Loralee Ryan
Max Wienandts
Ronan Fonseca
Autumn Dorsey

Introduction

This study tries to determine the most relevant factors for predicting coups d'état using data for 943 coup events from 1945 to 2019, using four different modeling techniques, logistic regression, decision trees, LightGBM and XGBoost. Models had similar accuracy performance in the test set, with Logistic Regression having the best performance in terms of classification accuracy on the testing set. Using SHAP metrics to compare variable impact between models, the results showed that all models selected the same set of predictors as their top four most important. These four factors were related to whether or not the coup was initiated by (1) a dissident group outside the current government, (2) military actors who are formally part of the government, (3) a faction inside the existing government, and (4) a popular revolt. Though each of the models did not exactly coincide in the order in which these factors were ranked, all of them considered the presence of a dissident group as a coup initiator to be far more important than the other predictors.

Motivation and Context

Coups are political events that significantly change the course of a country and its people. The overt-hrowing of the current government can signify the beginning of free, peaceful times, or the beginning of authoritarian periods marked by human rights violations. Given their impact, and the fact that coups still happen to this day, it is worth studying what factors are most relevant for determining a coup's success, so that we can better analyze those coups that happened in the past, and understand what can predict those that will happen in the future. This study is particularly meaningful for the members of this group, given the tumultuous processes for the latest presidential elections in Brazil and the United States.

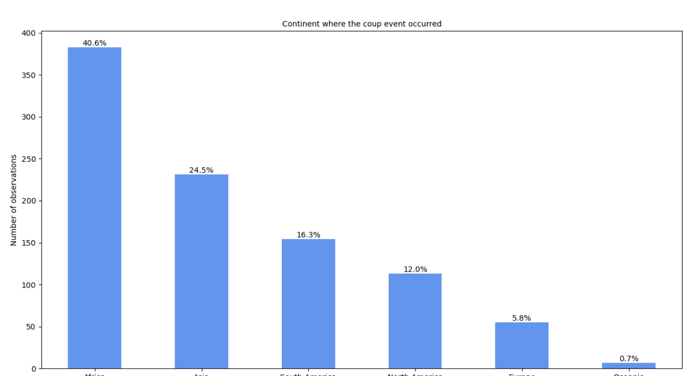
Problem Statement: Our problem statement is: **Which variables are most important to have a successful coup?**

Exploratory Data Analysis

Coups over Space and Time

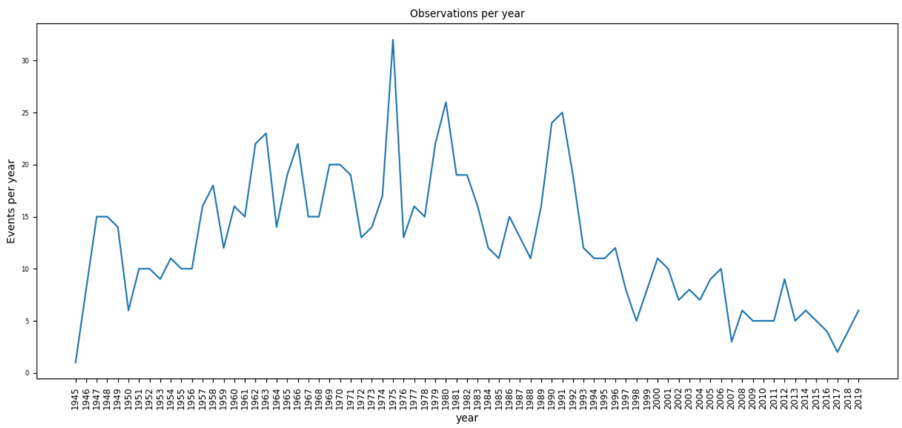
The bar chart below shows the distribution of coup events per continent. The distributions themselves do not seem particularly surprising, save for the 12.0% for North America, which would likely be different with further segmentation by including the Central American subcontinent.

Figure 1 - Continent where the Coup Event Occurred



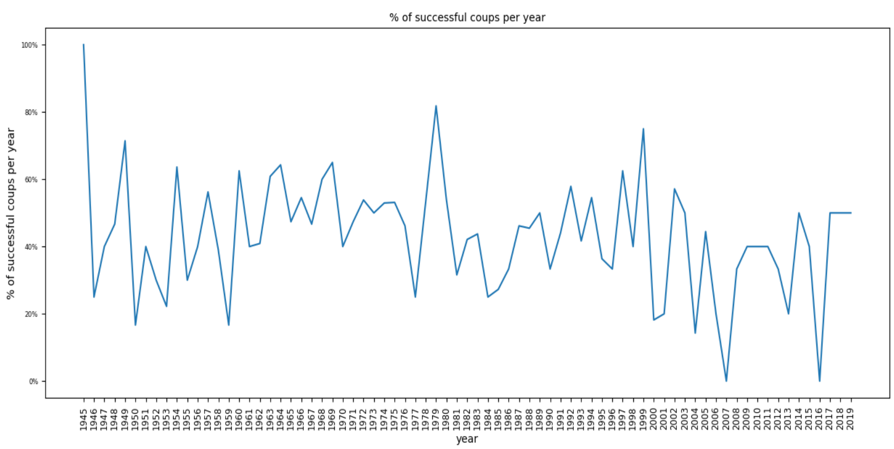
These coup events are showing a tendency of reduction, after having a generally upwards trend from 1945 all the way into the 1970s, with a (very chaotic and with tumultuous moments) trend of descent up until 2019.

