Master Thesis

MSc Economics and Business

Specialization: Data Science and Marketing Analytics

Bridging the gap between attitude and behavior in the sustainability field using the power of machine learning

Mohamed Ghazy | 476086

Supervisor: Martijn de Jong

Second assessor: Andreas Alfons

Date final version: 28-09-2020

Abstract

Climate change and global warming are some of the most critical issues of this century. Immediate action must be taken to tackle the issue and prevent the catastrophic events of climate change from happening. Thus, sustainable living has become more crucial than ever in order to revert the consequences of the issue at hand. When promoting sustainable living, policymakers and companies have been faced with the challenge of the attitude-behavior gap phenomenon. This phenomenon happens when people do not translate their positive attitude towards climate change and sustainability into action. To solve this issue and bridge the gap, it is important to investigate and identify the drivers behind this phenomenon. Thus, a study was conducted by surveying people from 12 different countries. Machine learning techniques, namely logistic lasso regression and random forest, were employed to analyze the data and determine the most important factors behind the gap. The study found that country of residence is the most important factor contributing to sustainable consumption. At the same time, there was not enough evidence regarding the role of gender or the duration of education in bridging the gap. Besides, perceived consumer effectiveness and perceived environmental threats were found to be associated with sustainable behavior. The study is concluded with recommendations for marketers and policymakers, which can assist them in their decision-making process when promoting sustainable living.

Keywords: climate change, sustainable consumption, attitude-behavior gap, decision-making, machine learning, logistic regression lasso, random forest

Table of Contents

1. Introduction

Climate change and sustainability are two of the most talked-about topics of the current century. According to NASA, earth has been through seven climate change cycles in the past 650,000 years; the last cycle was the ice age, which was around 11,700 years ago. These cycles were mainly due to the changes in the orbit of the earth that led to variations in the amount of energy coming from the sun to earth. The current modern climate is going through changes as well; however, this time, humans are the reason behind this global warming the earth is facing since the 20th century. The current global warming is behind the melting of icebergs in the Arctic, and the extinction of many species. Global warming has also led to an increase in sea levels; this increase can cause catastrophic flooding in the future, and it also threatens the global production of food. Recently, bushfires have been happening across the globe; two recent events were in the Amazon and Australia. According to the United Nations Environment Program, the Australian bushfire led to overwhelming consequences; thousands of homes were destroyed, and millions of animals were killed. Scott Morrison, the Australian prime minister, announced that global warming influenced this devastating event (Readfearn, 2020). The scope of climate change is global, and the scale of it is unprecedented. Immediate action must be taken to tackle the climate change issue; according to the United Nations, not taking drastic immediate actions will raise the cost of solving the issue and will result in making it more difficult. Due to these environmental issues, special attention is drawn towards environmental sustainability as consumers become more aware of the situation. Environmental sustainability is defined as fulfilling the needs and demands of the current generation without compromising future generations or the environment (Morelli, 2011). To achieve this, minimizing the personal negative impact on the environment is essential; this can be done through, for instance, recycling, reduction of meat consumption, the usage of non-toxic materials, and the reduction of waste production. As a result, companies are acknowledging the importance of sustainable living, and, therefore, they are striving to make their businesses more sustainable (White, Habib & Hardisty, 2019).

However, even with this increase in interest in sustainability, many studies have proven that this positive attitude towards sustainability and climate change is not always translated into pro-environmental behavior (Park & Ha, 2014); this phenomenon is known to researchers as the attitude-behavior gap. This gap has a significant negative effect on many entities, such as policymakers, companies, marketers, and non-profit organizations; the gap makes promoting sustainable living and pro-environmental behavior more challenging for the previously mentioned entities (White, Habib & Hardisty, 2019). This gap between attitude and behavior is a consequence of a complex mix of barriers; these barriers range from personal and social barriers to psychological ones. They constrain the person's ability to put their positive attitude towards the environment into actions (Antimova, Nawijn & Peeters, 2012). This study aims to understand the previously defined gap in order to bridge it and help save the environment. Hence, the following research question is answered throughout this study:

What are the determinants behind the attitude-behavior gap in the environmental sustainability field, and how can policymakers and marketers use these findings for better decision-making?

Many theories have tried to explain the attitude-behavior gap. However, each of these theories did not manage to fully explain it adequately; a thorough understanding of the attributes that influence the gap, such as knowledge, (personal) values, perceived consumer effectiveness and perceived environmental threat is not yet complete. Besides, most of these studies are qualitative. Indeed, some of them are quantitative; however, very few, if any, managed to use machine learning techniques. This study combines several theoretical frameworks and the power of machine learning techniques in order to understand the gap further. The aim is to add to the current research by examining survey results from 12 different global north and south countries. The usage of the machine learning techniques on this valuable data will provide policymakers and marketeers with a data-driven understanding of the gap, which can eventually help them in their decision-making process.

This paper starts with a literature review that discusses the theoretical background. Data and methodology sections follow this; in these two sections, the data is explained along with the methods used to prepare and analyze it. Next, the result of the analysis is presented. Finally, the study is concluded with a discussion of the obtained results, recommendations for policymakers and marketers, study limitations, and suggestions for future research.

2. Literature Review

The survey used in this paper in order to understand the attitude-behavior gap was created in a way that captures several determinants of this gap. In this section, the literature review aims to explain these determinants, and it expands a little beyond these attributes in order to get a thorough understanding of them. The survey starts with questions regarding education, gender, and country of residence; these questions aim to examine the effect of ‘demographics’ on the gap. The variable ‘knowledge’ is also included in the survey; for instance, the respondents were asked about their source of environmental knowledge and the entity (e.g., the media, and scientists) they trust the most. Furthermore, the survey contains questions on the number of days of extreme weather respondents experienced in the past year, and whether they expect more or less of these days in the future. These weather-related questions can help in studying the effect of ‘perceived environmental threats’ on the gap. Several questions in the survey were based on ‘social norms’ from the theory of reasoned action and its extension the theory of planned behavior. For example, to examine ‘social norms’, people were asked the following question: “According to you, what percentage of young people in your country are willing to spend money to tackle climate change?”. Besides, ‘perceived consumer effectiveness’ and ‘perceived behavior control’ were examined through asking respondents whether they are willing to spend money to reduce climate change, even if others do not. In addition, other variables were also included in the survey, such as ‘habits’, ‘values’, and ‘personal norms’; people were asked to agree or disagree with sentences, such as “I work very hard”, “I never hide my mistakes”, “I never steal”, “I never gossip”, and “I follow a routine”. The survey's conceptual framework is drawn in figure 1 below; the figure provides a simplified explanation of the attitude-behavior gap determinants that the survey is based on.

2.1 Theory of Planned Behavior

Many researchers across the years have confirmed the existence of the attitude and behavior gap, and several studies have investigated this phenomenon within the sustainability field (Zralek, 2017). Many of these attempts have been successful in explaining the gap; however, special attention was drawn towards the Theory of Reasoned Action (TRA) by Ajzen and Fishbein in 1980, and its extension the Theory of Planned Behavior (TPB) by Ajzen in 1991. Most of the current studies have followed these two theories in order to study the gap (Joshi & Rahman, 2015). According to the Theory of Reasoned Action, both attitude and social norms influence intention. In turn, intention affects whether a person will engage in a behavior or not. Theory of Planned Behavior built on TRA by adding perceived behavior control as an additional determinant of a person's behavior (Madden, Ajzen & Ellen, 1992). TPB suggests that in addition to subjective norm and attitude, perceived behavior control can influence intentions, which leads to behavior prediction (Ajzen, 1991). Perceived behavior control is an indication of a person's perception of their capabilities to undertake a specific action. It assumes that people are rational and consider the consequences of their behavior (Ramayah, Lee & Lim, 2012). TPB provides a comprehensive framework for researchers, which helps understand the factors that influence the decision process when engaging in pro-environmental behavior, such as recycling (Kumar, 2012).

2.2 Knowledge

Environmental knowledge refers to an individual's awareness level of the issues facing the environment. Researchers have identified two types of environmental knowledge; the first is about identifying the problem and the solutions. The second is about behavior knowledge, which can help in solving the issue (Kumar, 2012). Many studies indicated that environmental knowledge has a significant effect on both attitude and behavior (Braun, Cottrell & Dierkes, 2017). Hines, Hungerford, and Tomera (1987) suggest that people with strong knowledge of environmental issues are more likely to engage in sustainable behavior compared to those who lack this type of knowledge. Nevertheless, Kollmuss and Agyeman (2002) argue that environmental awareness and knowledge only contribute to a small part of sustainable behavior.

Moreover, in a study conducted by Carolan in 2006 on sustainable agriculture, he suggests a strong correlation between knowledge and trust. People validate the truth by associating it with a specific social network. If individuals believe this knowledge comes from people they trust, they will assume that this knowledge is true (Carolan, 2006). For example, mutual trust in knowledge transfer can lead to a decrease in the uncertainty surrounding recycling (Milchrahm & Hasler, 2002).

Furthermore, the gap between behavior and attitude exists partially because of the limited availability of trustworthy information on the sustainable characteristics of products (Getoor, Isley, London, and Tomkins, 2018). The availability of reliable and precise information is a critical factor when a consumer decides to purchase a product. Research shows that many consumers lack awareness regarding products' sustainable features and characteristics. Since sustainable products' benefits are usually communicated to consumers poorly, consumers tend to make uninformed decisions. In addition, consumers become uncertain regarding the choice of buying a specific product when the information about it is limited, inconsistent, or too complex. This uncertainty makes consumers resort to social information; they tend to evaluate the best outcome based on the information obtained from other people (Vermeir and Verbeke, 2006). Besides, consumers are faced with many choices in the store, and they usually decide on which product to buy by judging the labels, which assist them in evaluating both sustainable and non-sustainable products (Carrington, Neville, Whitwell, 2014). A study conducted in 2011 demonstrates that consumers' product assessment significantly changes when precise information becomes available. The study results show that consumers favor sustainable products over non-sustainable ones when non-sustainable products are obligated to include the negative impact of their harmful ingredients on the environment in their labels, and when sustainable products also highlight why their ingredients are eco-friendly (Borin, Cerf, and Krishnan, 2011).

H1: The higher the knowledge a person has regarding sustainable issues, the higher their probability of acting sustainability compared to those who lack such knowledge.

