Chapter 3

Exploratory Data Analysis

The following section illustrates the different steps undertaken to obtain a prediction model of pro-environmental behaviour.

In particular, 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 I obtain the final dataset, and therefore I 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 I have already mentioned in the Methodology (Chapter 2), in the first place the prediction is fitted to the entire dataset, and then, it is computed separately on two subsets: with the individuals that have a high-risk perception (qb2 greater or equal to 6) and with the individuals that have a low-risk perception (qb2 less than or equal to 5).

The last section of the chapter shows some characteristics at the macro-level. I 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 done.

**3.1 Unsupervised Machine Learning Algorithms: finding new patterns among citizens**

**Correlational Class Analysis**



Figure 1:CCA. Coding: qb4\_3 = benefits for companies, qb4\_5 = benefits for citizens, qb7 = renewable energy, qb8 = energy efficient, qb9 = greenhouse gas

CCA partitions the dataset into classes depending on the similarity of correlation patterns to apprehend unobserved heterogeneity (Rossoni et al., 2021). 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 has divided 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%). As previously stated, CCA aggregates individuals who agree about the cardinal arguments, although they do not coincide with the answers given or they do not concur with each other (Rossoni et al., 2020). Classes combine individuals who perform similar decision patterns between each pair of variables. Figure 1 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 more line is thicker, the more significant is the correlation. [[1]](#footnote-1)

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 that result if climate change is fought.

Group 3 has a slight correlation among all variables, except for benefits for citizens. Actually, *benefits for citizens* is always slightly correlated with *benefits for companies*, but it is not very significant.

In group 4 there is a strong positive correlation among *renewable energy*, *energy efficient*, and *greenhouse gas* and separately between *benefits for companies* and *benefits for citizens*. 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*; *benefits for citizens* and *energy efficient*; *benefits for citizens* and *greenhouse gas*. In this case, *benefits for companies* is completely isolated.

Figure 2: 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. I 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, I 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, I 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 governance of climate change. I remember 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, I 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 3. The best choice is 2.

Figure 3: Clustering silhouette

Observations are then divided into two clusters with PAM. The summary results of PAM clustering are shown in figure 4. 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 (the mode of *benefits for companies, benefits for citizens* is 2, while the mode of *renewable energy, energy efficient, greenhouse gas* is 1), I can recognize that 2 clusters resemble two different types of green-identity, which I call “moderate green identity”, “extreme green-identity”.

