Tilburg University

PREDICTING INDIVIDUAL ATTITUDES TOWARDS IMMIGRATION IN EUROPE

COMPARISON OF MACHINE LEARNING MODELS

GRZEGORZ POWICKI

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
OF TILBURG UNIVERSITY

STUDENT NUMBER

2149431

COMMITTEE

dr. Olier
ir. Lieber

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

December 1st, 2025

WORD COUNT

8713

ACKNOWLEDGMENTS

I would like to thank Dr. Olier for his valuable feedback and guidance throughout this project. I am also grateful to my family and friends for their constant support, especially those who accompanied me during my time in the Netherlands.

PREDICTING INDIVIDUAL ATTITUDES TOWARDS IMMIGRATION IN EUROPE

COMPARISON OF MACHINE LEARNING MODELS

GRZEGORZ POWICKI

Abstract

Public sentiment towards immigration remains a contentious issue within contemporary Europe. While extensive literature has examined the relationship between public's attitudes and socio-demographic and economic drivers, prior research has relied on explanatory regression models, overlooking predictive accuracy. This thesis addresses this methodological gap by applying machine learning techniques to assess the classification of individual immigration attitudes, additionally putting an emphasis on the predictive power of institutional and social trust. Utilizing data from the European Social Survey (ESS) Wave 11, this study evaluates the performance of Logistic Regression against ensemble methods. Namely, Random Forest, XGBoost, LightGBM, and CatBoost, and employs SHAP values for model interpretability.

Results indicate that while overall performance metrics are comparable, gradient boosting models demonstrate statistical superiority over classical approaches. In the pooled European model incorporating trust variables, LightGBM achieves an F1 score of 0.66 for immigration opponents and a ROC AUC of 0.79, surpassing predictive baselines established in earlier studies. Moreover, trust in other people and the European Parliament is a key predictor linked to positive attitudes, however, country-level models show that its influence varies across states. These findings demonstrate that immigration attitudes are driven by a complex relationship between psychological and contextual factors.

1 DATA SOURCE, ETHICS, CODE, AND TECHNOLOGY STATEMENT

This thesis uses data from the eleventh wave of the European Social Survey, ESS (European Social Survey European Research Infrastructure (ESS ERIC), 2025). The dataset was obtained from the official ESS website after online registration and is fully anonymised. Work on this thesis did not involve collecting new data from human participants or animals. The original owner of the data retains full ownership of the data during and after the completion of this thesis. The author of this thesis does not have any legal claim to the data. The dataset can be accessed at (https://ess.sikt.no/en/?tab=overview).

All figures and tables in this thesis were created by the author. The analytical code is based on open-source libraries. The software utilized in this study is documented in Appendix C, together with the corresponding references.

ChatGPT (https://openai.com/chatgpt) was used for coding assistance and improving the clarity of the author's original content. Grammarly (https://www.grammarly.com/) was used for spelling and grammar checks. No other external editing, typesetting, or writing services were used. The code is available at https://github.com/Grzegorz-Powicki/

2 INTRODUCTION

2.1 *Context*

In recent years, immigration has been considered one of the most pressing socio-economic and political challenges in Europe (Potančoková et al., 2021). When asked to name the two most concerning issues facing their countries and the European Union, EU citizens consistently mention immigration as one of them (European Commission, 2025). The importance of social attitudes towards immigration stems from their impact on policymakers, immigration strategies, and political cleavages (Dražanová et al., 2024; Javdani, 2020). These stances are also significant for the immigrants themselves. They determine the immigrants' social and economic integration into host countries, which can affect their mental well-being and social cohesion (Javdani, 2020; Verkuyten, 2021). Considering all stated factors, it is crucial to understand individual attitudes towards immigration due to their impact on both newcomers and host societies.

The attitudes of individuals towards immigration have been examined mainly from a statistical perspective to identify their socio-demographic and attitudinal determinants (Dražanová et al., 2024). These methods usually explain a limited amount of the variance, suggesting that this phenomenon might be more intricate and potentially influenced by non-linear relationships and interactions. Simultaneously, they are focusing on associations rather than providing predictive information on individual attitudes. Predictive research, conversely, remains scarce, even though it offers a complement to explanatory studies, by testing the predictive capabilities of their interpretations (Y. Chen et al., 2021). In addition, machine learning models applied within this predictive framework are specifically suited to capture more complex patterns, like nonlinearities and interactions, which are often missed by statistical methods (Dahiya et al., 2022). However, when applied, such approaches typically relied on a narrow set of models or predictors.

As per the outstanding literature, boosting methods, namely XGBoost, LightGBM, and CatBoost, have not been applied to predicting Europeans' individual immigration attitudes using sociodemographic and attitudinal data. Nguyen and Byeon (2024) demonstrated how such algorithms can be utilized in social science to predict life satisfaction. In a study by Yektansani et al. (2024) on predicting individual environmental attitudes among Europeans, Gradient Boosting outperformed models such as Logistic Regression, Random Forest, and Neural Networks. This raises the question of how accurately these algorithms can predict individual attitudes toward immigration exclusively in the European context.

2.2 *Scientific relevance*

This thesis holds scientific relevance in several areas. Firstly, it compares machine learning models previously unexplored in this domain (XGBoost and CatBoost), with those already examined (Logistic Regression, Random Forest, and LightGBM). It is worth noting that although LightGBM has been used in studies on immigration attitudes (Han, 2022), it has been applied to different types of target variables, making its use in this research another contribution to the literature. What is more, this work examines the predictive value of features that have not been used in a predictive context, such as institutional and social trust, to determine their relevance in explaining individual views on immigration in Europe. Lastly, this thesis uses the most recent European Social Survey conducted in 2023–2024. This particular version of the dataset has not been used in predictive research, which makes the use of the latest wave valuable for generating updated insights into current public immigration attitudes.

2.3 *Societal relevance*

Given the significance and scale of immigration in Europe, the societal relevance of this study lies in forecasting possible opponents of immigration and the determinants of their attitudes. Such knowledge can be useful for policymakers and assist them in constructing more accurately targeted integration strategies, considering the differences in support and opposition. These strategies may include diversity education and initiatives that foster contact between locals and newcomers, possibly resulting in reducing prejudice and improving social cohesion for the benefit of both immigrants and the host population.

2.4 *Research questions*

The main research question of this study is:

> *To what extent can different machine learning models predict individual attitudes towards immigration in Europe based on socio-demographic and attitudinal data, and which factors are the most influential for these predictions?*

The following sub-questions have been designed to enable a more in-depth investigation of this topic:

SQ1 *Among the evaluated models (Logistic Regression, Random Forest, LightGBM, XGBoost, CatBoost), which of them achieves the best results according to F1-score and ROC-AUC*

SQ2 *Which features contribute most to predicting immigration attitudes in the global baseline model without trust variables, according to SHAP values?*

SQ3 *To what extent do trust-related features contribute to predicting individual immigration attitudes in the global model?*

SQ3 *How biased is the best-performing global model across demographic subgroups?*

SQ4 *How do trust-related features alter country-level model performance and key predictors of immigration attitudes?*

2.5 *Main findings*

Among the evaluated algorithms, boosting models performed best, although the differences between them were negligible. Additionally, the study showed that adding trust-related features can slightly improve the LightGBM model using data from all European countries, and these features rank among the strongest predictors. However, when the model is trained separately on country-level data, the effects of trust-related features become much more heterogeneous across countries.

3 LITERATURE REVIEW

The objective of this section is to examine individual attitudes toward immigration and their determinants, especially in the European context.

The first part presents studies that employ statistical methods to achieve this objective. Subsequently, studies that concentrate on predictive approaches are summarized with emphasis on machine-learning algorithms used in this domain. Additionally, studies from other relevant social science fields are referenced as sources of potential approaches for this research. The last section outlines the identified gaps and explains how this study aims to address them.

3.1 *Correlational studies*

Goubin et al. (2022) analyzed Europeans' immigration attitudes with multilevel regression. Their findings show that negative perceptions are more common among older people, those with lower levels of education, higher religiosity, lower income, residents of rural areas, and less skilled job workers. Women seem to hold slightly more positive attitudes than men. Such positive views are also linked to greater political and interpersonal trust, interest in politics, and more left-wing political positions.

Palermo et al. (2022) reached mainly the same insights, using a linear regression (R^2 of approximately 0.15). However, in comparison to the work of Goubin et al. (2022), the level of religiosity was found to be not statistically significant. Moreover, Palermo et al. (2022) concluded that the effect of higher education on immigration attitudes is particularly important in rural areas and smaller cities compared to more urbanized areas, showing the possible existence of contextual differences.

Interestingly, Umansky et al. (2025) concluded that the positive effect of education is highly context dependent and shaped by country, with clear differences between Western and Eastern Europe, due to macroeconomic conditions. Furthermore, Rooduijn (2022) observed that the impact of education on immigration attitudes varies across countries and becomes more polarising in generous welfare states, where lower and medium educated groups are more negative towards immigrants.

Contextual dependencies were also found by (Schahbasi et al., 2021), who revealed that although women are generally less skeptical towards immigration than men, their attitudes become more critical with increasing age. Moreover, being more religious is associated with more positive stances, which contrasts with the findings of Goubin et al. (2022).

Halapuu et al. (2013), concentrated on the role of institutional and social trust on Europeans' immigration attitudes, which appeared to be among the strongest predictors. The OLS regression models achieved an R^2 score of approximately 0.30. Halapuu et al. (2013) attribute the importance of trust in institutions to the assurance that they act fairly and protect citizens, thereby reducing perceived risk of living in a culturally diverse society. The high significance of both social and institutional trust aligns with the results of Goubin et al. (2022). Social trust, as the strong driver of immigration attitudes, was also described by Mitchell (2021).

Hannuksela et al. (2024) broadened the understanding of individual perceptions of immigration by examining the interrelated role of political interest and ideology. According to their results, political engagement increases positive perceptions of immigrants, mainly among less conservative individuals. The significance of political interest is consistent with the findings of Goubin et al. (2022).

3.2 *Machine learning studies*

Recently, prediction has gained increasing importance as a necessary complement to explanation in social sciences (Y. Chen et al., 2021). Y. Chen et al. (2021) argue that the scientific rigor of sociology should be examined by testing its interpretations through predictive accuracy, which makes theoretical propositions more reliable. For this purpose, the growing utilization of machine learning is seen as a key development that enables validation through prediction.

Drouhot et al. (2023) discussed the usefulness of such methods in the context of immigration studies. Although their research primarily emphasized its applicability for digital trace data, it also underlined the possibility of utilizing such approaches to more traditional datasets, which is specifically relevant for this research, as it applies structured survey data.

Erhard and Heiberger (2023) predicted individual attitudes towards immigration using ESS data from 2002. The binary target variable was constructed based on three survey questions asking respondents about their willingness to accept immigrants from three different backgrounds. Because this thesis follows exactly the same approach, the detailed description of calculating this index is provided in 4.2. Erhard and Heiberger (2023) utilized classical approaches like Logistic and Ridge Regression and compared them to a more advanced machine learning algorithm – Random Forest. The performance of the models is as follows: Logistic Regression (F1 = 0.731, Accuracy = 0.6885), Ridge Regression (F1 = 0.735, Accuracy = 0.699), and Random Forest (F1 = 0.728, Accuracy = 0.6896), with F1 Scores reported for the positive immigration attitudes. The inclusion of feature interactions in Ridge Regression made its results slightly higher compared to the Logistic Regression, which indicates the presence of more intricate patterns in the immigration attitudes. Furthermore, Erhard et. al (2023), through Accumulated Local Plot, demonstrated possible non-linear relationships, showing the inverted U-shape pattern for religiosity. The authors highlight this as the main advantage of applying more advanced machine learning models, which can reveal nonlinearities in comparison to statistical approaches that assume linear relationships between variables.

Arcila Calderón et al. (2022) predicted Europeans' probabilities of supporting refugees. For that purpose, they employed Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K Nearest Neighbour, achieving macro F1 scores of 0.68, 0.52, 0.64, 0.68, and 0.41, respectively. Compared to the research by Erhard of al. (2023) (Erhard & Heiberger, 2023), the difference in performance between Logistic Regression and Random Forest is greater. This can be attributed to the fact that Arcila Calderón et al. (2022) used basic sociodemographic data as determinants, while Erhard and Heiberger (2023) placed more focus on attitudinal and ideological factors, which can form more intricate relationships in which more advanced models can succeed. Since this thesis incorporates an even broader range of determinants, integrating sociodemographics with attitudinal factors and adding trust-related measures, such advanced models are expected to perform even better.

Han (2022) examined attitudes of South Koreans towards immigrants using LightGBM and features mostly describing personal contacts with immigrants. Two independent indicators were analysed: openness to multiculturalism and support for preferential hiring of native Koreans in times of job scarcity. The former was treated as a continuous value, and a LightGBM regressor achieved an RMSE of 0.16, while the latter was a binary value, and a LightGBM classifier resulted in an F1 score of 0.97. Even though the excellent performance of LightGBM may be attributed to the specificity of targets and immigration-related features, it indicates that such an algorithm can be successful in the immigration context, motivating its evaluation on the target used in this research.

3.3 *Studies from other social domains*

Considering the limited application of predictive approaches in research on individual attitudes towards immigration, studies in related social areas are referenced to obtain a broader view of methodological options.

Yektansani et al. (2024), using ESS data, predicted individual environmental attitudes. Random Forest and Gradient Boosting generally performed better than Logistic Regression and Neural Networks, indicating their effectiveness in forecasting individual social attitudes. Environmental attitudes are largely driven by age and education level (Yektansani et al., 2024), which are also important drivers of immigration attitudes, motivating the application of boosting algorithms in this thesis.

Nguyen and Byeon (2024) predicted life satisfaction of South Koreans, comparing various tree-based algorithms with TabNet. Tree-based ensembles outperformed the standard TabNeT and were slightly worse than their version modified by the authors. Among the tree-based models, CatBoost was the top performer, followed by LightGBM, XGBoost, and Random Forest. Since Nguyen and Byeon (2024) demonstrated the usefulness of boosting algorithms on large-scale survey data similar to that used in the current study, it also motivates including such algorithms in this research.

3.4 *Gaps and contribution*

Based on the literature review, several research gaps are identified. Firstly, the majority of the research on individual immigration attitudes concentrates on correlational frameworks. Although they provide valuable knowledge regarding factors shaping these attitudes, they often explain a limited share of variance (Halapuu et al., 2013; Palermo et al., 2022), indicating that they struggle to fully capture the complexity of this phenomenon. They also reported meaningful interactions between variables, meaning that immigration attitudes are complex and multidimensional phenomena. Together, these observations motivate applying machine learning methods, which are able to implicitly detect interactions, hidden patterns, and nonlinear relationships often missed by conventional approaches (Dahiya et al., 2022). Additionally, Drouhot et al. (2023) emphasized the potential of machine learning in immigration studies. However, their application remains scarce. This thesis aims to fill these gaps by comparing machine learning models unexplored in this domain to the current state-of-the-art approaches.

The variety of features used in the existing predictive approaches in immigration research remains limited as well. This research seeks to bridge this gap by testing new variables as predictors, focusing on trust-related data. Despite being proven as strong determinants of attitudes towards immigration (Drouhot et al., 2023; Halapuu et al., 2013; Mitchell, 2021), they haven't been checked in a predictive setting, which could better substantiate their role and clarify their interpretation in shaping immigration attitudes (Y. Chen et al., 2021).

The limited use of machine learning models and features may stem from the fact that predictive studies have mainly aimed to showcase the potential of machine learning in immigration research (Erhard & Heiberger, 2023), rather than offering systematic comparisons.

4 METHODOLOGY

4.1 *Dataset descritpion*

This thesis uses data from the European Social Survey (ESS), Wave 11 (European Social Survey European Research Infrastructure (ESS ERIC), 2025). It is a cross-national survey, which is conducted every 2 years across Europe in the form of in-person interviews. Its main purpose is to assess the values, attitudes, and behavioural patterns among European citizens. In addition, sociodemographic information on respondents is collected, which together makes the dataset particularly suitable for predicting individual attitudes toward immigration.

The dataset consists of 46,162 instances, each representing an individual respondent, and 676 variables. Out of these variables, 25 were determined based on the literature review as the most relevant to the objectives of this study (See Appendix A, Table 13). It is worth mentioning that the large number of variables in the original dataset results from some questions being specified separately for each country, for example, those related to the education level. However, the dataset also includes standardized variables that allow for cross-country comparison, thereby enabling a more manageable structure for analysis.

4.2 *Data preprocessing*

Given the influence of the data quality on the performance of machine learning models, data preprocessing is a crucial part of performing such an analysis (Paranjape et al., 2022). It consists of several parts, including handling outliers, dealing with missing values, and feature engineering (Alasadi & Bhaya, 2017).

4.2.1 *Data cleaning and preparing*

Firstly, the raw dataset is prepared for the subsequent steps of data preprocessing. As mentioned, the dataset consists of numerical, ordinal, and categorical data types, with the categorical fields already label encoded in the original dataset. To improve interpretability during the mentioned steps and data visualisation, these fields were decoded to their descriptive category labels.

Furthermore, the dataset contains special labels indicating missing instances, including: 'Refusal', 'Don't know', 'No answer', and 'Not applicable'. To prepare data for the subsequent analysis, they were re-coded to missing values, represented as 'None'. The percentage of missing instances in each of the features, together with their types, is presented in Table 13, Appendix A. Following that, instances reflecting respondents from Israel were removed from the dataset, as this thesis focuses mainly on the European context.

During this step, an analysis was conducted to check for possible errors or inconsistencies in the dataset, which were found in a variable depicting the size of the respondent's household, as several instances had values of zero in this column. These rows were removed because, in practice, a household cannot have zero members, and they were present only in 6 observations, making the removal nondisruptive. Additionally, variable 'hincfel' describing feeling about the household income was reverse recorded, so that higher values indicate a more favourable perceived financial situation. After these steps, the dataset consists of 43746 instances.

4.2.2 *Target variable calculation*

The task of this thesis is a binary classification with a target reflecting individual attitudes towards immigration. Following the approach of Erhard and Heiberger (2023), it is calculated using three survey questions on immigration preferences: 'imsmetn' (allow many or few immigrants of the same race or ethnic group as the majority), 'imdfetn' (allow many or few immigrants of a different race or ethnic group from the majority), and 'impcntr' (allow many or few immigrants from poorer countries outside Europe). Firstly, instances with missing cases in these columns are removed from the dataset, given their low percentage, and because it was meaningful to retain the actual responses. Such an approach, called listwise deletion, is considered to be effective when the proportion of missing values is low (Zhang, 2016). Responses for the questions required to calculate the target value are coded on a scale with values from 1 to 4, where higher values indicate lower willingness to admit immigrants. Using a threshold of 2.5, observations with a higher mean across these 3 variables are coded as 1 to indicate a negative stance, while those with a lower mean are coded as 0 to reflect a positive immigration attitude (Erhard & Heiberger, 2023). Subsequently, the variables used for calculating this index are removed from the dataset to prevent data leakage. Class proportions of the target variable are presented in Figure 1. As illustrated, it is slightly imbalanced, with roughly 61% of instances holding positive attitudes towards immigration.

