



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Departamento de Ingeniería Comercial

**Data y Democracia: Estudio por Clustering de las Propuestas
Constitucionales 2022 y 2023 con un enfoque comparativo
Iberoamericano**

Memoria de Título presentada por

María Olga Egaña Muñoz

como requisito parcial para optar al título de la carrera de

Ingeniería Comercial

Profesor Guía

Dr. Pablo Nolasco Isla Madariaga

Profesor Correferente

Dra. Jocelyn Tapia Stefanoni

SANTIAGO, JUNIO, 2024

Agradecimientos

Con profunda gratitud, dedico esta tesis a aquellos que han brindado su apoyo incondicional a lo largo de este viaje académico.

A mi familia, les agradezco por las conversaciones interminables, donde compartieron sus experiencias, motivándome a esforzarme por alcanzar mis objetivos. Además, quiero reconocer su paciencia y comprensión, sobretodo en aquellos momentos de duda. De verdad, gracias por todo el amor, por siempre estar a mi lado y creer en mí.

Este logro también se lo dedico a mis amigos del CMAD, quienes durante meses me acompañaron en este proceso. Agradezco por su tiempo, que me hayan enseñado sin esperar recompensa, por acompañarme esas largas tardes en la universidad, mientras programábamos y por su apoyo incondicional incluso en mis peores momentos.

Además, quiero expresar mi profundo agradecimiento a mis profesores y a la Universidad, por brindarme la oportunidad de participar en esta investigación. Su dedicación y mentoría me permitieron sumergirme en un mundo nuevo, especialmente en el ámbito de la inteligencia artificial y el machine learning, proporcionándome las herramientas necesarias para fortalecer mi futuro laboral.

A todos y cada uno de ustedes, gracias de todo corazón por formar parte de este proceso y por hacerlo aún más memorable con su presencia y apoyo constante.

DATA Y DEMOCRACIA: ESTUDIO POR CLUSTERING DE LAS PROPUESTAS CONSTITUCIONALES 2022 Y 2023 CON UN ENFOQUE COMPARATIVO IBEROAMERICANO

María Olga Egaña Muñoz¹, Pablo Nolasco Isla Madariaga¹, Jocelyn Tapia Stefanoni¹

¹ Universidad Técnica Federico Santa María

Resumen

La presente tesis, titulada "Data y Democracia: Estudio por Clustering de las Propuestas Constitucionales 2022 y 2023 con un Enfoque Comparativo Iberoamericano", se desarrolla en el contexto del proceso constitucional chileno iniciado tras las movilizaciones sociales del año 2019. Este estudio tiene como objetivo principal realizar una comparativa semántica, utilizando técnicas de clustering, entre las propuestas constitucionales chilenas de los años 2022 y 2023, junto a las constituciones de 22 naciones iberoamericanas.

La metodología incluye la recolección y preparación de datos, limpieza y transformación de los textos constitucionales para generar vectores representativos y el uso de técnicas de clustering como K-means y dendrograma jerárquico. Se aborda la problemática de datos faltantes y se realizan visualizaciones como nubes de palabras y análisis de keyness para extraer información relevante.

Los resultados revelan que, a través de las técnicas de agrupamiento utilizadas, la propuesta chile del año 2023 se agrupa consistentemente con la constitución chilena de 1980 y la reforma cosntitucional de 2005, indicando una continuidad semántica significativa.

Las conclusiones de este estudio subrayan la identificación de patrones y similitudes entre las propuestas constitucionales chilenas y constituciones iberoamericanas, proporcionando una visión integral de las dinámicas políticas y legales que impulsan las reformas constitucionales en la región. Este análisis comparativo enriquece la comprensión de los textos constitucionales en el contexto contemporáneo de la información y la tecnología.

Palabras claves: Clustering, Constitución, Chile, Propuestas.

Índice

1. Introducción	4
2. Problemática	5
2.1. Estallido Social	5
2.2. Propuesta Constitucional 2022	5
2.3. Propuesta Constitucional 2023	8
3. Marco Teórico	9
3.1. Constitución de la República	9
3.2. Inteligencia Artificial	9
3.2.1. Natural Language Processing	10
3.3. Machine Learning	10
3.3.1. Modelos Supervisados	11
3.3.2. Modelos No Supervisados	11
3.3.3. Métodos de Agrupación	11
3.3.4. Métodos Jerárquico	12
3.3.5. Métodos No Jerárquico	13
3.4. Lenguaje Programación R	14
3.4.1. Quanteda: Quantitative Analysis of Textual Data	15
3.5. Enfoque Comparativo Iberoamericano	15
4. Objetivos	16
4.1. Objetivo General	16
4.2. Objetivos Específicos	16
4.3. Alcance del Estudio	16
5. Caso de Estudio	17
5.1. Metodología de la Investigación	17
5.1.1. Recolección de los Datos	17
5.1.2. Preparación de los Datos	18
5.1.3. Limpieza de Data	19
5.1.4. Document Feature Matrix	22
5.1.5. Determinación de Codo	22
5.1.6. Clustering K-Means	24
5.1.7. Dendrograma Jerárquico	26
5.2. Resultados	27
5.3. Clustering K-Means con k=4	27
5.4. Clustering K-Means con k=5	29
5.5. Dendrograma Jerárquico	30

6. Conclusiones	32
6.1. Conclusiones del caso	32
6.2. Conclusiones finales	33
7. Referencias bibliográficas	34

1. Introducción

A finales del año 2019, Chile presencié una serie de movilizaciones sociales y revueltas urbanas. Este evento marcó el inicio de un proceso trascendental en la historia constitucional del país, dando paso a la redacción de una nueva Carta Magna, una tarea monumental que respondía al clamor de la sociedad por un cambio estructural.

En respuesta a este llamado, se estableció la Convención Constitucional como el órgano encargado de elaborar la propuesta constitucional. Este proceso culminó con el rechazo de la propuesta en un plebiscito nacional. Sin embargo, el 12 de diciembre de 2022, el Acuerdo por Chile, anunciado por las autoridades legislativas, marcó el inicio de un nuevo proceso constitucional. Este acuerdo buscaba la redacción de una nueva Carta Fundamental para el país, reafirmando el compromiso de Chile con la construcción de una sociedad basada en principios democráticos y justicia social.

Es en este contexto de transformación política y jurídica que la presente tesis se sumerge en un análisis comparativo de las propuestas constitucionales presentadas durante los años 2022 y 2023 en Chile. Este estudio, en comparación con los textos constitucionales de 22 países de Iberoamérica, adopta una metodología central basada en técnicas de clustering. Esta aproximación permite la identificación de patrones y similitudes semánticas entre las diversas propuestas a nivel regional.

La relevancia de este enfoque comparativo iberoamericano radica en su capacidad para arrojar luz sobre las tendencias comunes y las diferencias presentes en las iniciativas constitucionales de distintos países de la región. Al ofrecer una visión integral, se busca apreciar tanto las convergencias como las particularidades que caracterizan estos documentos constitucionales en el contexto iberoamericano contemporáneo.

