
Entrega 3: Introducción y Resultados

José Saint Germain
joesg998@gmail.com

Índice

2	1. Introducción	3
3	1.1. Motivación	3
4	1.2. Estructura del documento	3
5	2. Marco Teórico	3
6	2.1. La Teoría de la Modernización y sus variantes	3
7	2.2. Teoría de la Dependencia y el Subdesarrollo	4
8	2.3. Datos Empíricos y Nuevas Perspectivas	4
9	3. Metodología	4
10	3.1. Algoritmos de predicción	4
11	3.2. Métrica de evaluación	5
12	3.3. Optimización de hiperparámetros	5
13	3.4. Block-time-series cross-validation	5
14	3.5. Valores Shapley	5
15	3.6. Ingeniería de atributos	6
16	3.7. Análisis Exploratorio de Datos	6
17	4. Resultados y discusión	9
18	4.1. Performance de los modelos	9
19	4.2. Análisis de variables	10
20	4.3. Discusiones	12
21	5. Conclusiones	12
22	6. Anexo	13
23	6.1. Código	13
24	6.2. Golpes	13
25	6.3. Gráficos y tablas adicionales	13

26 1. Introducción

27 El objetivo de este trabajo es lograr entrenar un modelo de aprendizaje automático que logre predecir
28 de manera aceptable la presencia de golpes de estado durante los años 2020 a 2022 en todos los
29 países del mundo a partir de la utilización de la base de datos provista por la fundación Varieties
30 of Democracy (V-Dem) (Coppedge, Gerring, Knutsen, Lindberg et al., 2024), así como tener una
31 noción acabada de las variables más importantes que los algoritmos utilizan para la predicción de la
32 variable objetivo.

33 1.1. Motivación

34 La motivación de este trabajo es dialogar con el artículo recientemente realizado por el Fondo
35 Monetario Internacionl (FMI) Cebotari et al., 2024. En el mismo se aborda el mismo objeto de estudio
36 utilizando diversas metodologías, siendo una de ellas la utilización de algoritmos de aprendizaje
37 automático. En este trabajo se replicó la metodología utilizada en esa sección; comparando los
38 mismos modelos, sus respectivos hiperparámetros y la métrica a maximizar durante su entrenamiento

39 La principal diferencia entre el paper del organismo y este trabajo radica en el origen de los datos.
40 Por un lado, el artículo del FMI utilizan 14 fuentes provenientes de diferentes organismos, de manera
41 de cubrir 5 grupos de variables sobre diferentes ámbitos (Desarrollo y demografía, Inclusión y
42 gobernanza, macroestabilidad, políticas públicas, estabilidad sociopolítica). En cambio, este trabajo
43 utilizará solamente la base de datos v-dem por dos motivos: en primer lugar, para abarcar solamente
44 variables que estén directamente ligadas a la situación política e insitucional de los países, excluyendo
45 en la medida de lo posible atributos ajenos a este ámbito. En segundo lugar, para realizar una
46 comparación con las nutridas y variadas fuentes del artículo citado. De esa manera, podemos tener
47 una noción del poder predictivo de atributos puramente político-institucionales frente a un abanico
48 más diverso de variables.

49 1.2. Estructura del documento

50 - Explicación de la estructura del trabajo (se realizará una vez que esté completado)

51 2. Marco Teórico

52 El estudio de los golpes de estado, así como los procesos de democratización han sido una preocupa-
53 ción central para la ciencia política moderna durante el siglo xx. Diversas teorías y contrateorías se
54 han desarrollado de manera de aprehender los causales de la democratización de un país así como de
55 su proceso inverso, ya sea una erosión democrática gradual o un golpe de estado autoritario; así como
56 los elementos sociales, culturales e institucionales que pueden evitar o disminuir la probabilidad de
57 que se produzcan estos fenómenos.

58 2.1. La Teoría de la Modernización y sus variantes

59 El estudio de los golpes de estado y la democratización ganó relevancia a mediados del siglo xx
60 con la teoría de la modernización. Entre los exponentes de esta teoría se encuentra Seymour Martin
61 Lipset quien con su artículo *"Some social requisites of democracy: economic developmente and
62 political legitimacy"* 1959. Desde un enfoque sociológico, argumenta que el grado de desarrollo
63 económico de una sociedad es una condición necesaria para el nacimiento y consolidación de un
64 régimen democrático, principalmente porque una sociedad dividida entre una masa empobrecida y
65 una élite rica es más propensa a generar una oligarquía (dictadura del estrato superior de la soicedad)
66 o una tiranía (dictadura basada en el estrato inferior).

67 Para medir el desarrollo económico, Lipset analiza y desgrega cuatro variables: el nivel de riqueza,
68 medido por pbi per cápita y por la cantidad de personas con vehículos de motor, radios, teléfonos y
69 diarios cada mil personas; el grado de industrialización, medido por el porcentaje de trabajadores
70 hombres en la agricultura y el nivel de energía utilizado per cápita (en toneladas de carbón); el
71 nivel de urbanización, medido en índices realizados previamente; así como el nivel educativo de la
72 población, del cual toma principalemnte la tasa de alfabetización. El autor subraya este último factor,
73 exponiendo que si no es una condición suficiente para la democracia, es una condición necesaria.

A su vez, Lipset describe cambios subyacentes en los diversos estratos sociales producto del desarrollo económico. En primer lugar, se desarrolla una suerte "lucha de clases" por parte de la clase baja, ya que mayores tasas de alfabetización y bienestar económico genera una visión más largoplacista y compleja de la política, desarrollando una ideología secular reformista y gradualista en la clase obrera. En segundo lugar, una clase media fortalecida y ensanchada por el crecimiento económico juega un papel mitigador del conflicto, penalizando extremismos y apoyando movimientos más moderados y democráticos. Por último, en una sociedad en donde las diferencias económicas entre clases sociales se moderan, se atenúan las percepciones negativas de las clases altas hacia las bajas, volviéndolas más tolerantes a compartir el poder y a otorgar derechos al resto de la sociedad. Por último, en una sociedad con mayor riqueza económica se expande la presencia de organizaciones intermedias e instituciones como fuentes de contrapeso al poder.

