CS109A Final Project

Modeling Coup d'Etat Data

Andrew Nickerson, Caleb Saul,
Drew Webster, Ayana Yaegashi

Motivation

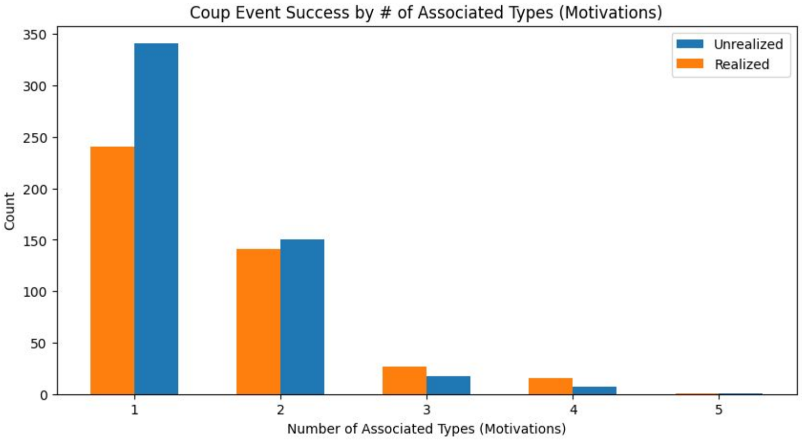
- Coups are very important in global events, even recently
    - Peruvian coup just days ago
    - Assassination of Japanese former prime minister over the summer
- Coups can change the course of a country’s history as well as the lives of the people in that country; it is important to be able to try to understand better what causes them

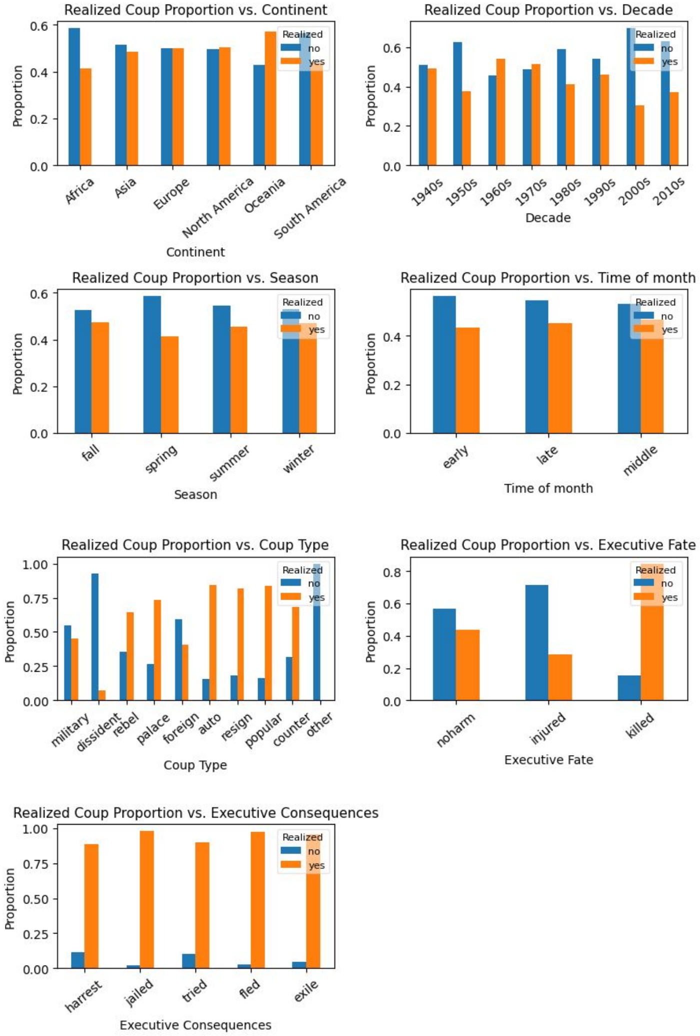
Dataset Overview

The Cline Center Coup D'état Project (CPD) Dataset "identifies coups, attempted coups, and coup plots/conspiracies in 136 countries (1945-2019). The data identifies the type of actor who initiated the coup (i.e.military, palace, rebel, etc.) as well as the fate of the deposed executive (killed, injured, exiled, etc.)."