Figure 2 - Coup Observations per Year



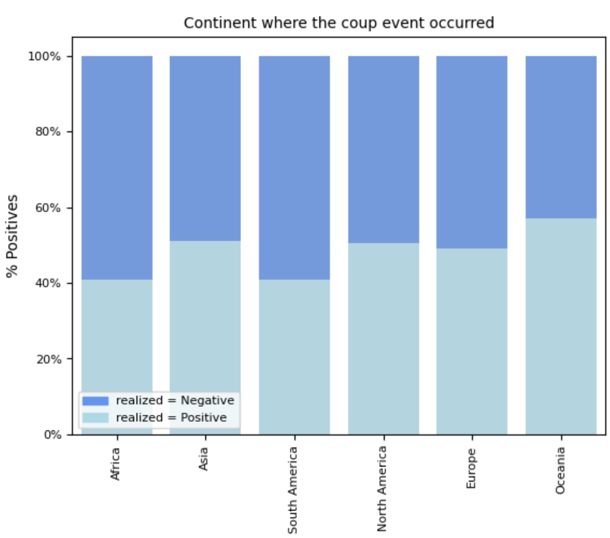
We started observing the chance of success by simply observing the evolution of successful coups over time, and the graph below does not seem to indicate that there’s any correlation between coup success and what time period it happened.

Figure 3 - Percentage of Successful Coups per Year



When observing aggregate data, the chances of a coup being successful are around 45%. This does change from continent to continent, however, with coups taking place in Africa or South America having a lower chance of success than those happening in Asia, which is the continent with the second-highest number of coup events.

Figure 4 - Continent where the Coup Event Occurred



Please see our Jupyter notebook for a far more in depth exploration of our data. We have only presented broad highlights here.

Methods

Dataset Description

Our group of Revolutionaries has chosen to work on the Coups D'état dataset, from the University of Illinois' Cline Center for Advanced Research. It consists of **exclusively categorical data (already neatly encoded as 0 and 1)** for 943 coup events, out of which 181 were conspiracies (thwarted before being attempted), 336 attempted coups (did not manage to remove the incumbent or to end their ability to direct national governance), and 426 realized coups. This data starts in the year 1945 and ends in 2019.

It is important to note the definition of coups d'état by the Coup D'état Project (CDP) is that of *organized efforts to effect sudden and irregular (e.g., illegal or extra-legal) removal of the incumbent executive authority of a national government, or to displace the authority of the highest levels of one or more branches of government*.

This definition may be considered somewhat broad because it includes not only coups initiated by the military and rebel groups, but also some cases of forced resignations, popular revolts, and counter-coups meant to reinstate the previous government. This is something that is worth mentioning because during data exploration the Brazilian members of the group (Max and Ronan) pointed out that the number of coups registered for Brazil is higher than the number of events that are traditionally considered as coups, in their understanding.

The project defines several different types of coups. These include "Military" (initiated by non-formal military actors), "Dissident" (initiated by small groups of discontents which may include types of influential leadership), "Rebel ("initiated by militarized groups within the existing government), "Palace" (initiated by members of a faction within the existing government), "Foreign-backed" (with a foreign power as a driving force), "Auto" (When the existing chief executives take extreme measures to eliminate other branches of government), "Forced resignations" (soft coups with no formal deposition), "Popular Revols" (irregular regime change driven by widespread popular dissatisfaction), "Counter Coup" (the post-coup leadership is removed by the prior regime within one month), and "Other".

Referring to the CDP Project codebook, we can find a description for each data column, but in the interest of space, we can summarize here that they are categorical, relate to the groups that started the coup and external influences (foreign support), what was the fate of the incumbent during the event, when the event took place, and what was the outcome (attempted, realized, conspiracy). Please see Table 1 in the appendix for a description of all of the data features.

Given this information, we used "realized" as a response variable. As independent variables, we used only the group of variables related to the circumstances surrounding the execution of the coup, plus the variable "country". Despite having information about the date of each coup, we did not consider this project as a time series problem. Moreover, we excluded the variables about the fate of the deposed executive because these variables may be considered as information leak about the result of the coup.

Considering that we are not interested in any specific country, we grouped the variable country into continents: Africa, Asia, Europe, North America, Oceania, and South America. After, we construct dummy variables to represent each continent. With this treatment, all the exogenous variables of the models turned into dummy variables.

The data set does not have missing values, but there were variables without representativity. Therefore, another treatment was the exclusion of variables that had less than 2% (19) observation in a category. Hence, the independent variables used in our models were continent_Africa, continent_Asia, continent_Europe, continent_North America, continent_South America, military, dissident, rebel, popular, counter, palace, foreign, and auto.

To test the impact of each variable on the outcome of a coup we tested 12 models:

1) *Logit;*
2) *Logit with Lasso;*
3) *Logit with forward stepwise feature selection;*
4) *Logit with interaction terms;*
5) *Logit with Lasso and interaction terms;*
6) *Logit with forward stepwise feature selection and interaction terms;*
7) *Decision Tree*
8) *Decision Tree with forward stepwise feature selection and interaction terms;*
9) *LightGBM;*
10) *LightGBM with forward stepwise feature selection;*
11) *XGBoost;*
12) *XGBoost with forward stepwise feature selection.*

Before testing any model, we separated our data into a training set and a test set. To do this we shuffled the data and sampled the data into 85% training (801 observations) and 15% test (142 observations).

To achieve our goal of determining the most important factors for a coup, we use variable importance as a metric. We used Shapley values (explained in more detail later), which are model-agnostic, to allow comparability of feature importance for all models.