Si bien el desarrollo económico, caracterizado en los párrafos anteriores, se torna una condición mínima para la consolidación democrática, Lipset subraya dos condiciones suficientes para lograr su estabilidad en el tiempo: la efectividad del sistema político -entendida como la performance del sistema político para resolver problemas- y la legitimidad -es decir, la capacidad de lograr la creencia de que la existencia de instituciones políticas es deseable para el conjunto de la sociedad. Una crisis de legitimidad, por lo tanto, es contemplada como un factor de inestabilidad para un sistema democrático. Este tipo de crisis, según el autor, pueden surgir de determinados cambios en la estructura social: cuando todos los grupos mayoritarios no se aseguran el acceso al sistema político de manera temprana en un período de transición, o cuando el estatus de las instituciones conservadoras es amenazado.

En los años 70 y 80, la atención se desplazó hacia las transiciones políticas y la estabilidad institucional. Samuel P. Huntington, en *Political Order in Changing Societies* Huntington, 1968, argumentó que las sociedades en transición son particularmente vulnerables a la inestabilidad y los golpes de estado. Huntington sugirió que el ritmo del cambio social y político puede superar la capacidad de las instituciones para adaptarse, creando un vacío de poder y oportunidades para golpes de estado. Esta perspectiva destaca la importancia de variables como la estabilidad política, la cohesión institucional y la capacidad del gobierno.

2.2. Teoría de la Dependencia y el Subdesarrollo

- Teorías de la dependencia y el subdesarrollo: Cardoso y Faletto

- Análisis sociales: elites y movimientos sociales (O'donnell y Schmitter)

2.3. Datos Empíricos y Nuevas Perspectivas

- Enfoques empíricos y cuantitativos: Powell y Thyne; Przeworski.

- Machine learning en la ciencia política. Usos recientes

- Síntesis y cierre del marco teórico / estado del arte

3. Metodología

Puesto que buscamos replicar el mismo trabajo realizado por el FMI (Cebotari et al., 2024) con diferentes datos, vamos a replicar las mismas técnicas de optimización de hiperparámetros, así como los mismos algoritmos de entrenamiento y de interpretación de resultados.

3.1. Algoritmos de predicción

Los algoritmos que se utilizarán serán Random Forest (Breiman, 2001) y XGBoost (Chen y Guestrin, 2016). Ambos algoritmos son modelos de ensamble basados en múltiples árboles de decisión. Un árbol de decisión individual es un modelo predictivo que divide los datos en subconjuntos cada vez más pequeños basándose en una serie de decisiones binarias sobre las características de los datos. En cada nodo del árbol, se selecciona una característica y un umbral para dividir los datos en dos grupos: aquellos que cumplen la condición y aquellos que no. Este proceso se repite de manera recursiva hasta que se alcanza una condición de parada, ya sea un mínimo de muestras en un nodo o una profundidad máxima del árbol.

121 El algoritmo Random Forest (bosque aleatorio) busca combinar múltiples árboles de decisión con
122 características disímiles, combinando sus predicciones mediante un promedio (en regresión) o
123 mediante votación (en clasificación). La variedad de árboles se logra mediante una selección aleatoria
124 de un subconjunto de los datos con remplazo, así como seleccionando una proporción aleatoria de
125 atributos del dataset. De esa manera, se reduce la varianza del modelo, se evita el sobreajuste y se
126 mejora la capacidad predictiva.

127 Por otro lado, XGBoost (Extreme Gradient Boosting) es un algoritmo de boosting que mejora las
128 predicciones combinando múltiples árboles de decisión débiles (de menor capacidad predictiva)
129 de manera secuencial. A diferencia de Random Forest, donde los árboles se entrenan de forma
130 independiente, en el boosting los árboles se entrenan uno tras otro, cada uno tratando de corregir los
131 errores cometidos por los árboles anteriores. Particularmente, XGBoost utiliza la técnica de gradient
132 boosting, donde cada árbol nuevo se ajusta a los residuos (errores) del modelo anterior utilizando el
133 gradiente del error. Adicionalmente, XGBoost incluye algunas mejoras como la regularización y el
134 manejo eficiente de datos faltantes.

135 3.2. Métrica de evaluación

136 Adicionalmente, para la evaluación de performance se utilizará el área bajo la curva ROC (AUC). La
137 curva ROC es construida trazando la tasa de verdaderos positivos (la sensibilidad) frente a la tasa
138 de falsos positivos (especificidad) en diferentes umbrales de decisión. El área total de esta curva es
139 la que se utilizará para evaluar la performance del modelo. Esta métrica toma valores entre 0.5 y 1.
140 Un valor de AUC de 0.5 indica que el modelo no tiene mayor capacidad predictiva que el puro azar,
141 mientras que un valor cercano a 1 indica que el modelo es un excelente predictor. Las ventajas de
142 esta métrica son que es insensible al desbalance de clases y que proporciona una evaluación única del
143 rendimiento del modelo en distintos umbrales de decisión.

144 3.3. Optimización de hiperparámetros

145 Con respecto al ajuste de hiperparámetros se utilizará la optimización bayesiana. La misma consistirá
146 en 100 iteraciones en donde se buscará el valor óptimo de los siguientes hiperparámetros:

- 147 ■ Random Forest: profundidad máxima de los árboles (max_depth) y la submuestra del ratio
148 de columnas a considerar cuando se construye cada árbol (max_features).
- 149 ■ XGBoost: la tasa de aprendizaje (learning_rate) y el término de regularización L2 en los
150 pesos (reg_lambda).

