Harvard John A. Paulson School of Engineering and Applied Sciences

Introduction to Data Science

Coup D'États

Authors:
Reem Al-Khalqi
Ana Vitoria
Adriana Trejo-Sheu

Final Project Report for CS209a

December 10, 2021

Abstract

History shows that coup d'états are usually the most important factor that lead to the sudden transfer of political power - an event that has strong consequences on the country's well-being. This study explores the factors that impact the success of coups using the data provided by the Cline Center for Advanced Social Research.

We explain our initial exploratory analysis to understand the data as well as our reasoning behind the selected predictors. We create a baseline model that we evaluate and improve through multiple means, namely, LogisticRegressionCV, RandomForest and Multi Perceptron classifier (MLP). The main research question of this study is "What constitutes a successful coup?" given the information from the data set. Additionally, we investigate how a coup's success might change based on a geographic sub-region or a time frame.

Contents

1 Introduction

Coup d'états are the sudden overthrowing or transfer of political power of a country's leader and it is an event that has consequences for a country's well-being. We use the data set provided by the Cline Center for Advanced Social Research from the University of Illinois Urbana-Champaign to analyze recorded characteristics of the coup and the fate of the leader. We are mainly interested in understanding what constitutes a successful coup. We also investigate if the geographic region of the coup affects its outcome and whether the requirements for a successful coup changes by region.

2 EDA Analysis

We perform a series of EDA analysis to explore the content of the dataset and this helps minimize our predictor list, do model selection, and determine how to extend past our baseline model. We did not need to impute any missing data and we are able to take the dataset as is. Our predictors are mostly categorical data that are nominal. We do have a couple of ordinal predictors such as year, month, and day.

2.1 Choosing the Response Variable

The main research question that we try to answer is what constitutes a successful coup given the predictors provided in the data. Our response variable should signify whether or not a coup was realized or unrealized. From the data given, 'event_type', realized and unrealized all describe the success or failure of a coup. The 'event_type' variable gives three outcomes ('coup', 'attempt', 'conspiracy'), which we can use to determine the success. The realized column gives an array of ones and zeros with one indicating a successful coup and zero indicating a failure ('attempt', 'conspiracy'). The unrealized column is the exact opposite of the realized column and gives a one to failed coups ('attempt', 'conspiracy').

From assessing these three different columns, we use the realized column as our response predictor. The 'event_type' can be used to supplement that column for understanding deeper what might have caused failures as well as for graphing purposes.

2.2 Cowcodes vs Country Names

The data provides both the 'cowcode' and the 'country' name. The 'cowcode' is a unique country code number based on the Correlates of War (COW) country code list. The data lists 136 countries and 134 'cowcode'. When inspected, we found that there were a couple of countries where each was listed twice as two different countries but share the same 'cowcode'. These countries were Cote d'Ivoire and Ivory coast ('cowcode' 437) and Kyrgyzstan and Kyrgyz Republic ('cowcode' 703). Essentially, these are the same country, simply imputed with different names. Therefore, we actually have 134 countries and not 136. Hence, we will keep 'cowcode' as a variable to indicate the country.

2.3 Realized vs Unrealized Coups

According to the Cline Center website, a coup is labeled as "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 the government.[1] A coup is considered to be 'realized' or 'unrealized'. A 'realized' coup would be a success in the removal of the incumbents or the removal of the incumbents ability to control the state. An 'unrealized' coup is considered to be a 'conspiracy' or an 'attempt' coup, meaning an unsuccessful removal of an incumbent's power.

Note: We will be using realized coup or successful coup interchangeably to mean a success. A coup technically includes conspiracies and attempts but these will be described as either unrealized, unsuccessful or as they are mentioned. Lastly, the 'event_type' column from the dataset notes a successful or realized coup as 'coup'.