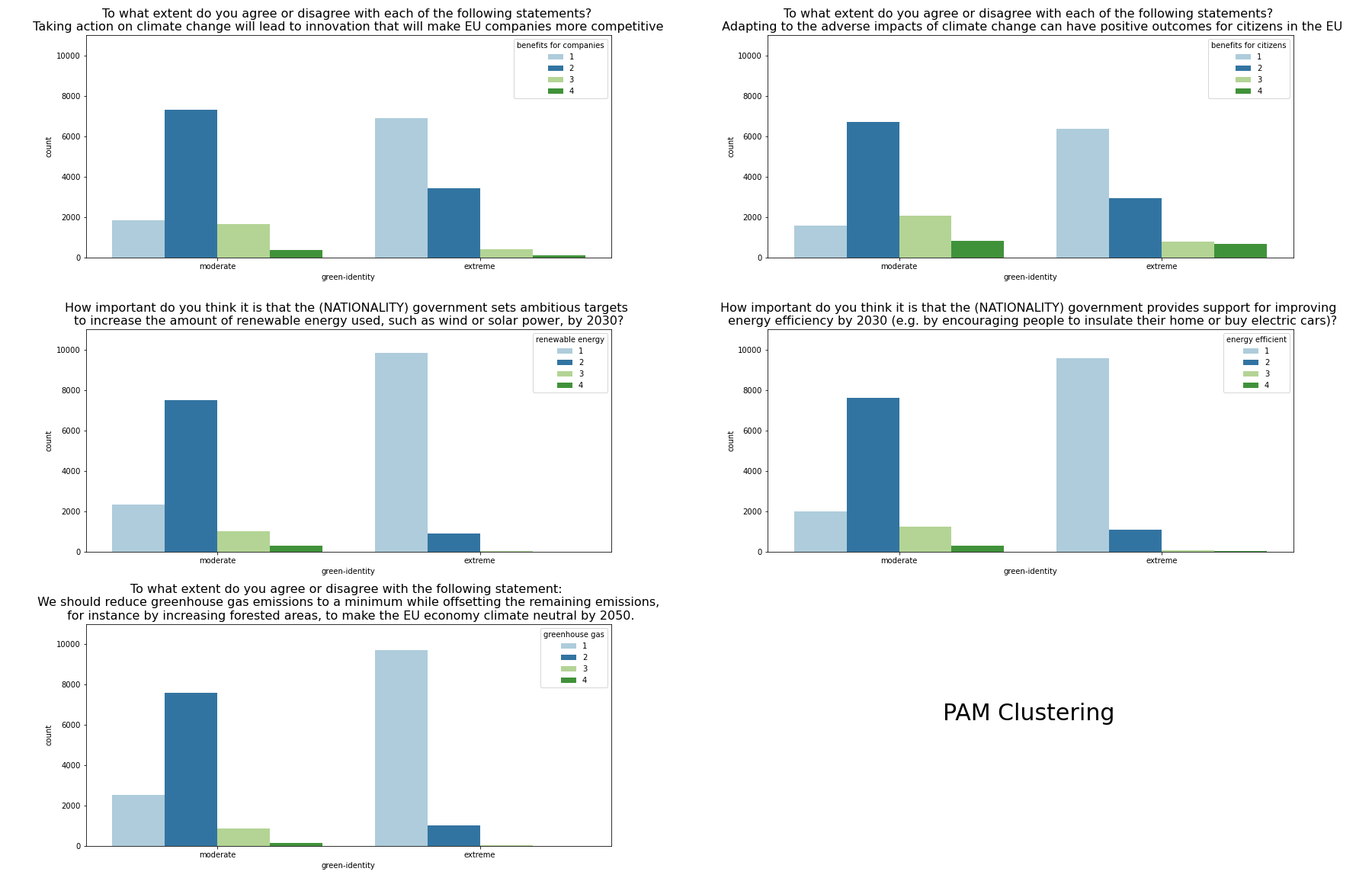


Figure 4: Summary of PAM Clustering

***Cluster 1: Moderate green identity***

* More than 75% answered questions (*benefits for companies, benefits for citizens, renewable energy, energy efficient, greenhouse gas* 9) with moderate answers (option 2 or 3);
* There is no difference among genders;
* 70% of citizens belong to centre, centre-left or centre-right;
* 57% stopped studying at the latest at 19, 35% stopped studying after 20, 6% are still studying;
* More than 65% of the citizens of Czech Republic, Estonia, Finland, Latvia, Poland belong to this cluster;
* the average age is 51 years old;
* the average climate change risk perception is 7.2;
* 56% take place a pro-environmental behaviour.

***Cluster 2: Extreme green-identity***

* More than 60% answered *benefits for companies* and *, benefits for citizens* questions with the maximum answer (1), and about 90% answered *renewable energy, energy efficient, greenhouse gas* questions with the maximum answer (1).
* There is no relevant difference among genders;
* 66% of citizens belong to centre, centre-left or centre-right
* 49% stopped studying at the latest at 19, 41% stopped studying after 20, 7% are still studying;
* More than 65% of the citizens of Cyprus, Denmark, Spain, United Kingdom belong to this cluster;
* The average age is 50 years old:
* The average climate change risk perception is8.6;
* 75% take place a pro-environmental behaviour.