151 Adicionalmente el parámetro que establece la cantidad de árboles creados (n_estimators) quedará
152 fijado en 1000.

153 3.4. Block-time-series cross-validation

154 Para evitar el data leakage, en cada iteración de la optimización bayesiana se utilizará la validación
155 cruzada. Sin embargo, como se trabajará con una base de datos de panel, conviene utilizar una versión
156 adaptada: el método *block-time-series cross-validation*, basado en Burman et al., 1994 y Racine,
157 2000. El método aplicado en este caso consiste en generar 5 pares de entrenamiento y validación:
158 1970 - 2009, 2010 - 2011; 1970 - 2011, 2012 - 2013; 1970 - 2013, 2014 - 2015; 1970 - 2015, 2016 -
159 2017; 1970 - 2017, 2018 - 2019. Por lo tanto, cada set de entrenamiento consiste en observaciones
160 desde 1970 hasta un año de corte (2009, 2011, 2013, 2015, 2017) y el set de validación contempla los
161 dos años siguientes del mismo. Una vez realizada la optimización bayesiana, se toman los valores de
162 hiperparámetros que lograron maximizar el AUC y se entrena el modelo con el set de entrenamiento
163 para intentar predecir los golpes de estado entre 2020 y 2022.

164 3.5. Valores Shapley

165 Para interpretar las variables más importantes en la predicción de golpes de estado, se utilizarán los
166 valores Shapley (Strumbelj y Kononenko, 2010; Lundberg y Lee, 2017). Basado en la teoría de
167 juegos, los valores Shapley consideran todas las posibles coaliciones de características y calculan
168 la contribución promedio de cada característica a través de todas las permutaciones posibles. En

169 otras palabras, determinan cuánto contribuye cada característica al valor de predicción del modelo,
170 considerando la interacción entre las características y evitando atribuciones injustas o redundantes.
171 Los valores Shapley proporcionan una forma intuitiva y sólida de interpretar y entender cómo las
172 características individuales afectan las decisiones del modelo, lo que los hace valiosos para explicar
173 modelos de aprendizaje automático complejos.

174 3.6. Ingeniería de atributos

175 Para dotar de mayor información a los algoritmos a la hora de predecir la variable objetivo, se crearon
176 nuevas variables a partir de las ya existentes. Fundamentalmente, se generaron variables llamadas
177 "lag" que toman el valor que obtuvo cierto país una x cantidad de años atrás. En este caso específico
178 se generaron lags para 1, 5 y 10 años anteriores. De esa manera, los algoritmos tienen algo más de
179 información sobre la tendencia temporal de las variables. Adicionalmente, se agregaron variables
180 binarias que informan sobre la región a la que pertenecen los países, especulando con que estas
181 variables pueden llegar a tener importancia si una región específica cuenta con muchos golpes de
182 estado en un momento determinado. Finalmente, se excluyeron todos los grupos de variables que
183 provengan de fuentes externas, con el objetivo de tener la certeza de contar con la mayoría de las
184 mismas en caso de querer repetir este experimento en años futuros; así como también se excluyeron
185 variables que no cuentan con información para ningún país en cierto punto de la serie (por ejemplo,
186 las variables históricas, que trabajan con datos anteriores al siglo xx).

187 3.7. Análisis Exploratorio de Datos

188 Como primera aproximación a la base de datos de Varieties of Democracy o V-Dem (Coppedge,
189 Gerring, Knutsen, Lindberg et al., 2024), pasaremos a explicar la manera en que se construye la misma.
190 Las variables centrales se obtienen a partir de encuestas suministradas a expertos sobre los distintos
191 países. Inicialmente, se busca que cada país cuente con al menos cinco expertos. Actualmente, la
192 institución cuenta con 22 expertos promedio por país y 7,1 expertos por combinación de variable
193 y país. Una vez obtenida las respuestas de los expertos, se pasa al proceso de agregación para así
194 conformar una base de datos donde cada fila corresponda a un país en un año específico. De esta
195 agregación obtienen diferentes versiones de la misma variable:

- 196 ■ Estimador del modelo (Variable sin sufijo): es la medida recomendada para su análisis.
197 Corresponde a obtener la mediana del valor de la variable entre los expertos, reescalado a
198 valores entre -5 a 5.
- 199 ■ Medidas de incertidumbre (*_codelow y *_codehigh): corresponden a un desvío estandar
200 por encima y por debajo del estimador del modelo. Usadas conjuntamente, construyen un
201 intervalo de confianza del 95 %.
- 202 ■ Escala original (*_osp): mediana de la variable, pero sin reescalar. Esta versión también
203 cuenta con sus medidas de incertidumbre correspondientes.
- 204 ■ Media simple (_mean): mediana de la variable, pero sin reescalar.
- 205 ■ Desvío estándar (_sd): desvío estándar de la variable.
- 206 ■ Media simple (_mean): media de la variable.
- 207 ■ Cantidades de expertos (_nr): cantidad de expertos que respondieron por país, año y variable.

208 Podemos mencionar que la base cuenta con 27734 filas y 4607 columnas. Como es una base de datos
209 de panel, se tiene información de 202 países durante 235 años. Las variables cuentan con un tipo de
210 codificación particular que permite identificar el origen de la variable. En primer lugar, el primer
211 prefijo es indicativo de si fue producido por V-Dem o no:

- 212 ■ v2: variables de V-Dem.
- 213 ■ v3: variables pertenecientes a la base V-Dem histórica.
- 214 ■ v2x_: Índices principales e índices componentes.
- 215 ■ v2x[indicador de dos letras]: Índices específicos de ciertas áreas (ver más abajo).
- 216 ■ e_: variables no generadas por V-Dem y variables V-Dem en versión ordinal.