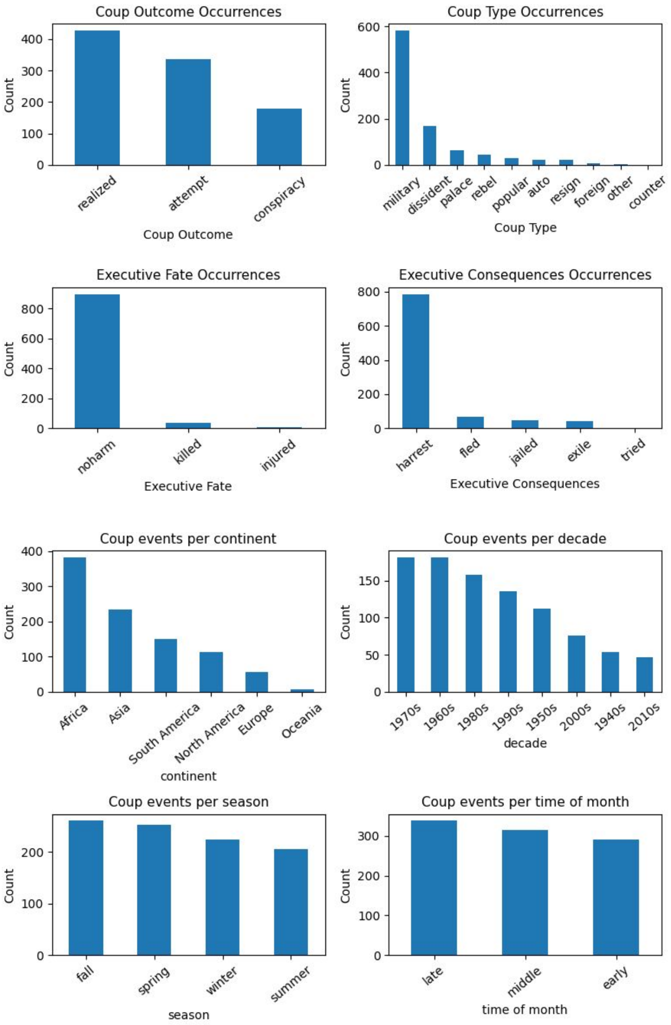
- Dropping unnecessary columns
  - **cowcode**, **coup_id**, **event_type**, and **unrealized** were dropped due to the fact that their data was already encoded in other columns and were therefore unnecessary
- Dropping "cheat" columns
  - **executive_fate**, **executive_consequence**
- Encoding predictors:
  - Country → continent
  - mm/dd/yyyy → time of month, season, decade
- One-hot-encoding categorical variables:
  - ex) continent → columns for Africa, South America, etc.

Exploratory Data Analysis





Exploratory Data Analysis



Research Questions

1. **Binary Prediction:** Can we predict whether a coup event will be successful (realized) based on some set of predictors? If so, which predictors are most useful for this task?
2. **Multi-Class Prediction:** Can we predict if an event was a conspiracy, an attempted coup, or an actual coup? If so, which predictors are most useful in this scenario?

Multi-Model Approach

- Baseline logistic regression (no regularization)
- Logistic regression with Lasso regularization
- Single Decision Tree
- Bagging
- Random Forest
- Boosting
  - AdaBoost
  - Gradient Boosting

Results (Binary Prediction)

