

# Taller de Tesis I – Entrega I – Año 2024

Estudiante: José Saint Germain

## Elección de conjunto de datos

La elección del conjunto de datos y del problema surge después de la lectura de un informe del Fondo Monetario Internacional (FMI) (Cebotari, y otros, 2024), el cual tienen como objetivo predecir la ocurrencia o no de un golpe de estado en un país y año determinado. Este artículo utiliza una gran variedad de fuentes, recogiendo variables sobre aspectos políticos, sociodemográficos, económicos y geográficos para obtener sus predicciones.

La idea de este trabajo de especialización es utilizar solamente una fuente de datos: la base de datos Varieties of Democracy (V-Dem) (Coppedge, y otros, 2024; Pemstein, y otros, 2024). La misma recoge la opinión de expertos de todos los países sobre distintos aspectos de la calidad institucional y democrática de un país en un año determinado. Únicamente se compartirá con el artículo del FMI la variable que indica la ocurrencia o no de un golpe de estado, proveniente de la base de datos construida por Powell y Thyne (2011).

## Elección de pregunta

En principio, el trabajo busca replicar la pregunta de investigación del informe: cuáles son los motivadores de la fragilidad política, concentrándose en los golpes de estado como sintomáticos de esa fragilidad (Cebotari, y otros, 2024). Este trabajo busca, entonces, predecir y analizar la misma ventana de tiempo, para los mismos países y utilizando la misma definición y categorización de golpes de estado (Powell & Thyne, 2011).

La motivación para replicar esta metodología con otra base de datos es relevar cuál es el poder que tienen variables exclusivamente institucionales y políticas para predecir la fragilidad institucional de un país, excluyendo variables más ligadas al devenir económico y demográfico del mismo. Como objetivo secundario, se aspira a alcanzar un poder predictivo por lo menos similar al expuesto en el artículo. Adicionalmente, mediante un análisis posterior a la predicción, se buscará identificar las principales variables que incrementan la probabilidad de la ocurrencia de un golpe de estado.

## Técnicas

Se replicarán las mismas técnicas de clasificación, de ajuste de hiperparámetros y de medida de performance utilizadas en el artículo, para así poder generar una comparación precisa. Los algoritmos que se utilizarán serán Random Forest (Breiman, 2001) y XGBoost (Chen & Guestrin, 2016). Para ajustar los hiperparámetros se utilizará la optimización bayesiana junto al método de block-time-series cross-validation. Por último, la métrica a optimizar y que se utilizará para comparar predicciones será el área bajo la curva (AUC).

Adicionalmente, se realizará un análisis exploratorio de datos de manera introductoria al trabajo y se buscará utilizar valores Shapley (Lundberg & Lee, 2017) para analizar los resultados de cada algoritmo.

## Bibliografía

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- 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.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.
- Coppedge, M., Gerring, J., Henrik Knutsen, C., Lindberg, S. I., Teorell, J., Altman, D., . . . Ziblatt, D. (2024). *V-Dem Dataset v14\* Varieties of Democracy (V-Dem) Project*. doi:<https://doi.org/10.23696/mcwt-fr58>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information processing systems*, 30.
- Pemstein, D., Marquadt, K. L., Tzelgov, E., Wang, Y.-t., Medzihorsky, J., Krusell, J., . . . von Römer, J. (2024). *The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data*. University of Gothenburg: Varieties of Democracy Institute, V-Dem Working Paper No. 21 9th edition.
- 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.
- Strumbelj, E. (2010). An efficient explanation of individual classification using game theory. *The Journal of Machine Learning Research*, 11, 1-18.