217 El nombre de la variable también permite identificar la área temática a la que pertenece:

- 218 ■ ca: Espacio cívico y académico
- 219 ■ cl: Libertad civil
- 220 ■ cs: Sociedad civil
- 221 ■ dd: Democracia directa
- 222 ■ de: Demografía
- 223 ■ dl: Deliberación
- 224 ■ el: Elecciones
- 225 ■ ex: Ejecutivo
- 226 ■ exl: Legitimación
- 227 ■ ju: Poder judicial
- 228 ■ leg: Legislatura
- 229 ■ lg: Legislatura
- 230 ■ me: Medios de comunicación
- 231 ■ pe: Igualdad política
- 232 ■ ps: Partidos políticos
- 233 ■ sv: Soberanía
- 234 ■ st: Estado
- 235 ■ x: Índice (calculado a partir de variables que también se incluyen en la base de datos)
- 236 ■ zz: Cuestionario posterior a la encuesta
- 237 ■ ws: Encuesta de sociedad digital

238 A la base original obtenida desde la librería de V-Dem, se le realizaron los siguientes filtros: en primer
239 lugar, se removieron todas las variables que no sean las principales, es decir, que no cuenten con
240 sufijo. De esa manera, se busca reducir el tamaño de la base y así poder agregar nuevas columnas
241 mediante ingeniería de atributos. En segundo lugar, se filtraron los años superiores a 1950, para
242 adecuarnos al periodo utilizado en el artículo del FMI. De esa manera, la base filtrada cuenta con
243 12208 filas y 1460 columnas. Por último, se remueven todas las variables de fuentes externas (cuyo
244 agrupador comienza con 'e'), las variables pertenecientes a la base histórica (agrupador 'hist') y las
245 de la encuesta de sistema de partidos políticos; en parte debido a que provienen de fuentes ajenas
246 a V-Dem que pueden comprometer la completitud futura de los datos y en parte porque algunas de
247 estas variables cuentan con alta tasa de nulos.

248 Realizando un análisis generalizado de los distintos grupos de variables de la base de datos, podemos
249 aprehender ciertos patrones sobre la presencia de nulos: En primer lugar, observamos variables que,
250 anteriormente a un año puntual, no cuentan con información. En este ejemplo caen las variables
251 sobre gobernanza otorgadas por el banco mundial (e7), las preguntas pertenecientes a la encuesta
252 de sociedad digital (wsmcio), variables referentes a la libertad en medios digitales (wsmdmf), las
253 referentes a la polarización en medios online (wsmomp) y las referentes a clivajes sociales (wsmsc).

254 En segundo lugar, figuran casos contrarios, en donde a partir de determinado año la cantidad de datos
255 faltantes salta a la totalidad de los casos. En este grupo figuran las variables asociadas a instituciones
256 y eventos políticos (e13), cuya fuente es un artículo de Przeworski de 2013; las variables cuya
257 fuente es la base de datos polity V (e14); las variables sobre educación (aumentan los nulos en
258 algunas variables) (eb1); las variables sobre recursos naturales (eb5), cuya fuente tiene datos hasta
259 2006; las variables sobre infraestructura (eb6); y las relacionadas a conflictos (eb8). En general, esta
260 discontinuidad sucede debido a que la información de estas variables provienen de fuentes externas
261 no gestionadas por V-Dem, las cuales finalizaron su serie en un año puntual. Por último figuran los
262 grupos de variables asociados a la base de datos histórica de v-dem (las que comienzan con hist), lo
263 cual es lógico puesto que esta base busca tomar datos previos a 1900.

264 Consecuentemente, quitaremos del grupo de variables a utilizar aquellas que sean de fuentes externas,
265 ya que de esa manera podemos asegurarnos que contaremos con todas las variables para predecir

golpes de estado en años futuros. También quitamos las variables provenientes de fuentes históricas y de las encuestas de sociedad digital, ya que no cuentan con información para toda la serie.

Haciendo foco en la variable objetivo, es importante aclarar que en este trabajo no estamos contando la cantidad precisa de golpes de estado sucedidos en un período de tiempo, sino que simplemente relevamos si al menos un golpe de estado sucedió en un país y año determinado. Por lo tanto, si un país sufrió más de un golpe de estado en un año, el mismo será contabilizado una sola vez. Adicionalmente, en este trabajo también se consideran los golpes de estado que no fueron exitosos, es decir, que no lograron derrocar al gobierno en cuestión. De allí se desprende que países como Argentina, que en total ha tenido seis golpes de estado exitosos, figure con el doble de golpes en la figura 1.

Para realizar un paneo general de la variable objetivo, es decir, la presencia de golpes de estado a lo largo de los años, generamos un conteo y lo visualizamos en un planisferio. Destacamos que la mayor presencia de golpes se encuentra en el continente africano, en América del Sur y parte del Caribe, Medio Oriente y el Sudeste Asiático, con algunos casos de apenas un golpe en España, Rusia, Ucrania y Corea del Sur; así como dos y tres golpes en Grecia y Portugal, respectivamente.