2.3 The influence of society and social norms

Many studies have shown the effectiveness and importance of social influence in convincing consumers to act more sustainable. Social influence happens when consumers' actions, opinions, and feelings are affected by society and the people around them (Abrahamse & Steg, 2013). In particular, social norms, as one of the facets of social influence, can explain the gap between attitude and behavior. Researchers were able to predict pro-environmental behavior, such as recycling, based on social norms (White, Habib & Hardisty, 2019). People tend to act according to social norms; they look at other's people behavior in order to understand social situations and to get a cue on how to act in specific circumstances. Consumers see the behavior of other people in society as an information source, which helps them in maximizing their social behavior effectiveness (Aarts & Dijksterhuis, 2002). They believe that following social norms help them gain approval from their society (Abrahamse & Steg, 2013). Based on the Focus Theory of Normative Conduct, which was developed by Cialdini, the social norm is divided into two categories, namely descriptive and injunctive norms. The descriptive norm refers to the shared behavior and how people act in society, while the injunctive norm is about the behavior that is approved by the community and has to be done (Cialdini et al., 2006). The effect of the descriptive norm was found to be relatively stronger when compared to the injunctive norm. For instance, people knowing that others are recycling is more effective than them knowing that others are expecting them to recycle (Thomas & Sharp, 2013). Therefore, the following hypothesis is proposed

H2: Social norms have a significant effect on the gap between attitude and behavior. If sustainable behavior is common in society, people will act more sustainably, and vice versa.

2.4 (Personal) Values

Over the years, social scientists have been studying the link between values and human behavior. This link is only recently studied by the marketing and consumer behavior investigators (McCarty & Shrum, 1994). Values are a potential factor in determining sustainable behavior, and they can play a significant part in the decision-making process of consumers (Vermeira & Verbekeb, 2007). Many researchers have found that behavior is an outcome of values and attitudes (Fritzsche, 1995). Thus, values are one of the factors that can explain the gap between behavior and attitude. Human values are defined as beliefs that are stable and behaviors that are socially and personally desirable. The Social Adaptation Theory considers values as a category of social cognition, which helps a person adapt to their surrounding environment (Fritzsche, 1995). Values aid both the interest of individuals and collectivities (Hofstede, 2001). They are, to some extent, shared by individuals in the same culture, and they can give an understanding of the psychological similarities among people of the same culture or even different cultures (Grunert & Juhl, 1995). Values tend to be more stable for people who are in a relatively stable environment. Many studies have linked personal values, such as self-control, self-direction, honesty, equality, and benevolence, to sustainable behavior (Vermeira & Verbekeb, 2007). Furthermore, the rewards of engaging in a sustainable behavior, such as recycling, for an individual are not realized immediately; thus, it is expected that principles are driving the engagement in such behavior (McCarty & Shrum, 1994). In addition, people are eager to perceive themselves positively; self-consistency for them is a way to achieve this. Reminding consumers of a situation when their behavior was conflicting with their personal values associated with sustainability can direct them to act in a way that aligns with those sustainable values (White, Habib & Hardisty, 2019).

2.5 Personal norms

Personal norm's differences between people are another significant factor in predicting sustainable behavior and can help in understanding the gap between attitude and behavior (White, Habib & Hardisty, 2019; Bashir, Khwaja, Turi & Toheed, 2019). In the Norm Activation Model by Schwartz (1977), personal norms are defined as a moral commitment in the process of decision making, and behavior occurs in response to these personal norms. The Norm Activation Model aims to understand altruistic behavior; however, it has also been used to explain the pro-environmental behavior (Lindenberg & Steg, 2007). In the context of environmental behavior, personal norms play a role in influencing the person's attitude and behavior concerning actions that improve or preserve the environment's health (Wynveen & Sutton, 2015). Wynveen and Sutton, in their study in 2015, found that people engaged in behavior that mitigates climate change negative effect on the reef ecosystem when their obligation sense is high. Based on the findings of this study, it can be suggested that when the personal norms of individuals are strong, they tend to purchase pro-environmental products because they feel the moral obligation to act accordingly (Kim & Seock, 2019). Two of the factors that activate personal norms are awareness of consequences (AC) and ascription of responsibility (AR). AC suggests that the activation of personal norms happens when the person becomes aware of his actions' consequences towards other people or the environment. AR suggests that personal norms will become active when the person thinks that s/he able to prevent these consequences (Schwartz, 1977). Furthermore, an extension of the Norm Activation Model is the theory of the Value-Belief-Norm (VBN) (Lindenberg & Steg, 2007). The VBN theory suggests that the activation of personal norms leads to sustainable behavior and that values, AC, and AR influence these personal norms (Wynveen & Sutton, 2015; Kim & Seock, 2019). Based on both values and personal norms, the following hypothesis is framed:

H3: High ethical and moral values are associated with higher sustainable behavior, and in turn, a lower gap.

2.6 Habits, sacrifice and commitment

Consumers find behaving in an environment-friendly way difficult due to their routines and habits. Based on research conducted by the University of Kentucky on Swedish consumers, people find it troublesome to try to alter their habits or to adapt to new routines. Routines and habits have a robust psychological power; they tend to make mentally exhausting and complicated choices easy. However, the struggle of adapting new habits and routines is not only psychological; changing a habit requires time. Giving that people tend to have busy lifestyles, saving time, and convenience are critical for many, which makes adapting sustainable habits even more challenging (Isenhour, 2010). Moreover, some consumers are not willing to sacrifice and commit in order to form a sustainable habit, which leads to a wider gap between their intention and their actual behavior. Consumers view the accessibility of ethical and sustainable products as relatively low compared to unsustainable ones, which makes the tradeoff demanding. Committing to long term sustainable habits sometimes means that consumers are giving up some of their purchasing power (higher prices), quality, and convenience, which creates mental and functionality barriers (Carrington, Neville, Whitwell, 2012). Therefore, the following hypothesis is proposed:

H4: Individuals with a busy lifestyle have a higher probability of exhibiting a gap between their attitude and behavior. Adapting new sustainable habits requires time, and those who have a busy lifestyle tend to find acting according to their sustainable attitude challenging.

2.7 Perceived consumer effectiveness

People with a positive attitude towards the environment end up not translating their attitude into action due to their belief that their actions are not very effective in resolving the issue. Thus, researchers have paid a great deal of attention to Perceived Consumer Effectiveness (PCE) as a significant variable in predicting environmental behavior (Kim, 2011). According to Berger and Corbin (1992), PCE is defined as the self-evaluation in an issue context. It is the belief that the individual effort can make a difference in solving a problem (Ellen, Wiener & Walgren, 1991). For example, a person can feel worried about an environmental problem, such as pollution; however, s/he feels helpless when it comes to having the ability to solve this issue through making changes in his or her consumption. PCE plays a significant role as a moderator between attitude and behavior (Berger & Corbin, 1992). People's PCE level has a significant effect on their willingness to behave in a pro-environmental way (Kim, 2011). Assume, for example, that some people are worried about the current environmental state; however, they believe that others, such as the government and corporates, are the ones that can solve this issue, and not them. This group of people tends to score high in attitude, low in perceived consumer effectiveness, and low in their pro-environmental behavior (Berger & Corbin, 1992). It is essential for consumers to be convinced that their behavior can help fight the degradation the environment is facing in order to motivate them to change their behavior (Roberts, 1996). The positive attitude towards the environment can have a more substantial effect on behavior when the person feels that their effort in changing their consumption habits contributes to the improvement of the environment (Kim, 2011). Therefore, in order to induce people to act on their positive attitude towards the environment, they must have a high PCE score (Wesley, Lee & Kim, 2012). Furthermore, PCE is closely associated with the perceived behavior control concept under the Theory of Planned Behavior; the confidence a person has in his or her ability to perform an action influence his or her behavior significantly (Kabaday, Dursun, Alan & Tuğer, 2015). Thus, the hypothesis below is expected:

H5: A high perceived consumer effectiveness level leads to a lower gap between attitude and behavior. People who believe their actions make a difference will behave sustainably and have a lower gap between their attitude and behavior.

2.8 Perceived environmental threats

Most people are present-oriented more than future-oriented, which makes behaving sustainably challenging for many. The results of sustainable behavior are not immediate and hard to measure; acting sustainably requires people to change some of their habits and put aside some of their interests for outcomes that they might not realize in their lifetime (White, Habib & Hardisty, 2019). One way to change their perception is by internalizing the threats and making it relevant to themselves and the people around them (Barr & Gilg, 2006). Researchers have studied the effect of perceived environmental threat on sustainable attitude and behavior in the past two decades. Special attention was drawn to this variable after a study conducted in 1992 by Baldassare and Katz, where they studied the sustainable engagement level when an individual knows that environmental issues cause a threat to their health and welfare. This study concluded that perceived environmental threat has a significant effect on sustainable behavior (Milfont, Duckitt & Wagner, 2010). The perceived environmental threat is defined as an individual's perceived probability of a threat caused by environmental issues. The idea of environmental threat perception is associated with the Protection Motivation Theory (PMT); This theory suggests that threatening events influence behavior. PMT has two main factors, namely vulnerability, and severity. Vulnerability is about the perceived probability of threat exposure, while severity is about the probability of this threat being serious. If the level of both factors is high, individuals tend to have a more influential sustainable behavior (Lim & Moon, 2020). Individuals worried about the threats caused by environmental issues are more likely to recycle, save water, and purchase green products (Baldassare & Katz, 1992). As a result, the below hypothesis is examined

H6: The higher the level of the current and potential threat, the lower the gap between attitude and behavior. If the consequences of climate change threaten people, they tend to behave more sustainably in order to reduce or eliminate the threat.

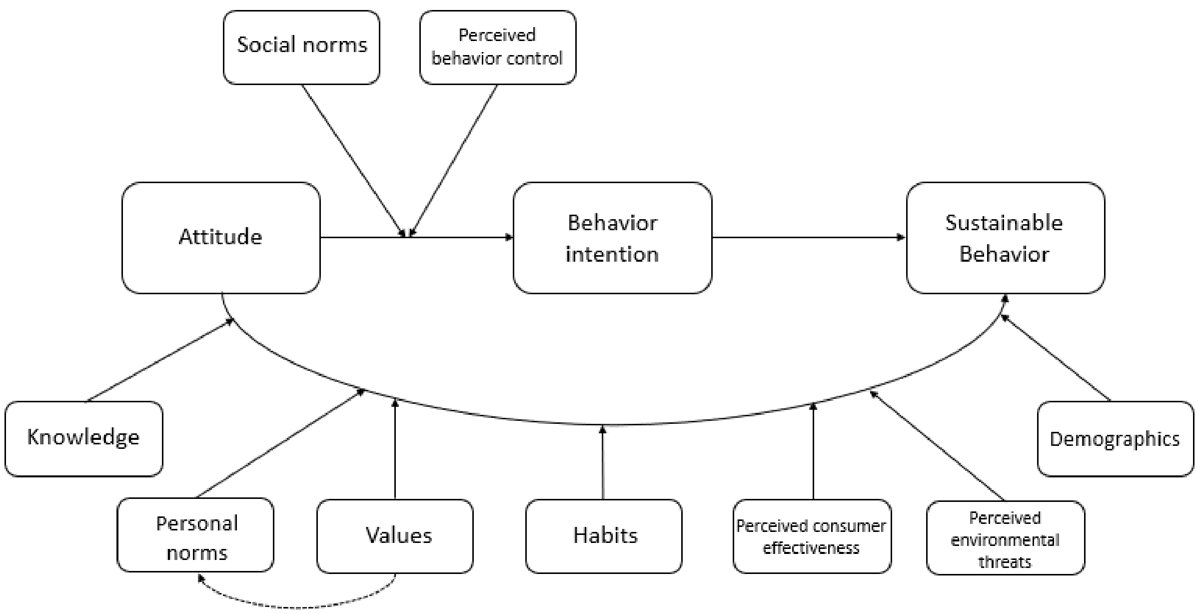
2.9 Demographics

Several researchers have studied socially responsible consumers, and they tried to classify them based on demographics features (Pelsmacker, Driesen & Rayp, 2003). Despite some inconsistency in these studies' results, some researchers managed to find a correlation between the demographics features and pro-environmental behavior (Roberts, 1995; White, Habib & Hardisty, 2019). Several marketers have classified socially responsible consumers as those individuals with high income, and respected occupations (Roberts, 1995). Moreover, gender and education influence both sustainable behavior and attitude. The environmental knowledge women possess is usually less than men; however, women are more concerned about the environment and are more engaged emotionally. They also display more willingness to change for the sake of the environment. Furthermore, longer education is correlated to better awareness of the environmental issues. Nevertheless, longer education does not always lead to environmental behavior (Kollmuss & Agyeman, 2010). Another study by Anderson and Cunningham in 1972 has found that young people are more environmentally conscious; however, the effect of income was not significant, and the effect of education was not apparent. Thus, the below hypotheses are anticipated.

