

**Department of Mathematics**

Master in Data Science

*Is Our House on Fire?*

*Analysis and Prediction of European Attitudes towards Climate Change.*

|  |  |
| --- | --- |
| *Thesis Supervisor:*  *Giuseppe A. Veltri*  *Co-Supervisor:*  *Elena Pavan* | *Master Student:*  *Claudia Giuspoli* |

Academic Year 2019/2020

Contents

[Introduction 1](#_Toc65832338)

1. [Literature Review 4](#_Toc65832339)

[1.1 Climate Change Risk Perception 6](#_Toc65832340)

[1.1.1 The Concept of Risk Perception 6](#_Toc65832341)

[1.1.2 Sociodemographic Dimension 7](#_Toc65832342)

[1.1.3 Cognitive Dimension 9](#_Toc65832343)

[1.1.4 Experiential Dimension 9](#_Toc65832344)

[1.1.5 Socio-cultural Dimension 10](#_Toc65832345)

[1.2 Pro-environmental Behaviour 10](#_Toc65832346)

[1.2.1 Types of Pro-environmental Behaviour 11](#_Toc65832347)

[1.2.2 Factors influencing Pro-environmental Behaviour 11](#_Toc65832348)

[1.2.1 Self-reported Behaviour 15](#_Toc65832349)

[1.3 From Climate Risk Perception to Pro-environmental Behaviour 15](#_Toc65832350)

2. [Data and Method 19](#_Toc65832351)

[2.1 Research Questions 19](#_Toc65832352)

[2.2 Data Description 21](#_Toc65832353)

[2.2.1 Data Cleaning 22](#_Toc65832354)

[2.3 Methodology 24](#_Toc65832355)

[2.3.1 Unsupervised Machine Learning Algorithms 24](#_Toc65832356)

[2.3.2 Supervised Machine Learning Algorithms 26](#_Toc65832357)

3. [Exploratory Data Analysis 31](#_Toc65832358)

[3.1 New segmentations among citizens 32](#_Toc65832359)

[3.2 Explanatory Data Analysis 38](#_Toc65832360)

4. [Analysis & Prediction 47](#_Toc65832361)

[4.1 Evaluating Classification Models 47](#_Toc65832362)

[4.2 Prediction of the Behavior on the Entire Dataset 48](#_Toc65832363)

[4.3 Prediction of the Behavior based on the Individual Risk Perception Level 54](#_Toc65832364)

5. [Conclusion 62](#_Toc65832365)

[5.1 Summary results 62](#_Toc65832366)

[5.2 Limitations 65](#_Toc65832367)

[5.3 Further studies 66](#_Toc65832368)

[Appendix 68](#_Toc65832369)

[Bibliography 86](#_Toc65832370)

# Introduction

“You are never too small to make a difference” remarked Greta Thunberg in her famous speech “Climate Justice Now” at the 2019 climate conference in Poland. Greta is a young activist who inspired an international movement, *FridaysForFuture*, to fight climate change.

Climate change is a social and ecological problem and a substantial political issue globally, and that it continues to grow in intensity and complexity. According to the World Health Organization (2008), human activities are a primary cause. The warming of the planet leads to extreme weather events such as floods, storms, and droughts. These phenomena can endanger health by altering the functioning of some of the essential elements for the planet's life: food security, clean air or water, and freedom from illness (*Protecting Health from Climate Change*, 2008). For this reason, younger adults protest during the Climate Strike marches led by Greta, scientists continue to affirm the severity of climate change, politicians feel the pressure of public opinion to adapt strategies to fight it.

Understanding the individual environmental attitudes is a fundamental key to promote citizen engagement in pro-environmental behavior and support green development (Liu et al., 2014; Zhou et al., 2020). Over the last several decades, social scientists have found different explanations and interpretations of citizens' attitudes towards climate change. Socio-cultural, natural, political, sociodemographic, and contextual factors can influence the attitudes of individuals.

The study focuses on the prediction and classification of European citizens self-reported pro-environmental behaviour. A pro-environmental performance can be regarded: a walk, recycling, and use public transport. Scholars have identified several factors that shape pro-environmental behaviour, such as sociodemographic, individual (as climate change risk perception), experiential, socio-cultural variables (Xie et al., 2019). These dimensions are not necessarily assumed to be independent. For example, some experiences can interact indirectly with risk perception, and successively affect behaviour. Therefore, the first aim of the research is to understand what factors influence pro-environmental behaviour. Additionally, the study wants to examine the role that climate change risk perception plays in the prediction. More attention is needed to risk perceptions since there is a value-action gap: worried people do not always perform any ecological actions. The second broader aim of the research is to study the relationship between climate change risk perception and pro-environmental behaviour. What factors influence pro-environmental behavior in worried citizens? And in unworried citizens? There are some differences?

To sum up, the dissertation, firstly, aims to explore the general phenomenon, creating a traditional self-reported pro-environmental behaviour model. Successively, it focuses on discovering some factors’ differences in determining behavior between worried and unworried citizens.

Chapter 1 presents the essential concepts to understand climate change risk perception and pro-environmental behaviour. These two notions are independent of each other partially. One does not automatically involve the other. Climate change risk perception is one of the main predictors of predicting pro-environmental behaviour, and for this reason, it is important also to stress the factors that shape it. Pro-environmental behaviour is the concept key of the analysis. Understanding the factors related to different explanations of pro-environmental behaviour, is particularly essential.

Chapter 2 completely focuses on the methodology. Primarily, the research questions are led by the described theoretical framework. The Special Eurobarometer 91.3 data are introduced, and the data cleaning process is also precisely reported. At the end of the chapter, the techniques performed to conduct these analyses are described.

Chapter 3 starts with some exploratory analysis. The fundamental aim of this part is to find new segmentations among attitudes’ citizens. Partition Around Medoids (PAM) clustering and Correlational Class Analysis (CCA) are adopted for looking at unknown and hidden patterns. These unsupervised machine learning algorithms aim to divide citizens into different groups, and they create two explanatory variables renamed respectively as green-identity and cultural schemas. Afterward, these new categories are adopted as predictors in the pro-environmental behaviour models. The second step of this chapter is to explore data, in particular the distributions of the explanatory variables.

The purpose of Chapter 4 is to examine self-reported pro-environmental behaviour at the individual level in the European context. Several machine learning algorithms are performed to search for a well-fitting model. The analysis emphasizes the most important factors that shape pro-environmental behaviour. Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting are implemented. The predictions are divided into two steps. In the first place, we will predict pro-environmental behaviour on the entire dataset. This pro-environmental behaviour model is remarkably traditional; in fact, we insert as predictors: risk perception, some sociodemographic information, the type of green self-identity and cultural schemas, and, lastly, country. The second part of the analysis pursues to examine the factors that determine the behaviors of worried and unworried citizens separately. The aim is to cover the gap in the literature related to the action-value gap. In light of this, the second block of predictions is more complex and tries to understand this discrepancy. Therefore, are the characteristics of citizens with low-risk perceptions the same as those who have high-risk perceptions in the prediction of pro-environmental behaviour? We always adopt the same classifiers and the same predictors (except for the exclusion of risk perception, already used to split data), examining this question.

Finally, Chapter 5 concludes this study. The results, the limitations, and the opening questions for future research are adequately exposed.

Chapter 1

# Literature Review

In recent years, scholars have increasingly paid attention to public opinion and attitudes about environmental issues in the United States (Brulle et al., 2012; Driscoll, 2019; Egan & Mullin, 2017; O’Connor et al., 1999; Shwom et al., 2015), in the European Countries (Lorenzoni & Pidgeon, 2006; Vainio & Paloniemi, 2013) and around the world (Lee et al., 2015; Sun & Han, 2018). “Public opinion on climate change is multidimensional, dynamic, and differentiated. […] It includes, among others, beliefs about anthropogenic climate change, perceptions of climate change risks, concern about its seriousness, and thoughts on what, if anything, should be done to address it” (Shwom et al., 2015, p. 269).

Some lexical clarifications are reported between global warming versus climate change. Global warming refers to temperatures increase on the Earth's surface (Dunlap, 2014). Instead, climate change refers more generally to changing climatic conditions and their effects (Dunlap, 2014). Public opinion and media use these two terms interchangeably (Weber, 2016). In this study, only the climate change term is used, due to it is more adopted (and more accurate) by the scientific community in the last years (Dunlap, 2014).

Attitudes changes over time and space due to individual, socio-cultural, political, economic, habitat factors (Shwom et al., 2015). We use the term “climate change attitudes” to report values, beliefs, concerns, and worries of people about issues in the environmental domain. Another component of attitudes is behaviour. Scholars have found an important relationship between an individual’s “green behaviours”, and therefore all actions to safeguard the environment, and their attitudes regarding climate change (Lacasse, 2015; O’Connor et al., 1999; Vainio & Paloniemi, 2013). Environmental beliefs are an important mediator in explaining support for climate-friendly action, as trust in climate change is associated with engaging in eco-friendly behaviours (Vainio & Paloniemi, 2013).

Climate change tends to remain a complex, uncertain, and abstract phenomenon as most citizens get information from mass media and are hardly aware whether they are experiencing it directly (Vainio & Paloniemi, 2013). This uncertainty makes opinions even more ambiguous since people have difficulty evaluating the consequences of their actions and solutions or understanding risks. According to the Special Eurobarometer 91.3 entitled ‘Climate change’, only 20% of European citizens claim that climate change is the single most serious problem facing the world (European Commission, Brussels, 2019). Climate change is considered less important than hunger and poverty in the world. On the one hand, 20% of Europeans perceive the seriousness of climate change; on the other hand, 80% of them nonetheless perform some environmental-friendly actions to reduce the phenomenon. There is a discrepancy between concern and behaviour (Lacroix & Gifford, 2018; Vainio & Paloniemi, 2013). Probably, an individual performs actions only whether he/she knows that he/she can make a difference and if he/she knows that other citizens and governments are moving in the same direction to save the planet (Lorenzoni & Pidgeon, 2006). Therefore, the relationship between action and attitude also becomes very complex and not obvious.

To sum up, the following chapter focuses on presenting the topic of the research: environmental-friendly behaviour. Also, we evidence the important and complicated relationship between pro-environmental behaviour and attitudes toward climate change, focusing on climate change risk perception.

* 1. **Climate Change Risk Perception** 
     1. **The Concept of Risk Perception**

Slovic (1987) affirms that risk perception varies according to inter and intrapersonal, geographical, cultural, and social influences. In this way, an “objective” risk perception does not exist (Yu et al., 2019). Nevertheless, some factors define risk perception, such as “dread risk” and “unknown risk”: the more a phenomenon is considered unpredictable, uncontrollable, with catastrophic consequences, and mostly it is invisible, the more perceived hazard or risk increases (Slovic, 1987). For example, people judge nuclear technology as riskier than car accidents, since the first has catastrophic consequences and it is uncontrollable and invisible (Slovic, 1987). Consistency, perceived risk is a mental and social construction, created to help people tackle uncertainty or danger (Slovic & Weber, 2002). “It does not exist ‘out there’, independent of our minds and cultures” (Slovic & Weber, 2002, p. 4).

Risk perceptions are shaped and influenced by different environmental field factors, which can be grouped into four categories, suggested in van der Linden’s (2015) Climate Change Risk Perception Model (CCRPM): sociodemographic, cognitive, experiential, and socio-cultural. All these dimensions lead to change and shape the individual level of risk perception. In the next sections, these categories are explained in detail so that is possible to clarify the precursors to climate change risk perception which we consider as one of the main predictors in the model we propose to study pro-environmental behaviours. Therefore, it is important to theoretically describe what causes and conditions could be related to individual risk perception.

* + 1. **Sociodemographic Dimension**

Sociodemographic factors such as gender, education level, age, marital status, city/town size, income, and political orientation are related to climate change risk perception.

In literature, women tend to have more concerns than men for many hazards (Finucane et al., 2000) also in environmental issues (Davidson & Haan, 2012; Goldsmith et al., 2013; O’Connor et al., 1999; Zhou et al., 2020). One explanation of the gender gap could be that women are more likely to have “a stronger sense of social responsibility and affinity for taking others’ perspectives” (Goldsmith et al., 2013, p. 6), and for this reason, they are more concerned than men.

Also, some studies find that education level is positively correlated with risk perception (Meyer, 2015): more educated people tend to be more informed on the topic and more aware of the consequences of their actions (Meyer, 2015; Sun & Han, 2018).

Recent studies demonstrated that younger adults are more worried about the consequences of climate change than older individuals (Echavarren et al., 2019; Sun & Han, 2018; Weber, 2016). However, the reasons for this divergence could be two: the *ageing effect* and the *cohort* *effect*, which are the result of being at a certain age point or belonging to a particular generation (Torgler & García-Valiñas, 2007). This last option could explain the difference in attitudes between two different generations due to generational variations in socialization and lifestyle (Torgler & García-Valiñas, 2007).

Regarding marital status, some observers suggest that it may influence environmentalism as especially married individuals and/or with children are more worried about climate change as a consequence of their “not being alone”, thus as a form of care for the others with whom they share their lives (Torgler & García-Valiñas, 2007).

The relationship between the place where individuals live and risk perception remains a debated factor. On the one side, citizens who are located in rural areas should be more in contact with nature and, therefore, they are expected to have more concerned with environmental issues. On the other side, those who live in a city tend to be more active concerning environmental policies (Torgler & García-Valiñas, 2007).

The economic situation is also correlated with environmental attitudes and therefore, with risk perception. In general, wealthier people expect a clean and healthy planet (Torgler & García-Valiñas, 2007). For this reason, low-income individuals appeared to be less knowledgeable of climate change than higher-income individuals (Shi et al., 2015). On the one hand, wealthier people desire a healthy planet, and they are worried about their beautiful homes disappearing. On the other hand, high-income individuals are more educated, and they are aware of the seriousness of the problem.

One last factor is political learning. American literature suggests that Democrats and Liberals are more likely to believe and to concern about climate change than Republicans and Conservatives (Davidson & Haan, 2012; Egan & Mullin, 2017; Fielding et al., 2012; Liu et al., 2014; McCright, 2011). McCright, Dunlap, and Marquart-Pyatt (2016) extend these findings in the European Union. However, in Europe, the distinction between Democrats (left) and Republicans (right) is not clear and uniform. In Western Europe, the left is concerned with change and equality; while in former Communist countries, this identification cannot be found (McCright et al., 2016). According to the authors, the polarization of climate change (Liberals are more worried than Conservatives) is found in only Western European countries: citizens on the right are unlikely to recognize the phenomenon as a serious issue than those on the left (McCright et al., 2016). Anyhow “the effect of left-right ideology in Western Europe is considerably weaker than the effect of political ideology (and party identification) in the USA” (McCright et al., 2016, p. 13). Instead, citizens of Eastern Europe are not divided from an ideological point of view on that topic, since the irrelevance of climate change as a political issue and then the difference ideologization and identification of left-right (McCright et al., 2016).

To summarize more educated, liberal (in Western Europe), wealthy, and married young women are more likely to show a higher risk perception level than older, less educated, conservative, poor and, single men.

* + 1. **Cognitive Dimension**

Knowledge is the main factor in the cognitive dimension related to climate change risk perception (Bradley et al., 2020; Hidalgo & Pisano, 2010; O’Connor et al., 1999). Van der Liden (2015) suggests that knowledge about the causes or impact of climate change can improve individuals’ concerns. The author underlines that we cannot make this critical distinction “between an individual's ‘subjective’ knowledge (i.e., what people think is true) and the actual ‘evidence’ (insofar a clear scientific consensus exists, e.g., that burning fossil fuels contributes to climate change)” (van der Linden, 2015, p. 114). However, people with accurate knowledge of the phenomenon seem to perceive it as a serious problem and, at the same time, they want to fight it (Bradley et al., 2020; Hidalgo & Pisano, 2010).

* + 1. **Experiential Dimension**

This section aims at shedding light on the importance of emotions and personal experiences with natural disasters in risk perception.

Looking at emotions first, it is important to stress that “risk as feeling” refers to an instinctive reaction to danger: people immediately judge a potential problem as positive or negative feelings (Slovic & Peters, 2006). The more the immediate feeling is negative, the more risk perception increase. When an individual begins to use this feeling as the first influencer of behaviour, it means that emotion is called “the affect heuristic” (Slovic & Peters, 2006). Some researchers evidence that emotion is a predictor of climate change risk perception, and that negative feelings increase concern (van der Linden, 2015). On the contrary, Taylor et colleagues (2014) declare that extreme negative emotions can create the opposite effect: fear and anxiety can lead to greater psychological distance and apathy towards climate change.

Similarly, personal experiences with a hazard or extreme weather events, such as extraordinarily hot or cold weather, storms flooding, and forest fires, can increase perceived risk (van der Linden, 2015). Familiarity with extraordinary weather events makes risk more concrete and real, increasing concern and decreasing the distance from danger psychologically (Akerlof et al., 2013; Bradley et al., 2020; Taylor et al., 2014; van der Linden, 2015).

* + 1. **Socio-cultural Dimension**

One of the most important approaches to risk perception is the cultural theory defined by Mary Douglas in the 1960s. According to theory, risks are a social construction, and they depend on: “(a) the form of social relationships people maintain; (b) cultural biases such as shared values ​​and beliefs including views on human nature, views on society, risk perceptions, and so-called myths of nature, which especially refer to biases toward environmental risks; and (c) preferred behavioural strategies” (Steg & Sievers, 2000, p. 251). Cultural theory suggests that people can be divided into four groups based on their worldview and values: fatalists, hierarchists, individualists, and egalitarians, based on their attitudes and perception (Steg & Sievers, 2000). Fatalists perceive the reality as the product of chance, and it is out of human control, hierarchists appreciate hierarchies and institutional values, and for them, nature can be safeguarded by regulations; individualists focus attention on personal freedom and they see nature as benign, lastly, egalitarians emphasize group welfare, and for them nature is fragile (Taylor et al., 2014; Wildavsky & Dake, 1990).

Various studies have found a significant relationship between “cultural worldview” and attitudes toward climate change (Steg & Sievers, 2000; Taylor et al., 2014). For example, the values of egalitarians are positively correlated with environmentalism, while individualists’ values are negatively correlated with it (Steg & Sievers, 2000).

* 1. **Pro-environmental Behaviour**

“Pro-environmental behaviour is most commonly defined as ‘intentionally reducing the negative impact that an action can have on the environment” (Dono et al., 2010, p. 178). Generally, scientists mean pro-environmental behaviour like walking, recycling, energy saving. It is an intent-oriented definition, which is different from an impact-oriented one: the first highlights the action as such, it may not produce an environmental impact, the second must necessarily have a sustainable effect (Stern, 2000). When scholars focus on individual attitudes or motives to understand behaviour, as in this case, they adopt an intent-oriented definition (Stern, 2000).

* + 1. **Types of Pro-environmental Behaviour**

According to Stern (2000), there are four different environmental behaviour types: environmental activism, nonactivist behaviours in the public sphere, private-sphere environmentalism, other environmentally significant behaviours.

Environmental activism concerns citizens' active involvement in manifestation, organizations and demonstrations (e.g., active participation in environmental organizations as the famous activist Greta Thunberg. She adopts several behaviours in both public and private spheres to save the planet) (Dono et al., 2010).

Nonactivist behaviours in the public sphereconcern individuals who support public policies or environmental citizenship, their actions have a positive, but indirect, impact on the environment (e.g., petitioning on environmental topics, participating, and contributing to environmental institutions as Fridays for Future marches by students).

Private-sphere environmentalismrefers simply to green consumers in the home and personal field. These customers purchase or use products that have an ecological impact (e.g., to buy an electric car).

Lastly, other environmentally significant behaviours mention to individuals who positively impact the environment even if other behaviours, such as a worker, can influence pro-environmental actions of the organization to which he/she belongs (e.g., engineers may create manufactured products in a more or less environmentally approach).

* + 1. **Factors influencing Pro-environmental Behaviour**

Sociologists suggest that pro-environmental behaviour is not only motivated by environmental attitudes (Stern, 2000). There are also psychological, social, and economic factors that can influence and mitigate behaviour (Whitmarsh & O’Neill, 2010). We can organize these factors into three main groups: individual, contextual, and again, sociodemographic dimension.

***Individual Dimension***

Individual factors are motivation, environmental knowledge, self-identity, emotion (Kollmuss & Agyeman, 2002). One of the most important theories which aim to predict behaviour is the theory of planned behaviour (TPB). It affirms behaviour is determined by attitudes towards that action, subjective norms, and perceived behavioural control (Oreg & Katz-Gerro, 2006; Whitmarsh & O’Neill, 2010). Therefore, actual behaviour is determined by behavioural intention, which has its turn is influenced by both attitudes and social, or normative, pressures (Kollmuss & Agyeman, 2002).