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**

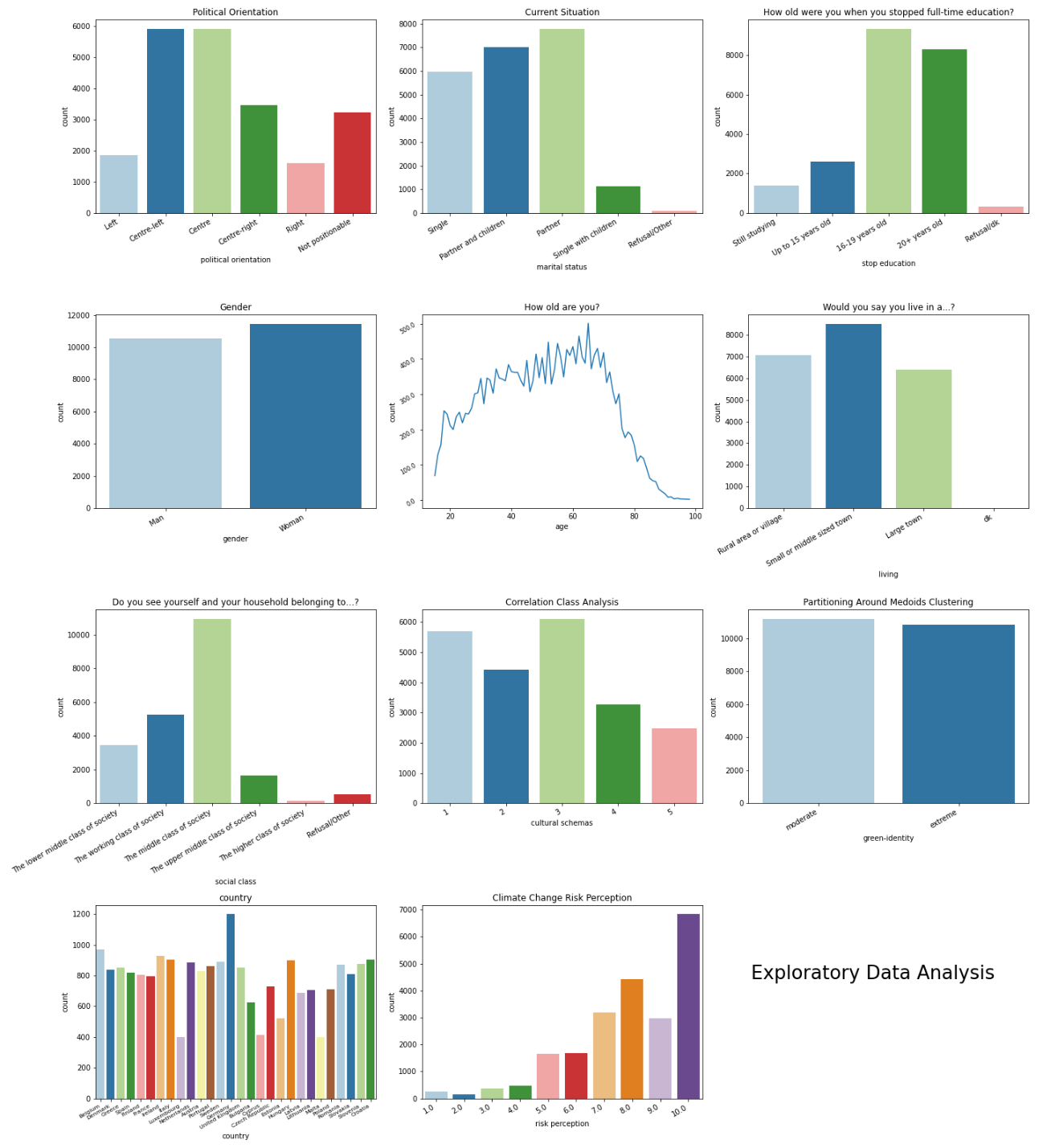
****

Figure 5: Countplot of explanatory variables-the entire dataset

Before starting with the analysis it is important to describe the dataset. Figure 5 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 and 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.