2.4 Preliminary Visuals

In our initial exploration, we looked at how each predictor impacted the outcome of our response variable, realized. In this case we plotted each predictor that was identified against 'event_type' rather than realized. This decision was in order to get a deeper understanding as to how each predictor affected each of the potential coup outcomes. It is an important distinction that we will use later on in the project when assessing the impact or influences of predictors. An example of a sample of our plots are shown below:

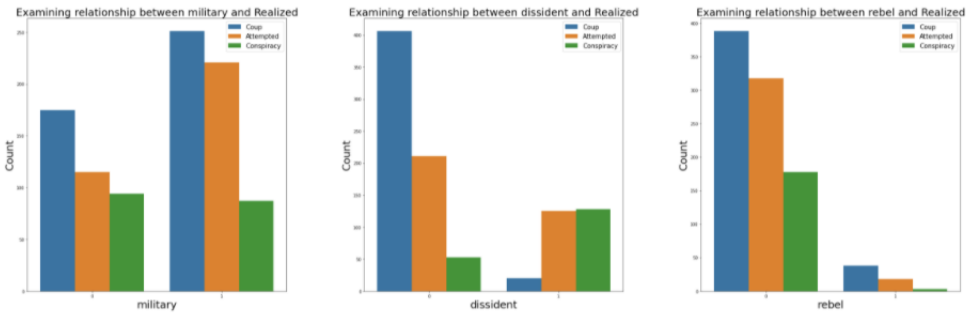

Figure 1: Sample of Subplots showing each Predictor against Event Type

Moreover, we looked at the collinearity of all predictors in the dataset to better identify highly correlated variables.

We can see from the collinearity heatmap that realized and unrealized have a collinearity of -1 which supports our decision in using one as our response variable and removing the other from the predictors list. Moreover, we can see that, unsurprisingly, 'noharm' and 'killed' have a collinearity of -0.84. The reason that they don't have a -1 collinearity is

[1]Cline Center Website

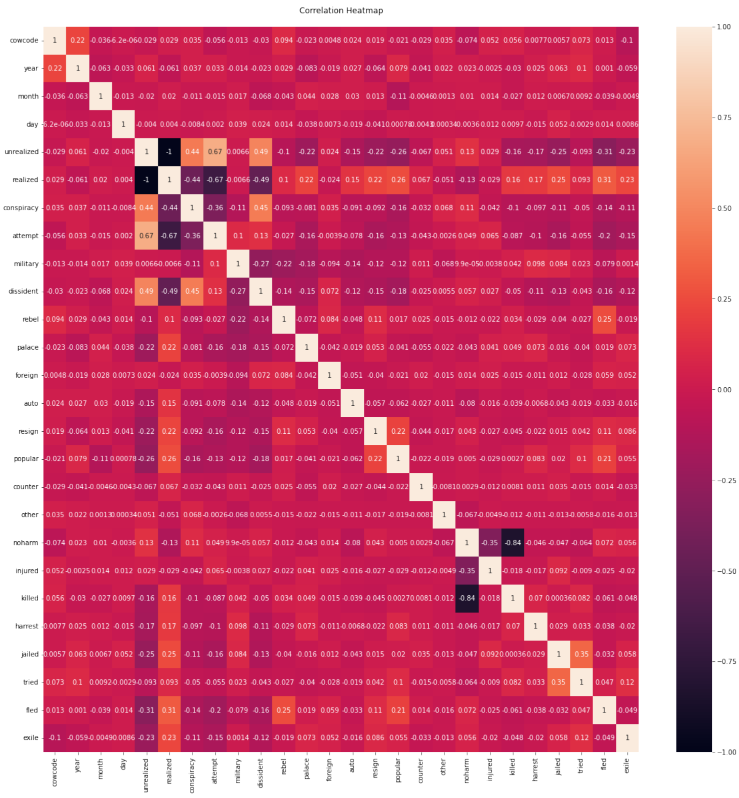

Figure 2: Collinearity heatmap of all predictors in Coup d'états dataset

because there are other fates considered for the leader including 'harrest', 'jailed', 'exiled' and 'fled'. There are several other predictors that have a noticeably high collinearity and being able to see the collineartiy between all these predictors will allow us to streamline the feature selection process.

3 Baseline Model

From our initial EDA and understanding of the variables, we have identified a list of problematic predictors, meaning that they could be revealing the outcome of the response. We identified these from their definitions within the codebook because they described what happened to the deposed executive, meaning that there was a realized coup. The list of predictors that are associated with a deposed target are 'noharm', 'injured', 'killed', 'harrest', 'jailed', 'tried', 'fled', 'exile', 'resign'. We decided to take an in-depth look at these predictors because we believe that they should always correlate with a realized coup or a response variable of 1.