The logit model, as described by the formula 4.4 in *An Introduction to Statistical Learning with applications in R* [3], calculates the log odds with the formula:

$$log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n \quad (1)$$

Where:

$p(x)$: the probability of a coup being a success;

X: design matrix with the independent variables;

n: denotes the number of variables in the mode.

The betas of this model are found using the Maximum Likelihood of the function:

$$l\left(\beta_0, \beta_1, \cdots, \beta_n\right) = \prod_{i:y_i=1} p\left(x_i\right) + \prod_{i':y_{i'}=0} \left(1 - p\left(x_{i'}\right)\right) \quad (2)$$

Considering that all the variables in our model are dummy variables, we calculated the impact of each feature in the odds ratio associated to the probability of a successful coup by: e^{β_k}, where β_k is the coefficient associated with a particular dummy variable.

Considering that we want to know only the variables that most impact the probability of having a successful coup, we also tried a model with Lasso regularization. This shrinkage method has the advantage of performing a variable selection by setting to zero the coefficients of the least important features. This regularization is done by changing the expression 2 into:

$$l\left(\beta_0, \beta_1, \cdots, \beta_n\right) = \prod_{i:y_i=1} p\left(x_i\right) + \prod_{i':y_{i'}=0} \left(1 - p\left(x_{i'}\right)\right) + \lambda \sum_{j}^{n} \left|\beta_j\right| \quad (3)$$

However, the expression 3 does not have a closed formula. Therefore, to find the best betas and maximize this log-likelihood scikit-learn uses the algorithm coordinate descent [4].

Another methodology used to decrease the number of variables in our model was the forward stepwise feature selection. This method is described in algorithm 6.2 in *An Introduction to Statistical Learning with applications in R* [3]. This method tries $1 + \frac{p(p+1)}{2}$ different models, where p is the number of features.

Before applying the forward stepwise selection algorithm, we separated our train set into train and validation using bootstrap. The validation set was constructed using all observations from our first training set that were not selected in the bootstrap sample of our second training set. We made 100 different training and validation sets using bootstrap and applied the forward stepwise selection algorithm for each train set. After a model was selected in the train set, we calculated the accuracy in the validation set.

Given that for each pair of training and validation sets the forward stepwise selection algorithm resulted in a different accuracy in the validation set, we verified the boxplot of these accuracies for all the best models selected using only, 1, 2, …, n features, where ‘n’ is the total number of variables available. We decided the best number of variables to be used in the model by comparing the median of each boxplot with the model’s performance boxplot when using one less feature.

For the model Logit with forward stepwise feature selection, 9 variables were selected: auto, continent_South America, counter, dissident, foreign, military, palace, popular, and rebel. For the model Logit with forward stepwise feature selection and interaction terms, 11 variables were selected: auto, continent_South America, counter, dissident, foreign, military, military dissident, military palace, palace, popular, and rebel. For the model Decision Tree with forward stepwise feature selection and interaction terms, 8 variables were selected: auto, continent_South America, counter, dissident, military, palace, popular, and rebel. For the model LightGBM with forward stepwise feature selection, 9 variables were selected: auto, continent_South America, counter, dissident, foreign, military, palace, popular, and rebel. For the model XGBoost with forward stepwise feature selection, 9 variables were selected: auto, continent_South America, counter, dissident, foreign, military, palace, popular, and rebel.

The interaction terms were calculated by multiplying each independent variable by another. After this multiplication we disregarded any interaction term that had less than 2% (19) observations. This resulted in a total of 35 variables.

The logit function is great when considering the interpretability of a model. However, this function does not capture the relationship among variables. Therefore, we used interaction terms to diminish this problem. Nonetheless, even with interaction terms, there may be some relationships with 3 or more variables that were not captured by the model. Hence, we also tried Decision Tree models.

With the right hyperparameters, the Decision Tree models have the advantage of trying all the possible combinations of the binary variables in our data set. However, this model can easily overfit if not configured correctly.

To calculate how much gain a tree split represents in the model it was used the Gini Index. This index is represented in function 8.6 in *An Introduction to Statistical Learning with applications in R* [3]:

$$G = \sum_{k=1}^{K} \hat{p}_{mk}\left(1 - \hat{p}_{mk}\right) \tag{4}$$

Where:

$\hat{p}_{mk}$: the proportion of training observations in the mth region that are from the kth class.

The Gini Index is a measure of purity. A value of zero means that a node contains only observations of the same class.

To avoid overfit, we tested several hyperparameters using cross-validation. The hyperparameters tested were:

- min_samples_split: from 5 to 100;
- min_samples_leaf: from 1 to 100; and
- max_depth: from 1 to 10.

The hyperparameter **min_samples_split** avoids that a split is done based on only a few observations. The hyperparameter **min_samples_leaf** avoids that a leaf is based on only a few observations. The hyperparameter **max_depth** avoids that the tree grows too much resulting in leaves that explain only the train set but does not generalize well to new data.

For each hyperparameter we plot the train accuracy, the validation mean accuracy based on a 5-fold cross validation, and a region expressing minus two and plus 2 standard deviations from the validation mean accuracy. After, we choose the best hyperparameter based on the highest validation mean accuracy in which the train accuracy was inside the confidence interval of 2 standard deviations from the validation mean accuracy.

For the Decision Tree model, the final hyperparameters were min_samples_split = 52, min_samples_leaf = 16, and max_depth = 5.

