Forecasting Election Outcomes in Contemporary Democracies

UNIVERSITY OF BIRMINGHAM

Connor Boyd-Lyon, Evelina Ivanova, Matthew Randall, Yarkın Yorulmaz, Yeni Jung, Zehn-Ul-Abideen Sharif

Table of Contents

1. Introduction

Democracy has become the gold standard of government in the 21st Century. In 1950, 15% of countries were classified as democracies; now more than half of UN countries are governed by a form of democracy [17]. Within these countries, the most common form of democracy is representative democracy, which Raikar [18] describes as “a system where citizens of a country or political entity may vote for representatives [to rule on their behalf]”.

Opinion polling is the direct surveying of the public via random sampling. Pollsters, the institutions that conduct polls, calculate public opinion on divisive matters and inform interested parties of the results. Kuha [19] notes that polling acts “as a feedback mechanism”, allowing political parties to modify policy choices and campaign approaches to best represent their electorate. Polling has been a significant factor in election campaigns over the last 100 years; the American Institute of Public Opinion brought polling into the mainstream in 1936 by correctly predicting that Franklin Roosevelt would defeat Alf Landon in the U.S. Presidential election [4]. Since then, polling has become integral to driving political campaigns. However, opinion polling is flawed – Brown [20] points out that in 2016 polls predicted Hillary Clinton to prevail over Donald Trump, while Brexit polls predicted a ‘Remain’ result. Both predictions were wrong.

An alternative approach to opinion polling is using a country’s summary statistics to predict voting habits. Chernov et al. [21] propose using financial data to bet on election outcomes. This was previously impossible, but due to accessible index data in the digital age, learning from a country’s social, economic, and political indicators is now a feasible methodology. This approach is based on rational choice theory, the assumption that individuals vote in line with their needs, rather than remaining loyal to one political party [22]. If a country is in recession, people are likely to desire a change in government to better their current circumstances.

This project aims to create a machine learning (ML) model that predicts a change in government in a general election. Predictions are made based on social, economic and political index scores, rather than polling. A key feature of the final model is its simplicity; rather than predicting who will win an election (which would require advanced knowledge of a country’s political landscape), the model predicts whether the incumbent government will remain in power or not. By simplifying the problem to a single-class classification task, model complexity is minimised. This simplicity enables the model to be used by a wide range of stakeholders, such as election observers, policy analysts, and civil society groups, particularly in contexts where polling data is limited or unavailable.

To collect training data, data on election results and indexes for each year were collated. Most election results are available via public databases, and where an existing database was unavailable, data was scraped using HTML parsing and regex. Index values over time were publicly available via online data banks including the OECD, The United Nations and The World Bank. Index values scored countries based on a range of issues including standards of healthcare, education, conflict and freedom. After exploratory data analysis (EDA), principal component analysis (PCA) helped reduce the dimensionality of the index data, resulting in a less complex feature set. Hierarchical Clustering was then performed to form four distinct clusters of countries, which were used to create the final four sub-models. Once the data was prepared, Logistic Regression, Random Forest (RF) and XGBoost models were compared. A final model was made that successfully predicted the results of elections using unseen data.

Figure 1.1 illustrates the overall workflow of the project, outlining the key stages from data collection and preprocessing to feature engineering, dimensionality reduction, modelling and final analysis.

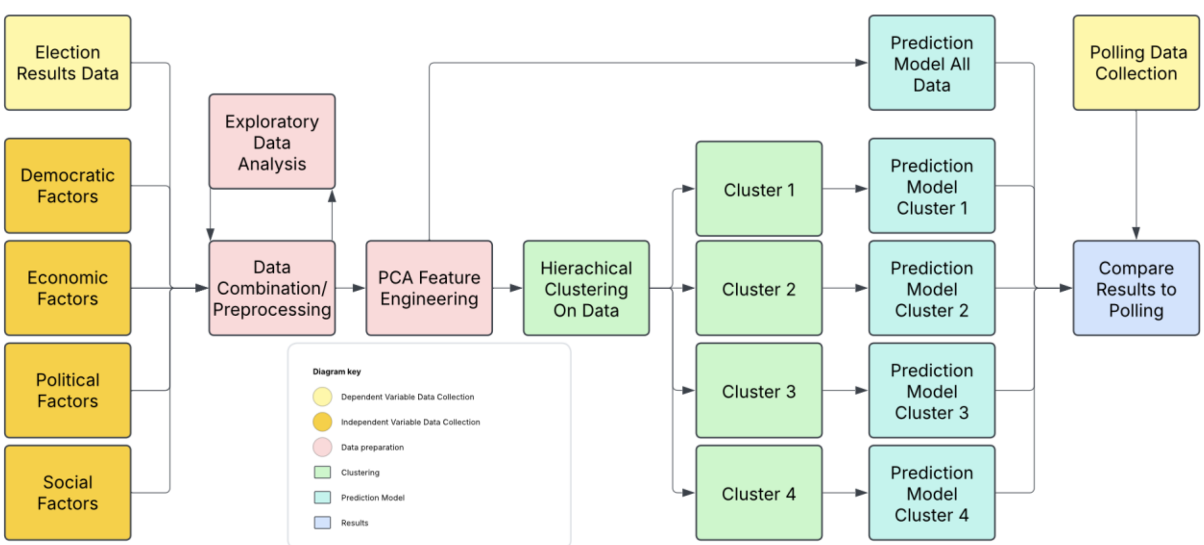

Figure 1.1: Project Overview

2. Background Research

Election forecasting traditionally relies on opinion polling, but consistent failures raise concerns about its accuracy [1]. Alternative methods are being explored; Econometric models and prediction markets have been compared to polling-based methods [2], while the predictive potential of social media sentiment has been reviewed [3]. This section identifies traditional and novel election forecasting techniques and then establishes the context for applying Machine Learning (ML) to predict election outcomes.

The adoption of polling into the mainstream is traced to the American Institute of Public Opinion [4], the first organisation to conduct opinion polling on a national scale. By correctly predicting the outcome of the 1936 U.S. presidential election, polling became the status quo of statistical methods for gathering information. Gallup's method incorporated random sampling, which only requires a small representative sample to find the views of the entire population.

We argue that in the 20th Century Gallup's method [4] was feasible because there was no better method. Advancements in technology in the 21st Century, alongside a rising population and more complex voting habits [6] have all impacted the success of opinion polling. Opinion polling has fared poorly in recent years, including failures in predicting the outcomes of the Brexit referendum poll and the 2016, 2020 and 2024 U.S. presidential elections [5].

John Barnhart [6] presents an early hypothesis that the economy is a significant factor in voting, stating that a bad economy may leave voters in a "receptive frame of mind". Morris Fiorina [7] describes this phenomenon as "retrospective voting", where incumbents vote based on how the current government has performed. Econometric models formalise this theory with some success [2]. There are few attempts to predict election outcomes using social factors. Yero [8] successfully used the Human Development Index to predict how various municipalities would vote in the 2018 Brazilian Presidential Election. Importantly, this is one of the only successful attempts to use ML to predict election outcomes, rather than traditional methods. These models each rely on Rational Choice Theory. Riker [9] describes Rational Choice Theory as "the assumption that [individuals] know what they want and can order their wants transitively". Josep Colomer [10] follows this by confirming that voters tend to maximise general satisfaction and minimise polarity and loyalty.

ML techniques are growing in popularity for forecasting tasks. An early application of neural networks to predict the outcome of the Iranian parliamentary elections was attempted with some success [11]. Sentiment analysis is another very popular method; a seminal 2002 paper [12] introduces the natural language processing approach to sentiment analysis; sentiment analysis on Twitter messages was explored in [13], where each message was classified as either 'positive' or 'negative' based on an input query. This method was used in 2011 in an attempt to predict elections [14], but a major limitation of these methods is the presence of a silent majority [15].

A recent review of machine learning techniques for election prediction found that Random Forest (RF) achieved the best balance between accuracy and model simplicity compared to neural networks [16]. The study highlighted model complexity as a key challenge in predicting election outcomes.

3. Question Development

The authors have experience in a range of disciplines, including computer science, engineering, economics, politics and mathematics. We sought to take advantage of this by using one or more of these fields as a foundation for this study. The authors studied literature from within these disciplines, with a focus on finding an unoptimized process that a data science-oriented approach could improve.

A discussion was held and the ideas of each author were scrutinised. We were particularly interested in a review of econometric models to predict elections [2]. These models benchmarked somewhat successfully but we criticised that these models assume that the quality of a country and its government can be inferred from its economy alone. Additionally, the models presented were designed to make predictions for only one country.

With a more complex model, the authors hypothesised that a fairer and more accurate prediction could be made. Social and political factors were considered as well as econometrics. Additionally, by transforming the model features to a normal scale, we suggested a 'modular' model which could make predictions for any country, given sufficient training data for that country or countries similar to it.

A major issue with the current research area was its high complexity; it was argued that "*each country has many political parties with different agendas, and political systems are not the same around the world"*. To simplify the research question two key changes were made. Firstly, countries with alternative or complex political systems weren't considered for analysis. Secondly, instead of attempting to predict who would win an election, our model would focus on whether or not the incumbent government would stay in power. This means that the model wouldn't require context on the political parties and system for the country in question.

This led to the development of our question, "Can a change in government be predicted using the democratic, economic, political and social factors of that country?"

4. Methodology

This section describes the methods employed to obtain the final model. Firstly the various data sources are given. The cleaning process is then described, followed by exploratory data analysis. Then, a correlation analysis led to the engineering of composite features. Clustering was performed to separate countries into groups and an initial model was made. By experimenting with different architectures, removing insignificant features and tuning hyperparameters, the final model was produced.

4.1 Data Collection

The dependent variable of the model was the outcomes of elections in countries of interest. Election results were extracted from the Parlgov Database. Parlgov is a political science infrastructure containing data on “all EU and most OECD countries, [containing] around 1000 elections between 1900 and 2023”. The database is stored as a repository on the Harvard Dataverse. Other election results including the U.S. presidential elections were extracted from online databases using web scraping and regex techniques. Data was manually imputed where it didn’t exist in an accessible format. By merging these data frames by year, the final data frame was made (*Table 4.1*).

The explanatory variables were found using various data banks, including The World Bank, OECD and UNESCO. The full list of data sources can be found in Appendix 1. We only used reputable data sources that are often used by other authors in the literature.

4.2 Data Cleaning

To simplify the election results data frame, each valid entry in a column was compared to the previous, returning a boolean. This was interpreted as *True* for a change in government or *False* otherwise (*Table 4.2*).

For the explanatory variables, various methods were explored to format the data most efficiently. The melt function formatted each dataset into country, year, and variable columns. Each database had unique naming rules and schemas, leading to compatibility issues when merging. For example, “Costa Rica” was formatted as “*costa_rica*”, “*costarica*” and “*Costa Rica*” across the different data sets. A function was created that employed regex techniques to unify the formatting of each data set. Concatenation was then possible, using country and year as the keys. The final dataset contained each explanatory variable as a column, with a multi-index for observations of country and year.