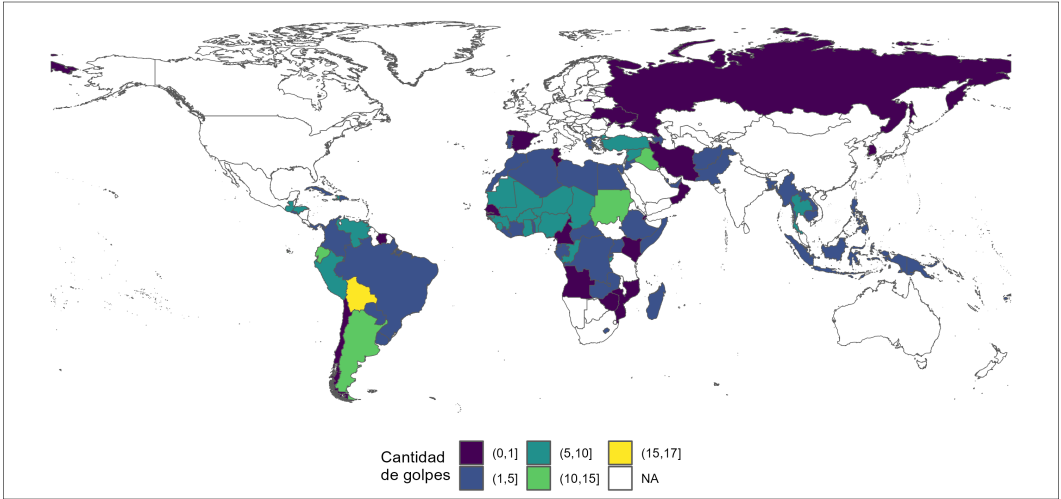


Figura 1: Golpes de estado período (1950-2023) Fuente: Powell y Thyne, 2011

Con mayor precisión, observamos que la región del Sahel se destaca con respecto a sus vecinos africanos. Los países en donde más golpes de estado se han producido son Bolivia (17), Sudán (14), Argentina (13), Ecuador (11), Iraq (11), Siria(11), Guatemala (10) y Tailandia (10).

Desagregando por década se observan algunos cambios, así como la persistencia en algunas regiones. La región del Sahel y varias naciones circundantes fueron persistentemente afectadas por golpes de estado desde los años 60. En América del Sur, en cambio, la presencia casi total de situaciones golpistas en la región se fue acotando a partir de los años 80 hasta finalmente desaparecer en el siglo xxi. Para observar con más detalle y discriminado por años y países se puede ver la figura 5.

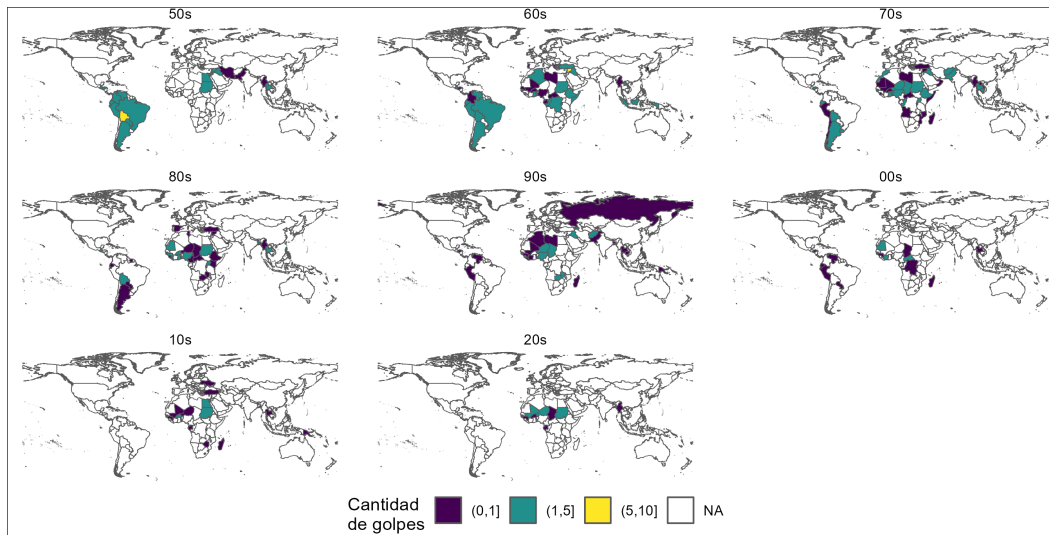


Figura 2: Conteo de golpes por década Fuente: Powell y Thyne, 2011

4. Resultados y discusión

4.1. Performance de los modelos

En primer lugar, se realizó la optimización bayesiana de ambos modelos según lo indicado en la metodología. En el caso de XGBoost se pudo realizar las 100 iteraciones sin mayores inconvenientes, tomando los valores óptimos de hiperparámetros para el entrenamiento final. Con respecto a Random Forest, en cambio, se alcanzaron 53 iteraciones, debido a que cada iteración consumía una gran cantidad de tiempo (en promedio una hora) y no se observaban mejoras significativas en el AUC. De las iteraciones generadas, se tomó los hiperparámetros del segundo mejor AUC, puesto que la diferencia con el ganador en el score era insignificante, pero el tiempo de cómputo era menos de la mitad.

Una vez seleccionado los mejores hiperparámetros, se procede a entrenar los modelos en el conjunto de entrenamiento final, el cual abarca los registros desde 1950 hasta 2019; así como a evaluar el desempeño del mismo en los años 2020, 2021 y 2022 para emular el trabajo realizado por el FMI.

Es importante destacar que existen dos enfoques para evaluar el modelo en los años de testeo: por un lado se pueden evaluar todos los años en su conjunto utilizando como datos de entrenamiento los registros hasta el año anterior del primer año de validación. Una opción alternativa es ir entrenando el modelo hasta el año anterior al de validación para cada año individualmente, de manera de poder utilizar todos los años anteriores y no perder performance. Para este trabajo utilizamos el primer enfoque, es decir que entrenamos los modelos hasta 2019 y los evaluamos en todos los años de evaluación a la vez, de manera de aprehender de manera geenal la importancia de cada variable en la predicción de la variable objetivo.

En el cuadro 1 observamos el desempeño de los modelos en los años de testeo. Por un lado figura el AUC individual de cada año por separado y por el otro observamos el AUC acumulada es decir, evaluando en ese año junto con los anteriores.

Año	XGBoost		Random Forest	
	AUC	AUC acumulada	AUC	AUC acumulada
2020	1.000000	1.000000	1.000000	1.000000
2021	0.750000	0.785714	0.830443	0.855718
2022	0.666667	0.750000	0.666667	0.799051

Cuadro 1: Área bajo la curva ROC por año puntual y acumulado (XGBoost y Random Forest)

Lo primero que podemos observar es que ambos modelos logran una performance perfecta para el año 2020, lo cual resulta esperable ya que cuentan con información del año inmediatamente anterior. También esperable, la performance decae en los años siguientes, lo cual impacta en el valor del AUC acumulada. Lo más destacable es que Random Forest logra una mejor performance que XGBoost en el resto de años, alcanzando un AUC de casi 0.8 y 0.75, respectivamente. Con esta información, se tomó la decisión de continuar el análisis de resultados con Random Forest.