For the Decision Tree with forward stepwise feature selection and interaction terms, the final hyperparameters were min_samples_split = 5, min_samples_leaf = 5, and max_depth = 7.

The Decision Tree models have an advantage that we can plot the tree and easily calculate the prediction of an event. On the other hand, just plotting the tree is not enough to have an idea of which variables are the most important. Therefore, we calculated the feature importance of the models using SHAP values.

The SHAP (Shapley Additive exPlanation) values for feature importance, explained by Reza Bagheri in your article *Introduction to SHAP Values and their Application in Machine Learning*, is based on the Shapley values that is a concept in game theory introduced by Lloyd Shapley in 1951. This is a model-agnostic method that calculates the importance of the variables by simulating the results of a model with and without values for a determined variable.

Finally, to try to achieve a better model performance, we tried two boosting algorithms: LightGBM [6], and XGBoost [7]. Both models are based on decision trees, and on the Gradient Boosting algorithm. However, the LightGBM is based on a leaf-wise tree growth, and the XGBoost is based on a level-wise tree growth.

Both models are prone to overfit. Hence to a better performance, using the same methodology as used in the Decision Tree models, we tested several hyperparameters.

For the LightGBM models, the tested hyperparameters were:

- num_iterations: from 10 to 500;
- min_data_in_leaf: from 5 to 100;
- feature_fraction: from 0.05 to 1;
- max_depth: from 1 to 10; and
- learning_rate: from 0.1 to 0.7.

The hyperparameter **num_iterations** is the number of models used in the algorithm. Differently from bagging and Random Forest, boosting models tend to overfit if used a large number for num_iterations. The hyperparameter **min_data_in_leaf** has the same meaning as the hyperparameter min_samples_leaf in the Decision Tree model. The hyperparameter **feature_fraction** is used to limit the number of features used in each tree model. This makes the several models calculated in this algorithm to be less correlated. The hyperparameter **learning_rate** is a shrinkage rate used to shrink the features weights to avoid overfit.

For the LightGBM model, the final hyperparameters were num_iterations = 50, min_data_in_leaf = 10, feature_fraction = 0.45, max_depth = 4, and learning_rate = 0.111.

For the LightGBM with forward stepwise feature selection model, the final hyperparameters were num_iterations = 440, min_data_in_leaf = 5, feature_fraction = 0.35, max_depth = 3, and learning_rate = 0.031.

For the XGBoost models, the tested hyperparameters were:

- max_depth: from 1 to 15;
- subsample: from 0.4 to 1;
- colsample_bytree: from 0.4 to 1; and
- learning_rate: from 0.01 to 0.5.

The hyperparameter **subsample** determines the proportion of the total observations to be used in each model. The hyperparameter **colsample_bytree** has the same meaning as the hyperparameter feature_fraction in the LightGBM model.

For the XGBoost model, the final hyperparameters were max_depth = 5, subsample = 0.4, colsample_bytree = 0.95, and learning_rate = 0.081.

For the XGBoost with forward stepwise feature selection model, the final hyperparameters were max_depth = 4, subsample = 0.9, colsample_bytree = 0.6, and learning_rate = 0.091.

To compare which model had the best performance we used the accuracy. However, we also calculated the precision, sensitivity, specificity, F1-score, and the ROC AUC.

Accuracy represents the percentage of how many observations the model predicted correctly. Precision calculates the percentage of how many observations predicted as positive by the model are correct. Sensitivity represents the percentage of how many observations that are from the class positive were correctly predicted by the model. Specificity represents the percentage of how many observations that are from the class negative were correctly predicted by the model. F1-score is the harmonic mean between precision and sensitivity. ROC-AUC is the area under the curve of the Receiver operating characteristic curve.

Please see the appendix for visualizations of variable number and hyperparameter selection

Results

Performance Comparison

Based on the four types of models chosen (Logistic Regression, Decision Tree, LightGBM, and XGBoost), we compared the performances of the version of it with the best accuracy on the test set. Out of these four groups, the best-performing models were those that underwent forward stepwise variable selection, save for the XGBoost model.

The best-performing model, Logistic Regression, achieved this result through superior performance in the precision metric compared to the others.

Table 2 - Model Performance Comparison Table for the Test Set

Model	Accuracy	Precision	Sensitivity	Specificity	F1 Score	ROC AUC
Logit Stepwise	0.7606	0.8125	0.6094	0.8846	0.6964	0.8266
LightGBM stepwise	0.7535	0.7959	0.6094	0.8718	0.6903	0.8300
Decision Tree stepwise	0.7465	0.8043	0.5781	0.8846	0.6727	0.8295
XGBoost	0.7465	0.7692	0.6250	0.8462	0.6897	0.8299

Relevant Factors

The models all converged on the same set of important key features for predicting the success of a coup.