H7: Women exhibit a lower gap between their sustainable attitude and behavior. As previously discussed, women tend to be more concerned and emotionally engaged with environmental issues.

H8: The educational level does not have a significant effect on reducing the gap between attitude and behavior. Previous studies have shown conflicting results regarding education; as previously discussed, education can lead to higher environmental knowledge; however, it does not necessarily lead to sustainable behavior.

Figure 1: Survey's conceptual framework



3. Data

In order to study the attitude-behavior gap in sustainability, a survey was conducted at the end of 2019. Erasmus School of Economics and KidsRights Foundation collaborated in creating the survey, and the mean of distribution was Facebook. After preparing and cleaning the data, 6070 responses remained; the respondents in this sample come from 12 countries, namely the United States (USA), Mexico (MEX), Brazil (BRA), the Netherlands (NLD), the United Kingdom of Great Britain and Northern Ireland (GBR), South Africa (ZAF), Egypt (EGY), Liberia (LBR), Nigeria (NGA), India (IND), Indonesia (IDN), and the

Philippines (PHL). The responses contain only people who have English as their primary default language on Facebook. The survey consists of questions related to demographics, ethics, morals, personal behavior, and climate change. Furthermore, the survey is divided into two parts; the first part is obligatory, while the second part is optional. A list of these questions and the short names used throughout the analysis are displayed in Appendix A.

In addition to the survey, country-level cross-sectional data is obtained; the corresponding data contains the latest available statistics on each of the 12 countries. This data can provide a broader understanding of some attributes that can influence people's behavior. Some of these attributes can also give an indication of people's behavior in a specific country. The data is obtained from various sources, namely the World Bank, the United Nations Development Program, and Our World in Data (a scientific publication handled by researchers from Oxford University).

4. Methodology

Machine learning methods are used in this paper in order to analyze the survey data. The problem at hand is a classification one, which means the dependent variable to be predicted is a categorical one (whether the respondent exhibits a gap between his/her attitude and behavior or not). Many machine learning techniques are available to solve the problem at hand; choosing which method to use can depend on the interpretability level and the accuracy rate desired. First, logistic lasso regression is applied; this method provides an improvement over logistic regression by combining regularization and variable selection. Logistic lasso regression leads to an improvement in the interpretability level of logistic regression by only including the important variables for prediction in the model while shrinking the unimportant ones to zero.

In an aim to improve the accuracy of prediction further, a more sophisticated machine learning method is going to be applied next, namely random forest. Applying random forest to the dataset might improve the accuracy rate; however, this will be at the expense of the level of interpretability. Random forest has the ability to handle different data types, high dimensions, interactions that are complex, and non-linearity in the data (Stekhoven & Bühlmann, 2012). It is chosen over decision trees because the latter method suffers from accuracy issues and high variance. Random forest also provides an improvement over other ensemble methods, namely bagging, as it has the ability to decorrelate the trees. Bagging, however, might result in highly correlated trees, which also might lead to accuracy issues. Furthermore, when comparing random forest to other powerful classification techniques, such as support vector machine, applying random forest to the data at hand resulted in a relatively higher prediction accuracy rate. In this section, the steps taken to prepare the survey's data are presented; this is followed by explaining logistic lasso regression, random forest, and the results' validation methods.

4.1 Survey data Preparation

4.1.1 Missing values and data quality

Initially, 10900 respondents took the survey, and their responses contain 173,585 missing values. These missing values can be a result of people taking only the first part of the survey or quitting the survey at any stage without finishing it or simply due to a technical error. Many of the machine learning methods cannot handle missing data; therefore, to perform these methods, a complete set of data is needed. However, removing all these missing values from the survey will cause a significant loss of relevant information. Thus, when cleaning the data from missing values, the aim was to retain as many responses as possible to avoid losing this valuable information. Consequently, the missing values were cleaned and handled in two steps. First, only two missing values were allowed in the first part of the survey, meaning that only responses with approximately 90% progress in the first part were included in the analysis. Besides, no missing values were allowed in the dependent variables (the first part of the survey contains the dependent variables). Second, given that missing values can cause significant issues in the analysis, imputation is applied in order to replace these missing values with predicted values.

It is crucial to choose the imputation method carefully in order to obtain proper and accurate estimates of the missing values. Many methods, such as K-nearest neighbors, can be used for imputation; however, most of them can only be applied to either categorical or continuous data. Given the nature of the survey data at hand, an imputation method that can handle the existence of both types of data at the same time is needed. Two of the most used methods for such data is random forest and the multivariate imputation by chained equations (MICE). MICE is a parametric method that requires specific data distribution assumptions, such as assuming the data is normally distributed, which can lead to some uncertainty. Besides, the method depends on parameter tuning; if the parameter tuning is not done accurately, it will affect the result of the imputation significantly. As a result, random forest is chosen over MICE in order to impute the missing values. Random forest is a non-parametric method that can handle different types of data at the same time. Besides, unlike all the previously mentioned methods, it has the ability to handle nonlinearly in the data, and also it can handle complex interactions (Stekhoven & Bühlmann, 2012). It also takes into account the relationship between all the variables. Random forest was proven to outperform all other imputation methods regardless of the data type, the missing values amount, and dimensionality of the data (Stekhoven & Bühlmann, 2012; Kokla, Virtanen, Kolehmainen, Paananen & Hanhineva, 2019).

Before performing imputation, a data matrix X is required with dimensions $n \times p$, where $X = (X_1, X_2, \dots, X_p)$. Then, for a random variable X_s that contains missing values, the data can be split into four parts. The first part is variable X_s observed values, which is represented by y^s_{obs}. The second part is the missing data in the same variable X_s, which is represented by y^s_{mis}. The third part is x^s_{mis}, which represents all other variables with i^s_{mis} observations, where $i^s_{mis} \subseteq \{1, \dots, n\}$. The last part contains all other variables' observations $i^s_{obs}$where $i^s_{obs} = \frac{\{1,\dots,n\}}{i^s_{mis}}$, which is represented by x^s_{obs}.

To start the imputation process of missing values for each variable, Random forest is trained with a dependent variable y^s_{obs} and independent variables x^s_{obs}. Then the trained model is applied to x^s_{mis} in order to predict y^s_{mis}. This process is repeated several times until a stopping criterion γ is reached; this stopping criterion is reached when an increase starts to happen for the first time in the difference between the old imputed data and the new one.

This difference for the continuous (N) is defined in equation 1

$$\Delta N = \frac{\sum_{j \in N} (X_{new}^{imp} - X_{old}^{imp})^2}{\sum_{j \in N} (X_{new}^{imp})^2} \qquad (1)$$

The difference for categorical variables (F) is defined in equation 2 below

$$\Delta F = \frac{\sum_{i=1}^{n} I_{X_{new}^{imp} \neq X_{new}^{imp}}}{\#NA} \qquad (2)$$

In equation 2 above, NA represents the number of categorical variables' missing values.

Finally, an out of bag (OOB) error is obtained for each of the imputed variables. This OOB error is a very close approximation of the actual imputation error; on average, the OOB error rate only deviates from the actual imputation error by only 10% to 15%. The OOB error is explained further under section 4.3.2. In appendix B, the OOB error estimate per imputed variable is presented.

In addition to handling the missing values, the following question is used to eliminate inadequate quality Responses: "Did you understand all the questions?". If the answer selected is "No", the corresponding response is eliminated from the data sample. This step is necessary because if the respondent does not understand all the questions, it is unclear if they should be included. Accordingly, removing these responses from the dataset will lead to better data quality.

4.1.2 Defining the dependent variable

The dependent variable in this study is defined based on two questions in order to account for attitude and behavior. The first is "Do you think climate change is a serious problem?". This question defines the attitude and is used to segment people into two categories, namely people who display attitude towards sustainability and people who do not. The possible answers to this question are based on a Likert scale with values from one to seven with one corresponding to "Not at all", and seven corresponding to "Very much". Those who chose an answer of four or above are considered to have an attitude towards sustainability.

People who chose an answer below four are considered to have no attitude, and, therefore, eliminated from the analysis. The second question is “In the past 12 months, did you and/or your family do any of the following?”. This question defines behavior; respondents are considered to behave according to their attitude when they choose the following two responses simultaneously: “Avoided eating meat or reduced my meat consumption*.”* and *“*Reused plastic shopping bags and/or recycle plastic bottles.”. This category of people is denoted by “Yes”. All other respondents are donated by “No”, which means they are not acting sustainably, and that they exhibit a gap between their attitude and behavior.

4.1.3 Imbalanced dependent variable

Another issue is that the dependent variable is imbalanced to some extent. Imbalanced data means that one of the dependent variable classes is rare in the dataset sample (Lunardon, Menardi & Torelli, 2014). In the dataset of this study, people who behave sustainably (denoted by “Yes”) consist of 39% of the total responses, while people who act otherwise (denoted by “No”) consist of 61% of the total Responses. Given this distribution of the dependent variable classes, the performance of the models used for prediction in this study might yield a lower prediction accuracy rate than desired. A solution for this is balancing the dependent variable using both under-sampling and random over-sampling techniques. Under-sampling randomly deletes data from the majority class, while oversampling balances the dependent variable by creating new random samples from the minority class based on the attributes of this class. Applying these two methods can make the prediction models more reliable and can improve the prediction accuracy rate.

