On the Classification of the Cuban Political Regime: 1959 - 2015

HarvardX Data Science Professional Certificate

Ignacio Vera

2024-06-12

Summary: Classifying the political regime that has prevailed Cuba over the past sixty-five years is challenging, yet essential for assessing the legacies of Fidel and Raúl Castro, interpreting their policy choices, and understanding the island's disproportionate influence on international relations. Ultimately, this classification enriches our understanding of Cuban history and its broader significance. In this project, I address the classification of governance forms in Cuba from 1959 to 2015, treating it as a multi-class classification problem. Using data from four sources, I first train a Naive Bayes classifier, followed by a Random Forest model. Common evaluation metrics reveals that the Naive Bayes model underperforms compared to the random forest. Notably, the Naive Bayes produces unlikely results given the non-democratic nature of the Cuban political regime, while the random forest yields more sensible outcomes. More specifically, the Random Forest results indicate a shift from a personalist rule (1959-2007) to a military rule (2008-2015) during Fidel and Raúl Castro's tenures, respectively. However, this outcome markedly differs from Anckar and Fredriksson's (2020) classification of the entire period as a single-party rule.

I. Introduction

Table 1 below presents an overview of the diverse forms of governance in Cuba from 1902 to 1959, leading up to the rise of the Castros to power.

Table 1: Cuba: Forms of Governance (1902 - 1959)

Period	Years	Classification
1902 - 1905, 1916 - 1933	22	Personalist rule
1934 - 1939, 1952 - 1959	14	Military rule
1940 - 1952	13	Semi-presidentialism
1909 - 1915	7	Presidentialism
1906 - 1908	3	US Military occupation

Note: Based on the Anckar and Fredriksson (2020) dataset

Over this period, the two most frequent forms of the political regimes in Cuba were as follows:

- **Personalist Rule:**, which spanned 22 years in total. This rule is characterized by the dominance of a single leader who centralizes power and often rules autocratically. In Cuba, this occurred during two periods: from 1902 to 1905 and again from 1916 to 1933.
- **Military Rule:**, which occurred for a total of 14 years and is distinguished by the control of the government by military leaders. Even within this classification, the leadership was often centered around a prominent figure or caudillo. A notable example is Fulgencio Batista, who dominated Cuban politics during two key periods, the second of which was a strict military dictatorship (1952 - 1959).

Both personalist and military rules in Cuba shared a common feature: the centralization of power in the hands of a single, dominant leader or caudillo. During the military rule from 1952 to 1959, Fulgencio Batista exemplified this parallel. As a caudillo, Batista maintained undisputed political leadership, consolidating military and political power in his hands, similar to leaders during personalist rule periods. Batista's reign began with a coup in 1952, and he ruled as a dictator until the Cuban Revolution in 1959. This period highlights how, even under the guise of military governance, personalist traits were prominent, with Batista exercising significant personal control over the state.

Not surprisingly, characterizations of the Cuban political regime often reflect the leader's personality, shaping perceptions of the entire political system.

The main objective of this project is to provide a simple yet systematic classification of the Cuban political regime from 1959 to 2015. This timeframe encompasses both Fidel Castro's tenure of more than 49 years (1959 - 2008) and the initial eight years of Raúl Castro's government (2008 - 2015). Fidel's tenure is one of the longest since the 1850s, matched only by Emperor Pedro II of Brazil, and surpassed by three monarchs: Nikola I Petrović-Njegoš of Montenegro, George I of Greece, and Francis Joseph I of Austria (see Table 2 below and Figure 1 for context).

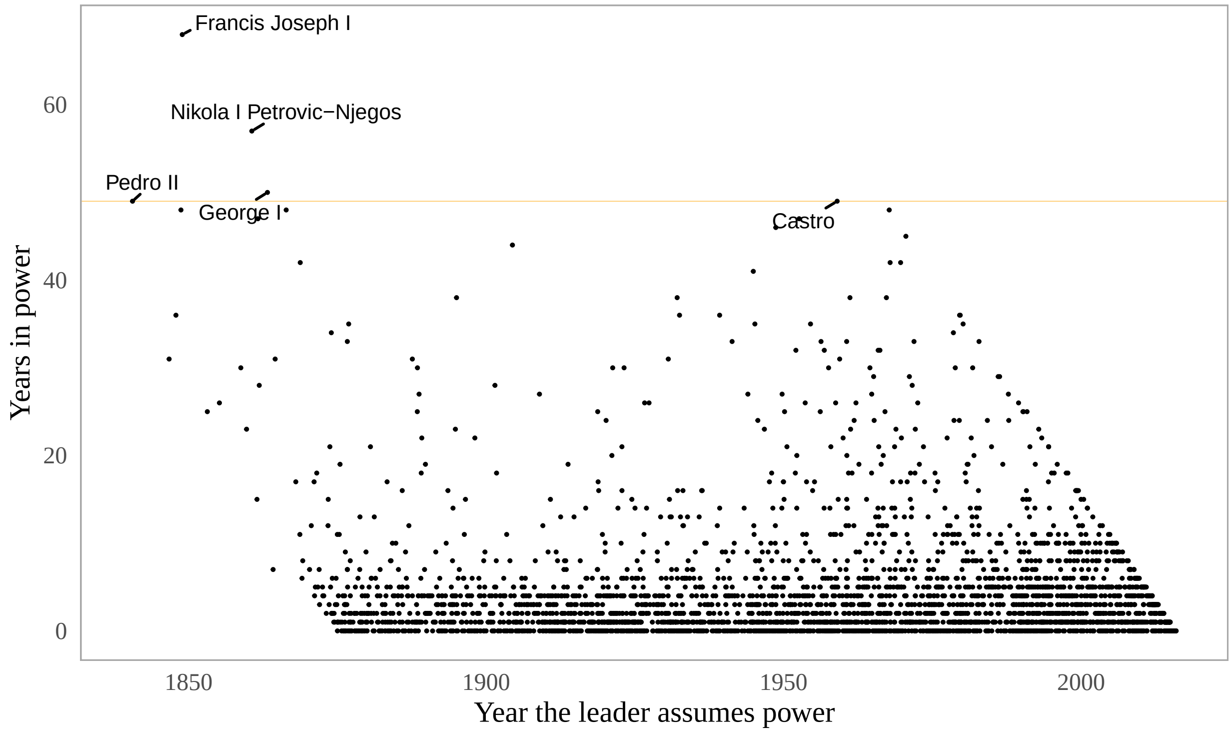
Table 2: The world: Top Five Longest Political Tenures (1840 - 2015)

Leader	Entry date	Exit date	Years in power
Castro	1959-01-02	2008-02-24	49
Pedro II	1840-07-23	1889-11-15	49
Francis Joseph I	1848-12-02	1916-11-21	68
Nikola I Petrovic-Njegos	1860-08-13	1917-01-17	57
George I	1863-03-30	1913-03-18	50

Note: Based on the Goemans et al (2019) dataset

Figure 1: The long tenure of Fidel Castro in historical context

(The orange line intersects the vertical axis at 49, the approximate number of years Fidel Castro held power)



Source: Goemans et al (2019)

Fidel Castro's extraordinarily long rule overshadowed Raúl's presence despite Raúl serving as Vice President, Second Secretary of the Communist Party and Minister of Defense.

On 31 July 2006, Raúl Castro was *"provisionally"* [1] handed power after his brother, Fidel Castro, stepped aside due to illness. The emergence of a softer form of non-democratic regime with a more pluralistic approach to governance became theoretically possible when Fidel Castro announced he had delegated his responsibilities as the *"principal promoter"* [2] of the Health, Education, and National Energy Revolution programs to six loyalists from his inner circle. However, as long as his health allowed, Fidel most likely continued to influence most important political decisions behind the scenes.