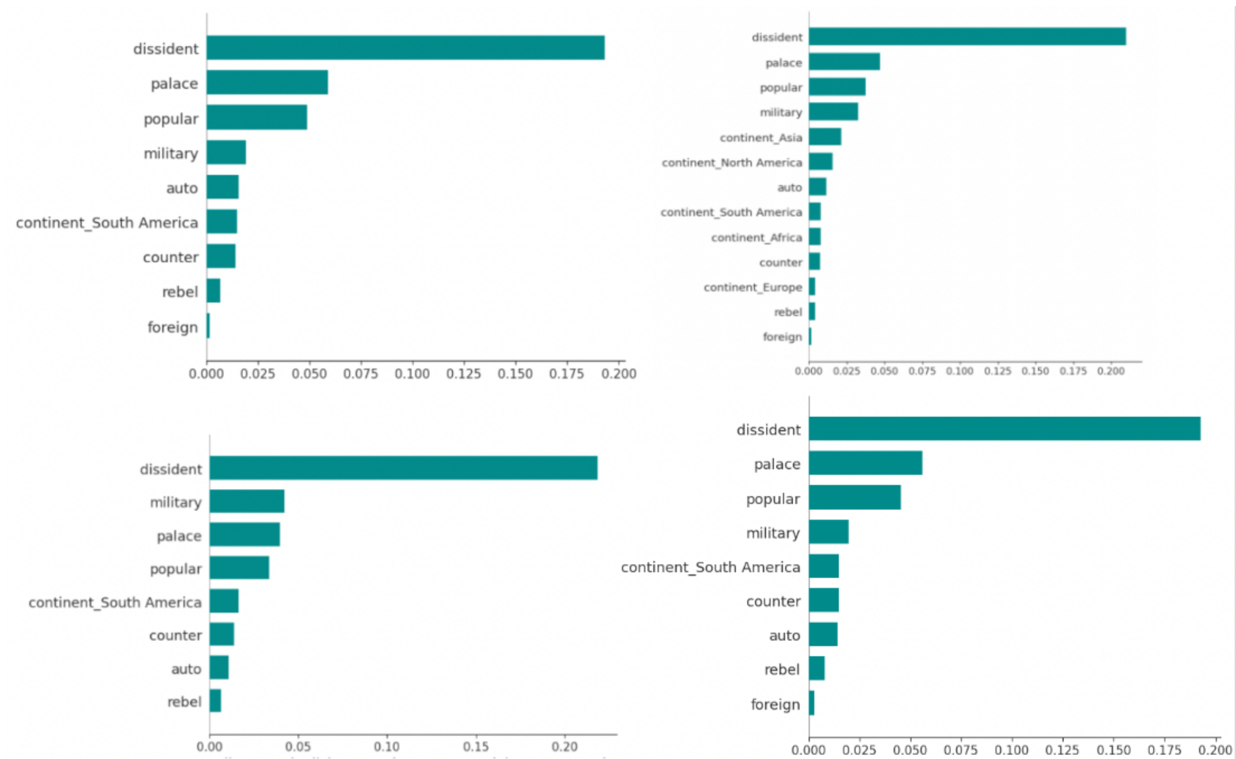


Figure 4 - SHAP Results from Four Different Models

Four models all converged on the same top features. From top left, going clockwise: Logit Regression with forward stepwise feature selection, XGBoost, LightGBM with forward stepwise feature selection, and Decision Tree with forward stepwise feature selection

Logistic Regression with forward stepwise feature selection

The Logistic Regression model used the following variables after stepwise selection: **auto, continent_South America, counter, dissident, foreign, military, palace, popular, rebel.** The SHAP metric results place dissident, palace, popular and military as the four most impactful predictors, with dissident at the top. This exact selection of features, in that level of importance, is also observed on LightGBM and XGBoost.

This level of importance assigned to auto coups and to dissident coups is different for the other three models, which, using the SHAP metric, place dissident coups as the most important predictor.

XGBoost

Unlike the other models, XGBoost was the only one that had superior test accuracy scores without using forward stepwise feature selection, making use of the entire set of 13 variables provided to the model, without using interaction terms. Again, the variables for dissident, palace, popular, and military figured as the top 4 in terms of SHAP-measured featured importance, with foreign being at the bottom.

LightGBM with forward stepwise feature selection

The stepwise LightGBM model selected the following 9 variables: **auto, counter, dissident, palace, military, popular, rebel, continent_South America, and foreign.**
This model shares the same variables used as Logistic Regression, with foreign also featuring as a relevant predictor. The order for the top four predictors is the same as the ones for Logistic Regression and XGBoost.

Decision Tree with forward stepwise feature selection
The stepwise Decision Tree Model selected the following 8 variables: **auto, continent_South America, counter, dissident, military, palace, popular, and rebel.** Compared to the Logistic Regression, the only variable that is not present is foreign, relative to the presence of foreign assistance or intervention in the coup. This model differs from the others because it assigns greater importance to the military predictor, placing it as the second most important. However, like the other models, it still ranks the palace predictor as of higher importance than the popular predictor.

Feature Ranking Comparison Table
The top four variables are the same for every model, ordered differently on the Decision Tree

Table 3 - Feature Ranking Comparison Table (Top 8)

Variable Rank	Logit Stepwise	Decision Tree Stepwise	LightGBM Stepwise	XGBoost
1	dissident	dissident	dissident	dissident
2	palace	military	palace	palace
3	popular	palace	popular	popular
4	military	popular	military	military
5	auto	continent_South America	continent_South America	continent_Asia
6	continent_South America	counter	counter	continent_North America
7	counter	auto	auto	auto
8	rebel	rebel	rebel	continent_ South America

The table comparison above synthesizes the findings for variable importance. For all models, dissident, palace, military, and popular factors were always among the top four, with dissidence in the first place, and variations only for the Decision Tree model, which placed military influence as the second most important. On the other end of the ranking, the predictors for coups initiated by rebels and for coups with foreign assistance were consistently selected as the two least important for all models.