4.2 Logistic lasso regression

In this study, Logistic Lasso regression is used to predict whether a person is going to behave sustainably. Logistic lasso applies both regularization and variable selection, leading to a better accuracy rate and a better interpretation of the obtained results. The model is a combination of logistic regression and lasso regression.

4.2.1 Logistic regression

Given the nature of the dependent variable in this study, binary logistic regression is used. The model assigns a probability to the dependent variable, meaning the dependent variable's probability to belong to a specific category. The odd ratio natural logarithm (Logit) is the central mathematical idea behind logistic regression. The value of the odds is between 0 and ∞. A low probability takes a value close to zero, while a high one is assigned a value close to ∞. Commonly, the model tests the hypotheses around the relationship between the dependent variable, given that this variable is categorical, and the independent variables; the independent variables can be continuous or categorical (Pendg, Lee & Ingersoll, 2010). The binary nature of the dependent variable in this study takes values either 1 or 0. The value one is assigned to people who act sustainably while the value 0 is assigned to people who behave otherwise.

Logistic regression is defined in equation 3 below.

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} \qquad (3)$$

In equation 3 above, $p(x)$ is the probability and returns an output between 0 and 1. β_0 and β_1 are the coefficients.

The coefficients β_0 and β_1 are unknowns, and they need to be estimated, and in order to do this, the model is trained using the maximum likelihood method. This method aims to find the coefficients' values that maximize the likelihood (Bewick, Cheek & Ball, 2005). The coefficients are estimated in a way that the predicted probability $\hat{p}(X)$ takes a number that is as close as possible to the observed value. Thus, when substituting $\widehat{\beta_0}$ and $\widehat{\beta_1}$ in equation 3, a number close to 1 is obtained for those who behave sustainably and 0 for those who behave otherwise.

The maximum likelihood can be illustrated in equation 4 below.

$$l(\beta_0 + \beta_1) = \prod_{i:y_i=1} p\,(x_i) \prod_{i\prime:y_{i\prime}=0} (\,1 - p(x_{i\prime})) \qquad (4)$$

The choice of $\widehat{\beta_0}$ and $\widehat{\beta_1}$ aims to maximize the likelihood equation above.

4.2.2 Lasso regression

The dataset used in this study contains many predictors; thus, applying a shrinkage method can improve the prediction accuracy and interpretation of the logistic regression model. Shrinking methods can reduce variance; however, in return, it introduces some bias. One of the most used shrinkage methods is lasso regression.

The Lasso is defined using the loss function below.

$$L(b_1,\dots,b_m) = \underbrace{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{m} x_{ij}\, b_j)^2}_{\text{Regression term}} + \underbrace{\lambda \sum_{j=1}^{m} |b_j|}_{\text{Penalty term}} \qquad (5)$$

In equation 5 above, b_j is an unknown regression weight for predictor $j = 1,\dots,m$. x_{ij} are elements in the predictor variables x matrix $n \times m$. y_i is a notation for the response variable for elements $i = 1,\dots,n$. λ is the penalty parameter, which is positive.

Lasso performs variable selection by penalizing the coefficients; the model sets some predictors to be precisely zero. For this to happen, λ needs to be large enough. As a result, the obtained model will contain a smaller number of variables, making it less complex and easier to interpret.

4.3 Random Forest

Logistic lasso regression returns results that are highly interpretable; however, the accuracy rate might not be optimal. Therefore, random forest is also going to be used in this study; the method can result in higher accuracy rate at the expense of the interpretability level. Random forest is one of the most robust machine learning techniques; it is a tree-based method that can provide an accurate classification prediction to the problem at hand by creating several trees. Random forest for classification problems predicts the classes by assigning every observation in the training data set to the most commonly occurring class. A classification tree uses a recursive binary splitting (top-down) approach. It commences at the top of the tree, where all the responses are in one region, and then it continues by splitting the predictor space. This splitting process is repeated in order to identify the most suitable predictor; however, this time, the splitting happens to the regions that were identified earlier instead of the predictor space. Random forest works by building several decision trees using the bootstrapped training dataset. Bootstrap is a resampling method, which samples the data set with replacement (James, Witten, Hastie & Tibshirani, 2013).

Furthermore, random forest works in a way that decorrelates the trees, which leads to a better prediction accuracy compared to other tree-based methods. This means that m number of predictors are selected randomly out of the complete set of p predictors as candidates whenever a split is taken into account, where $m \approx \sqrt{p}$. This approach is particularly useful when the dataset contains a strong predictor. Typically, when such a strong predictor exists among other predictors that are not as strong, this predictor will always be considered first and then used at the top of each split, meaning that the trees' similarity will be very high. As a result of this strong correlation between the trees, the model will suffer from high variance and less accuracy. Random forest overcomes this issue by forcing the algorithm to ignore the strong predictor in the sample, which gives moderately robust predictors a chance. Thus, the model ends up with lower variance and becomes more reliable (James, Witten, Hastie & Tibshirani, 2013).

4.3.1 Out of Bag error rate (OOB)

Random forest provides a straightforward estimation of the test error. Similar to cross-validation, random forest divides the training sample into two parts, the first sample consists of 2/3 of the observation, and used to fit the model, and the rest of the observations are not used in the initial prediction; the remaining third of the observations are called out of bag. The prediction of a single observation happens by using the trees, where this particular observation was OOB. This will result in $B/3$ predictions for this specific observation, and then the majority vote is taken to obtain a single prediction for the observation. Subsequently, a single out of bag prediction is obtained for the response; the process is repeated for all the observation in the training sample (James, Witten, Hastie & Tibshirani, 2013).

4.3.2 Interpretation of random forest

Random forest model provides a high accuracy rate; however, in return, interpretability becomes challenging. Thus, the method is considered a black box method as it is not feasible to examine each tree and understand the prediction process. One way to solve this is by using variable importance measures. These measures provides the opportunity to understand the most important variables for prediction. Gini index and mean decrease in accuracy can help obtain an overall summary of each variable's importance. The first approach measures the decrease in node purity using the Gini index. The node purity decrease is calculated by measuring how much the Gini index has decreased as a result of the splits over a specific variable, and then this decrease is averaged over all the B number of trees. The second approach measures the average decrease in the OOB sample prediction accuracy when excluding a predictor from the model (James, Witten, Hastie & Tibshirani, 2013).

4.4 Validation of the results

In this study, the results are validated using cross validation. Two types of cross validation are used when performing both logistic lasso regression and random forest, namely the validation set and k-fold approaches. The validation set approach involves randomly splitting the dataset into a training set and a test set; both random forest and logistic lasso regression are trained using the training set, and then each trained/fitted model is applied to the test set in order to predict the dependent variable. Finally, a test error is obtained, which indicates how well the model performed.

Furthermore, k-fold cross validation is also used when fitting random forest and logistic lasso regression using the training set. This approach is mainly used in order to tune the hyperparameters of the models. k-

fold cross validation approach works by dividing the training data set into k groups; the splitting of the training set is random, and the groups have a similar size. This method works by leaving out the first group as a test set, and then training the model using the remaining of the groups, $k-1$. The method is repeated k number of times, and each time a new group is left out as a test set. For logistic lasso, this method is used to determine the best lambda, while, for random forest, it is used in order to tune the number of trees and the number of the randomly selected variables m. Tuning the parameters instead of using the default values can increase the accuracy of the models' prediction (James, Witten, Hastie & Tibshirani, 2013).

4.5 Confusion matrix

In order to get more insights into how well the classification model performed, a confusion matrix is created. It is a table that indicates how well the classification model performed by returning the number of responses that were classified (in)correctly. The confusion matrix returns some measures that help assess the model, such as accuracy, specificity, and sensitivity. The accuracy gives a percentage that indicates how the model succeeded overall in classifying the dependent variable. Sensitivity measures the ability of the model to determine people who act sustainably correctly. In contrast, specificity measures how well the model did in predicting people who do not act sustainably correctly.

5. Results

5.1 Survey

The aim of this research is to study the attitude behavior gap in sustainability. As mentioned earlier, the survey contains 6070 total responses after cleaning the low-quality data. The respondents come from 12 different countries, and they answered questions that evaluates different aspects that potentially can lead to the gap, such as demographics, social norms, values, etc. All the 6070 respondents exhibited an attitude towards sustainability and climate change as they rated climate change as an important issue (they gave it a score of four or above). However, only 39% of the people (before balancing the data) exhibited sustainable behavior; those people have reduced their consumption of meat and they also used less plastic or recycled the plastic they used.

One of the first determinants is demographics (table 1 and 2); it is measured based on gender, age, education, country of residence, and financial status. Gender is quite balanced with 54% females, 45% males, and only 1% are out of this spectrum. Furthermore, over 60% of the respondents are between the age of 16 and 21, and less than 1% are above 24. Most of them have either an education up to 18 or a university degree with 20% and 37%, respectively. They are also financially stable to some extent as they claim to have enough money for now and the future.

Additionally, the sort of knowledge a person has was considered in this survey (table 3). When respondents were asked about their relative knowledge compared to other young people in their community, 87% of them claimed to know much more than their peers. Moreover, they were asked about their source of knowledge, and who or what taught them the most about climate change and sustainability. News leads the source of knowledge with 85%, followed by people outside their inner circle (other than their teachers, parents and friends), friends and movies with 72%, 64%, and 63%, respectively. Besides, trust is also used as part of the learning process (table 3). The most trusted entities are scientists, followed by media with 56% and 40%, respectively.

Perceived consumer threats and values were also examined (Table 4 and 5). The perceived consumer threat was examined through asking about the number of days with extreme weather the respondent experienced and their expectations of the weather in the future. Most of the respondents experienced extreme weather conditions with 39% experienced more than 20 days of extreme weather. Almost 62% of the survey takers expect to experience more extreme weather in the future. Also, when it comes to values, respondents are characterized with high strong values; such as honestly, law obedience, helping others without expecting something in return, etc.