Generally, as just seen, motivation, values, knowledge are interconnected. Motivation (unconscious or conscious) drives action and it could be shaped also linked by environmental knowledge and awareness (Kollmuss & Agyeman, 2002). Also, values, influenced by social networks (family, peer-groups, education), shape motivation and behaviour. “The more strongly individuals subscribe to values beyond their own immediate interests, that is, self-transcendent, prosocial, altruistic or biospheric values, the more likely they are to engage in pro-environmental behaviour” (Steg & Vlek, 2009, p. 311).

One last factor is self-identity. Generally, people who perceived themselves as a green-identity act accordingly pro-environmentally (Gatersleben et al., 2014; Whitmarsh & O’Neill, 2010). For example, an individual who perceives himself/herself as a typical recycler is more likely to recycle than those who do not recognize himself/herself as a recycler (Whitmarsh & O’Neill, 2010).

Finally, emotions. Emotional involvement is shaped by knowledge about the topic since, as we have already shown, climate change is an abstract and complex problem (Kollmuss & Agyeman, 2002). However, some individuals experience climate change (extreme atmospheric phenomena) directly and then they feel fear, anger, guilt (Kollmuss & Agyeman, 2002). These extreme negative feelings can lead to refusal to accept reality, rational distancing from the problem, apathy, and delegation of personal responsibility (Kollmuss & Agyeman, 2002). Therefore, when the emotion is too strong and extreme can prevent and block behaviour.

We can summarize that mainly attitudes, values, self-identity, or experiences have a powerful influence on behaviour. However, negative emotions, as in this case risk perception, are not always positively related to pro-environmental action. Sometimes, whether the perceived risk is too high or whether the lived experience is extremely negative, the opposite effect is obtained.

***Contextual Dimension***

Pro-environmental behaviours are affected by the context where individuals are embedded. According to Kollmuss & Agyeman (2002), institutional, economic, and socio-cultural factors influence individuals’ behaviour. Firstly, “many pro-environmental behaviours can only take place if the necessary infrastructure is provided (e.g. recycling, taking public transportation)” (Kollmuss & Agyeman, 2002, p. 248). It is evident that if there is no public transport, an individual can not take place an environmental-friendly action.

Then, economic factors are essential in the decision-making process regarding whether or not to perform pro-environmental behavior. People could be influenced partially by monetary motives to behave pro-environmentally; therefore, if an eco-friendly product is too expensive, it can lead to a lack of pro-environmental action (Stern, 2000).

Cultural norms and cross-cultural differences represent an influential role in shaping people’s performance (Kollmuss & Agyeman, 2002; Oreg & Katz-Gerro, 2006). According to cognitive sociology, “individuals are socialized into various thought communities, or cultures, via cognitive norms that specify appropriate ways of perceiving, focusing attention, and signifying” (Markle, 2019, p. 3). These socio-cognitive norms control the rules of thinking and interpretation of reality, how people think or determine what is relevant, and what is ignored. For example, according to Markle (2019), Americans consider recycling as emblematic pro-environmental performance, engaging selective attention. Therefore, for Americans, the only possible green action is recycling. They are aware of the problem, but they choose to focus only on the recycling practice, limiting their exposure to other types of actions. Socio-cognitive norms can be called “cultural schemas”. These cultural schemas may interact with eco-friendly actions, as just demonstrated with the example of Americans and recycling.

The last factor is the influence of nationality. Socially accepted behaviour varies by country and culture and can impact behavioural patterns at the individual level. Cultural values may shape and change the behaviour’s outcome. Researchers emphasize that Eastern cultures draw upon ethics of care, while Western cultures enforce individualist values (Simga-Mugan et al., 2005). The individualism dimension of culture refers to the low degree of emotional attachment to groups and the centrality of self-sufficiency or self-centeredness (e.g., USA). Instead, the collectivism dimension refers to the centrality role of communities (e.g., Japan) (Nagy & Konyha Molnárné, 2018). In collectivist societies, some pro-environmental behaviors are more likely to act than individualistic ones (Bonera et al., 2017; Nagy & Konyha Molnárné, 2018). The reason is that individuals living in the collectivistic culture are willing to achieve a group’s goals while sacrificing their personal ones (Cho et al., 2013). Generally, individualism is spread in Europe and North America, while collectivism is found in the rest of the world (Ilieș & Zahid, 2019).

***Sociodemographic Dimension***

According to Larson and colleges (2011), the effect of sociodemographic characteristics on pro-environmental behaviour has not been satisfactorily investigated, instead of the relationship on attitudes, as widely discussed above with perceived risk. Traditionally, poor, and uneducated citizens show lower pro-environmental behaviour than rich and highly educated ones (Larson et al., 2011). Being more concerned, women more likely to behave sustainably (Larson et al., 2011; Vicente-Molina et al., 2018). As we have already explained, this gap could be due to gender socialization theory: women should be more cooperative, empathic, and protective (in this case towards nature) than men (Vicente-Molina et al., 2018). Therefore, similar to risk perception, green consumers, or more in general environmentally activists, are “young, female, well educated, liberal and wealthy” (Gilg et al., 2005, p. 484).

## **1.2.1 Self-reported Behaviour**

In social research, a standard measurement of pro-environmental behaviour is based on respondents’ self-reports through questionnaire items (Steg & Vlek, 2009). Self-reported data reports what individuals believe they have done, and it is in contrast with behavioural data (Veltri, 2019). A problem arises as to individuals may not give an accurate and truthful answer to their actual behaviour (Gatersleben et al., 2002). “Self-reported behaviour reflects perceptions or beliefs about people’s behaviour rather than their actual behaviour. Factors such as social desirability and other types of (conscious or unconscious) response bias may result in inaccurate reports of actual behaviour” (Gatersleben et al., 2002, p. 337). Therefore, when an interviewer asks about behaviour, it is probably that interviewee refers to an intention rather than his/her concrete behaviour (Chao & Lam, 2011). Social responsibility and social desirability can lead to provide inaccurate information on behaviour (Chao & Lam, 2011; Veltri, 2019). Whether on one side the amount of individuals’ pro-environmental behaviour could be overestimated and not entirely precise, on the other side dichotomized questions about self-reported behaviours (“I do” or “I don’t”) result more accurate and reliable (Kaiser et al., 2003).

**1.3 From Climate Risk Perception to Pro-environmental Behaviour**

In the previous sections, we described two main concepts standing at the heart of my research: climate change risk perception and pro-environmental behaviour. These two concepts are partially separated from each other. One does not automatically involve the other. Literature is not clear about their relationship. Some research demonstrates the importance of climate change risk perception to predict pro-environmental behaviour (Xie et al., 2019; Yu et al., 2019; Zhou et al., 2020). Higher risk perception is positively correlated with pro-environmental action, becoming the main predictor and intermediary on behavioural (Xie et al., 2019). When people became aware and then concerned about the issue, they are more likely to behave eco-sustainably to mitigate and fight the environmental problem (Zhou et al., 2020). Instead, Stern (2000) and O’Connor and colleagues (1999) suggest that risk perception could fail to conduct pro-environmental behaviours. Two individuals who have the same level of concern may react by having completely divergent behaviour (Zeng et al., 2020). The reason is that other factors influence the decision-making process. Eco-friendly behaviour may be motived simply by financial interests and not by a high-risk perception (Stern, 2000). Otherwise, as we have seen in the paragraph above, high-risk perception may lead to apathy or to reject reality and therefore no pro-environmental actions are implemented (Kollmuss & Agyeman, 2002). This discrepancy is called the value-action gap (Lacroix & Gifford, 2018).



Figure 1: A schematic overview of Protection Motivation Theory (explained by Bubeck et al., 2018)

An explanation for this value-action gap is provided by the protection motivation theory (PMT), explained by Bubeck end colleagues (2018) and shown in figure 1, which has become popular to explain “the risk-reducing behaviour of residents against natural hazards” (Bubeck et al., 2018, p. 1239). According to the theory, the decision to behave pro-environmentally or not is driven by two different cognitive processes: threat appraisal (or referred to as “risk perception”) and coping appraisal. When a threshold of risk perception (threat appraisal) is exceeded, the individual begins to adopt a possible measure to reduce the threat, called coping appraisal. Coping appraisal refers to the individual’s evaluation of his/her capacity in responding to the perceived threat (Shafiei & Maleksaeidi, 2020). The interaction between risk perception and coping appraisal affects behaviour. If an individual has a high level of risk perception and high coping appraisal, then he/she will have pro-environmental behaviour. Otherwise, if he/she has a high-risk perception but a low coping appraisal, then he/she will have a nonprotective response. Keshavarz & Karami (2016) extend this framework to explain the farmers' pro-environmental behavior under drought. Even if all the farmers have a high-risk perception (high threat appeal) due to direct experience with drought, not all have carried out a pro-environmental behaviour, such as trying to prevent environmental pollution, The reason is that if a farmer thinks that he has no control over drought on the farm or that he/she cannot reduce its effects (low coping appraisal), then he/she will not perform any pro-environmental actions.

Nowadays, there is the revised theory introduced by Rogers, who adds some variables that influence risk perception and coping appraisal: environmental and intrapersonal sources (Bubeck et al., 2018). Prior experiences, sociodemographic, personal attitudes, and contextual factors may influence or modify risk perception and coping appraisal, affecting the behavioural response.

Chapter 2

# Data and Method

## **2.1 Research Questions**

The first chapter explored the main theoretical aspects that are related to attitudes towards climate change. Many elements are presented: climate change risk perception, pro-environmental behaviour, and their respective factors that influence them. All these concepts are interconnected with each other. Thus, firstly it is important to understand the aim of the research. The study focuses on the prediction and classification of self-reported pro-environmental behaviour in European countries. As explained in the literate review, contextual, individual, and sociodemographic conditions can lead to encouraging eco-friendly behaviour. Therefore, the study wants to understand what of these factors and predictors directly influence self-reported pro-environmental behaviour. Thus, we examine intent-oriented behaviour considering attitudes and motivations and try to clarify how they come along under specific conditions. Another important point to highlight is that there is no distinction between different types of pro-environmental behaviours. All actions, from activism to environmentalists in the private sphere, are taken into account.

Based on the literature review, we select several sociodemographic features: gender, age, education level, residence location, economic and marital status, political orientation, and country where the individual belongs. Climate Change risk perception, types of green identities, and cultural schemas (2.1) are added as predictors.

Given the complexity of the phenomenon, a brief summary is made of the factors that are considered in the analysis and that influence pro-environmental behaviour, and at the same time, climate change risk perception.

We start with sociodemographic variables. Being liberal, woman, married or mother/father, wealthy, and young positively influences risk perception. The same relationship, with these attributes just mentioned, occurs with pro-environmental behaviour. Afterward, for the individual dimension, people who perceive themselves as a green-identity and are seriously worried about climate change act pro-environmentally. Finally, we can examine the country of origin of the citizens and the shared cultural schemas for the contextual dimension. Generally, European citizens are characterized by an individualism dimension of culture, and they are more likely to act less environmentally than collectivist culture (i.e., Asian countries). Instead, cultural schemas can control the rules of thinking and interpretation of climate change topics. Different cultural schemas, or socio-cognitive norms, could modify and influence the way we perceive climate change’s phenomenon. Divergent rules of perception can lead to divergent behavioral responses.

Overall, the next chapter wants to answer the following questions: what are the most significant characteristics of those who behave environmentally? What is the role of climate change risk perception?

Therefore, according to the existing literature, it is assumed that:

Hp1: higher individual climate change risk perception positively influences and predicts pro-environmental behaviour;

Hp2: an extreme green-identity positively influences pro-environmental behaviour;

Hp3: different citizens’ cultural schemas towards climate shape contrasting behaviour;

Hp4: Demographic characteristics effect on pro-environmental behaviour, in particular:

1. Higher education has a positive effect on pro-environmental behaviour;
2. Women are more likely to perform pro-environmental behaviour;
3. Higher-income has a positive effect on pro-environmental behaviour;
4. Youngsters are more likely to perform pro-environmental behaviour;
5. left-wing individuals are more likely to perform pro-environmental behaviour.

In the second part of the analysis, we still want to understand the main factors to predict behaviour, but separately between two groups of those who worry about climate change and those who not. The illustrated literature especially shows the importance of risk perception in predicting action. However, it suggests also that there is a value-action gap. High levels of worry do not always lead to environmental behavior. Sometimes, worried citizens do not perform any pro-environmental actions, or unworried citizens act ecologically. Some controversial examples are explained to evidence that thus a causal relationship between risk perception and behaviour does not exist. We examine this complicated relationship between these two concepts in depth. No study has tried to understand what variables influence and vary behaviour whether the level of risk perception changes. This part aims to denote what features and factors emerge in predicting behaviour, but among those who have the same level of climate change risk perception. Therefore, the next questions are: what are the most significant features that classify and predict the behaviour of worried citizens? What are unworried citizens influenced by to act ecologically? The relevance of the investigation is to cover the existing gap in the literature, answering these questions.

It is assumed that:

Hp5: worried citizens and unworried citizens are influenced in pro-environmental behaviour by different factors.

## **2.2 Data Description**

As aforementioned, the research studies the self-reported pro-environmental behaviour of European citizens. The main data used in this project come from one wave of Eurobarometer survey. The Eurobarometer is a public opinion research institution in the European Union to examine a variety of topics and attitudes. The European Commission conducts Standard & Special Eurobarometer periodically. We used the Special Eurobarometer 91.3 dataset, entitled “Climate Change”, made available by the Eurobarometer Open Data website. This survey is collected in April 2019 using face-to-face interviews. There are 27655 respondents from 28 countries of the European Union. The Eurobarometer data are publicly available from GESIS (European Commission, Brussels, 2019). Eurobarometer 91.3 asks some questions about environmental issues and some sociodemographic information. Some relevant items about climate change and sociodemographic variables are selected.[[1]](#footnote-1)

### **2.2.1 Data Cleaning**

The first step before performing the analysis is data cleaning. To obtain an accurate analysis, some observations are dropped or recoded.

The missing data of our dependent variables, *pro-environmental behaviour*, is dropped since the analysis is based on the predictions of a dichotomous outcome (coded as 1 = Yes, 0= No). Even if this is a self-reported behaviour, it is considered valid and accurate according to the literature, being a dichotomous variable.

*Risk perception* is measured on a 1-10 scale, and no answers are dropped to keep the variables as a metric, as some previous research had done (Echavarren et al., 2019). The question does not directly about the perceived risk but it is referred to the seriousness of the phenomenon in the present moment and it is a one-dimension of climate change risk perception (Echavarren et al., 2019).

Successively, other questions regarding climate change are selected (they are briefly renamed as follows: *benefits for companies, benefits for citizens, renewable energy, energy efficient,* and *greenhouse* *gas).* They areexpressed on a 4-point Likert scale. Also, in this case, missing or refusal data is removed. The reason why PAM clustering and CCA does not accept missing data and therefore the entire observation must be removed.

Instead, sociodemographic variables are for the most part categorical and therefore *refusal* or *dk* (don’t know) are kept among the answer choices. However, some transformations are adopted in these variables.

*Political orientation* is originally presented in a 10-points Likert scale (1 = left to 10 =right). It is transformed into a categorical variable: the answers 1-2 are become “left”, 3-4 “centre-left”, 5-6 “centre”, 7-8 “centre-right”, 9-10 “right” and *dk* or *refusal* “not positionable”.

For the *marital status*, some new categories are created depending on whether an individual has declared that he/she lives with “partner”, “partner and children” or he/she is “single” or “single (and he/she lives) with children”.

The *stop education* variable, which means when he/she finished studying, has been converted from continuous to categorical. According to scholars (Abu-Omar & Rütten, 2008; Loyen, 2016), five categories are created: “up to 15 years”, “16-19 years”, “20+years”, “still studying” and “refusal/other”.

*Gender* and *age* are not manipulated since nobody answered with “other” and therefore the first variable is a dichotomous “male” and “female” option, while the second one is maintained as continuous.

For the *living* variable, the categories proposed by the Eurobarometer are kept. Respectively, the first has the following classes: “rural area or village”, “small or middle-sized town”, “large town” and “dk” (don’t know).

Also, for the *social class* variable, the Eurobarometer’s categories are kept, using the options: “the working class of society”, “the lower middle class of society”, “the middle class of society”, “the upper-middle class of society”, “the higher class of society”.

Lastly, the *country* variable is considered. Eurobarometer surveys collected about 1000 interviews on average for each country, except for small nations, such as Malta and Luxembourg. Only manipulation is computed: West and East Germany are combined into one country “Germany”. [[2]](#footnote-2)

To sum up, the final dataset has 21978 respondents (out of 27655). [[3]](#footnote-3)

## **2.3 Methodology**

As already explained, the research mainly consists of two different parts: in the first, unsupervised learning algorithms are adopted to obtain citizens' profiles, and in the second, supervised learning algorithms are used to classify and predict pro-environment behaviour.

### **2.3.1 Unsupervised Machine Learning Algorithms**

The first set of methods focuses on identifying some citizens' profiles using different types of unsupervised learning techniques: Partition Around Medoids (PAM) clustering and Correlational Class Analysis (CCA). Unsupervised learning techniques look for unknown and hidden patterns.

PAM is a type of clustering, used primarily for categorical features (Shendre, 2020). It seeks to identify a finite set of clusters or subgroups to describe data (Fonseca, 2013; James et al., 2013). This method creates some subgroups to maximize both the similarity within clusters and the differences among other groups. This algorithm adopts Gower distance which calculates the distance between two objects whose properties are a mix of categorical and quantitative values (Shendre, 2020). Clustering is used to obtain profiles of citizens with similar attitudes toward climate change.

CCA identifies such “cultural schemas” in survey data, in particular in public opinion data (Boutyline, 2017; Rossoni et al., 2020). This technique is an implementation of Relational Class Analysis (RCA) developed by Goldberg (2011) and it “seeks to parse out groups, or classes, of like-minded individuals. Unlike these methods, however, it uses relationality to compare these individuals not on their attitudes per se but on the patterns of relations between their attitudes” (p.1399). Therefore, the goal of RCA is to partition individuals into groups that shared “cultural classes” (Rossoni et al., 2020). The shared “cultural schemas” “does not imply having identical attitudes or behaviours. Rather it suggests agreeing on the structures of relevance and opposition that make actions and symbols meaningful” (Goldberg, 2011, p. 1402). Therefore, it tries to find patterns of associations between attitudes or behaviours in terms of “relationality”. It tries to find relationships both between individuals and between variables, combining clustering analysis and multidimensional scaling or factor analysis (Goldberg, 2011). The difference between RCA and CCA lies in the concept of “relationality”. While Goldberg (2011) uses linear dependency between two singular vectors of answers to find the shared cultural schemas, CCA suggests adopting Pearson’s correlation (Boutyline, 2017). Boutyline (2017) demonstrated that CCA produces more accurate results. In this case, there is no relationship between cases as in clustering, rather than between variables.

Only some climate change questions are selected to fit these two methods and they are *benefits for companies, benefits for citizens, renewable energy, energy efficient, greenhouse gas.* The questions deal with different aspects of climate change, such as actors (companies or citizens) who would benefit if the phenomenon were fought and what kinds of strategies (use of renewable energy, improving energy-efficient or reducing greenhouse gas) governments should adopt.[[4]](#footnote-4) The responses were on a 4-point Likert scale, with the following gradations and labels (the latter change according to the questions):

1 = Totally agree/ Very important

2 = Tend to agree/ Fairly important

3 = Tend to disagree/ Not very important

4 = Totally disagree/ Not at all important

Some summary statistics are described.[[5]](#footnote-5) The averages of *benefits for companies* and *benefits for citizens* variables areslightly greater than the other three, respectively they are 1.7 and 1.9. While the averages of *renewable energy, energy efficient, greenhouse gas* variables are all 1.5. However, if we examine the two extremes of the questions’ answers (1 = totally agree/ Very important to 4 = 1 = totally agree/ very important), we can declare that citizens give more importance to the different strategies to be adopted to fight climate change. The reason is that the averages of *renewable energy, energy efficient, greenhouse gas* variables as are closer to 1 than *benefits for companies* and *benefits for citizens*.

The purpose of this part of the analysis is to partition some similar types of citizens, called clusters or classes, that better describe the data used. In fact, through these techniques, some new segmentations of citizens could be identified and then they could help to find new explanations to the phenomenon studied. Theoretically, using these two different types of segmentation of citizens, the results should be opposite. On the one side, the traditional clustering profiles the data according to similar attitudes, therefore we will find different types of *green identity*. On the other side, CCA finds shared *cultural schemas*, structure of thought. Eventually, the classes obtained from PAM clustering and CCA are used as predictors in the subsequent classifications.

### **2.3.2 Supervised Machine Learning Algorithms**

The second set of methods focuses on predicting self-reported pro-environment behaviour using different types of supervised learning techniques and classifiers. Classification is used when a categorical variable is predicted (James et al., 2013). “The methods used for classification first predict the probability of each of the categories of a qualitative variable” (James et al., 2013, p. 127).

The different techniques of classifiers are briefly presented below. We have decided to use different methods because we want to compare the results obtained and, we want to find an optimal model that best predicts the behavior. These classifiers randomly split all data into training (70%) and testing sets (30%), and the output labels are *yes* or *no*, depending on whether the environmental action has been taken place.