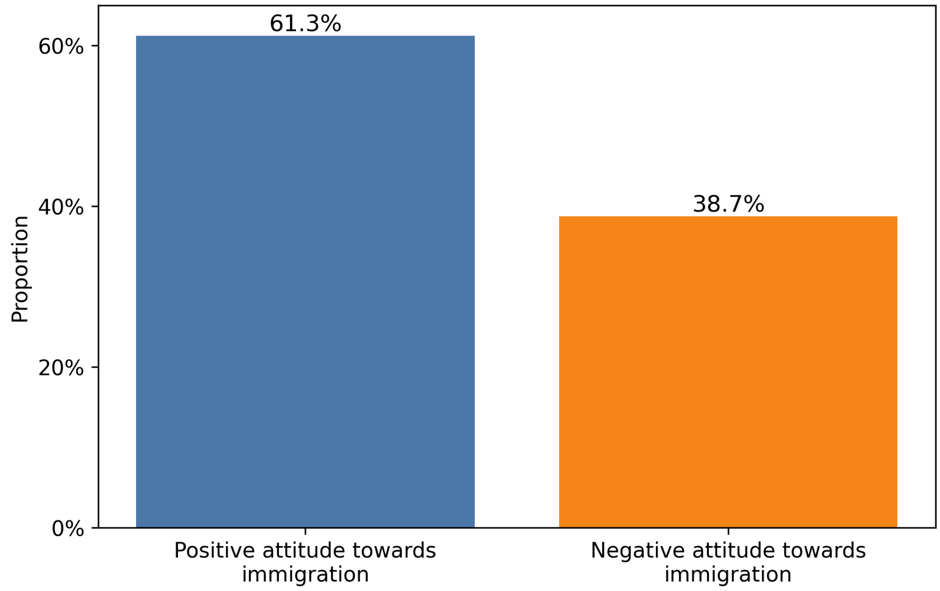


Figure 1: Target distribution

4.2.3 *Dataset split*

The dataset is split into training (80%) and test (20%). Firstly, it was done on a country level with stratification to preserve class proportions within one country. The resulting country-specific training and test sets were then combined into European training and test sets, so that the dataset and each individual country retained approximately the same class balance in the training and test data.

This procedure was performed before subsequent pre-processing steps to prevent data leakage into the training process.

4.2.4 *Outliers detection*

Considering the characteristics of the dataset, the occurrence of outliers is very unlikely, as most variables are on a Likert scale, limited to a fixed range. Thus, their presence will be examined in variables without fixed ranges, namely, respondents' age and household size, using the Interquartile Range method. No outliers were detected in the 'age' variable. On the other hand, the size of households with values of 5 or more were revealed as outliers. However, in practice, five or more people can live in the same household, thus, rows with these values, even though flagged as outliers, were retained for further analysis.

4.2.5 *Missing values*

The dataset consists of numerical and categorical data types (See Table 13). The missing values for the first were calculated using the median method, while for the latter, the mode method was utilized. Both of these methods are commonly applied in the literature due to their clarity (Jerez et al., 2010). Importantly, both imputations are performed separately within each country, that is, missing values are replaced using medians and modes computed on the subsample of respondents from the same country.

4.2.6 *Features creation*

This study follows the method of Erhard and Heiberger (2023) in calculating two predictors, namely self-transcendence and conservation. They are calculated based on responses to questions regarding the importance of personal values (PVQ) (See Appendix A, Table 14). First, all these items are re-coded so that higher values indicate greater importance of a specific value. Next, scores for the five basic value types (universalism, benevolence, security, conformity, and tradition) are computed and person mean-centered by subtracting each respondent's average across all 21 PVQ items to control for individual response style. Finally, the centered basic values are averaged into two indices: Conservation and Self-Transcendence. Positive scores mean that this value orientation is more important than usual for a given respondent, while negative scores mean it is less important. Subsequently, the columns required for the construction of these features are removed from the dataset to prevent data leakage.

Furthermore, drawing from Arcila Calderón et al. (2022), the feature 'hhmmb' reflecting the size of the respondent's household is transformed. Instead of representing a continuous number, it is categorized into four groups: living alone, two-person household, three to five persons in the household, and six or more.

4.2.7 *Exploratory data analysis*

A crucial part of any research analysis lies in an exploratory investigation of the data (Salciccioli et al., 2016). Exploratory data analysis (EDA) aims to provide researchers with an essential understanding of the dataset and consists of techniques such as data visualization (Patel et al., 2022).

Firstly, the potential impact of categorical variables is analysed using normalized stacked bar charts (See Figure 2). The importance of the country is particularly noticeable, as individuals from Nordic and Western countries tend to hold less negative immigration attitudes. Education level also follows a clear pattern, as more educated people are more positively oriented towards immigration.

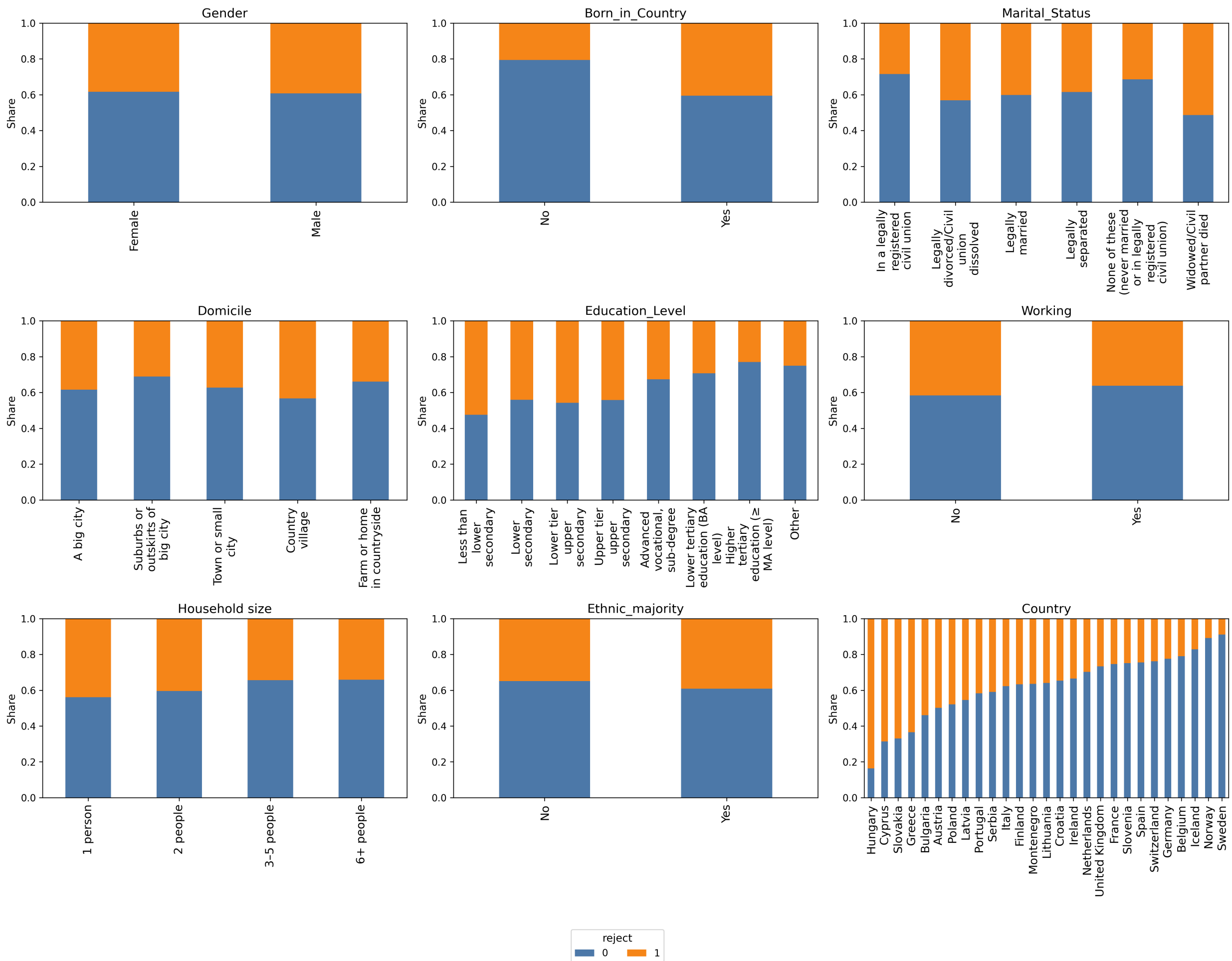


Figure 2: Categorical features

Subsequently, the influence of the numerical features on the target variable is examined by analysing the horizontal bar chart (See Figure 3). It indicates the strongest correlation with negative immigration attitudes for self-placement on the left-right scale, followed by age and religiosity level. All the other variables tend to influence these attitudes positively. As such, self-transcendence is the strongest correlate, while among the trust-related variables, trust in foreign institutions emerges as the most influential.

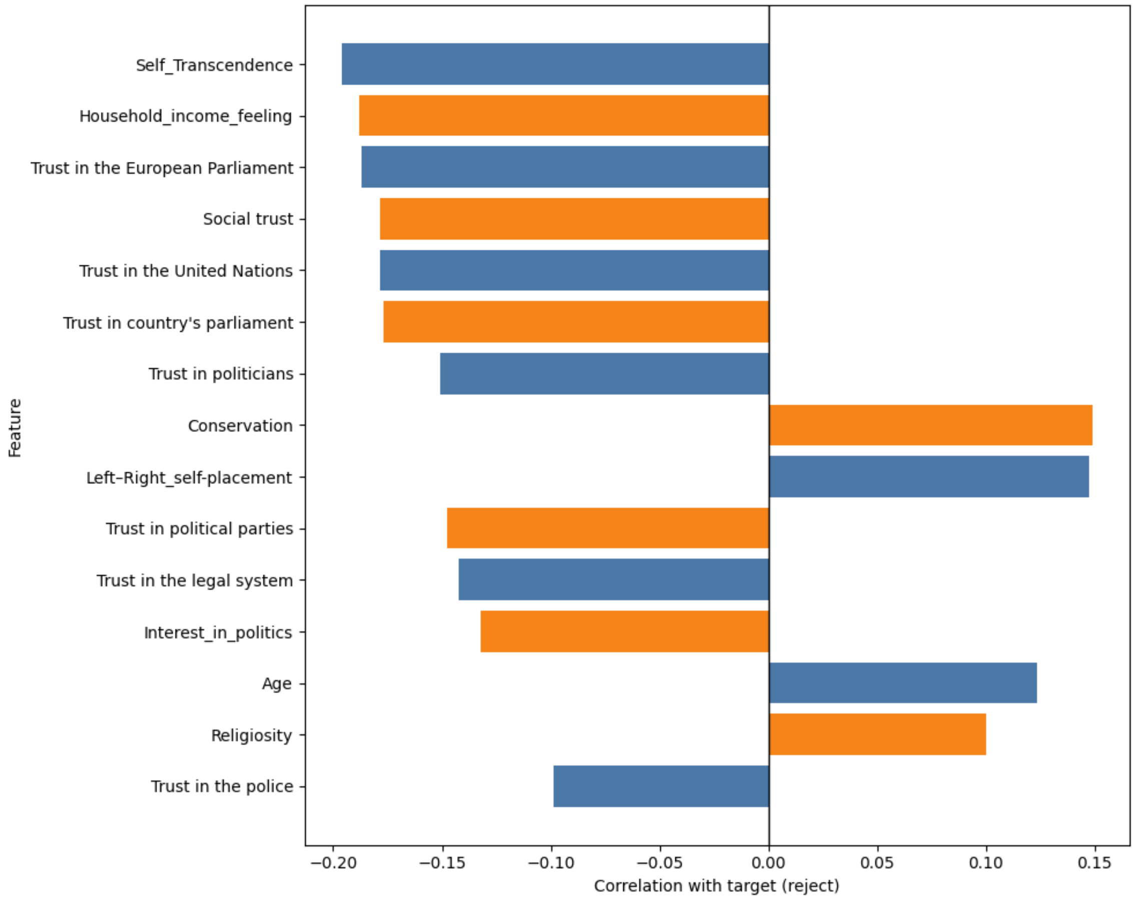


Figure 3: Correlations with target variable

4.2.8 *Encoding categorical variables*

Machine learning models in general operate on numerical input, thus, categorical variables need to be properly encoded to ensure compatibility (Gnat, 2021). It is worth reminding that the raw dataset was already label-encoded. However, this method implicitly imposes an order among categories, which is not always appropriate. Following this, the decision was made to use One-Hot encoding for encoding categorical features in this research.

4.2.9 *Data scaling*

Some machine learning frameworks, including tree-based algorithms, generally do not require this data scaling to work properly (Garcia-Carretero et al., 2021). This contrasts with the models like Logistic regression, which may be sensitive to the input scales (Cao et al., 2021). Feature scaling can also increase the convergence time of algorithms based on gradient descent (Wan, 2019). Because the present research employs models from both of these groups, the feature scaling is executed to maximize their performance. All numerical features were normalized using the Min-Max scaling approach. This method rescales every observation to lie within the range from 0 to 1.

4.3 *Models training*

4.3.1 *Nested cross-validation*

As previously stated, the dataset has been split such that 80% of the data is allocated for the training set. As such, model tuning and selection are performed with Nested Cross-Validation (NCV) using 5 outer and 3 inner folds. The inner folds are used to select the optimal hyperparameters, which are later evaluated on the corresponding outer folds. This approach may provide a more reliable estimate of model performance compared to flat cross-validation (Wainer & Cawley, 2021). To further reduce randomness, the entire Nested Cross-Validation procedure is repeated across five different random seeds (0, 7, 42, 123, 999). As a result, each model yields 25 outer-fold F1 scores and ROC-AUCs, which are averaged and reported with their standard deviations. The best-performing model, along with the hyperparameter configuration identified across the entire Nested Cross-Validation, is then trained on the full training set and evaluated on the remaining 20% of the data used as the test set.

4.3.2 *Hyperparameters optimization*

In this research, hyperparameter tuning was performed using the Optuna framework (Akiba et al., 2019). Optuna uses the Tree-structured Parzen Estimator (TPE) algorithm to dynamically refine the search space by learning from previous trials. Together with the pruning method that removes less promising trials during the optimization, computation time can be significantly reduced (Akiba et al., 2019). These features make Optuna more efficient than Random Search or Grid Search (Shekhar et al., 2021). In this study, Optuna was run within the inner loop of each outer fold in the Nested Cross-Validation, and for each inner loop, it performed 50 hyperparameter trials for a given model, optimizing ROC-AUC. The hyperparameter search spaces for each of the employed algorithms are described in the subsequent section.

4.4 *Models*

4.4.1 *Logistic regression*

Logistic Regression (LR) is a common choice for a binary classification, due to its proficiency and straightforwardness (Abdumalikov et al., 2024; Zaidi & Al Luhayb, 2023). Nevertheless, this method assumes the linear relationships between the predictors and the dependent variable (J. K. Harris, 2021), which might not always be the case in reality. The selected hyperparameters of this model, along with their ranges, are presented in Table 1. Particularly, the 'class_weight' is set to 'balanced', given the slight imbalance of the target value. Penalty consists of 'l2' and None, based on the previous research (Arcila Calderón et al., 2022; Erhard & Heiberger, 2023), and max_iter reflects the maximum number of iterations for the model to converge. The C stands for the regularisation strength, and therefore is tested only on the settings with L2 penalty.

Table 1: Hyperparameter search space for Logistic Regression

Hyperparameter	Values
solver	liblinear, lbfgs, saga
penalty	L1, L2, none
C	$[10^{-4}, 100]$ (Log Scale)
max_iter	(100, 5000)
class_weight	balanced

4.4.2 *Random forest*

Random Forest Classifier is an ensemble algorithm that averages the predictions from multiple decision trees built on random samples and features to mitigate overfitting and enhance predictive performance (Pedregosa et al., 2011; Salman et al., 2024). This algorithm is also acknowledged for its effectiveness in binary classification, sometimes proving to be more efficient than Logistic Regression (Boulesteix et al., 2012). Table 2 displays the ranges for the optimisation of the selected hyperparameters. In this case, 'n_estimators' reflects the number of decision trees used by the algorithm, and max_depth reflects the maximum depth of each of them. Together, they control the size and complexity of the forest. Parameter 'class_weight' is set to 'balanced to account for the class imbalance.

Table 2: Hyperparameter search space for Random Forest

Hyperparameter	Search space
n_estimators	(100, 1500)
max_depth	(3, 30)
min_samples_split	(2, 40)
min_samples_leaf	(1, 20)

4.4.3 *XGBoost*

Following the successful application of gradient boosting to the prediction of individual environmental attitudes using ESS data (Yektansani et al., 2024), XGBoost, along with LightGBM and CatBoost, was selected for this thesis. XGBoost (Extreme Gradient Boosting), introduced by T. Chen et al. (2015), similarly to Random Forest, is a tree-based ensemble method, but uses gradient boosting in place of bagging. In contrast to Random Forest, which trains many trees independently in parallel, XGBoost builds a large number of shallow trees in a sequence, with each new tree focusing on the remaining errors from the previous ones (Fatima et al., 2023). Furthermore, the gradient boosting approach is able to capture subtle patterns in the data (Yektansani et al., 2024), which makes it an appropriate choice for this thesis. The selected hyperparameters and their ranges are visible in Table 3. In particular, 'scale_pos_weight' is set to the ratio of negative to positive instances to account for the class imbalance, which is also applied in CatBoost. The parameter 'min_child_weight' is set to 1, to enable the algorithm to detect more intricate patterns, and the same choice is applied in both LightGBM and CatBoost. Other ranges are designed so that all the boosting algorithms can be tuned over a wide hyperparameter range while keeping the models efficient.