Table 1: Demographics

Age	% Frequency	Country based	% Frequency	Education	% Frequency	Gender	% Frequency
13	1.8	IND	42.8	University	37.2	Female	54.3
14	4.0	GBR	14.3	Edu up to age 18	20.1	Male	44.6
15	7.3	IDN	11.1	Edu after 18 (non-uni)	10.0	Other	1.1
16	11.3	ZAF	7.0	Edu up to age 16	20.8		
17	13.3	MEX	5.0	Edu up to age 14	9.4		
18	11.4	BRA	4.8	Edu up to age 12	2.1		
19	12.2	NGA	4.8	No formal EDU	0.2		
20	10.9	USA	3.9				
21	10.1	NLD	3.4				
22	6.8	EGY	1.0				
23	6.6	PHL	1.0				
24	4.0	LBR	0.9				
Older than 24	0.2						

Table 2: Income

Variable	Possible Values	Min	Max	Mean	SD
Enough money	1=Strongly disagree, 7=Strongly agree	1.0	7.0	4.1	1.7
Bills	1=Strongly disagree, 7=Strongly agree	1.0	7.0	3.9	1.8
Future money	1=Strongly disagree, 7=Strongly agree	1.0	7.0	3.9	1.8

Table 3: Knowledge

Variable	Possible Values	Min	Max	Mean	SD
Learn teacher	1=Nothing, 7= Very much	1.0	7.0	4.5	1.9
Learn parents	1=Nothing, 7= Very much	1.0	7.0	4.1	1.9
Learn friends	1=Nothing, 7= Very much	1.0	7.0	4.6	1.8
Learn other	1=Nothing, 7= Very much	1.0	7.0	5.0	1.7
Learn news	1=Nothing, 7= Very much	1.0	7.0	5.8	1.5
Learn movie	1=Nothing, 7= Very much	1.0	7.0	4.6	1.9
Trust politicians	1=Strongly disagree, 7=Strongly agree	1.0	7.0	2.3	1.3
Trust media	1=Strongly disagree, 7=Strongly agree	1.0	7.0	3.4	1.7
Trust scientists	1=Strongly disagree, 7=Strongly agree	1.0	7.0	6.0	1.2
Trust famous people	1=Strongly disagree, 7=Strongly agree	1.0	7.0	3.4	1.6

Table 4: Perceived environmental threats

Experience weather	% Frequency	Expect weather	% Frequency
Zero days	3.3	Fewer days with extreme weather	17.2
1-5 days	17.3	Same number of days with extreme weather	20.6
6-10 days	18.1	More days with extreme weather	62.2
11-15 days	12.0		
16-20 days	10.1		
More than 20 days	39.2		

Table 5: Values

Variable	Possible Values	Min	Max	Mean	SD
Give up something	0=Not at all willing, 10=Very willing	0.0	10.0	6.5	3.2
Risk taker	0=Not at all willing, 10=Very willing	0.0	10.0	6.8	2.9
Help others	0=Not at all willing, 10=Very willing	0.0	10.0	5.7	3.5
Hide mistakes	1=Definitely disagree, 7=Definitely agree	1.0	7.0	5.0	1.6
Steal	1=Definitely disagree, 7=Definitely agree	1.0	7.0	6.0	1.6
Gossip	1=Definitely disagree, 7=Definitely agree	1.0	7.0	5.3	1.7
Obey laws	1=Definitely disagree, 7=Definitely agree	1.0	7.0	5.7	1.5
Eavesdrop	1=Definitely disagree, 7=Definitely agree	1.0	7.0	5.4	1.8
Work rights	1=Definitely disagree, 7=Definitely agree	1.0	7.0	2.1	1.8

5.2 Country-level data

In addition to the survey data, country-level data was obtained to better understand the characteristics of the 12 participating countries in the survey, and in turn, understand the attitude-behavior gap further. The descriptive statistics of these variables for all the 12 countries are presented in table 6 below. Besides, a correlation matrix is constructed in figure 2 below to evaluate the relationship between these ten variables, if any.

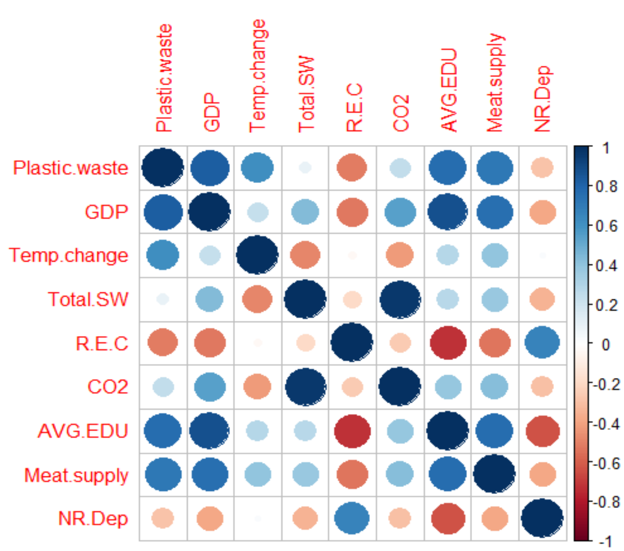
The correlation matrix shows that most of the variables are moderately to highly correlated. For instance, there is a moderate positive correlation between the amount of meat supplied and plastic waste. In contrast, a high positive correlation exists between the amount of CO_2 emission and total solid waste. Furthermore, the obtained country-level data shows some distinction between the global south, such as NGA, IND, and LBR and the global north countries, such as the USA, NL, and GBR. NGA, IND, and LBR have the lowest GDP per capita, and they are also characterized by the lowest mean years of schooling compared to the other 12 countries. These countries also have the lowest renewable energy consumption as a percentage of their total final consumption of energy. Besides, their meat supply per capita is relatively one of the lowest.

In contrast, the USA, GBR, and NL have the highest GDP per capita and mean years of schooling. Furthermore, the USA has the highest meat supply per capita among all the 12 countries; this is followed by BRA, GBR, and NLD, respectively. The Netherlands has witnessed relatively the highest temperature change, and it ranks second lowest in natural resource depletion. Moreover, the USA has the highest CO2 emission and total waste, while LBR ranks the lowest in both variables compared to the other 12 countries.

Table 6: Descriptive Statistics

Variable	Short name	Min	Max	Mean	SD
GDP per capita in US dollars	GDP	621.9	65280.7	16683.9	22801.2
Mean years of schooling (years)	AVG.EDU	4.7	13.4	9.0	2.8
CO2 emission in kiloton	CO2	1386.0	5006302.0	869588.0	1447247.0
Natural resource depletion (% of GNI)	NR.Dep	0.2	19.2	3.2	5.2
Plastic waste per capita (KG)	Plastic.waste	0.0	0.4	0.2	0.1
Meat supply per capita (KG)	Meat.supply	3.8	124.1	50.7	39.3
Temperature change (Celsius)	Temp.change	0.7	2.1	1.3	0.4
Total solid waste (tones)	Total.SW	564467.0	258000000.0	62276725.0	76676172.0
Renewable energy consumption (% of total final energy consumption)	R.E.C	5.7	86.6	30.8	28.7

Figure 2: Correlation matrix for the country-level cross-sectional data



5.3 Logistic lasso regression

Logistic lasso regression is the first prediction model used in this analysis in an aim to predict sustainable behavior and understand the gap. Before performing the model, the survey and the country-level data are merged. The obtained data is split into a 70% training set and a 30% test set; the model is fitted using the training sample and validated using the test sample. In addition, logistic lasso regression penalty term λ is determined, and to do this, k-fold cross validation with $k = 10$ is used. As a result, $\lambda_{min} \approx 0.0051$ is chosen; this value of λ returns the lowest RMSE. The outcome of λ is then used to fit the model; selected coefficients' estimates are displayed in figure 3 and table 7 below, while the full results can be seen in appendix C. The model predicts sustainable behavior with an accuracy rate of approximately 62%.

Figure 3 and table 7 below mostly show the most important variables in predicting sustainable behavior. Figure 3 shows the top ten variables that contribute to an increase in sustainable behavior, and a reduction in the attitude-behavior gap, while table 7 displays a mix of the most important variables that contribute to exhibiting sustainable behavior positively or negatively. Table 7 also displays some country-level attributes that influence the gap. Figure 3 shows that being a resident in the Netherlands will lead to the highest sustainable behavior increase. In addition, experiencing 11 to 20 days of extreme weather leads to a reduction in the gap between sustainable attitude and behavior. Furthermore, the higher the knowledge an individual has compared to their peers, the higher their probability of acting sustainably. Also, some other coefficients lead to a higher probability of translating attitude into action, such as being willing to help others without expecting anything in return, giving up something now for future benefits, never gossip, and not following a routine (Appendix C).

Besides, according to table 7, having no formal education affects behavior towards sustainability the most as it decreases the chance of acting according to the individual's positive attitude towards sustainability by approximately 1.7 units. Moreover, people who are residents of Egypt, Nigeria, or Indonesia are expected to have a lower probability of behaving according to their attitude than any other country in the survey.

Coming from these three countries reduces the log odds of sustainable behavior by approximately 1.11, 0.99, and 0.95 units, respectively. It is worth noting that many variables have been shrunk to zero, which means logistic lasso regression considers these variables insignificant for prediction. These variables include gender, social norms (perc.young.spend), habits (hard.worker) , and some levels of education, namely education up to age 12, education up to age 14, and education up to age 18 (Appendix C). Furthermore, some country characteristics from the country-level data have a significant effect on predicting the attitude behavior gap. For example, the higher the GDP, the higher the chances of people exhibiting more sustainable behavior. On the contrary, the higher the percentage of natural resource depletion in a country, the higher the chances that individuals in this particular country will act less sustainable, leading to a higher gap between their positive attitude and their actual behavior.

Figure 3: Logistic lasso variable importance (positive coefficients)