The pro-environmental behaviour prediction starts with a Logistic Regression. It is a form of binary regression and it explains relationships between a categorical outcome and some continuous or discrete predictors (Peng et al., 2002). It models the probability of being to a particular category (Peng et al., 2002; Stoltzfus, 2011).

The model requires some assumptions:

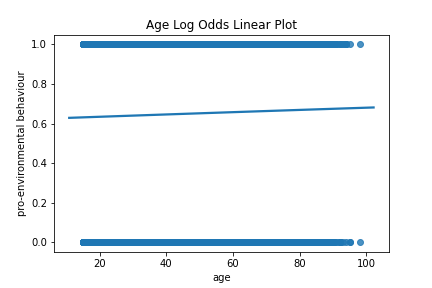
1. linearity in the logit for continuous independent variables;
2. the absence of multicollinearity among explanatory

variables;

1. the absence of extreme outliers

(Stoltzfus, 2011)

However, some assumptions are violated. All the assumptions are tested on continuous variables, thus in this case only on the age variable.

There is no present linearity in the logit for the age variable, as shown in figure 2 (there should be an “s” curve line).

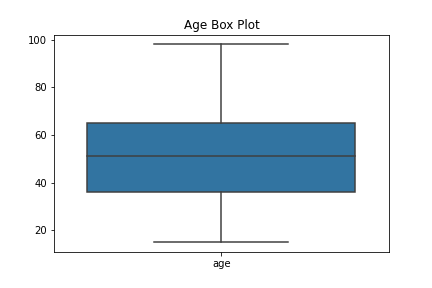
 Second, the absence of multicollinearity (independent variables) can not be checked, due to it is not possible to compute correlation with one variable alone.

Figure 2: Linearity in the Logit for Age

Lastly, lack of outliers, always for continuous variables.

Figure 3 displays that there are no outliers in the age variable.

Figure 3: Lack of outliers for Age

Despite the robustness of the Logistic Regression models, data cannot fully satisfy the assumptions, therefore the Decision Tree algorithm is fitted. Decision Tree is a “flow-chart-like hierarchical tree structure” (Jenhani et al., 2008, p. 786) and it is composed of three elements: nodes, edges, and leaves. Nodes represent attributes or variables, edges match the various potential attribute values and lastly, leaves include items that belong to the same class (Jenhani et al., 2008). The main advantages of Decision Tree are that it has not assumptions and especially it produces graphical representation, which makes it easier to read and interpret the model.

The analysis continues with another robust model: Random Forest, which is produced randomized multiple decision trees that work as an ensemble (Belgiu, 2016; Biau & Scornet, 2016). This classifier “can successfully handle high data dimensionality and multicollinearity, being both fast and insensitive to overfitting” (Belgiu, 2016, p.24). Another advantage is that it can be dealing with unbalanced data, as in this case (Belgiu, 2016).

The last classifier used is Gradient Boosting. It is similar to the Random Forest algorithm, but in this case, each new tree is been created using the previous ones, to correct mistakes made (James et al., 2013). Instead of fitting a large number of trees separately, it learns slowly from previous trees recursively.

The last two tree-based methods, producing multiple trees, have become more popular since they improve prediction accuracy but they lose the immediacy of interpretation (Belgiu, 2016; James et al., 2013).

To sum up, all these classifiers predict pro-environmental behaviour. Figure 4 shows the distribution of the complete dataset of the observations according to the dependent variable. We have 14327 (65%) individuals who declared to have taken any action to fight climate change over the past six months and 7651 (35%) who have not. In this dataset, the distribution is decidedly unbalanced.

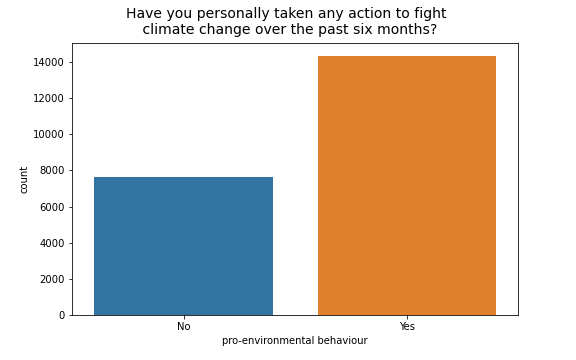


Figure 4: Pro-environmental behaviour distribution

Sociodemographic variables, including country, green-identity and cultural schemas (classes created respectively from PAM and CCA), and climate change risk perception are used as predictors. We want to investigate the main factors and predictors that influence pro-environmental behaviour. This process is achieved thanks to selected models, logistic regression, and tree-based methods, determining the importance of independent variables. This part is considered quite conventional according to the above-mentioned literature review: variables selected have already been used in previous studies, even if have mostly used more traditional techniques (and not machine learning techniques). An innovative point is PAM clustering and CCA’s use to create new groups where individuals belong to. Additionally, another originality of this research can be found in the next step.

Due to the theoretical relevance of climate change risk perception, in the pro-environmental behaviour’s prediction, two different analyses, according to the degree of this main explanatory variable, are performed. Two datasets are created: one with only the observations of individuals who declared very worried about the phenomenon (responses with a score greater than or equal to 6 are considered), and one with those who do not worry (score less than or equal to 5). The same algorithms, just described, are fitted for the two different subsets. The aim is to discover the divergent variables that predict actions and whether there are relevant differences between worried and unworried citizens.

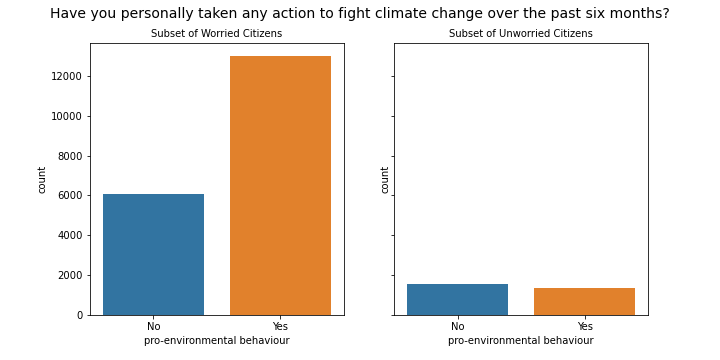


Figure 5: Pro-environmental behaviour distribution according to climate change risk perception

Figure 5 indicates the distribution of our dependent variable according to the two created subsets. The first subset with the observations (N= 19072) of those who have a high-risk perception is larger: 12988 worried citizens (68%) who engage any ecological behaviour and 6084 worried individuals (32%) who do nothing. This subset is quite similar to the complete dataset if we check the percentage of the two types of behaviours.

Instead, the second subset is composed of the observations (N= 2906) of those who do not worry about the environment. We have few cases, but they are balanced: 1339 (46%) unworried citizens who perform any environmental behaviours, and 1567 (54%) unworried citizens who do not perform any actions. In this case, for the subset of unworried citizens, the individuals who do not perform some eco-friendly actions are greater than those who do.[[6]](#footnote-6)

Chapter 3

# Exploratory Data Analysis

The following section illustrates the different steps undertaken to obtain a prediction model of pro-environmental behaviour. Primarily, the results of unsupervised machine learning algorithms are shown. Correlational Class Analysis and Partitioning Around Medoids Clustering are performed and the created classes are inserted in the final prediction models.

Performing these two algorithms we obtain the final dataset, and therefore we can continue with the second step: Exploratory Data Analysis. This analysis concerns a quick description of the complete dataset’s explanatory variables and then of the two distinct subsets based on individual climate change risk perception level. As we have already mentioned in the Methodology (Chapter 2), in the first place the prediction is fitted to the entire dataset. In the second place, the prediction is computed separately on two subsets: with the observations of individuals who have a high-risk perception (greater or equal to 6) and with the observations of those who that a low-risk perception (less than or equal to 5).

The last section of the chapter shows some characteristics at the macro-level. We have discussed that some of the factors that shape pro-environmental behavior could be regulations or habits of countries. For this reason, some descriptive analyses grouped per country are implemented.

**3.1 New segmentations among citizens**

***Correlational Class Analysis***

As previously stated in Chapter 2 (2.3.1), CCA is an unsupervised machine learning algorithm, which partitions individuals into separate classes, identifying different types of “cultural schemas”. The shared cultural schemas do not mean to group according to similar attitudes but rather to recognize a similar structure of thought among citizens (Goldberg, 2011). CCA aggregates individuals who perform similar decision patterns, although they do not coincide with the answers given (Rossoni et al., 2020). The aim is to find some shared cultural schemas among citizens. It can control the rules of thinking and interpretation the reality and in this case climate change topic. The algorithm is performed through the *corclass* package in the R software using climate change items (*benefits for companies, benefits for citizens, renewable energy, energy efficient, greenhouse gas*).

CCA divides the sample into five correlational classes: group 1 includes 5704 cases (26%), group 2 4415 (20%), group 3 6106 (28%), group 4 3271 (15%), group 5 2482 (11%). Figure 6 illustrates the individual modules or classes created as a network. Each node coincides with one item, while the edges reveal the statistically significant correlation between variables (Rossoni et al., 2021). The thicker the line, the more significant the correlation. [[7]](#footnote-7)



Figure 6: CCA. Coding: qb4\_3 = benefits for companies, qb4\_5 = benefits for citizens, qb7 = renewable energy, qb8 = energy efficient, qb9 = greenhouse gas

In group 1 the correlations between all pairs of variables are set to 1. Actually, group 1 is called “zero class”, due to these rows get 0 variance and to default, R sets the correlations between all pairs of 0 variance rows to 1. Therefore, there is no sharing cultural meaning in this group.

Group 2 obtains a high positive correlation (0.8) mainly between *benefits for companies* and *benefits for citizens*. These two questions concern the benefits that would be obtained if climate change were to be fought.

Group 3 has a slight correlation among all variables, except for *benefits for citizens*. On average, the correlation among these four variables (*benefits for companies, enewable energy, energy efficient, greenhouse gas*) is 0.5. Actually, *benefits for citizens* variableis always slightly correlated with only *benefits for companies* (0.07), but it is not significant as the others.

In group 4 there is a strong positive correlation between *renewable energy* and *energy efficient* (0.8); *renewable energy* and *greenhouse gas (0.8); energy efficient and greenhouse gas* (0.8).Separately between *benefits for companies* and *benefits for citizens* (0.9). The first block of variables mainly concerns the importance of the government decisions to fight climate change and the resulting benefits for companies or citizens; while the second block of variables, as we have already seen, concerns Europe’s role in renewable energy, energy-efficient, and greenhouse gas issues.

Lastly, group 5 has a strong correlation between the following pair of variables: *benefits for citizens* and *renewable energy* (0.7); *benefits for citizens* and *energy efficient* (0.7); *benefits for citizens* and *greenhouse gas* (0.8). In this case, *benefits for companies* variable is completely isolated.

Figure 7: CCA and Country

In the second place, it is interesting to analyse how CCA’s groups are distributed according to country, as shown in figure 2. The aim is to understand whether there are some shared cultural meanings among citizens of the same country. We examine the percentage of individuals belonging to the CCA’s groups for each country. Group 2, 4, and 5 are composed of a similar proportion of citizens from each country. Instead, we notice that over 30% of citizens from Cyprus, Ireland, Malta, Portugal, Slovakia, Spain, and the United Kingdom belong to group 1. Additionally, over 30% of citizens from Austria, Bulgaria, Czech Republic, Estonia, Finland, France, Germany, Greece Italy, Latvia, Luxemburg, and Sweden belong to group 3. In both cases, we can not detect some specific patterns, such as belonging to the same part of Europe. It is interesting to find similar shared meanings schemas among subjects in nations extremely different and opposite such as Italy and Estonia.

***Partitioning Around Medoids Clustering***

PAM clustering is an unsupervised method that looks for patterns without any information of the classification target. The aim is to partition citizens into classes (clusters) according to similar *green-identity*. Clustering groups the similar observations within each group, while the observations in different groups are different from other clusters. Grouping the attitudes of citizens in clusters means dividing them according to similar opinions about the climate change topic. We stress that also in this case, like CCA algorithm, the questions about climate change (*benefits for companies, benefits for citizens, renewable energy, energy efficient, greenhouse gas*) are fitted in that algorithm. Also for this PAM algorithm, We use the *cluster* package in the R software.

PAM requires, as a parameter, the number of clusters. “One of the most commonly applied methods for assessing cluster validity is silhouette width which encompasses two clustering criteria: separation (i.e., the average distance to the closest other clusters) and compactness (i.e., average within‐cluster distance)” (Lengyel & Botta‐Dukát, 2019, p. 13232). Silhouette width indicates how well each cluster divides observations, as shown in figure 8. The best choice is 2.

*Figure 8: Clustering silhouette*

Observations are then divided into two clusters with PAM. The summary results of PAM clustering are shown in figure 9. The distribution is balanced: 11171 (51%) individuals belong to cluster 1, and 10807 (49%) belong to cluster 2. Although overall the level of agreement or importance of these questions is elevated, se we have describe in Methodology (2.3.1) we can recognize that the two clusters resemble two different types of green-identity, which we call “moderate”, “extreme”. We evidence that citizens belonging to the moderate green-identity cluster have mainly answered questions with moderate answers (option 2 or 3). The opposite case happens in the extreme green-identity cluster. More than 60% of answers to *benefits for companies* and *benefits for citizens* questions yield the maximum answer (option 1); additionally, about 90% of answers to *renewable energy, energy efficient, greenhouse gas* questions have the maximum answer (1).

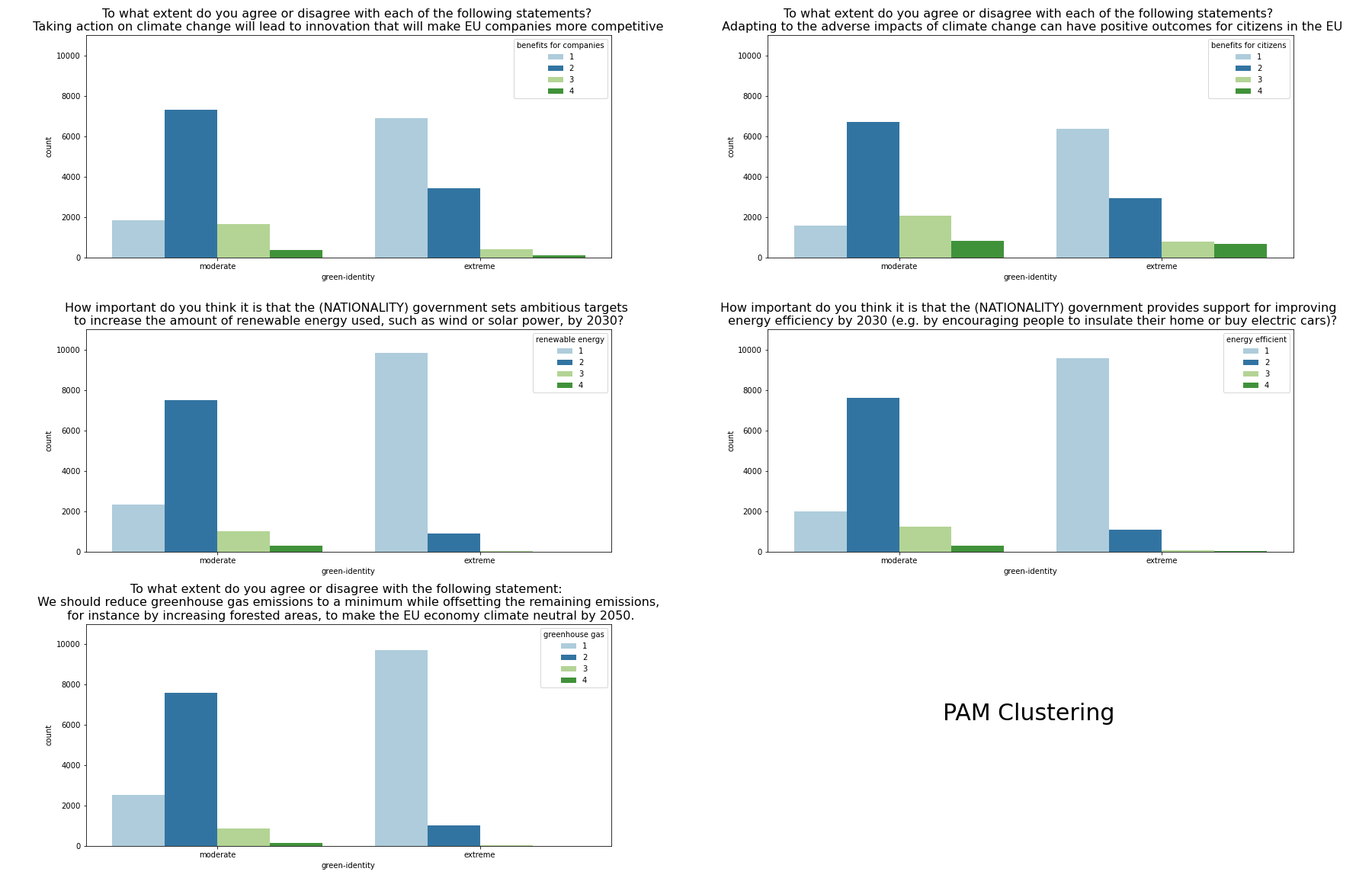


Figure 9: Summary of PAM Clustering (N=21978): extreme green identity cluster (n=10807) versus moderate green identity cluster (n=11171)

Table 1: Description of the clusters

|  |  |  |
| --- | --- | --- |
| **Variable** | **Cluster 1 (moderate)** | **Cluster 2 (extreme)** |
| **Position on green issues** | the mode in all questions is option 2 | the mode in all questions is option 1 |
| **Risk perception** | the average climate change risk perception is 7.2 | the average climate change risk perception is 8.6 |
| **Pro-environmental behaviour** | 56% perform a pro-environmental behaviour | 75% perform a pro-environmental behaviour |

If we analyse the citizens' profile into the clusters, we find that there is an opposite position both on green issues and in the variables of the individual dimension, as table 1 displays. The moderate green-identity sample is slightly less worried about climate change and less likely to engage in pro-environmental behaviour than the extreme green-identity sample. Additionally, we can analyse the sociodemographic characteristics of citizens according to the two opposite clusters. Other differences between the two clusters are that the citizens of the extreme green-identity sample are slightly more educated (41% of the citizens in this cluster stopped studying after 20 years old) than moderate one (the percentage drops to 35%). The citizens of the extreme sample are slightly younger (the average is 50) than the moderate one (51). Concerning political orientation and gender, there are no particular differences among classes. Finally country: more than 65% of the citizens of Czech Republic, Estonia, Finland, Latvia, Poland belongs to moderate green identity, insted of the citizens of Cyprus, Denmark, Spain, United Kingdom belong to extreme green-identity cluster.

To sum up, cluster 1 is composed of citizens that are more moderate in the answers to climate change items than the citizens of the opposite cluster. Additionally, cluster 2 is composed of slightly younger, more educated, more worried, and more active citizens.

## **3.2 Explanatory Data Analysis**

**The Entire Dataset**

Before starting with the analysis it is important to describe the dataset. Figure 10 shows a quick summary of the explanatory variables. Dataset is composed by:

* 80% of citizens have a central political orientation (between centre, centre-left, and centre-right);
* 70% of individuals live with partner and/or children;
* 42% of individuals stopped studying between 16-19 years old, while 38% after 20 years old;
* 52% of women;
* the average age is 50.5;
* 39% of individuals live in a small or middle-sized town;
* Germans are over-represented when compared to the number of other citizens of other countries;
* climate change risk perception average is 7.9, while the mode is the answer 10.