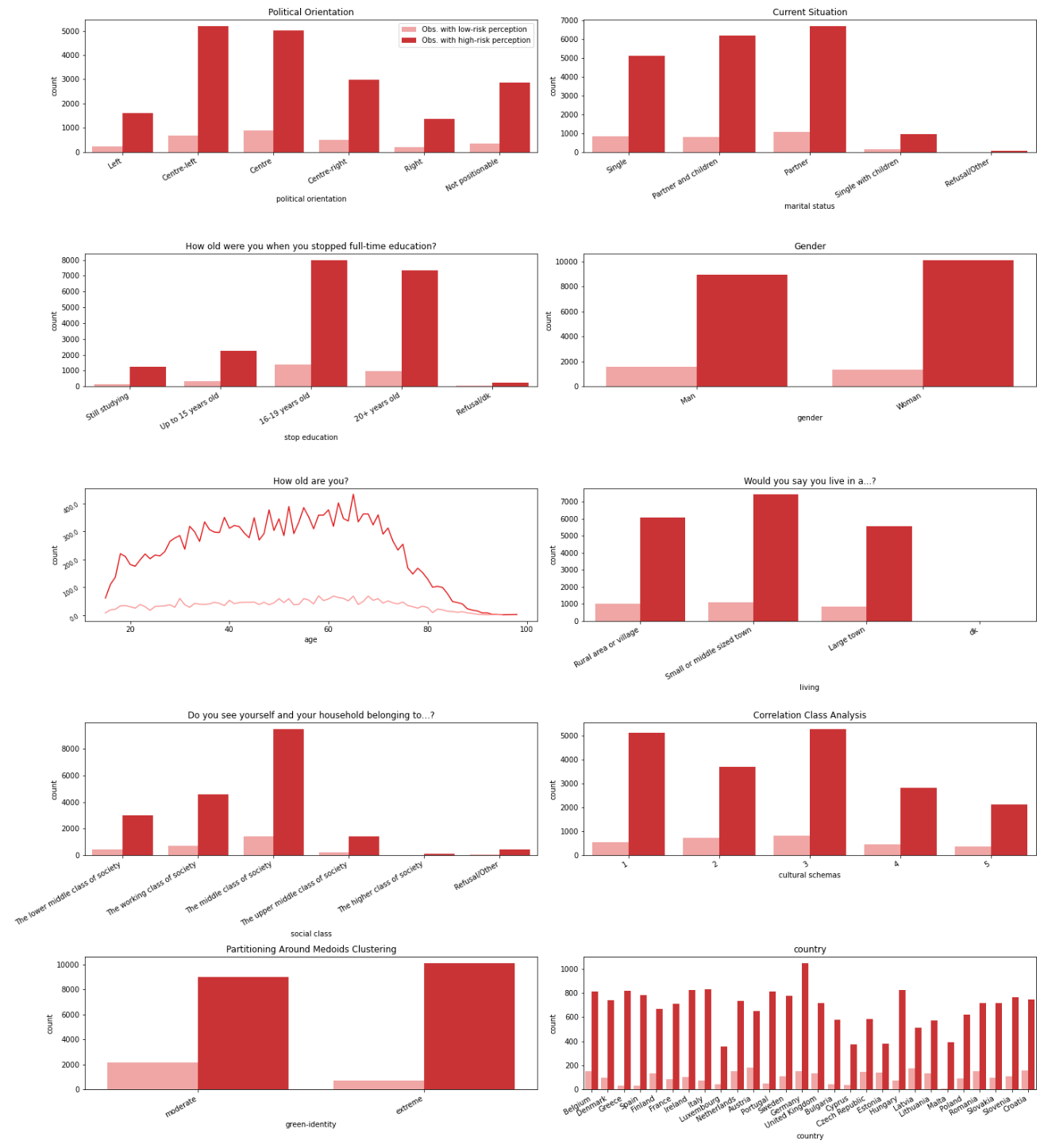
**The Datasets based on the Individual Risk Perception Level** 

Figure 6: Countplot of explanatory variables based on the individual risk perception level

Instead, figure 6 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.[[2]](#footnote-2) The subset with low-risk perception observations (in the graph it is represented with pink color) has 2906 cases (13%), while the high-risk perception one (red color) 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. Additionally, most of the citizens do not come from Germany but Austria. Lastly, as I 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.

**3.3 Macro-level factors**

Climate change attitudes do vary only between citizens but also between countries (Xie et al., 2019). As you can see in figure 7, 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 7: Single Most Serious Problem per Country

Another interesting example is the difference in behavior. As you can see in figure 8, more than 60% of the citizens of Romania declared that they are not behaving in favor of the environment. Conversely, Malta has 95% of citizens that take place 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. These possible characteristics are not taken into account, but I want to remind the reader of the hypothetical influence that these could have on the outcome of the behaviour.

Figure 8: Pro-environmental Behaviour per Country

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 should be significant mediators in explaining risk perception or pro-environmental behavior. Some indexes are considered with the sole purpose of remembering that they could affect and moderate the phenomenon studied. 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 macro levels 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 9 presents the score across European Union (EU). The best score is obtained from Denmark, while the worst from Bulgaria.

Figure 9: 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 10 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 10:The 2019 Climate Change Policy

In this way, It is important to remember that these macro-factors should affect also individual preferences.