Logistic Regression

No Regularization
Train Score: 0.7384
Test Score: 0.7619

LASSO
Train Score: 0.7517
Test Score: 0.7460

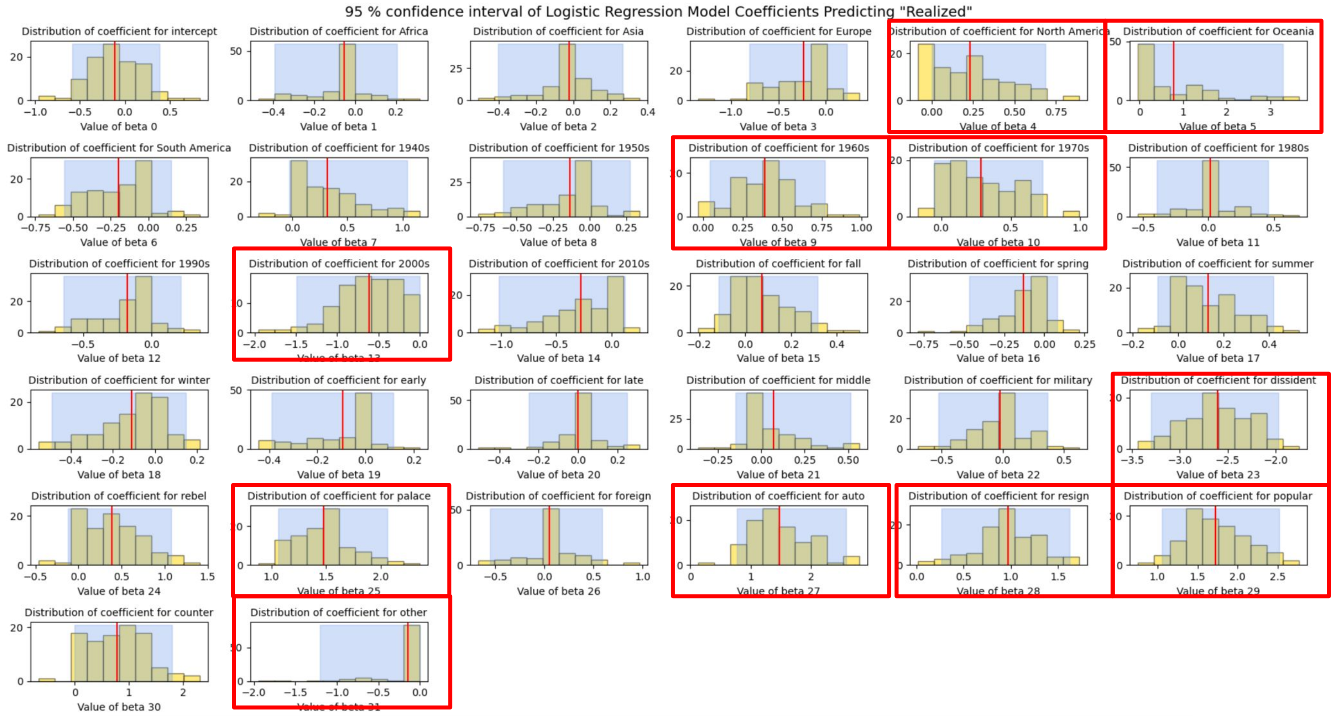
We found the following predictors were least important (zeroed out by Lasso): **Africa, Asia, late, military, foreign, other**

We note that the following coefficients did not include 0 in a 95% confidence interval after 100 bootstraps: **North America, Oceania, 1960s, 2000s, dissident, palace, auto, resign, popular, counter, and other**.

intercept	-0.2120
Africa	0.0000
Asia	0.0000
Europe	-0.2233
North America	0.2046
Oceania	0.5547
South America	-0.1533
1940s	0.3796
1950s	-0.0007
1960s	0.4104
1970s	0.3309
1980s	0.0002
1990s	-0.0592
2000s	-0.4636
2010s	-0.1717
fall	0.0399
spring	-0.0990
summer	0.1108
winter	-0.1196
early	-0.0555
late	0.0000
middle	0.0382
military	0.0000
dissident	-2.4630
rebel	0.2920
palace	1.4140
foreign	0.0000
auto	1.4594
resign	0.9653
popular	1.6418
counter	0.7712
other	0.0000

Results (Binary Prediction)