Next, null values were imputed. The decision to impute null values was made as there were concerns about model underfitting due to a small training set. Imputation was done using the best of three regression models. A linear regression model, a polynomial 2 and a polynomial 3 model were compared. Using Mean Absolute Percentage Error (MAPE) as the accuracy score, the best-fitting model was chosen on a case-by-case basis for all features, opting for the simplest viable model when possible. A MAPE threshold of 90% was set, and where models failed to pass the MAPE threshold nearest neighbour imputation was employed instead.

4.3 Exploratory Data Analysis

Exploratory data analysis (EDA) was conducted to better understand each feature and find relationships between them. For the dependent variable, the binary indicator was approximately uniform (*Figure 4.3*); 51.4% of elections resulted in a new government, while 48.6% kept the existing one. This ensured fairness for the final model.

The distributions and moments of each explanatory feature were scrutinised. The combined dataset included a binary indicator, "*Conflict*", representing whether or not the country was at war. This indicator had a significant skew as most observations were during periods of no conflict (*Figure 4.4*). Features with low variance or high skew were removed as features with non-ideal moments tend to be insignificant to the model output. *Figure 4.5* shows the distribution of features that were heavily skewed toward zero. These features were removed at this stage. The distributions of the remaining features suggested that they were informative (*Figure 4.6*). Many democratic indicators tended to be right-skewed; democratic countries should have high-scoring democratic indexes. Other indicators were more evenly distributed, which is likely where significant learning took place.

Temporal patterns of all indicators were examined (*Figure 4.7*). Democratic indicators appeared to have decreased after 2010, revealing a global decline in democratic standards over time. Contrastingly, the Human Development Index and urbanisation rates demonstrate continuous growth. It was observed that economic and democratic growth are not directly correlated. Gross Domestic Product (GDP) is the total economic output of an economy annually. [23]. The economic performance of a country can often be summarised via its GDP. Within the data, GDP fluctuated significantly over time, particularly in 2008 and 2020. These time-frames can be attributed to the 2008 financial crisis [24] and the Covid-19 pandemic [25]. *Media bias* and *journalist harassment* consistently decreased over time. This suggested an improvement in the freedom of the press. This was important for the final model as the media likely played a

key role in public opinion (and consequently voting habits) during the elections that the model was trained on.
Feature variance between countries was analysed (*Figure 4.8*). Some features had large variances, such as *parliamentary effectiveness* and features from the Human Development Index. Other features were more tightly clustered. *Figure 4.8* suggests that countries are alike in some ways and different in others. Features such as *infant mortality* and *life expectancy* have a fair spread which is ideal for the final model. The relationships between GDP, the Gini coefficient, and election outcomes were explored, revealing that higher GDP correlates with political stability (*Figure 4.9*).

Finally, a waterfall plot was made to demonstrate the effect each variable has on a change in government (*Figure 4.10*). This visualisation was chosen as its interpretation is clear and provides insights into how each variable may affect the model (linearly). Notably, economic factors had insignificant effects on government change, indicating that a non-linear approach could be more suitable for these variables. This was an early indicator that a machine-learning approach was viable with these factors.

4.4 Correlation Analysis & PCA

To further identify relationships between features, the correlation matrix was computed and highly correlated features were highlighted (*Figure 4.11, Figure 4.12*). High correlation indicated redundancy (if two variables were highly correlated, only one was required to convey the information). There were strong linear dependencies between variables from the same discipline. At this point, feature engineering was suggested to combine correlated variables into new ones. Principal Component Analysis (PCA) was employed to transform correlated features into independent principal components.

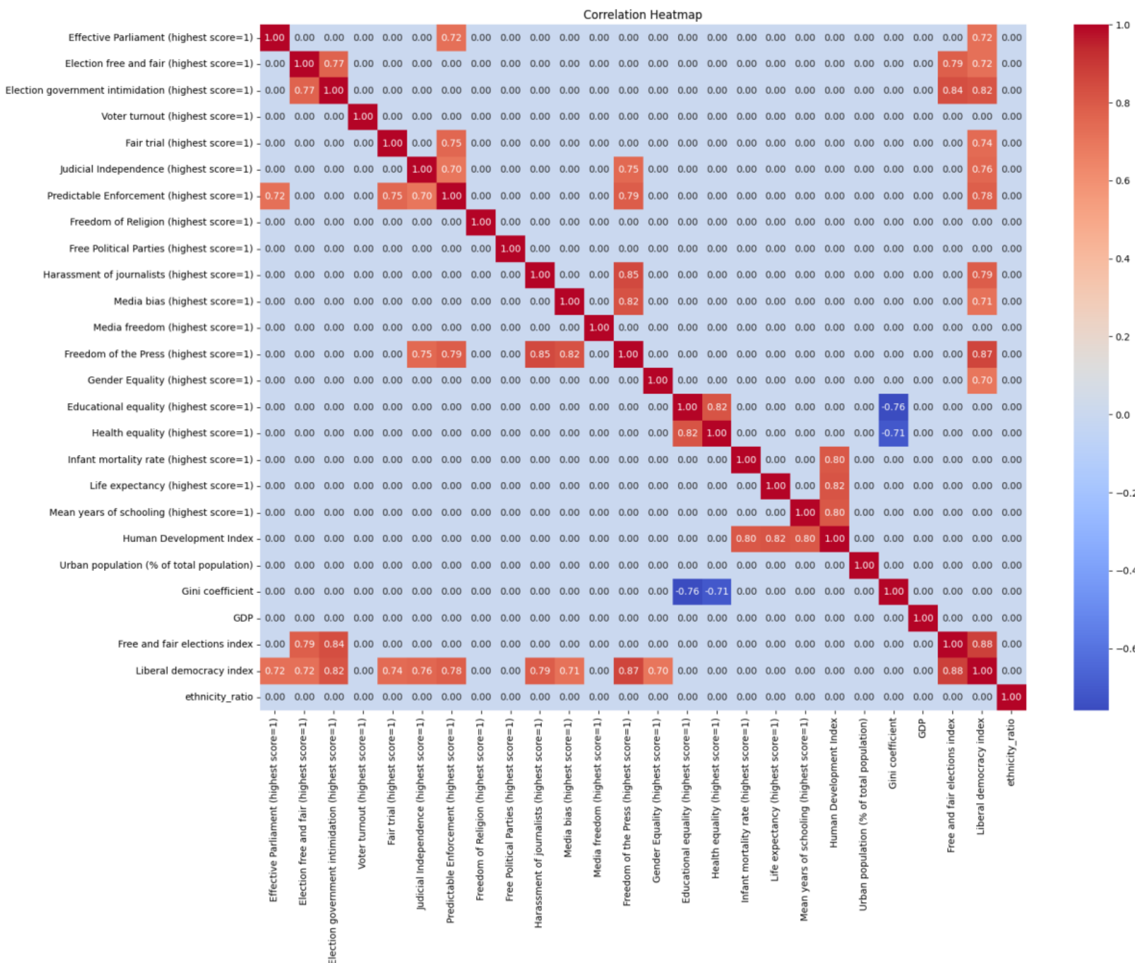

Figure 4.12: *Correlation Heatmap emphasising highly correlated features above 0.7*

After performing PCA on the combined feature groups, the correlation heatmap was recomputed (*Figure 4.13*). The new heatmap contained no `highly correlated features (≥ 0.7), indicating that` PCA was effective.

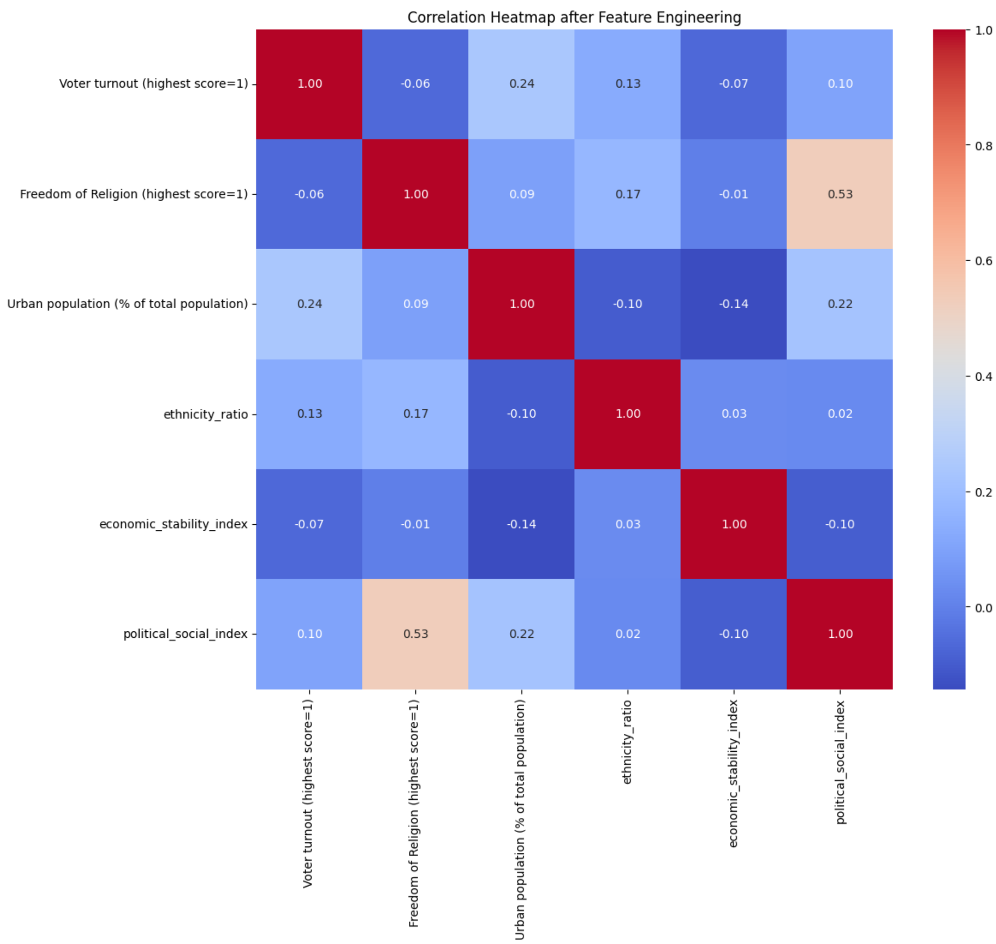

Figure 4.13: *Correlation Heatmap after Feature Engineering*

4.5 Feature Engineering

PCA was performed to extract the principal components that capture the most variance in each group (*Table 4.14*). Building on this, the feature engineering process focused on generating compound features that appropriately represent the highly correlated features found in *4.4*.

The standard approach is to retain multiple components of a whole dataset, based on explained variance. An alternative approach was employed; first, subsets were made of the highly correlated features; second, PCA was performed on each subset; finally, only the first principle component was retained, as it captured the most variance. The single remaining component for each subset represented a linear combination of the subset features.

Two composite indexes were created; the *'Political-Social Index'* and the *'Economic Stability Index'*, each representing groups of highly correlated features identified in *4.4*. The *'Political-Social Index'* explained 99.9% of the variance within its feature group, indicating an exceptionally strong linear relationship among the original variables. The *'Economic Stability Index'* explained 62.5% of the variance, which was considered sufficient to capture the dominant trend while reducing dimensionality. These indexes were subsequently used in place of the original correlated variables for modelling and further analysis.

