Electoral Vulnerability Index 2025
Technical report, May 2025

David Randahl[1,2], Maxine Leis[1,2], and Tim Gåsste[1,2]

[1]*Department of Peace and Conflict Research, Uppsala university*
[2]*UpSight QRA*

1   Introduction

During the spring 2025, the UpSight team has on behalf of the Kofi Annan Foundation (KAF) worked on an update of the Electoral Vulnerability Index (EVI), a forecasting tool for forecasting electoral violence. The EVI was originally developed in 2022 using state-of-the-art machine learning techniques. The system uses data from the Varieties of Democracy (V-DEM) project (Coppedge et al., 2021b), as well as data from the Digital Societies Project (DSP) (Mechkova et al., 2022), and the World Bank's World Development Indicators (WDI). Forecasts are made at the annual level for each country with at least one scheduled election in the specific year and made for two calendar years into the future from the last available data.

This technical report briefly outlines the definitions and methodology used by the EVI and highlights some changes in the methodology of the EVI since the 2024 EVI report. This report also shows the updated evaluation metrics of the tool and the forecasts for 2025 and 2026.

2   Defining electoral violence

The first task of the UpSight team was to develop an outcome measure for Electoral Violence, which is conceptually sound and has properties that make that allow us to produce forcasts that are practically useful for the KAF. In discussions with KAF about what types of forecasts are most useful in practice, and internal discussions about what data are possible to use, we landed in agreeing that an ordinal outcome for electoral violence was the most useful. Based on a review of the existing sources of electoral violence in Fjelde et al. (2021), we decided to use the Varieties of democracy (VDEM) project indicators of electoral violence. To code this outcome we used two indicators measuring electoral violence by non-government actors (v2elpeace) and intimidation and harassment by government actors (v2elintim) from the VDEM election-level data. These two indicators are themselves ordinal on a 5-point scale from 0 (most electoral violence/harassment) to 4 (least electoral violence/harassment). These indicators are coded by country experts, and we use the mean value across all coded values (Coppedge et al., 2021a). The full description of these two indicators and their levels can be found in Appendix A.

In order to make the forecasts more stable and to make a more conceptually clear outcome, we re-coded these two indicators into a three-point ordinal scale measuring 'electoral violence and/or harassment'. We did this by first re-coding the 5 point scales to a three point ordinal scales by merging the categories by coding values below 1.5 as 'severe electoral violence or harassment', values between 1.5 and 3 as 'limited electoral violence or harassment' and values of 3 or above as 'no electoral violence or harassment'. We then coded our outcome, 'electoral violence and/or harassment', as the most severe value across these two three-point indicators. This definition differs slightly from the definition used in the 2022 version of the EVI where the category of 'no electoral violence' was coded for elections with a value of 3.5 or above on the 5 point scales. This adjustment was made to be a bit more lenient in the coding of peaceful elections and thus to more effectively separate between the no violence and limited violence categories. Additional reflections on how this change affects the forecasts are outlined in the section on changes since the last EVI below.

One important caveat with regards to this outcome measure is that this indicator captures *both* electoral violence *and* intimidation/harassment by the government. One consequence of this is that elections where the government repression has been so severe as to cause an artificially calm election with no outbursts of visible violence are also coded as having 'severe' levels of electoral violence since the government repression

is so severe. Another implication of this definition is that countries with flawed elections are almost certain to experience electoral violence and thus yielding high levels of forecasted risk. To account for this, we present the true forecasts separately for countries deemed by the VDEM institute to be democracies separately from those deemed to be autocracies.

One alternative way of dealing with this potential problem is to separate the forecast into government harassment and/or intimidation as one category and electoral violence by other actors as a second category. Each country would, in that case, get a separate forecast for each of these two outcomes. This would make communication about the outcome more complex, but perhaps more in line with how electoral violence is conceptualized outwardly.

2.1   Forecasting the outcome

Since the outcome of electoral violence is ordinal the forecasts for the outcome show the probability that each election end up in the three categories of 'no electoral violence/harassment', 'limited electoral violence/harassment', and 'severe electoral violence/harassment'. In order to facilitate an easier interpretation of the results, two additional measures are also presented. First, the likelihood that *any* electoral violence occurs for the election, i.e. simply the sum of the probabilities of 'limited' and 'severe' electoral violence/harassment. Second, we also present a 'risk index' scaled from 0 to 1 where 0 indicates the least risk of electoral violence and 1 the highest risk of electoral violence. The risk index is constructed by simply taking the probability of 'severe' electoral violence and adding 0.5 x the probability of 'limited' electoral violence. We make these forecasts up to two years into the future.

3   Data and methodology

The forecasting methodology used for the project is anchored in the methodologies used by the Violence Early Warning System (ViEWS) to predict violence from armed conflict (Hegre et al., 2019; Hegre et al., 2021). To this end, train machine learning models on historical data with features (predictor variables) grouped into broad thematic *constituent models*. The predictions of these constituent models are then combined into an *ensemble* using a genetic algorithm to find the optimal weights. We use this ensemble to produce the final forecasts. In total, we tested 33 thematic constituent models grouped into five different overarching themes:

- Constituent models using features from election-level data from the last held election in the VDEM election-level data set
- Constituent models using features from the VDEM country year data set
- Constituent models using features from the digital societies project
- Constituent models using features from the World Bank's World Development indicators
- Constituent models using a combination of features from the above mentioned data sources

A list of all constituent models and which indicators each model contains can be found in Appendix B2, which also details which theme each constituent model corresponds to.

For each model, the features are taken as the last value in the last available calendar year. I.e. for forecasts one calendar year into the future, the values are taken from December of the previous year and for forecasts two calendar years into the future, the values are taken from December two calendar years back in time. Most features are only updated annually, and in these cases the features are simply lagged one or two years respectively, but for features such as election-related variables the values are taken to be the last observed value in the last year with available data. Missing data are replaced by filling the last observed value forward. This means that the EVI forecasting system can be considered a medium-term forecasting tool which forecasts the structural risk of electoral violence, but which does not take into account election dynamics in the individual election (e.g. the candidates records on encouragement of violence, or riots and retaliation escalation loops).

We use all national level elections to either the presidency or the lower house (or combinations of

elections which included at least one of these) in the period 1989-2024 (last available data) and coded by VDEM as the training data for the true forecasts, which we then produce for 2025 and 2026, i.e. two calendar years into the future. 2025 can thus be seen as a side-casted year since the model does not have access to the data from 2025 but we can evaluate how well the model did on the elections which have already been held. In the evaluation of the models (more below) we split the training data into different training and test periods to produce out-of-sample forecasts.

The process for producing the forecasts is detailed in Figure 1 below:
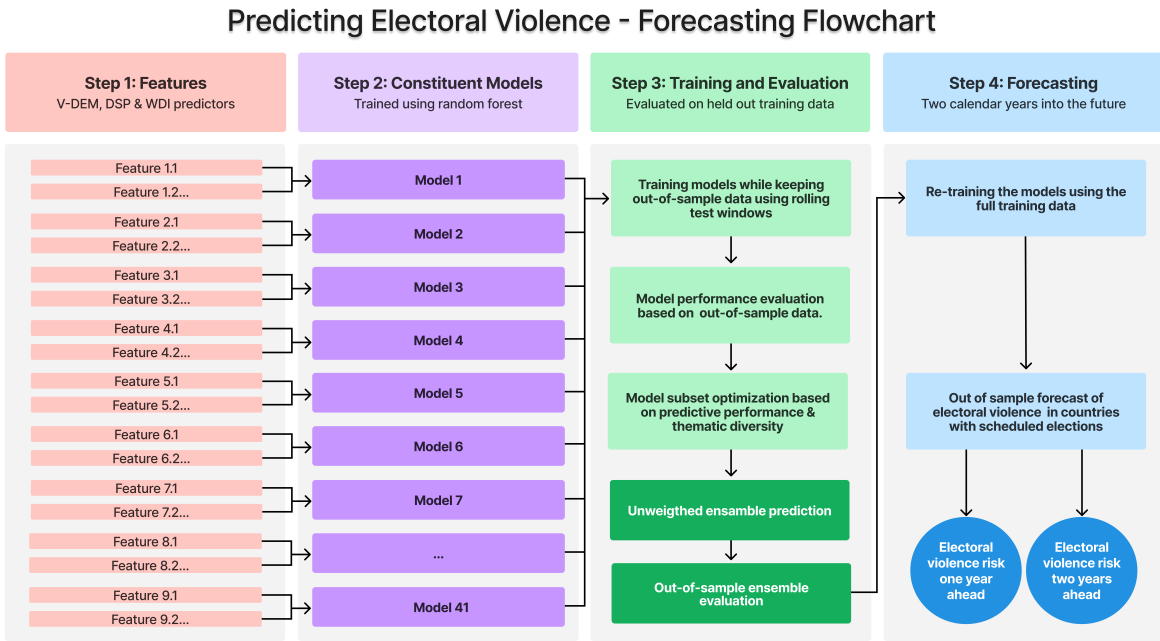

Figure 1. Forecasting flowchart

3.1   Forecasting models

To make the forecasts for electoral violence, we train the constituent models using a random forest classifier(Breiman, 2001). The goal is thereby to maximize the predictive performance of the model, rather than optimizing for inferences that can be drawn from the models. This type of model can handle effects that are non-linear and effects that differ depending on other characteristics in the model (interactions). Random forest algorithms have been shown to work well for predicting similar outcomes, for instance within the VIEWS project (Hegre et al., 2021). In a previous iteration of the EVI, we also used an extreme gradient boosted classifier (xgboost). However, due to similar predictive performance and other complexities in the xgboost model, such as the tuning of hyperparameters, we decided to proceed with only the random forest classifier for the EVI tool. The random forest classifier is less prone to overfitting than the xgboost model and is also less computationally demanding.

