Logo

Description automatically generated

[PROYECTO CAPSTONE]

[Uso de Manifesto Project para identificar patrones comunes en Programas de Gobierno para enfrentar la Corrupción]

POR: J. MATÍAS BUNSTER RABY

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del Desarrollo para optar al grado académico de Magíster en Data Science

Profesor guía:

DRA. MARÍA PAZ REVEAU MORALES

[Diciembre 2024]

[SANTIAGO-CHILE]

AGRADECIMIENTO

[Escriba aquí la dedicatoria para su trabajo, recuerde que ésta es opcional]

[Escriba aquí los agradecimientos a las personas que colaboraron con usted en la realización de este trabajo]

Tabla de contenido

[Resumen 1](#_Toc55504658)

[1. Introducción 2](#_Toc55504659)

[2. Trabajo Relacionado 4](#_Toc55504660)

[3. Hipótesis y Objetivos 5](#_Toc55504661)

[4. Datos y Metodología 6](#_Toc55504662)

[4.1. Datos 6](#_Toc55504663)

[4.2. Metodología 6](#_Toc55504664)

[5. Resultados 7](#_Toc55504665)

[6. Conclusión y Trabajo Futuro 8](#_Toc55504666)

[Bibliografía 9](#_Toc55504667)

# Resumen

Escriba un resumen del documento

# Introducción

Este informe responde a los requerimientos de la Universidad del Desarrollo para obtener el grado de Magíster en Data Science a través de un “Proyecto Capstone”.  
El objetivo principal del estudio es identificar patrones comunes entre los programas de gobierno de candidatos presidenciales o equivalentes (primer ministro, jefe de gobierno, etc.), ya sea que representen a un partido político o a una coalición. El análisis se focaliza en torno a una problemática que afecta transversalmente al mundo: la Corrupción.

Los programas de gobierno son, en esencia, una declaración de intenciones. Una vez que un candidato triunfa en una elección, la implementación del programa enfrenta diversos factores, como la realidad económica y política del país, y la necesidad de negociación legislativa. En este estudio, el análisis se centra exclusivamente en los programas de gobierno, dejando como posible extensión futura el análisis de su implementación.

La Corrupción es un fenómeno global presente en las sociedades a lo largo de la historia. En las últimas décadas, se ha señalado un aumento en su penetración en diversas instituciones, lo que ha llevado a los gobiernos a implementar políticas específicas para combatirla; sin embargo, no se sabe si las prioridades políticas en torno a ella permiten agrupar países en patrones comunes. Considerando esta problemática, este trabajo analiza los programas de gobierno de un conjunto amplio de países en un período de tiempo determinado, en torno al fenómeno de la Corrupción, con el objetivo de identificar patrones comunes y agrupar a los países en comunidades con características similares, tomando como base la estructura inherente de los datos.

Este es un primer paso para generar conocimiento estructurado sobre este y otros fenómenos. En futuros trabajos, la combinación de estos resultados con datos empíricos sobre la implementación y resultados de políticas contra la Corrupción podría servir como una base para el diseño de políticas públicas comparadas.

Los datos utilizados en este estudio provienen de la plataforma Manifesto Project, que recopila programas de gobierno desde 1920 hasta 2023 de una gran mayoría de países, facilitando análisis de preferencias y tendencias políticas. Se aplicó un enfoque de clustering y análisis de comunidades.

# Manifesto Project

(\*) Extraído de la página web de Manifesto Project (<https://manifesto-project.wzb.eu/information/documents/information>)

El ‘Manifesto Project’ (Proyecto Manifesto), analiza los manifiestos electorales de los partidos políticos para estudiar sus preferencias políticas. Desde octubre de 2009, el Proyecto Manifiesto ha sido financiado con una subvención a largo plazo de la Fundación Alemana para la Ciencia (DFG) como MARPOR (Manifiesto de Investigación sobre Representación Política). MARPOR continúa el trabajo del Grupo de Investigación Manifiesto (MRG 1979-1989) y el Proyecto Manifiestos Comparados (CMP 1989-2009). En 2003, el proyecto recibió el premio de la Asociación Estadounidense de Ciencias Políticas (APSA) al mejor conjunto de datos en política comparada.