Conclusions

Different modeling techniques all concentrated on the same set of four variables as the most significant in predicting the success of a coup. This “agreement” from different models is a strong indicator that these factors are in fact of increased relevance for predicting coups. These four variables all relate to the initiators of the coup, namely, dissenters outside of the government (dissent), opposing factions within the government (palace), military groups outside of the government structure (military), or the general population (popular). This is an interesting result, as the variables for geographical location were not the most significant, suggesting that there are certain elements predictive of a coup that are important regardless of historical conditions and cultural factors.

One potential limitation of the models created is the restriction of the dataset to qualitative variables. It is possible that quantitative variables relative to socioeconomic conditions at the time of the coup, such as average GDP growth in the years close to the event, may have some explanatory value to this model. This could improve the explanatory power of our models by adding contextual factors that could explain why a coup happened in the first place, and why it was successful or not, which could be related to general support for the coup.

There may be an explanation as to why the predictors are ranked in importance precisely as they are. Dissenters, according to the dataset codebook, include ex-military leaders, and religious or political leaders. Owing to their sociopolitical relevance, dissenters may have extra resources or support to ensure that their coups succeed. Palace coups may have that support precisely because they can use the government structure to maintain power, provided that there is no strong internal opposition to the coup. The military coup initiators are most capable of simply taking power through violent means, but they may not have the social backing to maintain power. Popular revolts may be the least impactful in success because of a lack of organization and resources to overthrow the government.

Future studies could focus on either including more contextual data to improve model performance and explainability, or also be more specific, trying to understand why there are significant differences between the different groups that start the coups, and how those differences impact the chances of a coup's success.

References

[1] Peyton, Buddy, Joseph Bajjalieh, Dan Shalmon, Michael Martin, and Jonathan Bonaguro. 2020. Cline Center Coup D'état Project Dataset. Cline Center for Advanced Social Research. V.2.0.0. November 16. University of Illinois Urbana-Champaign. doi: 10.13012/B2IDB-9651987_V3
[2] Peyton, Buddy, Joseph Bajjalieh, Dan Shalmon, Michael Martin, and Jonathan Bonaguro. 2021. "Cline Center Coup D'état Project Dataset Codebook". Cline Center Coup D'état Project Dataset. Cline Center for Advanced Social Research. V.2.0.2. February 23. University of Illinois Urbana-Champaign. doi: 10.13012/B2IDB-9651987_V3
[3] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021) *An Introduction to Statistical Learning with applications in R*, https://www.statlearning.com, Springer-Verlag, New York
[4] https://scikit-learn.org/stable/modules/linear_model.html#lasso (Accessed: 26 November 2022)
[5] https://towardsdatascience.com/introduction-to-shap-values-and-their-application-in-machine-learning-8003718e6827 (Accessed: 04 December 2022)
[6] https://lightgbm.readthedocs.io/en/v3.3.3/ (Accessed: 26 November 2022)
[7] https://xgboost.readthedocs.io/en/stable/ (Accessed: 26 November 2022)

[8] Albrecht, H., & Eibl, F. (2018). How to Keep Officers in the Barracks: Causes, Agents, and Types of Military Coups. *International Studies Quarterly*, *62*, 315–328. https://doi.org/10.1093/isq/sqx085
[9] Hiroi, T., & Omori, S. (2013). Causes and Triggers of *Coups d'etat*: An Event History Analysis. *Politics & Policy*, *41*(1), 39–64. https://doi.org/10.1111/polp.12001

Appendix

Table 1 - Description of Variables

Name of Variable	Description of Variable
coup_id	Unique number assigned to each event. It consists of the country's cowcode and the eight digit date of the event in MMDDYYYY.
cowcode	A unique country code number based on the Correlates of War (COW) country code list4. It is used to identify the country where a coup event occurred. Please note, these codes are slightly different from the canonical COW codes. For details, please see "Note on COW Country Codes," below.
country	Name of the country where the coup event occurred.
year	Year of the coup event.
month	Month of the coup event.
day	Day of the coup event.
event_type	Indicates whether the event is a coup, attempted coup, or conspiracy.
unrealized	A dummy variable where one indicates an unsuccessful coup or plot and zero otherwise.
realized	A dummy variable where one indicates a successful coup and zero otherwise.
conspiracy	A dummy variable where one indicates a coup conspiracy thwarted prior to execution and zero otherwise
attempt	A dummy variable where one indicates a coup was attempted by failed and zero otherwise.
military	A dummy variable where one indicates a military coup/attempt/conspiracy and zero otherwise.
dissident	A dummy variable where one indicates a dissident coup/attempt/conspiracy and zero otherwise.
rebel	A dummy variable where one indicates a rebel coup/attempt/conspiracy and zero otherwise.
palace	A dummy variable where one indicates a palace coup/attempt/conspiracy and zero otherwise.
foreign	a dummy variable where one indicates a foreign-backed coup/attempt/conspiracy and zero otherwise
auto	A dummy variable where one indicates an auto coup and zero otherwise.
resign	A dummy variable where one indicates a forced resignation and zero otherwise
popular	A dummy variable where one indicates a popular revolt and zero otherwise.
counter	A dummy variable where one indicates a counter-coup and zero otherwise.

other	A dummy variable where one indicates the coup event does not fit into any of the above categories or the actors were not identified and zero otherwise.
noharm	A dummy variable where one indicates the deposed executive was not harmed during the coup event and zero otherwise.
injured	A dummy variable where one indicates the deposed executive was injured during the coup event and zero otherwise.
killed	A dummy variable where one indicates the deposed executive was killed during the coup event and zero otherwise.
harrest	A dummy variable where one indicates the deposed executive was placed under house arrest and zero otherwise.
jailed	A dummy variable where one indicates the deposed executive was jailed and zero otherwise.
tried	A dummy variable where one indicates the deposed executive was tried and zero otherwise.
fled	A dummy variable where one indicates the deposed executive fled the country and zero otherwise.
exile	A dummy variable where one indicates the deposed executive was banished from the country and zero otherwise.

