Cross-country abortion travelers in Germany: What “Foreigners” in German abortion statistics can tell us

We present a classifier approach in order to estimate the origin of abortion seekers within the different states of Germany.

AUTHOR

Niklas Pawelzik, Justus v. Samson-Himmelstjerna, Alvaro Guijarro May

PUBLISHED

Dec. 8, 2022

Problem Introduction

There is a lot to be learnt from the Polish abortion laws about European morality policies. Morality policies are matters on which "at least a significant minority of citizens has a fundamental, first-principled conflict with the values embodied" (Mooney 1999), with abortion policies clearly constituting a textbook example. Despite its great overall impact on European politics, regarding morality policies, the European Union has remained mostly invisible; and morality policy-making in a Europeanized context is an underrepresented strand of research (Ayoub 2013) – partly because, in contrast to the European integration process, morality policies are seen as often non-economic in nature (Schwartz and Tatalovich 2019). In an indirect fashion, however, open Schengen borders and Europeanization of related policies have eased the access for European citizens to other morality policy regimes. In turn, this means that *national capacities to effectively govern their population can be undermined by the access to more liberal regimes in fellow EU-countries*: If the Polish government further restricts access to abortion care, Polish women will seek this care in more permissive neighboring countries.

To quantify this effect, however, is a difficult task, given the absence of official statistics or border controls, and the desire of abortion-seeking women for discretion. The current project, therefore, proposes to approach in an indirect manner, based on the assumption that, *if a tighter abortion regime in country A causes women to cross borders to seek abortion care in country B, this should result in a higher number of abortions performed on foreigners in country B*. Reviewing the literature in the field, researchers of the "Europe Abortion Access Project" conclude that the purpose of seeking abortion care abroad is a "phenomenon [that] remains poorly understood" ("Europe Abortion Access Project" 2020). They argue that the lack of scientific insight in this aspect of European abortion regimes is linked to a general lack of sufficient "quantitative and qualitative data" on the matter.

In line with these thoughts, we present a Logistic Regression Model for Binary Classification, a Random Forrest Classifier, and a Support Vector Machine Model that predict labels for abortions - regarding whether they were performed on foreigners or members of the national population. The labeling is based on state-level information about the distance to neighboring countries, about the population share with a migrant background, and about the foreign population in the respective state. For all three variables, in addition the restrictiveness of the respective country is factored in. Given the feasibility limits in terms of time and data access, this project developed an initial framework based on German data on abortion statistics of the year 2021. We conclude with several promising avenues how to expand this approach in future research.

Data

1. Abortions by foreign and local populations in Germany in 2021

The German Statistical Office annually publishes data on abortions performed annually at the federal level. The office furthermore offers information on the origin of the abortion seeker, distinguishing between the 16 German states and a 17th category, "foreign". Furthermore, information on the origin of the patient is listed for the individual procedures, whereby a distinction is made between the various federal states and the category "foreigners". The dataset contains information for all 16 federal states with about 100,000 abortions in all of Germany in 2021 [1] .

2. Abortion Index Score of European Countries

This Index is an initiative by the European Parliamentary Forum for Sexual and Reproductive Rights (EPF) and International Planned Parenthood Federation European Network (IPPF EN). In its in-depth analysis of European abortion policies for the Year 2021, 53 European countries and territories are scored on a scale from 0-100, based on the sub-categories "access", "legal status of abortion care", "clinical care and service delivery", and "information and online information" [2] .

3. Foreign Population within German States

The Federal Statistical Office publishes statistics on foreigners registered in Germany with state-level information on, among other things, their sex, years of age and country of origin [3] .

4. Germans with migrant background within German States

The so-called "Mikrozensus" of 2020, a census on a representation basis, provides estimations of the size German communities with migrant background. "German with migrant background", according to the census definition, are migrants of the first or second generation that are *also* German citizens. Since this census works with estimations, in order to provide reliable numbers, only information on groups of significant size are published. Therefore, not for every migrant community in every German state are numbers available [4] .