La amalgama de eventos que ha marcado la evolución constitucional en Chile, desde las movilizaciones sociales hasta los procesos de redacción y rechazo de propuestas, sitúa a esta investigación en el epicentro de una coyuntura histórica. El análisis comparativo y la aplicación de técnicas avanzadas, como el clustering y el procesamiento de lenguaje natural, se erigen como herramientas fundamentales para desentrañar la complejidad de estas propuestas constitucionales y, por ende, para ofrecer una perspectiva esclarecedora que contribuya al diálogo informado y reflexivo sobre el futuro político y jurídico de Chile.

2. Problemática

2.1. Estallido Social

El día 18 de octubre de 2019, en Chile se produjo un estallido social de gran magnitud, en el que diversos sectores de la sociedad se movilizaron para protestar y denunciar un conjunto de problemas estructurales que venían acumulando un profundo malestar en la ciudadanía (Avendaño, 2019). La respuesta institucional de la clase política al estallido y a la marcha del 25 del octubre de 2019 (la más grande de la historia del país) fue el ‘Acuerdo por la Paz Social y la Nueva Constitución’, que consistió en la realización de un plebiscito para decidir sobre el cambio de la Carta Fundamental (Avendaño, Osorio, 2021).

El día 15 de noviembre del 2019, diversos representantes, tanto del oficialismo como de la oposición, firmaron el “Acuerdo por la paz y la Nueva Constitución” que buscó agilizar el proceso de elaboración de una nueva Carta Constitucional. En este se estableció que el 26 de abril del 2020 los chilenos podrían participar de un plebiscito para votar por el “apruebo” o “rechazo” de la realización de una Asamblea Constituyente, así como – en caso de ser aprobada – por la composición de la misma entre una “convención mixta”, conformada por un 70 % de parlamentarios y un 30 % de civiles, o por una “convención constitucional”, donde el 100 % de los participantes deberán ser electos democráticamente (Urrejola, 2019).

2.2. Propuesta Constitucional 2022

Posterior al establecimiento del llamado "Acuerdo por la Paz Social y la Nueva Constitución", siguió un camino de reformas constitucionales aprobadas por el Congreso Nacional para establecer un itinerario cosntitucional y para regular la equidad de género y la reserva de escaños para los pueblos originarios (ChileAtiende, 2022).



Figura 1: Hitos del Proceso Constitucional 2022. Extraído desde Biblioteca del Congreso Nacional (BCN).

El día 25 de octubre de 2020 se sometió la decisión de la ciudadanía, la que votó por aprobar la redacción de una nueva Constitución y que esta se escribiera en una Convención integrada completamente por personas elegidas especialmente para el proceso. Luego, el 15 y 16 de mayo de 2021, se llevaron a cabo las elecciones de los 155 convencionales constituyentes, pertenecientes al organismo Convención Constituyente, quienes tenían por único objetivo redactar la nueva Constitución, no afectando las competencias y atribuciones de los demás órganos y poderes del Estado, disolviéndose una vez completa la tarea que le fue encargada (Chile Atiende, 2022).

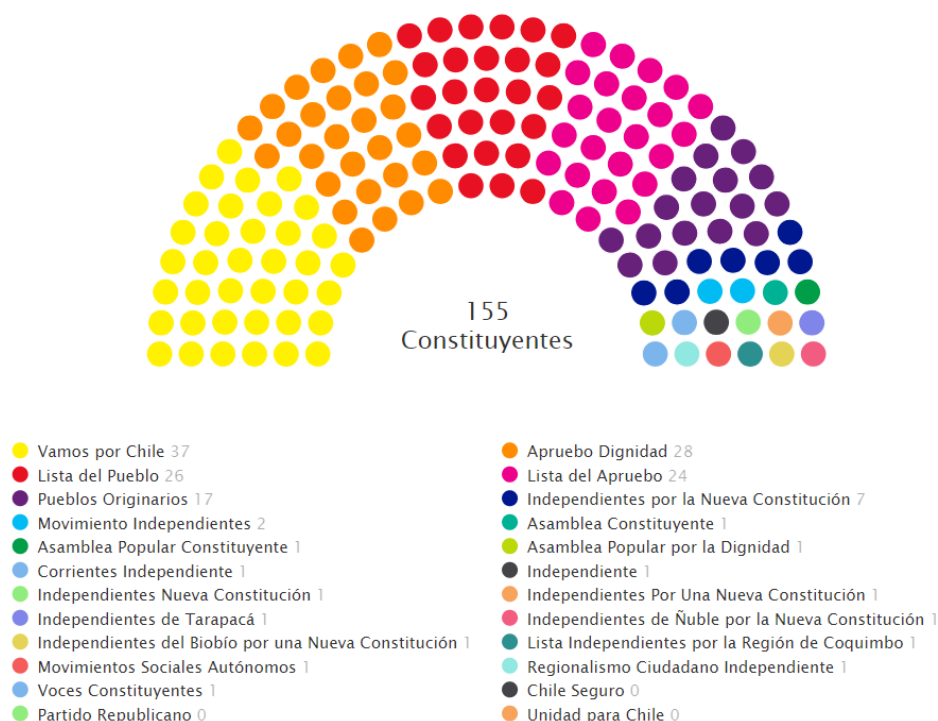


Figura 2: Convencionales Constituyentes proceso constitucional 2022. Extraído desde Biblioteca del Congreso Nacional (BCN).

Finalmente el día 04 de septiembre del año 2022 se llevó a cabo el Plebiscito de salida, donde en las urnas el resultado fue apabullante, más una altísima participación ciudadana gracias al sufragio universal obligatorio (conforme a las cifras dadas por el Servicio Electoral [SERVEL] votaron en total 13.021.063 personas [85.8 % del padrón electoral]: considerando los votos válidamente emitidos, éstos se inclinaron en un 61,86 % por el “rechazo” y 38,14 % por el “apruebo”). La opción “rechazo” ganó en 338 comunas y “apruebo” en 8 comunas. La participación ciudadana le dio voz y voto (sufragio universal obligatorio) a una “mayoría silenciosa”. Esa “mayoría silenciosa” encarna un cierto “sentido común” de la opción “rechazo”: deseaba cambios, pero no tantos, y en especial es muy crítica de los cambios contraculturales consignados en la propuesta de nueva Constitución. (Zúñiga, 2022)

2.3. Propuesta Constitucional 2023

Durante 2023, se desarrolló la apertura de un nuevo proceso constitucional, donde en esta oportunidad se crearon tres organismos: Consejo Constitucional, una Comisión Experta y un Comité Técnico de Adminisibilidad, los cuales participaron en la elaboración de la propuesta plebiscitada.