When taking a closer look at these variables, we found that noharm is in 94% of all data samples, and almost equally distributed across the unrealized and realized coups at 97% and 90% respectively. The rest of the predictors, we grouped together to analyze how the combination of these reveal the response. We found that the combination of these are present in 58% of successful coups whereas they are only present in 6% of unsuccessful coups. We feel that by revealing 58% of the successes is enough justification for their removal.

In the baseline model, we are using the selected variables from our initial feature selection process to predict the likelihood of a coup having success or not. We are not using any interaction terms or using the univariate feature selection process in this baseline.

We use LogisticRegressionCV with L1 penalization with the variables in the table below. The resulting **test score** is **73.852%** and the resulting training score is 75.152%.

coefficient	predictor	coefficient	predictor
0.8	intercept		
0	cowcode	0.9669	palace
-0.407	year	0	foreign
0	month	0.503	auto
0	day	1.326	popular
-0.2	military	0	counter
-2.5	dissident	0	other
0.035	rebel	-0.36	noharm

Table 1: Coefficients of Baseline Model.

We can interpret these a follows. **Year** - given that the coefficient is negative, we can interpret this to say that the more recent the attempt, the less likely is this going to have success. Meaning that attempt and conspiracies more years ago would be more likely to succeed compared to more recently in history. **Cowcode, month, day, foreign, counter, other** - these all have coefficient of 0. Meaning that the model does not consider these to be significant. **Remaining variables** - for the remaining variables on the table above (military, dissident, rebel, auto, popular, noharm): where the coefficient is negative, it means that the presence of that variable (variable is 1) decreases the odds of a successful coup. Where the coefficient is positive, it means that that variable (when it is 1) increases the odds of a successful coup.

From this baseline model, we will consider interaction terms and do further feature selection in order to improve our accuracy scores.

4 Feature Selection on Categorical Variables

Feature selection is important in modeling because it can greatly impact the accuracy and predictions a model makes. By understanding our predictors and their relationship to the response, we can make an informed decision on what predictors we plan to keep and drop. We will not be considering any of the variables that we had dropped in the baseline model, due to the explanation given prior.

We were also considering every two-way unique interaction terms from what predictors are left from the initial selection. It is important to see how these interaction terms could potentially give us more powerful insight into predicting the response. In total, we have 318 predictors that we are considering.

4.1 Interaction terms

To better understand the effect of the predictors on the outcome, we need to know if there are certain predictors that, when combined, affect the response variable. We create a list of all possible combinations of pairs for the predictors and check if the effect of one predictor variable depends on the value of another. If so, then we have a two-way interaction term. Additionally, we check for three-way interactions. That is, if any two-way interaction of β_a and β_b depends on the value of β_c, then we have a three-way interaction term.

To create the two-way interaction terms, we multiply the values/vectors of β_a and β_b for every combination of the selected predictors. Two create the three-way interaction terms, we multiply the values/vectors of β_a and β_b and β_c for every combination of the selected predictors. *The resulting vectors of the two-way interaction terms are not used to create the three-way interaction terms.* This gives us m new columns with the dimension of n where m is the number of possible combinations and n is the length of observations. Interaction term columns that have no values other than 0 were disregarded.

4.2 Univariate Feature Selection

We will be doing feature selection using a series of different statistical approaches to influence our decision our final predictor list that we will use to train the model. We are using univariate analysis in order to explore each variable in the data separately. We have chosen both Chi square and Mutual Information as methods to tests the inter-dependency between a predictor and the response. We have chosen these two methods because of our prior background experience using these statistics as well as because they were referenced in sklearn feature selection documentation. We will give the necessary background for each of the methods use and ultimately discuss the impact of their outcomes on our decision-making.

4.2.1 Chi Squared Background

Chi squared will test the relationship between features for distribution. Chi squares considers the degrees of Freedom and a selected alpha value to determine whether or not we reject or accept our hypothesis for independence. Independence of a predictor means it does not depend on the response, which we want to remove in this feature selection. Our null hypothesis is that a predictor and response are independent. The equation can be seen as:

$$\chi = \Sigma Z_i^2 \tag{1}$$

...where Zi is the standard normal variable.

$$Z_i = \sum_{j=0}^{N} \frac{(O_j - E_j)^2}{E_j} \tag{2}$$

...with O_i as observed values, E_i as expected values, N is the amount of observations.