Less than two years later, on 24 February 2008, Raúl Castro was formally vested with all powers. During his tenure, Raúl remained conspicuously unseen and astonishingly overshadowed by his brother's dominant personality. Raul's political speeches constantly invoked Fidel, since preserving the facade of traditional Fidelism not only aims to borrow political legitimacy but also serves as the most straightforward means of continuing to exercise absolute power.

In 2019, Raúl Castro handed over control to Miguel Díaz-Canel Bermúdez as President and, since 2021, as First Secretary of the Communist Party of Cuba.

Classifying the political regime governing Cuba over the past sixty-five years has been a challenging endeavor. As noted by Hoffman (2011), *"…Cuba since 1959 can not only be classified as a 'rebel regime,' 'military,' or 'one-party' rule, but also as a case combining charismatic leadership with forms of 'rationalized' bureaucratic authority in ways that have considerably changed over time."* [3]. Consequently, Fidel Castro's leadership style and governance have been subject to diverse interpretations, reflecting the multifaceted nature of the regime. Terms used to describe him include "tyrannic", "populist", "totalitarian", "dictatorial", "non-democratic",

[1] Castro (2008).
[2] Castro (2008).
[3] Hoffman (2011, p. 6)

"gerontocratic", "authoritarian", and the more colorful phrases such as *"genuine dictator"* [4] and *"movie star dictator"* [5], or even *"a dictator held up as a clown"* [6], with some depicting his regime as holding *"the most abject and unreformable of dictatorships"* [7]. Each of these terms highlights different aspects of his rule, from his authoritarian control and suppression of political opposition to his ability to inspire and mobilize support among the masses.

For some observers, the transfer of power from Fidel Castro to his brother Raúl Castro marked a significant change in Cuba's leadership style. As early as 2011, two authors characterized the Cuban political system as going through *"…the final phase of transition from charismatic leadership to institutionalized leadership."* [8] Surprisingly, five years later, one of these authors argued that *"Raúl Castro's presidency has completed a political transition from totalitarianism to post-totalitarianism."* [9] Despite the inherent contradiction in these statements, they suggest a transitions towards an undefined, though still autocratic, form of governance under Raúl Castro.

Ultimately, if the classification of Fidel Castro's era is challenging, that of Raúl's is even more elusive, perhaps reflecting Raúl's reserved and enigmatic nature. However, beyond the superficial differences in personalities and styles, Raúl has consistently shown one defining trait: his profound reverence for his brother. He looks up to Fidel more than any other loyalist, embodying one of the most surreal slogans in contemporary Cuban politics: *Yo soy Fidel (I am Fidel).*

The remainder of this document is organized around the following sections: Section II describes the data and methods used for obtaining and tranforming the dataset used in this project. Section III discusses two classification algorithms (Naive Bayes and Random Forest) used for the classification of the Cuban political regime(s) over the period under consideration (1959 - 2015). This discussion includes a brief overview of the relative advantages and disadvantages of these algorithms. Section IV presents the results obtained with these classification algorithms and assesses their performance. Section V wraps up the project and offers concluding remarks. Finally, section VI discusses limitations of the study and possible extensions for future research.

II. Data and Methods

Data Sources: In this project, I use variables extracted from four different datasets to predict the classification of the political regime in Cuba from 1959 to 2015. Three of these datasets are mostly used in Political Science:

- **Archigos** (Goemans et al, 2009): This dataset is a comprehensive compilation of political leaders and regimes worldwide covering the period 1875 - 2015 (version 4.1). It provides detailed information about the individuals who have held power in various countries, including their tenure, the circumstances of their coming to and leaving office, and the political context of their leadership. This dataset is particularly valuable for studying political stability, leadership changes, and regime types across different countries and historical periods.
- **CHISOLS** (Comparative Historical Indicators of State-Level Societies, Mattes et al, 2016): The CHISOLS dataset is a historical and sociopolitical database that includes a wide range of variables on state-level societies with populations of more than than half a million inhabitants spanning from 1919 to 2018. It encompasses information on social, economic, and political structures, allowing for comparative analysis across different historical periods and regions. The dataset is used for examining long-term trends and patterns in state development, governance, and societal change.

[4] Müllerson (2023, p. 871).
[5] Bardach (2002, Chapter 8).
[6] Pérez Firmat (2010, p. 168).
[7] Dalrymple (p. 145).
[8] Abrahams and López-Levy (2011, p. 92)
[9] López-Levy (2016).

- **Political Regimes of the World, Dataset v. 2.0** (Anckar and Fredriksson, 2020): This dataset provides a comprehensive classification of political regimes globally, including *…all countries that have been independent at any point in time since WWII"*, encompassing the years 1800 - 2019. It offers detailed categorizations and codifications of various regimes based on specific political characteristics and governance structures.

In addition, I use selected demographic variables published by the United Nations Population Division (United Nations, 2022). This dataset provides comprehensive demographic data and projections for countries and regions worldwide, including historical and projected population figures, fertility rates, mortality rates, migration patterns, and other vital statistics.

These datasets are well-documented and consistently coded, having been cleaned and maintained over many years. Consequently, data cleaning was kept to a minimum, for instance, to remove duplicates and missing data. Only minor feature engineering was required, such as extracting years from unique identifiers combining ISO3 country codes [10] with the year in which leaders took office, and estimating the approximate number of years leaders are in power. The most challenging aspects of data manipulation were the creation of a unique identifier to join the UN dataset with the three political datasets used, and the development of a function to fill missing years in two of the political science datasets used (Archigos and CHISOLS). [11]

The final assembled dataset spans from 1950 to 2008, containing 4129 rows and 58 columns, including the target variable.

I use the systematic classification provided by Anckar and Fredriksson (2020) as the target variable. This classification assigns 12 mutually exclusive classes to political regimes [12] Table 3 below presents these categories and their frequencies in the dataset.

Table 3: Distribution of the 12 classes in the target variable

Classes	Frequency	Proportion (%)
Parliamentarism	897	22
Presidentialism	713	17
Military rule	676	16
Personalist rule	458	11
Single-party rule	398	10
Absolute monarchy	302	7
Semi-presidentialism	287	7
Multi-party autoritarian rule	199	5
Missing	101	2
Other oligarchy	43	1
Semi-monarchy	34	1
Monarchic oligarchy	21	1

Note: Based on Anckar and Fredriksson (2020)

The following four classes, listed in the bottom portion of Table 3, each have relatively small proportions (less than 5 percent), resulting in an imbalanced target variable:

- Missing (2 percent)[13]

[10] ISO 3166-1 alpha-3 codes, also referred to as ISO3 country codes, are three-letter identifiers for countries, territories, and special geographical areas. They provide a standardized way to refer to locations and are issued by the International Organization for Standardization (ISO).

[11] Please refer to the code chunk titled *"function_for_filling_missing_years"* and Annex I for an explanation of how this function works.

[12] To the best of my knowledge, this variable (coded as "regimenarrowcat" in the Anckar and Fredriksson (2020) dataset) represents the most granular classification of political regimes available.

[13] This category includes military occupation, civil war or otherwise unclear categories (Anckar and Fredriksson (2020, p. 4).

- Other oligarchy (1 percent)
- Semi-monarchy (1 percent)
- Monarchic oligarchy (1 percent)

To address imbalanced classes in the target variable I used the inverse of the class weights thus allowing the classifies to pay more attention to the minority classes during training. [14]

Predictors: A total of 56 predictors were used in both the Naive Bayes and Random Forest classification algorithms. [15] Please refer to Table 4 in Annex II for details on these predictors, including a brief description, their class, and the data source from which they were selected.

III. Modeling Approach

Anckar and Fredriksson (2020) classify the the Cuban political regime from 1959 to 2015 as a *"Single party rule"*. However, I aim to predict the classification of the Cuban political regime for this period based on the **implicit assumption** that this classification may not be entirely accurate, given significant events that have taken place recently, notably the transfer of power from Fidel Castro to Raúl Castro in 2008.