Este segundo proceso constituyente, tuvo una serie de pasos para llegar a su redacción final: Primero, una Comisión de Expertos, compuesta por 24 miembros elegidos por el Congreso, que entregó un anteproyecto que sirvió como base para el trabajo del Consejo Constitucional (Maldonado, 2023). Una vez aprobado el anteproyecto por todos los integrantes de la Comisión, el Consejo Constitucional, formado por 51 miembros, debatió y aprobó la propuesta de nueva Constitución (BCN, 2023). Posteriormente, el proyecto final fue entregado al Presidente de la República, marcando la conclusión de la fase de redacción.

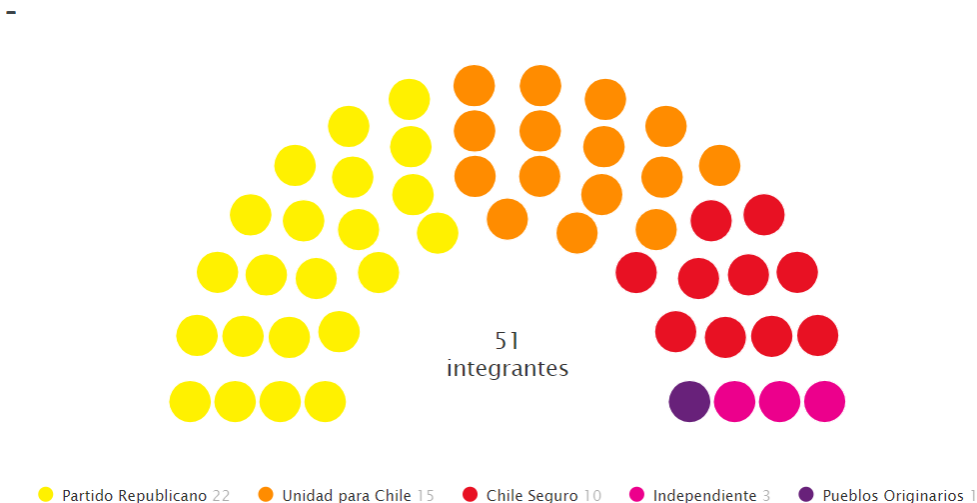


Figura 3: Consejo Constitucional proceso constitucional 2023. Extraído desde Biblioteca del Congreso Nacional (BCN).

Finalmente, el día Domingo 17 de diciembre se llevó a cabo el Plebiscito Ratificadorio para que ciudadanía se pronunciara a favor o en contra del texto de Nueva Constitución propuesto por el Consejo Constitucional. Proceso de carácter obligatorio, que convocó la asistencia de 13.014.963 electores, quienes llevaron a cabo su derecho ciudadano, donde 5.470.025 lo hicieron por la opción a favor lo que correspondió a un 44.24 %; mientras que 6.894.287 lo hicieron por la opción en contra lo que equivale al 55.76 % (SERVEL, 2023). Nuevamente, un documento constitucional rechazado por la ciudadanía.

3. Marco Teórico

3.1. Constitución de la República

La Constitución es una institución jurídica que limita el ejercicio del poder por medio del Derecho, que reconoce y consagra derechos fundamentales estableciendo los mecanismos de tutela y protección de los mismos. La Constitución no es más que el reflejo del acuerdo social en un momento histórico determinado sobre un mínimo o básico. Un acuerdo sobre lo fundamental. (Garrote, 2024).

En términos simples, una Constitución es la norma fundamental de carácter estructural que permite organizar a un Estado, siendo la guía que orienta su gobernación. Esta norma es también conocida bajo las expresiones de “Carta Magna”, “Carta Fundamental”, “Ley Suprema”, “Ley Superior” o “Código Político”. (BCN, 2024).

El análisis y comparación de propuestas constitucionales requieren el uso de herramientas avanzadas para manejar grandes volúmenes de texto y extraer patrones semánticos complejos. En este contexto, la inteligencia artificial (IA) y sus diversas técnicas juegan un papel crucial. Estas tecnologías permiten procesar y analizar datos textuales de manera eficiente, facilitando la identificación de similitudes y diferencias entre distintos textos constitucionales.

3.2. Inteligencia Artificial

Según Marvin Minsky, quien fue cofundador del laboratorio de IA del Instituto de Tecnología de Massachusetts (MIT), la inteligencia artificial es una disciplina que se encarga de crear máquinas programadas que sean capaces de hacer cosas que requieren la misma inteligencia que si fuesen hechas por humanos (como se citó en Ramírez, 2020). Una técnica computacional que permite a una computadora imitar el comportamiento humano y razonar. La figura 4 ilustra la interrelación que existe entre cuatro conceptos de Inteligencia Artificial.

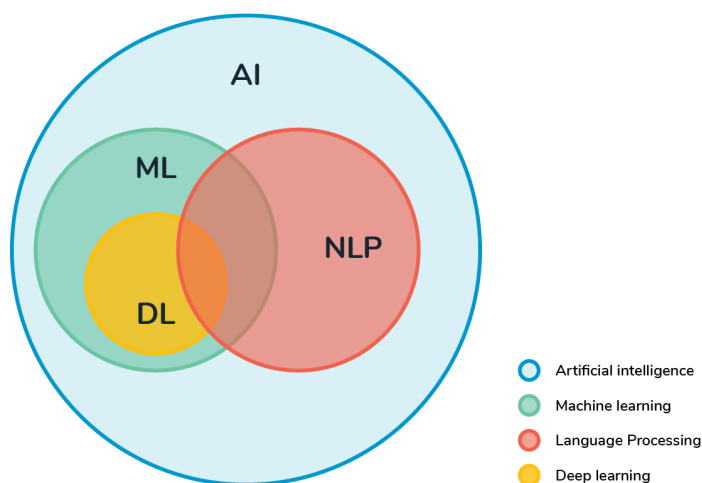


Figura 4: Interrelación entre Inteligencia Artificial. Fuente Extraído desde Sathiyakugan 2018.

3.2.1. Natural Language Processing

El Procesamiento del Lenguaje Natural (NLP) es un campo de la inteligencia artificial que utiliza técnicas computacionales para aprender, comprender y generar contenido en lenguaje humano. Los primeros enfoques de NLP se centraron en automatizar el análisis de la estructura lingüística, así como en desarrollar tecnologías fundamentales como la traducción automática, el reconocimiento del habla y la síntesis del habla (Hirschberg & Manning, 2015). Estos avances sentaron las bases para aplicaciones más complejas, permitiendo a las computadoras interpretar y manipular el lenguaje natural de manera cada vez más sofisticada.

Las técnicas de NLP son esenciales para el análisis comparativo de textos constitucionales, ya que facilitan la extracción de patrones semánticos y la identificación de similitudes y diferencias entre distintos documentos. Entre las herramientas más comunes de NLP se encuentran la tokenización, el etiquetado de partes del discurso, el análisis de sentimiento y la generación de resúmenes automáticos, todas las cuales permiten una comprensión más profunda y detallada de grandes volúmenes de texto (Jurafsky Martin, 2020).

