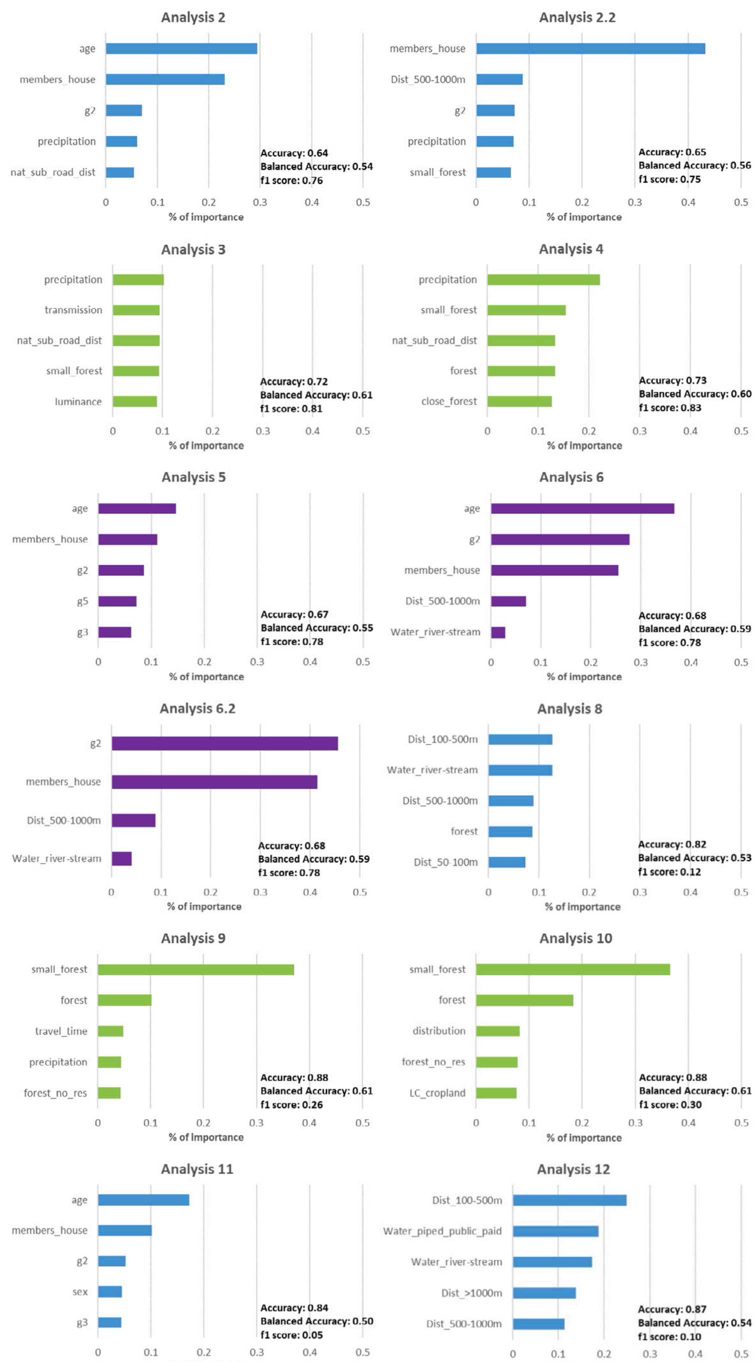

Fig. A.8. Top five variables for each analysis made.