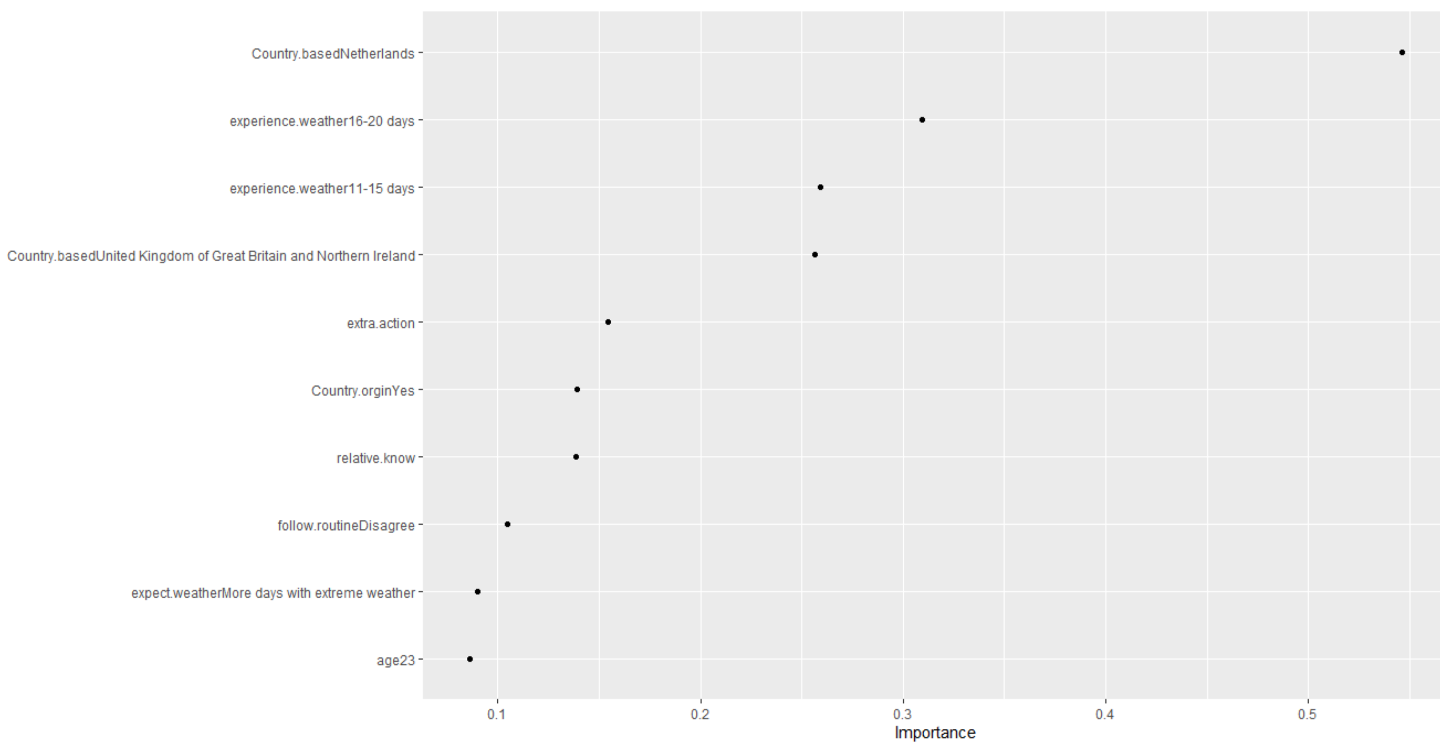


Table 7: Logistic lasso regression coefficients

Coefficient	Estimate	Coefficient	Estimate
No formal education	-1.689	Country based Netherlands	0.546
Country based Egypt	-1.113	Experience weather16-20 days	0.309
Country based Nigeria	-0.994	Experience weather11-15 days	0.259
Country based Indonesia	-0.945	Country based GB and Ireland	0.256
Country based South Africa	-0.502	Extra action	0.154
Country based USA	-0.365856215	Relative know	0.139
NR.Dep	-0.023	GDP	0.005

5.4 Random Forest

The accuracy rate obtained from the logistic lasso model is not optimal; applying a more sophisticated machine learning technique can improve the prediction accuracy rate and provide better results. Therefore, in this section, random forest is going to be applied to the survey data. Similar to logistic lasso regression, the training sample is going to be used in order to fit the model, and the test set is going to be used to validate the performance of the model; they contain 70% and 30% of the data, respectively.

First, random forest hyperparameters are tuned, namely the number of randomly selected variables (m) and the number of trees. The best m and the best number of trees that give the highest prediction accuracy rate are chosen; the result is validated using 10 folds cross validation. The best number of the randomly selected variables is 8 (figure 4), while the optimal number of trees to grow is 600. After fitting the model using the training sample and the hyperparameters, and also after validating the results using the test set, an overall prediction accuracy rate of 82% is obtained. Moreover, according to the confusion matrix, the model was able to predict respondents who act sustainably (according to their attitude) correctly with a sensitivity rate of approximately 80%, while it was able to classify those who act otherwise (in contrast to their attitude towards sustainability) with a specificity rate of approximately 84%.

Giving that random forest is a black box, and interpreting its results is challenging. The variable importance based on the mean decreased accuracy and Gini index is obtained. Figure 5 displays the top 30 variables based on their importance. Based on both the mean decreased accuracy and the Gini index, the top four most important variables are the country of residence (country.based), the age, the number of days of extreme weather the respondent experienced (experience.weather), and trusting famous people (trust.famousppl).

For the mean decreased accuracy, this is followed by trusting politicians (trus.politicians), and having enough money for the present and the future (bills, future.money, and enough.money), respectively. However, for Gini index, this is followed by, the level of education (education), trusting media (trust.media), enough money to pay the bills (bills), and not worrying about the money in the future (future.money), respectively.

Figure 4: Determining the number of randomly selected predictors

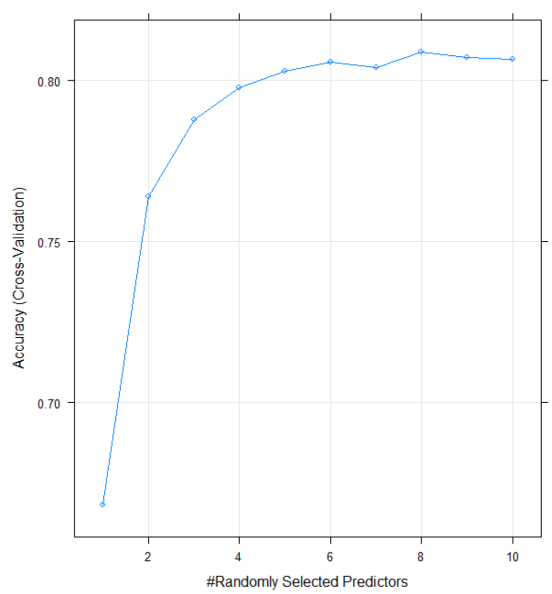
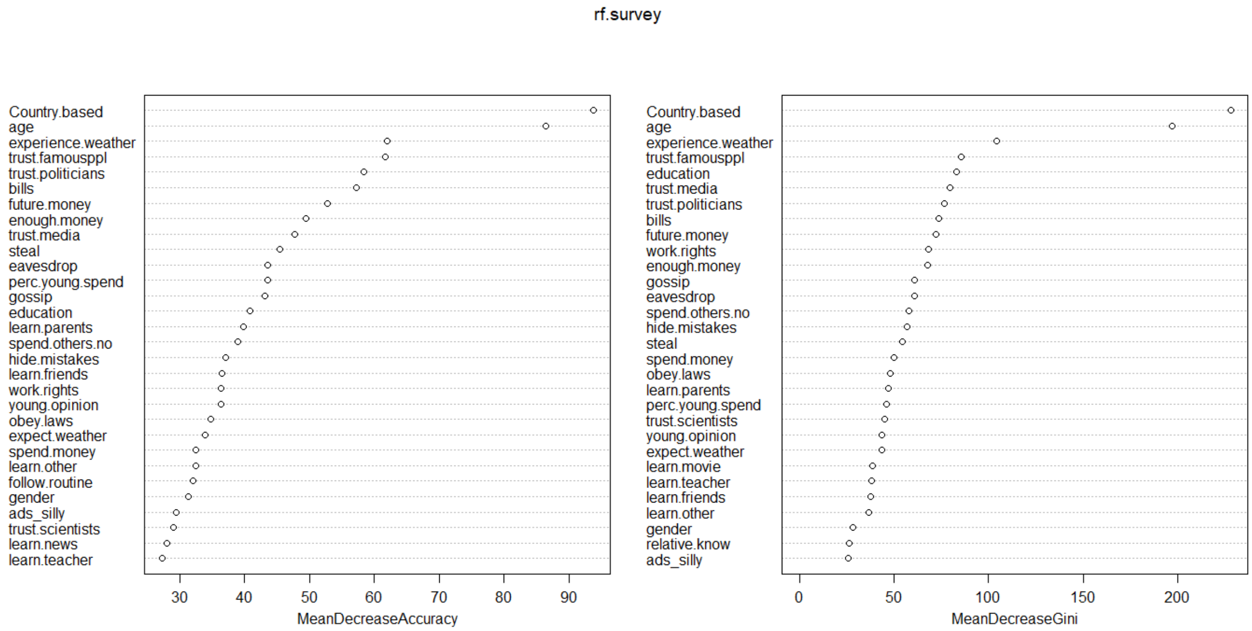


Figure 5: Random Forest Variable importance



6. Discussion

The survey results confirm the existence of the gap between sustainable attitude and behavior. Many attributes, such as demographics, have played a significant role in the existence of such a gap. Based on the random forest model, the country of residence is the most essential variable in determining the gap; it also has one of the highest effects on the probability of people acting sustainably, according to the logistic lasso regression model. Respondents from most global south countries in the survey, such as Nigeria and Egypt, tend to have a higher probability of exhibiting a gap between their sustainable attitude and behavior than those from the global north. Besides, age was the second-highest important variable in the random forest model; however, the effect of gender on the gap, according to the logistic lasso model, is ambiguous. Also, whether the respondent can pay his/her bills tend to affect their probability of acting sustainably. Logistic lasso model suggests that the higher the capability of people to pay their bills, the higher their chances of acting sustainably. The country-level data also support this; people in countries with relatively higher GDP will have a higher probability of behaving sustainably. Furthermore, unlike Kollmuss and Agyeman (2010) claim in their study, gender was found to have no significant effect on the gap, which means there is not enough evidence to support the hypothesis (H7) that women exhibit a lower gap.

Furthermore, in contrast to the literature review and hypothesis number eight (H8), education is one of the most critical determinants of the gap; people with no formal education will have a higher chance of exhibiting a gap between their attitude and behavior. However, whether or not longer education is correlated to the gap is ambiguous. It is worth noting that based on the country-level data, global north countries are characterized by higher educational attainment, while global south countries tend to have relatively lower mean years of education. However, according to the logistic regression model, this country-level variable does not have a significant effect on the gap.

Moreover, the number of days of extreme weather the respondent experienced is also one of the most important variables contributing to the gap. The logistic lasso model shows that experiencing extreme weather is associated with a lower gap. People who experienced between 11 and 20 days of extreme weather in the last 12 months have a higher probability of exhibiting a relatively lower gap between their attitude and behavior. It is worth noting that experiencing 16 to 20 days of extreme weather has a higher positive effect on the gap than experiencing 11 to 15 days of extreme weather only. Also, experiencing zero-days or more than 20 days of extreme weather has no significant effect on the gap. Besides, respondents who are expecting more days of extreme weather in the future tend to have a higher chance of having a lower gap. This supports the hypothesis (H6) about the perceived consumer threats; those exposed to a relatively higher threat (extreme weather) tend to behave more sustainably.

Additionally, hypothesis one (H1) regarding the individual's knowledge and its association with the gap is supported mainly based on the logistic lasso regression model. The higher the relative knowledge of an individual regarding climate change, the higher their probability of translating their attitude into action. Furthermore, the social norms hypothesis (H2) cannot be fully supported. The logistic lasso model shows that the percentage of other young people in a society willing to spend money to tackle climate change does not affect the respondent's decision to act sustainably. It is worth noting that one of the most important variables in determining sustainable behavior according to the random forest model is trusting famous people and politicians. These variables can examine the social influence of role models on people's desire to act more sustainable. Based on the logistic lasso model, people who trust famous people are less likely to behave sustainably. Besides, trusting scientists is associated with an increase in the log-odds of sustainable behavior.