3.3. Machine Learning

El Machine Learning (aprendizaje automático) se desarrolló como un subconjunto de la inteligencia artificial, donde los programas de computadora aprenden de los datos sin estar programados explícitamente para hacerlo (Marr, 2016). Siendo este un conjunto de métodos que pueden detectar automáticamente patrones en los datos y utilizar esos patrones para hacer predicciones (Murphy,

2012:1)

3.3.1. Modelos Supervisados

Los algoritmos de aprendizaje supervisado basan su aprendizaje en un juego de datos de entrenamiento previamente etiquetados. Por etiquetado entendemos que para cada ocurrencia del juego de datos de entrenamiento conocemos el valor de su atributo objetivo. Esto le permitirá al algoritmo poder “aprender” una función capaz de predecir el atributo objetivo para un juego de datos nuevo (Vallalta, 2024).

3.3.2. Modelos No Supervisados

El aprendizaje no supervisado, también conocido como machine learning no supervisado, utiliza algoritmos de machine learning para analizar y agrupar conjuntos de datos no etiquetados. Estos algoritmos descubren patrones ocultos o agrupaciones de datos sin necesidad de intervención humana (IBM, 2024).

En función de la naturaleza del problema, es factible emplear cinco variantes de modelos de aprendizaje automático con el propósito de formular predicciones. Estas modalidades se categorizan en dos conjuntos según el grado de intervención humana requerido durante la fase de entrenamiento del modelo. Dichas modalidades abarcan el aprendizaje supervisado, el aprendizaje no supervisado, el aprendizaje semisupervisado, el aprendizaje por refuerzo y el aprendizaje por transferencia (Furbush, 2018).

3.3.3. Métodos de Agrupación

Cuando se dispone de un grupo de objetos y se está interesado en analizar la similitud existente entre ellos, existen dos acercamientos al problema. Uno de ellos consiste en considerar cada objeto como un grupo aislado con identidad propia e ir incrementando su tamaño hasta obtener un único grupo formado por el total de todos los objetos. Este procedimiento es el denominado aglomerativo. El grupo aumenta de tamaño y, simultáneamente, decrece la similitud entre los objetos. Existe otro acercamiento al problema y es justamente el inverso, considerar que inicialmente todos los objetos forman un grupo. Este grupo inicial se va dividiendo a medida que se exige mayor similitud entre los objetos para que constituyan subgrupos. Este procedimiento se denomina divisivo (Larrechí, 2004).

La clasificación más habitual entre los métodos de análisis de agrupaciones consiste en hablar de métodos jerárquicos y métodos no jerárquicos.

3.3.4. Métodos Jerárquico

Los métodos jerárquicos se basan en la construcción de una estructura en forma de árbol denominada árbol de clasificación o dendrograma (Romero, 2023).

En estos métodos se unen los objetos o grupos de objetos que están más próximos en el espacio m -dimensional hasta formar un solo grupo. Son muchos los algoritmos jerárquicos [Kaufman 1990, Bratchell 1989, Ward 1963] que permiten alcanzar este objetivo, y se diferencian en el criterio utilizado para decidir cuáles deben ser las uniones y la magnitud considerada para evaluar la similitud. Nosotros vamos a describir los empleados más frecuentemente (Larrechi, 2004).

Dendrograma Jerárquico

Es una representación visual de la jerarquía de clústeres resultante de un análisis de clustering jerárquico. El dendrograma muestra cómo los objetos se agrupan y cómo los clústeres se dividen y fusionan entre sí.

Se crea a partir de la matriz de distancias que se calcula durante el análisis de clustering jerárquico. Cada objeto se representa como una hoja en la parte inferior del dendrograma, y los clústeres se representan por las uniones de las ramas. La longitud de las ramas representa la distancia entre los clústeres o entre un objeto y un clúster, y la altura de las uniones indica cuánto se han separado los clústeres en el proceso de clustering jerárquico (Mora, 2023).

Se interpreta leyendo las ramas de abajo hacia arriba. En la parte inferior del dendrograma, cada objeto se representa como un clúster individual. A medida que se sube en el dendrograma, los objetos se agrupan en clústeres cada vez más grandes. Las uniones de las ramas indican cuándo se han fusionado dos clústeres en uno solo, y la altura de la unión indica cuánto se han separado los clústeres en el proceso de clustering jerárquico .

La elección del nivel de corte para dividir el dendrograma en clústeres discretos puede ser subjetiva y depende de los objetivos del análisis. Por lo general, se elige un nivel de corte tal que los clústeres sean lo suficientemente distintos entre sí y lo suficientemente homogéneos en cuanto a la similitud entre sus objetos.

Método de Ward

El método de Ward es una técnica de agrupamiento jerárquico aglomerativo utilizada para formar clústeres de datos con el objetivo de minimizar la varianza dentro de cada grupo. Este método se

basa en el principio de que, en cada etapa del proceso, se combinan los dos clústeres que resulten en el menor incremento en la suma total de los cuadrados de las diferencias dentro de los clústeres. Es decir, el método de Ward busca fusionar aquellos clústeres cuya unión cause el menor aumento en la varianza total, calculada como la suma de las diferencias al cuadrado de cada punto respecto al centroide del clúster (Gallardo, 2024).

Este enfoque se fundamenta en la idea de que los clústeres deben ser lo más homogéneos posible. Para lograrlo, el método mide el impacto de combinar cada par de clústeres y selecciona la fusión que minimiza el incremento en la varianza intra-clúster. Este proceso iterativo continúa hasta que todos los datos se agrupan en un solo clúster o hasta alcanzar el número deseado de clústeres.

3.3.5. Métodos No Jerárquico

Los métodos no jerárquicos, siguen una estructura completamente diferente a la de los métodos jerárquicos. Con estos métodos se pretende clasificar las observaciones en K clusters donde K ha sido fijado previamente (Romero, 2023).

Agrupación en Clústeres

La agrupación es una técnica de minería de datos que agrupa datos sin etiquetar en función de sus similitudes o diferencias. Los algoritmos de agrupación en clústeres se utilizan para procesar objetos de datos brutos y sin clasificar en grupos representados por estructuras o patrones en la información. Los algoritmos de agrupación en clústeres se pueden clasificar en unos pocos tipos, específicamente exclusivos, superpuestos, jerárquicos y probabilísticos (IBM, 2024).

En una variante de la selección aleatoria, se opta por elegir un punto al azar como el primer centroide, y el siguiente centroide se determina como un punto distante del primero. Esta variación garantiza que los centroides iniciales estén adecuadamente espaciados entre sí. Este método de selección contribuye a mitigar la posibilidad de convergencia a soluciones subóptimas al inicio del proceso de agrupamiento, permitiendo así una exploración más efectiva del espacio de características y una identificación más robusta de los grupos en el conjunto de datos. Este enfoque estratégico se alinea con la necesidad de establecer un punto de partida que favorezca la convergencia hacia soluciones óptimas en el análisis de agrupamiento (Sharma, 2019).