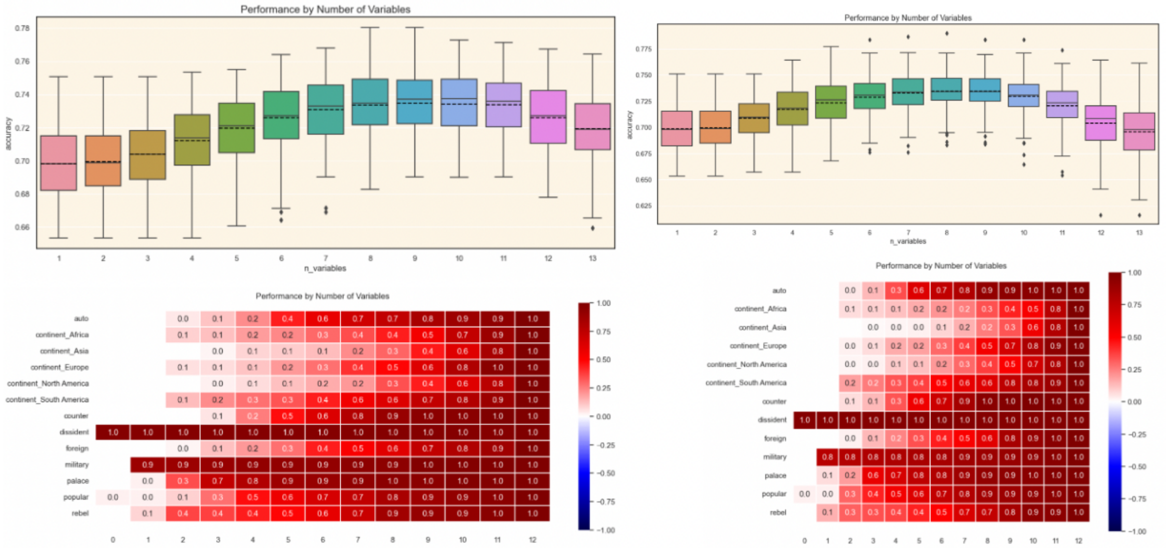


Figure 1 - Analysis of accuracy by number of variables of Logit regression with forward stepwise feature selection model (left) and Decision tree with forward stepwise feature selection model (right)

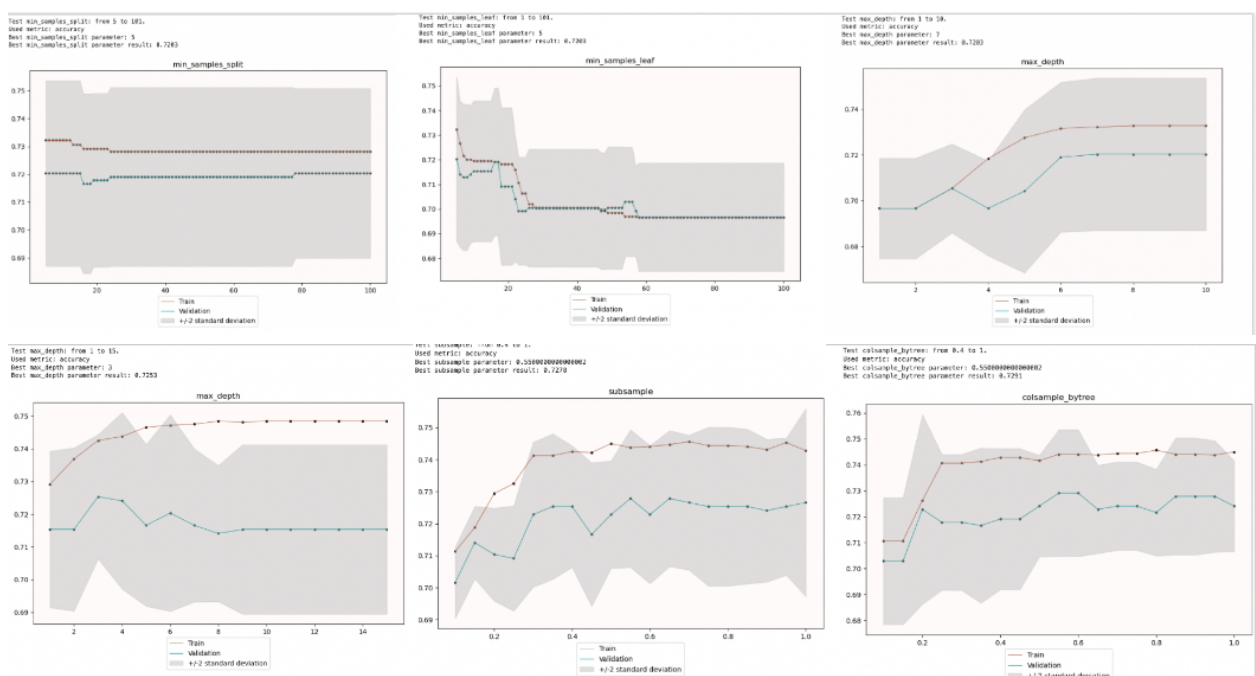


Figure 2 - Hyperparameter testing of Decision tree with forward stepwise feature selection model (top) and XGBoost model (bottom)