Chapter 4

Analysis and 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 it is divided into two groups of citizens based on the individual climate change risk perception. I 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 I will choose the best model is the maximization of these metrics. I 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 labeled 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, I decide to compromise: a fair 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, I explain how to understand what variables have a fundamental role in our models. I 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, *scikit-learn* implemented Gini Importance, also in this analysis, it is adopted.

**4.2 Prediction Behavior on the Entire Dataset**

In this part, the research study 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. The technique adopted for knowing the optimal hyperparameter is called *random search* (RandomizedSearchCV() in *scikit-learn*) (Benner, 2020).[[3]](#footnote-3) I 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.

I remember 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.

**4.2 Prediction of the Behaviour on the Entire Dataset**

It can be convenient, to summarize quickly, the predictive accuracy and macro-F1 of the models trained are indicated in table 1, while the comparison of feature importance is shown in figure 11.[[4]](#footnote-4) The accuracy and macro-F1 are slightly better in the Random Forest’s model, which are respectively 0.70 and 0.67.

Table 1: 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 11:Comparison of variables importance |  |

The Logistic Regression model has 0.67 of accuracy and 0.66 of macro f1-score. Different countries are the most important variables in this model. The coefficient plot shows that Romania, Bulgaria, Czech Republic, and Poland have a negative influence on predicting pro-environmental behaviour. it very interesting due to all these first countries are situated in Eastern Europe. Instead, Malta, Finland, and Portugal have a positive influence on predicting pro-environmental behaviour. [[5]](#footnote-5)

The Decision Tree model generates an accuracy of 0.66 and a macro-F1 of 0.53. Cluster 2, or called extreme green-identity, 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 cluster 2 and he/she has a level of risk perception greater or equal than 7,5 he/she is classified in the *yes-action* class.[[6]](#footnote-6) 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.[[7]](#footnote-7)

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. As shown in figure 11, analysing the feature importances of the model, I notice that age and, once again, climate risk perception are the most important variable 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, cluster 1, moderate green-identity, is more likely to do not behave in favour of the environment than cluster 2.

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. Cluster 1, moderate green identity, 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 tends to take place environmental action. This statement is confirmed in the data. If I analyse the percentage of those who perform actions within the level of perceived risk, I 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 taking place action increase in turn, as figure 12 indicates. Climate change risk perception is one of the predictors which align in almost all models. Therefore, I confirm that a higher individual climate change risk perception positively influences and predicts pro-environmental behaviour.

Figure 12: Crosstab between Risk Perception and Behaviour

Green identity, measured with clusters, is another important variable found in almost all models. 75% of individuals belonging to cluster 2 (extreme green-identity) have taken some eco-friendly actions. While 55% of individuals belonging to cluster 1 (moderate green-identity) perform pro-environmental behaviour. Extreme green-identity has a positive influence on predicting behaviour. Therefore, also the second hypothesis is confirmed.

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

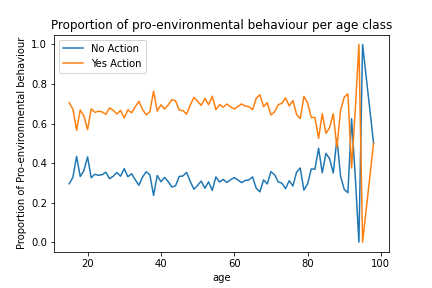
The last hypothesis concerns the importance of socio-demographic information. The Random Forest, our best model, suggests that the most important variable is age. However, it is not clear the relationship with the behaviour. I expect that most young people take place pro-environmental action. If I examine the percentage of those who perform pro-environmental actions within the age class, I find that not only youngsters behave pro-environmentally, as shown in figure 13. The trends are constant up to the age of 80.

Figure 13: Crosstab between age and behaviour

Also, 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.

However, I discover that a fundamental role in the prediction is the country of origin. Especially Romania is an important variable in all models. It is a negative effect on pro-environmental behaviour, since 66% of Romanians do not behave eco-friendly.

**4.4 Prediction of the behavior based on the Individual Risk Perception Level**

In the second part of the analysis, I divide the subset 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, I will call the first subset “of those who have a high-risk perception” and the second one “of those who have a low-risk perception”. At this point of the analysis, I 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, I want to understand what the most important predictors are if I change the level of risk perception. Also in this part, I implemented the same algorithms for both subsets: Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. Hyperparameters using random search are found.[[8]](#footnote-8)

We remember that 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 2 and 3 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 those who worry about climate change. This subset is particularly unbalanced, therefore macro-F1 is preferred to evaluate the model.