MARPOR aborda la recopilación y el análisis comparativo de contenido de los manifiestos de los partidos con el apoyo de codificadores de más de 50 países diferentes. El equipo del proyecto coordina la generación y el análisis de los datos. El conjunto de datos del Proyecto Manifiesto proporcionado para el análisis de las preferencias políticas cubre más de 1.000 partidos desde 1945 hasta la actualidad en más de 50 países en los cinco continentes. El conjunto de datos principal del proyecto se actualiza dos veces al año. Proporciona acceso instantáneo a textos de manifiestos y datos analíticos de contenido y ofrece formas accesibles de explorar y visualizar fácilmente los datos y el corpus de texto.

El proyecto MARPOR tiene como objetivo analizar sustancialmente el papel de los partidos en diferentes etapas del proceso político y examina específicamente la calidad de la representación programática. Estudia la oferta programática de los partidos, la relación entre partidos y votantes, el papel de los partidos en el parlamento y la traducción de los programas de los partidos en resultados políticos.

Referencia: <https://manifesto-project.wzb.eu/information/documents/information>

# Trabajo Relacionado

3.1 Manifesto Project Dataset/ Lehmann, Pola / Franzmann, Simon / Al-Gaddooa, Denise / Burst, Tobias / Ivanusch, Christoph / Regel, Sven / Riethmüller, Felicia / Volkens, Andrea / Weßels, Bernhard / Zehnter, Lisa (2024): The Manifesto Data Collection. Manifesto Project (MRG / CMP/ MARPOR). Version 2024a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (IfDem). <https://doi.org/10.25522/> manifesto.mpds.2024a

3.2 Manifesto Project Dataset - Codebook/ Lehmann, Pola / Franzmann, Simon / Al-Gaddooa, Denise / Burst, Tobias / Ivanusch, Christoph / Regel, Sven / Riethmüller, Felicia / Volkens, Andrea /Weßels, Bernhard / Zehnter, Lisa (2024): The Manifesto Project Dataset - Codebook. Manifesto Project (MRG / CMP / MARPOR). Version 2024a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (IfDem).

3.3 Manifesto Project Dataset – Coding Instructions (5th re-revised edition)/ [manifesto-communication@wzb.eu/](mailto:manifesto-communication@wzb.eu/) website: <https://manifesto-project.wzb.eu/>

3.4 Manifesto-Project Dataset/ List of Political Parties/ Version 2024a from March 2024/ Lehmann, Pola / Franzmann, Simon / Al-Gaddooa, Denise / Burst, Tobias / Ivanusch, Christoph / Regel, Sven / Riethmüller, Felicia / Volkens, Andrea / Weßels, Bernhard / Zehnter, Lisa (2024): The Manifesto Project Party List - Codebook. Version 2024a. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (IfDem). [manifesto-communication@wzb.eu/](mailto:manifesto-communication@wzb.eu/) website: <https://manifesto-project.wzb.eu/>

# Hipótesis y Objetivos

4.1 Hipótesis

Los programas de gobierno reflejan patrones comunes que permiten agrupar países en comunidades según sus prioridades temáticas, mostrando diferencias consistentes entre comunidades en términos de énfasis político.

4.2 Objetivo General

Identificar patrones comunes en los programas de gobierno relacionados con la corrupción que permitan agrupar países en comunidades, analizando las diferencias y similitudes temáticas entre las comunidades a lo largo del tiempo.

4.3 Objetivos Específicos

4.3.1 Explorar las variables temáticas relacionadas con la Corrupción: Analizar las variables del Manifesto Project relacionadas con la corrupción (Political Corruption, Law and Order, Traditional Morality) y cómo estas se distribuyen en los programas de gobierno de los países.

4.3.2 Detectar comunidades de países: Aplicar técnicas de clustering para identificar comunidades de países en base a sus programas de gobierno y analizar su composición en diferentes períodos de tiempo.