Table 3: Hyperparameter search space for XGBoost

Hyperparameter	Search space
n_estimators	(100, 1500)
learning_rate	(0.01, 0.5) (Log Scale)
max_depth	(3, 10)
subsample	(0.6, 1.0)
colsample_bytree	(0.6, 1.0)
gamma	(0.0, 5.0)
reg_alpha	(10^{-3}, 10) (Log scale)
reg_lambda	(10^{-3}, 10) (Log scale)
min_child_weight	1.0
scale_pos_weight	neg/pos

4.4.4 *LightGBM*

Another algorithm following the gradient boosting principle is LightGBM (Ke et al., 2017). LightGBM is particularly suited for large, high-dimensional datasets due to the novel optimisation techniques, Gradient-based One-Sided Sampling (GOSS) and Exclusive Feature Bundling (EFB). GOSS accelerates the training process by concentrating on instances with higher gradients. EFB decreases the complexity of training by merging similar features, which further quickens this process. These techniques help the LightGBM to decrease computation time, without compromising on accuracy (Ke et al., 2017). The search space of hyperparameters explored to optimize the performance of LightGBM is presented in Table 4. Due to the class imbalance, 'class_weight' is set to 'balanced'.

Table 4: Hyperparameter search space for LightGBM

Hyperparameter	Search space
n_estimators	(100, 1500)
learning_rate	(0.01, 0.5) (Log scale)
max_depth	(4, 20)
min_child_samples	(1, 30)
subsample	(0.5, 1.0)
colsample_bytree	(0.5, 1.0)
reg_alpha	(10^{-4}, 1.0) (Log scale)
reg_lambda	(10^{-4}, 1.0) (Log scale)
min_child_weight	1.0
class_weight	balanced

4.4.5 *CatBoost*

The gradient boosting models considered in this research conclude with CatBoost (Prokhorenkova et al., 2018). What distinguishes this model from XGBoost and LightGBM is that CatBoost places particular focus on reducing target leakage and prediction shift with the Ordered Boosting method. Additionally, Catboost is based on oblivious trees, which prevent overfitting, and is also able to automatically handle categorical features, due to Ordered Target Statistics technique (Prokhorenkova et al., 2018). The selected hyperparameter ranges for CatBoost optimization are presented in Table 5. Due to the encountered high computation time of CatBoost with its native handling of categorical features, the decision was made to use one-hot encoding for this model as well.

Table 5: Hyperparameter search space for CatBoost

Hyperparameter	Search space
iterations	(100, 1500)
learning_rate	(0.01, 0.5) (Log scale)
depth	(4, 10)
l2_leaf_reg	(10^{-4}, 1.0) (Log scale)
subsample	(0.5, 1.0)
colsample_bylevel	(0.5, 1.0)
scale_pos_weight	neg/pos

4.5 *Evaluation*

Because the target variable in this research is slightly imbalanced, the F1-score and ROC-AUC are used as main evaluation metrics. Unlike accuracy, the F1-score provides more informative results when the target classes are imbalanced, as it accounts for both precision and recall (Niu et al., 2025). In this study, the F1-score is computed for the class representing immigration opponents, since correctly identifying this group is of primary interest. The advantage of ROC-AUC lies in being a threshold-independent metric that measures how well the model distinguishes between the classes (Narkhede, 2018). It makes ROC-AUC a recommended choice in cases when the same problem exhibits different class imbalances dependent on context (Richardson et al., 2024). In this study, models are also trained separately for each country, and the class balance differs strongly between countries, which further supports the use of ROC-AUC.

4.6 *Feature importance*

To examine the predictive importance of the features, the SHAP (Shapley Additive exPlanations) method is used (Lundberg & Lee, 2017). The approach is grounded in game theory (Shapley et al., 1953) and assigns each feature its contribution to a particular prediction. It decomposes each prediction into a baseline equal to the model's expected output and a sum of feature contributions expressed as SHAP values. Each SHAP value captures the average marginal contribution of a feature across subsets of the other features. Summing the absolute SHAP values of a feature provides a global ranking of importance across the dataset. Positive SHAP values indicate increasing the predicted value of the model, while negative SHAP values decrease it (Lundberg & Lee, 2017; Molnar et al., 2020).

4.7 *Experimental procedure*

After preprocessing, several algorithms are compared with NCV on the training data without trust variables, and the best one is tuned with NCV to form the Europe-wide model without trust. The same algorithm is then tuned with NCV on the data with trust variables to form the Europe-wide trust model, which is first used for subgroup error analysis; finally, both models are replicated on country-level subsets and compared across countries. The summarised data science pipeline is presented in Figure 4.

5 RESULTS

5.1 *Model selection using Nested Cross-Validation (no trust-related variables)*

The mean F1 Scores and ROC-AUC of every model and their standard deviations, obtained after performing the Nested Cross-validation with 5 different random states, are presented in Table 6, and the optimal set of hyperparameters for each model is visible in Table 7. As presented, all the models achieved very similar results. Logistic regression with the mean F1 of 0.635 and a standard deviation of 0.007 is the relatively worst-performing model. However, due to the marginal differences in the models' performances, a formal statistical check is required to determine their significance.

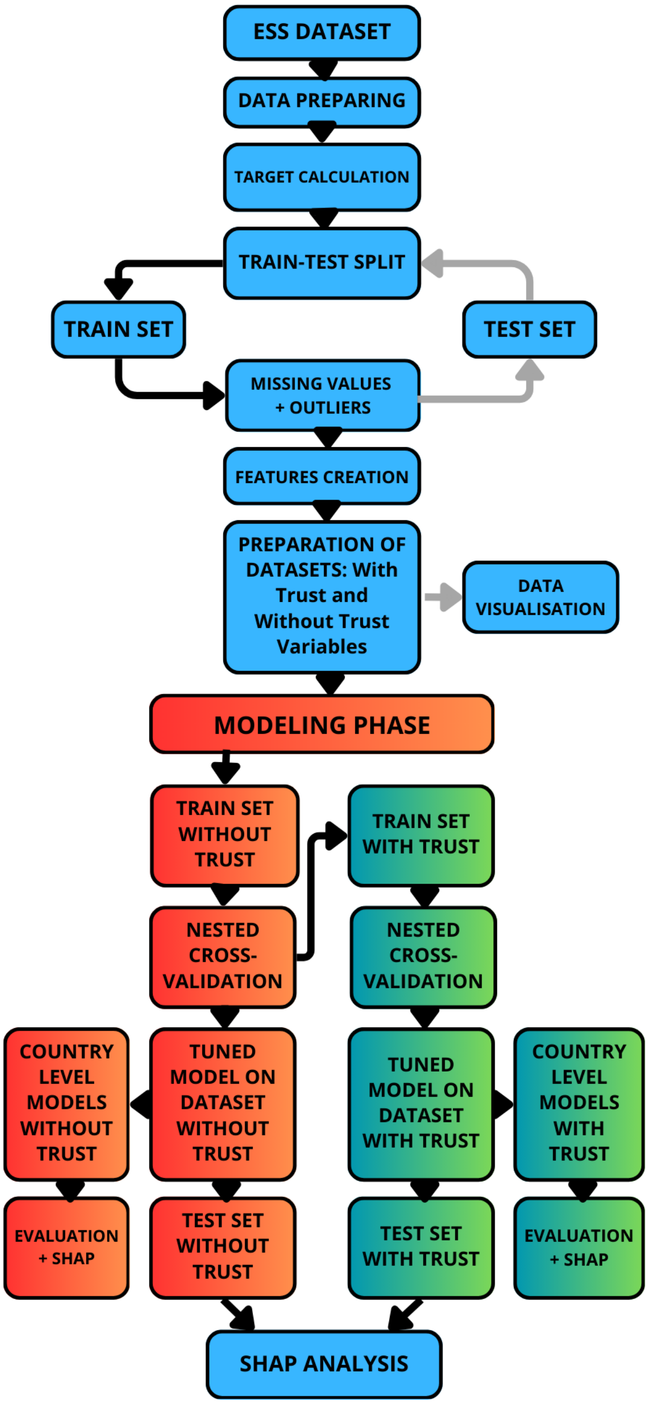


Figure 4: Flowchart of the experimental setup

Table 6: Mean F_1 and ROC-AUC scores of the evaluated models obtained with nested cross-validation.

Model	F_1 (mean ± std. dev.)	ROC-AUC (mean ± std. dev.)
Logistic Regression	0.6348 ± 0.0067	0.7632 ± 0.0051
Random Forest	0.6354 ± 0.0083	0.7757 ± 0.0049
XGBoost	0.6495 ± 0.0055	0.7805 ± 0.0042
LightGBM	0.6485 ± 0.0052	0.7794 ± 0.0046
CatBoost	0.649 ± 0.0049	0.7795 ± 0.0049

Table 7: Best hyperparameter configurations of the evaluated models obtained during nested cross-validation.

Model	Selected hyperparameters
Logistic Regression	penalty = `l2`, C = 1.0, class_weight = `balanced`
Random Forest	n_estimators = 1122, max_depth = 30, min_samples_split = 31, min_samples_leaf = 1, class_weight = `balanced`
XGBoost	n_estimators = 1264, learning_rate = 0.0101, max_depth = 10, subsample = 0.8851, colsample_bytree = 0.6072, gamma = 4.1231, reg_alpha = 0.0012, reg_lambda = 0.0221, min_child_weight = 1.0, scale_pos_weight = 1.5804
LightGBM	n_estimators = 811, learning_rate = 0.0129, max_depth = 7, min_child_samples = 16, subsample = 0.6592, colsample_bytree = 0.5027, reg_alpha = 0.0510, reg_lambda = 0.1541, class_weight = balanced, min_child_weight = 1.0
CatBoost	iterations = 1018, learning_rate = 0.0485, depth = 4, l2_leaf_reg = 0.3787, subsample = 0.7683, colsample_bylevel = 0.7150, scale_pos_weight = 1.5804

To determine whether differences in the models' results are statistically significant, the Friedman test is conducted on the F1 Scores obtained for all models across all outer folds and random states. The Friedman test is used to analyze the significance of differences in the performances of more than two models (Demšar, 2006). While primarily it is dedicated to assessing these differences on multiple datasets, it is commonly used to compare algorithms on different folds of a single dataset (Martinović et al., 2025; Rainio et al., 2024). Compared to ANOVA, it is a non-parametric test, which does not assume normal distributions and independence between groups. It makes it a recommended choice for comparing algorithms in machine learning research (Demšar, 2006; Rainio et al., 2024). The results of the Friedman test are presented in Table 8. Based on them, we can reject the null hypothesis, which means that at least one of the models performs significantly differently from the others.

Table 8: Friedman test comparing F1 scores of all evaluated models.

Test	Test statistic	p-value	Decision ($\alpha = 0.05$)
Friedman	78.11	4.37×10^{-16}	Reject H_0

To further examine which specific models differ from each other, a post-hoc Wilcoxon signed-rank tests with Bonferroni correction are applied to account for errors resulting from multiple pairwise comparisons (Demšar, 2006). According to the results, F1 Scores of logistic regression and random forest are significantly different than those of gradient boosting models (See Appendix A, Table 16). In contrast, the differences between the results of gradient boosting models are not statistically significant. When several models are statistically comparable, a common practice in selecting one is to prioritize computational efficiency (Ali et al., 2017; Cioch et al., 2025; Hall & Rasheed, 2025). The training time of Catboost was 10 hours, which was significantly longer than the training time of LightGBM and XGBoost. The training time of XGBoost took over 2 times longer than the training of LightGBM, thus LightGBM was chosen for further analysis. Additionally, in reality, the number of observations can be much higher, and in such settings, LightGBM is documented to scale more efficiently than XGBoost (Ke et al., 2017).

5.2 *Test set results - LightGBM without trust-related data*

Subsequently, the selected LightGBM was trained on the entire training set with the optimal set of hyperparameters and evaluated on the previously unseen test set. The results are presented in Table 9. As presented, both F1 and ROC-AUC scores are almost the same compared to their means obtained after NCV. It means that the algorithm did not encounter overfitting and generalized as expected. While the ROC-AUC and F1-score for class 1 remain the primary evaluation metrics, others were calculated to facilitate comparison with the existing research. In Figure 5, the confusion matrix shows that misclassifying respondents with positive immigration attitudes as negative is more common than the opposite error. Because this study is mainly concerned with identifying respondents with negative attitudes, this trade-off is reasonable.

Table 9: Performance of the LightGBM model on the test set

Class	Precision	Recall	F_1
Positive attitudes (class = 0)	0.7925	0.6933	0.7396
Negative attitudes (class = 1)	0.5950	0.7129	0.6486
Accuracy: 0.7009			
Macro F_1: 0.6941			
ROC-AUC: 0.7783			

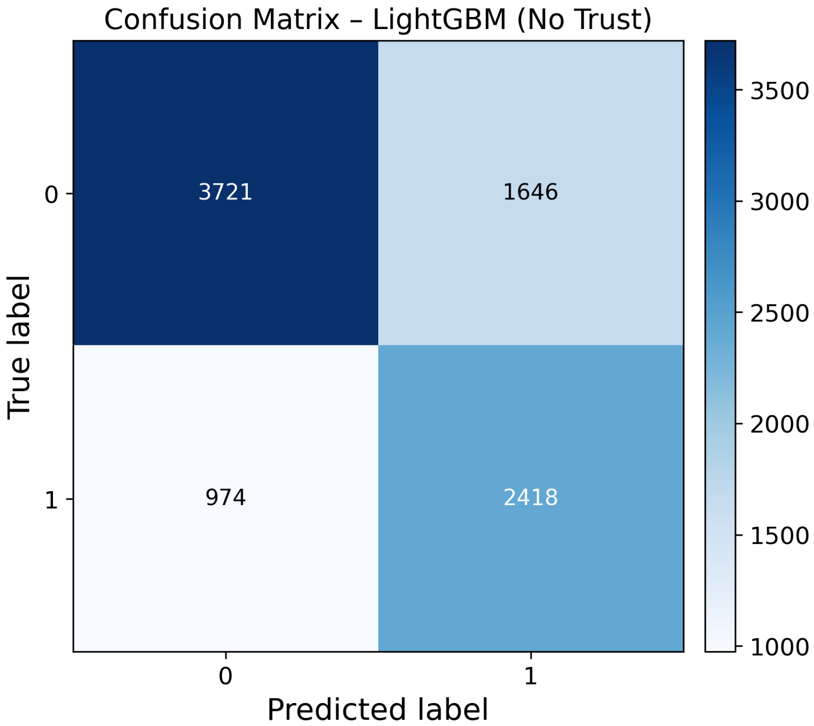


Figure 5: Confusion matrix of the LightGBM model without trust-related data

5.3 *Feature importance of the LightGBM model without trust-related data*

Figure 6 shows the SHAP values for the LightGBM without trust-related variables. The most important drivers are Self-Transcendence and Left-Right self-placement, followed by Household income feeling and Age. Important to note is the significant role of a country, as 8 out of the 20 most important features are country dummies, which means that the model relies heavily on where respondents live.

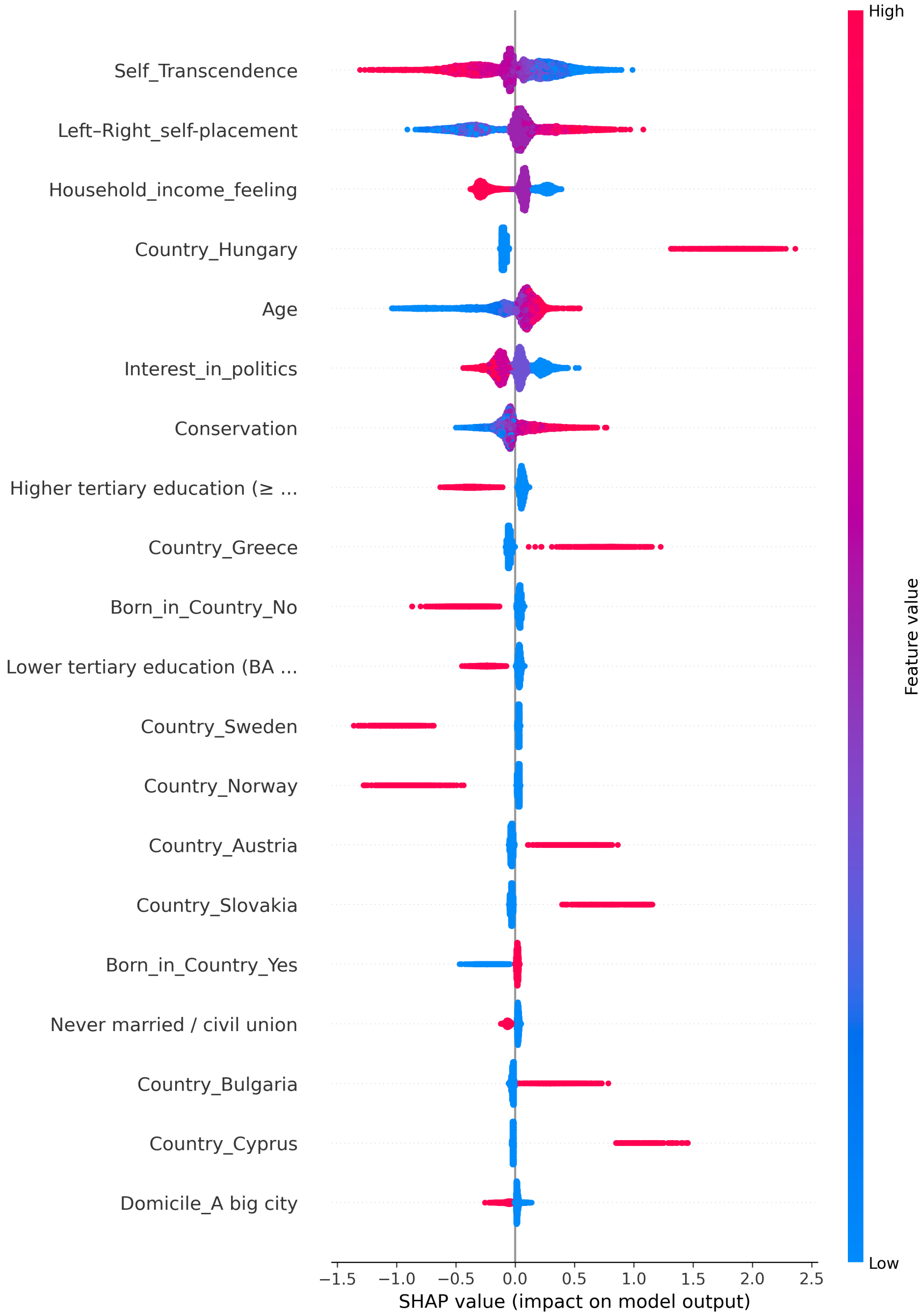


Figure 6: SHAP values for LightGBM model without trust-related data

5.4 *Analysis of the effect of trust-related data*

LightGBM was retrained with the addition of trust-related variables. It was tuned and evaluated using the same nested cross-validation procedure with the same 5 random states as the previous models to ensure results' comparability. As presented in Table 10, institutional and social trust variables slightly increased the model's results. This difference was further validated with the Wilcoxon signed-rank test (p-value < 0.05), performed on the F1-scores obtained from every outer fold and for each random seed, which revealed that adding these features significantly improves model performance. The optimal hyperparameters of the trust-extended model are summarised in Table 11.