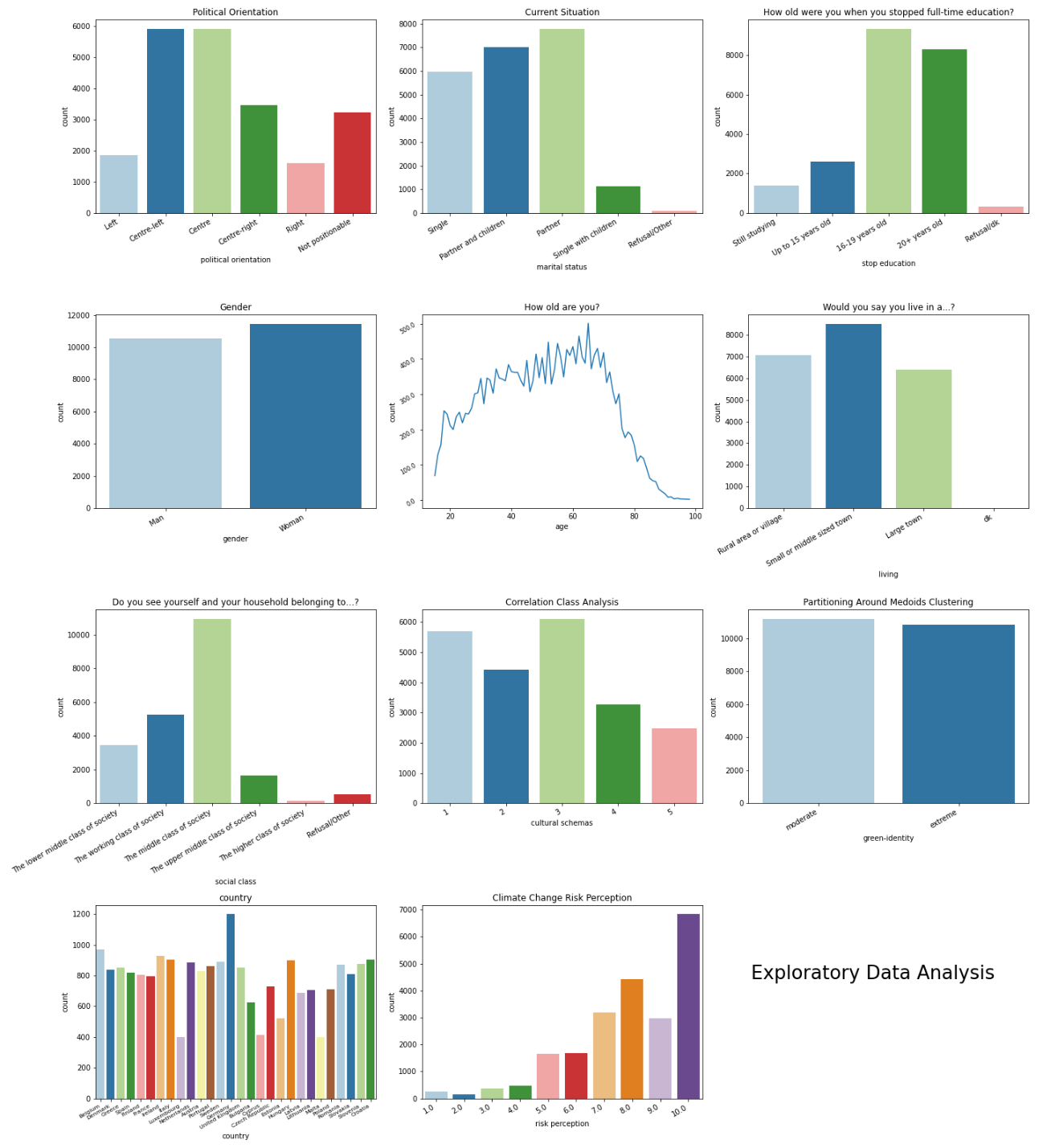
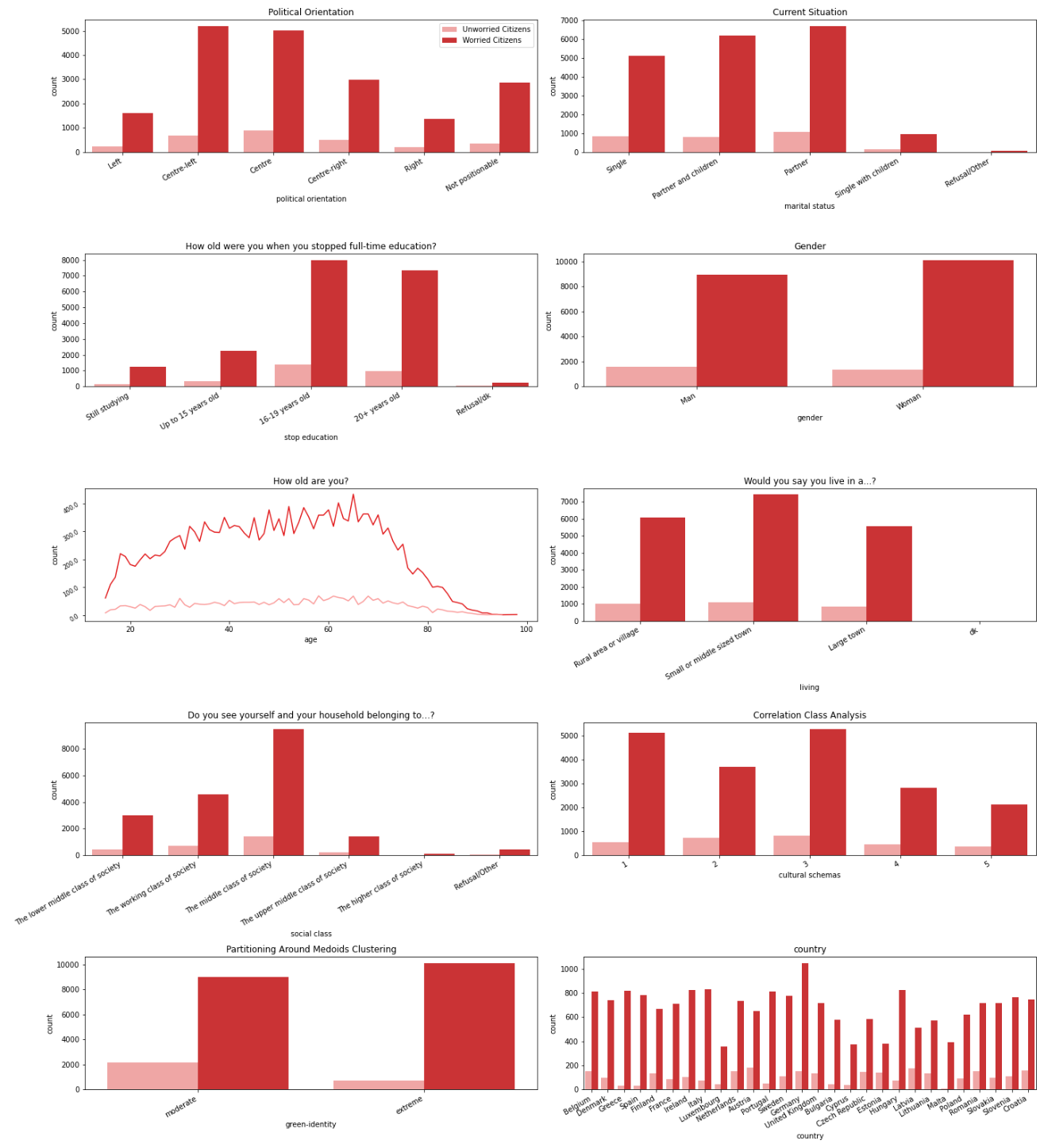
****

Figure 10: Countplot of the explanatory variables-the entire dataset (N=21978)

**The Datasets based on the Individual Risk Perception Level**

Instead, figure 11 displays the description of the two datasets based on the individual climate change risk perception level. As you can see, the two subsets are not balanced.[[8]](#footnote-8) The subset with unworried citizens (in the graph it is represented with pink colour) has 2906 observations (13%), while the subset of worried citizens (red colour) has 19072 cases (87%). The trends of all variables of the high-risk perception subset follow those of the complete dataset. Instead, the low-risk perception subset is slightly different. For example, individuals of this latter subset are slightly older and mostly men. Lastly, as we have already explained, the distribution of clusters between the two subsets is opposite: on the one side mostly individuals who belong at cluster 2 (extreme green-identity) have a high-risk perception, and on the other side, those who belong at cluster 1 have a low-risk perception.

*Figure 11: Countplot of the explanatory variables of subset of worried citizens (n= 19072) versus subset of unworried citizens (n= 2906). (N = 21978)*

## **3.3 Country-level statistical data**

Climate change attitudes do vary only between citizens but also between countries (Xie et al., 2019). As you can see in figure 12, the percentage of those who believe that climate change is the single most serious problem varies significantly according to country. For example, Bulgaria and Croatia obtain a smaller percentage, i.e., 11% of citizens consider climate change as the single most serious problem. On the contrary, about 1 out 2 of Swedish indicate climate change.

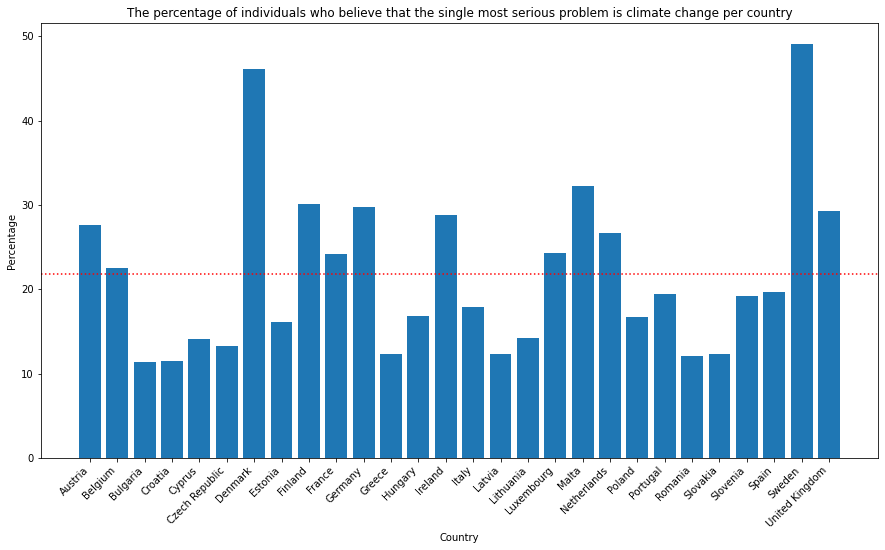


Figure 12: Single Most Serious Problem per Country (N = 21978)

Another interesting example is the difference in behavior. As you can see in figure 13, more than 60% of the citizens of Romania declared that they are not behaving in favour of the environment. Conversely, Malta has 95% of citizens that enact pro-environmental behavior. These differences may be due to a multitude of reasons, differences in economic incentives and taxes, way of socialization, social desirability that vary according to country.

Figure 13: Relative frequency of pro-environmental behaviour per country (N = 21978)

Attitudes among countries could so vary since they are influenced by different contextual factors (Echavarren et al., 2019; Krajhanzl, 2010). According to Echavarren and colleagues (2019), opinion, perception, and behavior could change due to different natural hazards and political contexts. For example, water deficit, temperature growth, or the level of environmentalism in the political activity of a given nation may shape citizens’ attitudes or behaviours (Echavarren et al., 2019). These macro-variables could be significant mediators in explaining risk perception or pro-environmental behavior. Some indexes are described and examined according to country. It is important to evidence the differences at the country-level due to could affect and moderate the individual attitudes towards climate change. Then, they are not inserted in the final models since only the multilevel analysis could be adopted. Besides, the aim of the research is not to evidence national or cultural differences, but on the contrary, it is to find patterns at individual levels, regardless of the place of origin. However, these differences at the country-level are presented.

For the natural hazards, the 2020 Environmental Performance Index (EPI) is used (the 2019 EPI is not available to use the same data of year of this survey) (Yale Center for Environmental Law & Policy, 2020). EPI quantifies numerically environmental health and ecosystem vitality around the world. This index helps us to describe with a score the severity of environmental problems in each country. Some indicators that composed the index are air pollution, drinking water quality, species protection. These phenomena could positively affect climate change concerns and opinions (Echavarren et al., 2019). Citizens should perceive biodiversity loss or temperature increases, leading to greater apprehension. Figure 14 shows the score across European Union (EU). The best score is obtained from Denmark, while the worst from Bulgaria.

Figure 14: The 2020 EPI

For the political context, the 2019 Climate Change Policy Performance is selected, which is a measurement of national and international climate policies (Burck, 2018) developed by organization Germanwatch. It is one of the indicators that belong to the Climate Change Performance Index (CCPI). The indicator constitutes the measurements taken by governments to reduce the current level of GHG emissions per capita or the use of renewable energy. Briefly, it is defined as a measure of countries’ progress and their capacity for climate protection (Burck, 2018). In the Climate Change Policy, the record goes to Portugal and Bulgaria gets the lowest score in all European Union, as figure 15 shows.

According to scholars (Echavarren et al., 2019; van der Linden, 2015), socio-cultural context influences individual attitudes towards climate change concerns. Therefore, the notable differences in attitudes across countries should be also due to these indicators. In fact, “sociological research suggests that contextual factors and processes can be powerful forces shaping how individuals and communities engage with the issue” (Lee et al., 2015, p. 1014). There are different ecological tax reforms or cultural habits that affect and shape individual climate change attitudes and behavior.

Figure 15:The 2019 Climate Change Policy

Chapter 4

# Analysis & Prediction

As mentioned in the literature review (Chapter 1), the main predictor of pro-environmental behaviour is climate change risk perception. However, other factors can shape the outcomes. The main focus is now to predict the behaviour of citizens in a dummy outcome in the entire dataset, thus using all the observations. This part aims to understand the actual role of climate change risk perception. The second step is to perform the pro-environmental behaviour prediction, but the prediction is divided into two groups of citizens based on the individual climate change risk perception. We want to understand whether factors that shape behaviour change according to risk perception level.

## **4.1 Evaluating Classification Models**

Before proceeding with the analysis, it is important to explain the ways the evaluation will be measured, since one of the main factors by which we will choose the best model is the maximization of these metrics. We consider two measurements of evaluations:

* **Accuracy**: the proportion of correct predictions to the total number of predictions (true positives and true negatives) given by the classifier (Battiti & Brunato, 2014). Accuracy is the most used metric that generally describes the goodness of a model. The formula is:

The range is 0-1, where 0 is the worst score and 1 the best score, all the inputs are predicted in the right way.

* **Macro F1-score average** (macro-F1 for short): combine precision and recall, where the precision is the fraction of the number of true positives among the total number of items labelled as positive, and the recall is the fraction of numbers of true positives among the total number of features that belong to the positive class (Shmueli, 2019). macro-F1 computes the harmonic mean, as shown in the formula:

The scoring range is 0-1, where 0 is the worst score and 1 is the best one. It is used, especially, when for the class imbalance problem due to it is more sensitive to data distribution, as in this case.

To select the best model for each classifier, we decide to balance between accuracy and macro-F1. The aim is to classify and predict as well both classes. On the one side, this compromise can sometimes lead to losing a few percentage points of accuracy but, on the other side, it improves accuracy within the classes.

In closing, we explain how to understand what variables have a fundamental role in our models. We compute feature importance by the different algorithms in *scikit-learn*. Logistic regression finds a set of coefficients to use in the weighted sum to produce a prediction. These coefficients can be used directly as a feature importance score. Instead, tree-based models can measure feature importance in two ways: Gini Importance or Mean Decrease Accuracy. Gini Importance counts the times a variable is employed to split a node. Mean Decrease Accuracy estimates how much accuracy the model losses by omitting each feature. However, Mean Decrease Accuracy is not implemented in the *scikit-learn* package, thus the Gini Importance method is adopted in the analysis as the measure of feature importance.

## **4.2 Prediction of the Behavior on the Entire Dataset**

In this part, the research aims to identify and discover the most significant variables and offer a good prediction of pro-environmental behaviour. As presented in the Methodology (Chapter 2), different classifiers are trained and implemented to predict behaviour. The analysis starts with Logistic Regression and continues with tree-based methods: Decision Tree, Random Forest, and Gradient Boosting.

These algorithms are implemented in *scikit-learn*, using the following functions in order: LogisticRegression(), *DecisionTreeClassifier(), RandomForestClassifier(), XGBClassifier().* For each classifier best tuning parameters, called hyperparameters, are fitted. Adopting a fitting set of hyperparameters is both significant in terms of model accuracy and computationally challenging. The technique adopted for knowing the optimal hyperparameter is called *random search* (RandomizedSearchCV() in *scikit-learn*) (Benner, 2020). We adopt, in turn, for random search 3 fold cross-validation, a resampling procedure to evaluate model performance. In practice, each possible random combination will train and evaluate for three different folds.[[9]](#footnote-9)

As mentioned above that the independent variables, at an individual level, of this part are:

* Risk perception
* Cultural schemas
* Green-identity
* Political orientation
* Marital status
* Stop education
* Gender
* Age
* Living
* Social Class
* Country

Each independent categorical variable is converted into a dummy, to fit the supervised algorithms. All explanatory variables are categorical, except for risk perception, which is maintained as a metric, and age. After the conversion, the dataset has 65 independent variables.

It can be convenient, to summarize quickly, the predictive accuracy and macro-F1 of the models trained are indicated in table 2, while the comparison of feature importance is shown in figure 16.[[10]](#footnote-10) The accuracy and macro-F1 are slightly better in the Random Forest’s model, which are respectively 0.70 and 0.67.

Table 2: Metrics comparison

|  |  |  |
| --- | --- | --- |
| **Models** | **Accuracy** | **Macro-F1** |
| **Logistic Regression** | 0.67 | 0.65 |
| **Decision Tree** | 0.64 | 0.62 |
| **Random Forest** | 0.70 | 0.67 |
| **Gradient Boosting** | 0.69 | 0.65 |

|  |  |
| --- | --- |
|  |  |
| Figure 16:Comparison of variables importance |  |

The Logistic Regression model yields 0.67 of accuracy and 0.66 of macro-f1. The behaviour varies according to different countries and is particularly not widespread in Romania, Bulgaria, and Czech Republic. It very interesting due to all these first countries are situated in Eastern Europe. Instead, Finland, Malta, and Portugal have a positive influence on predicting pro-environmental behaviour. [[11]](#footnote-11)

The Decision Tree model generates an accuracy of 0.66 and a macro-F1 of 0.53. The extreme green-identity cluster is the predictor variable used for the primary split. Risk perception is the predictor variable used for the second and the third split. Whether the individual belongs to the extreme green-identity cluster and he/she has a level of risk perception greater or equal than 7,5 he/she is classified in the *yes-action* class.[[12]](#footnote-12) If the individual does not belong to cluster 2 and he/she has a level of risk perception less or equal to 6.5, he/she is classified in the *no-action* class.

The accuracy and the macro-F1 of the Random Forest model greatly improve, they are respectively 0.70 and 0.66. It is the best model fitted. We notice that age and, once again, climate risk perception are the most important variables to classify and predict pro-environmental behaviour. Other important features are both created clusters, which correspond to the different green-identity. As we have already seen, the moderate green-identity cluster is more likely to do not behave in favour of the environment than the moderate one.

Gradient Boosting concludes the first part of the analysis. The model gets slightly worse results than the Random Forest classifier. The accuracy is 0.6 and also macro-F1 is 0.6. Again, the variable importance is checked and compared to the previous list. The moderate green-identity cluster is the best predictor. It is followed by Romania and climate change risk perception.

To sum up, most of the features with predictive importance are the same across models. Moreover, these predictors align with the hypotheses presented in section 2.1.

As referred to in the literature review (Chapter 1), the more an individual worries, the more he/she tends to perform some environmental actions. This statement is confirmed in the data. If we analyse the percentage of those who perform actions within the level of perceived risk, we discover that the more citizens decrease their climate change risk perception, the more likely they are not to behave ecologically. And vice versa, the more the risk increases, the more the percentage of performing action increase in turn, as figure 17 indicates. Climate change risk perception is one of the predictors which align in almost all models. Therefore, we confirm that a higher individual climate change risk perception positively influences and predicts pro-environmental behaviour.

Figure 17: Crosstab plot between Risk Perception and Behaviour

Self-green identity, measured with clusters, is another important variable found in almost all models. 75% of individuals belonging to the extreme green-identity cluster act ecologically. While 55% of individuals belonging to the moderate green-identity perform pro-environmental behaviour. Extreme green-identity has a positive influence on predicting behaviour. Therefore, also the second hypothesis is confirmed: the more the green-identity is strong, the more increase the likelihood to behave pro-environmentally.

However, the third hypothesis is rejected. Cultural schemas, measured with CCA classes, are not particularly relevant to the prediction.

The last hypothesis concerns the importance of sociodemographic information. Higher education has a positive effect on pro-environmental behaviour and it is one of the most important variables in the Decision Tree and the Gradient Boosting model. 72% of those who have declared to stop study after 20 years perform pro-environmental behaviour. The percentage is considerably higher than the other categories that have studied less or than students.

Gender, income, and political orientation seem not to be relevant.

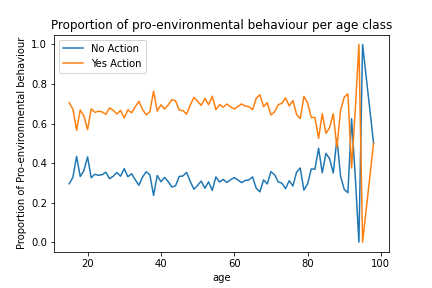
The Random Forest, our best model, suggests that the most important variable is age. However, it is not clear the relationship with the behaviour. We expect that younger individuals engage in pro-environmental action. If we examine the percentage of those who perform pro-environmental actions within the age class, we find that not only youngsters behave pro-environmentally, as shown in figure 18. The trends are constant up to the age of 80.

Figure 18: Crosstab plot between age and behaviour

To sum up, we can confirm only hypothesis 3 (a) concerning the role of education, while hypothesis 3 (b, c, d, e) can not be confirmed. However, we discover that a fundamental role in the prediction is the country of origin. Especially Eastern European countries are an important variable in all models, especially Romania. In these countries, pro-environmental behaviours are not widespread.