I use data from various political regimes worldwide between 1950 and 2015, as well as data for the Cuban political regime between 1950 and 1958, to train two classification algorithms: Naive Bayes and Random Forest.

The initial dataset was split into a training set (80 percent) and a testing set (20 precent) after filtering out the subset corresponding to the regimes of Fidel and Raúl Castro from 1959 to 2015.

To handle class imbalance (see Table 3 in Section II for the original distribution of classes in the target variable), I used the inverse of the class weights in training both classification algorithms. The inverse class weights were calculated as follows:

$$\text{Class weights}_i = \frac{\text{Total counts}}{\text{Class counts}_i}$$

Where Total counts is the total number of samples in the dataset and Class counts $_i$ is the number of samples in class i.

Using the inverse of the class weights helps mitigate the effect of class imbalance by ensuring that minority classes are better represented and learned during training. This approach may lead to a more effective model, improving performance on evaluation metrics like precision, recall, and F1-score, especially for the minority classes.

- Model Selection

I selected Naive Bayes and Random Forest because they differ significantly in their learning mechanisms and feature interaction handling. The Naive Bayes is computationally simpler, utilizing Bayes' Theorem to estimate the posterior distribution of the target under the assumption of conditional independence among predictors. This simplifying assumption makes the algorithm less likely to overfit but potentially prone to missing complex feature relationships.

Random Forest, on the other hand, is an ensemble method that builds multiple decision trees using bootstrapped aggregation (bagging). It selects random subsets of predictors for each tree, capturing complex interactions and non-linear relationships. This results in lower bias and higher variance, effectively handling complex patterns but risking overfitting if not properly tuned, especially with smaller datasets.

[14] Refer to section II for more details.

[15] Initially, the Random Forest algorithm utilized 56 predictors. However, following feature importance analysis, only 5 predictors remained significant. Refer to Table 5 in Section III for details.

Overall, Naive Bayes is suited for simpler problems where computing resources are limited, while Random Forest excels in capturing intricate patterns when feature interactions are crucial for more accurate classification.

- Hyperparameter tuning:

For the Naive Bayes algorithm I used a grid search to obtain an optimal combination of values for the following three hyperparameters:

- **fL** If needed, the algorithm employs this parameter to use Laplace correction fL = 1), which adjusts for missing classes in small datasets to prevent zero probabilities.
- **usekernel**, which determines whether kernel density estimation should be used for probability estimation. It is a binary parameter, with options being TRUE or FALSE. When set to TRUE, kernel density estimation is used, and when set to FALSE, a simple density estimation is employed.
- **adjust**, which adjusts the bandwidth of the kernel density estimation if it is used. Lower values result in narrower bandwidths, leading to more "localized" density estimates, while higher values produce broader bandwidths, and smoother density estimates.

The optimal combination of hyparameters for the Naive Bayes classifier was as follows:

Table 4: Tuned parameters for the Naive Bayes classifier

fL	usekernel	adjust
0	1	0

This combination of hyperparameters for the Naive Bayes classifier indicates that no Laplace correction was applied, kernel density estimation was used for continuous predictors, and the bandwidth of the kernel density estimator was set to half the default value, allowing for less smoothing and potentially capturing more detailed structure in the data.[16]

For the Random Forest algorithm I used a grid of values to search for the optimal **mtry** hyperparamenter. This hyperparameter specifies the number of predictors randomly sampled as candidates at each split. The specified grid includes values from 2 to 6, incremented by 1. The optimal*mtry* was 4.

I trained an initial Random Forest model using all 56 predictors to assess feature importance. Importance was measured by the improvement in accuracy and reduction in Gini impurity when a predictor was used in the trees. To obtain more stable results and improve generalization, I increased the number of trees to 1000 (ntree = 1000), compared to the default 500. [17] For the final Random Forest classifier, I also used 1000 trees, the default value for mtry (set to the square root of the number of predictors, which is r floor(sqrt(5))‘ for this dataset), and the five most important predictors, which are:

- Cross-Validation:

I used 5-fold cross-validation (CV), dividing the dataset into five equally-sized subsets. Each iteration of CV reserved one subset as the validation set while training on the remaining 4 subsets. This process repeated 5 times, with each subset serving once as the validation set.

[16]Kernel density estimation (KDE) is a non-parametric method used to estimate the probability density function of continuous variables. Instead of assuming a specific distribution, KDE uses the data to estimate the density function with a kernel function, which smooths the data points.This flexible approach can better capture the true distribution of continuous predictors compared to assuming a normal distribution.

[17]Kunh and Johnson (2013, pp. 199 - 200), *"…suggest using at least 1,000 trees…"*. James et al. (2023, p. 344), in discussing bagging, note that using *"a very large"* number of threes *"will not lead to overfitting"*. In his seminal paper, Breiman (2001), shows that Random Forests do not overfit due to increases in the number of trees (Theorem 1.2, p. 7).

Table 5: The top five most important predictors

Feature	Class	Brief description	Source
democracy	numeric	Binary variable for democracy	Mattes et al (2016)
PopSexRatio	numeric	Population Sex Ratio, as of 1 July	UN Population Division (2022)
CCode	numeric	Country code	UN Population Division (2022)
Years in office	numeric	Years the leader held the position	Goemans et al (2009)
PopDensity	numeric	Population Density, as of 1 July	UN Population Division (2022)

I also experimented with 10-fold cross-validation, following the same procedure as described above. However, I found that the results obtained from 10-fold CV were consistent with those from 5-fold CV. Therefore, I decided to use 5 folds only for computational efficiency. Additionally, using 5 folds allows for a larger proportion of the data to be used for training in each iteration, which can potentially lead to more stable and reliable performance estimates.

For both classifiers, I used parallel processing with the *doParallel* library thus enhancing computational efficiency.

IV. Results

Table 4 below presents the outcomes generated by both algorithms. Notably, Naive Bayes yields unexpected results when considering the observed attributes of the Cuban political regime spanning from 1959 to 2015. If there is one certainty about this political regime, it is its non-democratic nature. Hence, classifications such as *Presidentialism* (1969 to 1971; 1973 to 1978; and 1982 to 1991) and, surprisingly, the most democratic political regime, *Parliamentarism* (between 1992 and 2015, inclusive), push the boundaries of the most logical interpretation. Moreover, it is worth noting the frequent switches of classifications (seven in total), indicate a certain degree of instability in the results. This instability is particularly puzzling given the unchanging nature of both the regime and its key political actors.

The results obtained with the Random Forest classifier better align with the characteristics of the main actors and the perceived nature of the Cuban political regime. Specifically, this algorithm classifies the period between 1959 and 2007, corresponding to Fidel Castro's tenure, as a *Personalistic rule*, while categorizing the remaining years, corresponding to Raúl Castro's tenure, as a *Military rule*. Notably, the only change in classification occurs in 2008, coinciding with the transfer of power between Fidel and Raúl Castro.