5. Geographical distance from German States to European Countries

In order to calculate the distances between the 16 German States and the different European countries. Taking into account geographic and demographic information from a data set on simplemaps.com, we calculated the average distance between the German States' capitals and all available city data from each European country. This allowed us to create an average distance score from each German state to each European country [5] .

Proposed Method

The data available to us contains information on abortions per year per state. It distinguishes 'foreign' abortion-seekers (without further information on the distribution of countries of origin), which enables us to create a binary classification that distinguishes between 'foreign' and 'local'. Independent variables can be used to enable a Logistic Regression Model (LRM) and a Random Forrest Classifier (RFC) to estimate the probability that an event belongs to a specific class. In LRM, a statistical approach is used for classification problems and is used when the dependent variable (target) is categorical. The probability of an event being of the intended class (binomial in this case) is calculated by the fitted model, trained with a portion of the available data, and if the probability is under or over a set threshold, the model will assign the corresponding class. On the other hand, the RFC will generate the most frequent answer to several decision trees built with a random subset of features in order to select the predicted class.

The binary dependent variable we are trying to predict is the origin of the respective abortion seeker (foreigner / German citizen). In order to predict this dependent variable, we first used a Linear Regression Model with our three independent variables. The first group of variables involves the fraction of population of every German State with a migrant background. The second group of variables involves the fraction of foreign population residing in each German State. The third group of variables involve the average distance between the German states' capital cities and the primary cities of each European country. For all three variables we then created interaction terms between the respective variable and the countries' abortion scores from the abortion atlas.

Furthermore we implemented from the scikit-Learn library the internal "class_weight" function in all of the evaluated models and used the train-test-split function so that we could fit and evaluate the model on separate chunks of the dataset. This way, the training models can penalize the misclassification made towards the minority class (foreign patient) by setting a higher class weight and simultaneously reduce the weight for the majority class (domestic patient).

After experiencing some issues with parameter and hyperparameter tuning for our Logistic Regression we decided to focus on the Random Forest Classifier in order to further understand exactly what is being predicted. From sklearn.metrics we used accuracy_score, confusion_matrix, roc_auc_score, and classification_report. Since our first unweighted Random Forest Model was only predicting class 0, and hence forth completely ignoring the minority class in favor of the majority class, we decided to introduce a class weight to our model. We then further improved our AUROC value by using a Balanced Random Forest Model with random undersampling.

Lastly, in order to better compare our results from Logistic Regression Model and the Random Forest Classifier, we decided to run a Support Vector Machine algorithm as well. We reported the performance of this model by calculating the mean area under the ROC curve, by gathering three repeats of 10-fold cross-validations.

Results

As expected, our unweighted models would not be able to predict the minority class, since we knew that our data was heavily skewed and therefore, unbalanced. At a first glance, however, all weighted and balanced models yielded better results than we expected since the accuracy score for all of our weighted models were in the high 80s, which were promising high explanatory values. However, since accuracy is not a great measure of classifier performance when the classes are imbalanced, we also generated the area under the Receiver Operating Characteristic (ROC) Curve, measuring the overall summary of diagnostic accuracy. Since the AUC equals 0.5 when the ROC curve corresponds to random chance and 1.0 for perfect accuracy, we were quite content at first with our results in the high 80s and low 90s. However, as soon as we looked at our F1 Scores, as well as the feature importance of our model, we realized that these values did not measure the performance of our models accurately.