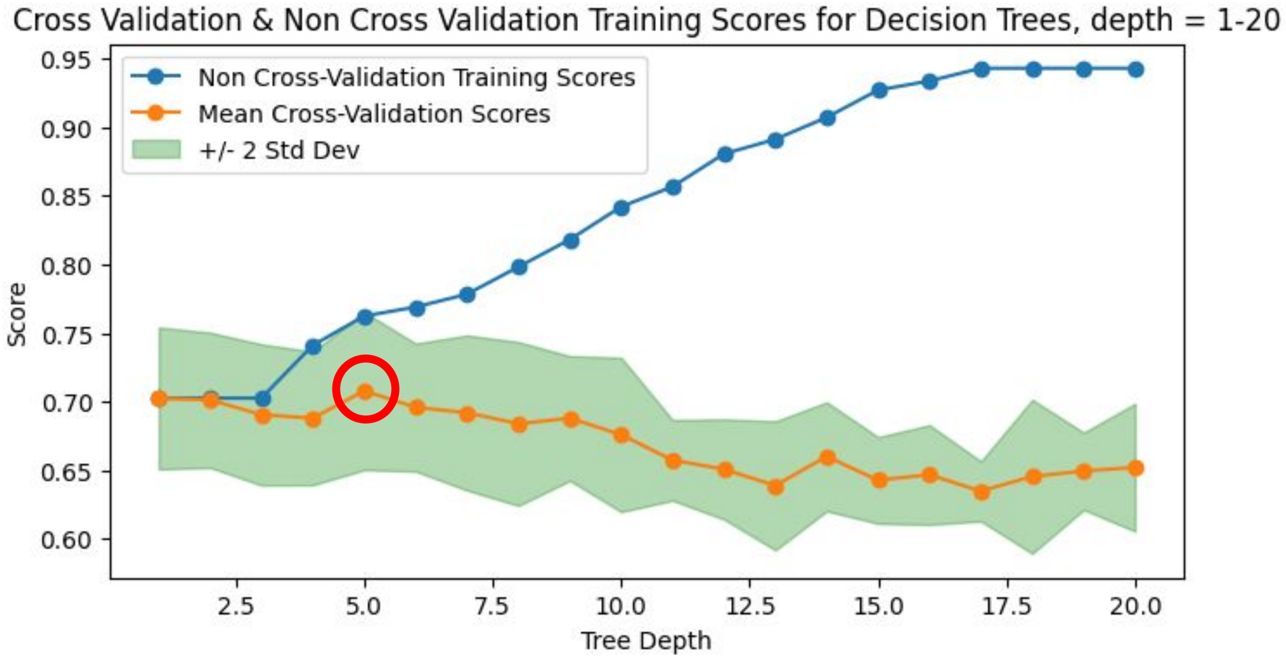
Logistic Regression



After bootstrapping, we had a better idea as to which predictors were most important. We note that the following coefficients did not include 0 in a 95% confidence interval after 100 bootstraps, sampling with replacement from our training data (753 data points): **North America, Oceania, 1960s, 2000s, dissident, palace, auto, resign, popular, counter, and other**.

Results (Binary Prediction)

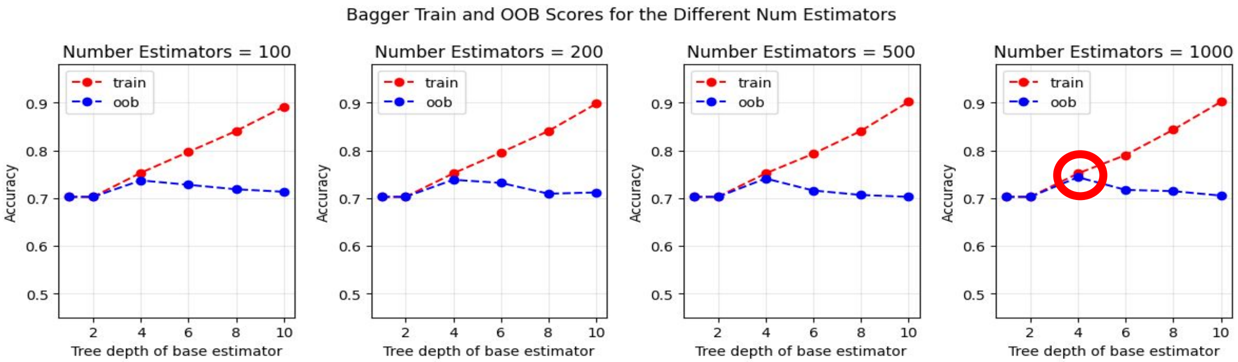
Single Decision Tree



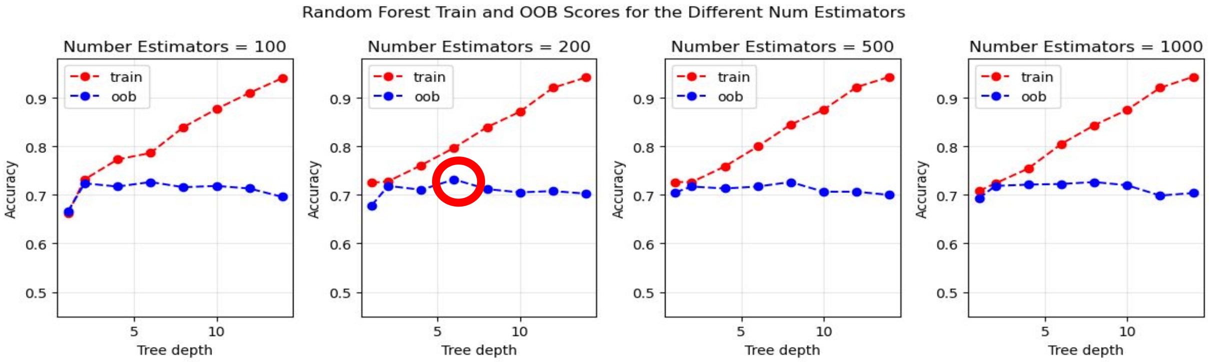
Best Depth: d = 5
Train Score: 0.7623
Test Score: 0.6931

Results (Binary Prediction)

Bagging & Random Forest Models



Best Depth: d = 4
Best Num Estimators = 1000
Train Score: 0.7530
Test Score: 0.7302



Best Depth: d = 6
Best Num Estimators = 200
Train Score: 0.7968
Test Score: 0.7460

Results (Binary Prediction)

Feature Importance: Bagging & Random Forest

Bagging
Best Depth = 4
Best Num Estimators = 1000
Train Score = 0.7530
Test Score = 0.7302

Random Forest
Best Depth = 6
Best Num Estimators = 200
Train Score: 0.7968
Test Score: 0.7460