Focalizando solamente en Random Forest, y puesto que los casos negativos (527 en los tres años de evaluación) fueron predichos de manera perfecta, podemos aprovechar para visualizar los casos positivos que son apenas diez casos, tantos los verdaderos positivos como los falsos negativos (Cuadro 2). En primer lugar, podemos notar que la predicción perfecta en el año 2020 se debe a que el modelo logró predecir correctamente el único golpe de ese año en Malí, en la región del Sahel. Después, en el año 2021 esta performance disminuye, al no lograr predecir los golpes de nuevo en Mali y en Níger, aunque sí predice golpes en Sudán, Guinea y Chad. A simple vista, no parece haber datos geográficos o históricos que permitan establecer por qué logra predecir algunos golpes y en otros no, en especial porque son países relativamente similares, de pocos años de independencia y dentro de la misma región. Adicionalmente, en este año también logró predecir el golpe de estado en Burma/Myanmar, una nación ubicada en una región alejada de África, en el sudeste asiático.

Año	País	¿Hubo golpe?	Predicción	Resultado
2020	Mali	Sí	Sí	Verdadero positivo
2021	Burma/Myanmar	Sí	Sí	Verdadero positivo
2021	Sudan	Sí	Sí	Verdadero positivo
2021	Guinea	Sí	Sí	Verdadero positivo
2021	Chad	Sí	Sí	Verdadero positivo
2022	Burkina Faso	Sí	Sí	Verdadero positivo
2021	Mali	Sí	No	Falso negativo
2021	Níger	Sí	No	Falso negativo
2022	Guinea-Bissau	Sí	No	Falso negativo
2022	Sao Tome and Principe	Sí	No	Falso negativo

Cuadro 2: Falsos negativos y verdaderos positivos (Random Forest)

Finalmente, el año 2022 expone la peor performance del modelo: si bien logra predecir un golpe de estado en Burkina Faso, falla al predecir golpes en Guinea-Bissau y Sao Tome y Principe, todos países en la misma región del continente africano. De manera general, podemos asociar esta baja en la performance a que el modelo deja de contar con información del año inmediatamente anterior al del conjunto de evaluación.

4.2. Análisis de variables

A continuación, pasaremos a evaluar la relevancia de las distintas variables del dataset para la predicción del modelo. De esa manera, podremos extraer elementos para determinar o reforzar los posibles causales de un golpe de estado en un territorio determinado. En primer lugar, utilizaremos la importancia de las variables según Random Forest, la cual se puede observar en la figura 3 (los nombres de las variables fueron traducidas y resumidas del libro de códigos de la base de datos para una vista amigable. Se puede verificar el nombre codificado y original de las variables en la tabla 3).

342 Las barras indican el porcentaje de importancia de las 10 variables con mayor peso. En total, estas
 343 diez variables representan alrededor del 50 % de la importancia. En general, todas las variables están
 344 relacionadas con la forma de gobierno, con la influencia de las fuerzas armadas en el mismo o con
 345 la misma variable objetivo en años anteriores. Entre el segundo y el cuarto lugar figuran variables
 346 que reflejan muy evidentemente una relación con la presencia de golpes de estado, como tener la
 347 legislatura cerrada o abortada o que el ejecutivo no sea más electo. El dato más interesante a destacar
 348 es que la variable con mayor importancia es la cantidad de días desde que comenzó el régimen. Se
 349 puede inferir de esto último que un régimen joven es más inestable y, por lo tanto, propensa a sufrir
 350 un nuevo cambio de régimen mediante un golpe.

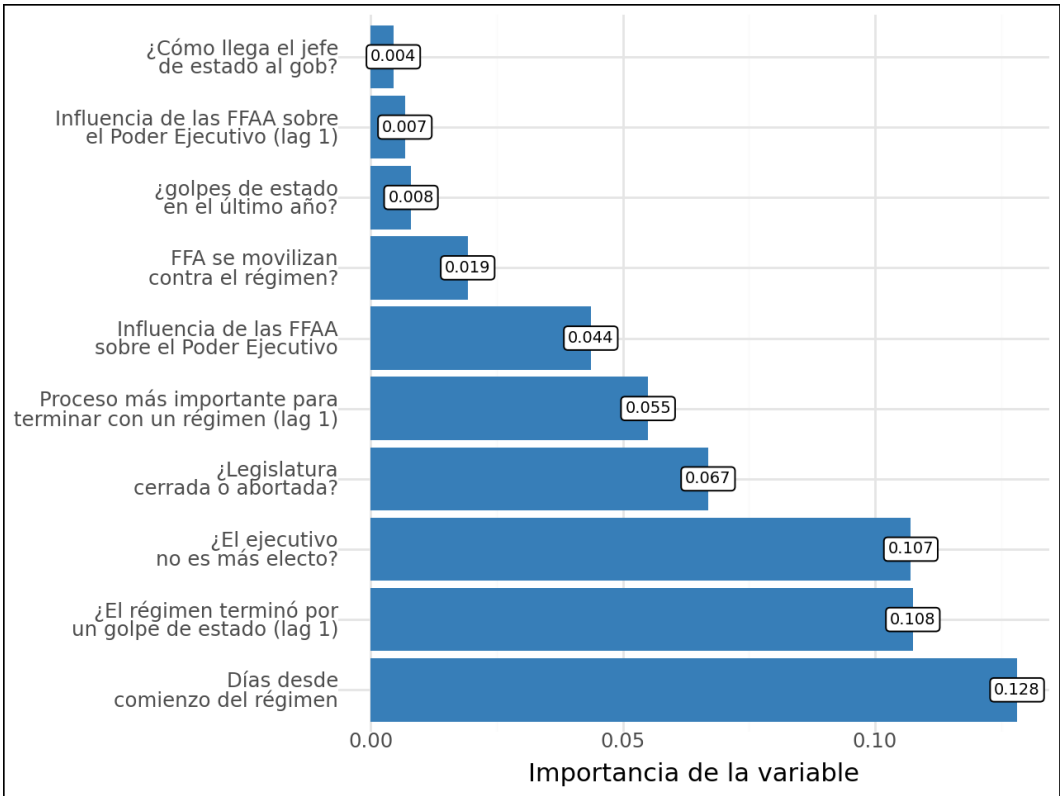


Figura 3: Importancia de las variables para predicción 2020-2022 (Random Forest)