Table 10: Nested cross-validation results for LightGBM with and without trust-related variables.

Model	F_1 (mean ± std. dev.)	ROC-AUC (mean ± std. dev.)
LightGBM without trust	0.6485 ± 0.0052	0.7794 ± 0.0046
LightGBM with trust	0.6592 ± 0.0065	0.7902 ± 0.0047

Table 11: Optimal hyperparameters for the trust-extended LightGBM model

Optimal hyperparameters
n_estimators: 1210, learning_rate: 0.0153, max_depth: 19, min_child_samples: 27, subsample: 0.5571, colsample_bytree: 0.5112, reg_alpha: 0.0004, reg_lambda: 0.0001, class_weight: balanced

Afterwards, the extended LightGBM with trust-related data was retrained on the training set and evaluated on the independent test set (See Table 12). The addition of trust-related variables improved the model's performance on the test set compared to the model without trust data, increasing the F_1 and the ROC-AUC by 0.013 and 0.0163, respectively. This mirrors the performance gains observed in the nested cross-validation results. The confusion matrix in Figure 7 shows that the model with trust-related variables correctly classifies 76 more respondents with negative attitudes (class 1) and 35 more respondents with positive attitudes (class 0).

Table 12: Classification metrics for LightGBM models with and without trust-related variables (test set)

Class	Without trust variables Precision	Recall	F_1	With trust variables Precision	Recall	F_1
Positive attitudes (class = 0)	0.7925	0.6933	0.7396	0.8017	0.7075	0.7517
Negative attitudes (class = 1)	0.5950	0.7129	0.6486	0.6097	0.7232	0.6616
Accuracy		0.7009			0.7136	
Macro F1		0.6941			0.7066	
ROC-AUC		0.7783			0.7946	

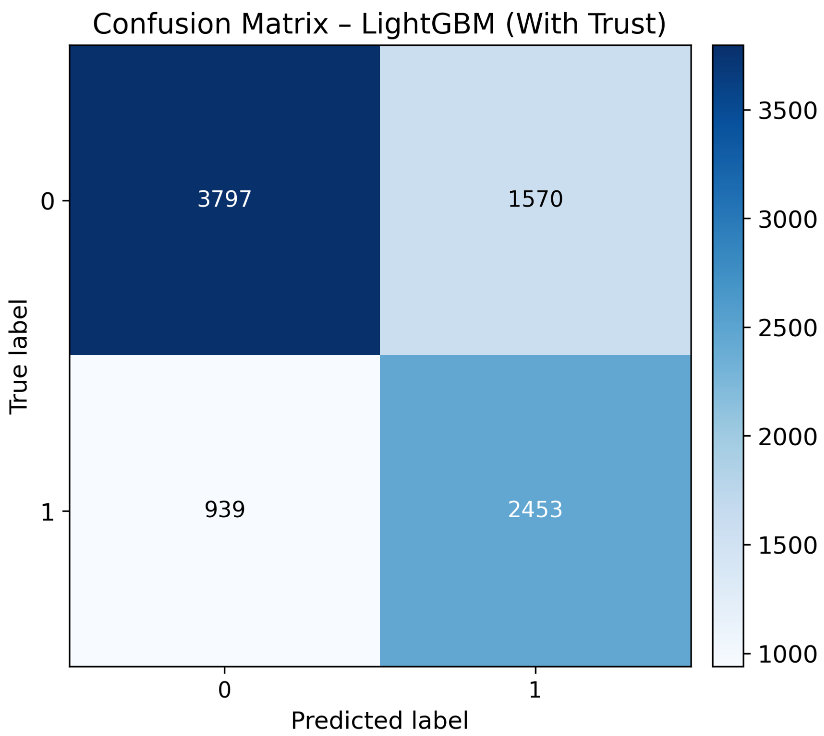


Figure 7: Enter Caption

5.5 *Feature importance of the LightGBM model with trust data*

Figure 8 shows the influence of the trust-related variables on the structure of model's feature importance. As presented, several of them became meaningful predictors. Particularly trust in other people and European Parliament. Compared to the model without trust data, several features shift in the ranking, indicating that predictive weight is redistributed once trust information is included. The contribution of country dummies appears smaller, but country still remains an important predictor.

Figure 8: SHAP with trust

5.6 *Error analysis in subgroups of the LightGBM model with trust data*

LightGBM with trust-related data underwent error analysis across different demographic groups to assess its fairness. Differences between F1 scores are particularly evident across countries (See Table 17, Appendix B), which is, however, understandable, due to significant differences between class proportions (See 15, Appendix A).

Across other subgroups, ROC-AUC results remained rather comparable, especially across gender and ethnic groups (See Table 18, Appendix B). Furthermore, slight disproportions in ROC-AUC are visible, mainly among respondents born locally and abroad. In terms of F1 scores, the model performed rather consistently across subgroups. Again, there is a more visible gap between respondents born in the country and those born abroad. F1 scores were also slightly worse for the youngest individuals and for those living in the suburbs of big cities and in the countryside.

5.7 *Cross-country comparison of model performance and feature importance*

Considering the strong role of countries in the European model and clear cross-country differences in errors, LightGBM models were trained separately for each country, both without and with trust-related variables, to better compare feature importance patterns and assess the impact of trust on immigration attitudes. As per Figure 9, adding trust-related variables generally improves model performance across countries, especially where the baseline ROC-AUC was relatively low, like in Sweden. There is also a visible significant improvement of the results in Montgenero. However, the effect is not uniform: in some countries, like the United Kingdom or Latvia, ROC-AUC slightly decreases. It suggests that the role of trust is context-dependent and can have different effects across countries, often improving model performance, but in some cases introducing additional noise.

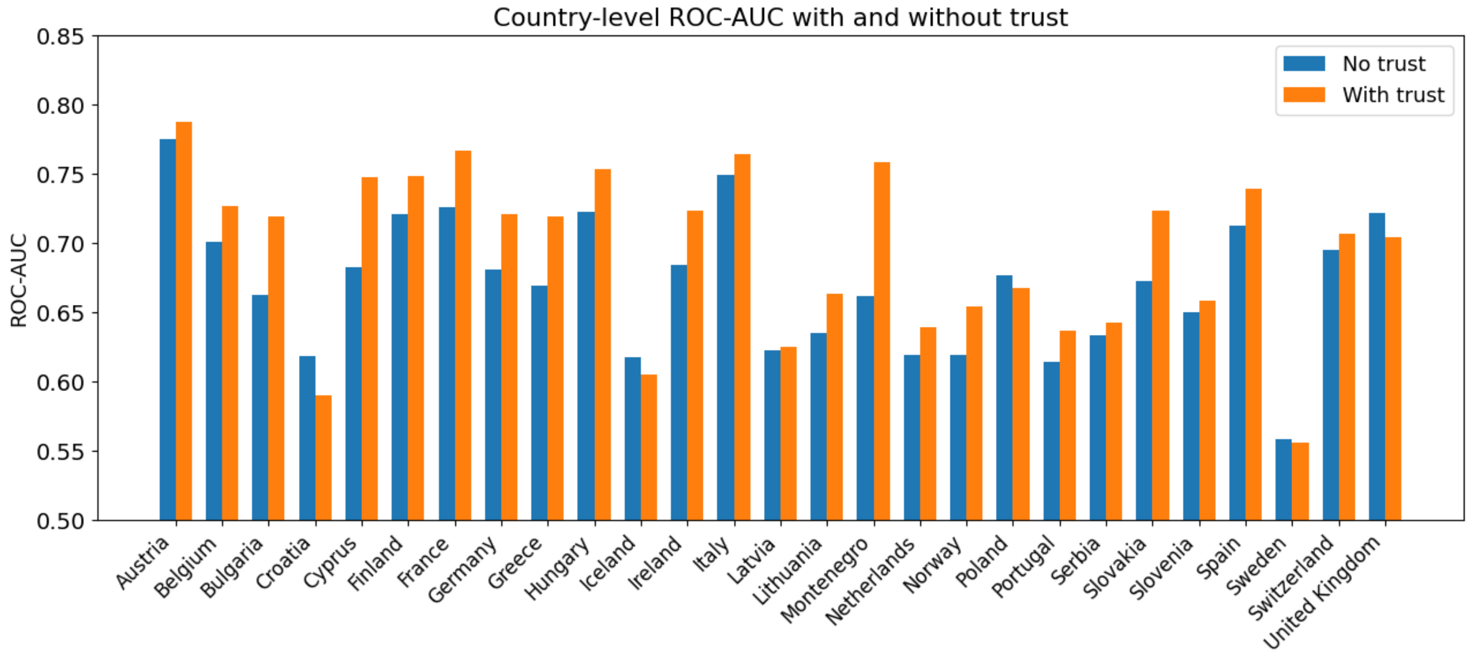
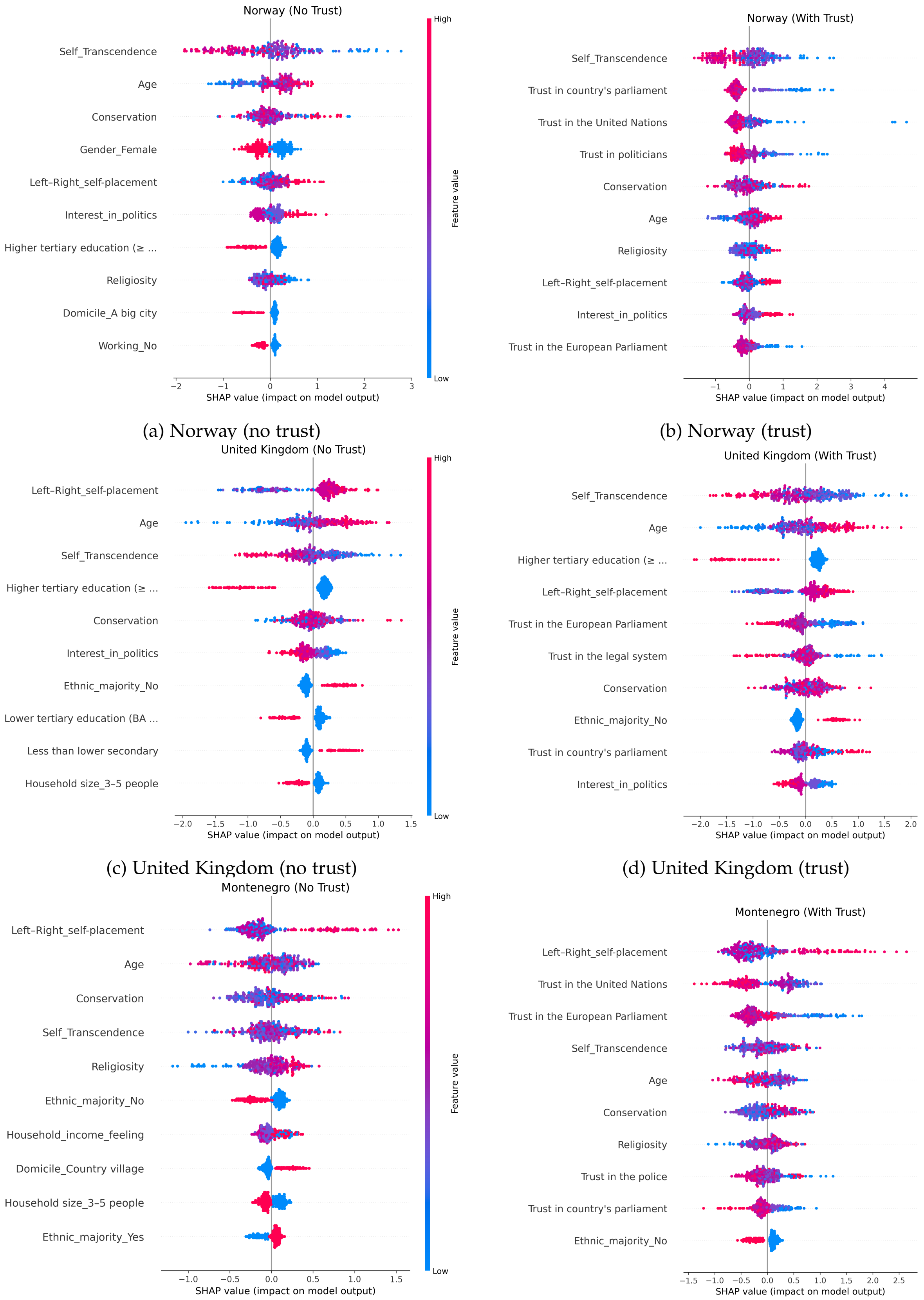


Figure 9: ROC-AUC for country-level models with and without trust variables

Figure 10 presents the comparison between SHAP values for models estimated without and with trust variables across countries. The United Kingdom and Montenegro were selected, as they reacted differently to the inclusion of trust. Norway was selected because it has a strong class imbalance, yet it performed better than Sweden, which shows a similar imbalance (See Figure 9). Across the three countries without trust variables, ideological and value-related features are central in all models. Interestingly, gender in Norway plays an important role compared to other countries. Worth to note is the unevenly distributed pattern of SHAP dots, especially evident in Montenegro in the first five features, suggesting that they drive predictions in an ambiguous way.

After adding trust, these patterns diverge. In Norway and Montenegro, trust seems to push predictions towards positive immigration attitudes. However, Norwegians, who are more trusting of the legal system of their country, hold more negative stances. In the United Kingdom, trust features, especially trust in the legal system, look much more scattered around zero and lack a stable trend. This could partially explain the drop in performance and suggests that, in this context, trust introduces noise instead of providing a coherent signal.

The comparisons between feature importance with and without trust-related data for the remaining countries are presented in Appendix C. In most countries, higher levels of trust are associated with more positive attitudes towards immigration, but this depends strongly on both the type of trust and the country. For example, in Belgium, almost all trust measures are linked to more positive immigration attitudes, while in Italy, people who trust politicians more tend to hold more negative views.



(a) Norway (no trust)

(b) Norway (trust)

(c) United Kingdom (no trust)

(d) United Kingdom (trust)

(e) Montenegro (no trust)

(f) Montenegro (trust)

Figure 10: Comparison of SHAP values for selected countries, for models estimated without and with trust-related variables.

6 DISCUSSION

6.1 Resarch goal

The main goal of this research was to identify the best-performing machine learning model for predicting individual immigration attitudes in Europe and their key determinants. It also aimed to examine how adding institutional and social trust variables affects predictive performance and the feature importance. Given the strong role of country-level differences, the study includes a separate assessment of model performance and main predictors within each country.

SQ1 *Among the evaluated models (Logistic Regression, Random Forest, LightGBM, XGBoost, CatBoost), which of them achieves the best results according to F1-score and ROC-AUC?*

The NCV results suggest (See Table 6) that all models employed in this study predict individual immigration attitudes very similarly, with F1 scores between 0.63 and 0.65 and ROC-AUC values between 0.76 and 0.78. However, executed statistics tests indicate that LR and RF performed significantly worse than boosting methods. The superiority of more advanced ML models indicates that immigration attitudes may be more intricate in nature and consist of nonlinear patterns and interactions. Nevertheless, the absolute performance gains of the boosting models are relatively small, suggesting that these patterns are not extremely complex and that even advanced models can only improve predictions to a limited extent.

Findings of this study contrast with the research of Erhard and Heiberger (2023), who found Ridge Regression to be the best model (F1 = 0.735, Accuracy = 0.699), followed by Random Forest (F1 = 0.728, Accuracy = 0.690) and Logistic Regression (F1 = 0.731, Accuracy = 0.689), with F1 scores reported for positive immigration attitudes. In this thesis, LightGBM was selected for evaluation on the test set, as among boosting methods, it had the shortest computation time. For the same class (positive immigration attitudes), it achieved slightly better results (Accuracy = 0.70, F1 = 0.7396) than the models reported by Erhard and Heiberger (2023)

LightGBM also outperformed the scores reported for Logistic Regression (macro F1 = 0.68, ROC-AUC = 0,68), SVM (macro F1 = 0.68, ROC-AUC = 0,68), and Random Forest (macro F1 = 0.64, ROC-AUC = 0.64), in the study by Arcila Calderón et al. (2022), achieving a macro F1 of 0.6941 and a ROC-AUC of 0.7783. Overall, comparisons show that advanced machine learning methods provide clear but still modest gains, and that immigration attitudes among Europeans remain difficult to predict. XGBoost, LightGBM, and CatBoost, which are applied for the first time in this specific context, offer a strong performance benchmark. Results indicate that they can capture more of the patterns present in the complex domain of immigration attitudes than the more traditional methods. Nevertheless, logistic regression, which in this study performed only slightly worse and has been identified as a leading approach in earlier research, remains attractive due to its simplicity and ease of interpretation.

SQ2 *Which features contribute most to predicting immigration attitudes in the global baseline model without trust variables, according to SHAP values?*

As per Figure 6, the most important predictors of individual immigration attitudes are attitudinal factors, like self-transcendence and self-placement on the left-right scale. The latter, together with subjective household income, is positioned significantly higher compared to the findings of Erhard and Heiberger (2023). In contrast to the study of Arcila Calderón et al. (2022), household size does not rank among the 20 most contributing factors. However, respondents' countries seem to carry more importance, which aligns with Arcila Calderón et al. (2022). Being born in a country seems important as well, which means that individuals who are not native to the country and thus share a similar non-native position with immigrants tend to hold more positive attitudes towards them. Overall, this pattern suggests that immigration attitudes are shaped mainly by deeper value orientations, political worldviews, and perceived economic security, with age and education also remaining important predictors in line with previous research. The strong presence of country effects further indicates that national political climates and public discourse reinforce these individual-level patterns.

SQ3 *To what extent do trust-related features contribute to predicting individual immigration attitudes in the global model?*

As depicted in Table 10, the inclusion of social and institutional trust-related variables resulted in a modest improvement in LightGBM's NCV performance of around 0.01, which was shown to be statistically significant according to the Wilcoxon signed-rank test. The same effect was observed in its performance on the test set, as F1 score and ROC-AUC increased by around 0.01, indicating a slight, but stable enhancement.

According to SHAP analysis, the majority of additional trust-related variables were found among the 20 most important predictors (See Figure 6). Especially important seems to be trust in the European Parliament. It may indicate that people who place confidence in the European Union assume that it is capable of dealing with immigration issues in an effective manner, which makes them more positive towards immigration. Worth noting is also the significant role of trust in other people, which aligns with Goubin et al. (2022) and Mitchell (2021). It indicates that people who generally trust others are also more willing to extend this trust to immigrants, suggesting that strengthening social cohesion may support a smoother integration process of immigrants.