**4.3 Prediction of the Behavior based on the Individual Risk Perception Level**

At this point of the analysis, we can confirm the hypothesis suggests from the literature review: climate change risk perception is one of the main factors for predicting pro-environmental behaviour. Now, we want to understand what the most important predictors are if we change the level of risk perception. Thus, in the second part of the analysis, we split the data into two: one with the only observations of those who have a high-risk perception (level greater than or equal to 6) and those who have a low-risk perception (level less than or equal to 5). For convenience, we will call the first subset “worried citizens” and the second one “unworried citizens”.

Also in this part, we implement the same algorithms for both subsets: Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. Hyperparameters using random search are found.[[13]](#footnote-13)

The independent variables, at an individual level, of this part are:

* Cultural schemas
* Green-identity
* Political orientation
* Marital status
* Stop education
* Gender
* Age
* Living
* Social Class
* Country

Tables 3 and 4 summarize the performance of all models. In both cases, the Random Forest model has yielded the best performance when compared to other classifiers. However, the Gradient Boosting model has the best macro-F1 score in the subset of worried citizens. This subset is particularly unbalanced, therefore macro-F1 is preferred to evaluate the model.

*Table 3: Metrics Comparison-**Subset of worried citizens.*

|  |  |  |
| --- | --- | --- |
| **Models** | **Accuracy** | **Macro-F1** |
| **Logistic Regression** | 0.65 | 0.63 |
| **Decision Tree** | 0.63 | 0.60 |
| **Random Forest** | 0.70 | 0.60 |
| **Gradient Boosting** | 0.66 | 0.64 |

Table 4: Metrics Comparison-Subset of unworried citizens.

|  |  |  |
| --- | --- | --- |
| **Models** | **Accuracy** | **Macro-F1** |
| **Logistic Regression** | 0.63 | 0.63 |
| **Decision Tree** | 0.60 | 0.58 |
| **Random Forest** | 0.64 | 0.63 |
| **Gradient Boosting** | 0.61 | 0.60 |

|  |  |
| --- | --- |
| Subset of worried citizens | Subset of unworried citizens |
| Logistic Regression | |
|  |  |
| Decision Tree | |
|  |  |

|  |  |
| --- | --- |
| Random Forest | |
|  |  |
| Gradient Boosting | |
|  |  |

Figure 19: Comparison of variables importance based on the individual risk perception level

Figure 19 displays the comparison among models of the first 20 feature importance for predicting pro-environmental behaviour based on the different levels of climate change risk perception. [[14]](#footnote-14)

The accuracies of the two Logistic Regression models are 0.65 for the subset of worried citizens, and 0.63 for the subset of unworried citizens. The important variables in the two different models are similar. For both models, Romania has a negative influence on predicting pro-environmental behaviour, while Malta has a positive influence.

The accuracies of the Decision Tree models are 0.63 and 0.60. These results are the worst when compared with the other models. For both trees, the root is the green-identity. On the one hand, in the first tree the root is occupied by the extreme green-identity cluster, on the other hand, the root is employed by the moderate green-identity cluster. We have another confirmation of the hypothesis formulated. A strong green-identity, mixed with a high-risk perception, influences positively pro-environmental behaviour. The opposite happens in the second model: a moderate green-identity mixed with a low-risk perception influences anti-environmental behaviour.[[15]](#footnote-15)

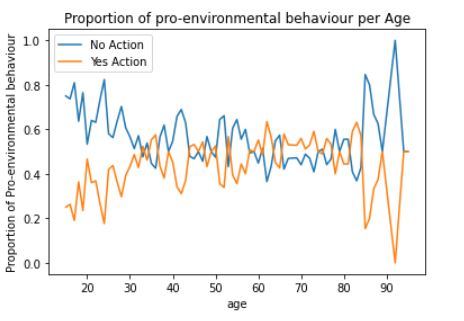
The Random Forest model yields the best accuracies in both models, 0.70 and 0.65. In both cases, age is the most important variable for the prediction of behaviour. The trend of age in the subset with high-risk perception is similar to the complete model. On average 60% of individuals of each age group perform eco-friendly actions. Another time, this percentage drops from 80 years old. More interesting is the trend in the subset of unworried citizens, as figure 20 shows. On average 70% of younger individuals do not behave in favour of the environment. This percentage drops with the increase of years, and the trend is reversed. This trend is opposite to the sociological theories explained in Literature Review (Chapter 1). Then, the subset of worried citizens has many important variables but in a much lesser way, such as the small or middle-sized town, 68% of this class in this subset perform environmental-friendly action.

Figure 20: trend of age in the Subset with low-risk perception observations.

Accuracies of the Gradient Boosting models yield 0.67 and 0.65. For the subset of worried citizens, the macro-F1 is 0.64, which is the best result. In both models, there are two main important variables: the moderate green-identity cluster, and Romania. In the subset of worried citizens, there is also the high-education level as an important variable. In the second subset of unworried citizens, we find two new variables: Slovenia and right (political orientation). Slovenia has a positive influence on pro-environmental behavior, due to 70% of Slovaks in this subset perform eco-friendly actions. Then, for the first time, political orientation turns out to have an important role. If we compute the proportion in this subset 50% of radicals engage in pro-environmental behaviour, while the other orientations get a smaller percentage.

To sum up, in this second part of the analysis we do not discover any significant differences in the comparison between the two opposite models. The important variables remain roughly the same. The high level of education is an important variable for worried citizens. In this subset, 75% of high-educated citizens perform pro-environmental behaviour. Therefore, high-education mixed with high-risk perception may positively influence to behave pro-environmentally. Another relevant variable is the extreme-identity. Thus, once again, the more strong is the green-identity of the individual, mixed with a high-risk perception, the more likely he/she behaves pro-environmentally. The opposite case is the subset of unworried citizens. In this case, Romania and the moderate green-identity cluster, are the variables found most often in models. 85% of Romanians who do not worry about climate change do not perform any actions. Therefore, this variable has a negative influence on predicting pro-environmental behaviour. Also, moderate green-identity could have a negative influence on the prediction of pro-environmental behaviour, albeit in a less direct way. 58% of those who do not worry and with a moderate green-identity do not behave pro-environmentally. However, we find different relevant predictors comparing the two best models: Gradient Boosting for the subset of worried citizens with Random Forest for the subset of unworried citizens. The predictors that positively influence the pro-environmental behaviour of worried citizens are extreme-green identity and the high level of education. Instead of unworried citizens are age. In this case, pro-environmental behaviours are more widespread in unworried adults than unworried young individuals. Instead, for worried individuals, there is no difference between young individuals and adults in performing pro-environmental behaviour; only senior citizens tend to not perform any actions. Finally, we can confirmed hypothesis 5, having discovered a slight difference in the predictors between worried and unworried citizens.

Chapter 5

# Conclusion

This chapter concludes the dissertation, summarising results, arguing limitations, and proposing further research.

## **5.1 Summary results**

Before concluding, we must go back to all steps taken in the analysis. Like any analysis, let us get started with data cleaning and data exploration. We select and recode the chosen variables concerning sociodemographic information and several opinions on citizens’ climate change. Successively, unsupervised machine learning algorithms are adopted to profile citizens and create some new patterns or classes, which we then insert into the classifiers as explanatory variables. Correlational Class Analysis and Partitioning Around Medoids clustering are fitted through using some climate change questions. The first method partitions citizens according to similar cultural schemas, identifying the different ways to interpret reality. The second algorithm clusters citizens according to similar attitudes, and it creates two different types of self-green identity, which we call *extreme green-identity* and *moderate green identity*. At this point of the analysis, we have all the independent variables to insert the models. To sum up, the variables selected, aside from pro-environmental behaviour, are climate change risk perception, political orientation, marital status, education, gender, age, residence, social class, country, cultural schemas, and self-green identity. Supervised machine learning is useful to classify and predict citizens’ behaviour. We use Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting to improve the model in terms of evaluation metrics and to compare the results of the different important variables. For each classifier best tuning parameters, called hyperparameters, are fitted to get the best results in terms of predictive accuracy and macro-F1. Anyhow data do not satisfy all the Logistic Regression' assumptions, but it is a useful starting point to predict behaviour. The prediction models continue to the other three non-parametric tree models. The advantage of these algorithms is that they are more robust than Logistic Regression. Being more robust, the tree-based models slightly improve the performance in terms of predictive accuracy and macro-F1. Or rather, Decision Tree is the worst model, while Random Forest and Gradient Boosting models have, in general, a better and satisfactory performance.

Now we can briefly summarize the results’ content concerning the pro-environmental behaviour models, which satisfies almost all the research questions described in the methodology (2.1).

In all non-parametric models, climate change risk perception is one of the main important factors in shaping pro-environmental behaviour. We have explained that a higher individual climate change risk perception positively influences and predicts pro-environmental behaviour. High levels of worry encourage environmental action, and in the opposite case, low levels of worry lead to apathy in terms of behaviour.

Another appealing factor that shapes pro-environmental behaviour is extreme green-identity. Individuals who have some powerful green values and who share the importance of fighting climate change are more likely to behave ecologically.

As opposite to green-identity, no cultural schemas created with CCA influence behaviour.

Lastly, sociodemographic variables are examined. In the research questions (2.1), we assume that younger adults, women, liberals, more-educated and wealthy people are more likely to perform pro-environmental behaviour. The results do not confirm all these hypotheses. The only two relevant variables are age and education. However, the relationship between age and pro-environmental behaviour remains a debated factor in the analysis. Not only youngsters act eco-friendly, but in all age groups, there is a high percentage of those who perform some pro-environmental behaviors. Only senior citizens (over 80 years old) tend to perform environmentally unfriendly actions. Concerning education, the relationship is more comprehensible: higher education has a positive effect on pro-environmental behaviour.

The contribution of the dissertation is to find some strategies for promoting citizens’ engagement in pro-environmental behaviour. Public policy should encourage education or information about climate change to increase risk perception and create an extreme-green identity. All these factors lead to an active engagement of European citizens.

However, as the literature review explains, there are some value-action gaps, also in this data. On the one hand, more worried citizens about climate change are more likely to behave ecologically. On the other hand, not all citizens that declare to be worried perform any pro-environmental actions. The opposite case also happens, as citizens who proclaim they are not worried still engage in pro-environmental behaviour. For this reason, we decide to continue to investigate this relationship with other blocks of predictions (the same algorithms are fitted). We divide the dataset into two subsets: one with only the observations of those who have a high-risk perception level; and the other one with only the observations of those who have a low-risk perception level. Once again, Random Forest and Gradient Boosting yield the best performances in both subsets. For the subset of worried citizens, the relevant variables remain the same as the complete model. Extreme green-identity and higher education have a positive effect on pro-environmental behavior. The relationship between age and behaviour remains unclear. Instead, for the subset of unworried citizens, the 30-60 years old age range has a positive effect on pro-environmental behaviour. Therefore, some surprising findings are found: the youngest unworried adults (15-30 years old) have a negative impact, and they are less likely to behave eco-friendly. Another important predictor is moderate green-identity. Besides, in this case, worried citizens with a moderate green-identity have a negative effect on pro-environmental behaviour. Hence summing, the important variables that predict the worried and unworried citizens’ pro-environmental behaviour are slightly different.

Lastly, analysis reveals some new interesting findings. Belonging to several Eastern countries (i.e., Bulgaria, Czech Republic, Romania, and Poland) negatively influences pro-environmental behaviour. These results can be found both in the complete and traditional model (with all observations) and in the subset of unworried citizens.

## **5.2 Limitations**

The first limitation we need to look at is that some variables would be missing to add to the models. In the literate review, we discuss several dimensions that shape pro-environmental behavior: sociodemographic, individual, and contextual factors. In this analysis, only some sociodemographic variables and climate change risk perception, for the individual dimension, could be included in the model. Data available from the Eurobarometer survey are used, but it is not possible to insert all factors explained in the literature review. For example, in the individual dimension, there are present: personal experiences, motivation, environmental knowledge, self-green identity, emotion. Only self-green identity is created using PAM clustering. The cluster groups similar citizens among different attitudes toward climate change. However, the created distinction between extreme green-identity and moderate green-identity may not reflect reality, as there is no confirmation from the citizens themselves. Variables concerning the external dimension are not present in the survey, except for the interviewees' country and the cultural schemes, always realized through CCA. The limit is the same as for PAM clustering. It is possible that the unsupervised algorithm does not find the appropriate shared cultural schemas of citizens. The reason is that, again, the input data is not recognized and is not specified in advance by citizens.

Another limitation is the simple measure of pro-environmental behaviour (dummy outcome: yes or no action). As mentioned above in the literature review, pro-environmental behaviour has multiple dimensions. For example, Stern (2000) classify into four macro-groups: environmental activism, nonactivist behaviours in the public sphere, private-sphere environmentalism, other environmentally significant behaviours. It is different to recycle daily due to some legal and social obligations, from organizing an international march, like Greta Thunberg. Ecological impact and intention are significantly different. However, the study does not diversity the types of behaviours due to the nature of the Eurobarometer’s question. Additionally, as we have explained in the literature review (1.2.3) self-reported behaviour should present some bias and inaccuracy. For example, respondents might be influenced by social desirability, and in this way, pro-environmental behaviour could be overestimated and not entirely precise (Veltri, 2019).

## **5.3 Further studies**

This research wants to be an exploratory study of the self-reported pro-environmental behaviour model. Therefore, many opportunities and many questions are yet to be answered. Future research should continue to analyse and examine factors that shape pro-environmental behaviour. Some open questions are described.

Firstly, it could be interesting to examine the actual behaviour to perform a more accurate analysis, as self-reported behaviour could be imprecise.

Second, some research should be done to diversify the types of pro-environmental behaviour. It might be interesting to understand what factors determine the different behaviour, primarily activism. Promoting citizens' commitment to environmental activism is the key to sensitizing people to the problem. At the same time, it is necessary to guide individuals to take action with a significant impact on the environment.

Additionally, as we stress in the limitations (5.2), for this analysis is not possible to adopt all the dimensions’ variables in the models. Especially contextual factors might be mostly examined and inserted in the analysis. A strategy could be to join different types of data into the Eurobarometer survey and adopt multilevel analysis.

Another open question is the role of each country. Predictions of the behaviour could be divided according to the country to examine particular strategies at a national level to promote pro-environmental behaviour. Understanding, especially, the specific factors that encourage citizens’ engagement in Eastern countries should be fundamental due to the high percentage of citizens who do not perform any pro-environmental behaviours.

# **Appendix**

The complete code is available at the following link GitHub:

LINK

**APPENDIX A. Survey Question Wording and Coding**

|  |  |  |
| --- | --- | --- |
| **Variables** | **Questions** | **Coding** |
|  | ***Question about Climate Change issues*** |  |
| Risk Perception | And how serious a problem do you think climate change is at this moment? Please use a scale from 1 to 10, with '1' meaning it is "not at all a serious problem" and '10' meaning it is "an extremely serious problem" | 1-10 scale: 1= Not at all a serious problem to 10= An extremely serious problem |
| Benefits For Companies | To what extent do you agree or disagree with each of the following statements? Taking action on climate change will lead to innovation that will make EU companies more competitive | 1-4 scale: 1= Totally agree to 4 = Totally disagree |
| Benefits For Citizens | To what extent do you agree or disagree with each of the following statements? Adapting to the adverse impacts of climate change can have positive outcomes for citizens in the EU | 1-4 scale: 1= Totally agree to 4 = Totally disagree |
| Pro-Environmental Behaviour | Have you personally taken any action to fight climate change over the past six months? | 1= Yes; 0= No |
| Renewable Energy | How important do you think it is that the (NATIONALITY) government sets ambitious targets to increase the amount of renewable energy used, such as wind or solar power, by 2030? | 1-4 scale: 1= Very important to 4= Not at all important |
| Energy Efficient | How important do you think it is that the (NATIONALITY) government provides support for improving energy efficiency by 2030 (e.g. by encouraging people to insulate their home or buy electric cars)? | 1-4 scale: 1= Very important to 4= Not at all important |
| Greenhouse Gas | To what extent do you agree or disagree with the following statement: We should reduce greenhouse gas emissions to a minimum while offsetting the remaining emissions, for instance by increasing forested areas, to make the EU economy climate neutral by 2050. | 1-4 scale: 1= Very important to 4= Not at all important |
|  | ***Sociodemographic information*** |  |
| Political Orientation | In political matters people talk of "the left" and "the right". How would you place your views on this scale? | 1-10 scale: 1= left to 10= Right |
| Marital Status | Which of the following best corresponds to your own current situation? | Categorical |
| Stop Education | How old were you when you stopped full-time education? | Number in actual years |
| Gender | Gender | Female; Male |
| Age | How old are you? | Number in actual years |
| Living | Would you say you live in a...? | Categorical |
| Social Class | Do you see yourself and your household belonging to…? | Categorical |
| country | Country | Categorical |

**APPENDIX B. Summary Statistics.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Obs.** | **Mean** | **Std. Dev.** | **Min** | **Max** |
| **Pro-Environmental Behaviour** | 21978 |  |  |  |  |
| *Yes* | *14327* |  |  |  |  |
| *No* | *7651* |  |  |  |  |
| **Risk Perception** | 21978 | 7.93 | 2.02 | 1 | 10 |
| **Benefits For Companies** | 21978 | 1.74 | 0.71 | 1 | 4 |
| **Benefits For Citizens** | 21978 | 1.90 | 0.87 | 1 | 4 |
| **Renewable Energy** | 21978 | 1.52 | 0.65 | 1 | 4 |
| **Energy Efficient** | 21978 | 1.56 | 0.68 | 1 | 4 |
| **Greenhouse** | 21978 | 1.50 | 0.62 | 1 | 4 |
| **Political Orientation** | 21978 |  |  |  |  |
| *Left* | *1853* |  |  |  |  |
| *Centre-Left* | *3856* |  |  |  |  |
| *Centre* | *7968* |  |  |  |  |
| *Centre-Right* | *3470* |  |  |  |  |
| *Right* | *1603* |  |  |  |  |
| *Not Positionable* | *3228* |  |  |  |  |
| **Marital Status** | 21978 |  |  |  |  |
| *Partner* | *7791* |  |  |  |  |
| *Partner And Children* | *7000* |  |  |  |  |
| *Single* | *5975* |  |  |  |  |
| *Single With Children* | *1120* |  |  |  |  |
| *Refusal/Other* | *92* |  |  |  |  |
| **Stop Education** | 21978 |  |  |  |  |
| *Still Studying* | 1405 |  |  |  |  |
| *Up To 15 Years Old* | *2598* |  |  |  |  |
| *16-19 Years Old* | *9358* |  |  |  |  |
| *20+ Years Old* | *8298* |  |  |  |  |
| *Refusal/Dk* | *319* |  |  |  |  |
| **Gender** | 21978 |  |  |  |  |
| *Man* | *10527* |  |  |  |  |
| *Woman* | *11451* |  |  |  |  |
| **Age** | 21978 | 50.51 | 17.88 | 15 | 98 |
| **Living** | 21978 |  |  |  |  |
| *Rural Area Or Village* | *7068* |  |  |  |  |
| *Small Or Middle Sized Town* | *8510* |  |  |  |  |
| *Large Town* | *6396* |  |  |  |  |
| *Dk* | *4* |  |  |  |  |
| **Social Class** | 21978 |  |  |  |  |
| *The Higher Class Of Society* | *154* |  |  |  |  |
| *The Lower Middle Class Of Society* | *3456* |  |  |  |  |
| *The Middle Class Of Society* | *10942* |  |  |  |  |
| *The Upper Middle Class Of Society* | *1630* |  |  |  |  |
| *The Working Class Of Society* | *5276* |  |  |  |  |
| *Refusal/Otherju* | *520* |  |  |  |  |
| ***Green-identity*** | 21978 |  |  |  |  |
| **extreme** | 10807 |  |  |  |  |
| **moderate** | 11171 |  |  |  |  |
| **Cultural schemas** | 21978 |  |  |  |  |
| **1** | 5704 |  |  |  |  |
| **2** | 4415 |  |  |  |  |
| **3** | 6106 |  |  |  |  |
| **4** | 3271 |  |  |  |  |
| **5** | 2482 |  |  |  |  |

**APPENDIX C. Sample Composition**

|  |  |
| --- | --- |
| **Country** | **Obs.** |
| Austria | 830 |
| Belgium | 970 |
| Bulgaria | 626 |
| Croatia | 904 |
| Cyprus | 411 |
| Czech Republic | 729 |
| Denmark | 839 |
| Estonia | 520 |
| Finland | 807 |
| France | 797 |
| Germany | 1200 |
| Greece | 854 |
| Hungary | 900 |
| Ireland | 928 |
| Italy | 905 |
| Latvia | 687 |
| Lithuania | 704 |
| Luxembourg | 399 |
| Malta | 397 |
| Netherlands | 883 |
| Poland | 710 |
| Portugal | 863 |
| Romania | 869 |
| Slovakia | 810 |
| Slovenia | 874 |
| Spain | 820 |
| Sweden | 890 |
| United Kingdom | 852 |
| **Total** | **21978** |