Classification Model	Accuracy		AUROC	Compute Time (sec)
	Balance	Unbalance		
Logistic Regression (unweighted)	0.500	0.994	0.902	9.418
Logistic Regression (weighted)	0.845	0.861	0.902	21.41
Random Forrest (unweighted)	0.994		0.902	6.11
Random Forrest (weighted)	0.861		0.845	5.683
Random Forrest with Random Undersampling	0.861		0.904	2.116
Support Vector Machine	0.882		0.858	411.716

Figure 1: Results of Evaluated Models

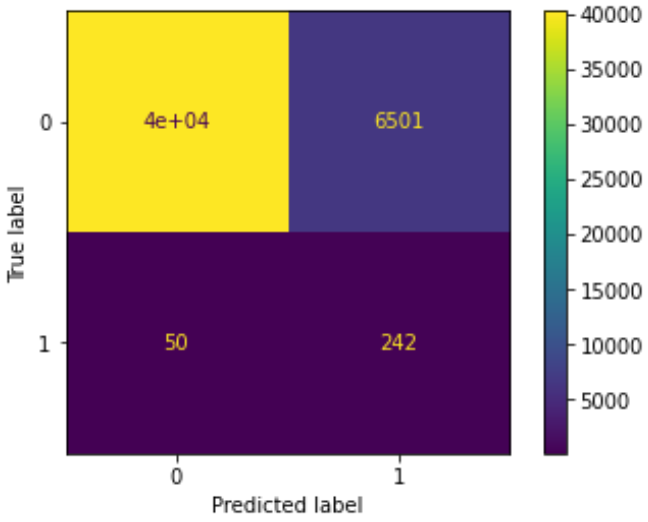

Figure 2: Confusion Matrix of Random Forrest Classifier (weighted)

All of the confusion matrixes look like the weighted Random Forrest Classifier model. Models that include computationally generated weights assign both labels, and with more than 40.000 abortions correctly predicted as performed on Germans, more than 80% of the labels per class were attached correctly. Thanks to the weights attached to the classes, in a similar manner more than 80% of the almost 300 abortions performed on foreigners were correctly predicted. Given that all of our variable operate exclusively on the state-level, we consider this explanatory capacity of 80% per class of our models a great success, especially compared to their unweighted counter parts.

The poor performance of our model on these metrics shows that the low number of German states involved in combination with abortion indices for only a single year lays too heavily on the explanatory power of our models. We anticipated that this problem, which is inherent in our available data, might constitute a challenge for our project. Our models provide a satisfying answer to the class imbalance; however, the problems in the data structure itself could neither be avoided nor solved by the current means at our hand. Given that we have had only state-level data from the 16 German states available, and that we could therefore only combine a total of 32 value combinations.

We, therefore, consider the current state of our project a partial success nevertheless, given that we detected the hidden performance issues, and were able to identify and explain the issue as outlined in the analysis below. This in turn allowed us to identify clear strategies for solutions and further expansions in more resourceful project settings in the conclusive remarks.

Analysis

The particularities of our data discussed in earlier chapters are reflected in the results our models produce. Given the extreme imbalance of our data, unbalanced models behave essentially like dummy models: Aiming for high accuracy scores, they label every abortion as domestic without exception. This results in accuracy scores that are logically equal to the class distribution: less than 1% of our data represents abortions performed on foreigners, thus labeling every abortion as domestic is a successful strategy in more than 99% of the cases. In order to produce meaningful results, however, we have to focus on foreign abortions: Mislabeled German (domestically performed) abortions are less problematic than mislabeled foreign abortions. Our balanced models have more explanatory power because they can provide precisely this added value in comparison to the unbalanced classifiers. Following this focus on labeling foreign abortions, our preliminary factor analysis becomes relevant: While keeping the flaws of our models in mind, the most important factors among the interaction terms are pointing in a clear direction: 13 of the 15 most important independent variables are distance-based, if one considers predictive power for abortions on foreigners. Even though these insights must be taken with caution, they are in line with the initial intuition: Proximity to neighboring countries in combination with the level of restrictiveness within these countries´ abortion regimes is a strong predictor for a high number of foreign abortions.