To validate the composite features, comparative analyses were made between their yearly mean values (averaged across all countries) and those of the original features before the PCA transformation (*Figure 4.15*). Each index reliably captures the overall temporal pattern of its feature group.

4.6 Clustering Analysis

Hierarchical clustering was performed to explore underlying patterns among the 37 countries in the dataset. This method groups countries based on their similarity across all input features. Hierarchical clustering was chosen due to its ability to visualise cluster formation explicitly, allowing for intuitive interpretation of cluster relationships and flexible determination of cluster numbers. Unlike methods such as K-means, hierarchical clustering does not require specifying the number of clusters upfront and can reveal nested cluster structures, which is valuable given the complexity of country-level economic and social features. Additionally, Principal Component Analysis (PCA) was applied to all features to visualise the data in the $\mathbb{R}^2$ space.

First, the data set was standardised using StandardScaler to ensure equal weighting of all features. Since the data set spans a range of years, values were aggregated over time for each country. Each country was then represented by a single value, allowing for clustering; this approach was necessary because clustering algorithms require a fixed feature space for all entities. By focusing on long-term country-level characteristics, we can identify broader structural similarities among countries rather than short-term fluctuations. One disadvantage of this method was that averaging across years removed seasonality.

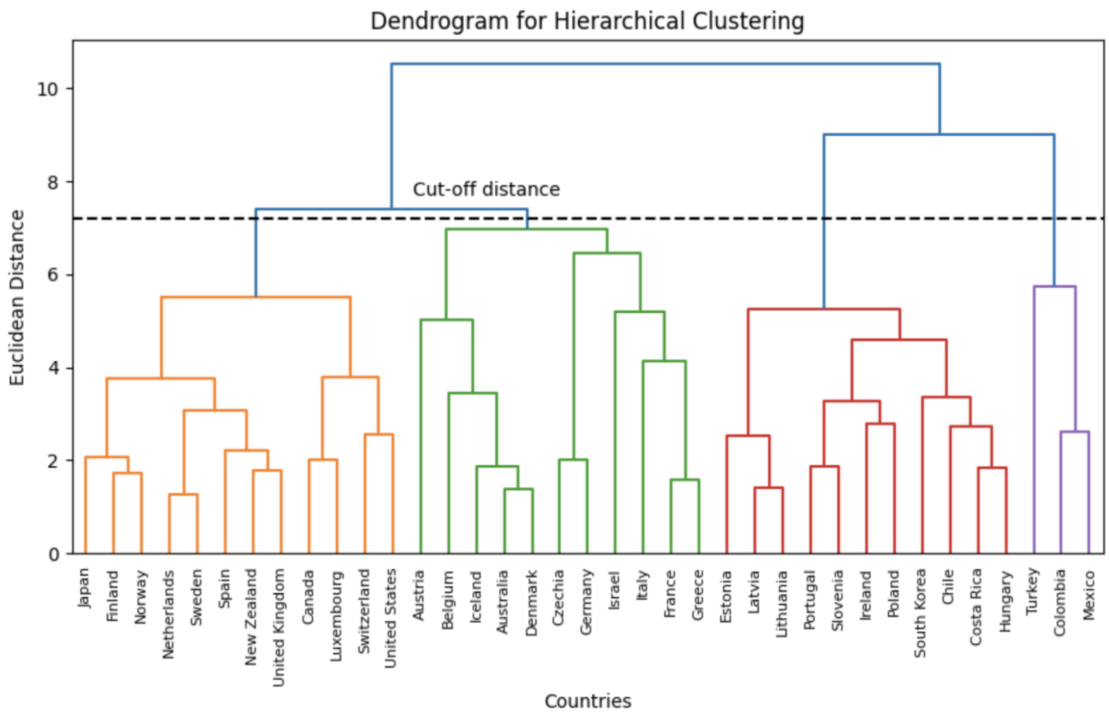

Figure 4.16: *Hierarchical Clustering Result*

Hierarchical clustering was applied using Ward's method (*Figure 4.16*). Ward's method was chosen to minimise intra-cluster variance. Euclidean distance was used to measure the similarity between countries. Four distinct clusters were found:

- Cluster 1: Japan, Finland, Norway, Netherlands, Sweden, Spain, New Zealand, United Kingdom, Canada, Luxembourg, Switzerland, United States
- Cluster 2: Austria, Belgium, Iceland, Australia, Denmark, Czechia, Germany, Israel, Italy, France, Greece
- Cluster 3: Estonia, Latvia, Lithuania, Portugal, Slovenia, Ireland, Poland, South Korea, Chile, Costa Rica, Hungary
- Cluster 4: Turkey, Colombia, Mexico

4.7 Further Cleaning

Modelling was challenging with input data containing ~75% NaN values. To resolve this, null values were imputed with the previous non-null value. To further simplify the data set, years were transformed into year 'ranges' (*Table 4.17*).

Table 4.17: An example row demonstrating the format of the election results data, after transforming the years into range.

country	range	features	change	year
Australia	2004-2007	...	True	2007

- **country**: The name of the country.
- **range**: The duration, in years, that the government remained in power.
- **features**: All the collected variables relevant to the analysis.
- **change**: Indicates whether governmental power changed during the election.
- **year**: The year in which the election was held.

4.8 Modelling

The following classification algorithms were selected for implementation:

- **Logistic Regression**: Chosen for its simplicity, interpretability, and strong baseline performance in binary classification problems. It provides probabilistic outputs and clear insights into feature importance through coefficients.

- **Random Forest**: Selected for its ability to handle high-dimensional data and capture complex, non-linear relationships. It is also robust to overfitting due to its ensemble nature and provides useful feature importance metrics.

- **XGBoost**: Utilised for its high predictive accuracy, scalability, and efficiency in handling imbalanced datasets. It combines gradient boosting with regularisation, making it suitable for structured tabular data and capable of managing feature interactions effectively.

Each of the model types was implemented and compared. Accuracy, loss and complexity were considered when comparing models.

4.9 Model Tuning: Feature Importance

After PCA (*4.4*), the dataset was reduced to six composite features. First, a model was made using the full dataset, to explore how these features influence election outcomes. The goal was to identify which features have the most impact, how well they predict election results, and how selecting the right features can improve model accuracy.

Second, four sub-models were made based on the clusters defined in *4.6*. By analysing feature importance within each group, the differences between clusters were better understood based on which features matter most in each. The best-performing combination of features was selected for the final models. This approach allowed for both an overall view and deeper insight into what affects voting habits across different types of countries.

Feature selection was conducted exclusively on the Logistic Regression and Random Forest Classifier models, as these algorithms support interpretable methods for evaluating feature importance. The architecture of XGBoost doesn’t support feature selection.

- **Logistic Regression**: Feature selection was performed using **permutation importance**.

This method was chosen because Logistic Regression does not inherently provide reliable feature importance scores when PCA-transformed features are used, as the model's coefficients lose their original interpretability. Permutation importance addresses this by measuring the decrease in model performance when each feature is randomly shuffled. This approach offers a model-agnostic, performance-based evaluation, making it well-suited for understanding the contribution of each principal component.

- **Random Forest Classifier**: Feature importance was extracted using the model’s **built-in feature importance function**, which calculates the mean decrease in impurity (Gini importance) across all trees in the ensemble.

This method is efficient, directly supported by the algorithm, and provides a straightforward way to assess the relative importance of each feature.

These two methods ensure that feature selection is both technically appropriate and aligned with the characteristics of each model.

5. Results

5.1. Predictive model results

The accuracy of each model was compared; cluster three's model achieved the highest accuracy (Table 5.1). Model accuracy is also compared before and after feature selection (Appendix Table *5.2)*.

The models demonstrated that machine learning (ML) can be used in predicting election outcomes. The Random Forest (RF) classifier emerged as the best-performing model overall, achieving a maximum accuracy of 66% in the full model using six key features. These include: *political_social_index*, *economic_stability_index*, *urban population (% of total population)*, *voter turnout*, *freedom of religion*, and *ethnicity ratio* — ordered by feature importance. Modelling clustered countries was more successful on every occasion than modelling all countries together.

Table 5.1*: Summary statistics for the full model and the four sub-models.*

Cluster	Countries Included	Model Used	Accuracy	Feature Importance (Sorted from most important to least)	Excluded Features
Full Model	All 38 countries	Random Forest Classifier	66%	1. political_social_index, 2. economic_stability_index, 3. urban population (%), 4. voter turnout, 5. freedom of religion, 6. ethnicity_ratio	None
First Cluster	Australia, Austria, Belgium, Czechia, Denmark, France, Germany, Greece, Iceland, Israel, Italy Random Forest Classifier	Random Forest Classifier	73%	1. economic_stability_index, 2. urban population (%), 3. political_social_index, 4. voter turnout	• freedom of religion, • ethnicity_ratio
Second Cluster	Colombia, Mexico, Turkey	Random Forest Classifier	75%	1. economic_stability_index, 2. urban population (%), 3. political_social_index, 4. voter turnout, 5. freedom of religion, 6. ethnicity_ratio	None
Third Cluster	Chile, Costa Rica, Estonia, Hungary, Ireland, Latvia, Lithuania, Poland, Portugal, Slovenia, South Korea	Logistic Regression	82%	1. voter turnout, 2. ethnicity_ratio, 3. economic_stability_index, 4. political_social_index, 5. urban population (%), 6. freedom of religion	None
Forth Cluster	Canada, Finland, Japan, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, United Kingdom, United States	Random Forest Classifier	73%	1. political_social_index, 2. economic_stability_index	• urban population (%), • voter turnout, • freedom of religion, • ethnicity_ratio

Table 5.2: *Model Accuracy Before and After Feature Importance (FI) by Cluster and Algorithm*

Logistic Regression									
Full Model		**1st Cluster**		**2nd Cluster**		**3rd Cluster**		**4th Cluster**	
Before F.I.	**After F.I.**	**Before F.I.**	**After F.I.**	**Before F.I.**	**After F.I.**	**Before F.I.**	**After F.I.**	**Before F.I.**	**After F.I.**
51%	*52%*	*32%*	*37%*	*33%*	*33%*	*82%*	*82%*	*50%*	*50%*
Random Forest Classifier									
64%	*66%*	*60%*	*73%*	*50%*	*75%*	*69%*	*69%*	*53%*	*73%*
XGBoost									
61%		*50%*		*66%*		*55%*		*40%*	

5.2. Case Study: Cluster 4

Justification for Feature Importance

The reliance on *political_social_index* and *economic_stability_index* in Cluster 4 underscores the institutional maturity and electoral stability of these countries. In such environments, voters are more responsive to nuanced shifts in political freedoms, democratic quality, and macroeconomic performance — as opposed to identity-based factors or access-based inequalities.