La agrupación en clústeres se rige como una herramienta valiosa en las etapas iniciales de las tareas de machine learning. Este enfoque estratégico facilita la comprensión y el descubrimiento de patrones intrínsecos en los conjuntos de datos, sirviendo como fundamento para la toma de decisiones informada y la formulación de estrategias en el análisis de datos complejos. Por estos motivos, la

agrupación en clústeres se utiliza a menudo en las primeras fases de las tareas de machine learning para explorar los datos y descubrir correlaciones inesperadas (Lu, 2021).

Agrupamiento en Clústeres basado en centroides

K-means (K- medias) es un método de análisis de conglomerados que agrupa observaciones minimizando las distancias euclidianas entre ellas. Las distancias euclidianas son análogas a medir la hipotenusa de un triángulo, donde las diferencias entre dos observaciones de dos variables (x e y) se introducen en la ecuación pitagórica para resolver la distancia más corta entre los dos puntos (longitud de la hipotenusa). Las distancias euclidianas se pueden extender a n dimensiones con cualquier número n, y las distancias se refieren a diferencias numéricas en cualquier variable continua medida, no solo distancias espaciales o geométricas. Por lo tanto, esta definición de distancia euclidiana requiere que todas las variables utilizadas para determinar la agrupación utilizando k-medias sean continuas (MacQueen, 2024).

Cuando configura un modelo de agrupación mediante el método K-means, debe especificar un número de destino k que indique la cantidad de centroides que desea en el modelo. El centroide es un punto representativo de cada grupo. El algoritmo K-means asigna cada punto de datos entrante a uno de los grupos minimizando la suma de cuadrados dentro del grupo (Microsoft, 2024).

Método Elbow

El método Elbow es un método heurístico de interpretación y validación de consistencia dentro del análisis de cluster, diseñado para ayudar a encontrar el número apropiado en un conjunto de datos. El método Elbow estipula que el número óptimo de cluster es aquel en el cual adicionando mayor número de los mismos no se reduce significativamente la varianza entre clusters (Gill, 2019). Este método utiliza el valor WCSS, es decir, la suma de la distancia al cuadrado de cada una de las observaciones de su centroide (López, 2019).

3.4. Lenguaje Programación R

R es un lenguaje de programación especialmente orientado al análisis estadístico y a la representación gráfica de los resultados obtenidos. Es un proyecto de GNU por lo que los usuarios son libres de modificarlo y extenderlo. R se distribuye como software libre bajo la licencia GNU y es multiplataforma (hay versiones para plataformas Windows, Mac y Linux, y de hecho algunas distribuciones de Linux lo tienen incorporado), lo que también ha facilitado su adopción y la existencia de una comunidad muy activa su entorno, con constantes desarrollos de nuevas funcionalidades y versiones mejoradas de las existentes. Es un lenguaje basado en comandos, en lugar de hacer clic y arrastrar iconos o menús con el mouse, escribe comandos o instrucciones que se ejecutan. Una secuencia de instrucciones o comandos R que implementa un flujo de trabajo para realizar una tarea se denomina

script o script R. (Manrique, 2014).

Dentro de su estructura, nos podemos encontrar a los paquetes en R, que son colecciones de funciones y conjunto de datos desarrollados por la comunidad. Estos incrementan la potencialidad de R mejorando las funcionalidades base en R, o añadiendo de nuevas (Sánchez, 2018).

3.4.1. Quanteda: Quantitative Analysis of Textual Data

Proporciona funcionalidad para la gestión de corpus, la creación y manipulación de tokens y n-gramas, la exploración de palabras clave en contexto, la formación y manipulación de matrices dispersas de documentos por características y coocurrencias de características, así como también, analizar palabras clave, calcular similitudes y distancias de características, aplicar diccionarios de contenido, aplicar aprendizaje automático supervisado y no supervisado, representar visualmente texto y análisis de texto, y más (Benoit, 2018).

Quanteda facilita la gestión de textos en forma de corpus, definido como una colección de textos que incluye variables a nivel de documento específicas de cada texto, así como metadatos. Quanteda incluye herramientas para hacer fácil y rápida la manipulación de los textos de un corpus, realizando las operaciones más comunes como tareas de Natural Language Processing (NLP) de forma sencilla y rápida, como tokenizar, derivar o formar agrupaciones (Watanabe, 2018).

3.5. Enfoque Comparativo Iberoamericano

La presente investigación adopta un enfoque comparativo Iberoamericano para analizar la evolución constitucional en América Latina y España, centrándose en la identificación y comparación de documentos constitucionales de los 21 países latinoamericanos y España. La metodología se orienta hacia la elucidación de similitudes y divergencias entre las constituciones vigentes en estos territorios y las propuestas de nueva Constitución en Chile, con especial atención a los años 2022 y 2023.

En aras de recopilar un conjunto representativo de documentos constitucionales en vigencia, se recurrió a la plataforma *Constitute.org*, reconocida por ser una fuente confiable y accesible de textos constitucionales a nivel global. Se procedió a la extracción sistemática de los textos constitucionales de los 21 países de América Latina y España, abarcando diversas tradiciones jurídicas y modelos de gobierno presentes en la región.

4. Objetivos

4.1. Objetivo General

Realizar una comparativa semántica, por medio de técnica de clustering, entre las diversas constituciones chilenas y propuestas constitucionales chilenas 2022 y 2023, agrupándolas con las constituciones iberoamericanas.

4.2. Objetivos Específicos

1. Preparación de los datos constitucionales provenientes de las propuestas de Chile, así como de los documentos de otras 22 naciones iberoamericanas, asegurando la coherencia y consistencia de la información para su posterior análisis.
2. Realizar el preprocesamiento de los datos mediante técnicas adecuadas de limpieza y transformación, con el fin de generar vectores representativos de los textos constitucionales, lo que permitirá su análisis computacional.
3. Investigar y abordar la problemática de los datos faltantes en las propuestas constitucionales, implementando técnicas específicas para su manejo. Además, se llevará a cabo la extracción de información relevante para la generación de visualizaciones como word clouds y análisis de keyness.
4. Utilizar técnicas de agrupamiento (clustering) para clasificar las constituciones de acuerdo con similitudes semánticas, lo que facilitará la identificación de patrones y tendencias comunes en los documentos analizados.
5. Elaborar un dendrograma que represente visualmente las relaciones jerárquicas entre las propuestas constitucionales agrupadas, brindando una comprensión intuitiva de la estructura y afinidades entre los textos.
6. Identificar y analizar los patrones y similitudes emergentes entre las propuestas constitucionales de distintos países iberoamericanos.

4.3. Alcance del Estudio

Esta investigación, titulada "Data y Democracia: Estudio por Clustering de las Propuestas Constitucionales 2022 y 2023 con un Enfoque Comparativo Iberoamericano", se propone analizar de manera exhaustiva las propuestas constitucionales presentadas en Chile durante los años mencionados. A continuación, se delimitan los aspectos fundamentales que definen el alcance metodológico y temático de este trabajo.