The higher the Chi-square value the more dependent the value is on the response, but we must compare the chi value to the chosen associated p value in order to deny or accept our hypothesis. We can also plot the outputted p values to see what the highest p values are. With high p values, the predictor is more likely to be independent of the response and cannot be considered for model training. The returned p value is correlated to each test and tells us the probability of getting a chi-square statistic as large or larger than the ran experiment. If the p value is less than 0.05, we reject our null hypothesis and say that the response and predictor are not independent. This means predictors with p values higher than 0.05 should be dropped.

4.2.2 Mutual Information Feature Selection Background

This works to determine the dependency two random, categorical, non-negative variables. The outputted statistic will return a zero for independent pairs. The higher the statistic means higher dependency between the variables. Mutual information measures how much information that two variables share and how knowing one variable will reduce the uncertainty about the other. This can show us independence between variables and allow us to make decisions of which variables to remove. If a predictor is independent of the response, it means that the predictor does not give any information about the response, showing independence. The equation for Mutual Information can be seen as:

$$log(\frac{p_{X_i,y}(X_i, y)}{p_{X_i}(X_i)p_y(y)} = log1 = 0 \tag{3}$$

The numerator is the joint distribution for the predictor and response variable whereas the denominator is the individual distributions. When the fraction is equal to 1, we get a log value of zero, which means independence.

4.3 Feature Selection Implementation & Findings

We build the list of predictors with the selected 14 variables. Next, we create all possible two-way and three-way interaction terms as described in section 4.1. This results in a new data frame with a total of 318 possible variables including the selected 14 variables.

Our selection process is as follows. First, we apply LogisticRegressionCV to select which ones the model considers significant or not equal to zero. Next, we apply Chi Squared and Mutual Information Classification on these selected 34 variables from LogisticRegressionCV, inorder to remove independent predictors, resulting in 28 found in table 2 below. The improved modelling will use the 28 predictors in LogisticRegressionCV, RandomForest and Multi Perceptron classifier (MLP) to improve the accuracy score found in the baseline.

5 Improved Modelling

In improved modelling we are going to use the 28 predictors kept from feature selection. Here we will use three models: Logistic Regression, Random Forest and MLP.

5.1 LogisticRegressionCV

After having selected the aforementioned 28 variables, we fit and predict with these with a well tuned Logistic Regression model with L1 penalization (we use LogisticRegressionCV). The resulting test score is 74.91% and the training score is 79.39%. This is an improvement compared to the testing performance of the Baseline Model with no interaction terms.

5.1.1 Coefficients and Interpretations

The coefficients from the LogisticRegressionCV model are listed in Table 2. It is worth noting that only the predictors that are considered important are listen in this table. Important predictors are ones with coefficients greater than or less than zero.

coefficient	predictor	coefficient	predictor
0.524	intercept		
5.2066	month military popular	-0.4260	month noharm
4.0693	year rebel	-0.5264	year month noharm
2.6827	palace	-0.6165	military noharm
2.4636	month auto	-0.8102	day dissident
2.3639	year popular	-0.8637	year noharm
2.2309	military dissident	-1.1630	dissident
2.2252	cowcode noharm	-1.3150	month dissident noharm
2.0253	military counter	-1.6848	day rebel
1.0509	month military	-1.7805	month rebel
0.7613	day military noharm	-2.0460	palace foreign noharm
0.5225	palace noharm	-2.2106	dissident noharm
0.0727	auto noharm	-3.0211	cowcode military noharm
0.0067	popular noharm	-3.7303	day dissident foreign
-0.0407	military popular	-4.3495	year palace

Table 2: Coefficients of Logistic Regression with Cross Validation and Interaction Terms

In depth interpretation of all 28 predictors that are considered significant will take more pages than what this paper aims for. Therefore, we will interpret only the interaction term with the lowest coefficient and the interaction term with the highest coefficient.

Military Counter - First, it is worth noting how the feature selection process impacted the predictors coefficients. In the baseline model,'counter' had no influence on the outcome and 'military' had a low predictive power of -0.2. However, when both events occur, that is, if the coup is a counter coup of the realized coup and is initiated by the military actors who are not a formal part of the governing apparatus of the current regime then the log-odds increase by 2.0253.