Table 2: Metrics Comparison-Subset of those who have a high-risk perception.

|  |  |  |
| --- | --- | --- |
| **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 3: Metrics Comparison-Subset of those who have a low-risk perception.

|  |  |  |
| --- | --- | --- |
| **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 those who have a high-risk perception. | Subset of those who have a low-risk perception. |
| Logistic Regression | |
|  |  |
| Decision Tree | |
|  |  |

|  |  |
| --- | --- |
| Random Forest | |
|  |  |
| Gradient Boosting | |
| Figure 14: Comparison of variables importance based on the individual risk perception level |  |

Figure 13 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. [[9]](#footnote-9)

The accuracies of the two Logistic Regression models are 0.65 for the subset of those who have a high-risk perception, and 0.63 for the subset of those who have a low-risk perception. 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 cluster 2, the extreme green-identity, on the other hand, the root is employed by cluster 1, the moderate green-identity. I 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.

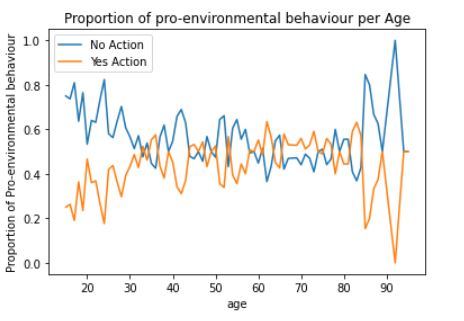
The Random Forest model yields the best accuracies in both models, 0.70 and 0.65. However, it does not happen in macro-F1 in the model of those who have a high-risk perception. In both cases, age is the most important variable for the prediction of behaviour. The trend of age in the subset with high-risk 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 second subset, as figure 15 shows. On average 70% of younger have negative behaviour in favor 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 first subset 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 15: trend of age in the Subset with low-risk perception observations.

Accuracies of the Gradient Boosting models yield 0.67 and 0.65. The model of those who have a high-risk perception macro-F1 is 0.64, which is the best result. In both models, there are two main important variables: cluster 1, moderate green-identity, and Romania. In the first subset, there is also the high-education level as an important variable. In the second one, I 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 I compute the proportion in this subset 50% of radicals take place pro-environmental behaviour, while the other orientations get a smaller percentage.

To sum up, in this second part of the analysis I 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 those who worry about climate change. 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.

The opposite case is the subset of those who do not worry about climate change. In this case, Romania and cluster 1, moderate green-identity, 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.

**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 those who have a high-risk perception***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 those who have a low-risk perception***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 for complete model**

***Logistic Regression***

{'solver': ['newton-cg', 'lbfgs', 'liblinear'],

'penalty': ['l2'],

'C': [100, 10, 1.0, 0.1, 0.01]}

1. **Complete model**

LogisticRegression(C=0.1, class\_weight={0: 1.43631197, 1: 0.76700502}, 'solver': 'lbfgs', 'penalty': 'l2')

1. **Model with observations of those who have a high-risk perception**

LogisticRegression(C=0.1, class\_weight={0: 1.56837406, 1: 0.73400044}, 'solver': 'lbfgs', 'penalty': 'l2')

1. **Model with observations of those who have a high-risk perception**

Model with Low Risk Perception Observations

LogisticRegression(C=10, class\_weight={0: 0.92119565, 1: 1.09354839},solver='newton-cg', 'penalty': 'l2')