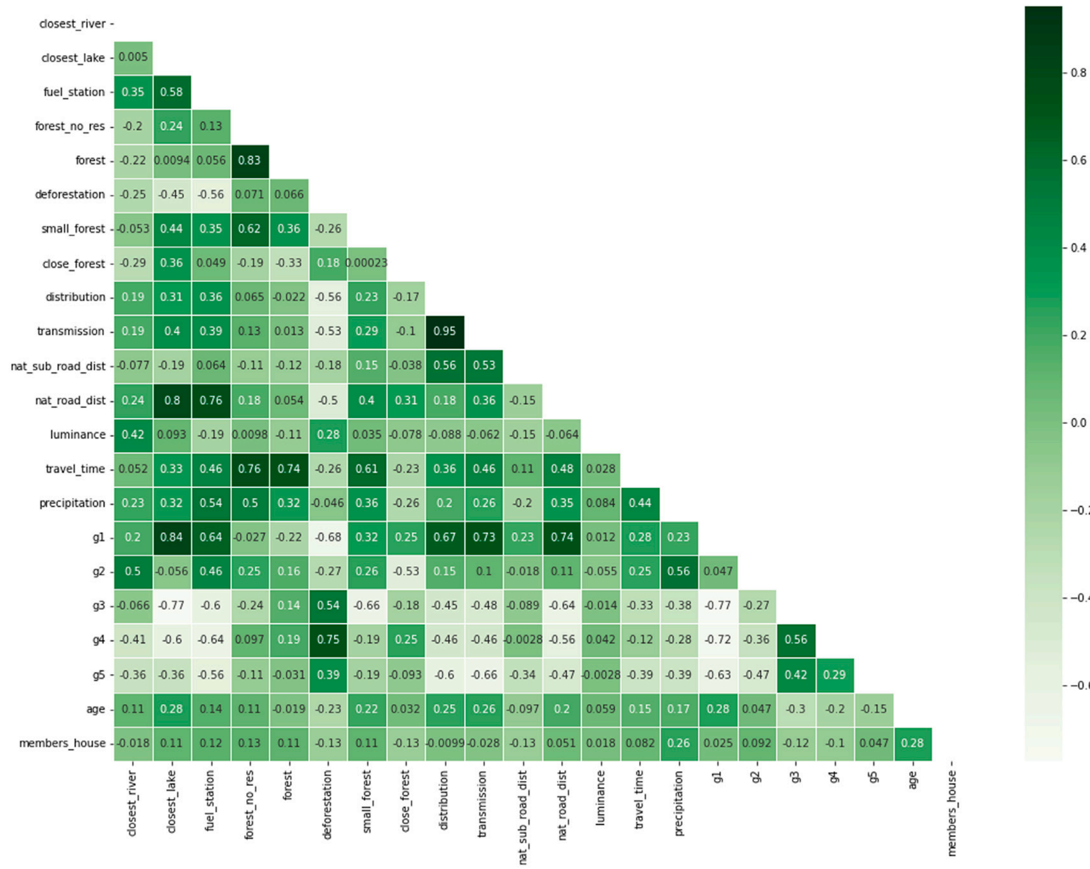

Fig. D.9. Correlation matrix with numerical variables (using Pearson's coefficient).

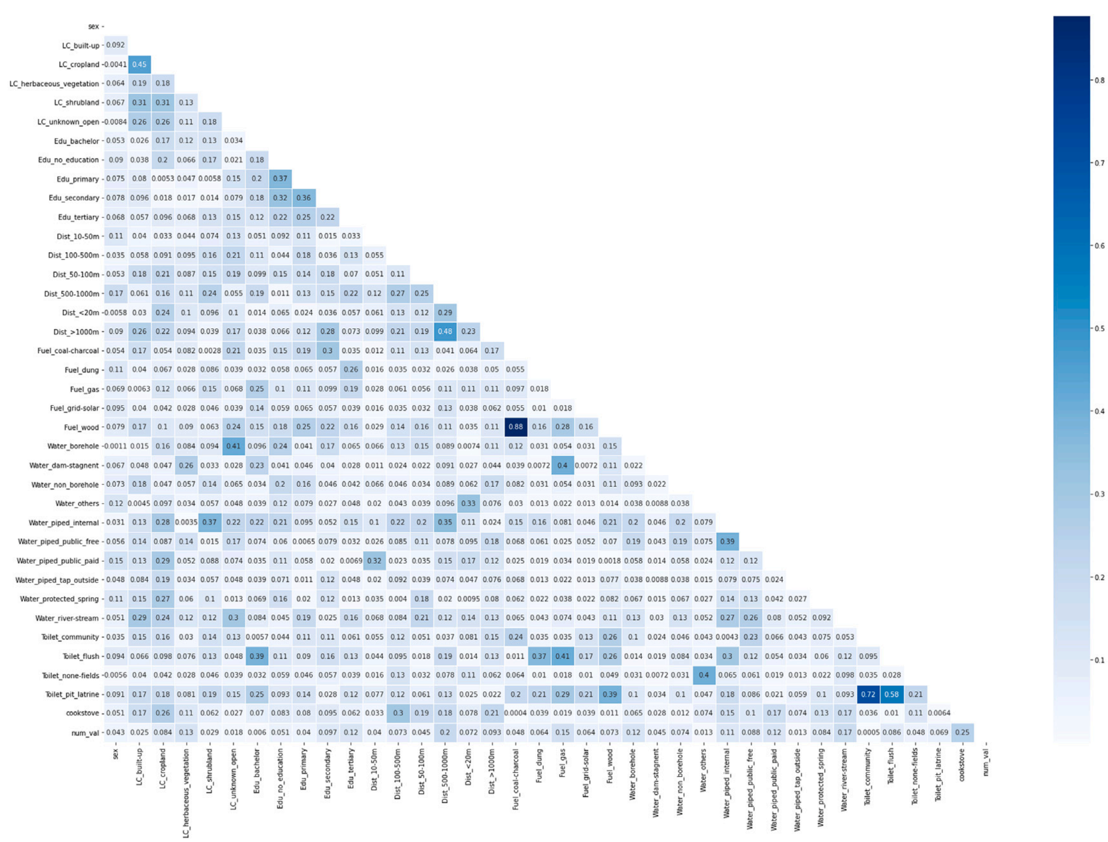

Fig. D.10. Correlation matrix with categorical variables (using Cramer's V coefficient).