**APPENDIX D. Correlation matrix for each CCA’s group**

Group 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **benefits for companies** | **benefits for citizens** | **renewable energy** | **energy efficient** | **greenhouse gas** |
| **benefits for companies** | 1 | 1 | 1 | 1 | 1 |
| **benefits for citizens** | 1 | 1 | 1 | 1 | 1 |
| **renewable energy** | 1 | 1 | 1 | 1 | 1 |
| **energy efficient** | 1 | 1 | 1 | 1 | 1 |
| **greenhouse gas** | 1 | 1 | 1 | 1 | 1 |

Group2

|  | **benefits for companies** | **benefits for citizens** | **renewable energy** | **energy efficient** | **greenhouse gas** |
| --- | --- | --- | --- | --- | --- |
| **benefits for companies** | 1 | 0.8209258 | 0.34825902 | 0.1660830 | 0.22869773 |
| **benefits for citizens** | 0.8209258 | 1 | 0.48758501 | 0.3021902 | 0.13889897 |
| **renewable energy** | 0.34825902 | 0.48758501 | 1 | 0.3923910 | -0.01232148 |
| **energy efficient** | 0.1660830 | 0.3021902 | 0.3923910 | 1 | -0.14657797 |
| **greenhouse** | 0.22869773 | 0.13889897 | -0.01232148 | -0.14657797 | 1 |

Group 3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **benefits for companies** | **benefits for citizens** | **renewable energy** | **energy efficient** | **greenhouse gas** |
| **benefits for companies** | 1 | 0.07469395 | 0.4400537 | 0.5437509 | 0.5195602 |
| **benefits for citizens** | 0.07469395 | 1 | -0.1859863 | -0.1112473 | -0.1150310 |
| **renewable energy** | 0.44005368 | -0.18598635 | 1 | 0.4965637 | 0.4023176 |
| **energy efficient** | 0.54375089 | -0.11124727 | 0.4965637 | 1 | 0.5111735 |
| **greenhouse gas** | 0.51956019 | -0.11503102 | 0.4023176 | 0.5111735 | 1 |

Group 4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **benefits for companies** | **benefits for citizens** | **renewable energy** | **energy efficient** | **greenhouse gas** |
| **benefits for companies** | 1 | 0.9001736 | -0.2393692 | -0.2281035 | -0.2411191 |
| **benefits for citizens** | 0.9001736 | 1 | -0.1607610 | -0.1477551 | -0.1715444 |
| **renewable energy** | -0.2393692 | -0.1607610 | 1 | 0.8869407 | 0.8548438 |
| **energy efficient** | -0.2281035 | -0.1477551 | 0.8869407 | 1 | 0.8476377 |
| **greenhouse gas** | -0.2411191 | -0.1715444 | 0.8548438 | 0.8476377 | 1 |

Group 5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **benefits for companies** | **benefits for citizens** | **renewable energy** | **energy efficient** | **greenhouse gas** |
| **benefits for companies** | 1 | -0.1514635 | -0.05905457 | -0.03462609 | -0.1298436 |
| **benefits for citizens** | -0.15146353 | 1 | 0.75879238 | 0.73159084 | 0.8669082 |
| **renewable energy** | -0.05905457 | 0.7587924 | 1 | 0.54913637 | 0.6663558 |
| **energy efficient** | -0.03462609 | 0.7315908 | 0.54913637 | 1 | 0.6540521 |
| **greenhouse gas** | -0.12984356 | 0.8669082 | 0.66635581 | 0.65405206 | 1 |

**APPENDIX E. Summary composition of the two subsets**

1. ***Dataset with observations of worried citizens***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Obs.** | **Mean** | **Std. Dev.** | **Min** | **Max** |
| **Pro-Environmental Behaviour** | 19072 |  |  |  |  |
| *Yes* | *12988* |  |  |  |  |
| *No* | *6084* |  |  |  |  |
| **Risk Perception** | 19072 | 8.53 | 1.35 | 6 | 10 |
| **Benefits For Companies** | 19072 | 1.69 | 0.68 | 1 | 4 |
| **Benefits For Citizens** | 19072 | 1.87 | 0.87 | 1 | 4 |
| **Renewable Energy** | 19072 | 1.46 | 0.60 | 1 | 4 |
| **Energy Efficient** | 19072 | 1.50 | 0.63 | 1 | 4 |
| **Greenhouse Gas** | 19072 | 1.45 | 0.58 | 1 | 4 |
| **Political Orientation** | 19072 |  |  |  |  |
| *Left* | *1617* |  |  |  |  |
| *Centre-Left* | *5210* |  |  |  |  |
| *Centre* | *7968* |  |  |  |  |
| *Centre-Right* | *2972* |  |  |  |  |
| *Right* | *1383* |  |  |  |  |
| *Not Positionable* | *2867* |  |  |  |  |
| **Marital Status** | 19072 |  |  |  |  |
| *Partner* | *6704* |  |  |  |  |
| *Partner And Children* | *6204* |  |  |  |  |
| *Single* | *5116* |  |  |  |  |
| *Single With Children* | *967* |  |  |  |  |
| *Refusal/Other* | *81* |  |  |  |  |
| **Stop Education** | 19072 |  |  |  |  |
| *Still Studying* | *1242* |  |  |  |  |
| *Up To 15 Years Old* | *2268* |  |  |  |  |
| *16-19 Years Old* | *7988* |  |  |  |  |
| *20+ Years Old* | *7323* |  |  |  |  |
| *Refusal/Dk* | *251* |  |  |  |  |
| **Gender** | 19072 |  |  |  |  |
| *Man* | *8968* |  |  |  |  |
| *Woman* | *10104* |  |  |  |  |
| **Age** | 19072 | 50.32 | 17.80 | 15 | 98 |
| **Living** | 19072 |  |  |  |  |
| *Rural Area Or Village* | *6075* |  |  |  |  |
| *Small Or Middle Sized Town* | *7422* |  |  |  |  |
| *Large Town* | *5572* |  |  |  |  |
| *Dk* | *3* |  |  |  |  |
| **Social Class** | 19072 |  |  |  |  |
| *The Higher Class Of Society* | *134* |  |  |  |  |
| *The Lower Middle Class Of Society* | *3007* |  |  |  |  |
| *The Middle Class Of Society* | *9520* |  |  |  |  |
| *The Upper Middle Class Of Society* | *1400* |  |  |  |  |
| *The Working Class Of Society* | *4561* |  |  |  |  |
| *Refusal/Other* | *450* |  |  |  |  |
| **Cultural Schemas** | *19072* |  |  |  |  |
| *1* | *5144* |  |  |  |  |
| *2* | *3697* |  |  |  |  |
| *3* | *5296* |  |  |  |  |
| *4* | *2823* |  |  |  |  |
| *5* | *2112* |  |  |  |  |
| **Green-Identity** | 19072 |  |  |  |  |
| *1* | *8982* |  |  |  |  |
| *2* | *10090* |  |  |  |  |

1. ***Dataset with observations of unworried citizens***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Obs.** | **Mean** | **Std. Dev.** | **Min** | **Max** |
| **Pro-Environmental Behaviour** | 2906 |  |  |  |  |
| *Yes* | *1339* |  |  |  |  |
| *No* | *1567* |  |  |  |  |
| **Risk Perception** | 2906 | 4.06 | 1.31 | 1 | 5 |
| **Benefits For Companies** | 2906 | 2.08 | 0.82 | 1 | 4 |
| **Benefits For Citizens** | 2906 | 2.15 | 0.88 | 1 | 4 |
| **Renewable Energy** | 2906 | 1.94 | 0.83 | 1 | 4 |
| **Energy Efficient** | 2906 | 1.95 | 0.83 | 1 | 4 |
| **Greenhouse Gas** | 2906 | 1.84 | 0.78 | 1 | 4 |
| **Political Orientation** | 2906 |  |  |  |  |
| *Left* | *236* |  |  |  |  |
| *Centre-Left* | *692* |  |  |  |  |
| *Centre* | *899* |  |  |  |  |
| *Centre-Right* | *498* |  |  |  |  |
| *Right* | *220* |  |  |  |  |
| *Not Positionable* | *361* |  |  |  |  |
| **Marital Status** | 2906 |  |  |  |  |
| *Partner* | *1087* |  |  |  |  |
| *Partner And Children* | *796* |  |  |  |  |
| *Single* | *859* |  |  |  |  |
| *Single With Children* | *153* |  |  |  |  |
| *Refusal/Other* | *11* |  |  |  |  |
| **Stop Education** | 2906 |  |  |  |  |
| *Still Studying* | *163* |  |  |  |  |
| *Up To 15 Years Old* | *330* |  |  |  |  |
| *16-19 Years Old* | *1370* |  |  |  |  |
| *20+ Years Old* | *975* |  |  |  |  |
| *Refusal/Dk* | *68* |  |  |  |  |
| **Gender** | 2906 |  |  |  |  |
| *Man* | *1559* |  |  |  |  |
| *Woman* | *1347* |  |  |  |  |
| **Age** | 2906 | 51.75 | 18.32 | 15 | 95 |
| **Living** | 2906 |  |  |  |  |
| *Rural Area or Village* | *993* |  |  |  |  |
| *Small or Middle Sized Town* | *1088* |  |  |  |  |
| *Large Town* | *824* |  |  |  |  |
| *Dk* | *1* |  |  |  |  |
| **Social Class** | 2906 |  |  |  |  |
| *The Higher Class of Society* | *20* |  |  |  |  |
| *The Lower Middle Class of Society* | *449* |  |  |  |  |
| *The Middle Class of Society* | *1422* |  |  |  |  |
| *The Upper Middle Class of Society* | *230* |  |  |  |  |
| *The Working Class of Society* | *715* |  |  |  |  |
| *Refusal/Other* | *70* |  |  |  |  |
| **Cultural Schemas** | 2906 |  |  |  |  |
| *1* | *560* |  |  |  |  |
| *2* | *718* |  |  |  |  |
| *3* | *810* |  |  |  |  |
| *4* | *448* |  |  |  |  |
| *5* | *370* |  |  |  |  |
| **Green-Identity** | 2906 |  |  |  |  |
| *1* | *2189* |  |  |  |  |
| *2* | *717* |  |  |  |  |

**APPENDIX F. Grid of Parameters and Hyperparameters**

***Logistic Regression***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameters** | **Parameter Grid** | **Hyperparameters** | | |
| **Complete Dataset** | **Subset of Worried Citizens** | **Subset of Unworried Citizens** |
| 'solver' | 'newton-cg', 'lbfgs', 'liblinear' | 'lbfgs' | 'lbfgs' | 'newton-cg' |
| 'penalty' | ‘l2' | ‘l2' | ‘l2' | ‘l2' |
| 'c' | 100, 10, 1.0, 0.1, 0.01 | 0.1 | 0.1 | 10 |

***Decision Tree***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameters** | **Parameter Grid** | **Hyperparameters** | | |
| **Complete Dataset** | **Subset of Worried Citizens** | **Subset of Unworried Citizens** |
| 'criterion' | 'gini', 'entropy' | 'entropy' | 'gini' | 'entropy' |
| 'max\_depth' | range(1, 10) | 8 | 7 | 5 |
| 'min\_samples\_split' | range(1, 10) | 4 | 6 | 8 |
| 'min\_samples\_leaf' | range(1, 5) | 1 | 3 | 2 |
| ccp\_alpha |  | 0.00099 | 0.00222 | 0.00939 |

***Random Forest***

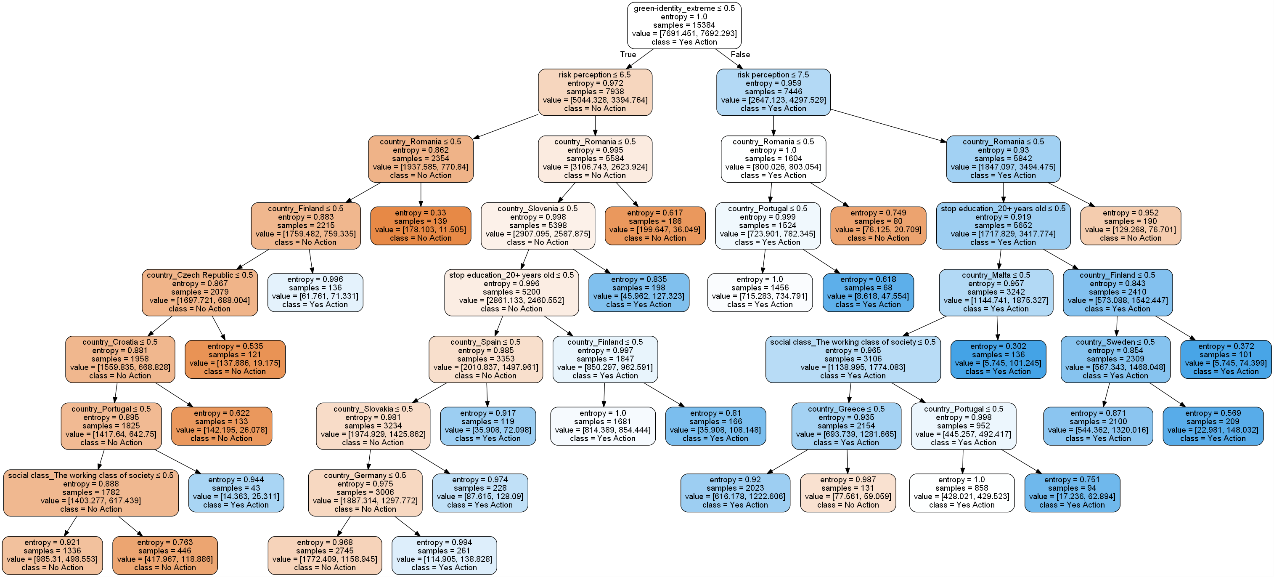
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameters** | **Parameter Grid** | **Hyperparameters** | | |
| **Complete Dataset** | **Subset of Worried Citizens** | **Subset of Unworried Citizens** |
| 'n\_estimators' | 50, 120, 190, 260, 330, 400 | 400 | 400 | 120 |
| 'max\_depth' | 5, 8, 15, 25, 30, None | 25 | 30 | 25 |
| 'min\_samples\_split' | 2, 5, 10, 15, 100 | 5 | 2 | 100 |
| 'min\_samples\_leaf' | 1, 2, 5, 10 | 2 | 1 | 5 |
| 'max\_features' | 'auto', 'sqrt', 'log2' | 'log2' | 'auto' | 'log2' |
| 'bootstrap' | True, False | True | True | False |

***Gradient Boosting***

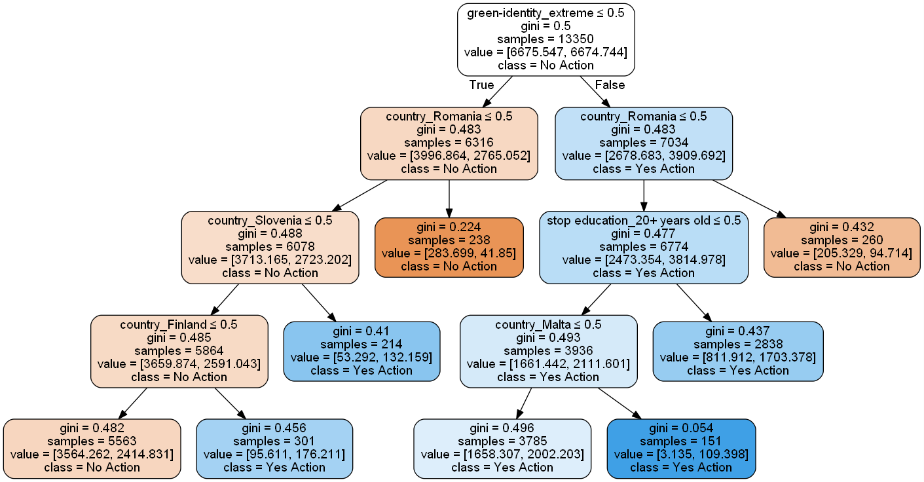
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameters** | **Parameter Grid** | **Hyperparameters** | | |
| **Complete Dataset** | **Subset of Worried Citizens** | **Subset of Unworried Citizens** |
| 'max\_depth' | 3, 4, 5 | 3 | 4 | 120 |
| 'learning\_rate' | 0.1, 0.25, 0.5, 0.75, 1 | 0.1 | 0.25 | 0.1 |
| 'n\_estimators' | 50, 100, 150, 250 | 150 | 150 | 150 |
| 'gamma' | 0.5, 1, 1.5, 2 | 2 | 1 | 1 |
| 'min\_child\_weight' | 1, 5, 10 | 5 | 1 | 1 |
| scale\_pos\_weight | 0.1, 0.25, 0.5, 0.75, 1 | 0.75 | 0.5 | 1 |