4.3.3 Calcular y visualizar las diferencias promedio en las variables relacionadas con la corrupción y las áreas temáticas principales entre las comunidades detectadas.

4.3.4 Visualizar coincidencias entre países: Construir heatmaps y redes de coincidencias para entender cuántas veces los países comparten un mismo clúster en diferentes períodos

4.3.5 Caracterizar las comunidades en base a sus prioridades temáticas: Explorar si las comunidades detectadas reflejan diferencias estructurales relacionadas con características políticas, económicas o sociales.

# Datos y Metodología

## Datos

Los datos utilizados corresponde a los del dataset de Manifesto Project disponible a Marzo 2024. El dataset es el resultado de un trabajo de etiquetado de los programas de gobierno de partidos o coaliciones que postulan al cargo político principal de los países o a la cámara baja. En el caso de elecciones presidenciales, el criterio ha sido etiquetar aquellos programas que han obtenido el 5% de los votos o más en la primera vuelta. En el caso de postulaciones a la cámara baja, se han etiquetado aquellos programas que han conseguido 2 o más asientos en el caso de países de Europa central y del este, y América del Sur, y uno o más asientos en el caso de Australia, Japón, Nueva Zelanda, Norte América, Corea del Sur y Europa del oeste). También incluye los programas de partidos que fueron actores relevantes en el pasado pero que ya no cumplen con el criterio de mínimo de votos.

En términos generales el dataset reúne información de:

* 67 países (principalmente desde 1945 en adelante ó desde su primera elección democrática),
* 1.387 partidos políticos (según el criterio indicado)
* 856 procesos eleccionarios
* 5.151 programas de gobierno obtenidos de fuentes públicas.

5.1.1 Etiquetado

Cada programa de gobierno ha sido procesado por un equipo entrenado de codificadores (‘etiquetadores’ entrenados en un conjunto de reglas de etiquetado), de manera de asegurar la compatibilidad y la comparabilidad entre programas.

Cada programa es separado en las frases (“quasi sentences” según las Coding Instructions) que lo componen, y cada frase es etiquetada con una y sólo una etiqueta.

La pregunta central en el proceso de etiquetado es cuál es el mensaje que el partido ó el candidato están tratando de transmitir a los votantes, y cuáles son los problemas que considera importantes.

Las etiquetas posibles son respecto de 56 categorías que a su vez se organizan en 7 “Dominios”. Existen sub categorías para análisis especiales pero éstas son agregadas en las categorías principales.

Una vez que un programa ha sido etiquetado, se cuentan las etiquetas que recibió cada categoría y se divide por el número total de etiquetas. Este es el valor que queda registrado para cada programa, en cada categoría. Dada esta construcción, la suma de los valores de cada categoría de un programa cualquiera es 100% .

5.1.2 Países

A group of countries/regions with names

Description automatically generated

5.1.3 Categorías y Dominios

Nota: en lo sucesivo, cuando se mencione a una “variable” se debe entender lo mismo que una “categoría”.

A table with text and numbers

Description automatically generated

A blue and white box with black text

Description automatically generated

A blue and white box with black text

Description automatically generated

A screenshot of a table

Description automatically generated

A screenshot of a table

Description automatically generated

A screenshot of a table

Description automatically generated

A blue and white text with black text

Description automatically generated

5.1.4 Temporalidad

El registro más antiguo del dataset corresponde a Estados Unidos en 1920, siendo el único para el cual hubo datos ese año. El registro más reciente corresponde a Montenegro en Junio 2023.