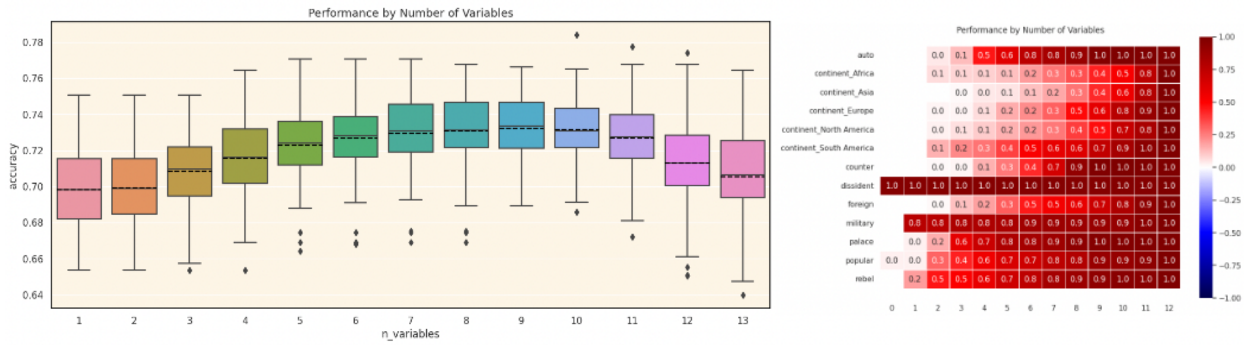


Figure 3 - Analysis of accuracy by number of variables of LightGBM model

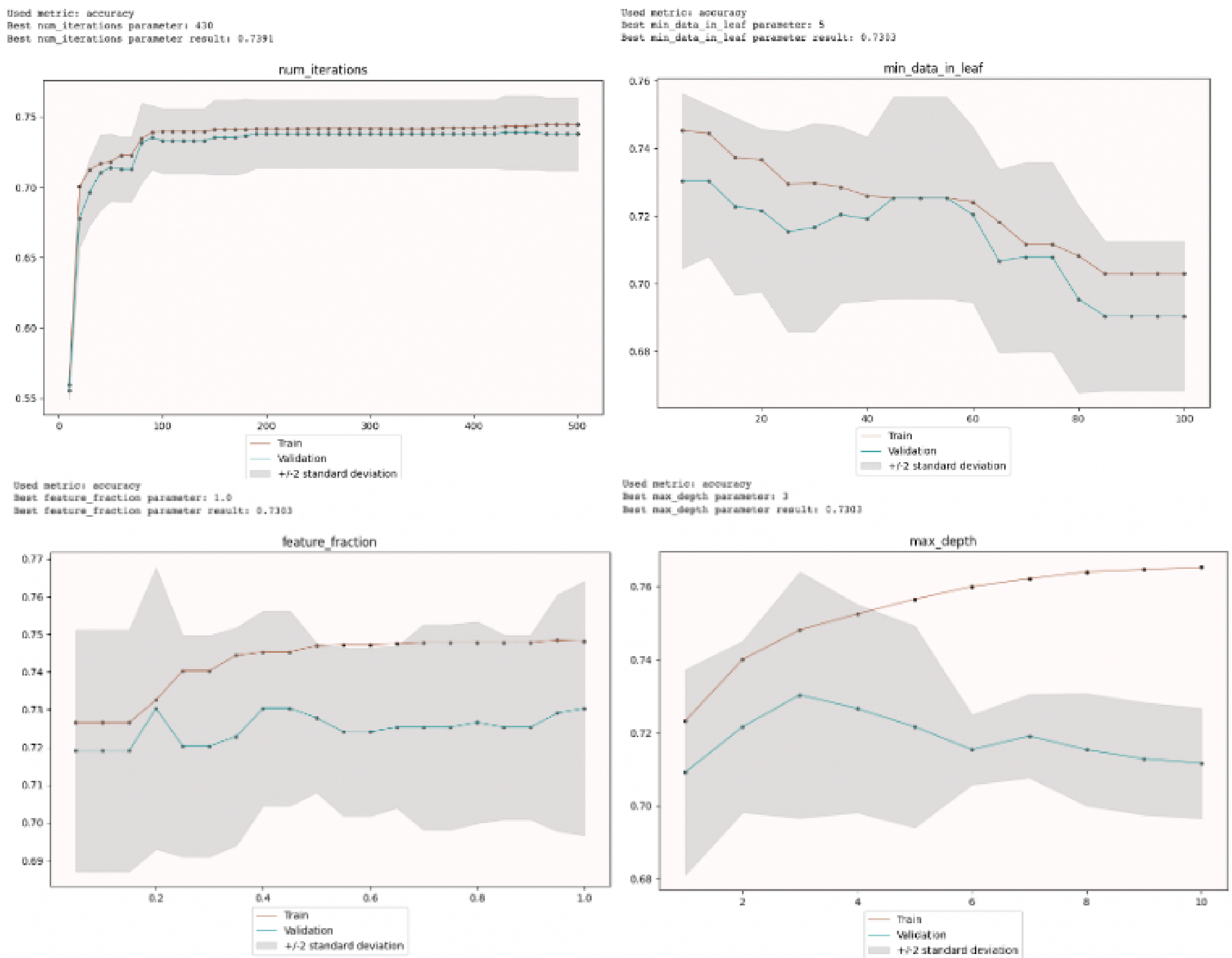


Figure 4 - Hyperparameter testing of LightGBM model