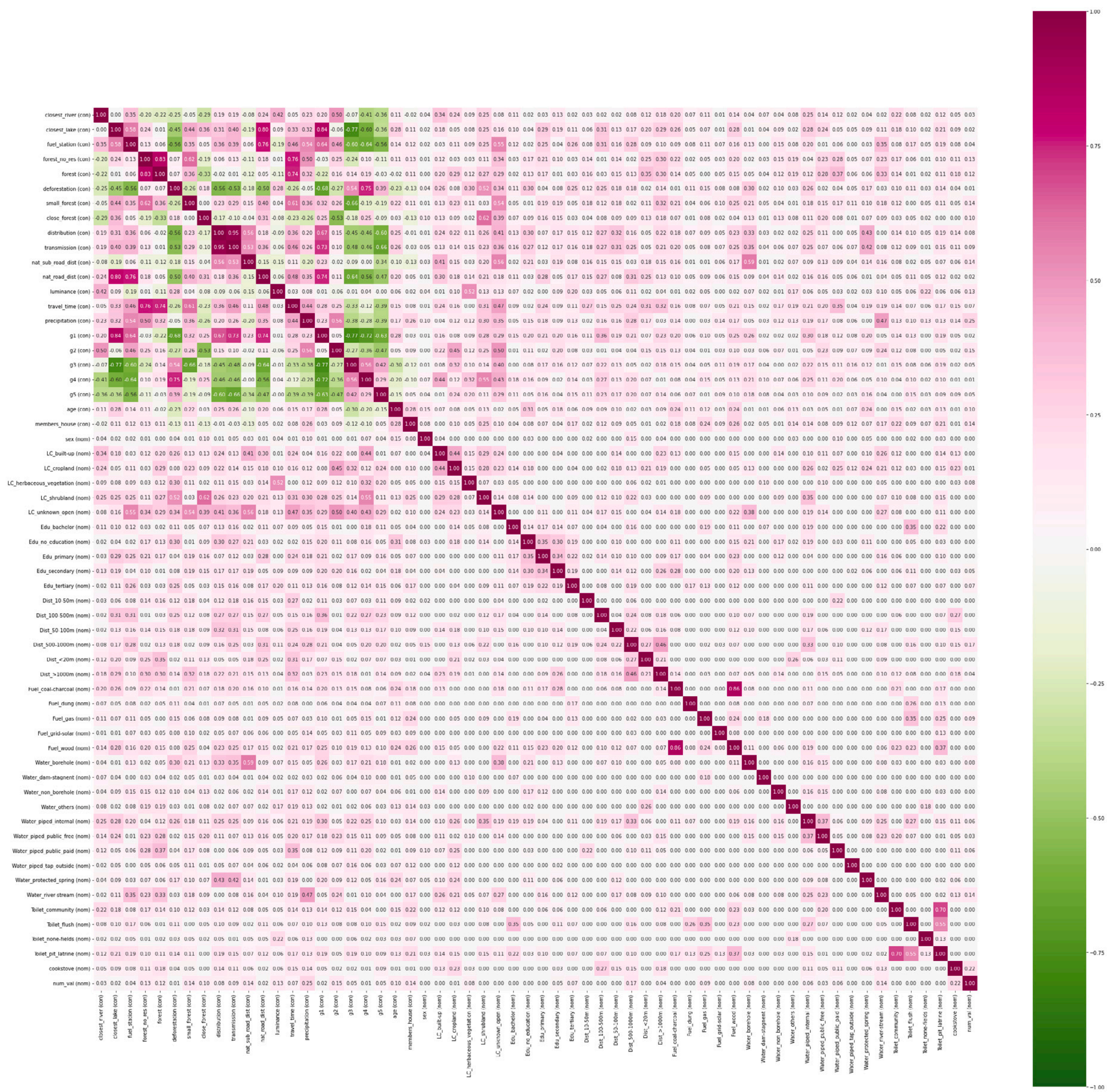

Fig. D.11. Correlation ratios between all variables.

Table B.9
Variables definitions, keys, and types. Note that distances are taken from the recorded location of the interviewee.