351 Ahora incorporaremos los Shapley Values para identificar variables importantes a la hora de predecir
 352 la presencia de golpes de Estado, como se expone en la figura 4. En el eje Y figuran las primeras 11
 353 variables con mayor valor de Shapley y en el eje X figura el valor Shapley, visualizando la distribución
 354 de los casos en forma de violín y los outliers como puntos. Finalmente, el color de los violines y de
 355 los puntos indica el valor de la variable en cuestión.

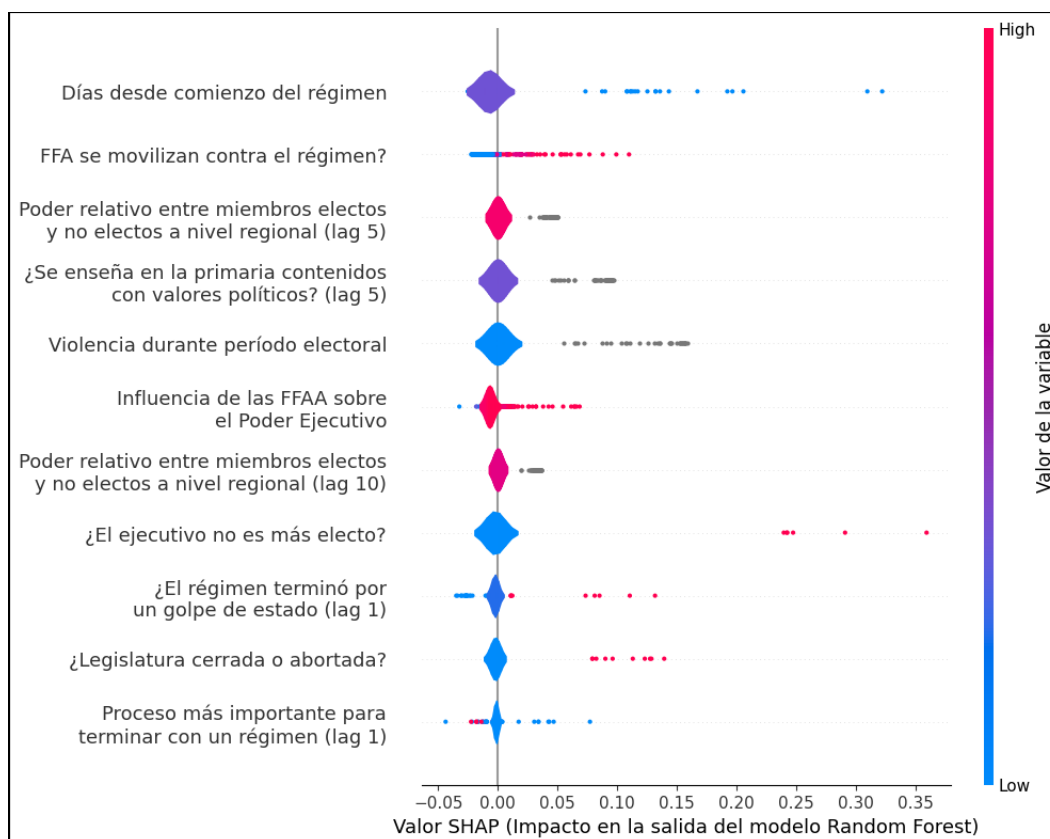


Figura 4: Shapley values para predicción 2020-2022 (Random Forest)

Si bien en este gráfico algunas variables figuran también en el gráfico de la importancia de las variables, podemos destacar algunas diferencias. Primero, figura la enseñanza de valores políticos en la escuela, en cuyo valores nulos tienen alto valor Shapley. También destacan los datos nulos en el poder relativo entre miembros electos y no electos a nivel regional, en su misma versión hace 10 años (lag 10) y en la presencia de violencia durante el período electoral.

Para comprender qué significan estos datos nulos, es de utilidad recurrir al libro de códigos de la base de datos. Por ejemplo, un valor faltante en el poder relativo entre oficiales electos y no electos significa que todos o casi todos de los funcionarios electos son subordinados de algún otro poder que no surgió de las urnas (a nivel regional).

Otros temas a tratar en los resultados en la próxima entrega: - Valores Shapley en cada año individual y/o en cada país - Análisis histórico de las variables destacadas

4.3. Discusiones

- Vinculación de resultados con estado del arte y marco teórico
- Comparación de performance y de variables importantes con el artículo del FMI
- Limitaciones

5. Conclusiones

- Resumen de los hallazgos principales
- Conclusiones generales y su relación con los objetivos del trabajo

374 - Recomendaciones para futuros trabajos

375 **6. Anexo**

376 **6.1. Código**

377 La totalidad del código y entregas en latex y PDF se encuentran en un repositorio abierto de Github
378 de José Saint Germain (Acceso al repositorio). En el mismo se describe la secuencia de códigos a
379 correr para la obtención de datos, la ingeniería de atributos, la optimización bayesiana, la corrida
380 final, el análisis exploratorio de datos y el análisis de resultados de los algoritmos.