Year and Palace - The predictor 'year' by itself has no influence on the outcome. However, when combined with palace, it has the highest negative contribution to the chances of a successful coup and as the value for 'year' increases, the chances for a successful coup decreases. This means that coup attempts by dissidents in the past were more likely to be successful compared to ones performed more recently in history.

Month, Military and Popular - Neither military nor popular nor month are considered significant independently. However, when combined, they become the highest positive contributors to the chances of a successful coup. To better understand this three-way interaction term, we look at each two-way interaction term combination. Since 'month popular' is not on the table we can consider it being insignificant. On the other hand, 'military popular' has a negative coefficient of -0.0407 indicating that if the coup was initiated by military actors who are not a formal part of the governing apparatus and the coup attempt was considered popular then it lowers the chances of a successful coup. Finally, 'month military' has a negative coefficient of 1.0509 indicating that if the coup was initiated by military actors who are not a formal part of the governing apparatus and it happened later in the year, then the chances of a successful coup are increased. We can then say that with 'month military popular' having a high positive coefficient, then we have evidence that military has a strong effect on the two terms popular and month. This three-way interaction term can then be interpreted as if a coup was a joint effort from a military group and a popular revolt, and if it happened later in the year, then it leads to a 5.2066 increase in the log-odds of a successful coup.

The most important takeaway of the remaining coefficients is that as the positive coefficient increases, its contribution towards a successful and realized coup increases. Likewise, as the negative coefficient decreases, its contribution towards a successful and realized coup decreases.

5.2 Random Forest

Random forest is an ensemble model that has a tree for every bootstrap. This method also uses a subset of the predictors for each bootstrap to avoid splitting on the same predicting, which allows for interpretations on the importance of many predictors because one will not dominate. In Random Forest, we typically use a depth that will make each individual tree overfit but by aggregating across these trees we can still get an ideal model that will not be overfitted. Random forest starts from the concept of high variance and low bias, to then decrease variance and increase bias through aggregation (averaging). Consequently, Random Forest is a strong instrument that will better avoid correlation between the trees.

We consider this model to be adapt for this data-set given that our problem is a classification problem, i.e. successful coup or unsuccessful. We apply Random Forest using the same aforementioned selected cross-interaction terms. First, we cross validate across 28 depths on a single Decision Tree to find the lowest mean validation score. This will indicate the worst possible depth among these that will be hyperparameter for our Random Forest.

The outcome of this model however doesn't prove to be appropriate as we thought. The test accuracy of our random forest is lower than our baseline mode. Its train accuracy is 99.24%, and on **the test is 72.438%**.

5.2.1 Most important predictors

Out of 100 bootstraps of our data, the predictors that seem to be the most important (how many times this feature was the top node in Random Forest), according to the model are:

Predictors	# top node	Predictors	# top node
day dissident	18	palace	5
dissident	17	popular noharm	3
year popular	13	palace noharm	3
dissident noharm	12	day dissident foreign	2
month dissident noharm	11	day military noharm	1
military dissident	7	day rebel	1
year palace	6	military counter	1

Table 3: List of Predictors and the amount of times they were split for the top node

From looking at the table, we can see what Random Forest saw as important predictors based on what predictors were considered as the top node across all bootstraps. It is important to see that day dissident was the most split on with dissident closely following, which we can interpret as coups initiated by small dissident groups are dominating predictions made within Random Forest.

6 209 Extension Model: Multi-Layer Perceptron

6.1 Background

For our extension, we chose to use Multi-Layer Perceptron, which is a Neural Network that trains iteratively at each step by updating the parameters due to calculations using loss function, which in our case is cross-entropy. Cross-entropy, in binary classification, can be calculated as:

$$-(ylog(p) + (1-y)log(1-p))$$

where p is the probability that the model thinks it is of class 1, y is the binary indicator (i.e. 0 or 1).

Neural networks are able to provide more insight to our data than logisitic regressions because they use layers, allowing for our model to "learn". Neural networks have an input and output layer, where we pass in our predictors and receive an output from each layer respectively. In between the input and output we have hidden layers that are connected and feed into each other by giving weights to the predictors.