The *political_social_index*, derived via PCA from 20 socio-political indicators (e.g., media freedom, judicial independence, gender equality), captures the overall health of democratic institutions. In stable democracies, even marginal declines in these areas can trigger electoral dissatisfaction, making this index a strong predictor of change. The *economic_stability_index*, synthesised from *GDP* and *Gini coefficient*, reflects both prosperity and income equity. In these countries, economic stagnation or growing inequality, even at moderate levels, tends to influence public perception of incumbent effectiveness — hence its importance as a feature. Together, these two PCA-derived indices offer a comprehensive view of democratic performance and economic confidence, which are critical determinants of electoral outcomes in developed liberal democracies.

The modelling approach, enhanced through PCA and cluster-specific strategies, demonstrates that political and economic conditions are not uniformly influential across contexts. For high-income democracies (Cluster 4), the political-social and economic dimensions are paramount, reflecting the electorate's sensitivity to institutional performance and macroeconomic stability. This insight validates the relevance of engineered feature indices in capturing latent structures driving democratic accountability.

5.3. Comparison to Polling

We selected a subset of countries with publicly available polling data: Australia, Canada, Czechia, France, Ireland, Norway, Poland, Spain, Sweden, and the USA. These countries span across clusters 1, 2, 3, and 4, allowing us to compare polling predictions with the performance of our machine learning models both before and after applying feature importance (FI). On average, both polling and the original model achieved 64% accuracy across all countries. However, after applying FI, the model's overall performance increased to 66%, with substantial improvements observed across most clusters. When comparing results on a country-by-country basis, the original model outperformed polling in three of the ten countries. (*Figure 5.3, Figure 5.4*)

After applying feature importance, the model consistently reached **higher accuracy** across most countries, often matching or surpassing polling accuracy. Poland and Ireland demonstrated the most remarkable improvements in polling accuracy, increasing from 56% and 55% to 82%. Similarly, France and Sweden saw substantial gains, with their accuracy rising from 57% and 50% to 75% and 73%, respectively. **Canada** and **Czechia** also saw clear gains, outperforming their polling figures. The only countries where polling remained noticeably more accurate were **Australia**, **Spain**, and the **United States**. However, the differences were minimal — just **2%** in the U.S., and **7%** and **8%** in Spain and Australia, respectively.

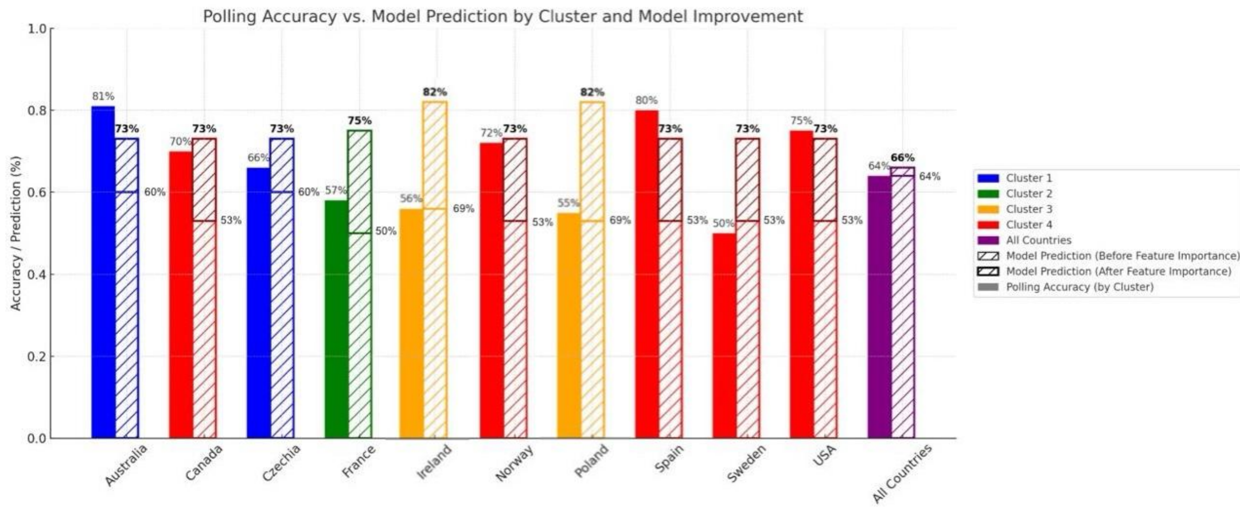

Figure 5.3: *Comparison of polling accuracy with model prediction by cluster and model improvement*

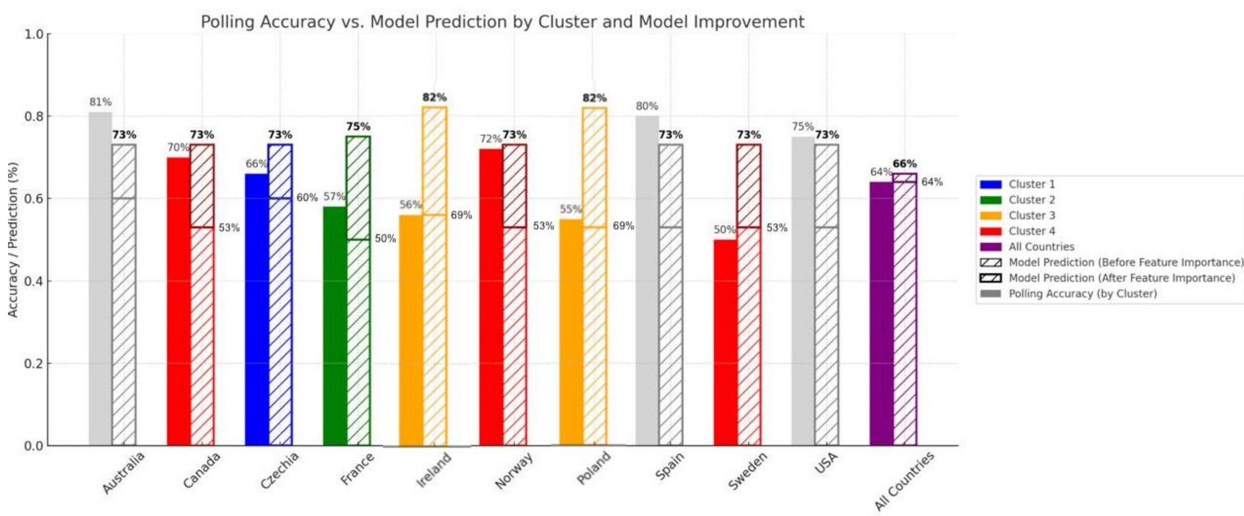

Figure 5.4: *Highlighted countries where model after feature importance outperformed polling*

6. Discussion & Limitations

Our findings demonstrate that traditional polling methods are surpassable by machine learning (ML) techniques in predicting election results. The ML techniques applied match or outperform polling accuracy in more than half of the countries in the dataset. The general model takes data from 38 countries and returns prediction rates between 66% and 82%. Using statistically similar countries to create a compound model is a novel approach; however, doubts remain due to previously poor results [26]. We hypothesise that different countries are similar enough to model collectively.

For all models, logistic regression (LR), random forest (RF), and XGBoost are tested. RF is the best model three times out of four, which aligns with [16]. RF is robust to outliers, especially in the full model. LR also shows some success, possibly due to the binary nature of the input. XGBoost is overly sensitive to outliers, resulting in poor performance. Additionally, the results of hierarchical clustering offer insight into the socio-economic structure of each country. Trends emerge beyond geographic or political boundaries. A geographic trend is visible, where neighbouring countries belong in the same cluster. For example, Estonia, Latvia, and Lithuania are highly statistically similar. Applying a geographical and historical context, these results are intuitive.

One limitation of the model was that the training data contained imputed data, which led to a lower accuracy on unseen data. Imputing data can lead to drawing invalid conclusions [27]. A major limitation of the methodology is the necessity of the boolean transformation of the dependent variable. Context on individual political parties becomes impossible to learn, making the final model blind to polarisation. Inspired by political literature [28], the model defines a political change as a change in the party with the most seats. This method ignores governmental structure, including the relationships between political parties. Some governments consist of coalitions of political parties; the model perceives coalitions as individual political parties, rather than a combination of the two. This method may draw criticism, however, the authors believe that it's a necessary methodology for the model to remain simple.

The countries in our dataset all contain either a purely parliamentary system or a hybrid form of government, with both parliamentary and presidential elections. Almost always, parliamentary elections are more relevant for governance [30]. Presidents generally deal with foreign affairs rather than domestic ones. An exemption to this is in the United States, hence presidential election results were used. While it isn’t desirable to not capture all aspects of democracy in our study, we believe parliamentary elections are the most feasible way to do so, except for the United States. Using different elections in the same dataset damages the replicability of our dataset, but is a worthwhile cost to include the United States, a globally important democracy.

7. Conclusion & Future Work

It's demonstrated that machine learning (ML) methods can match or outperform traditional polling techniques in predicting election outcomes. Predictions were generated across 38 countries with moderate or high accuracy.

In various electoral systems, political cultures, and socioeconomic contexts, our models scored consistently high in benchmarking, suggesting there are common factors that affect election outcomes globally. On the other hand, the variation in model performance across different clusters indicates there is value in tailoring machine learning approaches to localised contexts. This also broadens the model's potential application for stakeholders such as electoral observers, policymakers, and research institutions who require scalable tools to forecast outcomes in the absence of consistent polling data.

Further research may aim to improve the robustness and precision of our models by addressing limitations. Comprehensive election outcome data, particularly for coalition dynamics, could significantly enhance prediction accuracy. There is the opportunity to replicate our methodology with other election types, such as in local government. Improved input data would lead to even better predictions. Feature selection could also be optimised further using automated or unsupervised techniques which both reduce multicollinearity and increase model interpretability.

Alternative models, such as neural networks or ensemble meta-models, should be explored. Expanding the dataset to test the inclusion of psychological, technological and historical factors could potentially improve the model further.

Ultimately, this research has opened new pathways for computational approaches to political science. With further refinements, increased data quality, and interdisciplinary collaboration, machine learning approaches have the potential to become a mainstay in election forecasting and the broader analysis of democratic trends.

References

[1] C. Prosser and J. Mellon, “The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses,” Government and Opposition, vol. 53, no. 4, pp. 757–790, 2018, doi: 10.1017/gov.2018.7.

[2] A. Leigh and J. Wolfers, “Competing Approaches to Forecasting Elections: Economic Models, Opinion Polling and Prediction Markets,” Economic Record, vol. 82, pp. 325–340, 2006, doi: 10.1111/j.1475-4932.2006.00343.x.

[3] K. D. S. Brito, R. L. C. S. Filho and P. J. L. Adeodato, “A Systematic Review of Predicting Elections Based on Social Media Data: Research Challenges and Future Directions,” IEEE Trans. Comput. Social Syst., vol. 8, no. 4, pp. 819–843, Aug. 2021, doi: 10.1109/TCSS.2021.3063660.

[4] G. Gallup and C. Robinson, “American Institute of Public Opinion—Surveys, 1935–38,” Public Opinion Quarterly, vol. 2, no. 3, pp. 373–398, Jul. 1938, doi: 10.1086/265204.