Table 6: Classifiers predictions: Cuba's Governance forms (1959-2015)

Years	Naive Bayes	Random forest
1959	Personalist rule	Personalist rule
1960	Personalist rule	Personalist rule
1961	Personalist rule	Personalist rule
1962	Personalist rule	Personalist rule
1963	Personalist rule	Personalist rule
1964	Personalist rule	Personalist rule
1965	Personalist rule	Personalist rule
1966	Single party rule	Personalist rule
1967	Single party rule	Personalist rule
1968	Single party rule	Personalist rule
1969	Presidentialism	Personalist rule
1970	Presidentialism	Personalist rule

Table 6: Classifiers predictions: Cuba's Governance forms (1959-2015) *(continued)*

Years	Naive Bayes	Random forest
1971	Presidentialism	Personalist rule
1972	Single party rule	Personalist rule
1973	Presidentialism	Personalist rule
1974	Presidentialism	Personalist rule
1975	Presidentialism	Personalist rule
1976	Presidentialism	Personalist rule
1977	Presidentialism	Personalist rule
1978	Presidentialism	Personalist rule
1979	Semi presidentialism	Personalist rule
1980	Semi presidentialism	Personalist rule
1981	Semi presidentialism	Personalist rule
1982	Presidentialism	Personalist rule
1983	Presidentialism	Personalist rule
1984	Presidentialism	Personalist rule
1985	Presidentialism	Personalist rule
1986	Presidentialism	Personalist rule
1987	Presidentialism	Personalist rule
1988	Presidentialism	Personalist rule
1989	Presidentialism	Personalist rule
1990	Presidentialism	Personalist rule
1991	Presidentialism	Personalist rule
1992	Parliamentarism	Personalist rule
1993	Parliamentarism	Personalist rule
1994	Parliamentarism	Personalist rule
1995	Parliamentarism	Personalist rule
1996	Parliamentarism	Personalist rule
1997	Parliamentarism	Personalist rule
1998	Parliamentarism	Personalist rule
1999	Parliamentarism	Personalist rule
2000	Parliamentarism	Personalist rule
2001	Parliamentarism	Personalist rule
2002	Parliamentarism	Personalist rule
2003	Parliamentarism	Personalist rule
2004	Parliamentarism	Personalist rule
2005	Parliamentarism	Personalist rule
2006	Parliamentarism	Personalist rule
2007	Parliamentarism	Personalist rule
2008	Parliamentarism	Military rule
2009	Parliamentarism	Military rule
2010	Parliamentarism	Military rule
2011	Parliamentarism	Military rule
2012	Parliamentarism	Military rule
2013	Parliamentarism	Military rule
2014	Parliamentarism	Military rule
2015	Parliamentarism	Military rule

Performance of the Naive Bayes classifier:

As previously mentioned, the Naive Bayes classifier predicts unexpected outcomes when considering the observed attributes of the Cuban political regime from 1959 to 2015. Its overall accuracy is approximately 0.65 (see Table 5 below), with macro precision at 0.63, macro recall at 0.68, and macro F1 at 0.62. [18]

Table 7: Naives Bayes: Accuracy and some related statistics

Accuracy	Kappa	Lower bound	Upper bound	Null	p-value
0.654	0.596	0.62	0.686	0.218	0

The performance metrics in Table 5 above represent overall statistics referring to accuracy and its statistical significance. In particular,

- **Accuracy** measures the proportion of correctly classified instances out of the total instances. It indicates that only about 65 percent of the predictions made by the model were correct.
- **Kappa** measures the agreement between the observed accuracy and the expected accuracy that can be obtained by random guessing. For instance, kappa = 0 indicates agreement equivalent to chance, while kappa = 1 indicates perfect agreement. Therefore, a kappa value of about 0.6 suggests a moderate level of agreement beyond what would be expected by chance alone.

Lower and **Upper** bounds define the confidence interval for the accuracy metric at the 95 percent level.

Null represents the accuracy of a naive classifier that always predicts the majority class. It indicates that the baseline accuracy for the dataset is around 21.8 percent.

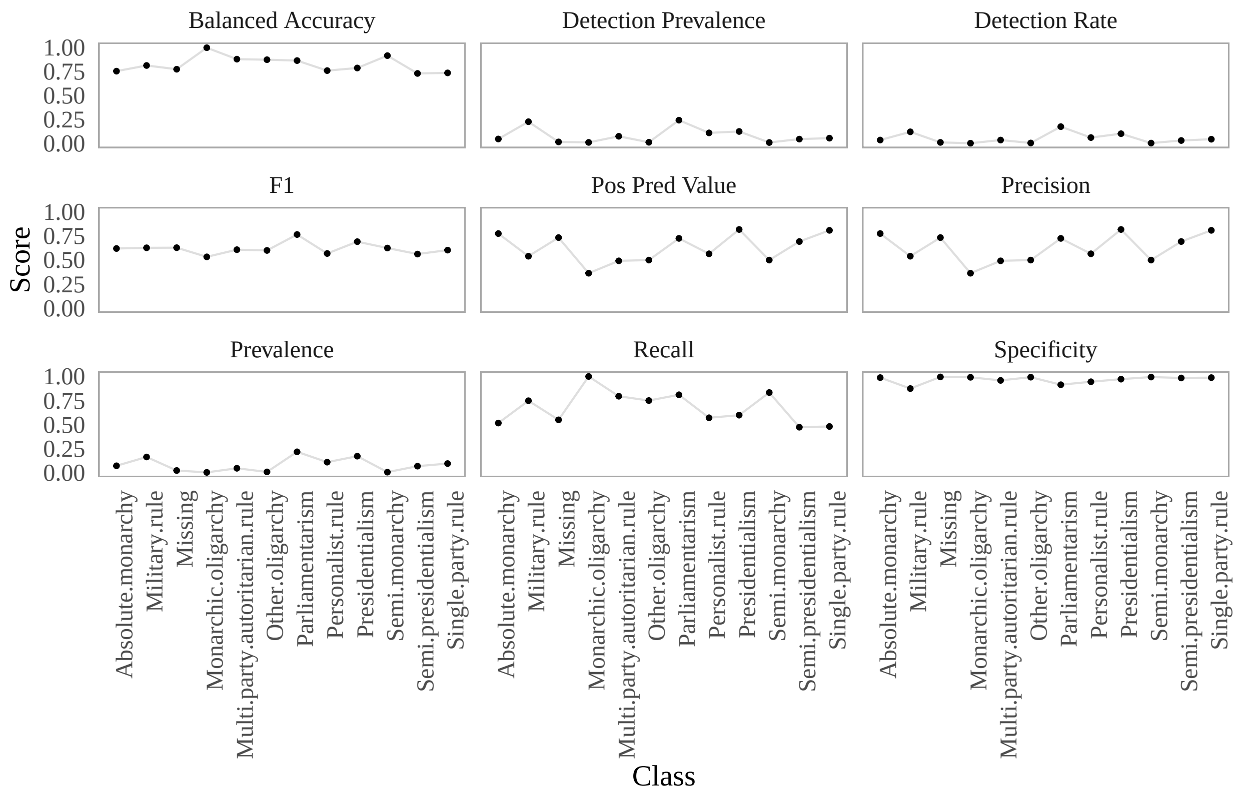
p-value tests the null hypothesis that the model's accuracy is equal to the null accuracy. A very small p-value (close to zero) indicates that the model's accuracy is significantly better than the baseline accuracy.

Figure 2 below focuses on nine out of the eleven metrics provided by the confusion matrix output generated by caret. [19]

[18] All "macro" metrics are class-level averages based on the target variable. Refer to Annex III for concise explanations of these metrics.

[19] Please refer to Annex for the complete set of 11 metrics and to Annex III for explanations on some of these metrics.

Figure 2: Naive Bayes − Confusion matrix by metric



As seen in Figure 2, Specificity tends to be high across most target classes, indicating the model's ability to identify true negatives. However, Recall (also known as Sensitivity), Precision, and the F1 score exhibit more variation, indicating challenges in correctly identifying and accurately predicting certain classes. Balanced accuracy, on the other hand, generally remains high for most classes, suggesting that, despite some variability, the model performs reasonably well across different metrics. The variability in metrics suggests that while the model performs well at avoiding false positives (high Specificity), it sometimes faces challenges with Recall and Precision in specific classes, reflecting the inherent difficulties in multi-class classification tasks. Figure 2 also suggests that while the model performs well for some classes, such as Monarchic Oligarchy (with very high Balanced Accuracy, Sensitivity/Recall, and Specificity), it encounters difficulties with others, such as Personalist Rule and Single-Party Rule, where Sensitivity/Recall are relatively low.