As the project uses a multitude of constituent models, these need to be weighted together in order to produce a final prediction. We do this by using a genetically weighted ensemble which optimizes the *Brier* score of the ensemble in the rolling test window. The genetic algorithm is run separately for each year and horizon, producing differently weighted ensembles for each year and horizon. In the 2025-iteration of the EVI the both the one and two-year ahead ensemble 10 constituent models with non-zero weights, while the two-years ahead ensemble uses The models included in the ensemble, including their relative weights, are shown in tables B1-B2 in Appendix B. The genetically weighted ensembles rely heavily on the irregularities and/or characteristics of the last election, including the reported level of electoral violence for the last election. The remaining weights are distributed among models that include a range of different structural

features, such as VDEM mid- and low-level indices, WDI structural indicators, and models which contain features from the digital society project (DSP). The use of genetically weighted ensembles was introduced in the 2024 iteration of the EVI, in contrast to the first report, where we made use of an unweighted ensemble of the best 9 models among our constituent models.

3.2 Evaluation of models

The predictive performance of the constituent models and the ensembles were evaluated using a range of standard evaluation statistics. More specifically, the accuracy, brier, area under the precision recall curve (AUPR) and area under the receiver operator characteristic curve (AUROC) scores were computed. For the AUPR and AUROC scores, the scores are the unweighted averages (i.e. macro) of the three one-against-all AUPR/AUROC scores. Accuracy measures the proportion of cases which are correctly classified on our three-level ordinal scale when the most likely level of electoral violence is taken as the prediction. The brier score, on the other hand, is a measure of the squared error, in terms of (decimal) probability for the model. AUPR measures the performance of the model when trading off the precision, i.e. the proportion of correctly classified cases among predicted positive cases, with recall, i.e. the proportion of all positive cases correctly predicted. AUROC is similar to AUPR but measures the performance of the model when trading off the recall with the false positive rate, i.e. the proportion of predicted positive cases which are in fact negative. Accuracy, AUPR and AUROC all theoretically range from 0 to 1 where 0 is the worst performing model and 1 is the best performing model. The Brier score also range from 0 to 1, but for this score a lower value indicate better performance (Zhou et al., 2021).

Crucial when evaluating machine learning models such as the random forest models is that the evaluation happens on data which the model has not seen, since the models tend to overfit (i.e. perform artificially well) on data which are within the sample. To ensure that the constituent models were evaluated on out-of-sample data, we evaluated the performance using a rolling test window. Here, the period 2011-2024 was treated as the test period. For each year, the models were trained using all years prior, and predictions were made one year into the future. I.e. for 2011 the period 1989-2010 was used as the training data, while for 2024 the period 1989-2023 was used as training data.

3.3 Evaluation results

The results of the evaluation of the individual constituent models as well as the genetically weighted algorithm can be seen in Appendix C Tables B3 and B4 for one and two years ahead predictions respectively. The lists in Appendix C1 are sorted on the *brier-score* metric, where a lower value indicates a better performance. The results show that the best performing models have accuracies between 77-83%, brier scores of 0.12-0.14. and AUPR and AUROC scores of 0.82-0.85 and 0.92-0.94 for both, the one and two year ahead forecasts.

Worth noting here is that the best performing models all include variables relating to the characteristics of the last held election, including the level of violence of that election. This is in line with results from other conflict forecasting efforts, such as the ViEWS project, where conflict history is usually the best predictor of future violence. The genetically weighted ensemble has the best performance on the brier core, AUPR, and AUROC metrics and among the highest performance on the accuracy metrics.

In Figure 2 below, we show the performance metrics for the one and two years ahead forecasts in the rolling test window (rw1 and rw2) specification for the genetically weighted ensemble as well as the full digital society project model.
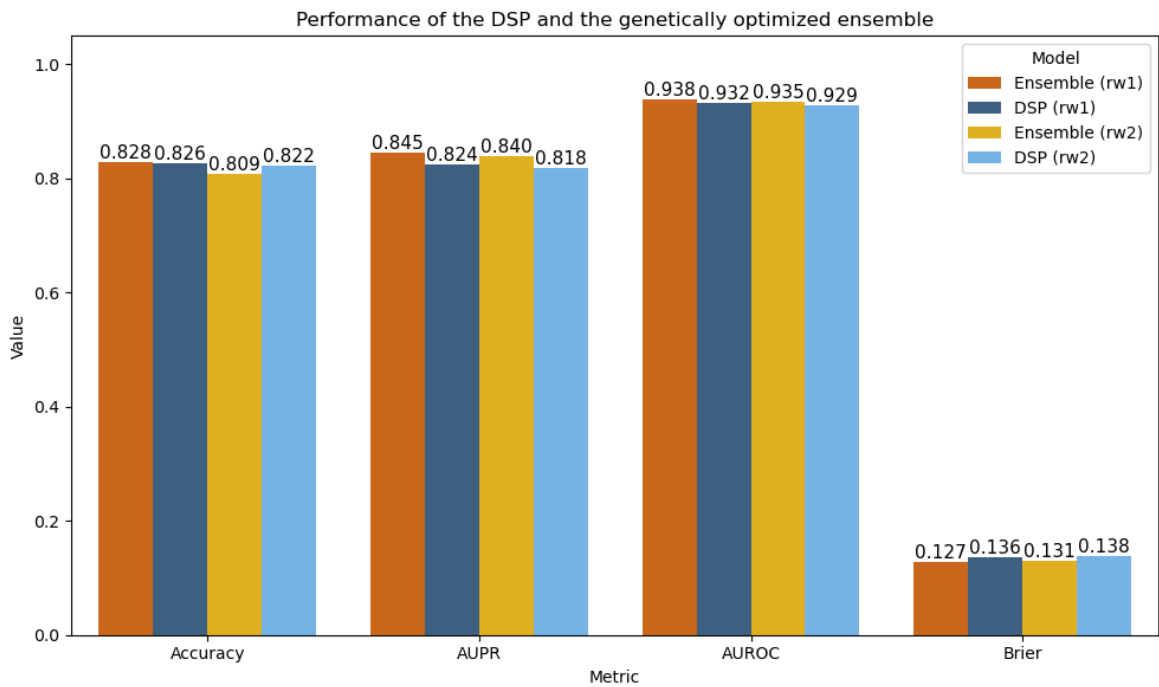
Figure 2. Performance comparison of random forest models for the one (rw1) and two (rw2) years ahead forecasts in the rolling test window

4 Methodological changes in the 2025 iteration of the EVI

There are two main methodological updates in the 2025 iteration of the EVI compared to the 2024 version: 1. A revision of how the genetic algorithm is trained during the evaluation period. 2. A transition from the `mlr3` forecasting framework to a custom, manually coded R implementation.

4.1 Revising the training of the genetic algorithm during the evaluation period

In previous versions of the EVI, the genetic algorithm used to weight the ensemble was trained on the full evaluation period. In the 2025 iteration, we modify this procedure to ensure that, for each year in the evaluation period, the ensemble weights are learned using only data available prior to that year. Additionally, the genetic algorithm is now run separately for each evaluation year, rather than once for the entire period.

Motivation: This change ensures that the evaluation of the EVI better reflects the real-world forecasting conditions under which the model operates, where future data is not available at the time of prediction. Running the genetic algorithm year-by-year with only prior data strengthens the validity of the evaluation metrics by eliminating potential look-ahead bias. It also better aligns the evaluation setup with the operational forecasting setup.