[5] A. Gelman, “Failure and Success in Political Polling and Election Forecasting,” Statistics and Public Policy, vol. 8, no. 1, pp. 67–72, 2021, doi: 10.1080/2330443X.2021.1971126.

[6] J. D. Barnhart, “Rainfall and the Populist Party in Nebraska,” American Political Science Review, vol. 19, no. 3, pp. 527–540, 1925. doi:10.2307/2939131

[7] M. P. Fiorina, “Economic retrospective voting in American elections: a microanalysis,” American Journal of Political Science, vol. 22, no. 2, pp. 426–443, 1978.

[8] E. J. Huerta Yero, N. C. Sacco and M. do C. Nicoletti, “Effect of the Municipal Human Development Index on the results of the 2018 Brazilian presidential elections,” Expert Systems with Applications, vol. 168, 2021, Art. no. 114305, doi: 10.1016/j.eswa.2020.114305.

[9] W. H. Riker, “The Political Psychology of Rational Choice Theory,” Political Psychology, vol. 16, no. 1, pp. 23–44, 1995, doi: 10.2307/3791448.

[10] J. M. Colomer, “The Median Voter Maximizes Satisfaction and Minimizes Polarization: Worldwide evidence since the 19th century,” Higher Council of Scientific Research (CSIC) and University Pompeu Fabra (UPF), Barcelona, 2000.

[11] A. Esfandiari, H. Khaloozadeh, M. Esfandiari and Z. J. Fard, “Machine Learning in Parliament Elections,” Research Journal of Applied Sciences, Engineering and Technology, vol. 4, no. 19, pp. 3732–3739, 2012. [Accessed: Mar. 27, 2025]. Available: Airiti Library.

[12] B. Pang, L. Lee and S. Vaithyanathan, “Thumbs up? Sentiment Classification using Machine Learning Techniques,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79–86.

[13] A. Go, R. Bhayani and L. Huang, “Twitter Sentiment Classification Using Distant Supervision,” CS224N Project Report, Stanford University, 2009, pp. 1–12.

[14] A. Bermingham and A. F. Smeaton, “On using Twitter to monitor political sentiment and predict election results,” in Proc. Workshop on Sentiment Analysis, 2011.

[15] G. Gao, B. N. Greenwood, R. Agarwal and J. S. McCullough, “Vocal minority and silent majority,” MIS Quarterly, vol. 39, no. 3, pp. 565–590, 2015.

[16] A. Argandoña-Mamani, T. Ormeño-Alarcón, O. Iparraguirre-Villanueva, C. Paulino-Moreno and M. Cabanillas-Carbonell, “Predicting Election Results with Machine Learning—A Review,” in Proc. 8th Int. Congr. Information and Communication Technology (ICICT), Singapore, 2023, vol. 695, pp. 1030–1043, doi: 10.1007/978-981-99-3043-2_82.

[17] V-Dem Institute, *Liberal Democracy Index*, 2024. [Online]. Available: https://v-dem.net/vdemds.html

[18] A. Raikar, Introduction to Political Representation and Electoral Systems, Oxford University Press, 2025.

[19] J. Kuha, “Polling as a Feedback Mechanism in Political Communication,” Political Science Review, vol. 45, no. 1, pp. 55–68, 2022.

[20] T. Brown, “Why Opinion Polls Keep Getting It Wrong,” The Guardian, Feb. 9, 2024. [Online]. Available: https://www.theguardian.com/politics/2024/feb/09/why-opinion-polls-keep-getting-it-wrong

[21] A. Chernov, B. Levitt, and S. Omar, “Using Financial Data to Predict Electoral Outcomes: A Rational Choice Perspective,” Journal of Political Economy, vol. 132, no. 3, pp. 741–765, 2024, doi: 10.1016/j.jpe.2024.01.009.

[22] M. Amadae, Rational Choice Theory and Democratic Decision-Making, Cambridge University Press, 2025.

[23] D. Slesnick, *Measuring Economic Well-Being: GDP and Beyond*, Brookings Institution Press, 2020.

[24] S. Chen, M. Mrkaic, and M. Nabar, "The Global Economic Recovery 10 Years After the 2008 Financial Crisis," *IMF Working Paper WP/18/197*, International Monetary Fund, 2018. [Online]. Available: https://www.imf.org/en/Publications/WP/Issues/2018/09/17/The-Global-Economic-Recovery-10-Years-After-the-2008-Financial-Crisis-46265

[25] E. Gagnon, S. B. Kamin, and J. Kearns, "The Impact of COVID-19 on the Global Economy: Insights from Central Banks," *Bank for International Settlements Bulletin No. 49*, 2023. [Online]. Available: https://www.bis.org/publ/bisbull49.htm

[26] J. Karp and J. Vowles, "Democracy and the Meaning of Elections: Comparing across Nations," *Electoral Studies*, vol. 44, pp. 241–251, 2016, doi: 10.1016/j.electstud.2016.07.005.

[27] H. Kang, "The prevention and handling of the missing data," Korean J Anesthesiol, vol. 64, no. 5, pp. 402–406, May 2013, doi: 10.4097/kjae.2013.64.5.402.

[28] A. Ake, "A Definition of Political Stability," Comparative Politics, vol. 7, no. 2, Jan. 1975. [Online]. Available: https://doi.org/10.2307/421552

[29] London School of Economics, "Balancing Act: How Coalition Governments Are Reshaping Democracy," LSE Blog, Jan. 30, 2023. [Online]. Available: https://blogs.lse.ac.uk/lseupr/2024/01/30/balancing-act-how-coalition-governments-are-reshaping-democracy/

[30] J. M. Carey, *Legislative Voting and Accountability*, Cambridge University Press, 2008.

[31] L. Vogt, A. Bormann, P. Rüegger, L. Cederman, K. Hunziker, and M. Girardin, "Integrating data on ethnicity, geography, and conflict: The Ethnic Power Relations dataset family," *Journal of Conflict Resolution*, vol. 59, no. 7, pp. 1327–1342, 2015. [Online]. Available: https://icr.ethz.ch/data/epr

[32] Uppsala Conflict Data Program (UCDP), *Armed Conflict Dataset*, 2024. [Online]. Available: https://ucdp.uu.se

[33] Natural Earth, *Public Domain Map Data*, 2023. [Online]. Available: https://www.naturalearthdata.com

[34] J. Klawans, “How accurate are political polls really?,” The Week, Nov. 2024. [Online]. Available: https://theweek.com/politics/2024-election-polls-accuracy

[35] W. Kulachai, U. Lerdtomornsakul, and P. Homyamyen, “Factors Influencing Voting Decision: A Comprehensive Literature Review,” Social Sciences, vol. 12, no. 9, Aug. 2023. [Online]. Available: https://doi.org/10.3390/socsci12090469

[36] A. Movsisyan et al., “Inflation and health: a global scoping review,” Science Direct, vol. 12, no. 6, pp. 1038–1048, Jun. 2024.

[37] T. Piketty, Capital in the 21st Century. Paris: Éditions du Seuil, 2013.

[38] Z. Quershi, “Rising inequality: A major issue of our time,” Brookings, May 16, 2023. [Online]. Available: https://www.brookings.edu/articles/rising-inequality-a-major-issue-of-our-time/

[39] S. Simon, “U.K.’s Conservative Party suffered its worst-ever defeat. What went wrong?,” NPR, Jul. 6, 2024. [Online]. Available: https://www.npr.org/2024/07/06/nx-s1-5026560/u-k-s-conservative-party-suffered-its-worst-ever-defeat-what-went-wrong

[40] R. Upward and P. Wright, “Income shocks, political support and voting behaviour,” Journal of Public Economics, vol. 239, Nov. 2024.

[41] R. Wike, M. Fagan, and L. Clancy, “Global Elections in 2024: What We Learned in a Year of Political Disruption,” Pew Research Center, Dec. 11, 2024. [Online]. Available: https://www.pewresearch.org/global/2024/12/11/global-elections-in-2024-what-we-learned-in-a-year-of-political-disruption/

[42] H. Döring and P. Manow, *Parliaments and Governments Database (ParlGov)*, University of Bremen, [Online]. Available: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2VZ5ZC

[43] International IDEA, *The Global State of Democracy Indices (GSoD), Version 7*, 2023. [Online]. Available: https://www.idea.int/gsod-indices

[44] United Nations Development Programme, *Human Development Reports*, 1990–2023. [Online]. Available: http://hdr.undp.org

[45] Transparency International, *Corruption Perceptions Index*, 1990–2023. [Online]. Available: https://www.transparency.org/en/cpi

[46] World Bank, *World Development Indicators*, [Online]. Available: https://data.worldbank.org

[47] OECD, *Global Revenue Statistics Database*, 2018. [Online]. Available: https://www.oecd.org/tax/tax-policy/tax-database.htm

[48] World Inequality Database, *Global Inequality Data*, [Online]. Available: https://wid.world

[49] M. Vogt, N.-C. Bormann, S. Rüegger, L.-E. Cederman, P. Hunziker, and L. Girardin, "Integrating data on ethnicity, geography, and conflict: The Ethnic Power Relations data set family," Journal of Conflict Resolution, vol. 59, no. 7, pp. 1327–1342, 2015.

[50] OECD, *OECD Health Statistics*, 2023. [Online]. Available: https://data-explorer.oecd.org

[51] UNESCO Institute for Statistics, *Education Statistics*, [Online]. Available: http://data.uis.unesco.org

[52] V-Dem Institute, *Liberal Democracy Index*, 2024. [Online]. Available: https://v-dem.net/vdemds.html

[53] Uppsala Conflict Data Program (UCDP), *Armed Conflict Dataset*, 2024. [Online]. Available: https://ucdp.uu.se

[54] Natural Earth, *Public Domain Map Data*, 2023. [Online]. Available: https://www.naturalearthdata.com

[55] A. Coppock and D. P. Green, "Is Voting Habit Forming? New Evidence from Experiments and Regression Discontinuities," American Journal of Political Science, vol. 60, no. 4, pp. 1044–1062, 2016.

Appendices

A.1 Data Sources

Sector	Variable	Sources
Election results	• Party/coalition with a parliamentary majority • Party winning a presidential election	• ParlGov • Various government websites
Democratic	• Liberal Democracy index • Democracy Index • Presence of armed conflicts • Voter turnout	• Our World in Data • International Institute for Democracy and Electoral Assistance
Economic	• GDP per capita • Economic Inequality	• The World Bank • OECD
Political	• The Global State of Democracy Indices • Human Development Index • Corruption Perceptions Index	• International IDEA • Resource Watch • Transparency International
Social	• Education • Population • Healthcare • Ethnic and religious diversity	• ETH Zurich • UNESCO • The World Bank • OECD

A.2 Figures and Tables

Table 4.1: An excerpt of the final data frame of election results. The table contains 124 rows (years) and 43 columns (countries). Much of the data was extracted from the Parlgov Database; the remaining data was scraped from reputable sources.