Variable	Value description	Key	Data type
Age	Interviewee age	Age	Numerical
Sex	0 = Female 1 = Male	Sex	Nominal
No. people in household	Number of people with whom the interviewee lives in the same house	members_house	Numerical
Education level	0 = Not completed/Never enrolled 1 = Completed primary school 2 = Completed secondary school 3 = Completed tertiary school 4 = Completed bachelor's degree	Edu_no_education Edu_primary Edu_secondary Edu_tertiary Edu_bachelor	Ordinal Ordinal Ordinal Ordinal Ordinal
Distance to fetch water	1 = Less than 20 meters 2 = 20 m - less than 50 m 3 = 50 m - less than 100 m 4 = 100 m - less than 500 m 5 = 500 m - less than 1 km 6 = More than 1km	Dist_<20m Dist_20-50m Dist_50-100m Dist_100-500m Dist_500-1000m Dist_>1000m	Ordinal Ordinal Ordinal Ordinal Ordinal Ordinal
Main type of fuel used for cooking in the household	1 = Wood 2 = Coal/charcoal 3 = Gas 4 = Dung 5 = electricity (grid or solar)	Fuel_wood Fuel_coal-charcoal Fuel_gas Fuel_dung Fuel_grid-solar	Nominal Nominal Nominal Nominal Nominal
Source of water mostly used in the household for drinking, bathing, and washing clothes	1 = Piped public tap/kiosk (free) 2 = Piped public tap/kiosk (paid for) 3 = Piped internal 4 = Piped tap outside 5 = Well (non borehole) 6 = Borehole/ hand pump 7 = Flowing river/stream 8 = Dam/stagnant water 9 = Protected spring 10 = Others	Water_piped_public_free Water_piped_public_paid Water_piped_internal Water_piped_tap_outside Water_non_borehole Water_borehole Water_river-stream Water_dam-stagnent Water_protected_spring Water_others	Nominal Nominal Nominal Nominal Nominal Nominal Nominal Nominal Nominal Nominal
Type of toilet used	1 = Flush toilet 2 = Pit latrine 3 = Community toilet 4 = None/use fields etc.	Toilet_flush Toilet_pit_latrine Toilet_community Toilet_none-fields	Nominal Nominal Nominal Nominal
Closest river	Straight line distance to the closest river (km).	closest_river	Numerical
Closest lake	Straight line distance to the closest lake (km).	closest_lake	Numerical
Closest fuel station	Straight line distance to the closest fuel station (km).	fuel_station	Numerical
Closest forest	Straight line distance to the edge of the nearest forest reserve (km).	small_forest	Numerical
Closest smallest forest within a radius	Straight line distance to the edge of the smallest forest reserve within ten km radius.	close_forest	Numerical
Density of forests	Average forest density in a range from 0 to 100 considering a radius of ten km from interviewee location.	Forest	Numerical
Density of forests (without reserves)	Average forest density, excluding forest reserves, in a range from 0 to 100 considering a radius of ten km from interviewee location.	forest_no_res	Numerical
Grade of deforestation	Average degree of deforestation from 0 (no loss) to 20 (total loss) from 2000 to 2021 considering ten km from interviewee location.	deforestation	Numerical
Closest distribution substation	Straight-line distance to the closest distribution substation (km).	distribution	Numerical
Closest transmission substation	Straight line distance to the closest transmission substation (km).	transmission	Numerical
Closest road	Straight-line distance to the closest national or subnational road (km).	nat_sub_road_dist	Numerical
Closest road (national)	Straight-line distance to the closest national road (km).	nat_road_dist	Numerical
Land Cover	Type of land cover present in the location of the interviewee.	LC_built-up LC_cropland LC_herbaceous_vegetation LC_shrubland LC_unknown_open	Nominal Nominal Nominal Nominal Nominal
Luminance	Luminance value between 3 and 255 perceived in the location of the interviewee.	Luminance	Numerical
Travel Time	Time in minutes that it takes to reach an urban conglomeration of more than 50,000 inhabitants according to the location of the interviewee.	travel_time	Numerical
Precipitation (mm)	Accumulated annual rainfall in mm at the location of the interviewee.	precipitation	Numerical
Wealth Index (Quintile 1) Wealth Index (Quintile 2) Wealth Index (Quintile 3) Wealth Index (Quintile 4) Wealth Index (Quintile 5)	Once the cluster of the DHS survey belonging to the location of each interviewee is determined, the number of people (in percentage) that belong to each quintile of the Wealth Index in that cluster is calculated; then, these values are assigned to each interviewee.	g1 g2 g3 g4 g5	Numerical Numerical Numerical Numerical Numerical

Table C.10
Correlation Ratios (CR) with 'TopVals' and 'cookstove'.