SQ3 *How biased is the best-performing global model across demographic subgroups?*

The best-performing model from the previous steps, LightGBM extended with the trust variables, was analysed in the context of potential bias. The F_1 scores suggest that the performance differs the most across countries. This could be explained by the strong differences in the class imbalances across them, as F_1 is strongly threshold dependent, but also by the different number of respondents from each country. The differences between ROC-AUC were relatively lower, meaning that the model is able to some extent differentiate between classes similarly across countries. Across other demographic groups, performance was much more stable, especially in terms of ROC-AUC. The model, however, performed worse for the individuals not born in their country of residence. Again, this may be caused by the dataset characteristics, as roughly 4 % of the respondents were born abroad. This analysis further emphasizes the role of the country in determining in shaping individual attitudes towards immigration.

SQ4 *How do trust-related features alter country-level model performance and key predictors of immigration attitudes?*

To better understand how the drivers of individual immigration attitudes differ between countries, the previously trained LightGBM models (with and without trust-related features) were evaluated at the country level. As presented in Figure 9, clear differences in ROC-AUC scores can be observed between countries. Without trust variables, ROC-AUC is lowest in countries with small shares of immigration opponents, such as Sweden. Interestingly in Norway, with a very similar class imbalance, ROC-AUC is visibly higher. This suggests that prediction difficulty depends not only on class proportions but also on how strongly the predictors relate to immigration attitudes in each country. This could, however, also be attributed to the use of a single model trained on the full European sample, which may not fully capture country-specific patterns. Adding trust variables improves ROC-AUC in most cases, especially in Montenegro, but slightly worsens it in Croatia and the United Kingdom, possibly due to added dimensionality and overfitting.

The SHAP values analysis revealed that in most European countries, attitudinal and ideological factors play the strongest role in shaping immigration attitudes. However, some differences were visible as well. For example, the effect of religiosity varied across countries. While in Greece it is rather associated with negative immigration attitudes, in Austria and Croatia this effect is subtle, whereas in Hungary it is slightly associated with more positive attitudes. These ambiguous findings align with Goubin et al. (2022), Palermo et al. (2022) and Schahbasi et al. (2021), who came to different conclusions regarding the influence of this factor. The strength of the education effect also depends on the country, which is consistent with the previous research (Rooduijn, 2022; Umansky et al., 2025). The trust-related variables became important in most of the countries, however, their effect is ambiguous as well. In Finland, both social and institutional trust are strong pro-immigration predictors, in Bulgaria, both forms of trust have only a weak and inconsistent, slightly anti-immigration effect, and in Greece, trust in police influences immigration attitudes negatively. This stands in contrast to Goubin et al. (2022), Halapuu et al. (2013), and Mitchell (2021), who concluded the positive role of trust when focusing on Europe as a whole. This thesis reaches a similar conclusion at the European level, but once the analysis moves to the country level, it becomes clear that the role of trust is not always straightforward. The reason for that could be cross-country differences in how institutions are framed in public debates and in how strongly they are linked to issues like migration and security.

6.2 *Contribution and societal relevance*

This study proposed novel methods for predicting individual immigration attitudes in Europe

6.3 *Limitations and future research*

This study encountered several limitations. Firstly, it relied on cross-sectional data, meaning that it could not determine a causal relationship. Future research could focus on tracking immigration attitudes over time, using panel or longitudinal data.

The target variable in this study was constructed based on survey questions about respondents' willingness to accept immigrants from different backgrounds. Future research could utilize machine learning to predict attitudes towards each immigrant group separately and compare both model performance and the main predictors across these groups.

Furthermore, Arcila Calderón et al. (2022) demonstrated the effectiveness of Support Vector Machines in a related study. Due to computational constraints, this model was not included in the comparison in this thesis. Future work could therefore benchmark SVM against the algorithms used in this research to assess its performance on the ESS data.

Although this study compared model performance between countries, the hyperparameters were tuned across the entire ESS dataset, which could affect their predictive abilities on much smaller country-level samples. Future research could focus on carefully tuning the models for each country separately, which may improve their results. Given the limited number of respondents from several countries, it is advisable to perform Nested Cross-Validation on each country separately, which could examine the robustness of the models' performances. Moreover, due to the significant differences in class proportions between countries, future work could explore the use of different sampling techniques, such as SMOTE, to assess whether they can improve the performance of the models' performance.

7 CONCLUSION

In conclusion, this thesis evaluated the capability of machine learning models to predict individual immigration attitudes across Europe using the ESS dataset. By introducing gradient boosting methods to this domain, the study provided a comprehensive comparison against established baselines. Results indicate that while XGBoost, LightGBM, and CatBoost capture complex patterns more effectively than Logistic Regression and Random Forest, the absolute performance gains remain modest. Furthermore, through SHAP value analysis, the study confirmed that deep-seated attitudinal factors (self-transcendence and political ideology) are the strongest determinants, which aligns with prior research. Crucially, this research addressed a gap in the predictive literature by incorporating institutional and social trust variables. The results demonstrate that these factors, particularly trust in other people and the European Parliament, significantly improve the predictive power of LightGBM and rank among the top predictors. However, country-level analysis revealed the complex nature of trust. In specific contexts, its inclusion introduced noise that slightly diminished model performance. Ultimately, this paper demonstrates that building social and institutional trust is essential for improving cohesion and overcoming skepticism towards immigration.

REFERENCES

Abdumalikov, S., Kim, J., & Yoon, Y. (2024). Performance analysis and improvement of machine learning with various feature selection methods for eeg-based emotion classification. *Applied Sciences*, *14*(22), 10511.

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2623–2631.

Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, *12*(16), 4102–4107.

Ali, R., Lee, S., & Chung, T. C. (2017). Accurate multi-criteria decision making methodology for recommending machine learning algorithm. *Expert Systems with Applications*, *71*, 257–278. https://doi.org/https://doi.org/10.1016/j.eswa.2016.11.034

Arcila Calderón, C., Jiménez Amores, J., & Stanek, M. (2022, October). Predicting support for refugees in europe: Using machine learning and synthetic populations to predict support for acceptance of asylum seekers in european regions. In *Data science for migration and mobility*. Oxford University Press. https://doi.org/10.5281/zenodo.17109255

Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(6), 493–507.

Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., Li, Z., & Tao, F. (2021). Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. *European Journal of Agronomy*, *123*, 126204.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., et al. (2015). Xgboost: Extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1–4.

Chen, Y., Wu, X., Hu, A., He, G., & Ju, G. (2021). Social prediction: A new research paradigm based on machine learning. *The Journal of Chinese Sociology*, *8*. https://doi.org/10.1186/s40711-021-00152-z

Cioch, M., Kulisz, M., & Gola, A. (2025). Comparison of machine learning methods in predictive maintenance of machines. *Advances in Science and Technology Research Journal*, *19*, 33–44. https://doi.org/10.12913/22998624/208284

Dahiya, N., Gupta, S., & Singh, S. (2022). A review paper on machine learning applications, advantages, and techniques. *ECS Transactions*, *107*(1), 6137.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, *7*(Jan), 1–30.

Dražanová, L., Gonnot, J., Heidland, T., & Krüger, F. (2024). Which individual-level factors explain public attitudes toward immigration? a meta-analysis. *Journal of ethnic and migration studies*, *50*(2), 317–340.

Drouhot, L. G., Deutschmann, E., Zuccotti, C. V., & Zagheni, E. (2023). Computational approaches to migration and integration research: Promises and challenges.

Erhard, L., & Heiberger, R. (2023). Regression and machine learning. In *Research handbook on digital sociology* (pp. 130–145). Edward Elgar Publishing.

European Commission. (2025). *Standard eurobarometer 103: Spring 2025* (Public opinion in the European Union). European Commission. https://europa.eu/eurobarometer/surveys/detail/3372

European Social Survey European Research Infrastructure (ESS ERIC). (2025). *Ess round 11 - 2023. social inequalities in health, gender in contemporary europe*. Sikt - Norwegian Agency for Shared Services in Education; Research. https://doi.org/10.21338/ess11-2023

Fatima, S., Hussain, A., Amir, S. B., Awan, M. G. Z., Ahmed, S. H., & Aslam, S. M. H. (2023). Xgboost and random forest algorithms: An in depth analysis. *Pakistan journal of scientific research*, *3*(1), 26–31.

Garcia-Carretero, R., Holgado-Cuadrado, R., & Barquero-Perez, O. (2021). Assessment of classification models and relevant features on nonalcoholic steatohepatitis using random forest. *Entropy*, *23*(6), 763.

Gnat, S. (2021). Impact of categorical variables encoding on property mass valuation. *Procedia Computer Science*, *192*, 3542–3550.

Goubin, S., Ruelens, A., & Nicaise, I. (2022). Trends in attitudes towards migration in europe. *A comparative analysis*.

Halapuu, V., Paas, T., Tammaru, T., & Schütz, A. (2013). Is institutional trust related to pro-immigrant attitudes? a pan-european evidence. *Eurasian Geography and Economics*, *54*(5-6), 572–593.

Hall, T., & Rasheed, K. (2025). A survey of machine learning methods for time series prediction. *Applied Sciences*, *15*(11). https://doi.org/10.3390/app15115957

Han, S. (2022). An analysis of koreans' attitudes towards migrants by application of algorithmic approaches. *Heliyon*, *8*(8).

Hannuksela, V., Söderlund, P., & Miklikowska, M. (2024). Political interest, political ideology, and attitudes toward immigration. *Frontiers in Political Science*, *6*, 1422364.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020).

Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Harris, J. K. (2021). Primer on binary logistic regression. *Family medicine and community health*, *9*(Suppl 1), e001290.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Javdani, M. (2020). Public attitudes toward immigrationâdeterminants and unknowns. *IZA World of Labor*, *None*(None), 473–473. https://doi.org/None

Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial intelligence in medicine*, *50*(2), 105–115.

joblib developers, T. (2025, May). *Joblib* (Version 1.5.1). Zenodo. https://doi.org/10.5281/zenodo.15496554

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Martinović, M., Dokic, K., & Pudić, D. (2025). Comparative analysis of machine learning models for predicting innovation outcomes: An applied ai approach. *Applied Sciences*, *15*(7), 3636.

McKinney, W. (2010). Data structures for statistical computing in python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (pp. 51–56).

Mitchell, J. (2021). Social trust and anti-immigrant attitudes in europe: A longitudinal multi-level analysis. *Frontiers in sociology*, *6*, 604884.

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B. (2020). General pitfalls of model-agnostic interpretation methods for machine learning models. *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, 39–68.

Narkhede, S. (2018). Understanding auc-roc curve. *Towards data science*, *26*(1), 220–227.

Nguyen, H. V., & Byeon, H. (2024). A hybrid self-supervised model predicting life satisfaction in south korea. *Frontiers in Public Health*, *12*, 1445864.

Niu, Q., Gui, R., Liu, H., Li, L., Shi, L., Jia, K., Li, P., & Wang, L. (2025). Automated sleep staging model for older adults based on cwt and deep learning. *Scientific Reports*, *15*(1), 22398.

Palermo, F., Sergi, B. S., & Sironi, E. (2022). Does urbanization matter? diverging attitudes toward migrants and europe's decision-making. *Socio-Economic Planning Sciences*, *83*, 101278.

pandas development team, T. (2020, February). *Pandas-dev/pandas: Pandas* (Version latest). Zenodo. https://doi.org/10.5281/zenodo.3509134

Paranjape, A., Katta, P., & Ohlenforst, M. (2022). Automated data preprocessing for machine learning based analyses. *Proceedings of the COLLA*.

Patel, H., Guttula, S., Mittal, R. S., Manwani, N., Berti-Equille, L., & Manatkar, A. (2022). Advances in exploratory data analysis, visualisation and quality for data centric ai systems. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4814–4815.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Potančoková, M., Stonawski, M., & Gailey, N. (2021). Migration and demographic disparities in macro-regions of the european union, a view to 2060. *Demographic Research*, *45*, 1317–1354.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in neural information processing systems*, *31*.

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, *14*(1), 6086.

Richardson, E., Trevizani, R., Greenbaum, J. A., Carter, H., Nielsen, M., & Peters, B. (2024). The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns*, *5*(6), 100994. https://doi.org/https://doi.org/10.1016/j.patter.2024.100994

Rooduijn, M. (2022). The education gap over immigration and socioeconomic security. *The British Journal of Sociology*, *73*(4), 699–705.

Salciccioli, J. D., Crutain, Y., Komorowski, M., & Marshall, D. C. (2016). Sensitivity analysis and model validation. *Secondary analysis of electronic health records*, 263–271.

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79.

Schahbasi, A., Huber, S., & Fieder, M. (2021). Factors affecting attitudes toward migrants—an evolutionary approach. *American Journal of Human Biology*, *33*(1), e23435.

Shapley, L. S., et al. (1953). A value for n-person games.

Shekhar, S., Bansode, A., & Salim, A. (2021). A comparative study of hyper-parameter optimization tools. *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 1–6.

Terpilowski, M. A. (2019). Scikit-posthocs: Pairwise multiple comparison tests in python. *Journal of Open Source Software*, *4*(36), 1169. https://doi.org/10.21105/joss.01169

Umansky, K., Weber, D., & Lutz, W. (2025). Revisiting the role of education in attitudes toward immigration in different contexts in europe. *Genus*, *81*(1), 1.

Verkuyten, M. (2021). Public attitudes towards migrants: Understanding cross-national and individual differences. *World Psychiatry*, *20*(1), 132.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wainer, J., & Cawley, G. (2021). Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Systems with Applications*, *182*, 115222.

Wan, X. (2019). Influence of feature scaling on convergence of gradient iterative algorithm. *Journal of physics: Conference series*, *1213*(3), 032021.

Yektansani, K., He, C., Azizi, S., Aftabi, A., & AzizKhani, M. (2024). Can machine learning predict environmental attitudes and beliefs? *International Journal of Environmental Studies*, *81*(4), 2012–2026.

Zaidi, A., & Al Luhayb, A. S. M. (2023). Two statistical approaches to justify the use of the logistic function in binary logistic regression. *Mathematical Problems in Engineering*, *2023*(1), 5525675.

Zhang, Z. (2016). Missing data imputation: Focusing on single imputation. *Annals of translational medicine*, *4*(1), 9.

APPENDIX A

Table 13: Overview of features and proportion of missing values.

Feature	Type	Description	Number missing	% missing
lrscale	Numerical (0–10)	Placement on left–right political scale	5998	13.3
trstun	Numerical (0–10)	Trust in the United Nations	3306	7.3
trstep	Numerical (0–10)	Trust in the European Parliament	2809	6.2
trstlgl	Numerical (0–10)	Trust in the legal system	945	2.1
trstprl	Numerical (0–10)	Trust in country's parliament	882	1.9
trstprt	Numerical (0–10)	Trust in political parties	807	1.8
trstplt	Numerical (0–10)	Trust in politicians	670	1.5
trstplc	Numerical (0–10)	Trust in the police	423	0.9
ppltrst	Numerical (0–10)	Most people can be trusted vs you can't be too careful	121	0.3
ipstrgva	Numerical (1–6)	Important that government is strong and ensures safety	1049	2.3
ipfrulea	Numerical (1–6)	Important to do what is told and follow rules	1065	2.4
impcntr	Numerical (1–4)	Allowance for immigrants from poorer countries outside Europe	984	2.2
imdfetn	Numerical (1–4)	Allowance for immigrants of a different race or ethnic group	951	2.1
imsmetn	Numerical (1–4)	Allowance for immigrants of the same race or ethnic group	935	2.1
iprspota	Numerical (1–6)	Important to get respect from others	922	2.0
ipbhprpa	Numerical (1–6)	Important to behave properly	880	1.9
ipudrsta	Numerical (1–6)	Important to understand different people	861	1.9
ipsucesa	Numerical (1–6)	Important to be successful	849	1.9
ipcrtiva	Numerical (1–6)	Important to think new ideas and be creative	812	1.8
ipadvnta	Numerical (1–6)	Important to seek adventures and an exciting life	783	1.7
ipmodsta	Numerical (1–6)	Important to be humble and modest, not draw attention	785	1.7
ipshabta	Numerical (1–6)	Important to show abilities and be admired	789	1.7
impdiffa	Numerical (1–6)	Important to try new and different things	760	1.7
ipeqopta	Numerical (1–6)	Important that people are treated equally and have equal opportunities	776	1.7
ipgdtima	Numerical (1–6)	Important to have a good time	770	1.7
impfuna	Numerical (1–6)	Important to seek fun and things that give pleasure	776	1.7
imptrada	Numerical (1–6)	Important to follow traditions and customs	758	1.7
impricha	Numerical (1–6)	Important to be rich	745	1.6
impfreea	Numerical (1–6)	Important to make own decisions and be free	731	1.6
marital_lb	Categorical	Legal marital or registered partnership status	726	1.6
iplylfra	Numerical (1–6)	Important to be loyal to friends and people close	739	1.6
impenva	Numerical (1–6)	Important to care for nature and environment	744	1.6
iphlppla	Numerical (1–6)	Important to help people and care for others' well-being	713	1.6
impsafea	Numerical (1–6)	Important to live in secure and safe surroundings	695	1.5
hincfel	Numerical (1–4)	Subjective feeling about household's income	586	1.3
feethngr	Categorical	Respondent belongs to ethnic majority in the country	481	1.1
agea	Numerical	Age of respondent in years	335	0.7
rlgdgr	Numerical (0–10)	Self-reported religiosity	312	0.7
hhmmb	Numerical	Number of people living in household	259	0.6
edulvl	Numerical (0–7)	Highest education level according to ES-ISCED	169	0.4
domicile	Categorical	Type of area where respondent lives (city, town, village, countryside)	93	0.2
polintr	Numerical (1–4)	Self-reported interest in politics	73	0.2
brncntr	Categorical	Whether respondent was born in the survey country	27	0.1
gndr	Categorical	Gender of respondent	0	0.0

Feature	Type	Description	Number missing	% missing
pdwrk	Categorical	Doing paid work in the last 7 days	0	0.0
cntry	Categorical	Country of interview (ESS country code)	0	0.0

Table 14: PVQ-21 items, their basic Schwartz values and higher-order dimensions used to construct Self-Transcendence and Conservation.