Weight	Feature
0.1868 ± 0.0747	dissident
0.0365 ± 0.0261	military
0.0222 ± 0.0114	popular
0.0164 ± 0.0493	decade
0.0048 ± 0.0145	resign
0.0026 ± 0.0071	rebel
0.0016 ± 0.0142	palace
0.0005 ± 0.0032	counter
0 ± 0.0000	other
0 ± 0.0000	spring
0 ± 0.0000	Asia
0 ± 0.0000	Europe
0 ± 0.0000	North America
0 ± 0.0000	Oceania
0 ± 0.0000	South America
0 ± 0.0000	fall
0 ± 0.0000	late
0 ± 0.0000	summer
0 ± 0.0000	winter
0 ± 0.0000	early
0 ± 0.0000	middle
0 ± 0.0000	foreign
0 ± 0.0000	Africa
-0.0021 ± 0.0052	auto

Bagging Model

Weight	Feature
0.1413 ± 0.0576	dissident
0.0317 ± 0.0189	popular
0.0217 ± 0.0397	military
0.0169 ± 0.0221	resign
0.0122 ± 0.0268	decade
0.0122 ± 0.0292	palace
0.0106 ± 0.0134	middle
0.0090 ± 0.0116	late
0.0090 ± 0.0134	rebel
0.0074 ± 0.0118	auto
0.0063 ± 0.0123	Africa
0.0053 ± 0.0000	other
0.0048 ± 0.0032	summer
0.0048 ± 0.0057	winter
0.0037 ± 0.0048	Asia
0.0016 ± 0.0068	foreign
0.0005 ± 0.0057	counter
0.0005 ± 0.0088	fall
0.0000 ± 0.0067	Europe
0 ± 0.0000	Oceania
-0.0026 ± 0.0127	spring
-0.0026 ± 0.0098	South America
-0.0079 ± 0.0118	North America
-0.0095 ± 0.0104	early

Random Forest Model

Results (Binary Prediction)

Boosting

AdaBoost
Best Depth: d = 4
Best Num Estimators = 200
Best Learning Rate = 0.005
Train Score: 0.7676
Test Score: 0.7354

Gradient Boosting
Best Depth: d = 2
Best Num Estimators = 800
Best Learning Rate = 0.005
Train Score: 0.7543
Test Score: 0.7249

Weight	Feature
0.1815 ± 0.0828	dissident
0.0376 ± 0.0286	military
0.0243 ± 0.0127	popular
0.0143 ± 0.0469	decade
0.0053 ± 0.0047	other
0.0053 ± 0.0116	South America
0.0053 ± 0.0183	palace
0.0037 ± 0.0134	resign
0.0026 ± 0.0098	rebel
0.0016 ± 0.0048	counter
0 ± 0.0000	spring
0 ± 0.0000	Asia
0 ± 0.0000	North America
0 ± 0.0000	Oceania
0 ± 0.0000	fall
0 ± 0.0000	late
0 ± 0.0000	summer
0 ± 0.0000	winter
0 ± 0.0000	middle
0 ± 0.0000	foreign
0 ± 0.0000	Africa
-0.0005 ± 0.0032	Europe
-0.0026 ± 0.0053	early
-0.0026 ± 0.0071	auto

AdaBoost

Weight	Feature
0.1889 ± 0.0716	dissident
0.0339 ± 0.0251	military
0.0217 ± 0.0138	popular
0.0111 ± 0.0515	decade
0.0037 ± 0.0171	resign
0.0005 ± 0.0138	palace
0.0005 ± 0.0032	counter
0 ± 0.0000	other
0 ± 0.0000	summer
0 ± 0.0000	Asia
0 ± 0.0000	Europe
0 ± 0.0000	North America
0 ± 0.0000	Oceania
0 ± 0.0000	South America
0 ± 0.0000	spring
0 ± 0.0000	late
0 ± 0.0000	winter
0 ± 0.0000	early
0 ± 0.0000	middle
0 ± 0.0000	rebel
0 ± 0.0000	foreign
0 ± 0.0000	Africa
-0.0011 ± 0.0042	fall
-0.0079 ± 0.0053	auto

Gradient Boosting

Results (Binary Prediction)

	train	test
Baseline Logistic Regression	0.7384	0.7619
Lasso Logistic Regression	0.7517	0.7460
Single Decision Tree Depth = 5	0.7623	0.6931
Bagging	0.7530	0.7302
Random Forest	0.7968	0.7460
Adaboost	0.7676	0.7355
Gradient Boosting	0.7543	0.7249