Variable	η with 'TopVals'	η with 'cookstove'
closest_river	0.03	0.05
closest_lake	0.02	0.09
fuel_station	0.04	0.08
forest_no_res	0.13	0.11
forest	0.12	0.18
deforestation	0.01	0.04
small_forest	0.14	0.05
close_forest	0.1	0
distribution	0.08	0.14
transmission	0.09	0.11
nat_sub_road_dist	0.14	0.06
nat_road_dist	0.02	0.02
luminance	0.13	0.06
travel_time	0.07	0.15
precipitation	0.25	0.14
g1	0.02	0.05
g2	0.15	0.02
g3	0.05	0.02
g4	0.01	0.01
g5	0.05	0.09
age	0.1	0.01
members_house	0.14	0.01
sex	0	0
LC_built-up	0	0.13
LC_cropland	0.01	0.23
LC_herbaceous_vegetation	0.08	0.03
LC_shrubland	0	0
LC_unknown_open	0	0
Edu_bachelor	0	0
Edu_no_education	0	0
Edu_primary	0	0
Edu_secondary	0.05	0.03
Edu_tertiary	0.07	0
Dist_10-50m	0	0
Dist_100-500m	0	0.27
Dist_50-100m	0	0.15
Dist_500-1000m	0.17	0.15
Dist_<20m	0	0
Dist_>1000m	0.04	0.18
Fuel_coal-charcoal	0	0
Fuel_dung	0	0
Fuel_gas	0.09	0
Fuel_grid-solar	0	0
Fuel_wood	0	0
Water_borehole	0.08	0
Water_dam-stagnent	0	0
Water_non_borehole	0	0
Water_others	0	0
Water_piped_internal	0.06	0.11
Water_piped_public_free	0.03	0.05
Water_piped_public_paid	0.06	0.11
Water_piped_tap_outside	0	0
Water_protected_spring	0	0.06
Water_river-stream	0.14	0.13
Toilet_community	0	0
Toilet_flush	0	0
Toilet_none-fields	0	0
Toilet_pit_latrine	0	0

References

[1] Fell MJ. Energy services: A conceptual review. Energy Res Soc Sci 2017;27:129–40.

[2] Clean Cooking Alliance. 2022 Clean cooking industry snapshot. Tech. rep., Clean Cooking Alliance; 2022, URL https://cleancooking.org/wp-content/uploads/2022/05/CCA-2022-Clean-Cooking-Industry-Snapshot.pdf.

[3] World Bank. Access to clean fuels and technologies for cooking (% of population) - Sub-Saharan Africa (excluding high income). 2021, URL https://data.worldbank.org/indicator/EG.CFT.ACCS.ZS?locations=ZF.

[4] World Bank. Access to clean fuels and technologies for cooking, rural (% of rural population) - Sub-Saharan Africa (excluding high income). 2021, URL https://data.worldbank.org/indicator/EG.CFT.ACCS.RU.ZS?locations=ZF.

[5] World Health Organisation. Population with primary reliance on polluting fuels and technologies for cooking (in millions). 2022, URL https://www.who.int/data/gho/data/indicators/indicator-details/GHO/population-with-primary-reliance-on-polluting-fuels-and-technologies-for-cooking-(in-millions).

[6] World Health Organisation. Household air pollution and health. 2022, URL https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health.

[7] Global Alliance for Clean Cooking. Igniting change: A strategy for universal adoption of clean cookstoves and fuels. Tech. rep., Global Alliance for Clean Cooking; 2011, URL https://cleancooking.org/reports-and-tools/igniting-change-a-strategy-for-universal-adoption-of-clean-cookstoves-and-fuels/.

[8] Bhanot J, Jha V. Moving towards tangible decision-making tools for policy makers: Measuring and monitoring energy access provision. Energy Policy 2012;47.

[9] Clean Cooking Alliance. Environmental impacts of cooking. Tech. rep., Clean Cooking Alliance; 2021.

[10] Bailis R, Drigo R, Ghilardi A, Masera O. The carbon footprint of traditional woodfuels. Nature Clim Change 2015;5.

[11] Njenga M, Gitau J, Mendum R. Women's work is never done: Lifting the gendered burden of firewood collection and household energy use in Kenya. Energy Res. Soc. Sci. 2021;77.

[12] Clean Cooking Alliance. Gender and clean cooking. Tech. rep., Clean Cooking Alliance; 2021, URL https://cleancooking.org/wp-content/uploads/2021/07/CCA-gender-sheet-ENGLISH.pdf.

[13] Global Alliance for Clean Cooking. Delivering on the SDGS through clean cooking. Tech. rep., Global Alliance for Clean Cookstoves; 2020, URL https://sdgs.un.org/sites/default/files/documents/11416Global%2520Alliance%2520for%2520Clean%2520Cookstoves%2520-%2520Delivering%2520on%2520the%2520SDGs%2520through%2520Clean%2520Cooking.pdf.

[14] Clean Cooking Alliance. Annual report 2021. Tech. rep., Clean Cooking Alliance; 2021, URL https://cleancooking.org/reports-and-tools/2021-annual-report/.