1. **Foco Temporal, Geográfico y Social** La investigación se centra exclusivamente en las propuestas constitucionales presentadas en Chile durante los años 2022 y 2023. El análisis se circunscribe a este periodo específico, excluyendo otras fases del proceso constitucional y no considera contexto político y social. Además el análisis no considera las diferentes interpretaciones que se pueden dar de las propuestas constitucionales.
2. **Metodología de Clustering y Machine Learning** El estudio se fundamenta en la aplicación de técnicas de clustering, particularmente el método K-Means, para identificar patrones y similitudes entre las propuestas constitucionales. La utilización de herramientas de machine learning se concentra en el análisis detallado de la diversidad y convergencia de estas propuestas, con énfasis en la cohesión intracluster.
3. **Marco Teórico y Enfoque Comparativo Iberoamericano** La fundamentación teórica se adscribe a modelos no supervisados y se enfoca en la data de documentos constitucionales. Además, se lleva a cabo un análisis comparativo a nivel iberoamericano a nivel macro, buscando identificar tendencias comunes sin profundizar en cada texto constitucional.

5. Caso de Estudio

5.1. Metodología de la Investigación

5.1.1. Recolección de los Datos

En una fase preliminar, se ejecutó la recopilación inicial de datos, obteniendo las constituciones Iberoamericanas de los países:

1. Argentina 1853
2. Bolivia 2009
3. Brasil 2009
4. Chile 1980
5. Propuesta Chile 2022
6. Propuesta Chile 2023
7. Reforma Chile 2005
8. Colombia 2015
9. Costa Rica 2020
10. Cuba 2002

11. Ecuador 2021
12. El Salvador 2014
13. España 2011
14. Guatemala 1193
15. Haití 2012
16. Honduras 2013
17. México 2015
18. Nicaragua 2014
19. Panamá 2004
20. Paraguay 2011
21. Perú 2021
22. Portugal 2005
23. República Dominicana 2015
24. Uruguay 2004
25. Venezuela 2009

Este procedimiento se materializó mediante el empleo de las plataformas de la Biblioteca del Congreso Nacional de Chile (BCN) y Constituteproject.org, plataforma que facilita las constituciones del mundo, para apoyar el diseño constitucional e informar a los ciudadanos. Estas plataformas actúan como fuentes primarias para la obtención de los textos constitucionales pertinentes.

Posterior a la adquisición de los documentos objeto de análisis, estos se agruparon dentro de una carpeta, que luego fue importada a la plataforma RStudio para dar inicio a la investigación.

5.1.2. Preparación de los Datos

1. **Utilización de RStudio:** Con el propósito de emplear el lenguaje de programación R para realizar el análisis textual, se procedió a la instalación de diversos paquetes en la plataforma, entre los cuales sobresalen "quanteda.textmodels", "quanteda.textplots" y "quanteda.textstats". Estas incorporaciones se llevaron a cabo con la finalidad de hacer uso de la biblioteca quanteda, la cual destaca como una herramienta fundamental para el estudio.

2. **Creación de Corpus Textual:** La creación del corpus constituye una fase esencial en el desarrollo de este estudio, representando una colección de textos constitucionales que se someterán al proceso de Procesamiento del Lenguaje Natural (NLP). La compilación del corpus se erige como un acto minucioso de selección y organización de documentos constitucionales provenientes de diversas jurisdicciones, abarcando un espectro representativo de la riqueza normativa y legal de la región Iberoamericana. Este conjunto, cuidadosamente curado, no solo refleja la diversidad jurídica presente en los países analizados, sino que también garantiza la representatividad necesaria para la posterior aplicación de algoritmos de aprendizaje automático.

5.1.3. Limpieza de Data

La fase de depuración textual se orienta a la eliminación de elementos carentes de relevancia temática, con el propósito de preservar únicamente la información pertinente.

1. **Vector de palabras:** Con el propósito de identificar y descartar elementos no pertinentes en el análisis, se llevó a cabo la confección de un vector que englobara aquellas palabras carecientes de relevancia para la investigación. Para la identificación de las palabras destinadas a ser almacenadas en este vector, se ejecutaron múltiples iteraciones mediante la generación de los gráficos de Wordcloud y Keyness.

La figura 5 presenta el gráfico de WordCloud, comúnmente denominado nube de palabras, que exhibe las palabras más frecuentes en los textos constitucionales. Este gráfico posibilita la identificación visual de la prevalencia de palabras, donde aquellas con mayor frecuencia se destacan mediante un tamaño proporcionalmente mayor, permitiendo una apreciación intuitiva de su relevancia en el conjunto analizado.

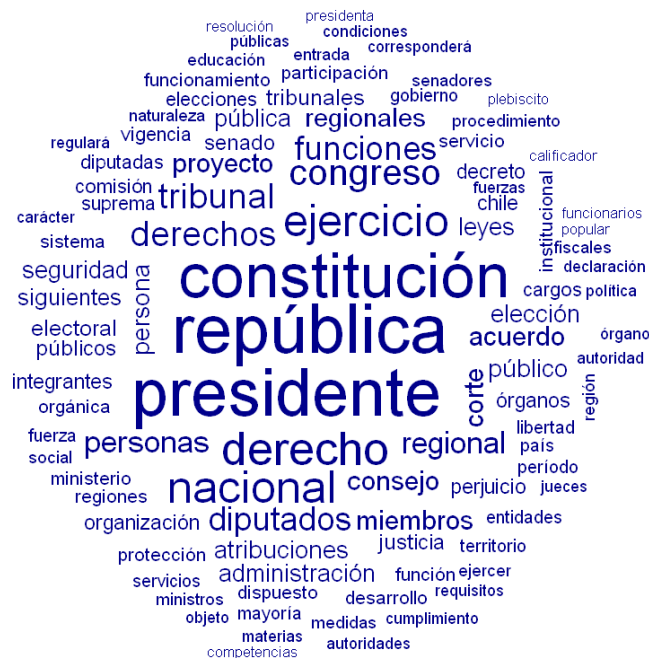


Figura 5: Gráfico WorldCloud con palabras contenidas en los textos constitucionales. Elaboración Propia.

La figura 6 ilustra un análisis Keynes, que se basa en encontrar términos que tengan alta probabilidad de ocurrencia en un texto -el texto objetivo- respecto de otro texto -texto base-. Las barras asociadas a los términos en el gráfico, corresponden a su probabilidad relativa. En el caso particular, se compararon las propuestas constitucionales chilenas del año 2022 y 2023.

