Discrimination in the EU

Project summary

By: Gina Tedesco, Ekin Kızıldaş, Aurora Sterpellone

01. Introduction

Support for transgender rights varies significantly across Europe, influenced by a complex interplay of cultural norms, legal frameworks, economic variables, and historical contexts. Western Europe generally exhibits more progressive policies and public support compared to Eastern Europe. This disparity is evident in societal attitudes and legal frameworks, with some countries adopting progressive laws that facilitate gender recognition and protect against discrimination, while others maintain restrictive policies that undermine the rights of transgender individuals.

This project aims to analyze the factors influencing support for transgender individuals across European countries, using data from the Special Eurobarometer 493 survey.

Our objective is twofold: first, to explain variations in support for transgender civil document changes by examining both individual-level factors (such as demographics, religion, and ideology) and country-level factors (including GDP, legal frameworks, and policy environments); and second, to develop a predictive model capable of forecasting support levels in other countries. In addition, we incorporate education and healthcare variables to assess how access to resources, institutional policies, and social services shape public attitudes. By considering these structural and individual determinants, our goal is to provide a comprehensive understanding of the factors that influence support, recognizing that national investments in education and healthcare can play a crucial role in fostering inclusivity and policy acceptance.

02. Data processing

To ensure our dataset was accurate, complete, and ready for analysis, we undertook a thorough data cleaning process. This included loading data with `read_csv()`, as well as addressing missing values using the `summary()` and `is.na()` functions to identify any gaps in the data. Based on the extent of missingness, we employed appropriate imputation techniques or, when necessary, removed rows or columns with excessive missing values. The process also included standardizing data types and column names (converted using functions like `as.factor()`, `as.numeric()`, and `rename()`), eliminating duplicates (using `distinct()` or `duplicated()`), treating outliers (using `boxplot()` and summary statistics), and performing feature selection and engineering.

Finally, data consistency was ensured by checking categorical variables and standardizing date formats. The cleaned data was then exported using `write_csv()` for analysis.

You can find more detailed information about the variables in the dataset in our codebook (also available in our GitHub and the folder).

03. Descriptive Analysis

Data Processing and Missing Values

Our analysis began with an examination of missing data. Visualizations and summary statistics revealed gaps, particularly in education expenditure variables. We employed mean imputation using the `MICE` package, ensuring data completeness while minimizing information loss. Post-imputation checks confirmed the absence of missing values.

Individual-Level Analysis

Individual-level analysis examined demographics, religion, political views, and attitudes towards diversity. Logistic regression models revealed that gender and religious affiliation were the most significant predictors of support for transgender rights. Specifically, women demonstrated significantly higher support compared to men, and individuals identifying as atheists or non-believers showed the strongest positive association. Political ideology also played a role, with individuals leaning towards progressive views showing slightly higher support, although this effect was less pronounced than that of gender or religion.

To further explore the interplay between these factors, interaction terms were introduced into the model. This revealed that gender and religion jointly interact to shape support levels. However, these interaction terms did not significantly improve the overall model fit, suggesting that the simpler model remained more effective.

A multinomial logistic regression was conducted to examine support across three categories: oppose, don't know, and support. This analysis reinforced the significance of gender, with women more likely to fall into the "don't know" or support categories. Religion continued to be a strong predictor, with atheists and non-believers displaying the highest levels of support, while political ideology primarily influenced uncertainty rather than direct opposition or support. To ensure the robustness of these findings, cross-validation was performed, confirming that the results held across various model specifications.

Country-Level Analysis

Country-level analysis explored economic, educational, health, and legal factors impacting support for transgender rights.

While education spending generally ranged between 4% and 6% of GDP, with some countries exceeding 7%, significant economic disparities were evident. Regression analysis revealed that GDP had a negligible impact on support, suggesting economic wealth alone does not drive public attitudes. Using aggregated data grouped by ISO3 codes, we were able to determine that primary and tertiary education expenditures showed a positive association with support, while pre-primary and secondary education expenditures exhibited a slight negative effect. Vocational education spending had a minor positive influence. Healthcare access analysis indicated that hormone replacement therapy and psychological care were more widely available than surgical procedures like vaginoplasty or phalloplasty, likely due to eligibility criteria and cost. Legal recognition policies varied considerably across countries. Self-determination was less common than policies requiring medical or psychological evaluations, and strict legal barriers like sterilization requirements, though less prevalent, persisted in some regions. Legal recognition for minors also differed significantly.

Overall, economic factors demonstrated a weak predictive capacity for support levels, while legal and healthcare access showed substantial variation. Gender and religion consistently emerged as the most influential factors, underscoring the dominance of cultural and ideological influences over purely economic considerations in shaping public attitudes towards transgender rights.