Results (Multi-Class Prediction)
Logistic Regression

No Regularization
Train Score: 0.6401
Test Score: 0.6032

LASSO
Train Score: 0.6401
Test Score: 0.6032

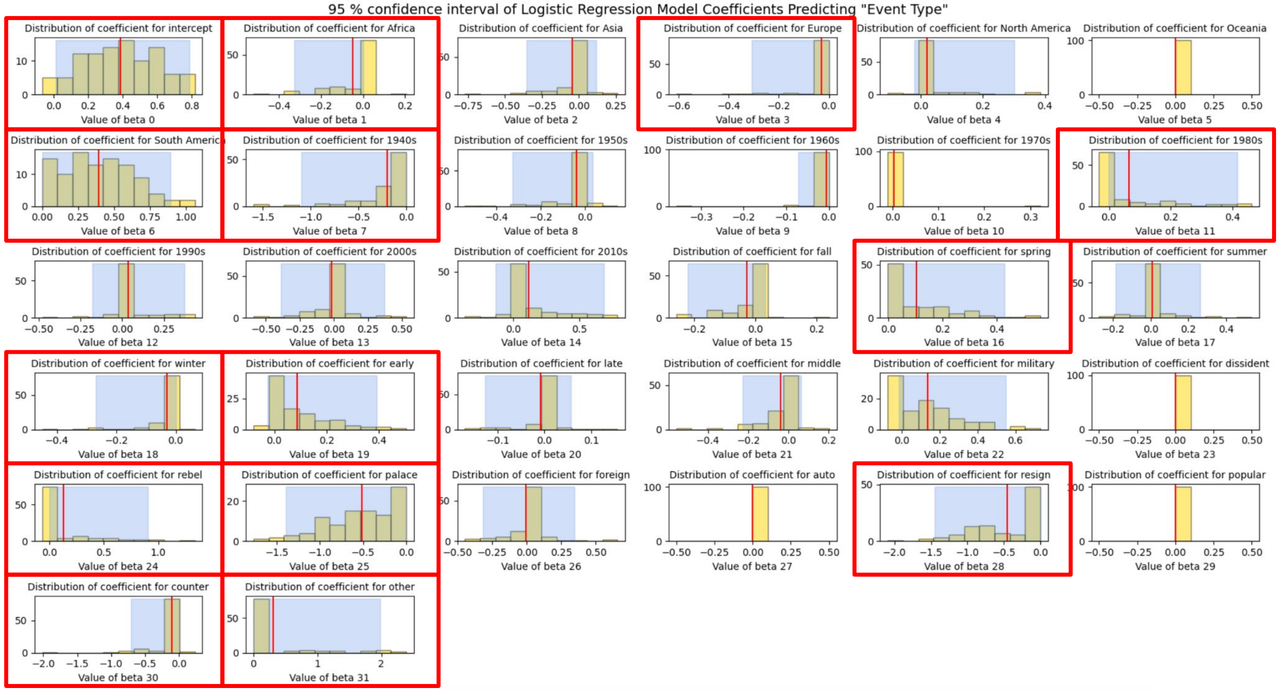
We found the following predictors were least important (zeroed out by Lasso): **1960s, 1970s, 1980s, 1990s, 2000s, summer, winter, late, dissident, rebel, counter**

We note that the following coefficients did not include 0 in a 95% confidence interval after 100 bootstraps: **Africa, Europe, Oceania, South America, 1950s, 1960s, 1980s, dissident, rebel, palace, auto, resign, popular, and other.**

intercept	0.5324
Africa	-0.0819
Asia	-0.0555
Europe	-0.0035
North America	-0.0017
Oceania	0.5071
South America	0.2737
1940s	-0.1267
1950s	-0.0076
1960s	0.0000
1970s	0.0000
1980s	0.0000
1990s	0.0000
2000s	0.0000
2010s	0.1701
fall	-0.0003
spring	0.1001
summer	0.0000
winter	0.0000
early	0.1423
late	0.0000
middle	-0.0274
military	0.2258
dissident	0.0000
rebel	0.0000
palace	-0.4481
foreign	-0.0042
auto	0.4199
resign	-0.3178
popular	0.4167
counter	0.0000
other	5.3492

Results (Multi-Class Prediction)