Furthermore, having high moral and ethical values does not always lead to a lower attitude-behavior gap. People who are willing to help others without expecting anything in return or believe that men and women should have equal job rights or claim that they never gossip have a higher probability of showing sustainable behavior. However, other examined personal norms and values, such as obeying laws and stealing, show some mixed results; some of these values either contribute negatively to the gap or have no significant effect. Thus, hypothesis two (H3) is not fully supported due to a lack of evidence. Hypothesis four (H4) regarding the negative effect of a busy lifestyle on the gap is also rejected. Random forest model considers variables linked to routines, habits, and busy lifestyles, such as working hard and following a routine, not important in predicting the gap. The logistic lasso model also considers working hard as an insignificant variable, while it considers those who do not follow a routine to have a higher chance to behave sustainably. The ‘follow routine’ variable, however, is not very informative as some habits and routines can be very sustainable, such as taking shorter showers.

Moreover, perceived consumer effectiveness presented in hypothesis five (H5) is supported. The higher the individual's willingness to give up something now for future benefits and to spend money to tackle environmental issues (even if others do not), the higher their probability of exhibiting a relatively lower gap. It is worth noting that supporting H5 also confirms the importance of perceived behavior control. This shows the individual's belief in their ability to perform the actions needed to tackle environmental issues.

6.1 Recommendations for policymakers and marketers

Policymakers and marketers can use this study's results to tackle the attitude-behavior gap phenomenon and get people to behave more sustainably. The study indicated that the country of residence is an essential determinant of the gap; this can help marketers and policymakers segment and target people. To promote sustainable products and services, marketers can target the countries where the gap tends to be smaller. Furthermore, when marketing sustainable products to people or when communicating a new pro-environmental policy, internalizing the threat of climate change, and making it more relevant and tangible to the individual can be very useful. Highlighting personal experiences with the impact of climate change, such as the events resulting from extreme weather, can shrink the gap between attitude and behavior, and lead people to act more sustainable; people will aim to change their behavior in order to reduce the chances of potential threats. Moreover, linking pro-environmental behavior to scientists instead of celebrities and politicians can lead to an increase in sustainable action and can make sustainable behavior more desirable. Besides, encouraging people to look beyond the self can have a positive impact on promoting sustainable behavior. Asking people to be future-oriented and making them aware that their actions, whether positive or negative, have an impact on environmental health can lead to a decrease in the gap between attitude and behavior. Finally, linking some of the values and personal norms to sustainable behavior can also lead to a reduction in the gap; emphasizing on the idea that giving up something now to benefit in the future, and also emphasizing on collectivity and helping others and the community without expecting anything in return can strengthen these values and lead to the sense of responsibility towards the environment.

6.2 Limitations and suggestions for further research

This study has many limitations that need to be taken into consideration when using its results. The study is not representative of the whole world, as only 12 countries participated in the survey. Global north countries are under-represented in the study as only three countries are global north, while the other nine are global south. The results should not be generalized to each country's overall population; each country's sample is very small and, therefore, not representative of the whole population. Furthermore, the survey's age groups are also not representative of the population, as only 1% of the respondents were above 24 years old. Another downside to this study is that the survey was in English, and only those who had their default settings set to English on their Facebook profile took the survey. Besides, giving that most of the respondents are between 13 and 24 years old, Facebook is not the most appropriate distributional channel for the survey. Many individuals between the ages of 13 and 17 have stopped using Facebook; in the United States, Facebook has witnessed a drop from 71% in 2015 to 51% in 2018 among these age groups (Solon, 2018). Moreover, the second part of the survey contains critical determinants of the attitude-behavior gap; however, this part was not mandatory, which resulted in many missing values. Also, some of the determinants need to be expanded by asking further questions to understand them better. For instance, based on the literature and the study results, routine is an important factor; however, it is important to distinguish between unstainable routines, such as taking long showers, and sustainable routines, such as taking short showers. In addition to the survey data, the country-level data obtained to complement the study is not comprehensive as the sample size is very small.

Further research should address all the previously mentioned limitations for better study results. First, better representative data should be collected. A more diversified number of countries should be surveyed, and the number of participants in each country should be sufficient in order for the study to be representative of the whole population or the target group. The survey should also be available in each country's local language to ensure the quality of responses and that people can understand all the questions. It will also be interesting to examine the differences between the age group, such as generation Z and millennials; this will allow a better understanding of the gap and help marketers and policymakers provide customized solutions. Also, to better target each of the age groups, especially young people, the survey should be distributed through other channels, such as Instagram and Tiktok. Also, each country has a unique culture, social norms, and values; a cross-cultural study might also be beneficial in understanding the gap and the differences between nations. Understanding sustainable behavior in each country separately can help establish better local and international laws.

7. Conclusion

The aim of this study was to examine the attitude-behavior gap in the sustainability field and its determinants in order to help policymakers and marketers promote sustainable behavior and bridge the gap. The study was conducted based on a survey taken by 6070 people from 12 different countries, and additional country-level data obtained to complement the survey results. Several determinants were studied, including demographics, values, personal and social norms, and other determinates based on the literature. Two machine learning techniques were used to analyze the data available at hand, namely logistic lasso regression and random forest. Based on these two prediction models, many determinates were found to be important in determining sustainable behavior, such as country of residence, perceived consumer effectiveness, and perceived environmental threats.

Generally, people who come from global south countries tend to be less sustainable. Gender was not found to be an important factor, while the effect of age is ambiguous. These demographics findings can help marketers segment countries and target them accordingly. Furthermore, respondents who believe in their action's effectiveness and who are also willing to spend money (even if others do not) to tackle climate change tend to have a higher probability of translating their attitude into action. Respondents who have specific values and personal norms, such as never gossip, had also higher probability of behaving sustainably; however, the effect of some other values and personal norms studied in the survey was ambiguous. When promoting sustainable consumption and policies, it is essential to highlight the message that each individual's action can make a difference. Also, encouraging people to be future-oriented is vital in promoting sustainable living. Besides, internalizing the threat of the consequences of climate change tends to help translate the positive pro-environmental attitude into action effectively; people who experience days of extreme weather tend to have a relatively higher probability of behaving sustainably. Finally, further research is still needed to better understand the determinants of the gap, such as values, personal norms, and routines. Collecting better data will enable researchers to effectively and efficiently bridge the gap between attitude and behavior, promote sustainable living, and eventually save the environment.

8. References

Anderson, W. T., & Cunningham, W. H. (1972). The Socially Conscious Consumer. *Journal of Marketing*, *36*(3), 23–31. https://doi.org/10.2307/1251036

Antimova, R., Nawijn, J., & Peeters, P. (2012). The awareness/attitude-gap in sustainable tourism: a theoretical perspective. *Tourism Review*, *67*(3), 7–16. https://doi.org/10.1108/16605371211259795

Baldassare, M., & Katz, C. (1992). The Personal Threat of Environmental Problems as Predictor of Environmental Practices. *Environment and Behavior*, *24*(5), 602–616. https://doi.org/10.1177/0013916592245002

Bashir, S., Khwaja, M. G., Turi, J. A., & Toheed, H. (2019). Extension of planned behavioral theory to consumer behaviors in green hotel. *Heliyon*, *5*(12), e02974. https://doi.org/10.1016/j.heliyon.2019.e02974

Berger, I. E., & Corbin, R. M. (1992). Perceived Consumer Effectiveness and Faith in Others as Moderators of Environmentally Responsible Behaviors. *Journal of Public Policy & Marketing*, *11*(2), 79–89. https://doi.org/10.1177/074391569201100208

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical care (London, England)*, *9*(1), 112–118. https://doi.org/10.1186/cc3045

Borin, N., Cerf, D. C., & Krishnan, R. (2011). Consumer effects of environmental impact in product labeling. *Journal of Consumer Marketing*, *28*(1), 76–86. https://doi.org/10.1108/07363761111101976

Braun, T., Cottrell, R., & Dierkes, P. (2017). Fostering changes in attitude, knowledge and behavior: demographic variation in environmental education effects. *Environmental Education Research, 24*(6), 899–920. https://doi.org/10.1080/13504622.2017.1343279

Carolan, M. S. (2006). Social change and the adoption and adaptation of knowledge claims: Whose truth do you trust in regard to sustainable agriculture? *Agriculture and Human Values*, *23*(3), 325–339. https://doi.org/10.1007/s10460-006-9006-4

Carrington, M. J., Neville, B. A., & Whitwell, G. J. (2014). Lost in translation: Exploring the ethical consumer intention–behavior gap. *Journal of Business Research*, *67*(1), 2759–2767.

Climate Change Evidence: How Do We Know? (n.d.). Climate Change: Vital Signs of the Planet. Retrieved August 26, 2020, from https://climate.nasa.gov/evidence/

De Pelsmacker, P., & Janssens, W. (2007). The effect of norms, attitudes and habits on speeding behavior: Scale development and model building and estimation. *Accident Analysis & Prevention*, *39*(1), 6–15. https://doi.org/10.1016/j.aap.2006.05.011

Ellen, P. S., Wiener, J. L., & Cobb-Walgren, C. (1991). The Role of Perceived Consumer Effectiveness in Motivating Environmentally Conscious Behaviors. *Journal of Public Policy & Marketing*, 10(2), 102–117. https://doi.org/10.1177/074391569101000206

Fritzsche, D. J. (1995). Personal values: Potential keys to ethical decision making. *Journal of Business Ethics*, *14*(11), 909–922. https://doi.org/10.1007/bf00882069

Gilg, A., & Barr, S. (2006). Behavioural attitudes towards water saving? Evidence from a study of environmental actions. *Ecological Economics*, *57*(3), 400–414. https://doi.org/10.1016/j.ecolecon.2005.04.010

Grunert, S. C., & Juhl, H. J. (1995). Values, environmental attitudes, and buying of organic foods. *Journal of Economic Psychology*, *16*(1), 39–62. https://doi.org/10.1016/0167-4870(94)00034-8

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)* (1st ed. 2013, Corr. 7th printing 2017 ed.). Springer.

Jody M. Hines , Harold R. Hungerford & Audrey N. Tomera (1987): Analysis and Synthesis of Research on Responsible Environmental Behavior: A Meta-Analysis, *The Journal of Environmental Education*, 18:2, 1- 8. http://dx.doi.org/10.1080/00958964.1987.9943482

Joshi, Y., & Rahman, Z. (2015). Factors Affecting Green Purchase Behaviour and Future Research Directions. *International Strategic Management Review*, *3*(1–2), 128–143. https://doi.org/10.1016/j.ism.2015.04.001

Kokla, M., Virtanen, J., Kolehmainen, M. et al. Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study. *BMC Bioinformatics* 20, 492 (2019). https://doi.org/10.1186/s12859-019-3110-0

Kim, S. H., & Seock, Y.-K. (2019). The roles of values and social norm on personal norms and pro-environmentally friendly apparel product purchasing behavior: The mediating role of personal norms. *Journal of Retailing and Consumer Services*, *51*, 83–90. https://doi.org/10.1016/j.jretconser.2019.05.023

Kim, Y. (2011). Understanding Green Purchase: The Influence of Collectivism, Personal Values and Environmental Attitudes, and the Moderating Effect of Perceived Consumer Effectiveness. *Seoul Journal of Business*, *17*(1), 65–92. https://doi.org/10.35152/snusjb.2011.17.1.003

Kollmuss, A., & Agyeman, J. (2002). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, *8*(3), 239–260. https://doi.org/10.1080/13504620220145401

Kumar, Bipul, 2012. "Theory of Planned Behaviour Approach to Understand the Purchasing Behaviour for Environmentally Sustainable Products," IIMA Working Papers WP2012-12-08, Indian Institute of Management Ahmedabad, Research and Publication Department.