Going back to the our starting assumption, this comes as no surprise: Berlin and Brandenburg, the closest states to Poland (a country with highly restrictive policies), have the highest number of foreign abortion seekers within Germany. States with big communities of Polish foreigners or Germans with Polish migrant backgrounds, however, such as North Rhine Westphalia, do not show higher numbers for abortions performed on foreigners, both in absolute and relative numbers. These two independent variables, people with migrant backgrounds and foreigners, seem to play only a minor role in explaining the cases of abortions performed on foreigners. While the data included needs to be expanded and diversified, there is a clear trend regarding what matters when it comes to explaining abortions performed on foreigners: proximity to neighboring countries and the degree of restrictiveness of their respective abortion regimes.

Conclusion

The project aimed to provide a classifier model that predicts labels for abortions in a binary fashion, as either being performed on German or non-German women. With models that include computationally generated weights, more than 80% of the labels per class were attached correctly. Given the exclusively state-level nature of our data, this explanatory capacity is a remarkable success. Preliminary factor importance analysis furthermore suggests that of the three considered explanatory variables, geographical proximity combined with abortion restrictiveness scores is the decisive factor for these predictions, and migrant background share and foreign population are of little relevance.

The poor performance indicated for instance by the F1 scores results from the flaws of our data structure. Variables that exclusively operate on a state level capture only 32 value combinations constitute a clear limitation to the models´ success. However, this clear identification of the source of problems - the low number of distinct value combinations - allows for three concrete suggestions for refined future research projects: A first option would be to include data that operate on a lower level, ideally on the level of the individual abortion. The binary classifier logic we developed for this project would allow an easy incorporation of such information since the dataframe already lists one observation per abortion. However, such individual data on abortions is not publicly available, the only way to gain access to this type of information would therefore be through non-public channels.

A second option would be to include updated data on abortion numbers and abortion scores in upcoming years. Including abortion numbers from past years currently already available would not be sufficient for this: In order to generate new value combinations, the independent variables would need to change. Unfortunately, only the abortion numbers - so the binary dependent variable - are updated on an annual basis. Only subsequent editions of the first abortion atlas could provide new value combinations for our current independent variables populated by the German state.

The third and most promising approach would therefore be to expand the developed approach beyond the German states and include data from further European countries. Including state-level data from further countries would multiply the amount of value combination for each independent variable and improve our model drastically. Unfortunately, accessing and processing the required census and abortion data from additional European countries would certainly not have been feasible within this project, given the nationally specific data structure and the extremely time-consuming wrangling process. However, the first promising results of the present project might justify a subsequent project in order to further pursue the developed pioneering approach to quantify abortion migration across European borders by using classifier models.

Footnotes

1. https://www-genesis.destatis.de/genesis//online?operation=table&code=23311-0006&bypass=true&levelindex=1&levelid=1664961110565#abreadcrumb [↩]
2. https://www.epfweb.org/node/857 [↩]
3. https://www-genesis.destatis.de/genesis/online#astructure [↩]
4. https://www.bamf.de/DE/Themen/Forschung/Veroeffentlichungen/Migrationsbericht2020/PersonenMigrationshintergrund/pernode.html [↩]
5. https://simplemaps.com/data/world-cities [↩]

References

Ayoub, Phillip M. 2013. “Cooperative Transnationalism in Contemporary Europe: Europeanization and Political Opportunities for LGBT Mobilization in the European Union.” *European Political Science Review* 5 (2): 279–310.

“Europe Abortion Access Project.” 2020. University of Barcelona. https://europeabortionaccessproject.org/.

Mooney, Christopher Z. 1999. “The Politics of Morality Policy: Symposium Editor’s Introduction.” *Policy Studies Journal* 27 (4): 675–80.

Schwartz, Midred A, and Raymond Tatalovich. 2019. “The Rise and Fall of Moral Conflicts in the United States and Canada.”

Corrections

If you see mistakes or want to suggest changes, please create an issue on the source repository.

Reuse