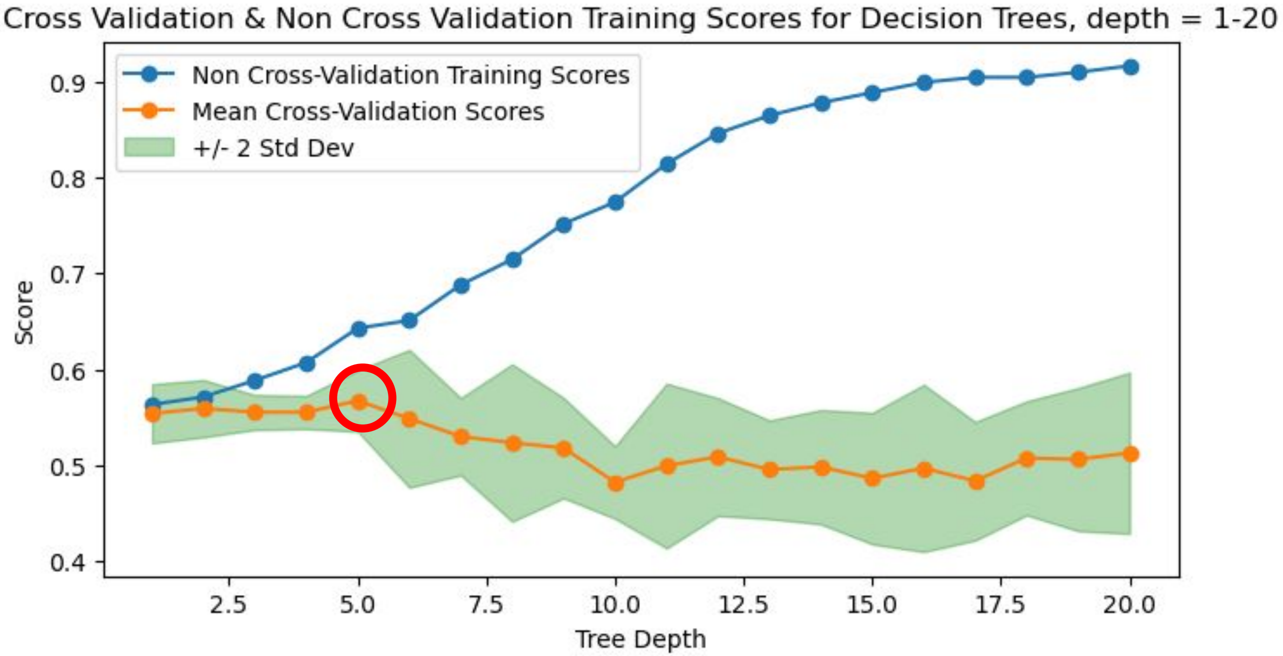
Logistic Regression



After bootstrapping, we had a better idea as to which predictors were most important. We note that the following coefficients did not include 0 in a 95% confidence interval after 100 bootstraps, sampling with replacement from our training data (753 data points): **Africa, Europe, Oceania, South America, 1950s, 1960s, 1980s, dissident, rebel, palace, auto, resign, popular, and other.**

Results (Multi-Class Prediction)