	Australia object Free Trade ... 10% 4 others 30% Missing 60%	Denmark object Liberal Party 30% Social Dem... 10% Missing 60%	Norway object Liberal Part... 20% 2 others 20% Missing 60%	Belgium object Catholic Pa... 45% Belgian Lab... 5% Missing 50%	France object Left Republ... 5% 4 others 20% Missing 75%	New Zealand obj... New Zealan... 20% Reform Party 10% Missing 70%
1900	nan	nan	Liberal Party of N...	Catholic Party	nan	nan
1901	Protectionist Party	Liberal Party	nan	nan	nan	nan
1902	nan	nan	nan	Catholic Party	Left Republican	New Zealand Lib...
1903	Free Trade Party	Liberal Party	Conservative Party	nan	nan	nan
1904	nan	nan	nan	Catholic Party	nan	nan
1905	nan	nan	nan	nan	nan	New Zealand Lib...
1906	Free Trade Party	Liberal Party	Liberal Party of N...	Catholic Party	Conservatives	nan
1908	nan	nan	nan	Catholic Party	nan	New Zealand Lib...
1909	nan	Social Democrats	Conservative Party	nan	nan	nan
1910	Australian Labor ...	Liberal Party	nan	Catholic Party	Republican Social...	nan
1911	nan	nan	nan	nan	nan	New Zealand Lib...
1912	nan	nan	Liberal Party of N...	Catholic Party	nan	nan
1913	Commonwealth L...	Social Democrats	nan	nan	nan	nan
1914	Australian Labor ...	nan	nan	Catholic Party	Republican Union	Reform Party

	Australia object True 25% False 15% Missing 60%	Denmark object False 20% True 20% Missing 60%	Norway object True 30% False 10% Missing 60%	Belgium object False 40% True 10% Missing 50%	France object True 20% False 5% Missing 75%	New Zealand obj... False 25% True 5% Missing 70%
1900	nan	nan	False	False	nan	nan
1901	False	False	nan	nan	nan	nan
1902	nan	nan	nan	False	False	False
1903	True	False	True	nan	nan	nan
1904	nan	nan	nan	False	nan	nan
1905	nan	nan	nan	nan	nan	False
1906	False	False	True	False	True	nan
1908	nan	nan	nan	False	nan	False
1909	nan	True	True	nan	nan	nan
1910	True	True	nan	False	True	nan
1911	nan	nan	nan	nan	nan	False
1912	nan	nan	True	False	nan	nan
1913	True	True	nan	nan	nan	nan
1914	True	nan	nan	False	True	True

Table 4.2: The cleaned data frame of election results. If an incumbent government is replaced in an election, "*True*" is returned. Otherwise, "*False*" is returned.

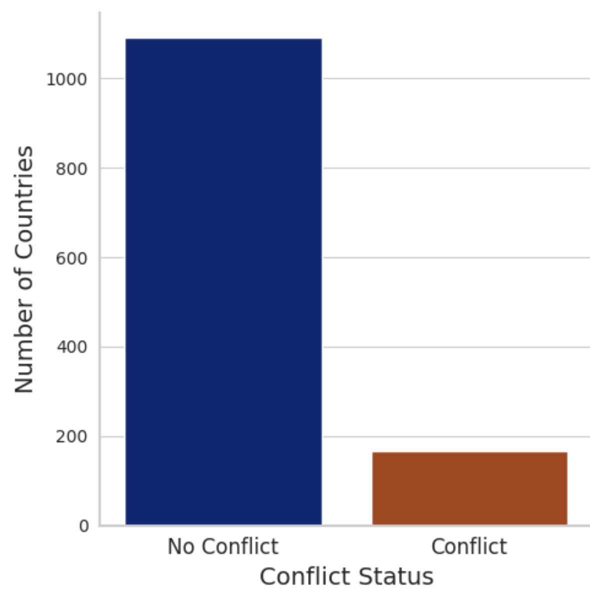

Figure 4.3: Distribution of Conflict Presence Across Countries
The bar chart illustrates the number of observations with and without conflict in the dataset. A substantial imbalance is evident, with most countries experiencing no conflict during the observation period.

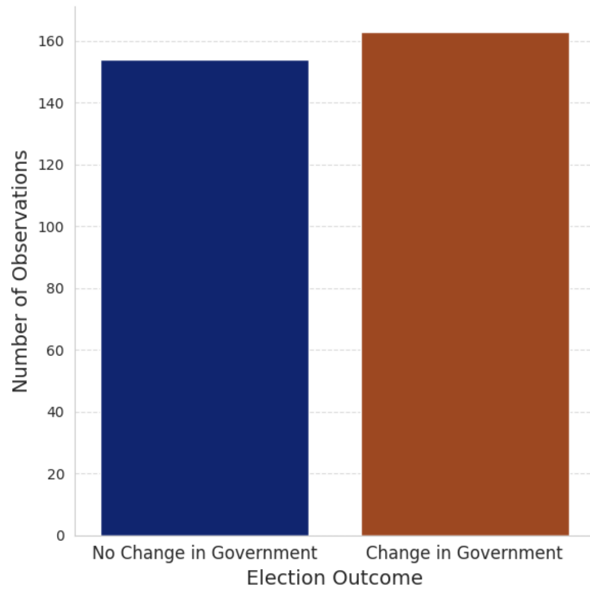

Figure 4.4: Distribution of Government Change Outcomes Post-Election
This bar chart shows the frequency of government change versus continuity following elections.

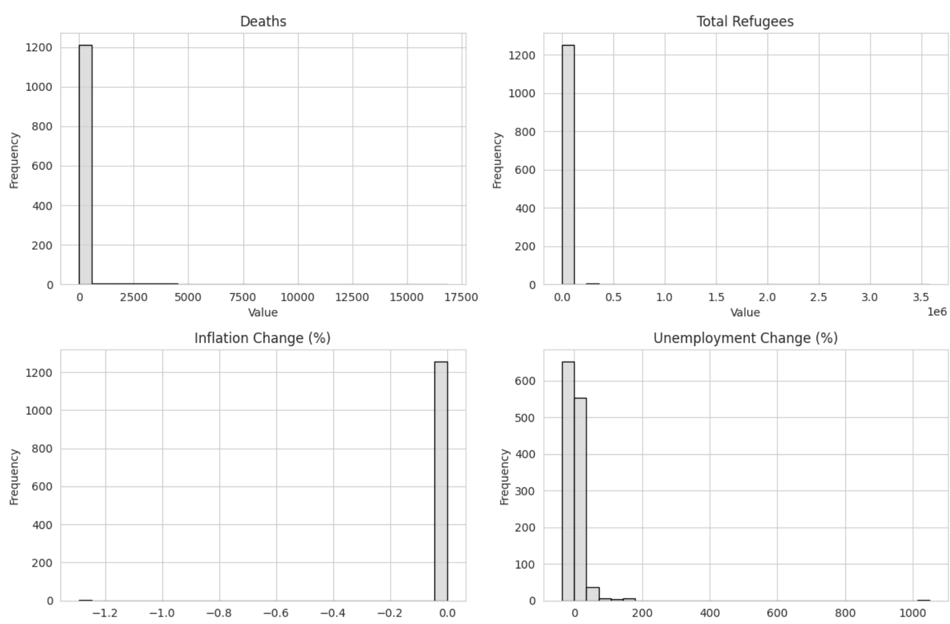

Figure 4.5: Distributions of Features Removed During Preprocessing
Histograms display the distribution of variables removed from the main analysis, including Deaths, Refugees, Inflation Change (%), and Unemployment Change (%).

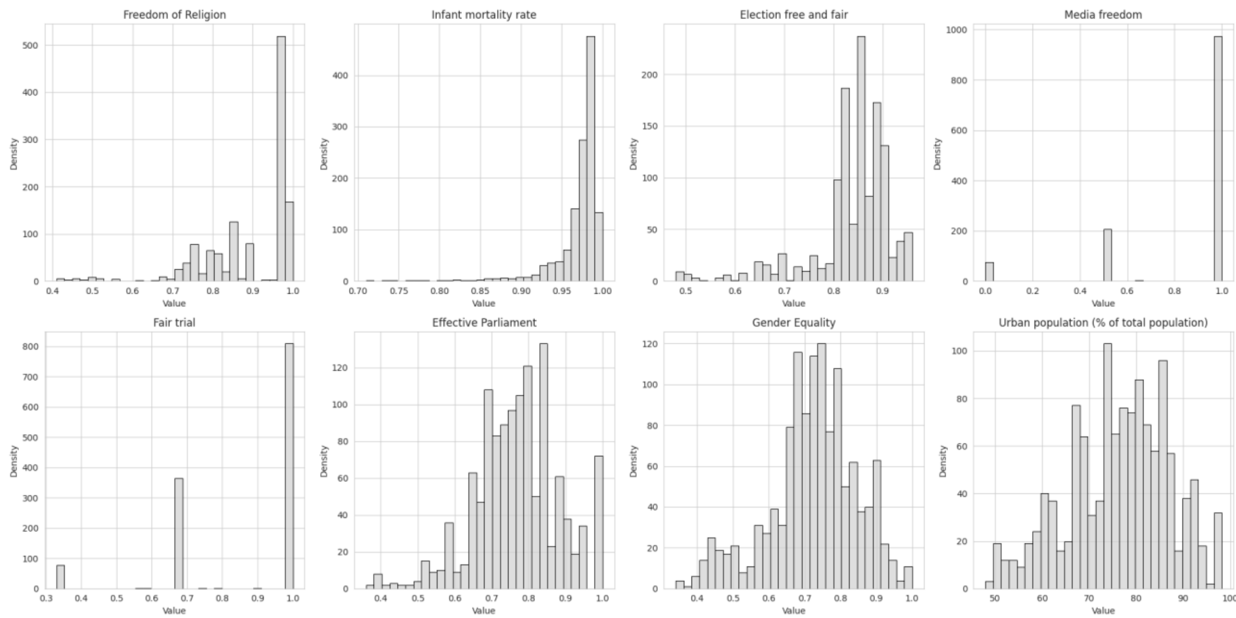

Figure 4.6: Distributions of Select Predictor Variables

The histograms show the distribution of key variables retained for modelling, including Freedom of Religion, Infant Mortality Rate, Election Free and Fair, Media Freedom, Fair Trial, Effective Parliament, Gender Equality, and Urban Population (%).