A close inspection of the confusion matrix for the Naive Bayes classifier (Refer to Annex IV) reveals that several classes are poorly predicted. For instance,

- *Military rule*: Despite a relatively high sensitivity (0.748), it has a low positive predictive value (0.540), indicating many false positives.
- *Personalist rule*: Both sensitivity (0.571) and positive predictive value (0.565) are moderate, indicating balanced but not excellent performance.
- *Single party rule*: Low sensitivity (0.481) and balanced accuracy (0.734) suggest difficulties in correct identification.

These metrics, all representing non-democratic forms of governance, clearly highlight the weaknesses of the Naive Bayes classifier. This underscores the need for more robust methods to improve classification accuracy.

Performance of the Random Forest model

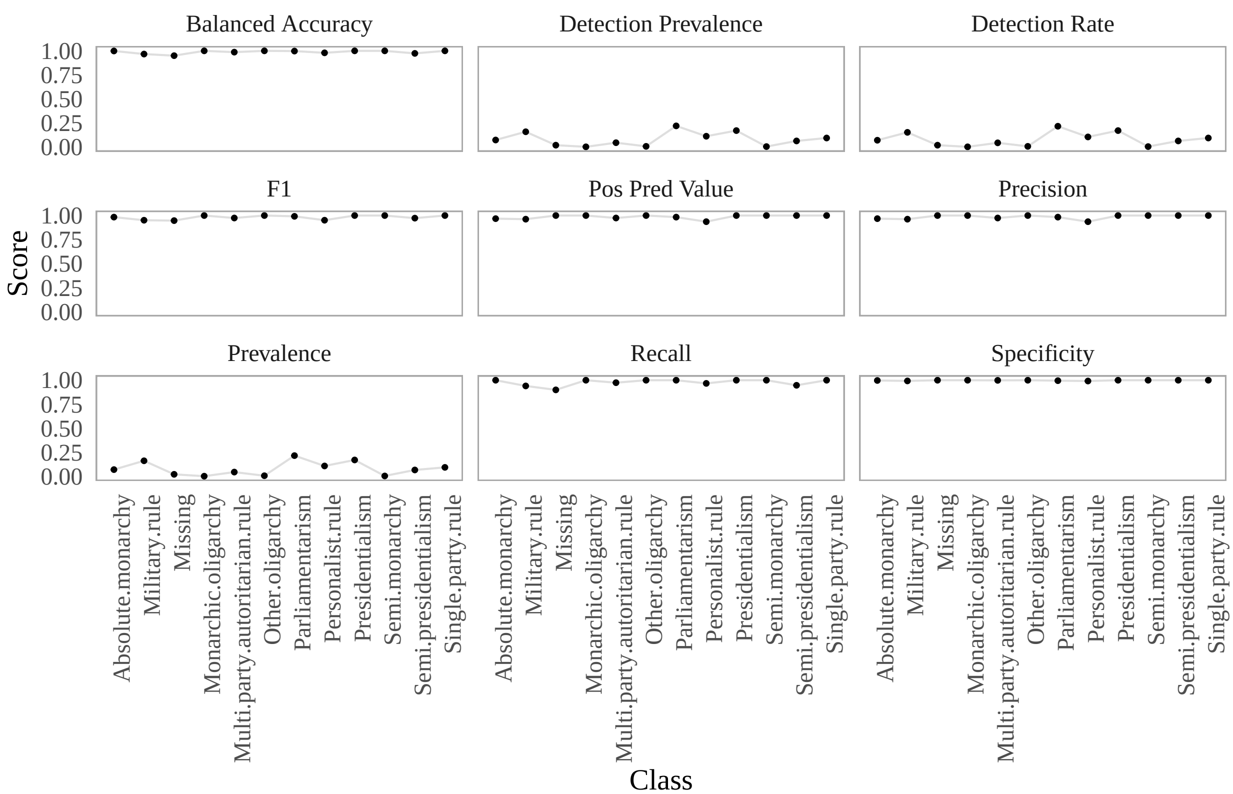
The Random Forest classifier outperforms the Naive Bayes classifier in several key overall performance metrics. Specifically, the Random Forest classifier achieves an accuracy score of 0.98, (95% CI (0.967, 0.988), and a Kappa score of 0.976. Furthermore, it also achieves a macro precision of 0.99, a macro recall of 0.98, and an macro F1 score of 0.98. These near-perfect metrics suggest that the Random Forest classifier accurately predicts the classes with minimal errors. Overall, the Random Forest model shows strong performance and high accuracy. Furthermore, this classifier performs well across most classes, demonstrating perfect or near-perfect metrics in several categories. The *Military rule* class, however, has slightly lower sensitivity and positive predictive value compared to others, indicating some room for improvement in this category.

Table 8: Random Forest: Accuracy and some related statistics

Accuracy	Kappa	Lower bound	Upper bound	Null	p-value
0.979	0.976	0.967	0.988	0.218	0

The superior performance of the Random Forest classifier, even with a limited number of predictors, underscores its robustness. However, a more sensible set of predictors could potentially enhance the performance and reliability of this model.

Figure 3: Random Forest − Confusion matrix by metric



The confusion matrix for the Random Forest classifier (Figure 3 and Annex IV) shows consistently high scores across various metrics for most classes. The metrics of interest are Balanced accuracy, F1 score, Positive predictive value, Precision, Recall, and Specificity.

The high scores for these metrics across multiple classes suggest that the model is performing well in terms of both Recall (Sensitivity) and Specificity, indicating its ability to correctly identify both positive and negative cases. The Positive predictive value and precision scores indicate the model's capability to accurately predict positive cases, while the recall score highlights its ability to capture true positives from the actual positive cases.

When interpreting these high scores, it's crucial to acknowledge the potential for overfitting. [20]. Overfitting transpires when the model memorizes the training data instead of extrapolating from it, leading to overly optimistic performance metrics.

However, a more plausible explanation is that the majority of the variability in the specified Random Forest model may be arising from the bootstrapping process. This is evident as the tuned *mtry* parameter closely approximates the total number of optimal features. Further refinement could potentially result in a broader array of features available for selection by the bagging mechanism.

V. Conclusion

Classifying the political regime that has governed Cuba over the past sixty-five years is inherently challenging. In this project, I tackle the issue of the classification of governance forms prevailing in Cuba from 1959 to 2015, treating it as a multi-class classification problem. After combining data from four different sources, I initially train a Naive Bayes classifier, followed by a Random Forest model. Evaluation of common metrics reveals that the Random Forest outperforming the Naive Bayes model. Notably, while the Naive Bayes model produces improbable results given the non-democratic nature of the Cuban political regime, the Random Forest classifier yields more plausible outcomes. Specifically, the Random Forest indicates a transition from personalist rule (1959-2007) to military rule (2008-2015) during the tenures of Fidel and Raúl Castro, respectively. However, this finding starkly contrasts with Anckar and Fredriksson's (2020) classification of the entire period as single-party rule.

VI. Limitations and further analysis

An important limitation of this project arises from the challenges associated with measuring the target variable, which is inherently unobservable and thus susceptible to subjectivity.

Admittedly, however, the primary constraints of this project stem from my limited expertise in both political science and data science.

Further analysis on classifying the Cuban political regime may involve exploring several avenues:

Expanding the dataset: Broadening the dataset to encompass a wider array of variables from diverse sources could provide a more comprehensive understanding of the Cuban political regime. This could involve considering additional predictors such as the leader's affiliation and their last held political position, which were not included in this project.[21]

Expanding the time coverage: Extending the time coverage of the analysis to include data for more recent years could offer additional insights into the evolution of the Cuban political regime.