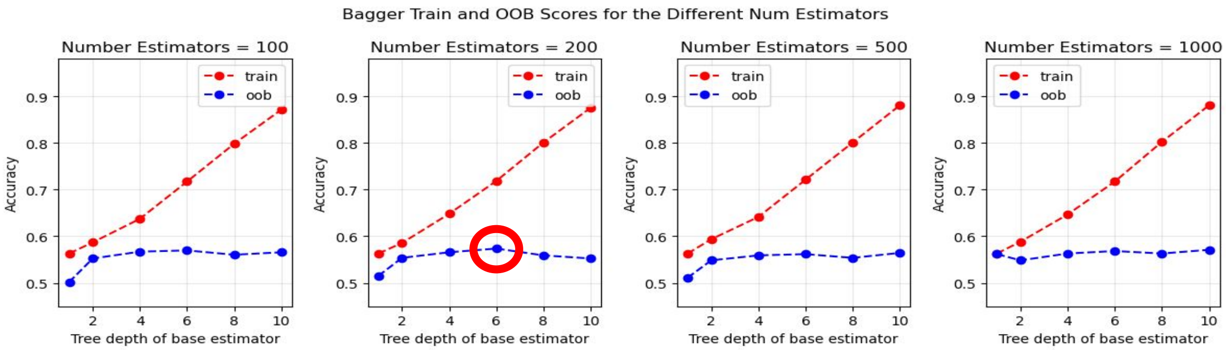
Single Decision Tree



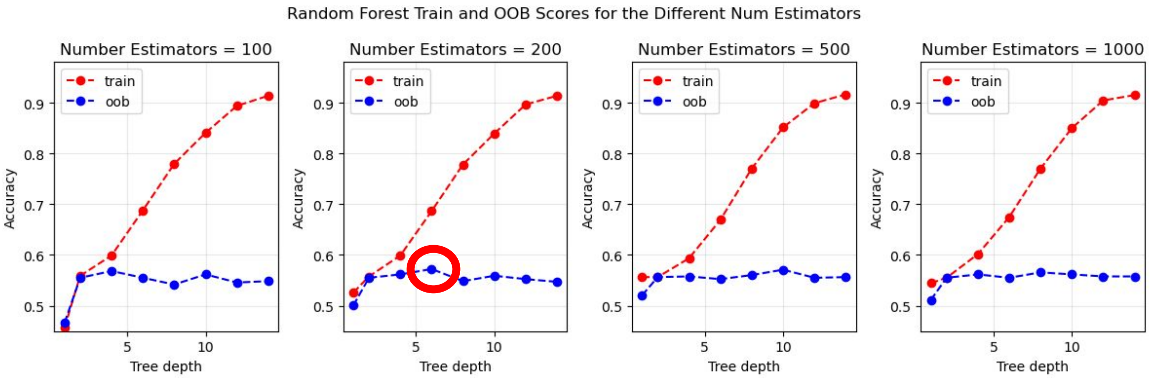
Best Depth: d = 5
Train Score: 0.6428
Test Score: 0.5714

Results (Multi-Class Prediction)

Bagging & Random Forest Models



Best Depth: d = 6
Best Num Estimators = 200
Train Score: 0.7185
Test Score: 0.5661



Best Depth: d = 6
Best Num Estimators = 200
Train Score: 0.6866
Test Score: 0.5820

Results (Multi-Class Prediction)
Boosting

AdaBoost
Best Depth: d = 4
Best Num Estimators = 200
Best Learning Rate = 0.005
Train Score: 0.7676
Test Score: 0.7354

Gradient Boosting
Best Depth: d = 2
Best Num Estimators = 800
Best Learning Rate = 0.005
Train Score: 0.7543
Test Score: 0.7249

Permutation Feature Importance

Weight	Feature
0.1286 ± 0.0568	dissident
0.0376 ± 0.0243	palace
0.0233 ± 0.0399	decade
0.0212 ± 0.0195	popular
0.0201 ± 0.0295	Africa
0.0190 ± 0.0127	winter
0.0132 ± 0.0118	auto
0.0090 ± 0.0068	North America
0.0085 ± 0.0269	military
0.0079 ± 0.0238	fall
0.0048 ± 0.0074	rebel
0.0037 ± 0.0083	foreign
0.0026 ± 0.0144	Europe
0.0011 ± 0.0079	Oceania
0.0000 ± 0.0206	resign
0 ± 0.0000	Asia
-0.0005 ± 0.0192	spring
-0.0011 ± 0.0063	counter
-0.0011 ± 0.0156	summer
-0.0016 ± 0.0106	middle
-0.0021 ± 0.0172	South America
-0.0032 ± 0.0052	early
-0.0053 ± 0.0082	other
-0.0169 ± 0.0221	late

AdaBoost

Weight	Feature
0.1270 ± 0.0568	dissident
0.0392 ± 0.0260	palace
0.0317 ± 0.0201	military
0.0286 ± 0.0202	Africa
0.0238 ± 0.0273	North America
0.0233 ± 0.0277	spring
0.0175 ± 0.0184	rebel
0.0169 ± 0.0226	resign
0.0169 ± 0.0507	decade
0.0153 ± 0.0145	auto
0.0143 ± 0.0095	middle
0.0138 ± 0.0185	popular
0.0132 ± 0.0228	South America
0.0095 ± 0.0063	counter
0.0079 ± 0.0127	winter
0.0063 ± 0.0063	foreign
0.0048 ± 0.0274	late
0.0042 ± 0.0250	early
0.0021 ± 0.0070	Europe
0.0016 ± 0.0177	Asia
0.0011 ± 0.0079	Oceania
-0.0016 ± 0.0177	summer
-0.0053 ± 0.0082	other
-0.0063 ± 0.0092	fall

Gradient Boosting

Results (Multi-Class Prediction)

	train	test
Baseline Logistic Regression	0.6401	0.6032
Lasso Logistic Regression	0.6401	0.6032
Single Decision Tree Depth = 5	0.6428	0.5714
Bagging	0.7184	0.5661
Random Forest	0.6865	0.5820
Adaboost	0.6215	0.5873
Gradient Boosting	0.6401	0.6031

Future Work

- **Imbalanced Class Representation**
  - Solution: Oversampling

- **Poor Predictors**
  - Solution: Expanded Dataset

Test Accuracy by Class

	coup	attempted	conspiracy
Baseline Logistic Regression	0.8023	0.4627	0.3889
Lasso Logistic Regression	0.8023	0.4627	0.3889
Single Decision Tree Depth = 5	0.7907	0.3731	0.4167
Bagging	0.8140	0.3731	0.3333
Random Forest	0.9302	0.3284	0.2222
Adaboost	0.8256	0.3582	0.4444
Gradient Boosting	0.8023	0.4627	0.3889