The hidden layers create these weights through an activation function, which creates the non-linearity in our model. The activation function is present in between each layer to adapt the previous layer's input based on the function. The activation function chosen for our MLP model is the logistic activation function because our dataset consists most of categorical variables that fall between 0 and 1. The logistic function uses the sigmoid function that covers these values of 0 and 1.

6.2 Fine tuning MLP Paramaters

We want to select the best hyper-parameters for our model through cross-validation in order to avoid overfitting to a single validation dataset. This selection can be done using GridSearchCV, which will return the best parameters after cross-validations are done. For our MLP model, we need to select the solver, amount of hidden layers, amount of neurons in the layers, and amount of iterations.

6.2.1 Solver: Adaptive Moment Estimation (Adam)

Through GridSearchCV, we found that Adam was the solver that was consistently performing the best compared to SGD and LBFGS. This is supported when plotting the curve of the loss function, shown in figure 3, where the Adam curve converges to a lower value at a faster rate. With these reasonings, Adam would be the best solver to use for our MLP model.

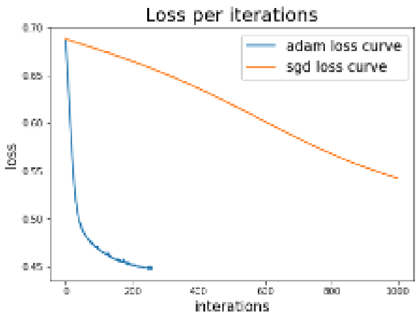

Figure 3: Adam Loss curve and SGD loss curve

Adam extends past stochastic gradient descent and works to take the advantages from the Adaptive Gradient Algorithm Root Mean Square Propagation. Throughout training, Adam keeps a single learning rate for weight updates and this learning rate does not change. However, there is an additional learning rate for each parameter, which is different from the learning rate of the model, that is separately adapted as the learning continues. Adam does this by using the average of the second moments or v_t of the gradient, also called the uncentered variance. Adam still calculates the first moment or m_t, which is used in combined with the second movement to create the exponential moving average of the gradient. The equations for the first and second moments are shown below:

First moment:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

Second moment:

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

This moving average is used to place a larger significance and weight on the recent data points to create a bias calculation for each moment.[2]

$$\hat{m_t} = \frac{m_t}{(1 - \beta_1^t)}$$
$$\hat{v_t} = \frac{v_t}{(1 - \beta_2^t)}$$

Adam will continue calculating the moments and moving averages until the bias converges to zero, and thus giving us our solution. [3]

6.2.2 Hidden layer sizes

We fine tuned the number of hidden layers and the number of neurons per hidden layer by trying several pairs. An example is: two hidden layers (in total 4 layers) where each hidden layer has 150 neurons, i.e. hidden_layer_sizes = (150, 150). Also, we tried to fine tune the number of iterations. However, with GridSearchCV this computation was time and computationally expensive, hence we use 1000 iterations. Adam ends up needing to use only 260 to reach the minima. Lastly, we are not fine tuning the learning rate (adaptive or constant) because this applied when the optimization method chosen is SGD, and not adam.

From GridSearchCV with a chosen list of hidden layer sizes for our MPL model, we found that the highest validation score is at 77% with 1 hidden layer and 300 neurons in this hidden layer, i.e. at hidden_layer_sizes = (300,).]. We then use those identified parameters to build our model.

6.3 Results

This model has a higher predictive power compared to the well fine tuned Logistic Regression with L1 penalization and compared to Random Forest. The best loss we reached is 0.447. MLP has a **test score** of **75.971%**. The model learns a lot about the data in each layer, and we can see that the model was able to learn about these cross

[2]These were referenced to summarize the equations: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/ & https://ruder.io/optimizing- gradient-descent

[3]This was just a brief summary, for more reference please refer to the original paper here: `https://arxiv.org/pdf/1412.6980.pdf`.

interaction terms more compared to what the LogisticRegressionCV model learned about the data. However, although the predictive power is higher, we loose interpretability.

7 Model Comparison

7.1 Accuracy Scores

Here we will report all of the train and test accuracy scores across all modelling.

Model	Train Score	Test Score
Baseline model	75.152%	73.852%
LogisticRegressionCV	79.394%	74.912%
Random Forest	98.636%	72.085 %
MLP	77.576%	75.972%

Table 4: Comparision of Accuracy Scores across the model

As shown above, each improvement from the baseline model has increased the value of the test score except for Random Forest, which was our lowest score. A potential reason for this is because all of our other models had a consistent random state value that we utilized in fine-tuning. Random state did not have this parameter, and, therefore, we could not keep those control the outcomes in the same way.