Utilizing ensemble methods: Exploring ensemble methods, like stacking, can enhance both robustness and accuracy by combining multiple models' predictions.

Integrating a more systematic approach: Incorporating a systematic theoretical framework by referencing relevant literature from both data science and political science fields could enhance the analysis's rigor and depth and informing modeling decisions.

Annex I - Function to fill missing years in the Archigos and CHISOLS datasets

The function *"fill_missing_years"* fills in the intermediate, non-existent years in datasets like Archigos, which only include records (rows) for the year when a leader enters (inaugurates term). In contrast, datasets

[20] Recent insights, such as those discussed in Molas (2022) blog post, suggest that Random Forests can indeed overfit, deviating from their generalization capabilities

[21] These predictors were not considered in this project due to exceeding the allowable limit of classes that a feature in caret, 53. Exploring techniques to reduce the number of classes, such as one-hot encoding or using forcats::fct_lump(), was not feasible within the time constraints of this study.

such as Anckar and Fredriksson (2020) and UN include records for all years within their covered periods.

To illustrate, let's assume we want to join the Archigos and UN datasets. Without any adjustments, this would mean losing all the information for the years present in the UN dataset but missing in Archigos. For example, consider Nixon and Carter, the American presidents that assumed office in 1969 and 1977, respectively. Joining the Archigos and Anckar and Fredriksson (2020) datasets as they are would only preserve information for the years 1969 and 1977 (See Table 5 below).

Table 9: Joining datasets without any adjustements

year	leader	yrborn	startdate	entry
1969	Nixon	1913	1969-01-20	Regular
1977	Carter	1924	1977-01-20	Regular

The function *"fill_missing_years"* addresses this issue by replicating the data from 1969 across the intermediate years (1970 to 1976) and adding these years to the dataset. This ensures that no data from the other two datasets are lost during the merge (See Table 6 below).

However, an important concern with this approach is the potential introduction of near-zero variance in some predictors when estimating probabilities with the Naive Bayes algorithm. [22], although this may not be an issue for tree-based methods like random forests. This function effectively duplicates the information for the predictors over the years. However, the only feature exhibiting near-zero variance is gender, and it was consequently removed.

[22] In Naive Bayes, near-zero variance can lead to probabilities that are extremely close to 0 or 1, resulting in numerical instability.

Table 10: Joining datasets using the function *fill_missing_years*

year	leader	yrborn	startdate	entry
1969	Nixon	1913	1969-01-20	Regular
1970	Nixon	1913	1969-01-20	Regular
1971	Nixon	1913	1969-01-20	Regular
1972	Nixon	1913	1969-01-20	Regular
1973	Nixon	1913	1969-01-20	Regular
1977	Carter	1924	1977-01-20	Regular

Table 11: Near-zero variance in one feature: gender

	freqRatio	percentUnique	zeroVar	nzv
gender	35.22	0.05	FALSE	TRUE

Annex II - Predictors in the initial dataset

Table 12: Selected predictors

Predictor	Class	Brief description	Source
yrborn	numeric	Year the leader was born	Goemans et al (2009)
entry	categorical	How the leader reaches power	Goemans et al (2009)
prevtimesinoffice	numeric	Number of times a leader has previously been in office	Goemans et al (2009)
fties	categorical	Family ties to a previous or future leader	Goemans et al (2009)
Years in office	numeric	Years the leader held the position	Goemans et al (2009)
numeric_year	numeric	Year leader enters office	Mattes et al (2016)
democracy	numeric	Binary variable for democracy	Mattes et al (2016)
CCode	numeric	Country code	UN Population Division (2022)
PopDensity	numeric	PopDensity	UN Population Division (2022)
PopSexRatio	numeric	PopSexRatio	UN Population Division (2022)
MedianAgePop	numeric	MedianAgePop	UN Population Division (2022)
NatChange	numeric	NatChange	UN Population Division (2022)
NatChangeRT	numeric	NatChangeRT	UN Population Division (2022)
PopChange	numeric	PopChange	UN Population Division (2022)
PopGrowthRate	numeric	PopGrowthRate	UN Population Division (2022)
Births	numeric	Births	UN Population Division (2022)
Births1519	numeric	Births1519	UN Population Division (2022)
CBR	numeric	CBR	UN Population Division (2022)
TFR	numeric	TFR	UN Population Division (2022)
NRR	numeric	NRR	UN Population Division (2022)
MAC	numeric	MAC	UN Population Division (2022)
SRB	numeric	SRB	UN Population Division (2022)
Deaths	numeric	Deaths	UN Population Division (2022)
DeathsMale	numeric	DeathsMale	UN Population Division (2022)
DeathsFemale	numeric	DeathsFemale	UN Population Division (2022)
CDR	numeric	CDR	UN Population Division (2022)
LEx	numeric	LEx	UN Population Division (2022)
LExMale	numeric	LExMale	UN Population Division (2022)
LExFemale	numeric	LExFemale	UN Population Division (2022)
LE15	numeric	LE15	UN Population Division (2022)
LE15Male	numeric	LE15Male	UN Population Division (2022)
LE15Female	numeric	LE15Female	UN Population Division (2022)

Table 12: Selected predictors *(continued)*

Predictor	Class	Brief description	Source
LE65	numeric	LE65	UN Population Division (2022)
LE65Male	numeric	LE65Male	UN Population Division (2022)
LE65Female	numeric	LE65Female	UN Population Division (2022)
LE80	numeric	LE80	UN Population Division (2022)
LE80Male	numeric	LE80Male	UN Population Division (2022)
LE80Female	numeric	LE80Female	UN Population Division (2022)
IMR	numeric	IMR	UN Population Division (2022)
LBsurvivingAge1	numeric	LBsurvivingAge1	UN Population Division (2022)
Under5Deaths	numeric	Under5Deaths	UN Population Division (2022)
Q5	numeric	Q5	UN Population Division (2022)
Q0040	numeric	Q0040	UN Population Division (2022)
Q0040Male	numeric	Q0040Male	UN Population Division (2022)
Q0040Female	numeric	Q0040Female	UN Population Division (2022)
Q0060	numeric	Q0060	UN Population Division (2022)
Q0060Male	numeric	Q0060Male	UN Population Division (2022)
Q0060Female	numeric	Q0060Female	UN Population Division (2022)
Q1550	numeric	Q1550	UN Population Division (2022)
Q1550Male	numeric	Q1550Male	UN Population Division (2022)
Q1550Female	numeric	Q1550Female	UN Population Division (2022)
Q1560	numeric	Q1560	UN Population Division (2022)
Q1560Male	numeric	Q1560Male	UN Population Division (2022)
Q1560Female	numeric	Q1560Female	UN Population Division (2022)
NetMigrations	numeric	NetMigrations	UN Population Division (2022)
CNMR	numeric	CNMR	UN Population Division (2022)

Annex III - Common perfomance metrics overview

Accuracy Accuracy measures the overall correctness of the model's predictions, taking into account both true positive (TP) and true negative (TN) predictions, and is calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Precision Precision measures the proportion of correctly predicted positive instances (TP) among all instances predicted as positive (TP + FP). It focuses on the accuracy of positive predictions, and the formula is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall (Sensitivity) Recall measures the proportion of correctly predicted positive instances (TP) among all actual positive instances (TP + FN). It focuses on the ability of the model to capture all positive instances, and the formula is:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-Score The F1-Score is the harmonic mean of precision and recall, providing a balance between these two metrics, and is calculated as:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Annex IV - Confusion matrices