**APPENDIX G. Decision Tree Plot**

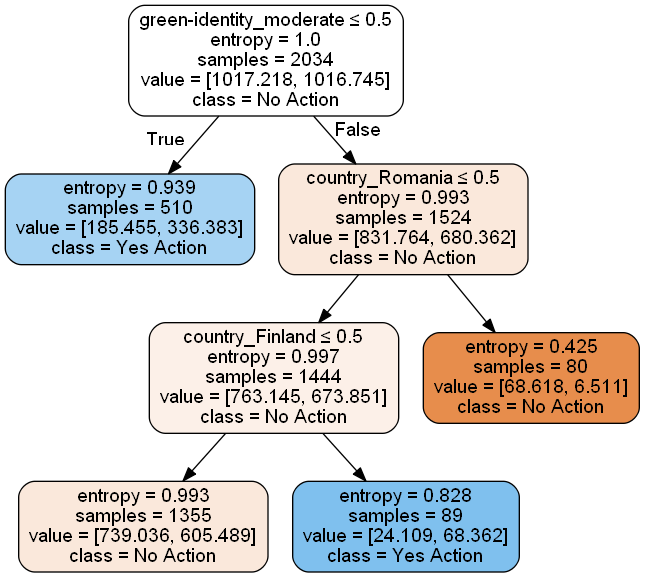
1. ***Complete Dataset***



1. ***Subset of worried citizens***



1. ***Subset of unworried citizens***



**APPENDIX H. Comparison of Feature importance**

1. ***Complete Dataset***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Logistic Regression (log-odds)** | **Logistic Regression (log-ratio)** | **Decision Tree** | **Random Forest** | **Gradient Boosting** |
| age | -0.1231 | 1.00169 | 0.0000 | 0.1183 | 0.0084 |
| country\_Austria | 0.0149 | 1.398890 | 0.0000 | 0.0078 | 0.0154 |
| country\_Belgium | -0.2186 | 0.9486 | 0.0000 | 0.0080 | 0.0061 |
| country\_Bulgaria | 0.0017 | 0.4049 | 0.0000 | 0.0120 | 0.0324 |
| country\_Croatia | -0.2483 | 0.6867 | 0.0103 | 0.0078 | 0.0257 |
| country\_Cyprus | 0.0226 | 0.7368 | 0.0000 | 0.0043 | 0.0088 |
| country\_Czech Republic | -0.0809 | 0.4921 | 0.0133 | 0.0093 | 0.0323 |
| country\_Denmark | -0.3844 | 1.265029 | 0.0000 | 0.0063 | 0.0200 |
| country\_Estonia | -0.6174 | 0.7801 | 0.0000 | 0.0053 | 0.0125 |
| country\_Finland | -0.1115 | 2.4090 | 0.0437 | 0.0127 | 0.0316 |
| country\_France | 0.0819 | 1.01496 | 0.0000 | 0.0070 | 0.0000 |
| country\_Germany | 0.2351 | 1.45909 | 0.0110 | 0.0090 | 0.0126 |
| country\_Greece | -0.9041 | 0.6037 | 0.0144 | 0.0107 | 0.0233 |
| country\_Hungary | 0.1867 | 1.0797 | 0.0000 | 0.0066 | 0.0068 |
| country\_Ireland | 0.1049 | 1.079711 | 0.0000 | 0.0073 | 0.0085 |
| country\_Italy | -0.3055 | 0.6560 | 0.0000 | 0.0081 | 0.0195 |
| country\_Latvia | -0.0862 | 0.5393 | 0.0000 | 0.0089 | 0.0315 |
| country\_Lithuania | -0.7090 | 0.6808 | 0.0000 | 0.0073 | 0.0188 |
| country\_Luxembourg | -0.0527 | 1.76024 | 0.0000 | 0.0048 | 0.0120 |
| country\_Malta | 0.0185 | 2.36044 | 0.0326 | 0.0108 | 0.0240 |
| country\_Netherlands | -0.0859 | 0.9506 | 0.0000 | 0.0060 | 0.0000 |
| country\_Poland | -0.0955 | 0.5024 | 0.0000 | 0.0094 | 0.0301 |
| country\_Portugal | -0.1292 | 2.18902 | 0.0359 | 0.0114 | 0.0262 |
| country\_Romania | 0.1727 | 0.2780 | 0.1384 | 0.0221 | 0.0589 |
| country\_Slovakia | 0.3778 | 1.38627 | 0.0143 | 0.0084 | 0.0108 |
| country\_Slovenia | 0.0060 | 2.11387 | 0.0259 | 0.0156 | 0.0300 |
| country\_Spain | -0.3065 | 1.93655 | 0.0132 | 0.0117 | 0.0281 |
| country\_Sweden | 0.3357 | 1.78669 | 0.0111 | 0.0071 | 0.0178 |
| country\_United Kingdom | -0.5046 | 0.7290 | 0.0000 | 0.0075 | 0.0142 |
| cultural schemas\_1 | -0.1191 | 1.0852 | 0.0000 | 0.0185 | 0.0128 |
| cultural schemas\_2 | 0.6609 | 0.8810 | 0.0000 | 0.0170 | 0.0075 |
| cultural schemas\_3 | -0.1266 | 0.9869 | 0.0000 | 0.0184 | 0.0088 |
| cultural schemas\_4 | 0.8588 | 0.8093 | 0.0000 | 0.0154 | 0.0069 |
| cultural schemas\_5 | -0.3161 | 0.8842 | 0.0000 | 0.0123 | 0.0048 |
| gender\_Man | 0.0767 | 0.7360 | 0.0000 | 0.0173 | 0.0125 |
| gender\_Woman | 0.0077 | 0.9174 | 0.0000 | 0.0174 | 0.0000 |
| green-identity\_extreme | -0.2545 | 1.1105 | 0.3633 | 0.0432 | 0.1104 |
| green-identity\_moderate | -0.3759 | 0.6080 | 0.0000 | 0.0368 | 0.0000 |
| living\_Large town | -0.0588 | 0.9014 | 0.0000 | 0.0186 | 0.0088 |
| living\_Rural area or village | -0.0931 | 0.9111 | 0.0000 | 0.0197 | 0.0074 |
| living\_Small or middle sized town | -0.0507 | 0.8159 | 0.0000 | 0.0202 | 0.0085 |
| living\_dk | -0.0132 | 1.0077 | 0.0000 | 0.0000 | 0.0000 |
| marital status\_Partner | 0.1012 | 0.9196 | 0.0000 | 0.0190 | 0.0108 |
| marital status\_Partner and children | -0.0839 | 1.0747 | 0.0000 | 0.0181 | 0.0064 |
| marital status\_Refusal/Other | -0.3475 | 0.9264 | 0.0000 | 0.0007 | 0.0057 |
| marital status\_Single | 0.5655 | 0.8037 | 0.0000 | 0.0178 | 0.0143 |
| marital status\_Single with children | -0.1241 | 0.9177 | 0.0000 | 0.0065 | 0.0000 |
| political orientation\_Centre | 0.1845 | 0.8945 | 0.0000 | 0.0188 | 0.0064 |
| political orientation\_Centre-left | -0.4976 | 1.0186 | 0.0000 | 0.0185 | 0.0121 |
| political orientation\_Centre-right | 0.7485 | 0.8788 | 0.0000 | 0.0147 | 0.0067 |
| political orientation\_Left | -0.2116 | 1.0060 | 0.0000 | 0.0106 | 0.0095 |
| political orientation\_Not positionable | -0.0900 | 0.9223 | 0.0000 | 0.0143 | 0.0052 |
| political orientation\_Right | -0.1038 | 0.9089 | 0.0000 | 0.0099 | 0.0062 |
| risk perception | 0.3266 | 1.1885 | 0.1796 | 0.1039 | 0.0525 |
| social class\_Refusal/Other | 0.5804 | 0.8833 | 0.0000 | 0.0042 | 0.0000 |
| social class\_The higher class of society | -0.1964 | 1.1065 | 0.0000 | 0.0012 | 0.0000 |
| social class\_The lower middle class of society | -0.0674 | 0.8878 | 0.0000 | 0.0142 | 0.0042 |
| social class\_The middle class of society | -0.2035 | 0.9140 | 0.0000 | 0.0190 | 0.0072 |
| social class\_The upper middle class of society | 0.7835 | 1.2052 | 0.0000 | 0.0093 | 0.0102 |
| social class\_The working class of society | 0.8792 | 0.7064 | 0.0332 | 0.0181 | 0.0241 |
| stop education\_16-19 years old | 0.0721 | 0.9348 | 0.0000 | 0.0162 | 0.0046 |
| stop education\_20+ years old | -1.2800 | 1.2025 | 0.0596 | 0.0204 | 0.0427 |
| stop education\_Refusal/dk | -0.4216 | 0.7753 | 0.0000 | 0.0026 | 0.0095 |
| stop education\_Still studying | -0.0764 | 0.9429 | 0.0000 | 0.0058 | 0.0000 |
| stop education\_Up to 15 years old | -0.6884 | 0.8217 | 0.0000 | 0.0099 | 0.0072 |
| age | -0.1231 | 1.0016 | 0.0000 | 0.1183 | 0.0084 |

1. ***Subset of worried citizens***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Logistic Regression (log-odds)** | **Logistic Regression (log-ratio)** | **Decision Tree** | **Random Forest** | **Gradient Boosting** |
| age | -0.0079 | 1.0020 | 0.0000 | 0.1706 | 0.0093 |
| country\_Austria | -0.6876 | 1.5271 | 0.0000 | 0.0076 | 0.0121 |
| country\_Belgium | -0.1913 | 0.8960 | 0.0000 | 0.0102 | 0.0109 |
| country\_Bulgaria | 0.1316 | 0.3990 | 0.0000 | 0.0101 | 0.0407 |
| country\_Croatia | -0.4443 | 0.6867 | 0.0000 | 0.0090 | 0.0337 |
| country\_Cyprus | -0.0810 | 0.7224 | 0.0000 | 0.0057 | 0.0066 |
| country\_Czech Republic | -0.1335 | 0.5028 | 0.0000 | 0.0082 | 0.0283 |
| country\_Denmark | 0.6082 | 1.1386 | 0.0000 | 0.0068 | 0.0132 |
| country\_Estonia | -0.3758 | 0.6413 | 0.0000 | 0.0061 | 0.0074 |
| country\_Finland | -0.1419 | 2.4496 | 0.0656 | 0.0090 | 0.0291 |
| country\_France | -0.0751 | 1.09409 | 0.0000 | 0.0094 | 0.0122 |
| country\_Germany | -0.4245 | 1.3536 | 0.0000 | 0.0099 | 0.0173 |
| country\_Greece | 0.1299 | 0.6800 | 0.0000 | 0.0106 | 0.0212 |
| country\_Hungary | -0.0081 | 1.0968 | 0.0000 | 0.0093 | 0.0092 |
| country\_Ireland | 0.1873 | 0.9222 | 0.0000 | 0.0095 | 0.0085 |
| country\_Italy | -0.3252 | 0.6541 | 0.0000 | 0.0086 | 0.0311 |
| country\_Latvia | -0.9188 | 0.5525 | 0.0000 | 0.0073 | 0.0272 |
| country\_Lithuania | 0.8960 | 0.6970 | 0.0000 | 0.0083 | 0.0196 |
| country\_Luxembourg | -0.1098 | 1.8282 | 0.0000 | 0.0055 | 0.0188 |
| country\_Malta | -0.0105 | 3.3218 | 0.0833 | 0.0079 | 0.0304 |
| country\_Netherlands | 0.0500 | 0.8910 | 0.0000 | 0.0077 | 0.0073 |
| country\_Poland | 0.1156 | 0.5142 | 0.0000 | 0.0093 | 0.0377 |
| country\_Portugal | 0.1998 | 2.3915 | 0.0000 | 0.0094 | 0.0208 |
| country\_Romania | 0.0020 | 0.2940 | 0.2159 | 0.0170 | 0.0651 |
| country\_Slovakia | 0.0925 | 1.4054 | 0.0000 | 0.0082 | 0.0114 |
| country\_Slovenia | -0.0491 | 2.2989 | 0.0676 | 0.0103 | 0.0307 |
| country\_Spain | -0.1193 | 1.9625 | 0.0000 | 0.0095 | 0.0271 |
| country\_Sweden | 0.4234 | 1.8370 | 0.0000 | 0.0072 | 0.0158 |
| country\_United Kingdom | 0.0899 | 0.6993 | 0.0000 | 0.0085 | 0.0158 |
| cultural schemas\_1 | 0.8719 | 1.1608 | 0.0000 | 0.0213 | 0.0174 |
| cultural schemas\_2 | 0.6742 | 0.8778 | 0.0000 | 0.0204 | 0.0087 |
| cultural schemas\_3 | 0.0526 | 1.1016 | 0.0000 | 0.0237 | 0.0081 |
| cultural schemas\_4 | -0.0113 | 0.8947 | 0.0000 | 0.0189 | 0.0084 |
| cultural schemas\_5 | -0.1113 | 0.9921 | 0.0000 | 0.0163 | 0.0064 |
| gender\_Man | -0.0698 | 0.8876 | 0.0000 | 0.0220 | 0.0117 |
| gender\_Woman | 0.2187 | 1.1225 | 0.0000 | 0.0218 | 0.0000 |
| green-identity\_extreme | -0.3856 | 1.4156 | 0.4796 | 0.0211 | 0.0729 |
| green-identity\_moderate | 0.3028 | 0.7038 | 0.0000 | 0.0217 | 0.0000 |
| living\_Large town | -0.0166 | 1.0539 | 0.0000 | 0.0240 | 0.0078 |
| living\_Rural area or village | 0.0968 | 0.9919 | 0.0000 | 0.0242 | 0.0085 |
| living\_Small or middle sized town | -0.0570 | 0.9276 | 0.0000 | 0.0258 | 0.0067 |
| living\_dk | 0.3476 | 1.0273 | 0.0000 | 0.0000 | 0.0000 |
| marital status\_Partner | -0.3513 | 0.9835 | 0.0000 | 0.0249 | 0.0051 |
| marital status\_Partner and children | -0.3577 | 1.1341 | 0.0000 | 0.0235 | 0.0085 |
| marital status\_Refusal/Other | 0.2806 | 0.9333 | 0.0000 | 0.0011 | 0.0000 |
| marital status\_Single | -0.0842 | 0.9215 | 0.0000 | 0.0229 | 0.0102 |
| marital status\_Single with children | 0.0379 | 1.0386 | 0.0000 | 0.0096 | 0.0097 |
| political orientation\_Centre | -0.1304 | 0.9895 | 0.0000 | 0.0247 | 0.0080 |
| political orientation\_Centre-left | 0.8325 | 1.1406 | 0.0000 | 0.0223 | 0.0139 |
| political orientation\_Centre-right | -1.2241 | 0.8677 | 0.0000 | 0.0193 | 0.0104 |
| political orientation\_Left | 1.2005 | 1.2211 | 0.0000 | 0.0139 | 0.0086 |
| political orientation\_Not positionable | -0.0817 | 0.9520 | 0.0000 | 0.0180 | 0.0093 |
| political orientation\_Right | 0.3404 | 0.8751 | 0.0000 | 0.0127 | 0.0077 |
| social class\_Refusal/Other | 0.0270 | 0.9193 | 0.0000 | 0.0058 | 0.0080 |
| social class\_The higher class of society | -0.6652 | 1.2059 | 0.0000 | 0.0020 | 0.0000 |
| social class\_The lower middle class of society | -0.2798 | 0.9446 | 0.0000 | 0.0178 | 0.0092 |
| social class\_The middle class of society | 0.1259 | 0.9508 | 0.0000 | 0.0224 | 0.0077 |
| social class\_The upper middle class of society | -0.5933 | 1.3238 | 0.0000 | 0.0098 | 0.0142 |
| social class\_The working class of society | -0.3609 | 0.7559 | 0.0000 | 0.0177 | 0.0193 |
| stop education\_16-19 years old | -0.0505 | 0.9326 | 0.0000 | 0.0163 | 0.0085 |
| stop education\_20+ years old | 0.6034 | 1.2445 | 0.0880 | 0.0154 | 0.0472 |
| stop education\_Refusal/dk | -0.0691 | 1.0512 | 0.0000 | 0.0036 | 0.0061 |
| stop education\_Still studying | 0.1491 | 0.9888 | 0.0000 | 0.0067 | 0.0065 |
| stop education\_Up to 15 years old | -0.1154 | 0.8259 | 0.0000 | 0.0111 | 0.0069 |

1. ***Subset of unworried citizens***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **Logistic Regression (log-odds)** | **Logistic Regression (log-ratio)** | **Decision Tree** | **Random Forest** | **Gradient Boosting** |
| age | -0.1729 | 1.0100 | 0.0000 | 0.1031 | 0.0184 |
| country\_Austria | -1.1101 | 0.9790 | 0.0000 | 0.0205 | 0.0000 |
| country\_Belgium | -0.1443 | 1.0235 | 0.0000 | 0.0044 | 0.0000 |
| country\_Bulgaria | -0.0042 | 0.2729 | 0.0000 | 0.0102 | 0.0213 |
| country\_Croatia | -0.5525 | 0.7319 | 0.0000 | 0.0144 | 0.0172 |
| country\_Cyprus | 0.1226 | 1.1628 | 0.0000 | 0.0021 | 0.0218 |
| country\_Czech Republic | -0.1649 | 0.3295 | 0.0000 | 0.0410 | 0.0341 |
| country\_Denmark | 0.4695 | 1.6501 | 0.0000 | 0.0185 | 0.0205 |
| country\_Estonia | -0.3121 | 0.5755 | 0.0000 | 0.0072 | 0.0136 |
| country\_Finland | -0.1089 | 2.7327 | 0.2137 | 0.0406 | 0.0280 |
| country\_France | 0.1673 | 0.6360 | 0.0000 | 0.0132 | 0.0127 |
| country\_Germany | -0.5053 | 1.4552 | 0.0000 | 0.0100 | 0.0190 |
| country\_Greece | 0.5009 | 0.2480 | 0.0000 | 0.0058 | 0.0262 |
| country\_Hungary | 0.4085 | 0.8101 | 0.0000 | 0.0097 | 0.0000 |
| country\_Ireland | 0.6703 | 11304884111431700.0000 | 0.0000 | 0.0023 | 0.0102 |
| country\_Italy | 0.1509 | 0.6033 | 0.0000 | 0.0140 | 0.0162 |
| country\_Latvia | -1.2988 | 0.4097 | 0.0000 | 0.0290 | 0.0292 |
| country\_Lithuania | 1.0052 | 0.7740 | 0.0000 | 0.0114 | 0.0308 |
| country\_Luxembourg | 0.0232 | 2.0856 | 0.0000 | 0.0061 | 0.0142 |
| country\_Malta | 0.1208 | 7.2485 | 0.0000 | 6.7475 | 0.0000 |
| country\_Netherlands | -0.6253 | 0.7911 | 0.0000 | 0.0058 | 0.0129 |
| country\_Poland | 0.0963 | 0.6773 | 0.0000 | 0.0052 | 0.0000 |
| country\_Portugal | 0.0528 | 6.5786 | 0.0000 | 0.0329 | 0.0266 |
| country\_Romania | 0.0100 | 0.1778 | 0.3601 | 0.1122 | 0.0591 |
| country\_Slovakia | -0.2106 | 1.7713 | 0.0000 | 0.0072 | 0.0145 |
| country\_Slovenia | 0.1041 | 2.4184 | 0.0000 | 0.0431 | 0.0365 |
| country\_Spain | -0.0965 | 3.3741 | 0.0000 | 0.0092 | 0.0105 |
| country\_Sweden | -0.0212 | 1.59915 | 0.0000 | 0.0060 | 0.0102 |
| country\_United Kingdom | -0.4525 | 0.5705 | 0.0000 | 0.0105 | 0.0225 |
| cultural schemas\_1 | 1.8838 | 1.0064 | 0.0000 | 0.0054 | 0.0077 |
| cultural schemas\_2 | 1.2161 | 1.1292 | 0.0000 | 0.0081 | 0.0159 |
| cultural schemas\_3 | 0.4824 | 1.0178 | 0.0000 | 0.0046 | 0.0063 |
| cultural schemas\_4 | 0.3875 | 1.0273 | 0.0000 | 0.0056 | 0.0121 |
| cultural schemas\_5 | 0.0270 | 0.8412 | 0.0000 | 0.0090 | 0.0142 |
| gender\_Man | 0.1625 | 0.9080 | 0.0000 | 0.0057 | 0.0151 |
| gender\_Woman | 0.2193 | 1.1010 | 0.0000 | 0.0072 | 0.0000 |
| green-identity\_extreme | -1.3942 | 1.4898 | 0.0000 | 0.0673 | 0.0579 |
| green-identity\_moderate | 0.3752 | 0.6710 | 0.4263 | 0.0962 | 0.0000 |
| living\_Large town | -0.0023 | 1.6200 | 0.0000 | 0.0082 | 0.0101 |
| living\_Rural area or village | 0.0177 | 1.5045 | 0.0000 | 0.0049 | 0.0120 |
| living\_Small or middle sized town | -0.0151 | 1.1821 | 0.0000 | 0.0087 | 0.0119 |
| living\_dk | 0.3987 | 0.3470 | 0.0000 | 0.0000 | 0.0000 |
| marital status\_Partner | -0.3990 | 0.9977 | 0.0000 | 0.0126 | 0.0115 |
| marital status\_Partner and children | -0.5613 | 1.0800 | 0.0000 | 0.0061 | 0.0274 |
| marital status\_Refusal/Other | -0.1437 | 1.2842 | 0.0000 | 0.0000 | 0.0000 |
| marital status\_Single | -0.0491 | 0.6665 | 0.0000 | 0.0159 | 0.0277 |
| marital status\_Single with children | 0.0805 | 1.0837 | 0.0000 | 0.0043 | 0.0111 |
| political orientation\_Centre | 0.1215 | 1.1284 | 0.0000 | 0.0054 | 0.0125 |
| political orientation\_Centre-left | 0.8831 | 0.9958 | 0.0000 | 0.0044 | 0.0110 |
| political orientation\_Centre-right | -1.7273 | 0.8969 | 0.0000 | 0.0045 | 0.0196 |
| political orientation\_Left | 1.9808 | 1.0542 | 0.0000 | 0.0057 | 0.0000 |
| political orientation\_Not positionable | -0.4057 | 1.1097 | 0.0000 | 0.0055 | 0.0228 |
| political orientation\_Right | 0.5718 | 0.8480 | 0.0000 | 0.0041 | 0.0346 |
| social class\_Refusal/Other | -1.0585 | 0.9521 | 0.0000 | 0.0032 | 0.0169 |
| social class\_The higher class of society | -0.3897 | 1.9548 | 0.0000 | 0.0028 | 0.0058 |
| social class\_The lower middle class of society | -0.4292 | 0.9851 | 0.0000 | 0.0050 | 0.0096 |
| social class\_The middle class of society | 0.0770 | 0.9671 | 0.0000 | 0.0061 | 0.0123 |
| social class\_The upper middle class of society | -0.8923 | 0.8661 | 0.0000 | 0.0049 | 0.0149 |
| social class\_The working class of society | -0.2562 | 0.6510 | 0.0000 | 0.0192 | 0.0228 |
| stop education\_16-19 years old | -0.0335 | 1.1764 | 0.0000 | 0.0086 | 0.0096 |
| stop education\_20+ years old | 0.7351 | 1.2451 | 0.0000 | 0.0324 | 0.0245 |
| stop education\_Refusal/dk | 0.2502 | 0.5351 | 0.0000 | 0.0088 | 0.0100 |
| stop education\_Still studying | 0.0064 | 1.4733 | 0.0000 | 0.0091 | 0.0058 |
| stop education\_Up to 15 years old | -0.2343 | 0.8656 | 0.0000 | 0.0048 | 0.0000 |
| age | -0.1729 | 1.0100 | 0.0000 | 0.1031 | 0.0184 |

# **Bibliography**

Abu-Omar, K., & Rütten, A. (2008). Relation of leisure time, occupational, domestic, and commuting physical activity to health indicators in Europe. *Preventive Medicine*, *47*(3), 319–323. https://doi.org/10.1016/j.ypmed.2008.03.012

Akerlof, K., Maibach, E. W., Fitzgerald, D., Cedeno, A. Y., & Neuman, A. (2013). Do people “personally experience” global warming, and if so how, and does it matter? *Global Environmental Change*, *23*(1), 81–91. https://doi.org/10.1016/j.gloenvcha.2012.07.006

Battiti, R., & Brunato, M. (2014). *The LION way: Machine learning plus intelligent optimization* (1. ed). CreateSpace.

Belgiu, M. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 8.