Consequences: This change has no impact on the actual forecasts for future elections, which are always based on data available at the time of prediction. However, it does make the evaluation results more principled and reliable. One implication of this change is that forecast comparability across years in the evaluation period is slightly affected, as part of the differences in model output may stem from this updated methodology rather than changes in the underlying data.

4.2 Transition to a custom forecasting pipeline

In the 2024 iteration, the EVI forecasts were produced using the `mlr3` machine learning framework. In the 2025 iteration, we have transitioned to a manually coded forecasting system in R.

Motivation: The switch to a custom implementation allows for greater flexibility and transparency in the forecasting pipeline. It also enables more direct control over output formatting, diagnostics, and ensemble

---

construction.

Consequences: This change has minimal impact on the forecasts themselves, as the modeling procedures remain largely the same. However, the custom pipeline simplifies internal development and debugging, and it facilitates a more streamlined integration of future methodological updates.

4.3   Brief note on previous methodological changes

The current methodological updates to the EVI in 2025 build on a foundation laid in earlier iterations of the index. In the 2024 version, key adjustments included a refinement of the definitions used to classify levels of electoral violence, the discontinuation of the xgboost classifier, and the replacement of the unweighted, naive ensemble with a genetically weighted ensemble. These changes were primarily aimed at: 1. Improving the conceptual clarity of electoral violence and better distinguishing cases with and without electoral violence. 2. Reducing the complexity of the forecasting system by eliminating redundant steps in the procedure without compromising performance. 3. Enhancing the performance of the ensemble and increasing the interpretability of the prediction drivers.

For a detailed overview of the methodological changes introduced prior to the 2025 iteration, we refer the reader to the corresponding section of the 2024 EVI report.

References

Breiman, Leo (2001). "Random forests". In: *Machine learning* 45, pp. 5–32.

Coppedge, Michael et al. (2021a). *V-Dem Codebook v11*.

Coppedge, Michael et al. (2021b). *V-Dem Country-Year/Country-Date Dataset v11*.

Fjelde, Hanne et al. (2021). *The Deadly Electoral Conflict Dataset (DECO) Codebook*.

Hegre, Håvard et al. (2019). "ViEWS: A political violence early-warning system". In: *Journal of peace research* 56.2, pp. 155–174.

Hegre, Håvard et al. (2021). "ViEWS2020: revising and evaluating the ViEWS political violence early-warning system". In: *Journal of peace research* 58.3, pp. 599–611.

Mechkova, Valeriya et al. (2022). *Digital Society Project Dataset v4*.

Zhou, Jianlong et al. (2021). "Evaluating the quality of machine learning explanations: A survey on methods and metrics". In: *Electronics* 10.5, p. 593.

Appendix A: The outcome of electoral violence

Definition of v2elintim:

Question: In this national election, were opposition candidates/parties/campaign workers subjected to repression, intimidation, violence, or harassment by the government, the ruling party, or their agents?

Clarification: Other types of clearly distinguishable civil violence, even if politically motivated, during the election period should not be factored in when scoring this indicator (it is dealt with separately).

Responses:

0: Yes. The repression and intimidation by the government or its agents was so strong that the entire period was quiet.

1: Yes, frequent: There was systematic, frequent and violent harassment and intimidation of the opposition by the government or its agents during the election period.

2: Yes, some. There was periodic, not systematic, but possibly centrally coordinated — harassment and intimidation of the opposition by the government or its agents.

3: Restrained. There were sporadic instances of violent harassment and intimidation by the government or its agents, in at least one part of the country, and directed at only one or two local branches of opposition groups.

4: None. There was no harassment or intimidation of opposition by the government or its agents, during the election campaign period and polling day.

Definition of v2elpeace:

Question: In this national election, was the campaign period, election day, and post-election process free from other types (not by the government, the ruling party, or their agents) of violence related to the conduct of the election and the campaigns (but not conducted by the government and its agents)?

Responses:

0: No. There was widespread violence between civilians occurring throughout the election period, or in an intense period of more than a week and in large swaths of the country. It resulted in a large number of deaths or displaced refugees.

1: Not really. There were significant levels of violence but not throughout the election period or beyond limited parts of the country. A few people may have died as a result, and some people may have been forced to move temporarily.

2: Somewhat. There were some outbursts of limited violence for a day or two, and only in a small part of the country. The number of injured and otherwise affected was relatively small.

3: Almost. There were only a few instances of isolated violent acts, involving only a few people; no one died and very few were injured.

4: Peaceful. No election-related violence between civilians occurred.

Definition of electoral violence target:

Recoding of v2elintim and v2elpeace into three level ordinal variables with values [0-1.5] corresponding to 'Severe electoral violence/intimidation', values (1.5-3) corresponding to 'limited electoral violence/intimidation', and value [3-4] corresponding to 'no electoral violence/intimidation' perpetrated by government actors (v2elintim) and non-government actors (v2elpeace) respectively. The target 'electoral violence' is then taken as the max of the two three level ordinal variables measuring electoral violence/intimidation.

Appendix B1: Constituent models in the final ensemble

The final ensemble for the predictions consists of a genetically weighted ensemble. The ensembles rely heavily on the irregularities and/or characteristics of the last election, including the reported level of electoral violence for the last election. The remaining weights are distributed among models that include a range of different structural features, such as VDEM mid- and low-level indices, WDI structural indicators, and models which contain features from the digital society project (DSP). The exact constituent models and their weights in the one- and two-years ahead forecasts for all years are shown in table B1-B2 below.

Table B1. Weights for the genetic algorithm 2015-2025 for the one year ahead forecasts. Average weight as the last column

	Model Name	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	Avg. w
1	History of electoral violence (full)	0.000	0.330	0.483	0.417	0.431	0.265	0.276	0.181	0.000	0.057	0.061	0.227
2	VDEM full model	0.000	0.000	0.058	0.067	0.191	0.256	0.288	0.071	0.197	0.181	0.076	0.126
3	Election Irregularities last election (long)	0.000	0.025	0.122	0.261	0.069	0.000	0.068	0.068	0.157	0.141	0.076	0.090
4	History of electoral violence (history only)	0.000	0.000	0.000	0.000	0.000	0.132	0.057	0.180	0.194	0.207	0.195	0.088
5	Election Irregularities last election (short)	0.000	0.091	0.046	0.000	0.102	0.145	0.056	0.068	0.088	0.000	0.198	0.072
6	VDEM Mid level indicies and WDI structural	0.094	0.022	0.000	0.026	0.000	0.000	0.000	0.156	0.060	0.176	0.188	0.066
7	Election Irregularities (last election), VDEM exclusion, and WDI structural	0.248	0.272	0.122	0.000	0.000	0.000	0.030	0.000	0.000	0.026	0.000	0.063
8	Election Characteristics last election (full)	0.000	0.000	0.063	0.055	0.027	0.042	0.042	0.030	0.000	0.048	0.000	0.028
9	Full VDEM, WDI, and DSP model	0.071	0.000	0.000	0.042	0.027	0.000	0.000	0.000	0.146	0.000	0.000	0.026
10	VDEM Mid level indicies, WDI structural, and DSP infrastructure	0.077	0.000	0.000	0.000	0.000	0.000	0.056	0.000	0.000	0.042	0.085	0.024
11	Election Irregularities (last election), VDEM civil liberties, and WDI structural	0.254	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023
12	VDEM mid level indicies	0.000	0.000	0.000	0.000	0.067	0.060	0.039	0.063	0.000	0.000	0.000	0.021
13	VDEM mid level indicies	0.000	0.113	0.000	0.000	0.086	0.000	0.000	0.000	0.000	0.000	0.029	0.021
14	DSP full model	0.000	0.000	0.000	0.000	0.000	0.032	0.036	0.000	0.138	0.000	0.000	0.019
15	VDEM Neopatrimonialism	0.000	0.035	0.039	0.069	0.000	0.027	0.028	0.000	0.000	0.000	0.000	0.018
16	VDEM High level indicies	0.044	0.075	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.015
17	VDEM Political Exclusion Indicies	0.000	0.036	0.000	0.000	0.000	0.000	0.000	0.057	0.000	0.000	0.051	0.013
18	DSP Social Media Climate, security	0.000	0.000	0.066	0.000	0.000	0.000	0.000	0.052	0.000	0.025	0.000	0.013
19	VDEM High level indivies and WDI structural	0.078	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.000	0.010
20	DSP Infra	0.092	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008
21	Election Characteristics last election (full)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.070	0.000	0.008
22	DSP Disinformation, social climate and usage	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.051	0.000	0.000	0.000	0.007
23	DSP Social media climate	0.000	0.000	0.000	0.000	0.000	0.040	0.000	0.022	0.000	0.000	0.000	0.006
24	WDI Structural	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
25	WDI full model	0.000	0.000	0.000	0.038	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
26	Election Characteristics last election, structural	0.000	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
27	DSP Monitoring	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
28	DSP Disinformation and social media usage	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	VDEM Accountability Indicies	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
30	VDEM Civil Liberties Indicies	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
31	VDEM Gender	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
32	WDI Education	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
33	WDI Resources	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table B2. Weights for the genetic algorithm 2015-2026 for the two year ahead forecasts. Average weight as the last column