***Decision Tree***

{'criterion': ['gini', 'entropy'],

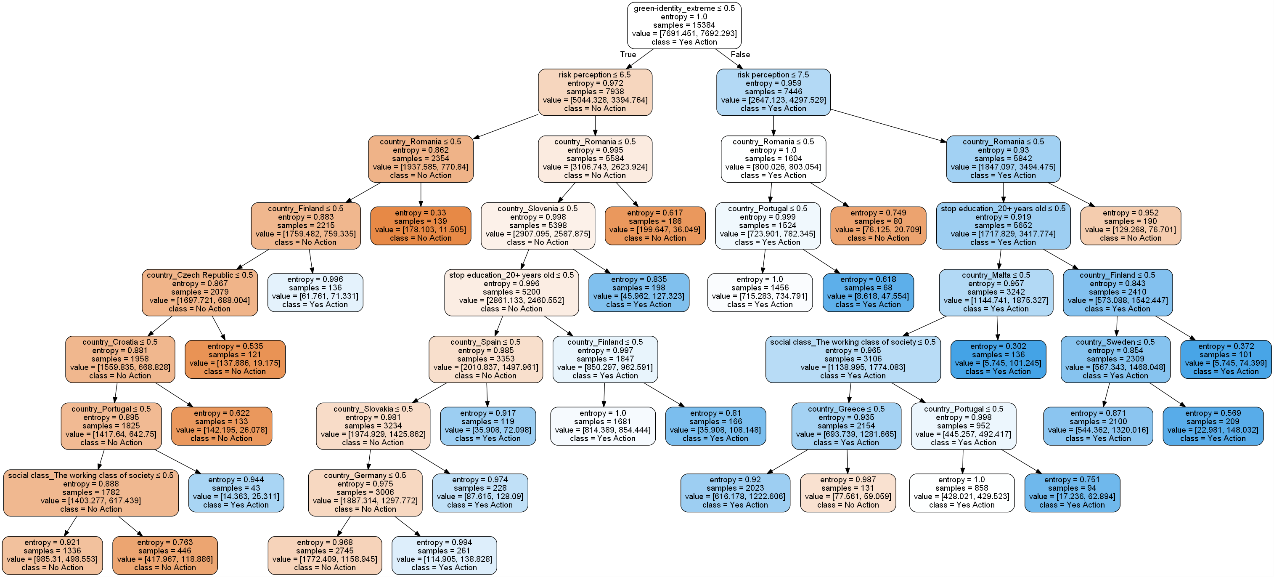
'max\_depth': range(1, 10),

'min\_samples\_split': range(0, 10),

'min\_samples\_leaf': range(1, 5)}

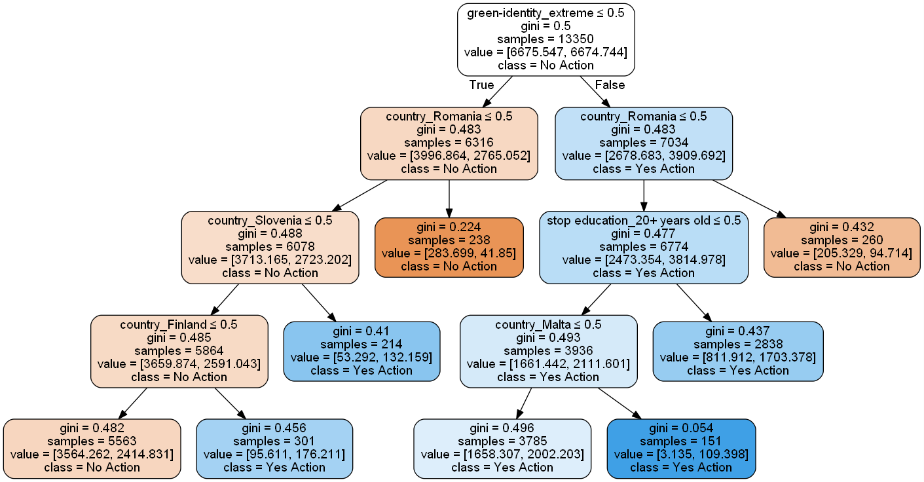
1. **Complete model**

DecisionTreeClassifier(ccp\_alpha=0.000993163342228487,class\_weight={0: 1.43631197, 1: 0.76700502},criterion='entropy', max\_depth=8, min\_samples\_split=4,random\_state=123)



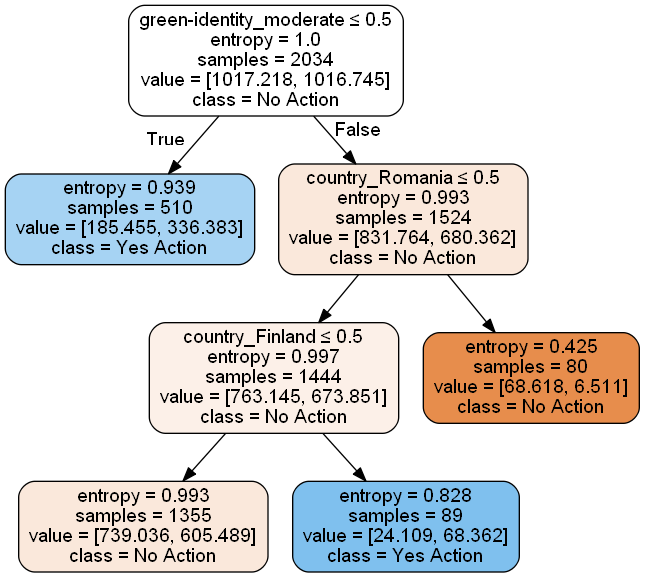
1. **Model with observations of those who have a high-risk perception**