[15] Boudewijns EA, Trucchi M, van der Kleij RM, Vermond D, Hoffman CM, Chavannes NH, et al. Facilitators and barriers to the implementation of improved solid fuel cookstoves and clean fuels in low-income and middle-income countries: An umbrella review. Lancet Planet Health 2022.

[16] Thakur M, Van Schayck C, Boudewijns E. Improved cookstoves in low-resource settings: A spur to successful implementation strategies. NPJ Primary Care Respir Med 2019;29(1):36.

[17] Hirmer SA. Improving the sustainability of rural electrification schemes: Capturing value for rural communities in Uganda [Ph.D. Thesis], Centre for Sustainable Development, Department of Engineering, University of Cambridge; 2017.

[18] Vigolo V, Sallaku R, Testa F. Drivers and barriers to clean cooking: A systematic literature review from a consumer behavior perspective. Sustainability 2018;10(4322).

[19] Boudewijns E, Trucchi M, van der Kleij Rianne MJJ, Vermond D, Hoffman CM, Chavannes NH, et al. Facilitators and barriers to the implementation of improved solid fuel cookstoves and clean fuels in low-income and middle-income countries: An umbrella review. Lancet Planet Health 2022;6.

[20] Barnes DF, Openshaw K, Smith KR, van der Plas R. What makes people cook with improved biomass stoves? Energy series, The World Bank; 1994.

[21] Narayanasamy N. Participatory rural appraisal: Principles, methods, and applications. International series of monographs on physics, New Delhi: SAGE Publications India Pvt Ltd; 2009.

[22] MartíCosta S, Serrano-García I. Needs assessment and community development. Prev Hum Serv 2010;2(4).

[23] Sandhu JS, Altankhuyag P, Amarsaikhan D. Serial hanging out: Rapid ethnographic needs assessment in rural settings. Human-Computer Interaction, Part I, HCII 2007 2007;4550.

[24] Leonard A. Scalable spatial design of electricity access systems [Ph.D. Thesis], Hertford College, University of Oxford; 2022.

[25] Manjula M, Gopi G. Universal access to clean cooking energy and the need for an inclusive policy: Evidence from analysis of cooking fuel use in Odisha and Tamil Nadu. Indian Inst Manag Calcutta 2017;44(3).

[26] Nnaemeka U. How pursuing clean cooking in Africa fights indoor air pollution and deforestation. 2020, URL https://ecowarriorprincess.net/2020/09/clean-cooking-africa-indoor-air-pollution-deforestation/.

[27] Schlag N, Zuzarte F, Schlag N. Market barriers to clean cooking fuels in Sub-Saharan Africa: A review of literature. Tech. rep., Stockholm Environment Institute; 2008, URL https://www.sei.org/publications/market-barriers-clean-cooking-fuels-sub-saharan-africa-review-literature/.

[28] Baker JL, Grosh ME. Poverty reduction through geographic targeting: How well does it work? World Dev 1994;22(7):983–95.

[29] De-Arteaga M, Herlands W, Neill DB, Dubrawski A. Machine learning for the developing world. ACM Trans Manag Inf Syst (TMIS) 2018;9(2):1–14.