7.2 False Positive Rates and False Negative Rates

Essentially, our model deals with a classification problem. Given a set of predictors, it learns what are the most significant predictors that contribute to the success of a coup and tries to classify them when tested. Thus, it is important to measure the performance of the classification model. To do so, we check the False Positive Rate (FPR) and False Negative Rate (FNR).

In our case, FPR means that our model predicts a realized coup when in reality the true data says it was unrealized. A FNR means our model predicts an unrealized coup when our true data says it was realized.

From here we can compare the FPR and FNR from each model, we consistently see our model favoring FPR over FNR, except for the MLP. By favoring FPR, it means our model is more likely to predict a realized coup when the data says there was an unrealized coup. The MLP model, however, favored FNR, meaning it predicted unrealized coups when the true outcome is realized.

Model	FPR	FNR
Baseline model	0.371	0.121
LogisticRegressionCV	0.371	0.121
Random Forest	0.302	0.242
MLP	0.189	0.306

Table 5: False Positive and False Negative Rates for each Model's Test Data

8 Further explorations

In addition to exploring what constitutes a successful coup, we also explore different geographic sub-regions and investigate if there could be differences in how a coup succeeds or fails in those regions. We aim to answer questions like is a coup more likely to be realized in Africa when their is a military force president? Or does it matter more if there it was initiated by members of a faction within the existing government (palace)? How does this compare to a country in Asia? Or in Latin America?

Additionally, we explore how coups have changed throughout time. As modern warfare has changed with technology, has how coups succeed also change? Could there be differences between time frames where decolonization happened? These questions we have in mind will guide us through these explorations.

The same response variable, realized, is used and the same list of problematic predictors that reveal the response variable is dropped.

8.1 Investigating Sub-Regions & Coup Success

The data is divided into 3 larger sub-regions. These sub-regions will be the Americas (North and Latin America), Africa, and Euro-Asia (includes the Middle East). Countries in the Carribean were grouped with the Americas. While we understand that some of those countries have ethnic ties to other regions, the geographic spacing was our main factor in this decision. The three regions were chosen in order to avoid any dimensionality issues between using interaction terms and the amount of observations for each data frame. We acknowledge that the political climates of these sub-regions may not completely align but we want to avoid overfitting while still have the ability to interpret the outcomes to compare the differences between these regions. We create a sub data frames for each region and perform the same Lasso LogisticRegressionCV to find the important predictors.

Results Seventy-three predictors were considered important by the LogisticRegressionCV model. In Figure 7, we plot 13 predictors where each predictor is considered significant by at least two regions. There were only 3 predictors that were deemed significant in all three regions. These predictors along with their coefficients are in Table 4.

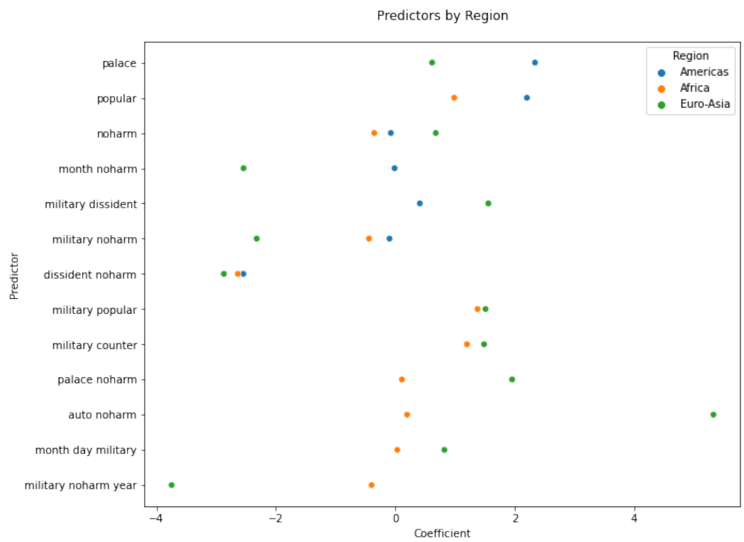

Figure 4: Selected Predictors from all Regions

Dissident and No Harm - Since all the coefficients for the two-way interaction term of 'dissident noharm' are negative, then it means that if a coup initiated by a small discontent group in any of the regions and the executive was not harmed during the coup, the chances of the coup being successful is decreased. The most negative contribution to the success of a coup is in Euro-Asia.