Lim, J. Y., & Moon, K.-K. (2020). Perceived Environmental Threats and Pro-Environmental Behaviors: Investigating the Role of Political Participation Using a South Korean Survey. *International Journal of Environmental Research and Public Health*, *17*(9), 3244. https://doi.org/10.3390/ijerph17093244

Lindenberg, S., & Steg, L. (2007). Normative, Gain and Hedonic Goal Frames Guiding Environmental Behavior. *Journal of Social Issues*, *63*(1), 117–137. https://doi.org/10.1111/j.1540-4560.2007.00499.x

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: a Package for Binary Imbalanced Learning. *The R Journal*, *6*(1), 79. https://doi.org/10.32614/rj-2014-008

Madden, T. J., Ellen, P. S., & Ajzen, I. (1992). A Comparison of the Theory of Planned Behavior and the Theory of Reasoned Action. *Personality and Social Psychology Bulletin*, 18(1), 3–9. https://doi.org/10.1177/0146167292181001

McCarty, J. A., & Shrum, L. J. (1994). The recycling of solid wastes: Personal values, value orientations, and attitudes about recycling as antecedents of recycling behavior. *Journal of Business Research*, *30*(1), 53–62. https://doi.org/10.1016/0148-2963(94)90068-x

Milchrahm, E., & Hasler, A. (2002). Knowledge Transfer in Recycling Networks: Fostering Sustainable Development. J-Jucs, 8(5), 546–556. https://doi.org/10.3217/jucs-008-05-0546

Milfont, T. L., Duckitt, J., & Wagner, C. (2010). A Cross-Cultural Test of the Value-Attitude-Behavior Hierarchy. *Journal of Applied Social Psychology*, *40*(11), 2791–2813. https://doi.org/10.1111/j.1559-1816.2010.00681.x

Morelli, J. (2011). Environmental Sustainability: A Definition for Environmental Professionals. *Journal of Environmental Sustainability*, *1*(1), 1–10. https://doi.org/10.14448/jes.01.0002

P. De Pelsmacker & L. Driesen & G. Rayp, 2003. "Are fair trade labels good business ? Ethics and coffee buying intentions," Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 03/165, Ghent University, Faculty of Economics and Business Administration.

Park, J., & Ha, S. (2014). Understanding Consumer Recycling Behavior: Combining the Theory of Planned Behavior and the Norm Activation Model. *Family and Consumer Sciences Research Journal*, *42*(3), 278–291. https://doi.org/10.1111/fcsr.12061

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, *96*(1), 3–14. https://doi.org/10.1080/00220670209598786

Ramayah, T., Lee, J. W. C., & Lim, S. (2012). Sustaining the environment through recycling: An empirical study. *Journal of Environmental Management*, *102*, 141–147. https://doi.org/10.1016/j.jenvman.2012.02.025

Readfearn, G. (2020, January 13). *Explainer: what are the underlying causes of Australia's shocking bushfire season?* The Guardian.

Roberts, J. A. (1995). Profiling Levels of Socially Responsible Consumer Behavior: A Cluster Analytic Approach and Its Implications for Marketing. *Journal of Marketing Theory and Practice*, *3*(4), 97–117. https://doi.org/10.1080/10696679.1995.11501709

Roberts, J. A. (1996). Green consumers in the 1990s: Profile and implications for advertising. *Journal of Business Research*, *36*(3), 217–231. https://doi.org/10.1016/0148-2963(95)00150-6

Solon, O. (2018, June 1). *Teens are abandoning Facebook in dramatic numbers, study finds*. The Guardian.

Stekhoven, D. J., & Buhlmann, P. (2011). MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Ten impacts of the Australian bushfires. (2020, January 22). UN Environment. https://www.unenvironment.org/news-and-stories/story/ten-impacts-australian-bushfires

Vermeir, I., & Verbeke, W. (2008). Sustainable food consumption among young adults in Belgium: Theory of planned behaviour and the role of confidence and values. *Ecological Economics*, *64*(3), 542–553. https://doi.org/10.1016/j.ecolecon.2007.03.007

Wesley, S. C., Lee, M.-Y., & Kim, E. Y. (2012). The Role of Perceived Consumer Effectiveness and Motivational Attitude on Socially Responsible Purchasing Behavior in South Korea. *Journal of Global Marketing*, *25*(1), 29–44. https://doi.org/10.1080/08911762.2012.697383

White, K., Habib, R., & Hardisty, D. J. (2019). How to SHIFT Consumer Behaviors to be More Sustainable: A Literature Review and Guiding Framework. *Journal of Marketing*, *83*(3), 22–49. https://doi.org/10.1177/0022242919825649

Wynveen, C. J., & Sutton, S. G. (2015). Engaging the public in climate change-related pro-environmental behaviors to protect coral reefs: The role of public trust in the management agency. *Marine Policy*, *53*, 131–140. https://doi.org/10.1016/j.marpol.2014.10.030

Zrałek, J. (2017). Sustainable Consumption in a Trap: Attitude-Behavior Gap and Its Rationalization. *Annales Universitatis Mariae Curie-Skłodowska, Sectio H, Oeconomia*, *51*(2), 281–288. https://doi.org/10.17951/h.2017.51.2.281

9. Appendix

Appendix A: Survey variables

Question	Short name
What is your age (in years)?	age
In which country do you currently live?	country.based
Were you born in this country?	country.origin
What is your gender?	gender
What is your highest completed education?	education
Do you think that extra actions are needed to tackle climate change?	extra.action
In the past 12 months, did you and/or your family do any of the following things?	
How much did you learn about climate change from the following sources?	learn.teacher learn.parents learn.friends learn.other learn.news learn.movie
To what extent do you disagree or agree with the following statements	trust.politicians trust.media trust.scientists trust.famousppl
To what extent do you disagree or agree with the following statements	spend.money spend.others.no
According to you, what percentage of young people in your country think climate change is a serious problem?	young.opinion
According to you, what percentage of young people in your country are willing to spend money to tackle climate change?	perc.young.spend
Compared to other young people in your country, how much do you know about climate change?	relative.know
How many days with extreme weather did you experience in the past 12 months?	experience.weather
Compared to this, how many days with extreme weather do you expect for the next 12 months?	expect.weather
In general, how willing are you to give up something today so you can benefit more in the future?	giveup.something

Appendix A: Survey variables (continued)

Question	Short name
In general, how willing are you to take risks?	risk.taker
In general, how willing are you to help others without expecting anything in return?	help.others
Please indicate for the following sentences whether you disagree or agree	hide.mistakes steal gossip obey.laws eavesdrop work.rights
Please indicate for the following sentences whether you disagree or agree	hard.worker eat.alot visit.places ads.silly follow.routine
In general, how willing are you to take risks?	risk.taker
In general, how willing are you to help others without expecting anything in return?	help.others

Appendix B: OOB imputation error

Imputed variable	Imputation error rate (%)	Imputed variable	Imputation error rate (%)	Imputed variable	Imputation error rate (%)
learn.news	0.075	trust.famousppl	0.027	expect.weather	0.000
trust.scientists	0.056	spend.money	0.019	help.others	0.063
learn.teacher	0.019	spend.others.no	0.019	enough.money	0.010
trust.politicians	0.060	young.opinion	0.054	bills	0.005
learn.movie	0.032	perc.young.spend	0.055	future.money	0.007
learn.friends	0.022	relative.know	0.026	hide.mistakes	0.043
trust.media	0.020	experience.weather	0.010	steal	0.055
learn.other	0.042	giveup.something	0.056	gossip	0.054
learn.parents	0.027	risk.taker	0.066	obey.laws	0.057
eavesdrop	0.054	eat.alot	0.000	follow.routine	0.000
work.rights	0.043	visit.places	0.000		
hard.worker	0.000	ads_silly	0.000		

Appendix C: Logistic lasso coefficients

Coefficient	Estimate	Coefficient	Estimate
Country.basedEgypt	-1.112939310	Country.orginYes	0.138812594
Country.basedIndia	.	genderMale	.
Country.basedIndonesia	-0.944615905	genderOther	.
Country.basedLiberia	.	educationEducation up to age 12	.
Country.basedMexico	0.012272972	educationEducation up to age 14	.
Country.basedNetherlands	0.546306511	educationEducation up to age 16	0.018468305
Country.basedNigeria	-0.993786205	educationEducation up to age 18	.
Country.basedPhilippines	.	educationNo formal education	-1.689829813
Country.basedUnited Kingdom	0.256237543	educationUniversity	.
Country.basedUSA	-0.365856215	extra.action	0.154125445
age14	-0.043325551	learn.teacher	-0.067839111
age15	-0.065859966	learn.parents	.
age16	.cg	learn.friends	-0.007307496
age17	0.029616150	learn.other	0.008004530
age18	.	learn.news	-0.005084347
age19	.	learn.movie	.
age20	.	trust.politicians	.
age21	-0.092786761	trust.media	-0.023837751
age22	.	trust.scientists	0.058813556
age23	0.086178311	trust.famousppl	-0.014589562
Age24	-0.070438028	spend.money	.
Age older than 24	.	spend.others.no	0.045698040
experience.weather11-15 days	0.259128980	young.opinion	-0.012722926
experience.weather16-20 days	0.309272055	perc.young.spend	.
experience.weather6-10 days	.	relative.know	0.138700713
experience.weathermore than 20 days	.	giveup.something	0.018898314
experience.weatherzero days	.	risk.taker	0.005425620

Appendix C: Logistic lasso coefficients (Continued)

Coefficient	Estimate	Coefficient	Estimate
expect.weatherMore days with extreme weather	0.089728854	help.others	0.048698550
expect.weatherSame number of days with extreme weather	.	enough.money	.
hard.workerDisagree	.	bills	0.025250103
eat.alotDisagree	-0.062815408	future.money	.
visit.placesDisagree	.	hide.mistakes	.
ads.sillyDisagree	-0.067118000	steal	-0.003304467
follow.routineDisagree	0.104516163	gossip	0.031955496
Plastic.waste	.	obey.laws	.
GDP	0.004780418	eavesdrop	.
Temp.change	.	work.rights	-0.005671825
Total.SW	.	AVG.EDU	.
R.E.C	-0.002500600	Meat.supply	0.003250040
CO2	.	NR.Dep	-0.023378477