Benner, J. (2020, agosto 6). Cross-Validation and Hyperparameter Tuning: How to Optimise your Machine Learning Model. *Towards Data Science*. https://towardsdatascience.com/cross-validation-and-hyperparameter-tuning-how-to-optimise-your-machine-learning-model-13f005af9d7d

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bonera, M., Corvi, E., Codini, A., & Ma, R. (2017). Does Nationality Matter in Eco-Behaviour? *Sustainability*, *9*(10), 1694. https://doi.org/10.3390/su9101694

Boutyline, A. (2017). Improving the Measurement of Shared Cultural Schemas with Correlational Class Analysis: Theory and Method. *Sociological Science*, *4*, 353–393. https://doi.org/10.15195/v4.a15

Bradley, G. L., Babutsidze, Z., Chai, A., & Reser, J. P. (2020). The role of climate change risk perception, response efficacy, and psychological adaptation in pro-environmental behavior: A two nation study. *Journal of Environmental Psychology*, *68*, 1–12. https://doi.org/10.1016/j.jenvp.2020.101410

Brulle, R. J., Carmichael, J., & Jenkins, J. C. (2012). Shifting public opinion on climate change: An empirical assessment of factors influencing concern over climate change in the U.S., 2002–2010. *Climatic Change*, *114*(2), 169–188. https://doi.org/10.1007/s10584-012-0403-y

Bubeck, P., Wouter Botzen, W. J., Laudan, J., Aerts, J. C. J. H., & Thieken, A. H. (2018). Insights into Flood-Coping Appraisals of Protection Motivation Theory: Empirical Evidence from Germany and France: Insights into Flood-Coping Appraisals of Protection Motivation Theory. *Risk Analysis*, *38*(6), 1239–1257. https://doi.org/10.1111/risa.12938

Burck, J. (2018). *CCPI, climate change performance index background and methodology*.

Chao, Y.-L., & Lam, S.-P. (2011). Measuring Responsible Environmental Behavior: Self-Reported and Other-Reported Measures and Their Differences in Testing a Behavioral Model. *Environment and Behavior*, *43*(1), 53–71. https://doi.org/10.1177/0013916509350849

Cho, Y.-N., Thyroff, A., Rapert, M. I., Park, S.-Y., & Lee, H. J. (2013). To be or not to be green: Exploring individualism and collectivism as antecedents of environmental behavior. *Journal of Business Research*, *66*(8), 1052–1059. https://doi.org/10.1016/j.jbusres.2012.08.020

Davidson, D. J., & Haan, M. (2012). Gender, political ideology, and climate change beliefs in an extractive industry community. *Population and Environment*, *34*(2), 217–234. https://doi.org/10.1007/s11111-011-0156-y

Dono, J., Webb, J., & Richardson, B. (2010). The relationship between environmental activism, pro-environmental behaviour and social identity. *Journal of Environmental Psychology*, *30*(2), 178–186. https://doi.org/10.1016/j.jenvp.2009.11.006

Driscoll, D. (2019). Assessing Sociodemographic Predictors of Climate Change Concern, 1994–2016. *Social Science Quarterly*, 1699–1708. https://doi.org/10.1111/ssqu.12683

Dunlap, R. E. (2014, aprile 22). Global Warming or Climate Change: Is There a Difference? *Gallup*. https://news.gallup.com/poll/168617/global-warming-climate-change-difference.aspx

Echavarren, J. M., Balžekienė, A., & Telešienė, A. (2019). Multilevel analysis of climate change risk perception in Europe: Natural hazards, political contexts and mediating individual effects. *Safety Science*, *120*, 813–823. https://doi.org/10.1016/j.ssci.2019.08.024

Egan, P. J., & Mullin, M. (2017). Climate Change: US Public Opinion. *Annual Review of Political Science*, *20*(1), 209–227. https://doi.org/10.1146/annurev-polisci-051215-022857

European Commission, Brussels. (2019). *Eurobarometer 91.3 (2019)Eurobarometer 91.3 (2019): Rule of Law, and Climate Change: Rule of Law, and Climate Change* (1.0.0) [Data set]. GESIS Data Archive. https://doi.org/10.4232/1.13372

Fielding, K. S., Head, B. W., Laffan, W., Western, M., & Hoegh-Guldberg, O. (2012). Australian politicians’ beliefs about climate change: Political partisanship and political ideology. *Environmental Politics*, *21*(5), 712–733. https://doi.org/10.1080/09644016.2012.698887

Finucane, M. L., Slovic, P., Mertz, C. K., Flynn, J., & Satterfield, T. A. (2000). Gender, race, and perceived risk: The «white male» effect. *Health, Risk & Society*, *2*(2), 159–172. https://doi.org/10.1080/713670162

Fonseca, J. R. S. (2013). Clustering in the field of social sciences: That is your choice. *International Journal of Social Research Methodology*, *16*(5), 403–428. https://doi.org/10.1080/13645579.2012.716973

Gatersleben, B., Murtagh, N., & Abrahamse, W. (2014). Values, identity and pro-environmental behaviour. *Contemporary Social Science*, *9*(4), 374–392. https://doi.org/10.1080/21582041.2012.682086

Gatersleben, B., Steg, L., & Vlek, C. (2002). Measurement and Determinants of Environmentally Significant Consumer Behavior. *Environment and Behavior*, *34*(3), 335–362. https://doi.org/10.1177/0013916502034003004

Gilg, A., Barr, S., & Ford, N. (2005). Green consumption or sustainable lifestyles? Identifying the sustainable consumer. *Futures*, *37*(6), 481–504. https://doi.org/10.1016/j.futures.2004.10.016

Goldberg, A. (2011). Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined. *American Journal of Sociology*, *116*(5), 1397–1436. https://doi.org/10.1086/657976

Goldsmith, R. E., Feygina, I., & Jost, J. T. (2013). The Gender Gap in Environmental Attitudes: A System Justification Perspective. In M. Alston & K. Whittenbury (A c. Di), *Research, Action and Policy: Addressing the Gendered Impacts of Climate Change* (pagg. 159–171). Springer Netherlands. https://doi.org/10.1007/978-94-007-5518-5\_12

Hidalgo, M. C., & Pisano, I. (2010). Determinants of risk perception and willingness to tackle climate change. A pilot study. *Psyecology*, *1*(1), 105–112. https://doi.org/10.1174/217119710790709595

Ilieș, A., & Zahid, R. M. A. (2019). The Impact of Europe’s Individualism/Collectivism on the International Trade. *European Journal of Marketing and Economics*, *2*(1), 6–20. https://doi.org/10.26417/ejme-2019.v2i1-59

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R* (Vol. 103). Spinger.

Jenhani, I., Amor, N. B., & Elouedi, Z. (2008). Decision trees as possibilistic classifiers. *International Journal of Approximate Reasoning*, *48*(3), 784–807. https://doi.org/10.1016/j.ijar.2007.12.002

Kaiser, F. G., Doka, G., Hofstetter, P., & Ranney, M. A. (2003). Ecological behavior and its environmental consequences: A life cycle assessment of a self-report measure. *Journal of Environmental Psychology*, *23*(1), 11–20. https://doi.org/10.1016/S0272-4944(02)00075-0

Keshavarz, M., & Karami, E. (2016). Farmers’ pro-environmental behavior under drought: Application of protection motivation theory. *Journal of Arid Environments*, *127*, 128–136. https://doi.org/10.1016/j.jaridenv.2015.11.010

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

Krajhanzl, J. (2010). *Environmental and Pro-environmental Behavior*. *School and Health*, 251–274.

Lacasse, K. (2015). The Importance of Being Green: The Influence of Green Behaviors on Americans’ Political Attitudes Toward Climate Change. *Environment and Behavior*, *47*(7), 754–781. https://doi.org/10.1177/0013916513520491

Lacroix, K., & Gifford, R. (2018). Psychological Barriers to Energy Conservation Behavior: The Role of Worldviews and Climate Change Risk Perception. *Environment and Behavior*, *50*(7), 749–780. https://doi.org/10.1177/0013916517715296

Larson, L. R., Whiting, J. W., & Green, G. T. (2011). Exploring the influence of outdoor recreation participation on pro-environmental behaviour in a demographically diverse population. *Local Environment*, *16*(1), 67–86. https://doi.org/10.1080/13549839.2010.548373

Lee, T. M., Markowitz, E. M., Howe, P. D., Ko, C.-Y., & Leiserowitz, A. A. (2015). Predictors of public climate change awareness and risk perception around the world. *Nature Climate Change*, *5*(11), 1014–1020. https://doi.org/10.1038/nclimate2728

Lengyel, A., & Botta‐Dukát, Z. (2019). Silhouette width using generalized mean—A flexible method for assessing clustering efficiency. *Ecology and Evolution*, *9*(23), 13231–13243. https://doi.org/10.1002/ece3.5774

Liu, X., Vedlitz, A., & Shi, L. (2014). Examining the determinants of public environmental concern: Evidence from national public surveys. *Environmental Science & Policy*, *39*, 77–94. https://doi.org/10.1016/j.envsci.2014.02.006

Lorenzoni, I., & Pidgeon, N. F. (2006). Public Views on Climate Change: European and USA Perspectives. *Climatic Change*, *77*(1–2), 73–95. https://doi.org/10.1007/s10584-006-9072-z

Loyen, A. (2016). European Sitting Championship: Prevalence and Correlates of Self-Reported Sitting Time in the 28 European Union Member States. *PLOS ONE*, 17.

Markle, G. (2019). Understanding Pro-Environmental Behavior in the US: Insights from Grid-Group Cultural Theory and Cognitive Sociology. *Sustainability*, *11*(2), 532. https://doi.org/10.3390/su11020532

McCright, A. M. (2011). Political orientation moderates Americans’ beliefs and concern about climate change: An editorial comment. *Climatic Change*, *104*(2), 243–253. https://doi.org/10.1007/s10584-010-9946-y

McCright, A. M., Dunlap, R. E., & Marquart-Pyatt, S. T. (2016). Political ideology and views about climate change in the European Union. *Environmental Politics*, *25*(2), 338–358. https://doi.org/10.1080/09644016.2015.1090371

Meyer, A. (2015). Does education increase pro-environmental behavior? Evidence from Europe. *Ecological Economics*, *116*, 108–121. https://doi.org/10.1016/j.ecolecon.2015.04.018

Nagy, S., & Konyha Molnárné, C. (2018). The Effects of Hofstede’s Cultural Dimensions on Pro-Environmental Behaviour: How Culture Influences Environmentally Conscious Behaviour. *Theory, Methodology, Practice*, *14*(1), 27–36. https://doi.org/10.18096/TMP.2018.01.03

O’Connor, R. E., Bard, R. J., & Fisher, A. (1999). Risk Perceptions, General Environmental Beliefs, and Willingness to Address Climate Change. *Risk Analysis*, *19*(3), 461–471. https://doi.org/10.1111/j.1539-6924.1999.tb00421.x

Oreg, S., & Katz-Gerro, T. (2006). Predicting Proenvironmental Behavior Cross-Nationally: Values, the Theory of Planned Behavior, and Value-Belief-Norm Theory. *Environment and Behavior*, *38*(4), 462–483. https://doi.org/10.1177/0013916505286012

Peng, C.-Y. J., So, T.-S. H., Stage, F. K., & John, E. P. S. (2002). *THE USE AND INTERPRETATION OF LOGISTIC REGRESSION IN HIGHER EDUCATION JOURNALS: 1988–1999*. 35.

*Protecting health from climate change: World Health Day 2008.* (2008). World Health Organization.

Rossoni, L., Gonçalves, C. P., da Silva, M. P., & Gonçalves, A. F. (2020). *Mapping Organizational Culture Schemes Based on Correlational Class Analysis* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/sf2v4

Rossoni, L., Gonçalves, C. P., Silva, M. P. da, & Gonçalves, A. F. (2021). Mapping Organizational Culture Schemas Based on Correlational Class Analysis: A Tutorial. *Revista de Administração Contemporânea*, *25*(1), e200096. https://doi.org/10.1590/1982-7849rac2021200096

Shafiei, A., & Maleksaeidi, H. (2020). Pro-environmental behavior of university students: Application of protection motivation theory. *Global Ecology and Conservation*, *22*, e00908. https://doi.org/10.1016/j.gecco.2020.e00908

Shendre, S. (2020, aprile 29). Clustering datasets having both numerical and categorical variables. *Towards Data Science*. https://towardsdatascience. com/clustering - datasets - having- both- numerical- and- categorical - variables-ed91cdca0677

Shi, J., Visschers, V. H. M., & Siegrist, M. (2015). Public Perception of Climate Change: The Importance of Knowledge and Cultural Worldviews: The Importance of Knowledge and Cultural Worldviews in Climate Change Perception. *Risk Analysis*, *35*(12), 2183–2201. https://doi.org/10.1111/risa.12406

Shmueli, B. (2019, luglio 3). Multi-Class Metrics Made Simple, Part II: the F1-score. *Towards Data Science*. https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1

Shwom, R. L., McCright, A. M., Brechin, S. R., Dunlap, R. E., Marquart-Pyatt, S. T., & Hamilton, L. C. (2015). Public Opinion on Climate Change. In R. E. Dunlap & R. J. Brulle (A c. Di), *Climate Change and Society* (pagg. 269–299). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199356102.003.0009

Simga-Mugan, C., Daly, B. A., Onkal, D., & Kavut, L. (2005). The Influence of Nationality and Gender on Ethical Sensitivity: An Application of the Issue-Contingent Model. *Journal of Business Ethics*, *57*(2), 139–159. https://doi.org/10.1007/s10551-004-4601-z

Slovic, P. (1987). Perception of risk. *Science*, *236*(4799), 280–285. https://doi.org/10.1126/science.3563507

Slovic, P., & Peters, E. (2006). Risk Perception and Affect. *Current Directions in Psychological Science*, *15*(6), 322–325. https://doi.org/10.1111/j.1467-8721.2006.00461.x

Slovic, P., & Weber, E. U. (2002, aprile). *Perception of risk posed by extreme events*. Paper presented at Risk Management Strategies in an Uncertain World Conference, Palisades, NY.

Steg, L., & Sievers, I. (2000, marzo). Cultural theory and individual perceptions of environmental risks. *Environment and Behavior*, *32*(2), 250–269.

Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology*, *29*(3), 309–317. https://doi.org/10.1016/j.jenvp.2008.10.004

Stern, P. C. (2000). New Environmental Theories: Toward a Coherent Theory of Environmentally Significant Behavior. *Journal of Social Issues*, *56*(3), 407–424. https://doi.org/10.1111/0022-4537.00175

Stoltzfus, J. C. (2011). Logistic Regression: A Brief Primer: LOGISTIC REGRESSION: A BRIEF PRIMER. *Academic Emergency Medicine*, *18*(10), 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Sun, Y., & Han, Z. (2018). Climate Change Risk Perception in Taiwan: Correlation with Individual and Societal Factors. *International Journal of Environmental Research and Public Health*, *15*(1), 1–12. https://doi.org/10.3390/ijerph15010091

Taylor, A. L., Dessai, S., & Bruine de Bruin, W. (2014). Public perception of climate risk and adaptation in the UK: A review of the literature. *Climate Risk Management*, *4–5*, 1–16. https://doi.org/10.1016/j.crm.2014.09.001

Torgler, B., & García-Valiñas, M. A. (2007). The determinants of individuals’ attitudes towards preventing environmental damage. *Ecological Economics*, *63*(2–3), 536–552. https://doi.org/10.1016/j.ecolecon.2006.12.013

Vainio, A., & Paloniemi, R. (2013). Does belief matter in climate change action? *Public Understanding of Science*, *22*(4), 382–395. https://doi.org/10.1177/0963662511410268

van der Linden, S. (2015). The social-psychological determinants of climate change risk perceptions: Towards a comprehensive model. *Journal of Environmental Psychology*, *41*, 112–124. https://doi.org/10.1016/j.jenvp.2014.11.012

Veltri, G. A. (2019). *Digital social research*. Polity Press.

Vicente-Molina, M. A., Fernández-Sainz, A., & Izagirre-Olaizola, J. (2018). Does gender make a difference in pro-environmental behavior? The case of the Basque Country University students. *Journal of Cleaner Production*, *176*, 89–98. https://doi.org/10.1016/j.jclepro.2017.12.079

Weber, E. U. (2016). What shapes perceptions of climate change? New research since 2010: What shapes perceptions of climate change? *Wiley Interdisciplinary Reviews: Climate Change*, *7*(1), 125–134. https://doi.org/10.1002/wcc.377

Whitmarsh, L., & O’Neill, S. (2010). Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *Journal of Environmental Psychology*, *30*(3), 305–314. https://doi.org/10.1016/j.jenvp.2010.01.003

Wildavsky, A., & Dake, K. (1990). Theories of Risk Perception: Who Fears What and Why? *Daedalus*, *119*(4), 41–60.

Xie, B., Brewer, M. B., Hayes, B. K., McDonald, R. I., & Newell, B. R. (2019). Predicting climate change risk perception and willingness to act. *Journal of Environmental Psychology*, *65*, 101331. https://doi.org/10.1016/j.jenvp.2019.101331

Yale Center for Environmental Law & Policy. (2020). *Environmental Performance Index 2020*. https://epi.yale.edu/

Yu, T.-K., Chang, Y.-J., Chang, I.-C., & Yu, T.-Y. (2019). A pro-environmental behavior model for investigating the roles of social norm, risk perception, and place attachment on adaptation strategies of climate change. *Environmental Science and Pollution Research*, *26*(24), 25178–25189. https://doi.org/10.1007/s11356-019-05806-7

Zeng, J., Jiang, M., & Yuan, M. (2020). Environmental Risk Perception, Risk Culture, and Pro-Environmental Behavior. *International Journal of Environmental Research and Public Health*, *17*(5), 1750. https://doi.org/10.3390/ijerph17051750

Zhou, Z., Liu, J., Zeng, H., Zhang, T., & Chen, X. (2020). How does soil pollution risk perception affect farmers’ pro-environmental behavior? The role of income level. *Journal of Environmental Management*, *270*, 1–10. https://doi.org/10.1016/j.jenvman.2020.110806

**Ringraziamenti**

Grazie al professor Veltri, che con i suoi consigli e la sua guida mi ha aiutato a realizzare questo progetto di tesi. Grazie alla professoressa Pavan che da alcuni anni ormai crede in me e non ha mai smesso di ricordarmelo. Grazie per i consigli e per avermi insegnato che a volte è sufficiente un cambiamento di prospettiva per vedere un po’ di luce.

Grazie ai miei genitori, instancabili sostenitori delle mie scelte e delle mie capacità. Grazie per avermi sempre incoraggiata dopo ogni caduta, incertezza, cambiamento o traguardo raggiunto. Grazie alle mie sorelle, Federica e Lara, che mi ricordano ogni giorno che non bisogna mai prendersi troppo sul serio e che una risata ti migliora sempre la giornata. E ovviamente grazie ai miei nonni, continua fonte di ispirazione per me. Con amore mi sostengono e non smettono mai di essere orgogliosi di me.

Grazie alle mie amiche e ai miei amici, che in questi anni hanno gioito e sofferto insieme a me, grazie per avermi guidata quando mi sembrava di aver perso l’orientamento e soprattutto grazie per avermi fatto sentire sempre a casa, nonostante la distanza. Grazie alle mie compegne e ai miei compagni di corso, per i pianti, le risate, i consigli, le correzioni di inglese e le serate spensierate passate insieme. Senza di voi questo percorso universitario non sarebbe stato lo stesso.

E infine, ma non meno importante, devo ringraziare Marco, che non ha mai smesso di scommettere e di credere in me. Grazie per l’infinita pazienza, amore e fiducia. Grazie per avermi insegnato la tenacia e la perseveranza. E soprattutto grazie perché insieme abbiamo imparato che non conta essere forti, ma sentirsi vivi e tornare a sorridere sì.

1. See Appendix A for the list of selected variables. [↑](#footnote-ref-1)
2. See Appendix C for the number of observations by country. [↑](#footnote-ref-2)
3. See Appendix B for the summary statistics. [↑](#footnote-ref-3)
4. See Appendix A for thesurvey question wording and coding. [↑](#footnote-ref-4)
5. See Appendix B for more detail about summary statistics. [↑](#footnote-ref-5)
6. See Appendix E for the summary statistics of the two subsets. [↑](#footnote-ref-6)
7. See Appendix D for the correlation of each class separately. [↑](#footnote-ref-7)
8. See Appendix E for summary composition of the two subsets. [↑](#footnote-ref-8)
9. See Appendix F for the grids of parameters and the hyperparamenters for each algorithm. [↑](#footnote-ref-9)
10. See Appendix G for the importance feature tables. [↑](#footnote-ref-10)
11. We stress that these coefficients are in log-odds terms the interpretation of log-odds terms is not so easy. For this reason, we convert the log-odds term into the odds ratio, which means the probability of an event occurring. When the odds ratio is greater than 1, it shows a positive relationship. If an odds ratio less than 1 means a negative relationship. See appendix G for more details about the log-odd table. [↑](#footnote-ref-11)
12. See Appendix G for Decision Tree plot. [↑](#footnote-ref-12)
13. See Appendix F for more details. [↑](#footnote-ref-13)
14. See Appendix H for the importance feature tables. [↑](#footnote-ref-14)
15. See Appendix F for Decision Tree plot. [↑](#footnote-ref-15)