PVQ item (ESS wording)	Basic value	Higher-order dimension
Important to think new ideas and being creative	-	-
Important to make own decisions and be free	-	-
Important to try new and different things in life	-	-
Important to seek adventures and have an exciting life	-	-
Important to have a good time	-	-
Important to seek fun and things that give pleasure	-	-
Important to show abilities and be admired	-	-
Important to be successful and that people recognise achievements	-	-
Important to be rich, have money and expensive things	-	-
Important to get respect from others	-	-
Important that people are treated equally and have equal opportunities	Universalism	Self-Transcendence
Important to understand different people	Universalism	Self-Transcendence
Important to care for nature and environment	Universalism	Self-Transcendence
Important to help people and care for others well-being	Benevolence	Self-Transcendence
Important to be loyal to friends and devote to people close	Benevolence	Self-Transcendence
Important to live in secure and safe surroundings	Security	Conservation
Important that government is strong and ensures safety	Security	Conservation
Important to do what is told and follow rules	Conformity	Conservation
Important to behave properly	Conformity	Conservation
Important to be humble and modest, not draw attention	Tradition	Conservation
Important to follow traditions and customs	Tradition	Conservation

Table 15: Class distribution of the target variable by country

Country	Negative attitudes	Positive attitudes	Total	Share Negative attitudes	Share Positive attitudes
Austria	1138	1133	2271	0.501	0.499
Belgium	1233	327	1560	0.790	0.210
Bulgaria	1004	1178	2182	0.460	0.540
Croatia	994	526	1520	0.654	0.346
Cyprus	209	456	665	0.314	0.686
Finland	962	558	1520	0.633	0.367
France	1255	426	1681	0.747	0.253
Germany	1802	523	2325	0.775	0.225
Greece	988	1718	2706	0.365	0.635
Hungary	333	1705	2038	0.163	0.837
Iceland	668	139	807	0.828	0.172
Ireland	1308	657	1965	0.666	0.334
Italy	1721	1047	2768	0.622	0.378
Latvia	635	530	1165	0.545	0.455
Lithuania	846	475	1321	0.640	0.360
Montenegro	985	564	1549	0.636	0.364
Netherlands	1167	494	1661	0.703	0.297
Norway	1178	143	1321	0.892	0.108
Poland	723	664	1387	0.521	0.479
Portugal	780	560	1340	0.582	0.418
Serbia	874	605	1479	0.591	0.409
Slovakia	462	936	1398	0.330	0.670
Slovenia	909	301	1210	0.751	0.249
Spain	1348	438	1786	0.755	0.245
Sweden	1093	107	1200	0.911	0.089
Switzerland	1013	317	1330	0.762	0.238
United Kingdom	1167	424	1591	0.734	0.266

Table 16: Pairwise Wilcoxon test with Bonferroni correction (F1 scores).

Model	RF	CatBoost	LogReg	LightGBM	XGBoost
RF	1.000000	0.000001***	1.000000	0.000001***	0.000001***
CatBoost	0.000001***	1.000000	0.000001***	1.000000	1.000000
LogReg	1.000000	0.000001***	1.000000	0.000001***	0.000001***
LightGBM	0.000001***	1.000000	0.000001***	1.000000	0.147220
XGBoost	0.000001***	1.000000	0.000001***	0.147220	1.000000

Note. Entries report Bonferroni-adjusted p-values from pairwise Wilcoxon signed-rank tests based on F1 scores.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

APPENDIX B: MODEL PERFORMANCES OF THE LIGHTGBM WITH TRUST-RELATED DATA ACROSS DIFFERENT DEMOGRAPHIC SUBGROUPS

Table 17: Model performance by country

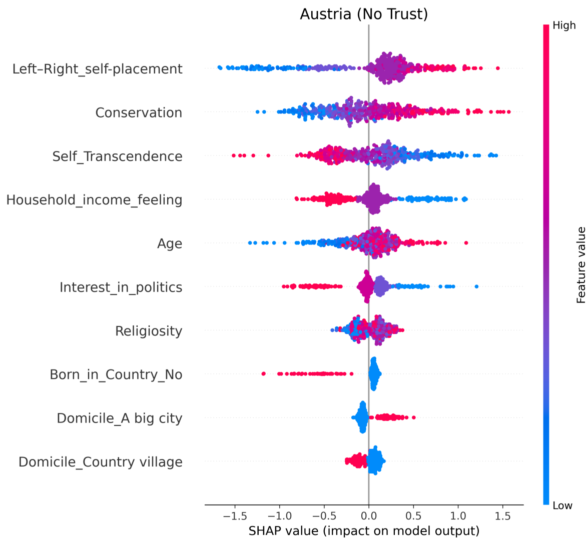
Countries (part 1)		
Country	F1 (class 1)	ROC-AUC
Austria	0.733	0.780
Belgium	0.372	0.725
Bulgaria	0.723	0.691
Croatia	0.516	0.668
Cyprus	0.821	0.788
Finland	0.635	0.803
France	0.539	0.802
Germany	0.393	0.759
Greece	0.787	0.703
Hungary	0.911	0.743
Iceland	0.150	0.693
Ireland	0.510	0.722
Italy	0.654	0.778
Latvia	0.609	0.627

Countries (part 2)		
Country	F1 (class 1)	ROC-AUC
Lithuania	0.538	0.695
Montenegro	0.563	0.716
Netherlands	0.477	0.693
Norway	0.111	0.718
Poland	0.717	0.752
Portugal	0.591	0.667
Serbia	0.559	0.662
Slovakia	0.819	0.745
Slovenia	0.427	0.695
Spain	0.462	0.772
Sweden	0.091	0.660
Switzerland	0.400	0.747
United Kingdom	0.526	0.781

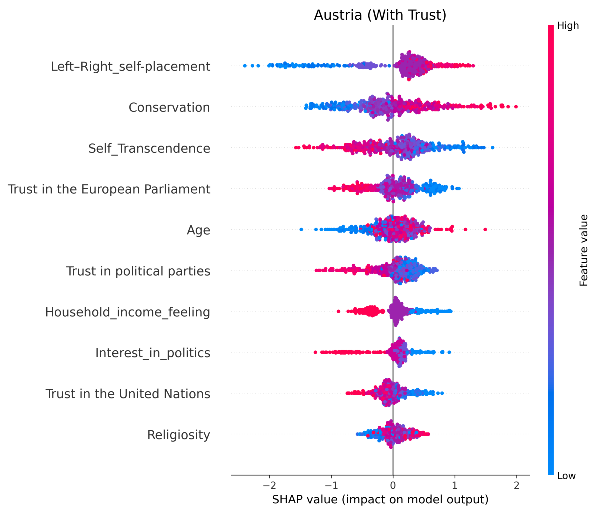
Table 18: Model performance across demographic subgroups

Subgroup	F1 (class 1)	ROC–AUC
Age groups		
15–24	0.575	0.806
25–34	0.622	0.805
35–44	0.609	0.776
45–54	0.651	0.779
55–64	0.663	0.794
65–74	0.694	0.772
75+	0.728	0.790
Born in country		
No	0.307	0.714
Yes	0.675	0.794
Domicile		
A big city	0.665	0.797
Country village	0.687	0.778
Farm or home in countryside	0.606	0.813
Suburbs or outskirts	0.602	0.813
Town or small city	0.647	0.793
Ethnic majority		
No	0.622	0.786
Yes	0.665	0.795
Gender		
Female	0.660	0.795
Male	0.664	0.794

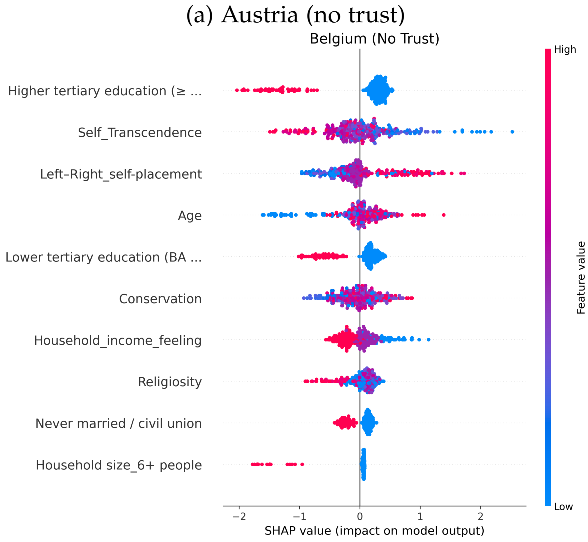
APPENDIX C



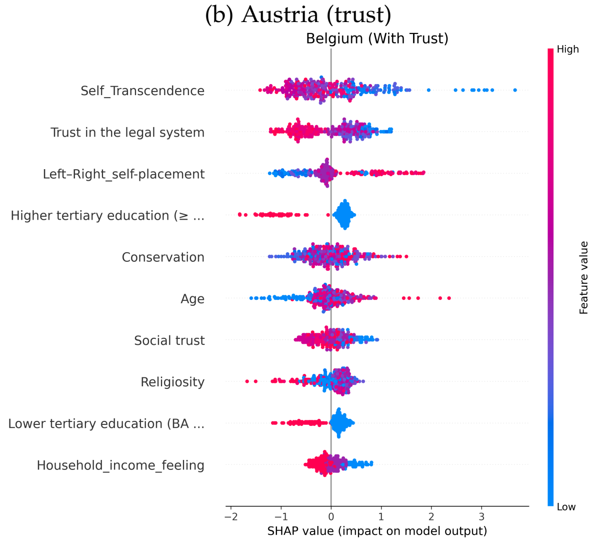
(a) Austria (no trust)



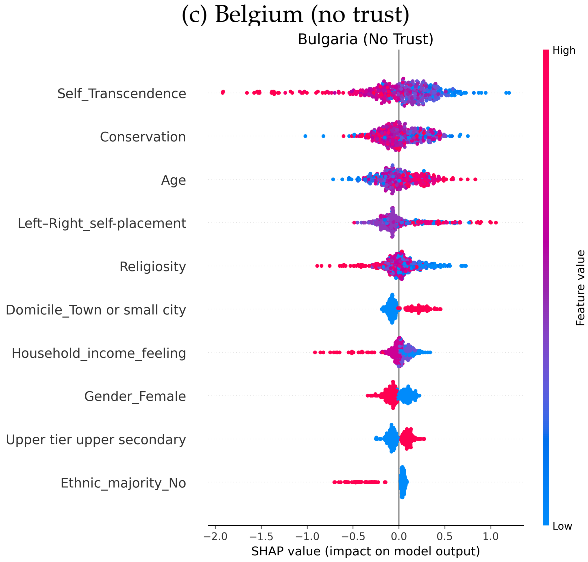
(b) Austria (trust)



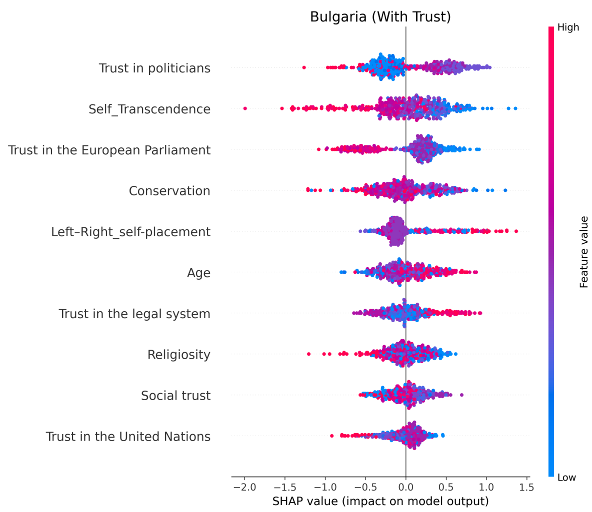
(c) Belgium (no trust)



(d) Belgium (trust)

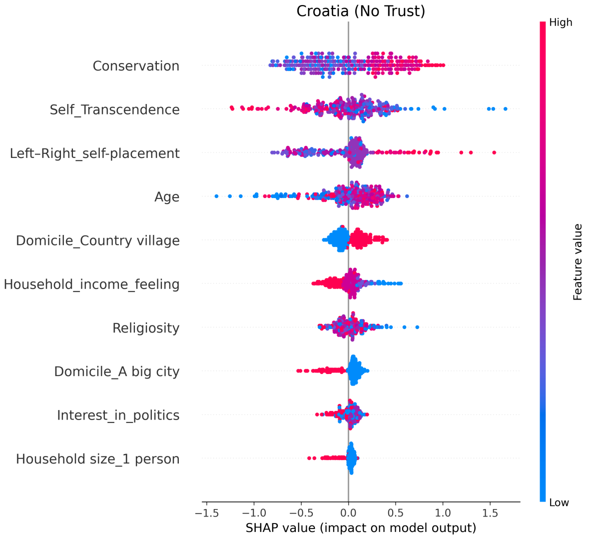


(e) Bulgaria (no trust)



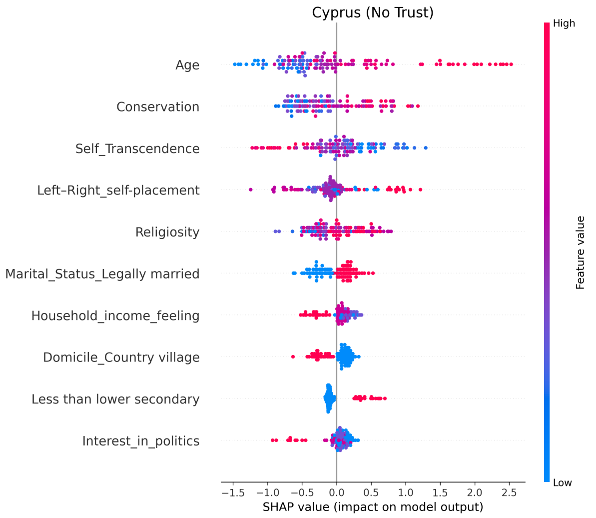
(f) Bulgaria (trust)

Figure 11: SHAP values for Austria, Belgium and Bulgaria, for models estimated without and with trust-related variables.



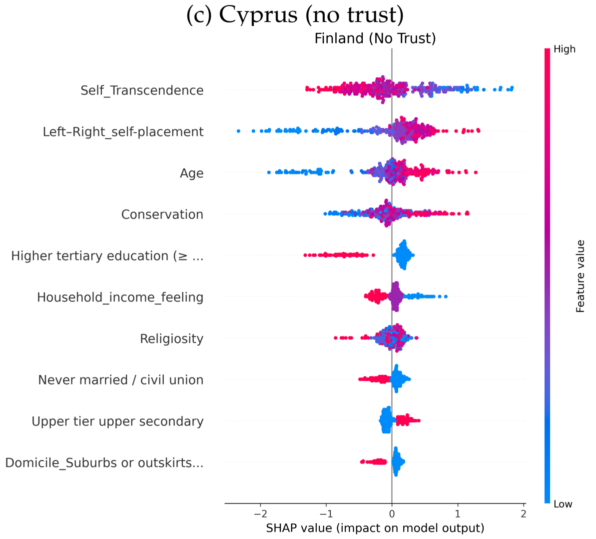
(a) Croatia (no trust)

(b) Croatia (trust)

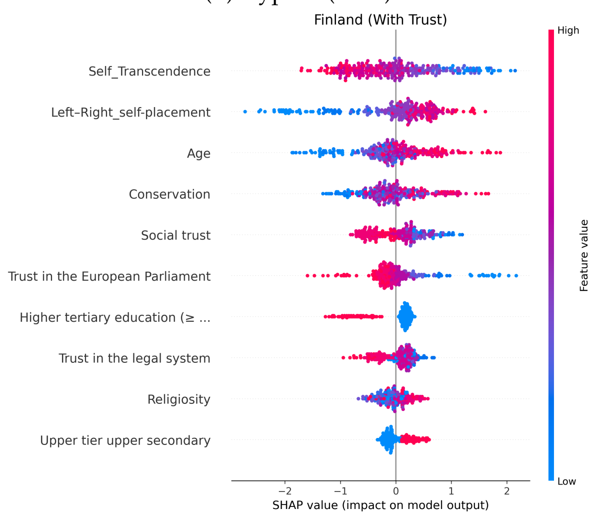


(c) Cyprus (no trust)

(d) Cyprus (trust)

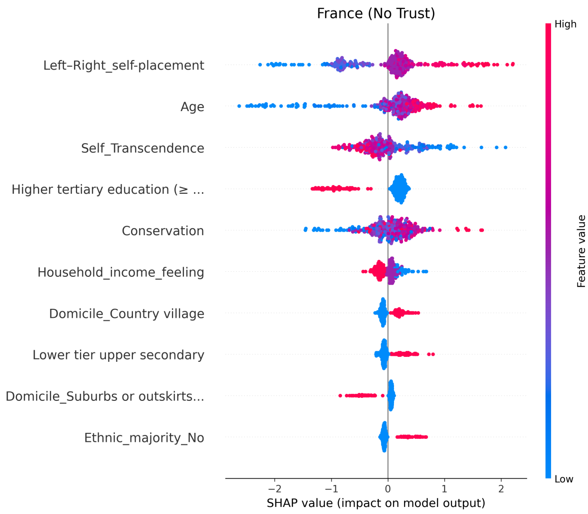


(e) Finland (no trust)

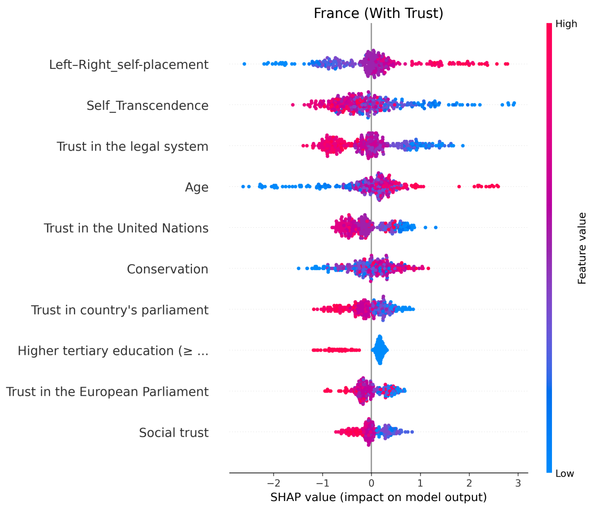


(f) Finland (trust)

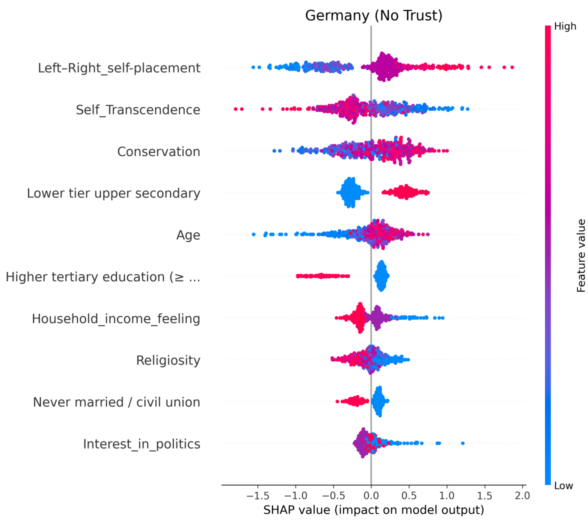
Figure 12: SHAP values for Croatia, Cyprus and Finland, for models estimated without and with trust-related variables.