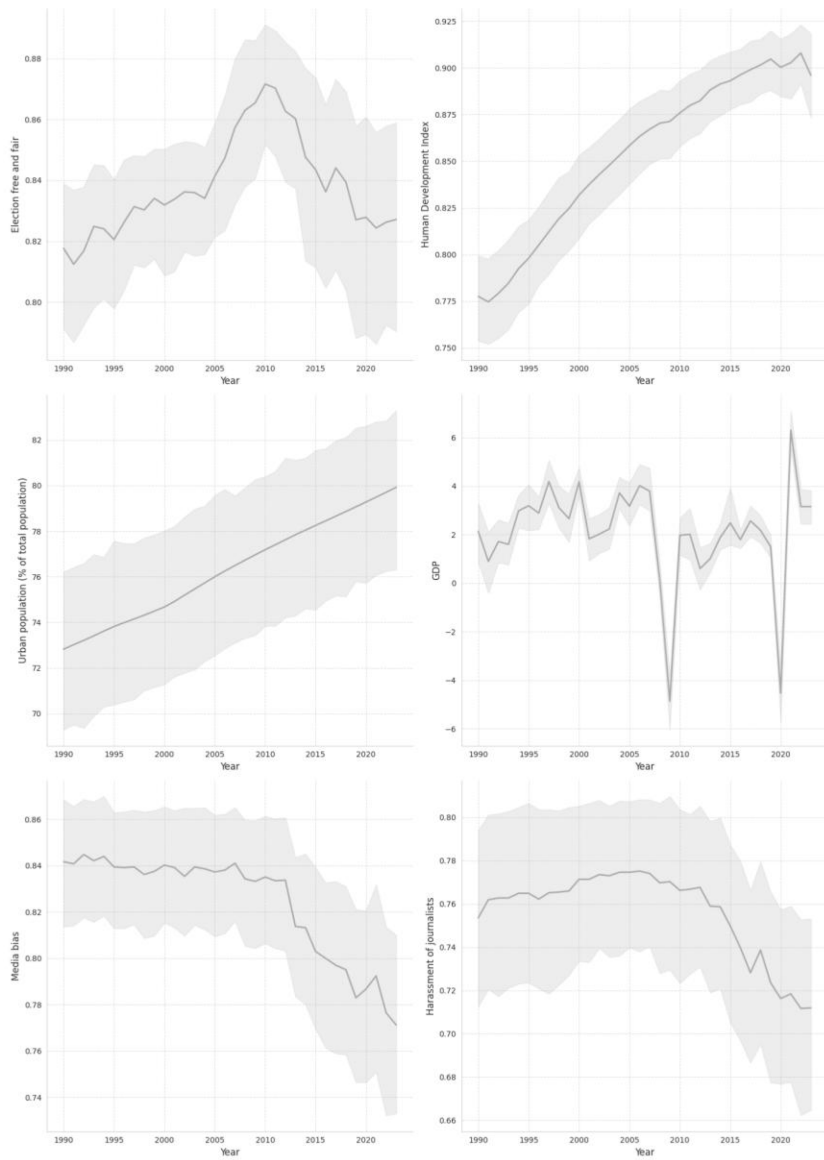

Figure 4.7: Temporal Trends in Political, Social, and Economic Indicators (1990–2023)

This time-series plot highlights trends for several features, including Election Free and Fair, Human Development Index, Urban Population, GDP, Media Bias, and Journalist Harassment, with 95% confidence intervals.

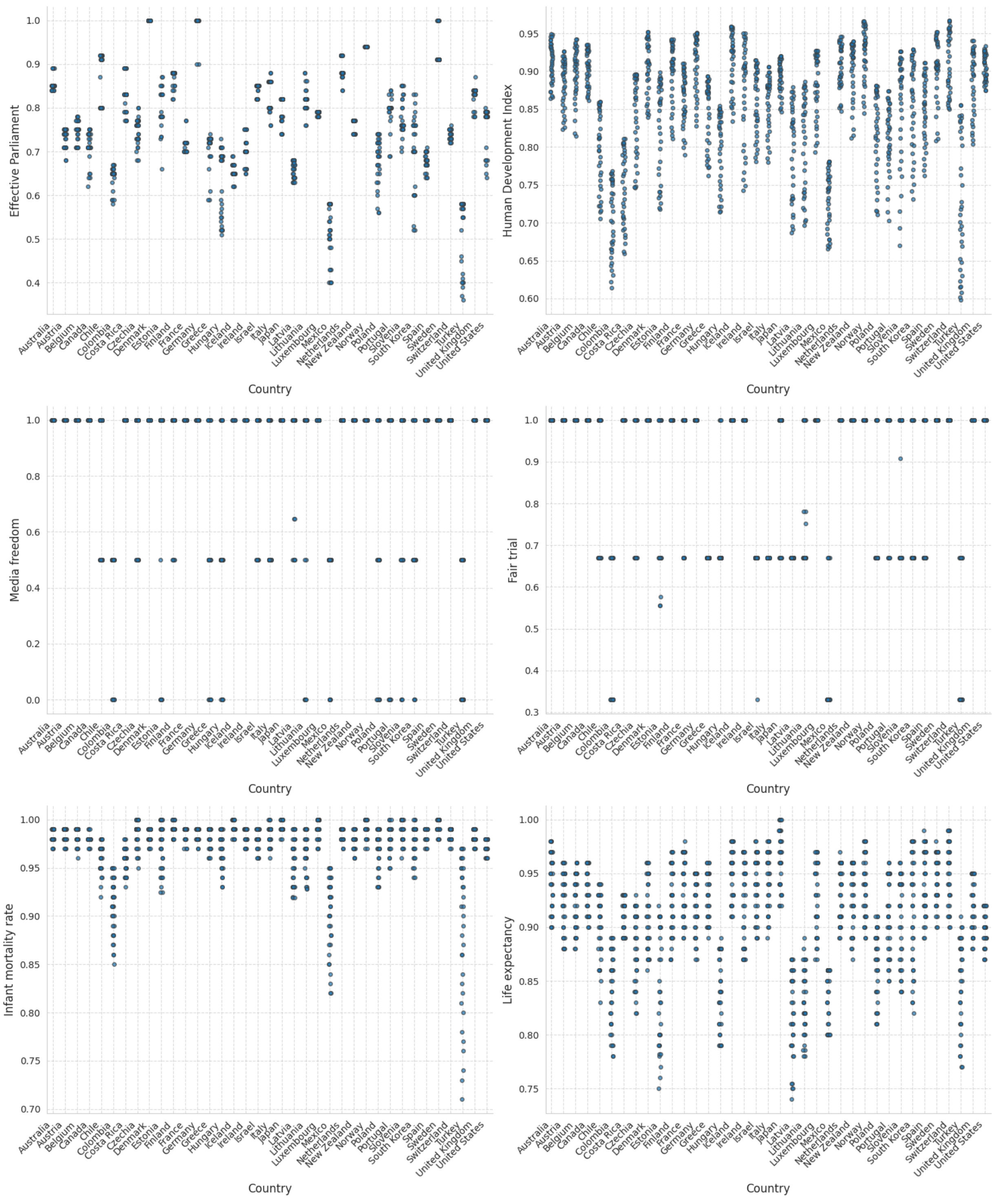

Figure 4.8: Country-Level Variation in Select Key Indicators

This strip plot visualises the distribution of various indicators (e.g., Effective Parliament, Human Development Index, Media Freedom, Infant Mortality Rate) across countries.

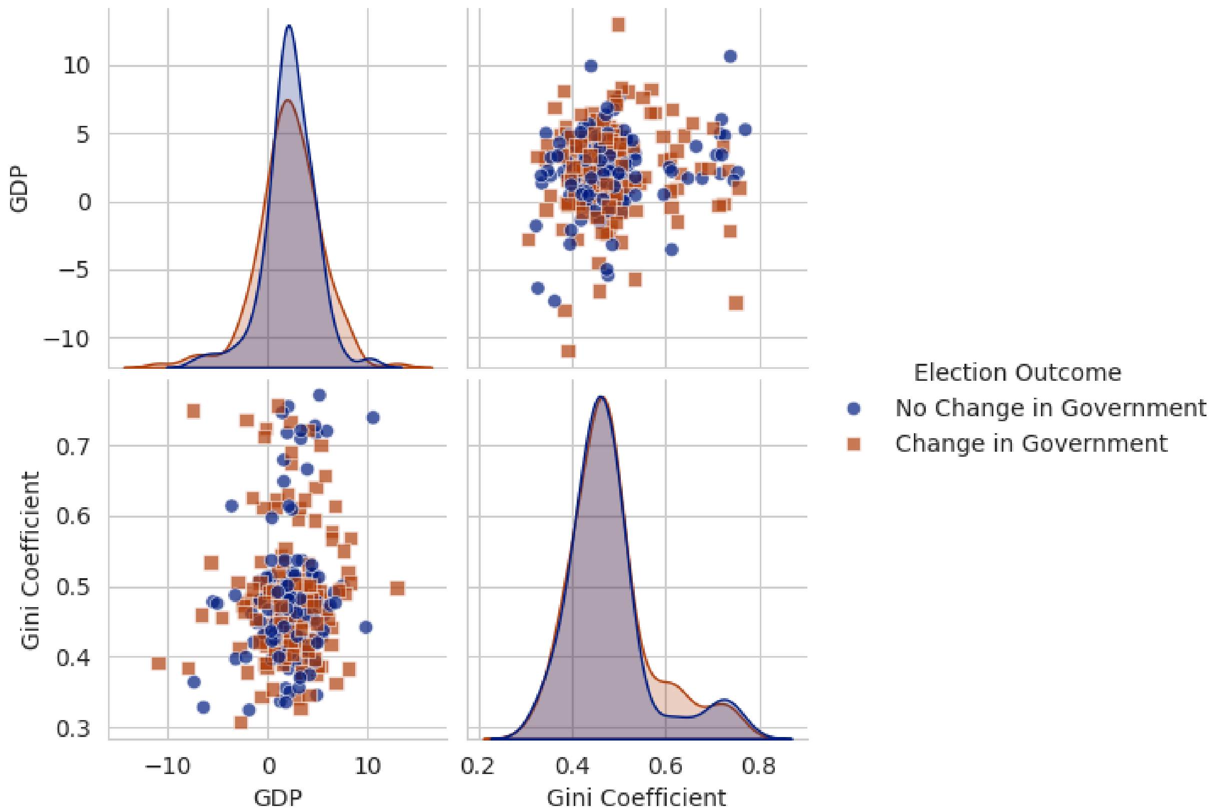

Figure 4.9: The joint distribution of GDP and the Gini coefficient, segmented by whether a government changes following an election. Countries with no government change have higher GDPs. Changes in government are more common in places with low or stagnant GDP growth. The authors hypothesise that economic uncertainty can cause a change in government.