381 **6.2. Golpes**

382 - Breve descripción de los 10 golpes que se buscaron predecir. Su contexto histórico, político y social.

383 **6.3. Gráficos y tablas adicionales**

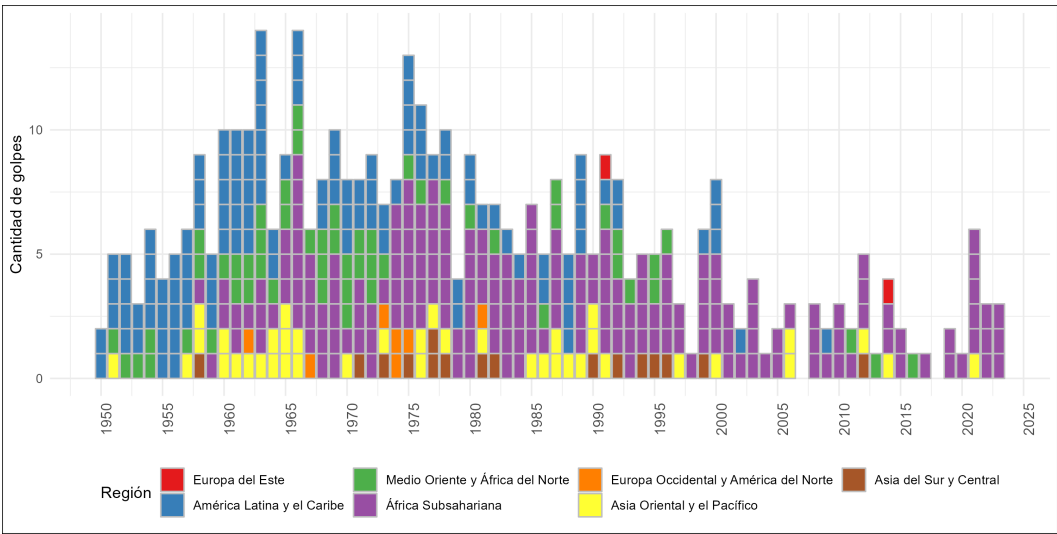


Figura 5: Conteo de golpes por año y región (Powell y Thyne, 2011)

Variable	Descripción
coup_lag_1	¿golpes de estado en el último año?
v2expathhs	¿Cómo llega el jefe de estado al gob?
v2regdur	Días desde comienzo del régimen
v2regoppgroupsact_5	FFA se movilizan contra el régimen?
v2elrgpwr_lag_5	Poder relativo entre miembros electos y no electos a nivel regional (lag 5)
v2edpoledprim_lag_5	¿Se enseña en la primaria contenidos con valores políticos? (lag 5)
v2elpeace	Violencia durante período electoral
v2x_ex_military	Influencia de las FFAA sobre el Poder Ejecutivo
v2x_ex_military_lag_1	Influencia de las FFAA sobre el Poder Ejecutivo (lag 1)
v2elrgpwr_lag_10	Poder relativo entre miembros electos y no electos a nivel regional (lag 10)
v2x_hosinter	¿El ejecutivo no es más electo?
v2regendtypems_0_lag_1	¿El régimen terminó por un golpe de estado? (lag 1)
v2xlg_leginter	¿Legislatura cerrada o abortada?
v2regendtype_lag_1	Proceso más importante para terminar con un régimen (lag 1)
coup_lag_1	¿Golpes de estado en el último año?

Cuadro 3: Nombre original de variables y su descripción

Referencias

- Huntington, S. P. (1068). *Political order in changing societies*. Yale university press.
- Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *American political science review*, 53(1), 69-105.
- Burman, P., Chow, E., & Nolan, D. (1994). A Cross-Validatory Method for Dependent Data. *Biometrika*, 81(2), 351-358. Consultado el 1 de mayo de 2024, desde <http://www.jstor.org/stable/2336965>
- Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics*, 99(1), 39-61. [https://doi.org/https://doi.org/10.1016/S0304-4076\(00\)00030-0](https://doi.org/https://doi.org/10.1016/S0304-4076(00)00030-0)
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <https://doi.org/http://doi.org/10.1023/A:1010933404324>
- Strumbelj, E., & Kononenko, I. (2010). An Efficient Explanation of Individual Classifications using Game Theory. *The Journal of Machine Learning Research*, 11, 1-18.
- Powell, J. M., & Thyne, C. L. (2011). Global Instances of Coups from 1950 to 2010: A New Dataset. *Journal of Peace Research*, 48(2), 249-259.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd interntional conference on knowledge discovery and data mining*, 785-794. <https://doi.org/https://doi.org/10.48550/arXiv.1603.02754>
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett, Eds.). 30. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- Cebotari, A., Chueca-Montuenga, E., Diallo, Y., Ma, Y., Turk, R., Xin, W., & Zavarce, H. (2024). *Political Fragility: Coups d'État and Their Drivers*. IMF Working Paper 24/34. <https://doi.org/https://doi.org/10.23696/mcwt-fr58>
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Marquardt, K. L., Medzihorsky, J., Pemstein, D., Fox, L., Gastaldi, L., Pernes, J., Rydén, O., von Römer, J., Tzelgov, E., Wang, Y.-t., & Wilson, S. (2024). "V-Dem Methodology v14" *Varieties of Democracy (V-Dem) Project* (Report). <https://v-dem.net/data/reference-documents/>
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., Angiolillo, F., Bernhard, M., Borella, C., Cornell, A., Fish, S. M., Fox, L., Gastaldi, L., Gjerlów, H., Glynn, A., God, A. G., Grahn, S., Hicken, A., Kinzelbach, K., ... Ziblatt, D. (2024). *V-Dem Dataset v14* Varieties of Democracy (V-Dem) Project* (Report). <https://doi.org/https://doi.org/10.23696/mcwt-fr58>