	Model name	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	Avg. w
1	History of electoral violence (full)	0.272	0.349	0.356	0.531	0.331	0.283	0.291	0.048	0.082	0.045	0.043	0.000	0.219
2	Election Irregularities last election (short)	0.000	0.000	0.160	0.104	0.294	0.165	0.208	0.174	0.119	0.367	0.133	0.200	0.160
3	Election Irregularities last election (long)	0.000	0.094	0.045	0.100	0.215	0.253	0.125	0.423	0.233	0.084	0.105	0.164	0.153
4	Full VDEM, WDI, and DSP model	0.000	0.060	0.000	0.026	0.038	0.000	0.000	0.070	0.406	0.145	0.122	0.071	0.078
5	Election Irregularities (last election), VDEM exclusion, and WDI structural	0.373	0.049	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.000	0.226	0.132	0.067
6	Election Characteristics last election (full)	0.000	0.000	0.162	0.060	0.000	0.148	0.084	0.000	0.000	0.067	0.119	0.064	0.059
7	History of electoral violence (history only)	0.000	0.276	0.191	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.025	0.044
8	VDEM Mid level indicies and WDI structural	0.000	0.000	0.000	0.000	0.030	0.000	0.073	0.046	0.000	0.106	0.132	0.105	0.041
9	DSP full model	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.106	0.060	0.048	0.065	0.116	0.033
10	WDI full model	0.023	0.075	0.032	0.071	0.049	0.000	0.000	0.070	0.068	0.000	0.000	0.000	0.032
11	VDEM Mid level indicies, WDI structural, and DSP infrastructure	0.000	0.000	0.000	0.000	0.000	0.072	0.000	0.000	0.000	0.000	0.032	0.100	0.017
12	DSP Infra	0.157	0.000	0.035	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016
13	VDEM mid level indicies	0.000	0.000	0.000	0.000	0.000	0.000	0.076	0.040	0.000	0.041	0.000	0.000	0.013
14	DSP Disinformation, social climate and usage	0.000	0.000	0.000	0.000	0.000	0.046	0.037	0.000	0.000	0.045	0.000	0.000	0.011
15	VDEM full model	0.000	0.034	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.052	0.000	0.000	0.010
16	Election Characteristics last election (full)	0.000	0.000	0.000	0.000	0.044	0.000	0.040	0.000	0.000	0.000	0.000	0.024	0.009
17	Election Characteristics last election, structural	0.092	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008
18	Election Irregularities (last election), VDEM civil liberties, and WDI structural	0.000	0.031	0.000	0.027	0.000	0.000	0.035	0.000	0.000	0.000	0.000	0.000	0.008
19	DSP Social Media Climate, security	0.000	0.000	0.000	0.081	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007
20	DSP Disinformation and social media usage	0.048	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.000	0.006
21	VDEM Gender	0.036	0.032	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006
22	DSP Social media climate	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.000	0.000	0.003
23	VDEM Neopatrimonialism	0.000	0.000	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
24	DSP Monitoring	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25	VDEM Accountability Indicies	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
26	VDEM Civil Liberties Indicies	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	VDEM Political Exclusion Indicies	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
28	VDEM High level indicies and WDI structural	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	VDEM High level indicies	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
30	VDEM mid level indicies	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
31	WDI Education	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
32	WDI Resources	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
33	WDI Structural	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Appendix B2: Model Performance

Rank	Model	Brier	Accuracy	AUROC	AUPR
1	Ensemble	0.127	0.828	0.938	0.845
2	VDEM Mid level indicies, WDI structural & DSP infrastructure	0.129	0.821	0.931	0.823
3	Election Irregularities last election (long)	0.129	0.824	0.934	0.836
4	Election Characteristics last election (full)	0.130	0.830	0.934	0.838
5	Election Irregularities last election (short)	0.130	0.826	0.934	0.840
6	VDEM Mid level indicies & WDI structural	0.130	0.828	0.928	0.822
7	Election Characteristics last election (full)	0.132	0.830	0.932	0.834
8	Election Irregularities (last election), VDEM exclusion & WDI structural	0.132	0.831	0.924	0.812
9	VDEM full model	0.132	0.824	0.933	0.843
10	Election Irregularities (last election), VDEM civil liberties & WDI structural	0.133	0.838	0.923	0.805
11	History of electoral violence (full)	0.133	0.819	0.929	0.825
12	VDEM mid level indicies	0.135	0.824	0.932	0.847
13	History of electoral violence (history only)	0.135	0.838	0.915	0.807
14	DSP full model	0.136	0.826	0.932	0.824
15	VDEM High level indivies and WDI structural	0.136	0.831	0.929	0.824
16	VDEM mid level indicies	0.144	0.805	0.919	0.816
17	Election Characteristics last election, structural	0.148	0.808	0.918	0.817
18	DSP Infra	0.150	0.793	0.913	0.793
19	DSP Social Media Climate, security	0.150	0.805	0.918	0.809
20	DSP Disinformation & social climate & usage	0.156	0.795	0.908	0.802
21	DSP Disinformation & social media usage	0.156	0.785	0.907	0.796
22	DSP Social media climate	0.156	0.791	0.906	0.795
23	VDEM Political Exclusion Indicies	0.158	0.765	0.908	0.800
24	VDEM High level indicies	0.158	0.789	0.905	0.786
25	VDEM Neopatrimonialism	0.169	0.767	0.886	0.760
26	Full model	0.169	0.781	0.866	0.784
27	DSP Monitoring	0.173	0.771	0.888	0.761
28	WDI Structural	0.174	0.753	0.887	0.709
29	VDEM Accountability Indicies	0.176	0.748	0.884	0.760
30	VDEM Civil Liberties Indicies	0.197	0.714	0.859	0.706
31	WDI full model	0.204	0.757	0.810	0.688
32	WDI Education	0.216	0.681	0.801	0.642
33	VDEM Gender	0.221	0.675	0.828	0.661
34	WDI Resources	0.234	0.640	0.799	0.636

Table B3. Performance of models in the one year ahead prediction task

Rank	Model	Accuracy	Brier	AUROC	AUPR
1	Election Irregularities last election (short)	0.130	0.821	0.934	0.840
2	Ensemble	0.131	0.809	0.935	0.840
3	Election Characteristics last election (full)	0.132	0.826	0.933	0.837
4	Election Irregularities last election (long)	0.132	0.811	0.931	0.840
5	VDEM Mid level indicies, WDI structural & DSP infrastructure	0.132	0.824	0.923	0.816
6	Election Characteristics last election (full)	0.134	0.824	0.931	0.831
7	VDEM Mid level indicies & WDI structural	0.135	0.819	0.922	0.812
8	Election Irregularities (last election), VDEM exclusion & WDI structural	0.135	0.819	0.924	0.809
9	Election Irregularities (last election), VDEM civil liberties & WDI structural	0.135	0.835	0.923	0.807
10	VDEM full model	0.136	0.822	0.929	0.830
11	DSP full model	0.138	0.822	0.929	0.818
12	VDEM mid level indicies	0.138	0.817	0.928	0.840
13	History of electoral violence (full)	0.140	0.811	0.927	0.826
14	VDEM High level indivies & WDI structural	0.141	0.802	0.922	0.820
15	VDEM mid level indicies	0.147	0.797	0.917	0.822
16	Election Characteristics last election, structural	0.148	0.796	0.915	0.803
17	DSP Social Media Climate, security	0.151	0.799	0.916	0.802
18	DSP Infra	0.153	0.798	0.906	0.783
19	History of electoral violence (history only)	0.154	0.797	0.904	0.782
20	DSP Disinformation & social climate and usage	0.156	0.787	0.907	0.794
21	DSP Disinformation & social media usage	0.156	0.799	0.906	0.797
22	DSP Social media climate	0.156	0.791	0.907	0.792
23	VDEM High level indicies	0.161	0.769	0.903	0.785
24	VDEM Political Exclusion Indicies	0.166	0.742	0.899	0.764
25	WDI Structural	0.168	0.748	0.894	0.716
26	Full model	0.172	0.781	0.863	0.780
27	VDEM Neopatrimonialism	0.176	0.742	0.879	0.747
28	DSP Monitoring	0.177	0.748	0.884	0.740
29	VDEM Accountability Indicies	0.178	0.744	0.878	0.746
30	WDI full model	0.198	0.757	0.825	0.700
31	VDEM Civil Liberties Indicies	0.199	0.700	0.855	0.703
32	VDEM Gender	0.215	0.688	0.836	0.682
33	WDI Education	0.219	0.676	0.798	0.623
34	WDI Resources	0.235	0.649	0.792	0.627

Table B4. Performance of models in the two year ahead prediction task

Appendix C1: Constituent models description for models in the final ensembles

This appendix briefly describes the 33 thematic constituent models, as well as the abbreviated names of the features included and the performance and average weight of the models in the final ensemble. The constituent models are ordered by their performance on the Brier score in the one-year ahead prediction task. The full names of the features are provided in Appendix A2.