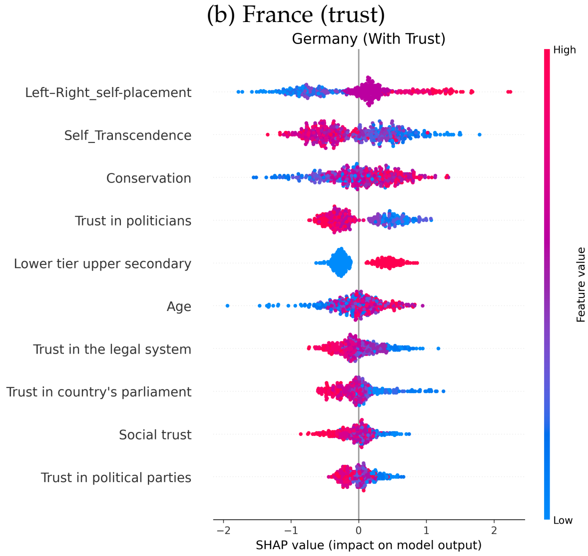
(a) France (no trust)



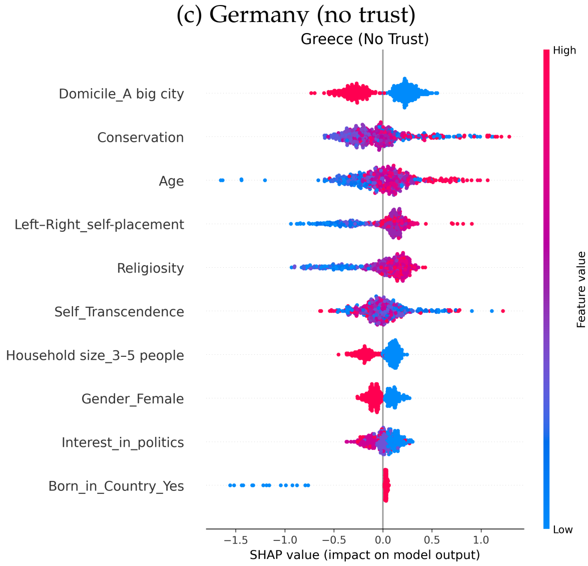
(b) France (trust)



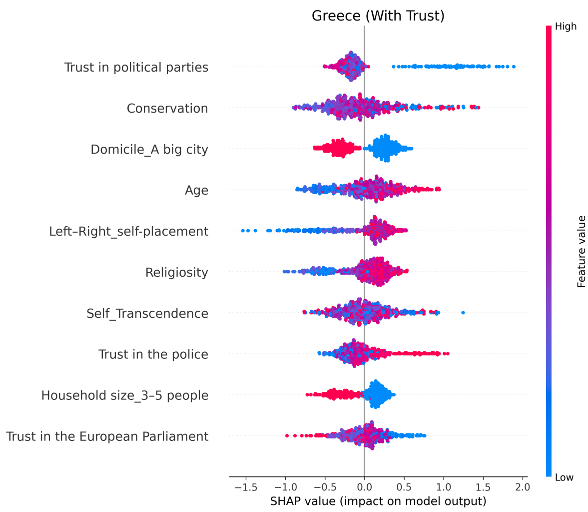
(c) Germany (no trust)



(d) Germany (trust)

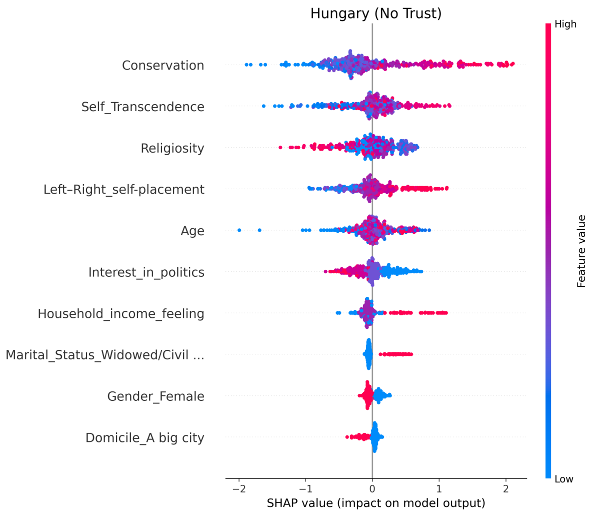


(e) Greece (no trust)

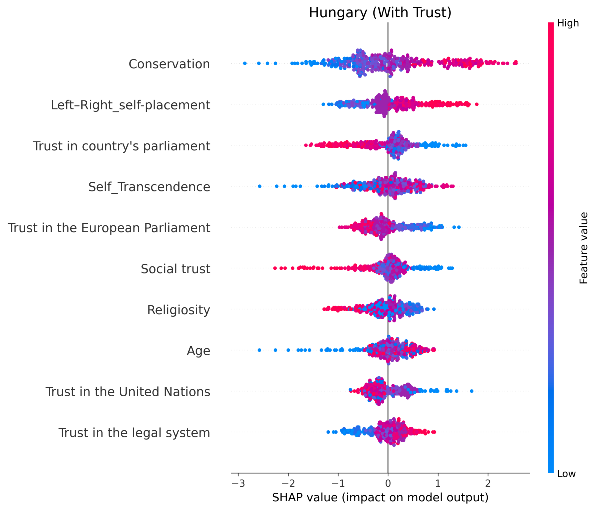


(f) Greece (trust)

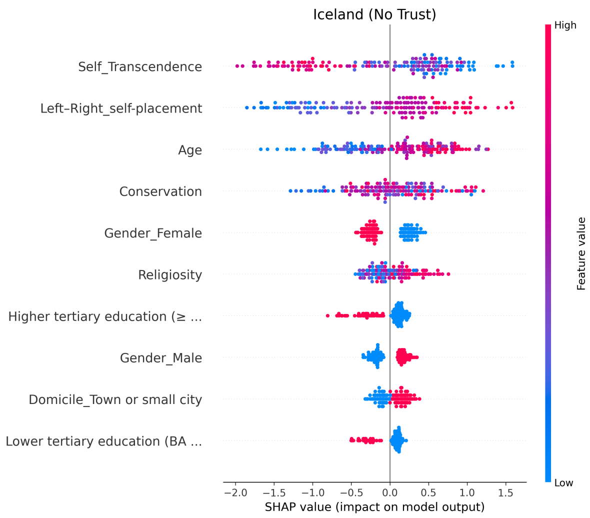
Figure 13: SHAP values for France, Germany and Greece, for models estimated without and with trust-related variables.



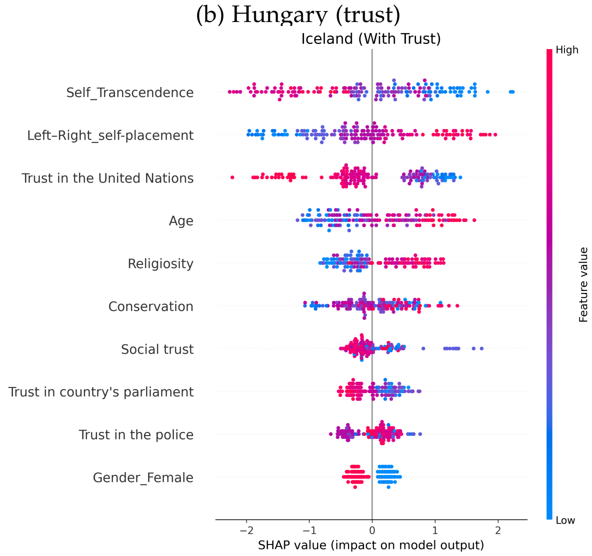
(a) Hungary (no trust)



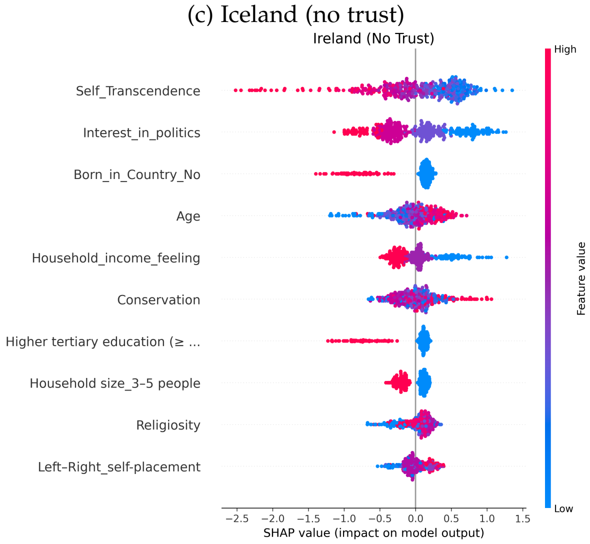
(b) Hungary (trust)



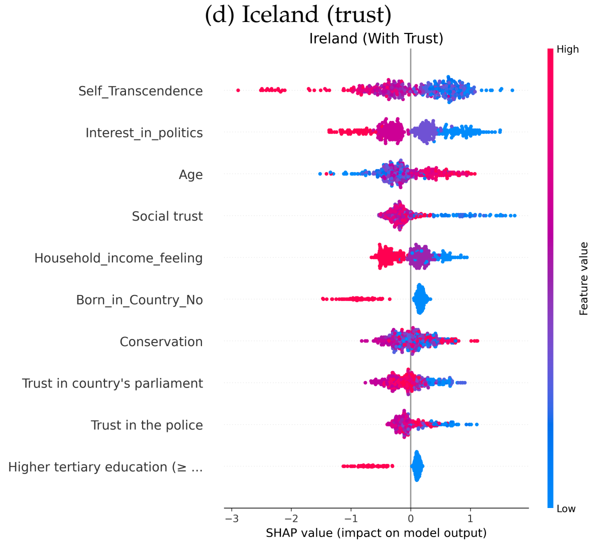
(c) Iceland (no trust)



(d) Iceland (trust)

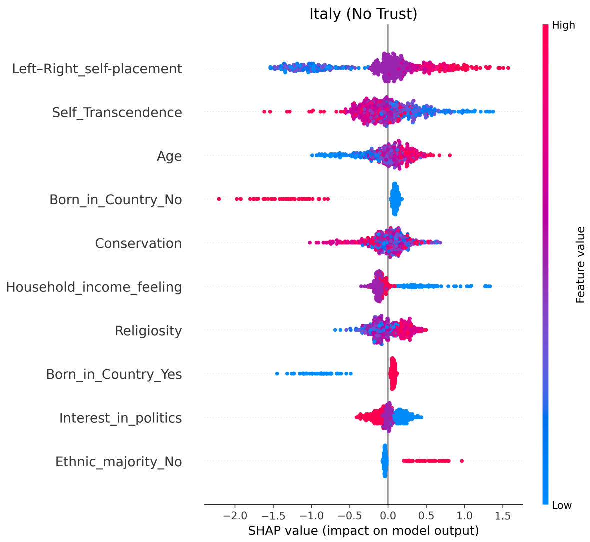


(e) Ireland (no trust)

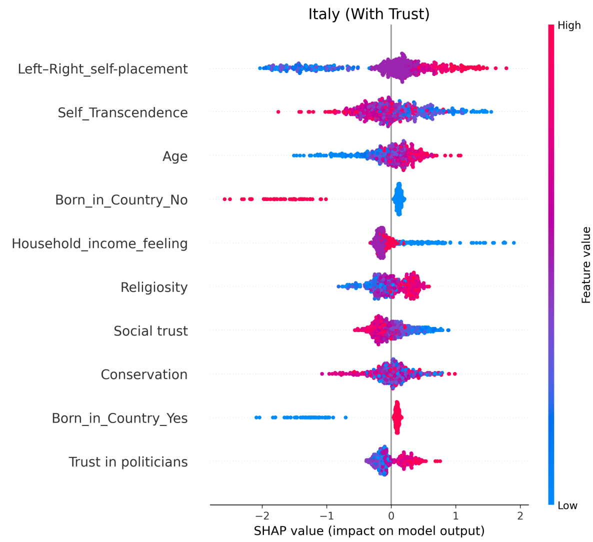


(f) Ireland (trust)

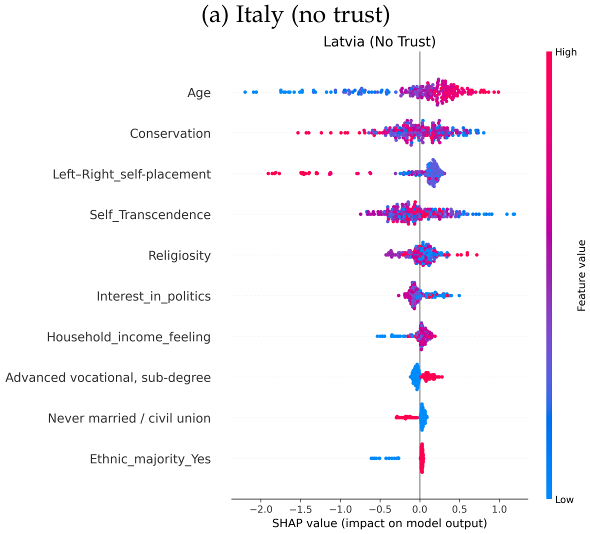
Figure 14: SHAP values for Hungary, Iceland and Ireland, for models estimated without and with trust-related variables.



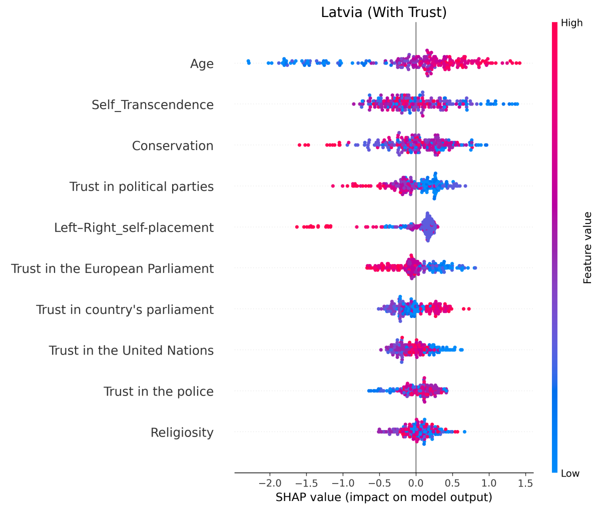
(a) Italy (no trust)



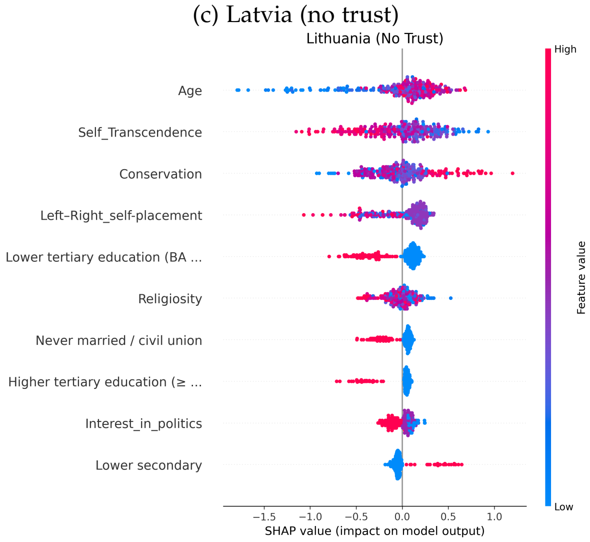
(b) Italy (trust)



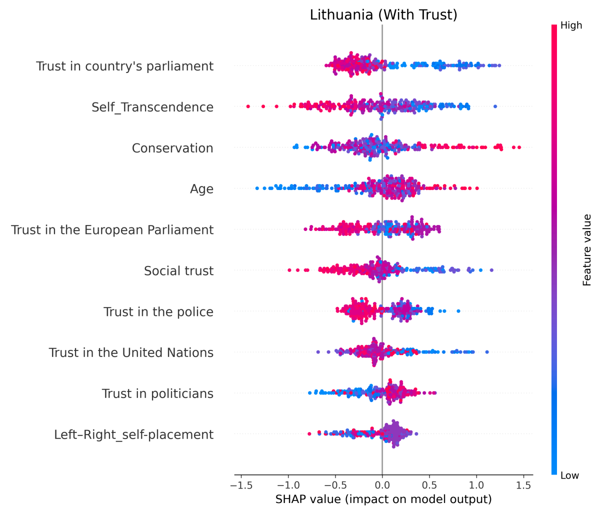
(c) Latvia (no trust)



(d) Latvia (trust)

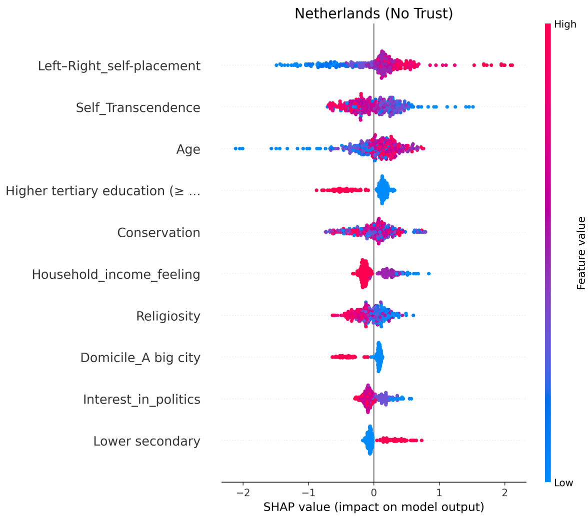


(e) Lithuania (no trust)

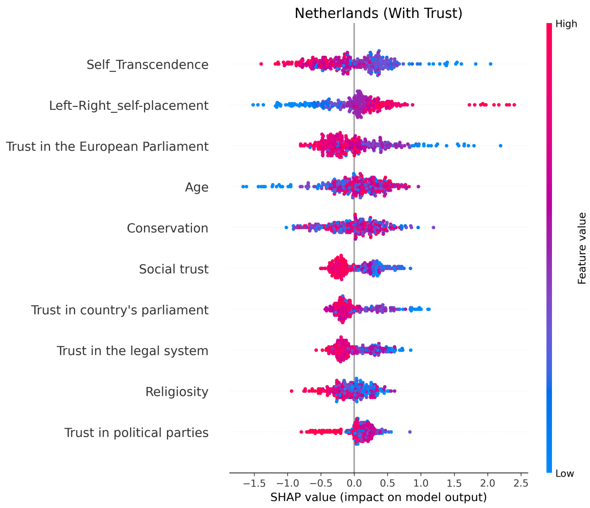


(f) Lithuania (trust)

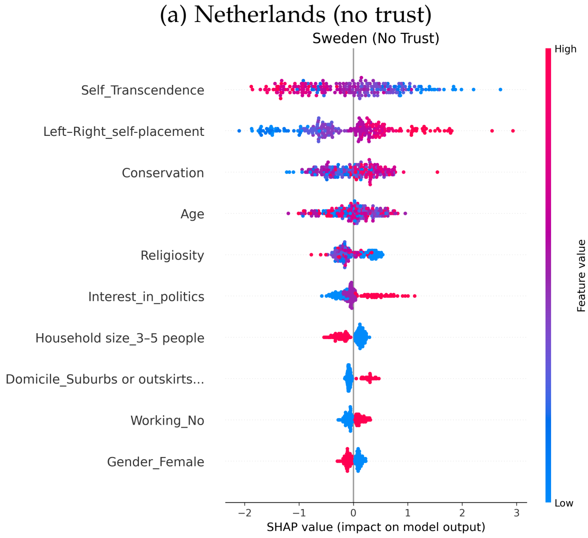
Figure 15: SHAP values for Italy, Latvia and Lithuania, for models estimated without and with trust-related variables.