Los países se fueron incorporando paulatinamente. De manera acumulada, en 1960 hay información de 25 países, en 1980 de 28, en 1990 de 45 y en 2003 de 67 países, según se puede apreciar en este gráfico:

Gráfico 1: Número de países por año presentes en la dataset

A graph with blue lines and numbers

Description automatically generated

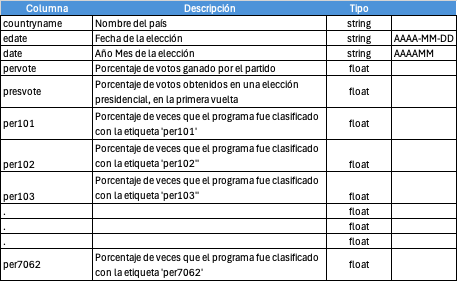
5.1.5 Columnas del dataset de Manifesto Project

El dataset de Manifesto Project contiene 5.151 registros con información de programas de gobierno según las siguientes columnas:

A document with text on it

Description automatically generated

Para el objetivo de este trabajo la columnas que se utilizó fue:



## Metodología

El método utilizado fue el siguiente:

5.2.1 Conocer el dataset y validar con la realidad conocida en Chile

5.2.2 Identificación de las variables para el análisis de programas con el foco en la Corrupción

Para analizar el fenómeno de la Corrupción, se seleccionó las siguientes variables:

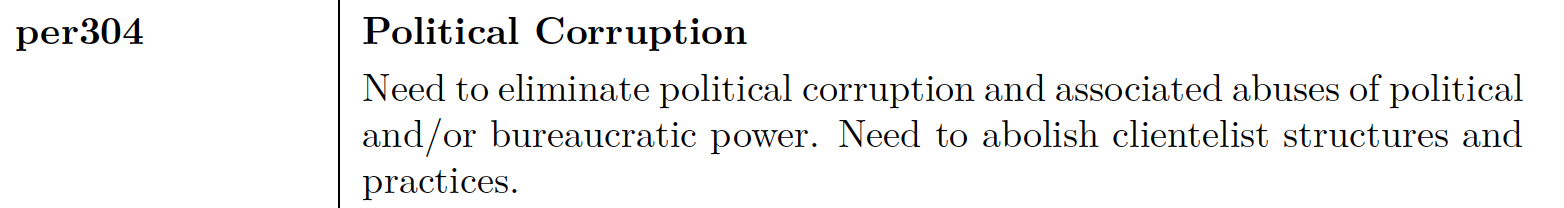
per304: Political Corruption

per603: Traditional Morality: Positive

per604: Traditional Morality: Negative

per605: Law and Order: Positive

La descripción de cada una de estas variables según el Manifesto Project es la siguiente:

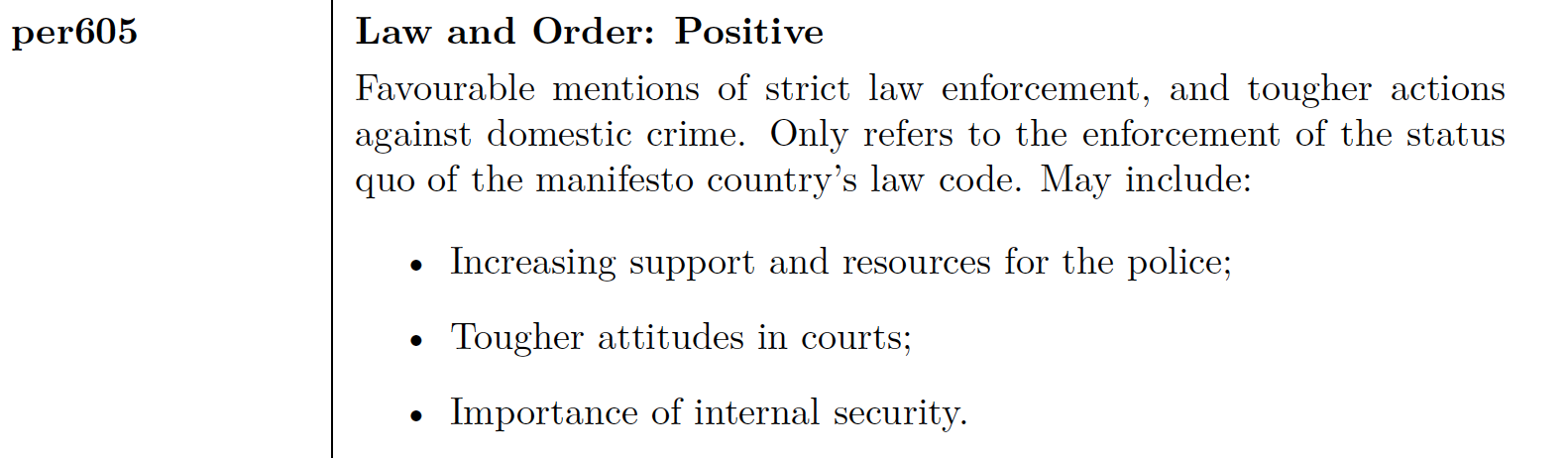


A close-up of a paper

Description automatically generated

A close-up of a text

Description automatically generated



5.2.2 Análisis exploratorio del dataset de las variables seleccionadas

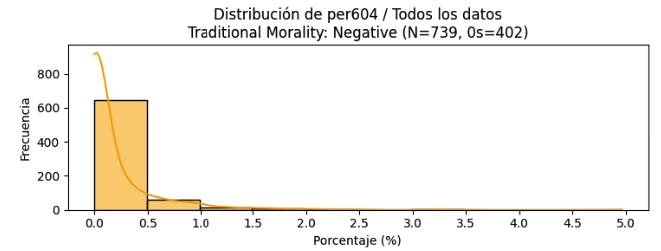
5.2.2.1 Distribución de valores en el dataset completo

A graph with numbers and lines

Description automatically generated

A graph with numbers and a line

Description automatically generated



A graph with numbers and a line

Description automatically generated

Como se puede observar, los gráficos están cargados a la izquierda, en torno al cero.

Esta distribución toma esta forma principalmente por el hecho de que la suma de los valores en cada fila -por diseño del dataset-, es 100%; a esto se le denomina “datos composicionales” y su distribución se llama distribución Dirichlet. En una distribución Dirichlet el análisis de la contribución de una de las variables sigue una distribución “Beta”, cuya forma es la siguiente: (\*\*) poner aquí una referencia a la función Dirichlet.

A graph with a blue line

Description automatically generated

5.2.1.1 Variables seleccionadas para el análisis del problema de la Corrupción

5.2.3 Normalización

# Resultados

Incluir acá análisis exploratorios, modelos, etc.

# Conclusiones

Incluir aquí las conclusiones el trabajo futuro y limitaciones.

# Bibliografía

# Anexo

1. Manifesto Project en detalle

El Proyecto Manifiesto analiza los manifiestos electorales de los partidos para estudiar sus preferencias políticas. Desde octubre de 2009, el Proyecto Manifiesto ha sido financiado con una subvención a largo plazo de la Fundación Alemana para la Ciencia (DFG) como MARPOR (Manifiesto de Investigación sobre Representación Política). MARPOR continúa el trabajo del Grupo de Investigación Manifiesto (MRG 1979-1989) y el Proyecto Manifiestos Comparados (CMP 1989-2009). En 2003, el proyecto recibió el premio de la Asociación Estadounidense de Ciencias Políticas (APSA) al mejor conjunto de datos en política comparada.

MARPOR aborda la recopilación y el análisis comparativo de contenido de los manifiestos de los partidos con el apoyo de codificadores de más de 50 países diferentes. El equipo del proyecto coordina la generación y el análisis de los datos. El conjunto de datos del Proyecto Manifiesto proporcionado para el análisis de las preferencias políticas cubre más de 1000 partidos desde 1945 hasta la actualidad en más de 50 países en los cinco continentes. El conjunto de datos principal del proyecto se actualiza dos veces al año. Proporciona acceso instantáneo a textos de manifiestos y datos analíticos de contenido y ofrece formas accesibles de explorar y visualizar fácilmente los datos y el corpus de texto.

El proyecto MARPOR tiene como objetivo analizar sustancialmente el papel de los partidos en diferentes etapas del proceso político y examina específicamente la calidad de la representación programática. Estudia la oferta programática de los partidos, la relación entre partidos y votantes, el papel de los partidos en el parlamento y la traducción de los programas de los partidos en resultados políticos.

Referencia: <https://manifesto-project.wzb.eu/information/documents/information>