[30] Aiken EL, Bedoya G, Coville A, Blumenstock JE. Targeting development aid with machine learning and mobile phone data: Evidence from an anti-poverty intervention in afghanistan. In: Proceedings of the 3rd ACM SIGCAS conference on computing and sustainable societies. 2020, p. 310–1.
[31] Aiken E, Bellue S, Karlan D, Udry C, Blumenstock JE. Machine learning and phone data can improve targeting of humanitarian aid. Nature 2022;603(7903):864–70.
[32] Awwad Y, Fletcher R, Frey D, Gandhi A, Najafian M, Teodorescu M. Exploring fairness in machine learning for international development. Tech. rep., CITE MIT D-Lab; 2020.
[33] United Nations. Do you know all 17 SDGs? 2022, URL https://sdgs.un.org/es/goals.
[34] Sustainable Energy for All. Clean cooking. 2022, URL https://www.seforall.org/clean-cooking.
[35] Electricity Regulatory Authority of Uganda. Energy minister launches reviewed electricity tariff structure. 2021, URL https://www.era.go.ug/index.php/media-centre/what-s-new/371-energy-minister-launches-reviewed-electricity-tariff-structure.
[36] National Planning Authority. Third national development plan (NDPIII) 2020/21 – 2024/25. Tech. rep., Republic of Uganda; 2020, URL http://www.npa.go.ug/wp-content/uploads/2020/08/NDPIII-Finale_Compressed.pdf.
[37] Orena AG. Energy minister launches reviewed electricity tariff structure. 2022, URL https://observer.ug/news/headlines/73411-era-targets-50-000-homes-to-shift-from-charcoal-to-power.
[38] Ogwok J, Naluwagga A, Abbo MS, Tesfamichael M. Uganda's cooking energy sector: A review. Tech. rep., Modern Energy Cooking Services (MECS); 2022, URL https://mecs.org.uk/wp-content/uploads/2022/06/Ugandas-cooking-energy-sector-a-review.pdf.
[39] Sidorchuk R. The concept of "value" in the theory of marketing. Asian Soc Sci 2015;11(9).
[40] Stanford Encyclopedia of Philosophy. Value theory. 2008, URL https://plato.stanford.edu/entries/value-theory/.
[41] Jaspers E. Oxford bibliographies. 2018, URL https://www.oxfordbibliographies.com/view/document/obo-9780199756384/obo-9780199756384-0182.xml.
[42] Doyle P. Marketing management and strategy. Harlow: Financial Times Prentice Hall; 2002.
[43] Oyserman D. Values, psychological perspectives. Elsevier; 2015.
[44] Schwartz SH. Are there universal aspects in the structure and contents of human values? J Soc Issues 1994;50(4).
[45] Nurkka P, Kujala S, Kemppainen K. Capturing users' perceptions of valuable experience and meaning. J Eng Des 2009;20(5).
[46] Wilson E, Godfrey Wood R, Garside B. Sustainable energy for all? Linking poor communities to modern energy services. Tech. rep., International Institute of Environment and Development; 2012, URL https://www.iied.org/16038iied.
[47] Conforti C, Hirmer S, Morgan D, Basaldella M, Or YB. Natural language processing for achieving sustainable development: The case of neural labelling to enhance community profiling. In: Proceedings of the 2020 conference on empirical methods in natural language processing. 2020, p. 8427–44.
[48] Hirmer S, Leonard A, Conforti S, Conforti C. Perceived value interviews and socio-economic survey data for communities in rural Uganda. Data Brief 2022;40.
[49] Leary J, Menyeh B, Chapungu V, Troncoso K. eCooking: Challenges and opportunities from a consumer behaviour perspective. Energies 2021;14.
[50] Bielecki C, Wingenbach G. Rethinking improved cookstove diffusion programs: A case study of social perceptions and cooking choices in rural Guatemala. Energy Policy 2014;66.
[51] Kapfudzaruwa F, Fay J, Hart T. Improved cookstoves in Africa: Explaining adoption patterns. Dev Southern Africa 2017;34(5).
[52] Urmee T, Gyamf S. A review of improved Cookstove technologies and programs. Renew Sustain Energy Rev 2014;33.
[53] Loo JD, Hyseni L, Ouda R, Koske S, Nyagol R, Sadumah I, et al. User perspectives of characteristics of improved cookstoves from a field evaluation in Western Kenya. Int J Environ Res Public Health 2016;13(167).
[54] Martin SL, Arney JK, Mueller LM, Kumakech E, Walugembe F, Mugisha E. Using formative research to design a behavior change strategy to increase the use of improved cookstoves in Peri-Urban Kampala, Uganda. Int J Environ Res Public Health 2013;10.
[55] van der Kroon B, Brouwer R, van Beukering PJ. The impact of the household decision environment on fuel choice behavior. Energy Econ 2014;44.
[56] Kulindwa YJ, Lokina R, Ahlgren EO. Driving forces for households' adoption of improved cooking stoves in rural Tanzania. Energy Strategy Rev 2018;20.
[57] Bonan J, Pareglio S, Tavoni M. Access to modern energy: A review of barriers, drivers and impacts drivers and impacts. Environ Dev Econ 2017;22(5).
[58] Jan I, Ullah S, Akram W, Khan NP, Asim SM, Mahmood Z, et al. Adoption of improved cookstoves in Pakistan: A logit analysis. Biomass Bioenergy 2017;103.
[59] Muneer SET, Mohamed EWM. Adoption of biomass improved cookstoves in a patriarchal society: An example from Sudan. Sci Total Environ 2003;307.
[60] Rural Senses. Demographic-rich qualitative UPV-Interviews dataset (DR-QI). 2022, URL https://github.com/gvanboven/DR-QI.