Gender and religion are the primary factors influencing support for transgender rights in Europe, overshadowing economic influences. Legal and healthcare accessibility vary significantly across countries, reflecting diverse national policies.

04. Statistical Models

The initial phase focused on data preparation, where the dataset was cleaned, missing values were handled, and feature types were correctly formatted. The target variable, representing support levels, was converted into a factor with three categories. The dataset was then split into training and testing sets to prevent data leakage during feature selection. Random Forest was employed to identify the top 20 most influential features, which were subsequently used to train various models.

SMOTE (Synthetic Minority Over-sampling Technique)

We utilized several machine learning algorithms, including Random Forest, Multinomial Logistic Regression, Decision Trees, XGBoost, Gradient Boosting Machine (GBM), and an ensemble method.
The Random Forest model achieved a perfect accuracy of 1.000, which is highly suspicious and indicative of potential overfitting or data leakage, requiring further validation.
Multinomial Logistic Regression demonstrated moderate predictive power with an accuracy of 0.6086, effectively capturing some data complexities but struggling with the "Don't Know" category.
The Decision Tree model performed similarly to logistic regression, achieving an accuracy of 0.609, highlighting the effectiveness of rule-based approaches.
XGBoost outperformed other models with an accuracy of 0.6296, showcasing its ability to capture complex patterns through boosted ensemble learning.
The GBM model, however, performed poorly with an accuracy of 0.5262, likely due to suboptimal parameter tuning and overfitting.
The ensemble method, which combined predictions from multiple models, also

underperformed, achieving an accuracy of 0.3766, suggesting that the combined predictions did not provide additional value.

A comparison of the models revealed that Random Forest and Multinomial Logistic Regression's perfect accuracy raised concerns about overfitting. XGBoost and Decision Trees offered moderate predictive power, while GBM and the ensemble method exhibited lower performance. The confusion matrix for the Random Forest model confirmed its perfect classification, further emphasizing the need for validation to ensure model integrity.

Results

Feature importance analysis from Random Forest identified key predictors, with "transgender_civil_dc1" showing an unusually high importance score, potentially indicating data leakage. Other significant features included school policies related to transgender and intersex inclusion, national identifiers, legal and medical requirements, and socio-political factors such as discrimination and ideology.

These findings highlight the complex relationship between institutional, social, and cultural factors influencing public support for transgender civil document changes. The project underscores the importance of rigorous validation and feature engineering to develop robust and reliable predictive models.

05. Challenges faced and solutions

Throughout the modeling process, several key challenges emerged that impacted both model performance and interpretability. One major issue was data leakage, particularly concerning the extreme importance of the `transgender_civil_dc1` variable in the Random Forest model, which led to artificially high accuracy. This required extensive debugging to ensure that the model was learning meaningful patterns rather than relying on unintended data artifacts. However, despite multiple attempts to mitigate this issue—such as adjusting feature selection, revalidating preprocessing steps, and tuning hyperparameters—we repeatedly encountered either extreme errors or models that reported 100% accuracy, making it difficult to fully resolve the concern.

Another significant challenge was class imbalance, where the distribution of the target variable was skewed, leading to models that struggled to accurately predict the minority class. To address this, SMOTE (Synthetic Minority

Over-sampling Technique) was applied, but its implementation required extensive preprocessing, including one-hot encoding to ensure numerical consistency across features. Even with this correction, some models continued to exhibit unexpectedly perfect performance, raising further concerns about potential hidden biases in the dataset.

Additionally, model instability was an issue, particularly with Gradient Boosting Machines (GBM), where the `multinomial` distribution setting was flagged as unstable in the package, potentially affecting reliability. Alternative boosting methods like XGBoost performed better but still faced tuning difficulties and inconsistent improvements.

The ensemble model also presented an unexpected challenge, as combining multiple models did not improve accuracy—instead, it underperformed compared to individual models, suggesting that weak base models limited the effectiveness of ensembling.

Finally, computational intensity was a concern, especially with Random Forest and XGBoost, requiring careful tuning of hyperparameters to balance performance and runtime efficiency. Despite our best efforts to adjust, retrain, and refine the models, no matter how we approached the issue, we either ran into significant errors or achieved unrealistic 100% accuracy. Given these persistent challenges, we have chosen to present the models as they currently stand, as they are the closest we could get them to functioning properly at this stage without all of them displaying perfect accuracy. These results highlight the complexities of working with real-world data and the need for further refinements to improve generalizability and validity in future iterations.