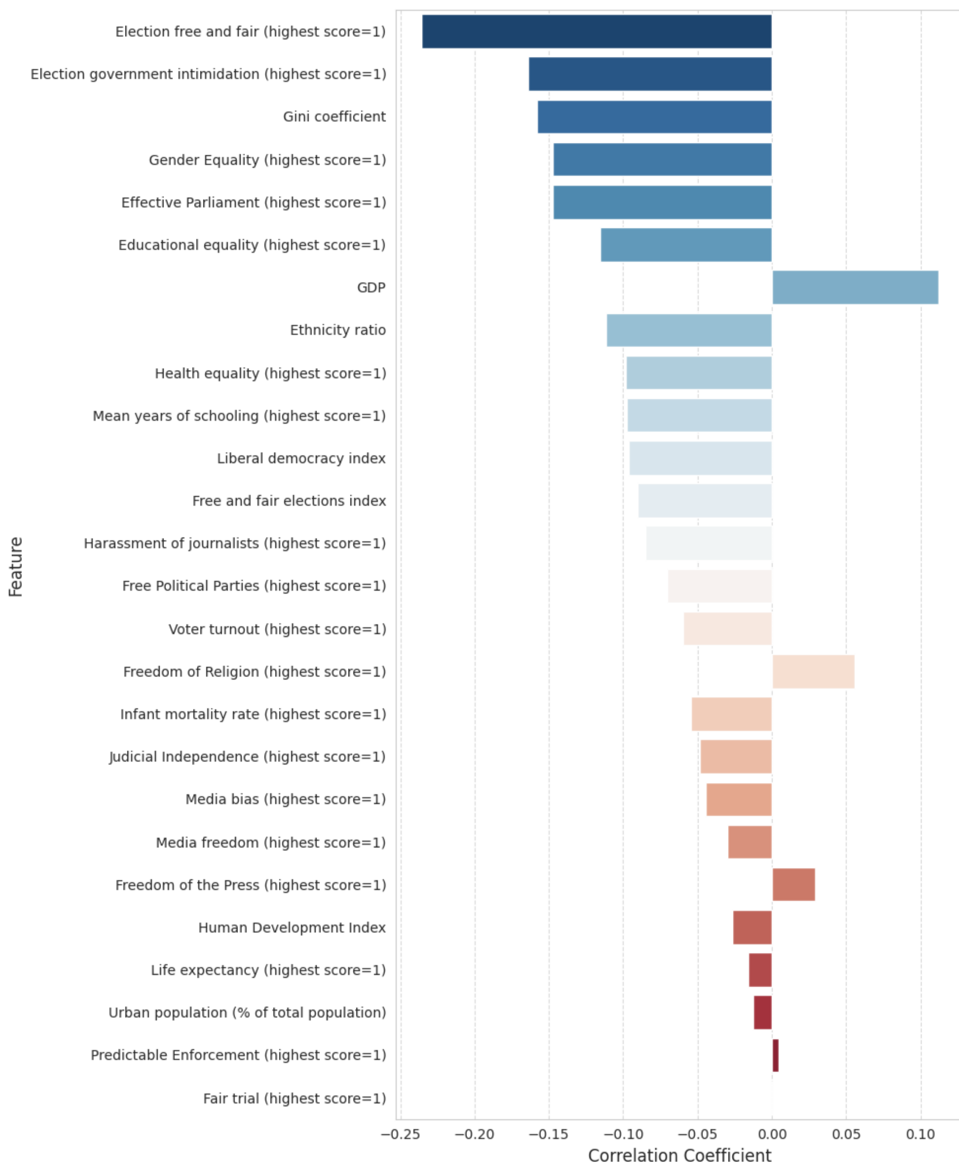

Figure 4.10: Waterfall plot showing the correlation between government change & selected predictors. It's clear that for a government to change, the election must be free and fair; the high-scoring features make sense here.

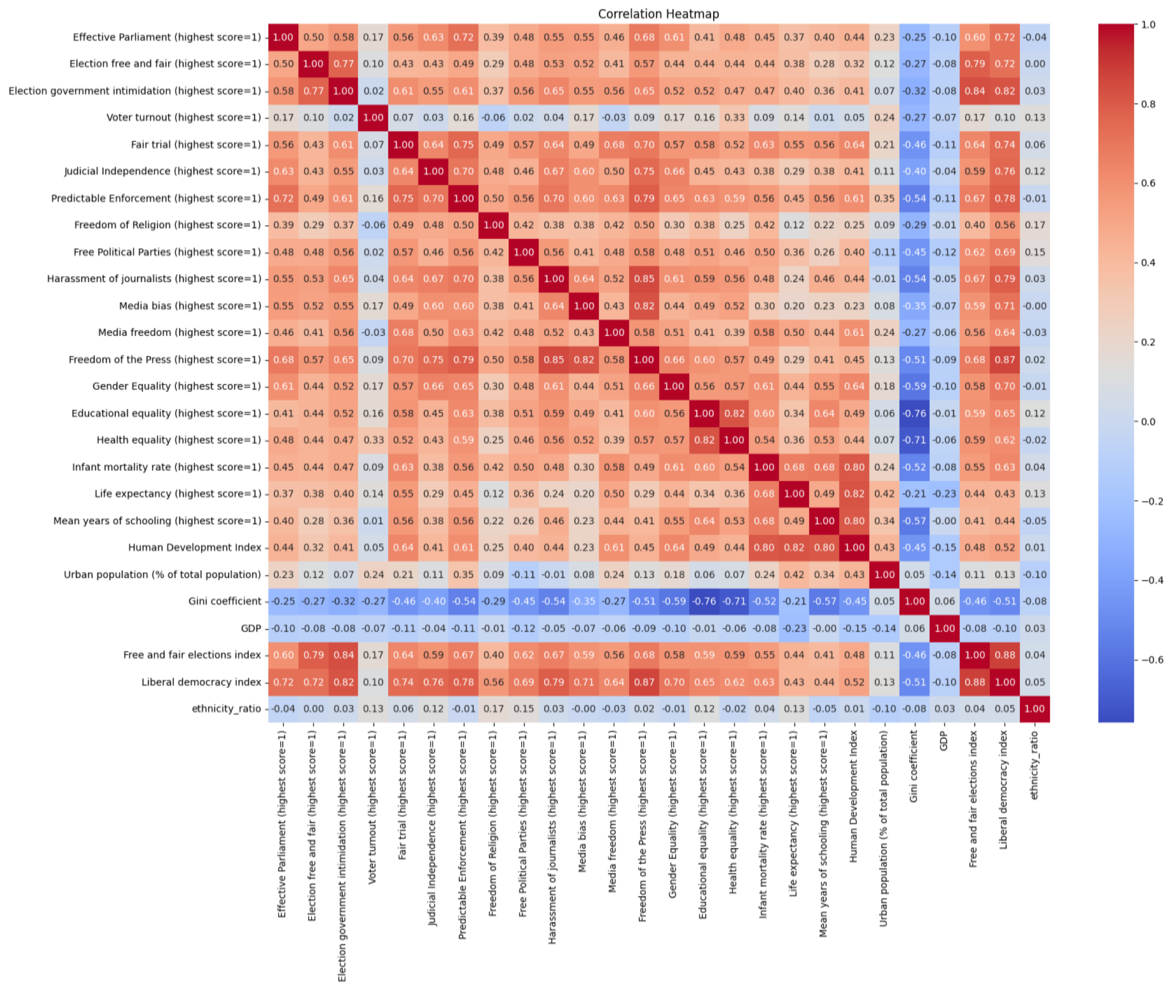

Figure 4.11: Correlation matrix heatmap of the combined dataset.

***Table 4.14*:** Highly correlated features grouped in the related area.

"economic_stability_index"	"Gini coefficient", "GDP"
"political_social_index"	"Effective Parliament (highest score=1)", "Election free and fair (highest score=1)", "Election government intimidation (highest score=1)", "Fair trial (highest score=1)", "Judicial Independence (highest score=1)", "Predictable Enforcement (highest score=1)", "Free Political Parties (highest score=1)", "Harassment of journalists (highest score=1)", "Media bias (highest score=1)", "Media freedom (highest score=1)", "Freedom of the Press (highest score=1)", "Free and fair elections index", "Liberal democracy index", "Gender Equality (highest score=1)", "Educational equality (highest score=1)", "Health equality (highest score=1)", "Infant mortality rate (highest score=1)", "Life expectancy (highest score=1)", "Mean years of schooling (highest score=1)", "Human Development Index"

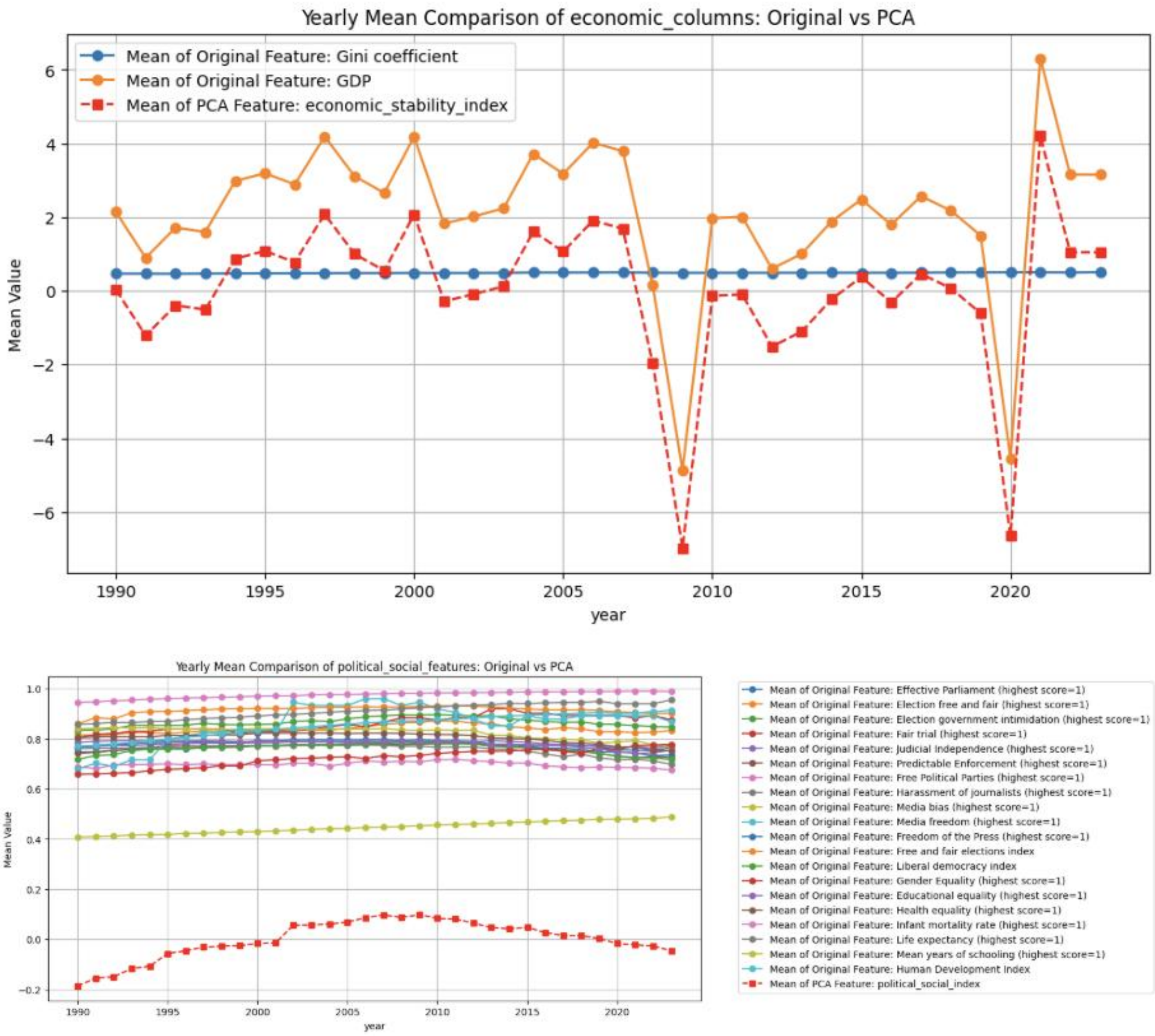

***Figure 4.15*:** Yearly Mean Comparison of original feature groups and new indices.