DecisionTreeClassifier(class\_weight={0: 1.56739773, 1: 0.73421448}, max\_depth=7,min\_samples\_leaf=3, min\_samples\_split=6, ccp\_alpha=0.0022213271836104)



1. ***Model with observations of those who have a high-risk perception***

DecisionTreeClassifier(ccp\_alpha=0.009396010830688362,class\_weight={0: 0.92727273, 1: 1.08510638},criterion='entropy', max\_depth=5, min\_samples\_leaf=2,min\_samples\_split=8, random\_state=50)



***Random Forest***

{'n\_estimators': [50, 120, 190, 260, 330, 400],

'max\_depth': [5, 8, 15, 25, 30, None],

'min\_samples\_split': [2, 5, 10, 15, 100],

'min\_samples\_leaf': [1, 2, 5, 10],

'max\_features': ['auto', 'sqrt', 'log2'],

'bootstrap': [True, False]}

1. **Complete model**

RandomForestClassifier(bootstrap=False,class\_weight={0: 1.43631197, 1: 0.76700502}, max\_depth=30, max\_features='log2', max\_leaf\_nodes=49,min\_samples\_leaf=10, min\_samples\_split=10, n\_estimators=136)

1. **Model with observations of those who have a high-risk perception**

RandomForestClassifier(class\_weight={0: 1.56739773, 1: 0.73421448},max\_depth=25, min\_samples\_split=5, n\_estimators=150)

***3.Model with observations of those who have a high-risk perception***

RandomForestClassifier(class\_weight={0: 0.9272727, 1: 1.08510638}, max\_depth=25,min\_samples\_leaf=5, n\_estimators=500)

***Gradient Boosting***

{'max\_depth': [3, 4, 5],

'learning\_rate': [0.1, 0.25, 0.5, 0.75, 1],

'n\_estimators': [50, 100, 150],

'gamma': [0.5, 1, 1.5, 2],

'min\_child\_weight': [1, 5, 10]}

1. **Complete model**

XGBClassifier(n\_estimators= 100, class\_weight={0: 1.43631197, 1: 0.76700502}, gamma=0.5, max\_depth=4, min\_child\_weight=10, scale\_pos\_weight=0.75, learning\_rate=0.1)

1. ***Model with observations of those who have a high-risk perception***

XGBClassifier(n\_estimators= 50, class\_weight={ 1.56726931, 0.73424266}, gamma=1, max\_depth=5, min\_child\_weight=5, scale\_pos\_weight=0.5, learning\_rate=0.25)

1. ***Model with observations of those who have a low-risk perception***

XGBClassifier(n\_estimators= 50, class\_weight={ 0: 0.92727273 , 1:1.08510638}, gamma=0.5, max\_depth=5, min\_child\_weight=1, scale\_pos\_weight=0.75, learning\_rate=0.1)

**Appendix G. 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. ***Dataset with observations of those who have a high-risk perception***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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. ***Dataset with observations of those who have a low-risk perception***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 |

**REFERENCES**

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

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

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

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

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

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

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

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

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

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/

1. See appendix D for the correlation of each class separately. [↑](#footnote-ref-1)
2. See appendix E for summary composition of the two subsets. [↑](#footnote-ref-2)
3. See appendix for the grids of parameters and the hyperparamenters for each algorithm. [↑](#footnote-ref-3)
4. See appendix G for the importance feature tables. [↑](#footnote-ref-4)
5. I 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-5)
6. See appendix F for Decision Tree plot. [↑](#footnote-ref-6)
7. [↑](#footnote-ref-7)
8. See appendix for more details. [↑](#footnote-ref-8)
9. See appendix G for the importance feature tables. [↑](#footnote-ref-9)