Confusion matrix for the Naive Bayes classifier

```
## Confusion Matrix and Statistics
##
##                                  Reference
## Prediction                        Absolute.monarchy Military.rule Missing
##   Absolute.monarchy                              31             2       0
##   Military.rule                                   8           101       5
##   Missing                                         0             0      11
##   Monarchic.oligarchy                             0             0       1
##   Multi.party.autoritarian.rule                   0             4       0
##   Other.oligarchy                                 0             2       0
##   Parliamentarism                                14             2       1
##   Personalist.rule                                6            12       1
##   Presidentialism                                 0             5       1
##   Semi.monarchy                                   0             0       0
##   Semi.presidentialism                            0             1       0
##   Single.party.rule                               1             6       0
##                                  Reference
## Prediction                        Monarchic.oligarchy
##   Absolute.monarchy                                 0
##   Military.rule                                     0
##   Missing                                           0
##   Monarchic.oligarchy                               4
```

```
##   Multi.party.autoritarian.rule                  0
##   Other.oligarchy                                0
##   Parliamentarism                                0
##   Personalist.rule                               0
##   Presidentialism                                0
##   Semi.monarchy                                  0
##   Semi.presidentialism                           0
##   Single.party.rule                              0
##                              Reference
## Prediction                    Multi.party.autoritarian.rule Other.oligarchy
##   Absolute.monarchy                                       0               0
##   Military.rule                                           4               2
##   Missing                                                 0               0
##   Monarchic.oligarchy                                     0               0
##   Multi.party.autoritarian.rule                          31               0
##   Other.oligarchy                                         0               6
##   Parliamentarism                                         0               0
##   Personalist.rule                                        0               0
##   Presidentialism                                         3               0
##   Semi.monarchy                                           0               0
##   Semi.presidentialism                                    0               0
##   Single.party.rule                                       1               0
##                              Reference
## Prediction                    Parliamentarism Personalist.rule
##   Absolute.monarchy                         1                5
##   Military.rule                             8               16
##   Missing                                   1                3
##   Monarchic.oligarchy                       4                0
##   Multi.party.autoritarian.rule             2                4
##   Other.oligarchy                           3                0
##   Parliamentarism                         145                2
##   Personalist.rule                          6               52
##   Presidentialism                           2                6
##   Semi.monarchy                             2                0
##   Semi.presidentialism                      5                2
##   Single.party.rule                         0                1
##                              Reference
## Prediction                    Presidentialism Semi.monarchy
##   Absolute.monarchy                         0             0
##   Military.rule                            14             0
##   Missing                                   0             0
##   Monarchic.oligarchy                       0             1
##   Multi.party.autoritarian.rule            13             0
##   Other.oligarchy                           0             0
##   Parliamentarism                          21             0
##   Personalist.rule                          8             0
##   Presidentialism                          85             0
##   Semi.monarchy                             0             5
##   Semi.presidentialism                      1             0
##   Single.party.rule                         0             0
##                              Reference
## Prediction                    Semi.presidentialism Single.party.rule
##   Absolute.monarchy                              0                 1
##   Military.rule                                 14                15
```

```
##   Missing                                       0                 0
##   Monarchic.oligarchy                           0                 1
##   Multi.party.autoritarian.rule                 3                 6
##   Other.oligarchy                               0                 1
##   Parliamentarism                              10                 5
##   Personalist.rule                              1                 6
##   Presidentialism                               2                 0
##   Semi.monarchy                                 0                 3
##   Semi.presidentialism                         27                 3
##   Single.party.rule                             0                38
##
## Overall Statistics
##
##                Accuracy : 0.6537
##                  95% CI : (0.62, 0.6862)
##     No Information Rate : 0.2183
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.5963
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Absolute.monarchy Class: Military.rule
## Sensitivity                           0.51667               0.7481
## Specificity                           0.98816               0.8745
## Pos Pred Value                        0.77500               0.5401
## Neg Pred Value                        0.96282               0.9463
## Prevalence                            0.07317               0.1646
## Detection Rate                        0.03780               0.1232
## Detection Prevalence                  0.04878               0.2280
## Balanced Accuracy                     0.75241               0.8113
##                      Class: Missing Class: Monarchic.oligarchy
## Sensitivity                 0.55000                   1.000000
## Specificity                 0.99500                   0.991422
## Pos Pred Value              0.73333                   0.363636
## Neg Pred Value              0.98882                   1.000000
## Prevalence                  0.02439                   0.004878
## Detection Rate              0.01341                   0.004878
## Detection Prevalence        0.01829                   0.013415
## Balanced Accuracy           0.77250                   0.995711
##                      Class: Multi.party.autoritarian.rule
## Sensitivity                                       0.79487
## Specificity                                       0.95903
## Pos Pred Value                                    0.49206
## Neg Pred Value                                    0.98943
## Prevalence                                        0.04756
## Detection Rate                                    0.03780
## Detection Prevalence                              0.07683
## Balanced Accuracy                                 0.87695
##                      Class: Other.oligarchy Class: Parliamentarism
## Sensitivity                        0.750000                 0.8101
## Specificity                        0.992611                 0.9142
```

```
## Pos Pred Value                      0.500000                  0.7250
## Neg Pred Value                      0.997525                  0.9452
## Prevalence                          0.009756                  0.2183
## Detection Rate                      0.007317                  0.1768
## Detection Prevalence                0.014634                  0.2439
## Balanced Accuracy                   0.871305                  0.8621
##                      Class: Personalist.rule Class: Presidentialism
## Sensitivity                            0.57143                 0.5986
## Specificity                            0.94513                 0.9720
## Pos Pred Value                         0.56522                 0.8173
## Neg Pred Value                         0.94643                 0.9204
## Prevalence                             0.11098                 0.1732
## Detection Rate                         0.06341                 0.1037
## Detection Prevalence                   0.11220                 0.1268
## Balanced Accuracy                      0.75828                 0.7853
##                      Class: Semi.monarchy Class: Semi.presidentialism
## Sensitivity                      0.833333                     0.47368
## Specificity                      0.993857                     0.98427
## Pos Pred Value                   0.500000                     0.69231
## Neg Pred Value                   0.998765                     0.96159
## Prevalence                       0.007317                     0.06951
## Detection Rate                   0.006098                     0.03293
## Detection Prevalence             0.012195                     0.04756
## Balanced Accuracy                0.913595                     0.72898
##                      Class: Single.party.rule
## Sensitivity                           0.48101
## Specificity                           0.98785
## Pos Pred Value                        0.80851
## Neg Pred Value                        0.94696
## Prevalence                            0.09634
## Detection Rate                        0.04634
## Detection Prevalence                  0.05732
## Balanced Accuracy                     0.73443
```