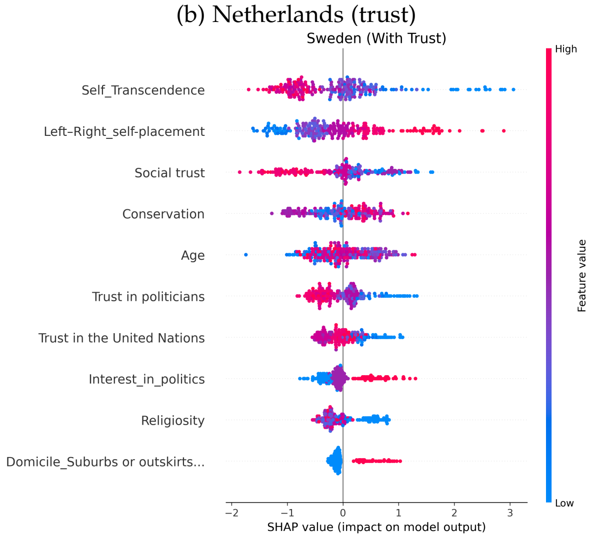
(a) Netherlands (no trust)



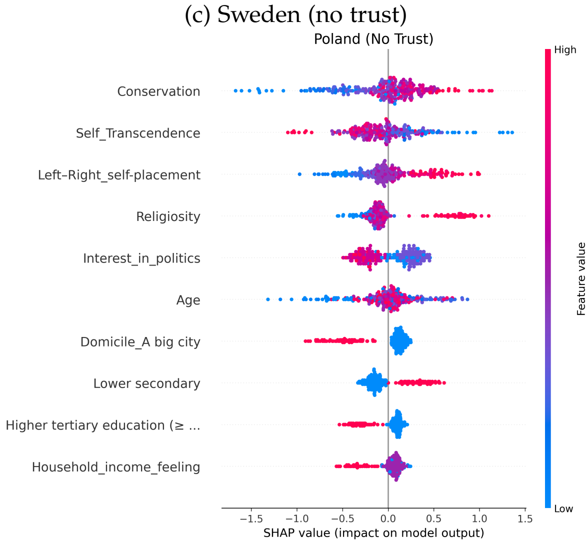
(b) Netherlands (trust)



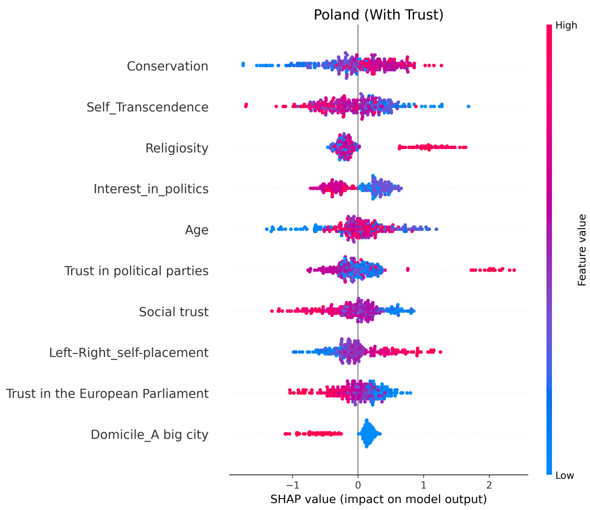
(c) Sweden (no trust)



(d) Sweden (trust)

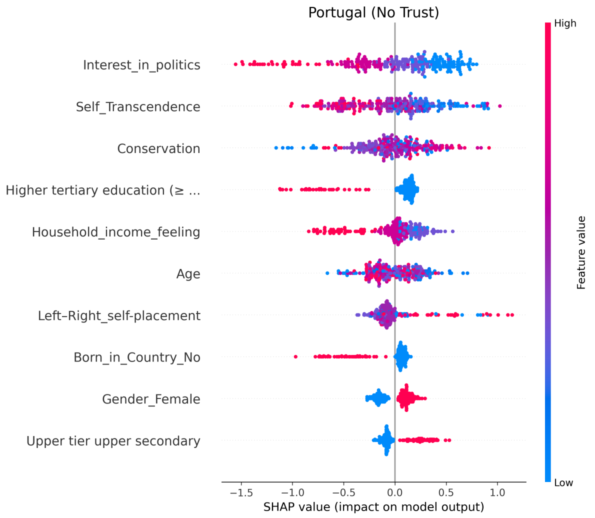


(e) Poland (no trust)

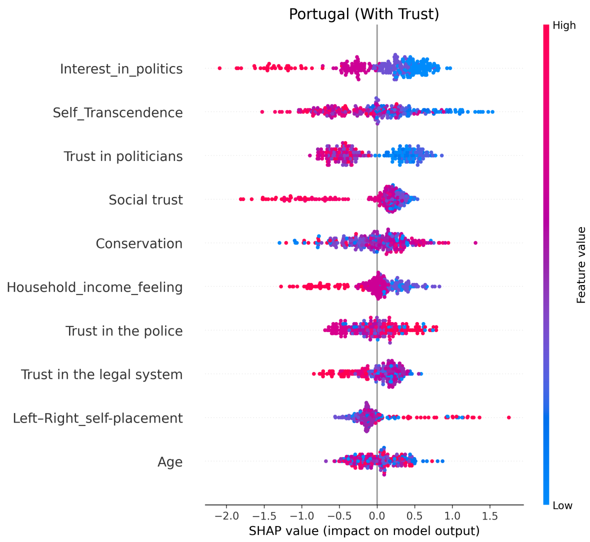


(f) Poland (trust)

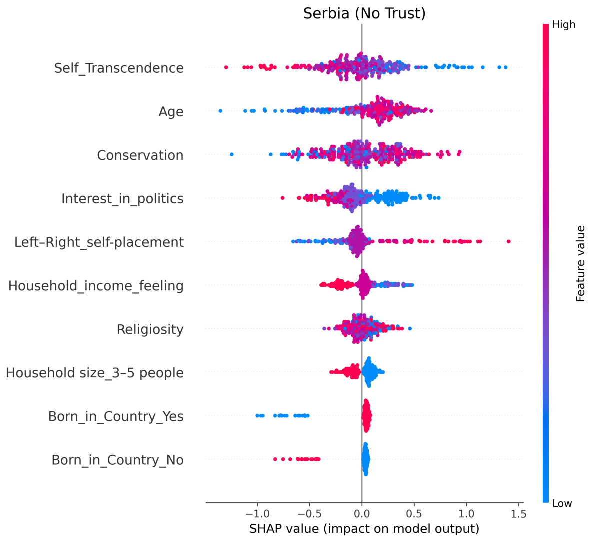
Figure 16: SHAP values for the Netherlands, Norway and Poland, for models estimated without and with trust-related variables.



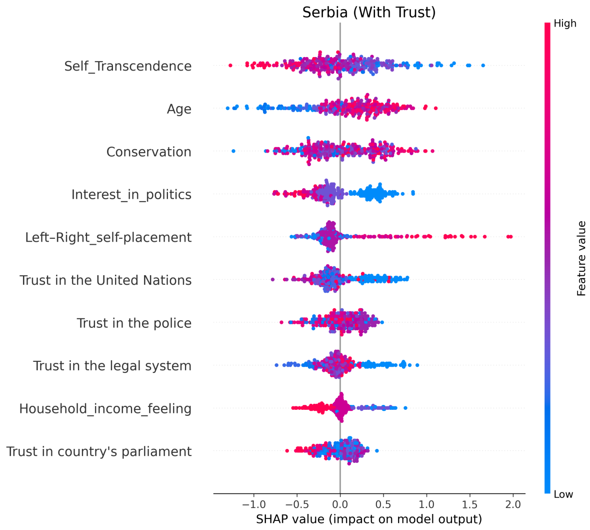
(a) Portugal (no trust)



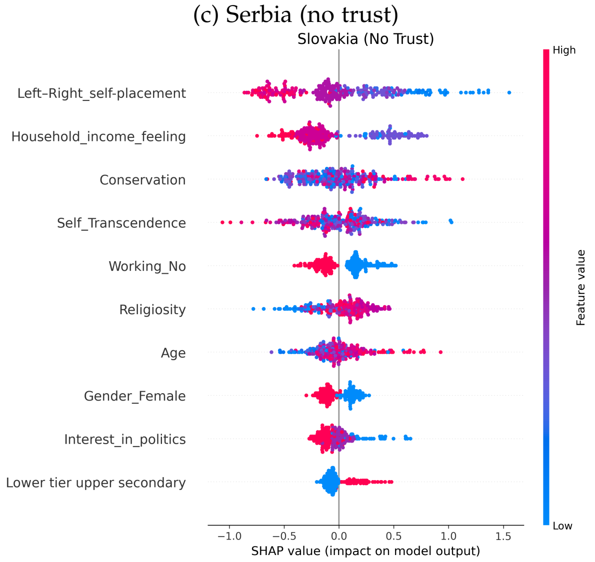
(b) Portugal (trust)



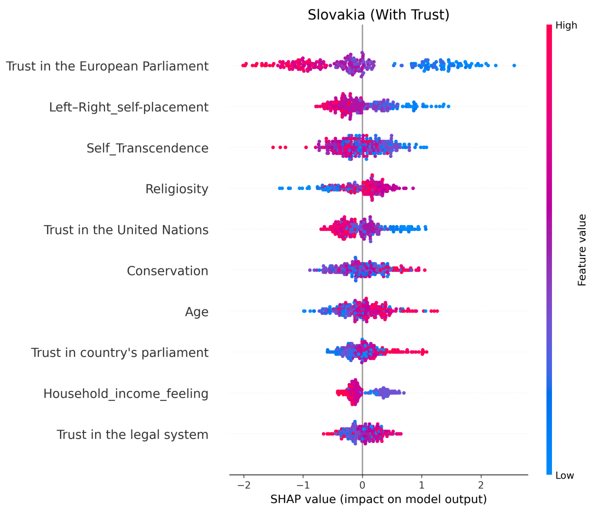
(c) Serbia (no trust)



(d) Serbia (trust)

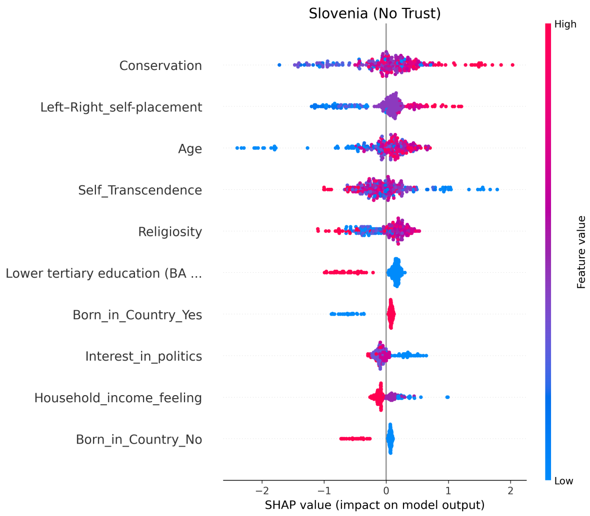


(e) Slovakia (no trust)

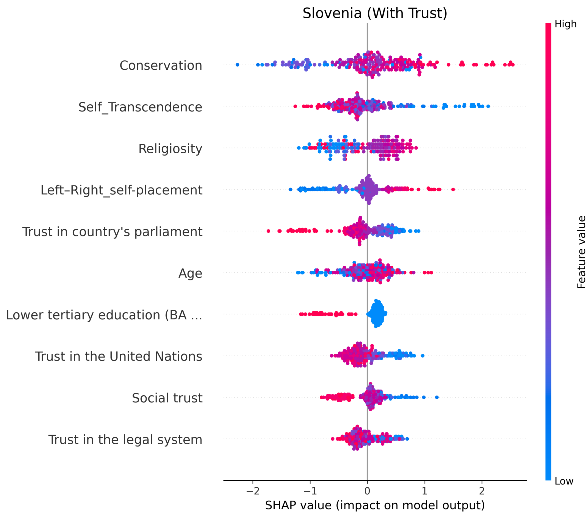


(f) Slovakia (trust)

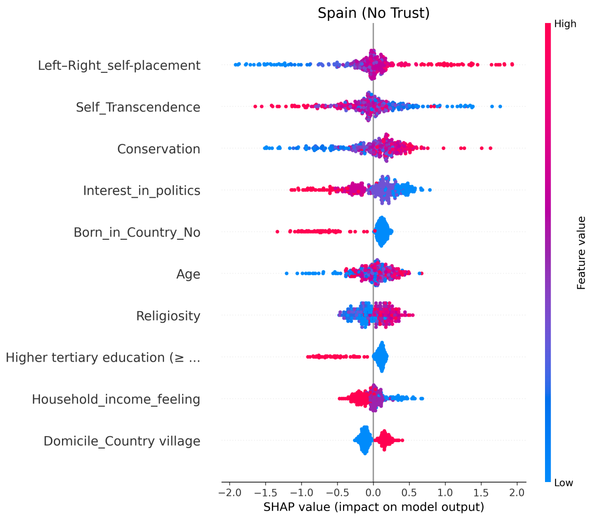
Figure 17: SHAP values for Portugal, Serbia and Slovakia, for models estimated without and with trust-related variables.



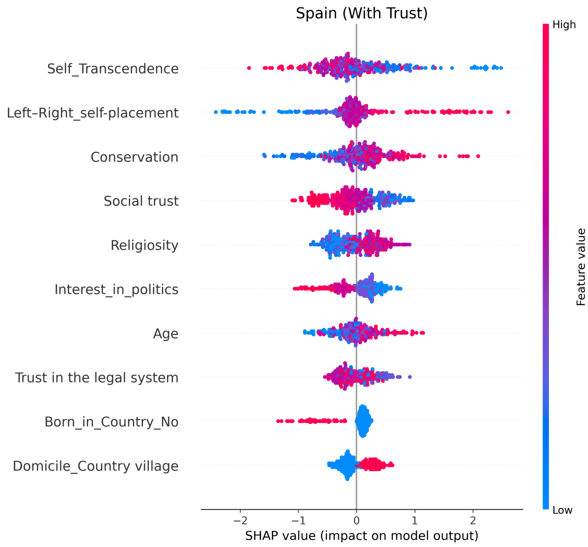
(a) Slovenia (no trust)



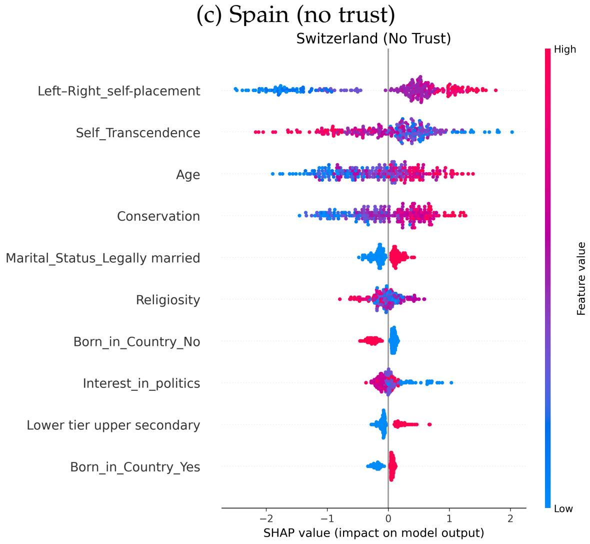
(b) Slovenia (trust)



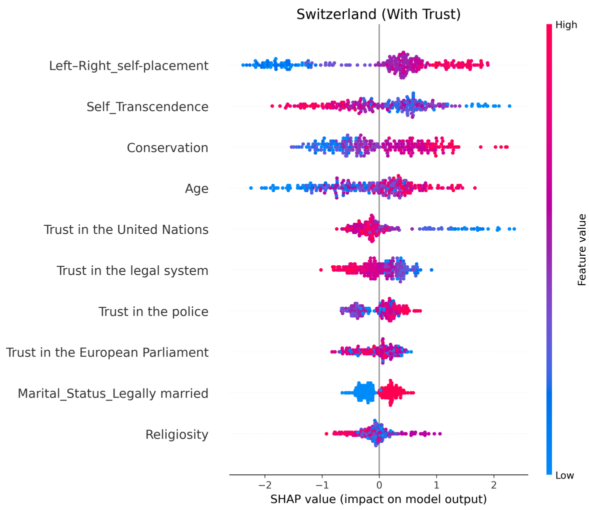
(c) Spain (no trust)



(d) Spain (trust)



(e) Switzerland (no trust)



(f) Switzerland (trust)

Figure 18: SHAP values for Slovenia, Spain and Switzerland, for models estimated without and with trust-related variables.

APPENDIX D

Table 19: Overview of software and libraries used in this thesis

Software / library	Reference
Python 3.12.4	–
NumPy	C. R. Harris et al., 2020
pandas	McKinney, 2010; pandas development team, 2020
Matplotlib	Hunter, 2007
scikit-learn	Pedregosa et al., 2011
SciPy	Virtanen et al., 2020
joblib	joblib developers, 2025
scikit-posthocs	Terpilowski, 2019
XGBoost	T. Chen et al., 2015
LightGBM	Ke et al., 2017
CatBoost	Prokhorenkova et al., 2018