1. VDEM Mid level indicies, WDI structural, and DSP infrastructure:

Description: This model is a combination of features from VDEM mid level indices, WDI structural, and DSP infrastructure. It includes a wide range of features related to democracy, governance, and development, and is aimed at capturing the overall state of a country's political, economic, and digital institutions.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.821	0.129	0.931	0.823	0.024
Two years ahead	0.824	0.132	0.923	0.816	0.017

Included features: v2x_accountability, v2x_neopat, v2x_civlib, v2x_gender, v2x_corr, v2x_rule, v2xcs_ccsi, v2xps_party, v2x_divparctrl, v2x_feduni, sp.pop.totl, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.imrt.in, sp.dyn.le00.in, sp.pop.0014.fe.zs, sp.pop.grow, sp.pop.65up.fe.zs, v2smonex, v2elfrcamp, v2mecrit, v2merange, v2elembaut, v2smgovcapsec, v2smpolcap, v2smregcap, v2smgovfilprc, v2smgovsmmon, v2smgovsmcenprc, v2smarrest, it.net.user.zs

2. Election Irregularities last election (long):

Description: This model contains a broad set of features related to election irregularities from the last election, as well as the history of electoral violence. It includes a wide range of features related to election violence, fraud, and irregularities, and is aimed at capturing the overall state of a country's electoral integrity.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.824	0.129	0.934	0.836	0.090
Two years ahead	0.811	0.132	0.931	0.840	0.153

Included features: cons_elect, peaceful_streak, violent_streak, lowviolent_streak, v2elintim_osp, v2elpeace_osp, v2elembaut, v2elembcap, v2elmulpar, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrfair, v2elmonden, v2elmonref

3. Election Characteristics last election (full):

Description: This model is a broad model encompassing all features related to election characteristics from the last election. It includes a wide range of features related to suffrage, electoral integrity, and campaign finance, and is aimed at capturing the overall state of a country's electoral system.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.830	0.130	0.934	0.838	0.008
Two years ahead	0.826	0.132	0.933	0.837	0.059

Included features: v2asuffrage, v2elcomvot, v2elgvsuflvl, v2eldonate, v2elpubfin, v2elembaut, v2elembcap, v2elmulpar, v2elrgstry, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrcamp, v2elpdcamp, v2elpaidig, v2elfrfair, v2eldommon, v2elintmon, v2elmonden, v2elmonref, v2elaccept, v2elasmoff, v2elvaptrn, cons_elect, peaceful_streak, violent_streak, lowviolent_streak, v2elintim_osp, v2elpeace_osp

4. Election Irregularities last election (short):

Description: This model contains a subset of features related to election irregularities from the last election. It includes a wide range of features related to election violence, fraud, and irregularities, and is aimed at capturing the overall state of a country's electoral integrity.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.826	0.130	0.934	0.840	0.072
Two years ahead	0.821	0.130	0.934	0.840	0.160

Included features: v2elembaut, v2elembcap, v2elmulpar, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrfair, v2elmonden, v2elmonref

Included features: v2elembaut, v2elembcap, v2elmulpar, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrfair, v2elmonden, v2elmonref

5. VDEM Mid level indicies and WDI structural:

Description: This model contains a broad set of features related to the V-Dem mid-level indices and the World Development Indicators structural indicators. It includes a wide range of features related to democracy, governance, and development, and is aimed at capturing the overall state of a country's political and economic institutions.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.828	0.130	0.928	0.822	0.066
Two years ahead	0.819	0.135	0.922	0.812	0.041

Included features: v2x_accountability, v2x_neopat, v2x_civlib, v2x_gender, v2x_corr, v2x_rule, v2xcs_ccsi, v2xps_party, v2x_divparctrl, v2x_feduni, sp.pop.totl, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.imrt.in, sp.dyn.le00.in, sp.pop.0014.fe.zs, sp.pop.grow, sp.pop.65up.fe.zs

6. Election Characteristics last election (full):

Description: This model is a broad model containing all features related to election characteristics from the last election. It includes a wide range of features related to suffrage, electoral integrity, and election violence, and is aimed at capturing the overall state of a country's electoral system.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.830	0.132	0.932	0.834	0.008
Two years ahead	0.824	0.134	0.931	0.831	0.009

Included features: v2asuffrage, v2elcomvot, v2elgvsuflvl, v2eldonate, v2elpubfin, v2elembaut, v2elembcap, v2elmulpar, v2elrgstry, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrcamp, v2elpdcamp, v2elpaidig, v2elfrfair, v2eldommon, v2elintmon, v2elmonden, v2elmonref, v2elaccept, v2elasmoff, v2elvaptrn, cons_elect, peaceful_streak, violent_streak, lowviolent_streak, v2elintim_osp, v2elpeace_osp

7. Election Irregularities (last election), VDEM exclusion, and WDI structural:

Description: This model is a combination model containing features related to election irregularities from the last election, V-Dem exclusion indices, and the World Development Indicators structural indicators. It includes a wide range of features related to election violence, electoral integrity, and exclusion, and is aimed at capturing the overall state of a country's electoral and economic system, as well as specifically the exclusion of certain groups from the political process.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.831	0.132	0.924	0.812	0.063
Two years ahead	0.819	0.135	0.924	0.809	0.067

Included features: v2asuffrage, v2elcomvot, v2elgvsuflvl, v2eldonate, v2elpubfin, v2elembaut, v2elembcap, v2elmulpar, v2elrgstry, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrcamp, v2elpdcamp, v2elpaidig, v2elfrfair, v2eldommon, v2elintmon, v2elmonden, v2elmonref, v2elaccept, v2elasmoff, v2elvaptrn, cons_elect, peaceful_streak, violent_streak, lowviolent_streak, v2elintim_osp, v2elpeace_osp, v2xpe_exlecon, v2xpe_exlgender, v2xpe_exlgeo, v2xpe_exlpol, v2xpe_exlsocgr, sp.pop.totl, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.imrt.in, sp.dyn.le00.in, sp.pop.0014.fe.zs, sp.pop.grow, sp.pop.65up.fe.zs

8. VDEM full model:

Description: This model is a broad model containing all v2x features from the Varieties of Democracy dataset. It includes a wide range of features related to democracy, electoral integrity, and exclusion, and is aimed at capturing the overall state of a country's political system.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.824	0.132	0.933	0.843	0.126
Two years ahead	0.822	0.136	0.929	0.830	0.010

Included features: v2x_freexp_altinf, v2x_frassoc_thick, v2x_suffr, v2x_delibdem, v2xel_frefair, v2x_elecoff, v2xcl_rol, v2x_jucon, v2xlg_legcon, v2x_cspart, v2xdd_dd, v2xel_locelec, v2xel_regelec, v2xdl_delib, v2xeg_eqprotec, v2xeg_eqaccess, v2x_veracc, v2x_diagacc, v2x_horacc, v2x_ex_confidence, v2x_ex_direlect, v2x_ex_hereditary, v2x_ex_military, v2x_ex_party, v2xnp_client, v2xnp_pres, v2xnp_regcorr, v2x_clphy, v2x_clpol, v2x_clpriv, v2xpe_exlecon, v2xpe_exlgender, v2xpe_exlgeo, v2xpe_exlpol, v2xpe_exlsocgr, v2x_corr, v2x_execorr, v2x_pubcorr, v2x_gencl, v2x_gencs, v2x_genpp, v2x_rule, v2xdd_cic, v2xdd_i_ci, v2xdd_i_rf, v2xdd_toc, v2xdd_i_pl, v2xdd_i_or, v2xcs_ccsi, v2x_EDcomp_thick, v2xcl_disc, v2xcl_dmove, v2xcl_slave, v2xex_elecleg, v2xme_altinf, v2xps_party, v2x_divparctrl, v2x_feduni, v2xca_academ