06. Conclusion

In this project, we explained the variations in support for transgender rights across Europe and developed a predictive model for support levels in other countries. Our analysis of individual-level factors demonstrated that gender and religion are the primary drivers of support, with women and non-believers showing higher levels of support. While political ideology plays a role, its influence is less pronounced. Country-level analysis indicated that economic factors, such as GDP, have a limited impact on support, whereas legal frameworks and healthcare access vary significantly across countries. Our statistical models highlighted the predictive power of Random Forest, Multinomial Logistic Regression, Decision Trees, and XGBoost, although some models, like Random Forest, raised concerns about potential overfitting. Feature importance analysis identified key predictors, including transgender civil document change indicators, school policies, legal/medical requirements, and socio-political factors. These findings underscore the complex interplay of individual, institutional, social, and cultural factors shaping public support for transgender rights. The project emphasizes the need for continued research and rigorous validation to foster a deeper understanding of these dynamics and to develop more robust and reliable predictive models.

References

Asher & Lyric. (2023, August 8). *Global Trans Rights Index: 203 countries ranked in 2023*. https://www.asherfergusson.com/global-trans-rights-index/

Calderon-Cifuentes, P. A. (2021). *Trans Discrimination in Europe. A TGEU analysis of the FRA LGBTI Survey 2019*. TGEU.

Datanovia. (2020, April 27). *Highchart Interactive World Map in R: The Essentials*. https://www.datanovia.com/en/lessons/highchart-interactive-world-map-in-r/

European Commission. (n.d.). *Legal aspects of LGBTIQ equality*. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/lesbian-gay-bi-trans-and-intersex-equality/legal-aspects-lgbtiq-equality_en

European Institute for Gender Equality. (n.d.). *Culture*. https://eige.europa.eu/gender-mainstreaming/policy-areas/culture?language_content_entity=en

Falck, F., & Bränström, R. (2023). The significance of structural stigma towards transgender people in health care encounters across Europe: Health care access, gender identity disclosure, and discrimination in health care as a function of national legislation and public attitudes. *BMC Public Health*, *23*(1). https://doi.org/10.1186/s12889-023-15856-9

Garrido, M. (2025, March 4). *The 15 Countries with the Best Education: 2025 ranking*. Global Citizen Solutions. https://www.globalcitizensolutions.com/countries-with-best-education/

ILGA-Europe. (n.d.). *Legal gender recognition - Rainbow Map*. https://rainbowmap.ilga-europe.org/categories/legal-gender-recognition/#:~\:text=Legal%20gender%20recognition%20is%20based,only%20in%20Germany%20and%20Iceland

Mosley, T. (2023, June 28). How school systems, educators and parents can support transgender children. *NPR*. https://www.npr.org/2023/06/28/1184805981/how-school-systems-educators-and-parents-can-support-transgender-children

O'Donoghue, S. (2023, October 30). Trans rights in Europe: Where does your country stand? *Euronews*. https://www.euronews.com/2023/10/28/trans-rights-in-europe-where-does-your-country-stand

TGEU - Transgender Europe. (2024a, January 23). *10 Years of Trans Rights Mapping in Europe & Central Asia: Decade-long Research Reveals Overall*. https://tgeu.org/10-years-of-trans-rights-map-europe-central-asia/

TGEU - Transgender Europe. (2024b, May 15). *Trans Rights Index & Map 2024 reveals polarisation in trans rights in Europe and Central Asia*. https://tgeu.org/tgeus-trans-rights-index-map-2024-reveals-polarisation-in-trans-rights-in-europe-and-central-asia/

TGEU - Transgender Europe. (2025, March 13). *Homepage*. https://tgeu.org/

The Williams Institute at UCLA School of Law. (2022, January 25). *The relationship between LGBT inclusion and Economic Development: Emerging Economies - Williams Institute*. Williams Institute. https://williamsinstitute.law.ucla.edu/publications/lgbt-inclusion-economic-dev/#:~:text=In%20particular%2C%20the%20analysis%20suggests,correlation%20with%20GDP%20per%20capita.

Toxigon. (2025, January 11). *Trans Rights and the Fight for Equality: A 2025 perspective*. Toxigon. https://toxigon.com/trans-rights-and-the-fight-for-equality

Trans Health Map 2024. (2024, November 2). Trans Health Map 2024. https://transhealthmap.tgeu.org/

Transition from Special Education to Inclusive Education Systems | European School Education Platform. (2024, November 5). https://school-education.ec.europa.eu/en/discover/publications/transition-special-education-inclusive-education-systems

Williamson, M. (2023). A Global Analysis of Transgender Rights: Introducing the Trans Rights Indicator Project (TRIP). *Perspectives on Politics*, *22*(3), 799–818. https://doi.org/10.1017/s1537592723002827