[61] Jagger P, Jumbe C. Stoves or sugar? Willingness to adopt improved cookstoves in Malawi. Energy Policy 2016;92.
[62] Troncoso K, Armendáriz C, Alatorre S. Improved cook stove adoption and impact assessment: A proposed methodology. Energy Policy 2013;62.
[63] Nguyen TTPT. Women's adoption of improved cook stoves in Timor-Leste: Challenges and opportunities. Dev Pract 2017;27(8).
[64] Asada S. The influence of food traditions and cooking methods on energy transition: High demand for charcoal in Kampala, Uganda. Nilo-Ethiopian Stud 2019;24.
[65] Rahut DB, Behera B, Ali A, Marenya P. A ladder within a ladder: Understanding the factors influencing a household's domestic use of electricity in four African countries. Energy Econ 2017;66.
[66] Lombardi F, Riva F, Sacchi M, Colombo E. Enabling combined access to electricity and clean cooking with pvmicrogrids: New evidences from a high-resolution model of cooking loads. Energy Sustain Dev 2019;49.
[67] Zhang J. Identifying key drivers for cooking energy consumption patterns in developing countries using machine learning. Zurich: Urban Energy Systems Laboratory, Swiss Federal Laboratories for Materials Science and Technology (EMPA); 2019.
[68] Efficiency for Access. Off- and weak-grid appliances impact assessment framework. 2022, URL https://efficiencyforaccess.org/publications/impact-assessment-framework.
[69] United States Agency for International Development. Demographic and socioeconomic surveys program. 2022, URL https://dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm.
[70] National Aeronautics and Space Administration. Giovanni user guide. 2022, URL https://giovanni.gsfc.nasa.gov/giovanni/doc/UsersManualworkingdocument.docx.html.
[71] National Aeronautics and Space Administration. Earth observatory. 2022, URL https://earthobservatory.nasa.gov/features/NightLights/page3.php.
[72] Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova, Turubanova SA, et al. Global forest change. 2022, URL https://glad.earthengine.app/view/global-forest-change.
[73] Buchhorn M, Lesiv M, Tsendbazar N-E, Herold M, Bertels L, Smets B. Copernicus global land cover layers—Collection 2. Remote Sens 2020;12(6).
[74] Malaria Atlas Project. Accessibility to cities. 2022, URL https://malariaatlas.org/research-project/accessibility-to-cities/.
[75] United Nations Environment Program - World Conservation Monitoring Centre. Discover the world's protected areas. 2022, URL https://www.protectedplanet.net/en.
[76] OpenStreetMap. Overpass turbo. 2022, URL https://overpass-turbo.eu/.
[77] ENERGYDATAINFO. Uganda - distribution substations. 2016, URL https://energydata.info/dataset/distribution-substations-2016.
[78] ENERGYDATAINFO. Uganda - transmission SubStations operational. 2016, URL https://energydata.info/dataset/transmission-substations-2016-operational.
[79] Dataug. Data.ug. 2022, URL http://catalog.data.ug/pages/about-data-ug.
[80] ENERGYDATAINFO. Uganda - rivers. 2014, URL https://energydata.info/dataset/uganda-rivers.
[81] ENERGYDATAINFO. Uganda - lakes. 2014, URL https://energydata.info/dataset/uganda-lakes.
[82] Mishra P. 5 outlier detection techniques that every "data enthusiast" must know. 2021, URL https://towardsdatascience.com/5-outlier-detection-methods-that-every-data-enthusiast-must-know-f917bf439210.
[83] Acuna E, Rodriguez C. The treatment of missing values and its effect the treatment of missing values and its effect. Classif, Clust, Data Min Appl 2004.
[84] Dython. Nominal - associations. 2022, URL http://shakedzy.xyz/dython/modules/nominal/.
[85] Hossin M, Sulaiman M. A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag Process (IJDKP) 2015;5(2).
[86] Tomitaka S, Kawasaki Y, Ide K, Akutagawa M, Ono Y, Furukawa TA. Stability of the distribution of patient health questionnaire-9 scores against age in the general population: Data from the national health and nutrition examination survey. Front. Psychiatry 2018;9.
[87] Wang Y, Bailis R. The revolution from the kitchen: Social processes of the removal of traditional cookstoves in Himachal Pradesh, India. Energy Sustain Dev 2015;27.
[88] Ruiz-Mercado I, Masera O, Zamora H, Smith KR. Adoption and sustained use of improved cookstoves. Energy Policy 2011;39.
[89] Wallmo K, Jacobson SK. A social and environmental evaluation of fuel-efficient cook-stoves and conservation in Uganda. Environ Conserv 1998;25(2).
[90] Ogunbunmi K, Mwando M. Africa's climate policies burned by firewood dependence. 2014, URL https://news.trust.org/item/20140530183509-63ekq.
[91] Ouedraogo B. Household energy preferences for cooking in urban Ouagadougou, Burkina Faso. Energy Policy 2016;34.
[92] O'Kelly J. Rain on seasoned firewood (does it matter?) 2020, URL https://fireplaceuniverse.com/rain-on-seasoned-firewood/.
[93] United Nations Children's Fund (UNICEF). MICS surveys. 2023, URL https://mics.unicef.org/surveys.
[94] Shell Foundation. Social marketing in India: Lessons learned from efforts to foster demand for cleaner cookstoves. Tech. rep., Shell Foundation; 2013.