9. Election Irregularities (last election), VDEM civil liberties, and WDI structural:

Description: This model is a combination model containing features related to election irregularities from the last election, V-Dem civil liberties, and the World Development Indicators structural indicators. It includes a wide range of features related to election violence, electoral integrity, and civil liberties, and is aimed at capturing the overall state of a country's electoral and economic system, and the respect for civil liberties.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.838	0.133	0.923	0.805	0.023
Two years ahead	0.835	0.135	0.923	0.807	0.008

Included features: v2asuffrage, v2elcomvot, v2elgvsuflvl, v2eldonate, v2elpubfin, v2elembaut, v2elembcap, v2elmulpar, v2elrgstry, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrcamp, v2elpdcamp, v2elpaidig, v2elfrfair, v2eldommon, v2elintmon, v2elmonden, v2elmonref, v2elaccept, v2elasmoff, v2elvaptrn, cons_elect, peaceful_streak, violent_streak, lowviolent_streak, v2elintim_osp, v2elpeace_osp, v2x_clphy, v2x_clpol, v2x_clpriv, v2x_civlib, sp.pop.totl, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.imrt.in, sp.dyn.le00.in, sp.pop.0014.fe.zs, sp.pop.grow, sp.pop.65up.fe.zs

10. History of electoral violence (full):

Description: This model is based on the full history of electoral violence data, including all features related to electoral violence. It is aimed at capturing the overall history of electoral violence in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.819	0.133	0.929	0.825	0.227
Two years ahead	0.811	0.140	0.927	0.826	0.219

Included features: cons_elect, peaceful_streak, violent_streak, lowviolent_streak, v2elintim_osp, v2elpeace_osp

11. VDEM mid level indicies:

Description: This model is based on the V-Dem mid level indices, including all features related to the V-Dem mid level indices. It is aimed at capturing the overall state of democracy in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.824	0.135	0.932	0.847	0.021
Two years ahead	0.817	0.138	0.928	0.840	0.013

Included features: v2x_api, v2x_mpi, v2x_freexp_altinf, v2x_frassoc_thick, v2x_suffr, v2x_delibdem, v2xel_frefair, v2x_elecoff, v2xcl_rol, v2x_jucon, v2xlg_legcon, v2x_cspart, v2xdd_dd, v2xel_locelec, v2xel_regelec, v2xdl_delib, v2xeg_eqprotec, v2xeg_eqaccess

12. History of electoral violence (history only):

Description: This model is based on the history of electoral violence data, including only the features related to electoral violence. It is aimed at capturing the overall history of electoral violence in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.838	0.135	0.915	0.807	0.088
Two years ahead	0.797	0.154	0.904	0.782	0.044

Included features: cons_elect, peaceful_streak, violent_streak, lowviolent_streak

13. DSP full model:

Description: This model is based on the full DSP model, including all features related to the DSP data. It is aimed at capturing the overall state of digital infrastructure in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.826	0.136	0.932	0.824	0.019
Two years ahead	0.822	0.138	0.929	0.818	0.033

Included features: v2smgovdom, v2smgovab, v2smpardom, v2smparab, v2smfordom, v2smforads, v2smgovfilcap, v2smgovfilprc, v2smgovshutcap, v2smgovshut, v2smgovsm, v2smgovsmalt, v2smgovsmmon, v2smgovsmcenprc, v2smgovcapsec, v2smregcon, v2smprivex, v2smprivcon, v2smregcap, v2smregapp, v2smlawpr, v2smdefabu, v2smonex, v2smonper, v2smmefra, v2smorgviol, v2smorgavgact, v2smorgelitact, v2smcamp, v2smarrest, v2smpolsoc, v2smpolhate

14. VDEM High level indicies and WDI structural:

Description: This model is based on the V-Dem high level indices and the WDI structural data, including all features related to the V-Dem high level indices and the WDI structural. It is aimed at capturing the overall state of democracy and structural economic indicators in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.831	0.136	0.929	0.824	0.010
Two years ahead	0.802	0.141	0.922	0.820	0.000

Included features: v2x_polyarchy, v2x_libdem, v2x_partipdem, v2x_delibdem, v2x_egaldem, sp.pop.totl, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.imrt.in, sp.dyn.le00.in, sp.pop.0014.fe.zs, sp.pop.grow, sp.pop.65up.fe.zs

15. VDEM mid level indicies (alternative):

Description: This model is based on the V-Dem mid level indices, including all features related to the V-Dem mid level indices. It is aimed at capturing the overall state of democracy in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.805	0.144	0.919	0.816	0.021
Two years ahead	0.797	0.147	0.917	0.822	0.000

Included features: v2x_accountability, v2x_neopat, v2x_civlib, v2x_gender, v2x_corr, v2x_rule, v2xcs_ccsi, v2xps_party, v2x_divparctrl, v2x_feduni

16. Election Characteristics last election, structural:

Description: This model is based on the structural election characteristics of the last election. It is aimed at capturing the structural characteristics of the last election in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.808	0.148	0.918	0.817	0.002
Two years ahead	0.796	0.148	0.915	0.803	0.008

Included features: v2asuffrage, v2elcomvot, v2elgvsuflvl, v2eldonate, v2elpubfin, v2elembaut, v2elembcap, v2elmulpar, v2elrgstry, v2elvotbuy, v2elfrcamp, v2elpdcamp, v2elpaidig, v2eldommon, v2elintmon, v2elvaptrn

17. DSP Infra:

Description: This model is based on the Digital Society Project infrastructure data. It is aimed at capturing the state of digital infrastructure in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.793	0.150	0.913	0.793	0.008
Two years ahead	0.798	0.153	0.906	0.783	0.016

Included features: v2smonex, v2elfrcamp, v2mecrit, v2merange, v2elembaut, v2smgovcapsec, v2smpolcap, v2smregcap, v2smgovfilprc, v2smgovsmmon, v2smgovsmcenprc, v2smarrest, it.net.user.zs

18. DSP Social Media Climate, security:

Description: This model is based on the Digital Society Project social media climate data, focusing on security. It is aimed at capturing the state of social media climate in a country, with a focus on security.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.805	0.150	0.918	0.809	0.013
Two years ahead	0.799	0.151	0.916	0.802	0.007

Included features: v2smgovdom, v2smpardom, v2smfordom, v2smonper, v2smmefra, v2smpolhate, v2smorgelitact, v2smcamp, v2smgovcapsec, v2smpolcap, v2smorgavgact

19. DSP Disinformation, social climate and usage:

Description: This model is based on the Digital Society Project disinformation data, focusing on social climate and usage. It is aimed at capturing the state of disinformation in a country, with a focus on social climate and usage.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.795	0.156	0.908	0.802	0.007
Two years ahead	0.787	0.156	0.907	0.794	0.011

Included features: v2smgovdom, v2smpardom, v2smfordom, v2smonper, v2smmefra, v2smpolsoc, v2smpolhate, v2smorgelitact, v2smcamp,

20. DSP Disinformation and social media usage:

Description: This model is based on the Digital Society Project disinformation data, focusing on social media usage. It is aimed at capturing the state of disinformation in a country, with a focus on social media usage.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.785	0.156	0.907	0.796	0.000
Two years ahead	0.799	0.156	0.906	0.797	0.006

Included features: v2smgovdom, v2smpardom, v2smfordom, v2smonper, v2smmefra, v2smpolsoc, v2smpolhate, v2smorgelitact, v2smcamp,

21. DSP Social media climate:

Description: This model is based on the Digital Society Project social media climate data. It is aimed at capturing the state of social media climate in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.791	0.156	0.906	0.795	0.006
Two years ahead	0.791	0.156	0.907	0.792	0.003

Included features: v2smgovdom, v2smpardom, v2smfordom, v2smorgelitact, v2smcamp, v2smonper, v2smmefra, v2smpolsoc, v2smpolhate

22. VDEM Political Exclusion Indicies:

Description: This model is based on the V-Dem Political Exclusion Indicies, which are based on the V-Dem dataset and focus on political exclusion.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.765	0.158	0.908	0.800	0.013
Two years ahead	0.742	0.166	0.899	0.764	0.000

Included features: v2xpe_exlecon, v2xpe_exlgender, v2xpe_exlgeo, v2xpe_exlpol, v2xpe_exlsocgr

23. VDEM High level indicies:

Description: This model is based on the V-Dem high level indices, which are based on the V-Dem dataset and focus on high level indicators of democracy. The model is aimed at capturing the state of democracy in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.789	0.158	0.905	0.786	0.015
Two years ahead	0.769	0.161	0.903	0.785	0.000