Confusion matrix for the Random Forest classifier

```
## Confusion Matrix and Statistics
##
##                                Reference
## Prediction                       Absolute.monarchy Military.rule Missing
##   Absolute.monarchy                             60             1       0
##   Military.rule                                  0           127       2
##   Missing                                        0             0      18
##   Monarchic.oligarchy                            0             0       0
##   Multi.party.autoritarian.rule                  0             1       0
##   Other.oligarchy                                0             0       0
##   Parliamentarism                                0             0       0
##   Personalist.rule                               0             6       0
##   Presidentialism                                0             0       0
##   Semi.monarchy                                  0             0       0
##   Semi.presidentialism                           0             0       0
##   Single.party.rule                              0             0       0
##                                Reference
```

```
## Prediction                     Monarchic.oligarchy
##   Absolute.monarchy                              0
##   Military.rule                                  0
##   Missing                                        0
##   Monarchic.oligarchy                            4
##   Multi.party.autoritarian.rule                  0
##   Other.oligarchy                                0
##   Parliamentarism                                0
##   Personalist.rule                               0
##   Presidentialism                                0
##   Semi.monarchy                                  0
##   Semi.presidentialism                           0
##   Single.party.rule                              0
##                                 Reference
## Prediction                     Multi.party.autoritarian.rule Other.oligarchy
##   Absolute.monarchy                                        0               0
##   Military.rule                                            1               0
##   Missing                                                  0               0
##   Monarchic.oligarchy                                      0               0
##   Multi.party.autoritarian.rule                           38               0
##   Other.oligarchy                                          0               8
##   Parliamentarism                                          0               0
##   Personalist.rule                                         0               0
##   Presidentialism                                          0               0
##   Semi.monarchy                                            0               0
##   Semi.presidentialism                                     0               0
##   Single.party.rule                                        0               0
##                                 Reference
## Prediction                     Parliamentarism Personalist.rule
##   Absolute.monarchy                          0                1
##   Military.rule                              0                2
##   Missing                                    0                0
##   Monarchic.oligarchy                        0                0
##   Multi.party.autoritarian.rule              0                0
##   Other.oligarchy                            0                0
##   Parliamentarism                          179                0
##   Personalist.rule                           0               88
##   Presidentialism                            0                0
##   Semi.monarchy                              0                0
##   Semi.presidentialism                       0                0
##   Single.party.rule                          0                0
##                                 Reference
## Prediction                     Presidentialism Semi.monarchy
##   Absolute.monarchy                          0             0
##   Military.rule                              0             0
##   Missing                                    0             0
##   Monarchic.oligarchy                        0             0
##   Multi.party.autoritarian.rule              0             0
##   Other.oligarchy                            0             0
##   Parliamentarism                            0             0
##   Personalist.rule                           0             0
##   Presidentialism                          142             0
##   Semi.monarchy                              0             6
##   Semi.presidentialism                       0             0
```

```
##   Single.party.rule                          0             0
##                                Reference
## Prediction                      Semi.presidentialism Single.party.rule
##   Absolute.monarchy                                0                 0
##   Military.rule                                    0                 0
##   Missing                                          0                 0
##   Monarchic.oligarchy                              0                 0
##   Multi.party.autoritarian.rule                    0                 0
##   Other.oligarchy                                  0                 0
##   Parliamentarism                                  3                 0
##   Personalist.rule                                 0                 0
##   Presidentialism                                  0                 0
##   Semi.monarchy                                    0                 0
##   Semi.presidentialism                            54                 0
##   Single.party.rule                                0                79
##
## Overall Statistics
##
##                Accuracy : 0.9793
##                  95% CI : (0.967, 0.9879)
##     No Information Rate : 0.2183
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9759
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Absolute.monarchy Class: Military.rule
## Sensitivity                           1.00000               0.9407
## Specificity                           0.99737               0.9927
## Pos Pred Value                        0.96774               0.9621
## Neg Pred Value                        1.00000               0.9884
## Prevalence                            0.07317               0.1646
## Detection Rate                        0.07317               0.1549
## Detection Prevalence                  0.07561               0.1610
## Balanced Accuracy                     0.99868               0.9667
##                      Class: Missing Class: Monarchic.oligarchy
## Sensitivity                 0.90000                   1.000000
## Specificity                 1.00000                   1.000000
## Pos Pred Value              1.00000                   1.000000
## Neg Pred Value              0.99751                   1.000000
## Prevalence                  0.02439                   0.004878
## Detection Rate              0.02195                   0.004878
## Detection Prevalence        0.02195                   0.004878
## Balanced Accuracy           0.95000                   1.000000
##                      Class: Multi.party.autoritarian.rule
## Sensitivity                                       0.97436
## Specificity                                       0.99872
## Pos Pred Value                                    0.97436
## Neg Pred Value                                    0.99872
## Prevalence                                        0.04756
## Detection Rate                                    0.04634
```

```
## Detection Prevalence                              0.04756
## Balanced Accuracy                                 0.98654
##                      Class: Other.oligarchy Class: Parliamentarism
## Sensitivity                        1.000000                 1.0000
## Specificity                        1.000000                 0.9953
## Pos Pred Value                     1.000000                 0.9835
## Neg Pred Value                     1.000000                 1.0000
## Prevalence                         0.009756                 0.2183
## Detection Rate                     0.009756                 0.2183
## Detection Prevalence               0.009756                 0.2220
## Balanced Accuracy                  1.000000                 0.9977
##                      Class: Personalist.rule Class: Presidentialism
## Sensitivity                           0.9670                 1.0000
## Specificity                           0.9918                 1.0000
## Pos Pred Value                        0.9362                 1.0000
## Neg Pred Value                        0.9959                 1.0000
## Prevalence                            0.1110                 0.1732
## Detection Rate                        0.1073                 0.1732
## Detection Prevalence                  0.1146                 0.1732
## Balanced Accuracy                     0.9794                 1.0000
##                      Class: Semi.monarchy Class: Semi.presidentialism
## Sensitivity                      1.000000                     0.94737
## Specificity                      1.000000                     1.00000
## Pos Pred Value                   1.000000                     1.00000
## Neg Pred Value                   1.000000                     0.99608
## Prevalence                       0.007317                     0.06951
## Detection Rate                   0.007317                     0.06585
## Detection Prevalence             0.007317                     0.06585
## Balanced Accuracy                1.000000                     0.97368
##                      Class: Single.party.rule
## Sensitivity                           1.00000
## Specificity                           1.00000
## Pos Pred Value                        1.00000
## Neg Pred Value                        1.00000
## Prevalence                            0.09634
## Detection Rate                        0.09634
## Detection Prevalence                  0.09634
## Balanced Accuracy                     1.00000
```

References

Abrahams and Arturo López-Levy (2011). *Raúl Castro and the New Cuba.* Mcfarland and Company Inc.

Anckar, Carsten and Cecilia Fredriksson (2019). *Classifying political regimes 1800–2016: a typology and a new dataset.* European Political Science, Vol. 18, pp. 84-96.

Anckar, Carsten and Cecilia Fredriksson (2020). *Political Regimes of the World* and Dataset, v. 2.0.

Breiman, Leo (2001). *Random Forests.* Machine Learning, 45, pp. 5–32.

Castro, Fidel (2008). *Proclamation of Cuban transfer of duties 2006.* https://en.wikisource.org/wiki/Proclamation_of_Cuban_transfer_of_duties_2006

Dalrymple, Theodore (2012) *The Wilder Shores of Marx: Journeys in a Vanishing World.* Monday Books. Hutchinson.

Goemans, Henk E., Kristian Skrede Gleditsch, Giacomo Chiozza (2009). *Introducing Archigos: A Data Set of Political Leaders*. Journal of Peace Research, Vol. 46, No. 2, March, pp.

Hoffmann, Bert (2011). *The International Dimensions of Authoritarian Legitimation.* German Institute for Global and Area Studies. Working Papers, No. 182

James, Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani and Jonathan Taylor (2023). *An Introduction to Statistical Learning with Applications in Python.* Springe Texts in Statistics.

Kuhn, Max and Kjell Johnson (2013). *Applied Predictive Modeling.* Springer, New York, Heidelberg, Dordrecht, London.

López-Levy, Arturo (2016). *"Cuba After Fidel: Economic Reform, Political Liberalization and Foreign Policy (2006–2014).* Univefsity of Denver. Electronic Theses and Dissertations. 1212.

Mattes, Michaela, Brett Ashley Leeds, and Naoko Matsumura (2016). *Measuring Change in Source of Leader Support: The CHISOLS Dataset.* Journal of Peace Research 53(2): 259-267.

Müllerson, Rein (2023). *E Pluribus Unum – A Dangerous Concept for the World since Not Always Those Who Are Not Like Us Are Against Us* European Law Open, Vol. 2, pp. 857 - 879.

Pérez Firmat, Gustavo (2010). *The Havana Habit.* Yale University Press. New Haven and London.

United Nations (2022). *World Population Prospects.* UN Department of Economic and Social Affairs, Population Division. Online Edition (https://population.un.org/wpp/Download/Standard/CSV/)