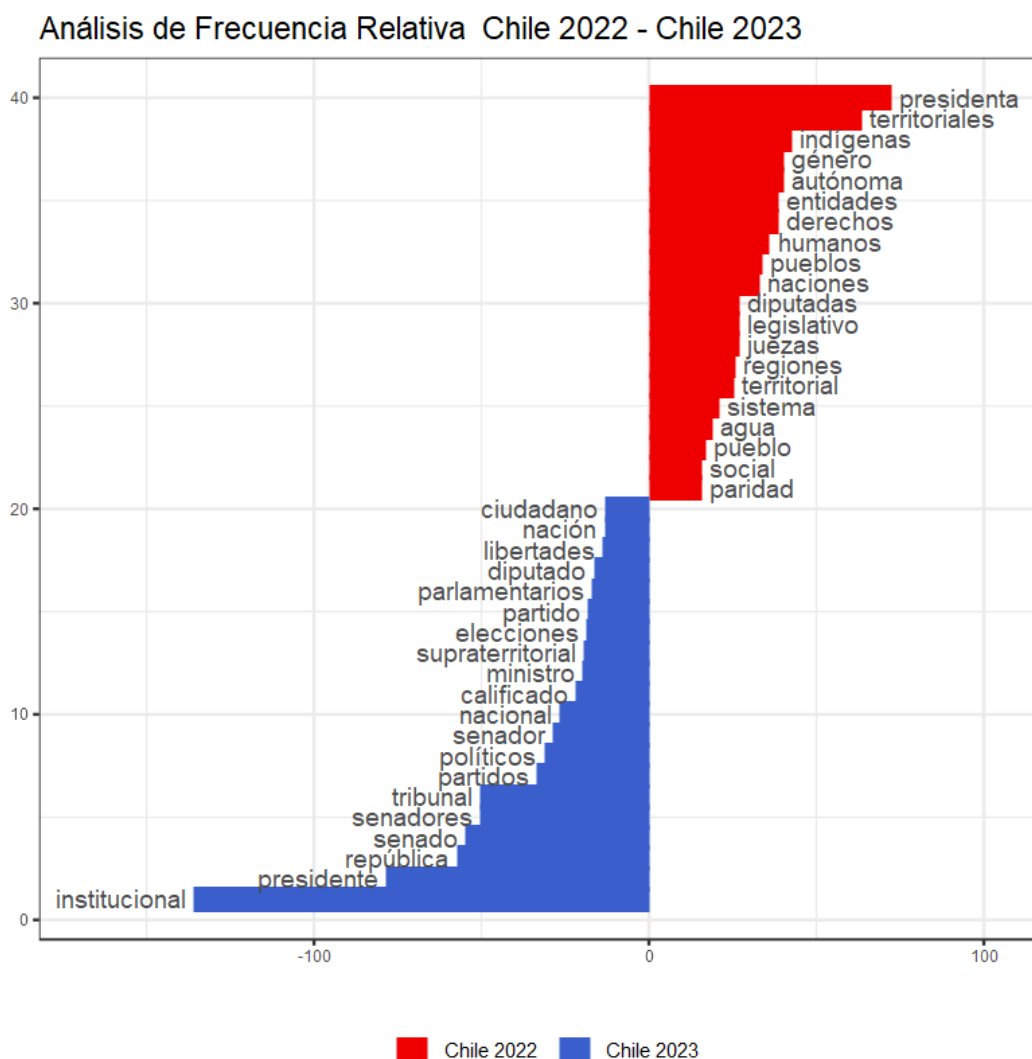


Figura 6: Gráfico Keynes con palabras contenidas en las propuestas constitucionales del año 2022 y 2023. Elaboración Propia.

Estos procedimientos permitieron esclarecer las palabras carentes de significado en el contexto de la investigación. Como resultado, se logró la eliminación de dichas palabras, posibilitando la continuidad del proceso de depuración del análisis.

2. **Tokenización:** Se llevó a cabo la supresión de puntuaciones, números, componentes irrelevantes y palabras con una longitud de caracteres inferior a tres.

5.1.4. Document Feature Matrix

Prosiguiendo con la investigación, se llevó a cabo la creación de un Document-Feature Matrix (DFM), una representación textual que exhibe la frecuencia de ocurrencia de las palabras dentro del corpus de documentos constitucionales.

Para procesar este DFM, se aplicaron diversas funciones:

- `dfm_weight`, la cual emplea el esquema de proporción. Esto implica asignar un peso específico a cada palabra en relación con su frecuencia de aparición en cada documento. Este enfoque resalta las palabras que son más distintivas y, posiblemente, más relevantes para comprender el contenido de cada documento constitucional.
- `textstat_dist`, calcula la distancia entre los documentos. Utilizando la distancia euclidiana, se midieron las diferencia entre dos conjuntos de datos.
- `as.dist`, donde convertimos el resultado de la distancia en una matriz que representa estas distancias. Esta matriz nos da una forma de comparar los documentos entre sí. Cada valor en la matriz nos dice qué tan diferentes son dos documentos entre sí en términos de las palabras que contienen.

Este paso adicional en el proceso de investigación aporta una dimensión analítica esencial al permitir la cuantificación y estructuración de la información contenida en los textos constitucionales recopilados.

5.1.5. Determinación de Codo

La determinación del número óptimo de clústeres para el análisis se realizó mediante la aplicación del método del codo, una técnica ampliamente empleada en el ámbito del análisis de agrupamiento para establecer el número más apropiado de clústeres en un conjunto de datos. Este proceso se llevó a cabo utilizando la matriz de distancias entre los documentos, previamente calculada.

Según lo observado en la figura 4, un punto crítico donde el descenso en la WSS (Within-Cluster Sum of Squares o suma de cuadrados dentro del clúster) se vuelve menos pronunciado, marcando el número óptimo de clústeres, se sitúa en los puntos 4 y 5. La identificación de estos puntos, comúnmente conocido como el "codo" en la gráfica, señala el equilibrio entre una partición más detallada y la preservación de la cohesión intracúster.