Military and No Harm - Likewise, all the coefficients for the two-way interaction term of 'military noharm' are negative. This means that if a coup was initiated by military actors who are not a formal part of the governing apparatus and the executive was not harmed during the coup, the chances of the coup being successful is decreased. The impact of this interaction term in the Euro-Asia model is 24x more significant than in Americas and 5x more significant than Africa.

No Harm - The noharm predictor has a negative coefficient in the Africa and Americas. This means that no harm to the person in power during the coup event means a -0.3527 and -0.0743 decrease in log-odds of a successful coup in Africa and Americas respectively. However, because the noharm predictor has a positive coefficient in Euro-Asia, it indicates no harm to the executive will contribute with a 0.6784 increase in the log-odds of a successful coup.

Region	dissident noharm	military noharm	noharm
Africa	-2.6365	-0.4380	-0.3527
Americas	-2.5440	-0.096	-0.0743
Euro-Asia	-2.8709	-2.3201	0.6784

Table 6: Common Predictors in all regions

8.2 Investigating Time-Frames & Coup Success

In this section, the data is divided into two sub time frames. The division is based on the number of observations - we decided on looking for an equal split in order to avoid dimensionality issues. The equal split happens at 1975, so we group all observations that take place between 1945 and 1975 as one time frame and 1976 - 2019 as the second time frame. We also decided to not focus on any historical time frames as the data's earliest observation are after the end of WWII and there aren't any other historical events afterwards that significantly impacted all of the countries in the data. We will consider historical events in our evaluation of the coefficients but not as a splitting metric.

Results For the time-frame between 1945-1975, only two predictors considered important. Those predictors were 'dissident' and 'palace'. For the time-frame between 1976-2019, 17 predictors were considered important. Those predictors were 'month', 'rebel', 'palace', 'popular', 'injured', 'cowcode month', 'month foreign', 'day foreign', 'military dissident', 'military popular', 'military noharm', 'dissident noharm', 'palace noharm', 'auto noharm', 'cowcode military noharm', 'month foreign noharm', and 'day foreign noharm'. The only commonality between both time frames is 'palace' with a coefficient of 0.0173 in the earlier time-frame and 0.5481 in the later timeframe. **Palace**-Considering the difference between the time-frames, the coefficient value tells us that it is more likely for coups initiated after 1975 to be successful if they were initiated by members of faction within the existing government (i.e. ministers, cabinet members, military personnel's (if members of ruling military junta) or other high ranking people in the executive branch) as a positive coefficient positively contributes to the chances of a successful/realized coup.

9 Conclusion

From our initial EDA, we were able to develop a research question and extension questions that we used to guide our modelling. We started with a baseline LogisticRegressionCV model with a reduced predictor list that was chosen passed naive analysis and understanding the definition of predictors. To improve our model, we do feature selection of all possible two way and three way interaction terms. We then pass this reduced list to another LogisticRegressionCV model that improved our test score. We also tried to use Random Forest to improve our model but found that it did not. The last model we tried was a Multi-Layer Perceptron that produced the best test score. Lastly, we investigated our extension questions and looked at how our predictors impact the success of coup across different sub-regions and time frames.

References

1: Coup d'état project (CDP). Coup D'état Project (CDP) — Cline Center. (n.d.). Retrieved December 11, 2021, from https://clinecenter.illinois.edu/project/research-themes/democracy-and-development/coup-detat-project-cdp.

2: Brownlee, J. (2021, January 12). Gentle introduction to the adam optimization algorithm for deep learning. Machine Learning Mastery. Retrieved December 11, 2021, from https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/.

3: Sebastian Ruder. (2020, March 20). An overview of gradient descent optimization algorithms. Sebastian Ruder. Retrieved December 11, 2021, from https://ruder.io/optimizing-gradient-descent/.

5: 'Adam: a method for stochastic optimization', Diederik P. Kingma, Jimmy Lei Ba. ICLR 2015.