Included features: v2x_polyarchy, v2x_libdem, v2x_partipdem, v2x_delibdem, v2x_egaldem

24. VDEM Neopatrimonialism:

Description: This model is based on the V-Dem Neopatrimonialism indices, which are based on the V-Dem dataset and focus on neopatrimonialism. The model is aimed at capturing the state of neopatrimonialism in a country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.767	0.169	0.886	0.760	0.018
Two years ahead	0.742	0.176	0.879	0.747	0.002

Included features: v2x_neopat, v2xnp_client, v2xnp_pres, v2xnp_regcorr

25. Full VDEM, WDI, and DSP model:

Description: This model is a combination of all available features from the VDEM mid level indicies, WDI structural, and DSP infrastructure datasets.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.781	0.169	0.866	0.784	0.026
Two years ahead	0.781	0.172	0.863	0.780	0.078

Included features: v2smgovdom, v2smgovab, v2smpardom, v2smparab, v2smfordom, v2smforads, v2smgovfilcap, v2smgovfilprc, v2smgovshutcap, v2smgovshut, v2smgovsm, v2smgovsmalt, v2smgovsmmon, v2smgovsmcenprc, v2smgovcapsec, v2smregcon, v2smprivex, v2smprivcon, v2smregcap, v2smregapp, v2smlawpr, v2smdefabu, v2smonex, v2smonper, v2smmefra, v2smorgviol, v2smorgavgact, v2smorgelitact, v2smcamp, v2smarrest, v2smpolsoc, v2smpolhate, v2x_freexp_altinf, v2x_frassoc_thick, v2x_suffr, v2x_delibdem, v2xel_frefair, v2x_elecoff, v2xcl_rol, v2x_jucon, v2xlg_legcon, v2x_cspart, v2xdd_dd, v2xel_locelec, v2xel_regelec, v2xdl_delib, v2xeg_eqprotec, v2xeg_eqaccess, v2x_veracc, v2x_diagacc, v2x_horacc, v2x_ex_confidence, v2x_ex_direlect, v2x_ex_hereditary, v2x_ex_military, v2x_ex_party, v2xnp_client, v2xnp_pres, v2xnp_regcorr, v2x_clphy, v2x_clpol, v2x_clpriv, v2xpe_exlecon, v2xpe_exlgender, v2xpe_exlgeo, v2xpe_exlpol, v2xpe_exlsocgr, v2x_corr, v2x_execorr, v2x_pubcorr, v2x_gencl, v2x_gencs, v2x_genpp, v2x_rule, v2xdd_cic, v2xdd_i_ci, v2xdd_i_rf, v2xdd_toc, v2xdd_i_pl, v2xdd_i_or, v2xcs_ccsi, v2x_EDcomp_thick, v2xcl_disc, v2xcl_dmove, v2xcl_slave, v2xex_elecleg, v2xme_altinf, v2xps_party, v2x_divparctrl, v2x_feduni, v2xca_academ, v2asuffrage, v2elcomvot, v2elgvsuflvl, v2eldonate, v2elpubfin, v2elembaut, v2elembcap, v2elmulpar, v2elrgstry, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrcamp, v2elpdcamp, v2elpaidig, v2elfrfair, v2eldommon, v2elintmon, v2elmonden, v2elmonref, v2elaccept, v2elasmoff, v2elvaptrn, cons_elect, sp.pop.totl, ms.mil.xpnd.zs, ms.mil.xpnd.gd.zs, dt.oda.odat.pc.zs, nv.agr.totl.kn, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.le00.in, se.enr.prim.fm.zs, se.enr.prsc.fm.zs, se.prm.nenr, sh.sta.maln.zs, sh.sta.stnt.zs, sl.tlf.totl.fe.zs, sm.pop.totl.zs, sp.dyn.imrt.in, sh.dyn.mort.fe, sp.pop.0014.fe.zs, sp.pop.1564.fe.zs, sp.pop.65up.fe.zs, sp.pop.grow, sp.urb.totl.in.zs, se.xpd.totl.gb.zs, se.xpd.totl.gd.zs, sl.uem.neet.zs, ny.gdp.petr.rt.zs, ny.gdp.totl.rt.zs, it.net.user.zs

26. DSP Monitoring:

Description: This model is based on the DSP features targeting digital monitoring. It is aimed at capturing the digital surveillance in the country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.771	0.173	0.888	0.761	0.000
Two years ahead	0.748	0.177	0.884	0.740	0.000

Included features: v2smregcap, v2smgovfilprc, v2smgovsmmon, v2smgovsmcenprc, v2smarrest

27. WDI Structural:

Description: This model is based on the World Development Indicators (WDI) features targeting the structural characteristics of the country. It includes features on population, GDP, mortality, and life expectancy.

Included features: sp.pop.totl, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.imrt.in, sp.dyn.le00.in, sp.pop.0014.fe.zs, sp.pop.grow, sp.pop.65up.fe.zs

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.753	0.174	0.887	0.709	0.004
Two years ahead	0.748	0.168	0.894	0.716	0.000

28. VDEM Accountability Indicies:

Description: This model is based on the V-Dem Accountability indices. It includes features on vertical accountability, horizontal accountability, and diagonal accountability.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.748	0.176	0.884	0.760	0.000
Two years ahead	0.744	0.178	0.878	0.746	0.000

Included features: v2x_accountability, v2x_veracc, v2x_diagacc, v2x_horacc

29. VDEM Civil Liberties Indicies:

Description: This model is based on the V-Dem Civil Liberties indices. It includes features on civil liberties, political liberties, and private liberties.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.714	0.197	0.859	0.706	0.000
Two years ahead	0.700	0.199	0.855	0.703	0.000

Included features: v2x_clphy, v2x_clpol, v2x_clpriv, v2x_civlib

30. VDEM Gender:

Description: This model is based on the VDEM gender indicies and is aimed at capturing the gender equality aspects of the country.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.675	0.221	0.828	0.661	0.000
Two years ahead	0.688	0.215	0.836	0.682	0.006

Included features: v2x_gencl, v2x_gencs, v2x_genpp, v2x_gender

31. WDI Education:

Description: This model is based on the World Development Indicators (WDI) education data. It includes features on primary education completion rate, primary education enrollment rate, government expenditure on education, and secondary education enrollment rate.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.681	0.216	0.801	0.642	0.000
Two years ahead	0.676	0.219	0.798	0.623	0.000

Included features: se.enr.prim.fm.zs, se.enr.prsc.fm.zs, se.prm.nenr, se.xpd.totl.gb.zs, se.xpd.totl.gd.zs

32. WDI Resources:

Description: This model is based on the WDI resources indicators, including on economic resources and foreign aid.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.640	0.234	0.799	0.636	0.000
Two years ahead	0.649	0.235	0.792	0.627	0.000

Included features: ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, dt.oda.odat.pc.zs, ny.gdp.petr.rt.zs, ny.gdp.totl.rt.zs

33. WDI full model:

Description: This model is based on the full set of World Development Indicators (WDI) data, including features on education, resources, and structural indicators.

Performance in the rolling test window and average weight for the genetic algorithm:

	Accuracy	Brier	AUROC	AUPR	Avg. ensemble weight
One year ahead	0.757	0.204	0.810	0.688	0.003
Two years ahead	0.757	0.198	0.825	0.700	0.032

Included features: sp.pop.totl, ms.mil.xpnd.zs, ms.mil.xpnd.gd.zs, dt.oda.odat.pc.zs, nv.agr.totl.kn, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.le00.in, se.enr.prim.fm.zs, se.enr.prsc.fm.zs, se.prm.nenr, sh.sta.maln.zs, sh.sta.stnt.zs, sl.tlf.totl.fe.zs, sm.pop.totl.zs, sp.dyn.imrt.in, sh.dyn.mort.fe, sp.pop.0014.fe.zs, sp.pop.1564.fe.zs, sp.pop.65up.fe.zs, sp.pop.grow, sp.urb.totl.in.zs, se.xpd.totl.gb.zs, se.xpd.totl.gd.zs, sl.uem.neet.zs, ny.gdp.petr.rt.zs, ny.gdp.totl.rt.zs, it.net.user.zs

Appendix C2: Feature description

This subsection of the appendix includes a brief description of all features included in the constituent models. All features are derived from the Varieties of Democracy and the World Bank's World Development Indicators. For extensive information on each feature see the VDem codebook at v-dem.net/documents/55/codebook.pdf or search the World banks database at databank.worldbank.org. For reference, all indicators starting with "v2.." in the table below are VDem features (such as v2x_libdem: Liberal democracy index), while all features separated by dot's (such as ny.gdp.pcap.kd: GDP per capita (constant 2015 US$)) are from the World Bank.

Table C1. Feature descriptions

Feature	Description
cons_elect	Number of consecutive elections in the data
dt.oda.odat.pc.zs	Net official development assistance received per capita (current US$)
it.net.user.zs	Individuals using the Internet (% of population)
lowviolent_streak	Number of consecutive elections with moderate electoral violence
ms.mil.xpnd.gd.zs	Military expenditure (% of GDP)
ms.mil.xpnd.zs	Military expenditure (% of government expenditure)
nv.agr.totl.kn	Agriculture, forestry, and fishing, value added (constant LCU)
ny.gdp.pcap.kd	GDP per capita (constant 2015 US$)
ny.gdp.pcap.kd.zg	GDP per capita growth (annual %)
ny.gdp.petr.rt.zs	Oil rents (% of GDP)
ny.gdp.totl.rt.zs	Total natural resources rents (% of GDP)
peaceful_streak	Number of consecutive elections with no electoral violence
se.enr.prim.fm.zs	School enrollment, primary (gross), gender parity index (GPI)
se.enr.prsc.fm.zs	School enrollment, primary and secondary (gross), gender parity index (GPI)
se.prm.nenr	School enrollment, primary (% net)
se.xpd.totl.gb.zs	Government expenditure on education, total (% of government expenditure)
se.xpd.totl.gd.zs	Government expenditure on education, total (% of GDP)
sh.dyn.mort.fe	Mortality rate, under-5, female (per 1,000 live births)
sh.sta.maln.zs	Prevalence of underweight, weight for age (% of children under 5)
sh.sta.stnt.zs	Prevalence of stunting, height for age (% of children under 5)
sl.tlf.totl.fe.zs	Labor force, female (% of total labor force)
sl.uem.neet.zs	Share of youth not in education, employment or training, total (% of youth population)
sm.pop.totl.zs	International migrant stock (% of population)
sp.dyn.imrt.in	Mortality rate, infant (per 1,000 live births)
sp.dyn.le00.in	Life expectancy at birth, total (years)
sp.pop.0014.fe.zs	Population ages 0–14, female (% of female population)
sp.pop.1564.fe.zs	Population ages 15–64, female (% of female population)
sp.pop.65up.fe.zs	Population ages 65 and above, female (% of female population)
sp.pop.grow	Population growth (annual %)
sp.pop.totl	Population, total
sp.urb.totl.in.zs	Urban population (% of total population)
v2asuffrage	Percentage of enfranchised adults older than minimal voting age
v2elaccept	Election losers accept results
v2elasmoff	Election assume office
v2elboycot	Election boycotts
v2elcomvot	Compulsory voting
v2eldommon	Election domestic election monitors
v2eldonate	Disclosure of campaign donations
v2elembaut	Election Management Body (EMB) autonomy
v2elembcap	Election Management Body (EMB) capacity
v2elfrcamp	Election free campaign media
v2elfrfair	Election free and fair
v2elgvsuflvl	Suffrage level
v2elintim	Election government intimidation
v2elintim_osp	Election government intimidation (Original Scale)
v2elintmon	Election international monitors
v2elirreg	Election other voting irregularities
v2elmonden	Election international monitors denied
v2elmonref	Monitors refuse to be present
v2elmulpar	Elections multiparty
v2elpaidig	Election paid interest group media
v2elpdcamp	Election paid campaign advertisements
v2elpeace	Election other electoral violence
v2elpeace_osp	Election other electoral violence (Original Scale)
v2elpubfin	Public campaign finance
v2elrgstry	Election voter registry
v2elvaptrn	Election VAP turnout
v2elvotbuy	Election vote buying
v2mecrit	Print/broadcast media critical

Continued on next page

Feature	Description
v2merange	Print/broadcast media perspectives
v2smarrest	Arrests for political content
v2smcamp	Party/candidate use of social media in campaigns
v2smdefabu	Abuse of defamation and copyright law by elites
v2smforads	Foreign governments ads
v2smfordom	Foreign governments dissemination of false information
v2smgovab	Government dissemination of false information abroad
v2smgovcapsec	Government cyber security capacity
v2smgovdom	Government dissemination of false information domestic
v2smgovfilcap	Government Internet filtering capacity
v2smgovfilprc	Government Internet filtering in practice
v2smgovshut	Government Internet shut down in practice
v2smgovshutcap	Government Internet shut down capacity
v2smgovsm	Government social media censorship in practice
v2smgovsmalt	Government social media alternatives
v2smgovsmcenprc	Government social media censorship in practice
v2smgovsmmon	Government social media monitoring
v2smlawpr	Defamation protection
v2smmefra	Online media fractionalization
v2smonex	Online media existence
v2smonper	Online media perspectives
v2smorgavgact	Average people's use of social media to organize offline action
v2smorgelitact	Elites' use of social media to organize offline action
v2smorgviol	Use of social media to organize offline violence
v2smparab	Party dissemination of false information abroad
v2smpardom	Party dissemination of false information domestic
v2smpolcap	Political parties cyber security capacity
v2smpolhate	Political parties hate speech
v2smpolsoc	Polarization of society
v2smprivcon	Privacy protection by law content
v2smprivex	Privacy protection by law exists
v2smregapp	Government online content regulation approach
v2smregcap	Government capacity to regulate online content
v2smregcon	Internet legal regulation content
v2x_EDcomp_thick	Electoral component index
v2x_accountability	Accountability index
v2x_api	v2x_api
v2x_civlib	Civil liberties index
v2x_clphy	Physical violence index
v2x_clpol	Political civil liberties index
v2x_clpriv	Private civil liberties index
v2x_corr	Political corruption index
v2x_cspart	Civil society participation index
v2x_delibdem	Deliberative democracy index
v2x_diagacc	Diagonal accountability index
v2x_divparctrl	Divided party control index
v2x_egaldem	Egalitarian democracy index
v2x_elecoff	Elected officials index
v2x_ex_confidence	Confidence dimension index
v2x_ex_direlect	Direct election dimension index
v2x_ex_hereditary	Hereditary dimension index
v2x_ex_military	Military dimension index
v2x_ex_party	Ruling party dimension index
v2x_feduni	Division of power index
v2x_frassoc_thick	Freedom of association thick index
v2x_freexp_altinf	Freedom of Expression and Alternative Sources of Information index
v2x_gencl	Women civil liberties index
v2x_gencs	Women civil society participation index
v2x_gender	Women political empowerment index
v2x_genpp	Women political participation index
v2x_horacc	Party dissemination of false information abroad
v2x_jucon	Judicial constraints on the executive index
v2x_libdem	Liberal democracy index
v2x_mpi	Multiplicative polyarchy index
v2x_neopat	Neopatrimonial Rule Index
v2x_partipdem	Participatory democracy index
v2x_polyarchy	Electoral democracy index

Feature	Description
v2x_rule	Rule of law index
v2x_suffr	Share of population with suffrage
v2x_veracc	Vertical accountability index
v2xca_academ	Academic Freedom Index
v2xcl_disc	Freedom of discussion
v2xcl_dmove	Freedom of domestic movement
v2xcl_rol	Equality before the law and individual liberty index
v2xcl_slave	Freedom from forced labor
v2xcs_ccsi	Core civil society index
v2xdd_cic	Citizen-initiated component of direct popular vote index
v2xdd_dd	Direct popular vote index ordinal
v2xdd_i_ci	Popular initiative index
v2xdd_i_or	Obligatory referendum index
v2xdd_i_pl	Plebiscite index
v2xdd_i_rf	Popular referendum index
v2xdd_toc	Top-Down component of direct popular vote index
v2xdl_delib	Deliberative component index
v2xeg_eqaccess	Equal access index
v2xeg_eqprotec	Equal protection index
v2xel_frefair	Clean elections index
v2xel_locelec	Local government index
v2xel_regelec	Regional government index
v2xex_elecleg	Legislature directly elected
v2xlg_legcon	Legislative constraints on the executive index
v2xme_altinf	Alternative sources of information index
v2xnp_client	Clientelism Index
v2xnp_pres	Presidentialism Index
v2xnp_regcorr	Regime corruption
v2xpe_exlecon	Exclusion by Socio-Economic Group
v2xpe_exlgender	Exclusion by Gender index
v2xpe_exlgeo	Exclusion by Urban-Rural Location index
v2xpe_exlpol	Exclusion by Political Group index
v2xpe_exlsocgr	Exclusion by Social Group index
v2xps_party	Party institutionalization index ordinal
violent_streak	Number of consecutive elections with severe electoral violence