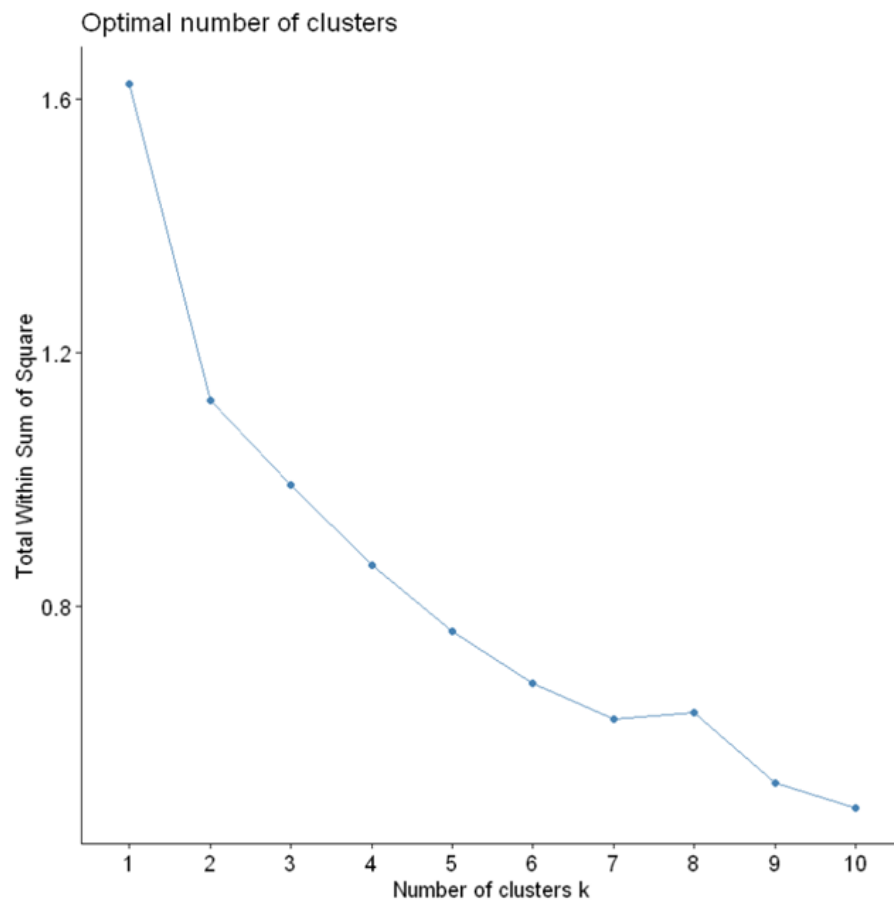


Figura 7: Número óptimo de clústeres $k=4$ y $k = 5$, mediante método del codo. Elaboración propia.

5.1.6. Clustering K-Means

K-Means con $k=4$

En este estudio, se empleó el método de agrupamiento Clustering K-Means, una técnica que segmenta los documentos en k grupos. Para este caso específico, se optó por $k = 4$. En términos sencillos, cada grupo o clúster puede considerarse como una categoría que agrupa documentos con similitudes significativas, asignando cada documento al clúster más afín en función de su proximidad semántica.

La elección de $k = 4$ se basó en una evaluación cuidadosa mediante la metodología del codo, que busca identificar un equilibrio óptimo entre la complejidad de la agrupación y la representatividad de los patrones identificados. La metodología del codo ayuda a determinar el número adecuado de clústeres al analizar la suma de las distancias cuadradas dentro de cada clúster y observar el punto donde la tasa de disminución comienza a estabilizarse. Este análisis indicó que cuatro clústeres proporcionan una estructura adecuada para interpretar las relaciones entre los documentos, manteniendo un balance entre la diferenciación de los grupos y la simplicidad del modelo.

Clúster	Países
Clúster n°1	• Colombia 1991 (rev. 2015) • Guatemala 1985 (rev. 1993) • Honduras 1982 (rev. 2013) • México 1917 (rev. 2015) • Paraguay 1992 (rev. 2011) • Perú 1993 (rev. 2021) • Propuesta Chile 2022 • República Dominicana 2015
Clúster n°2	• Argentina 1853 (reinst. 1983, rev. 1994) • Bolivia (Estado Plurinacional de) 2009 • Brazil 1988 (rev. 2017) • Cuba 1976 (rev. 2002) • España 1978 (rev. 2011) • Uruguay 1966 (reinst. 1985, rev. 2004)
Clúster n°3	• Chile 1980 • Propuesta Chile 2023 • Reforma Chile 2005
Clúster n°4	• Costa Rica 1949 (rev. 2020) • Ecuador 2008 (rev. 2021) • El Salvador 2014 • Haiti 1987 (rev. 2012) • Nicaragua 1987 (rev. 2014) • Panama 1972 (rev. 2004) • Portugal 1976 (rev. 2005) • Venezuela (República Bolivariana de) 1999 (rev. 2009)

Figura 8: Constituciones Iberoamericanas agrupadas en 4 clústeres. Elaboración propia.

K-Means con $k=5$

Como continuación de la metodología aplicada en este análisis, se consideró relevante explorar la variación de resultados al aumentar el número de centroides a $k = 5$. Esta decisión, surgió de la necesidad de una agrupación más amplia que permitiera una mejor captura de la complejidad de los datos constitucionales. Ejecutando nuevamente el proceso de Clustering K-Means, ahora con 5 centroides.

Clúster	Países
Clúster n°1	• Bolivia (Estado Plurinacional de) 2009 • Cuba 1976 (rev. 2002) • Costa Rica 1949 (rev. 2020) • Uruguay 1966 (reinst. 1985, rev. 2004)
Clúster n°2	• Chile 1980 • Reforma Chile 2005 • Propuesta Chile 2023
Clúster n°3	• Colombia 1991 (rev. 2015) • Guatemala 1985 (rev. 1993) • El Salvador 1983 (rev. 2014) • Honduras 1982 (rev. 2013) • México 1917 (rev. 2015) • Panamá 1972 (rev. 2004) • Paraguay 1992 (rev. 2011) • Perú 1993 (rev. 2021) • Propuesta Chile 2022 • República Dominicana 2015
Clúster n°4	• Argentina 1853 (reinst. 1983, rev. 1994) • Brasil 1988 (rev. 2017) • España 1978 (rev. 2011)
Clúster n°5	• Ecuador 2008 (rev. 2021) • Haití 1987 (rev. 2012) • Nicaragua 1987 (rev. 2014) • Portugal 1976 (rev. 2005) • Venezuela (República Bolivariana de) 1999 (rev. 2009)

Figura 9: Constituciones Iberoamericanas agrupadas en 5 clústeres. Elaboración propia.

La figura 9 muestra las constituciones iberoamericanas agrupadas en 5 clústeres. Este nuevo enfoque permitió una segmentación más detallada de los documentos constitucionales, revelando patrones entre los textos.

5.1.7. Dendrograma Jerárquico

En pos de fortalecer y profundizar la comprensión de los patrones identificados durante la ejecución de los dos clustering K-Means previos, se tomó la decisión de construir un dendrograma jerárquico con $k=5$.

El dendrograma jerárquico es una herramienta visual utilizada en el clustering jerárquico. En este método los datos se organizan en una estructura de árbol jerárquico, donde los clústeres se dividen y fusionan sucesivamente en función de la distancia entre los puntos de datos.

La representación jerárquica del dendrograma facilita una interpretación detallada de las relaciones de proximidad y distinción entre los grupos, ofreciendo una visión profunda de las dinámicas y patrones presentes en el conjunto de datos analizado.

5.2. Resultados

En esta sección, se presentarán los hallazgos obtenidos a partir del análisis realizado en la investigación, con foco en las propuestas constitucionales chilenas del año 2022 y 2023. Se incluirán los resultados de los diferentes enfoques metodológicos empleados, como el clustering K-Means con $k=5$ y $k=4$, así como el dendrograma jerárquico generado. Se proporcionarán detalles sobre la distribución de los documentos constitucionales en los clústeres identificados y se analizarán las similitudes y diferencias entre ellos.

5.3. Clustering K-Means con $k=4$

El análisis de K-means con $k=4$ identificó cuatro clústeres principales, observados en la figura 10.

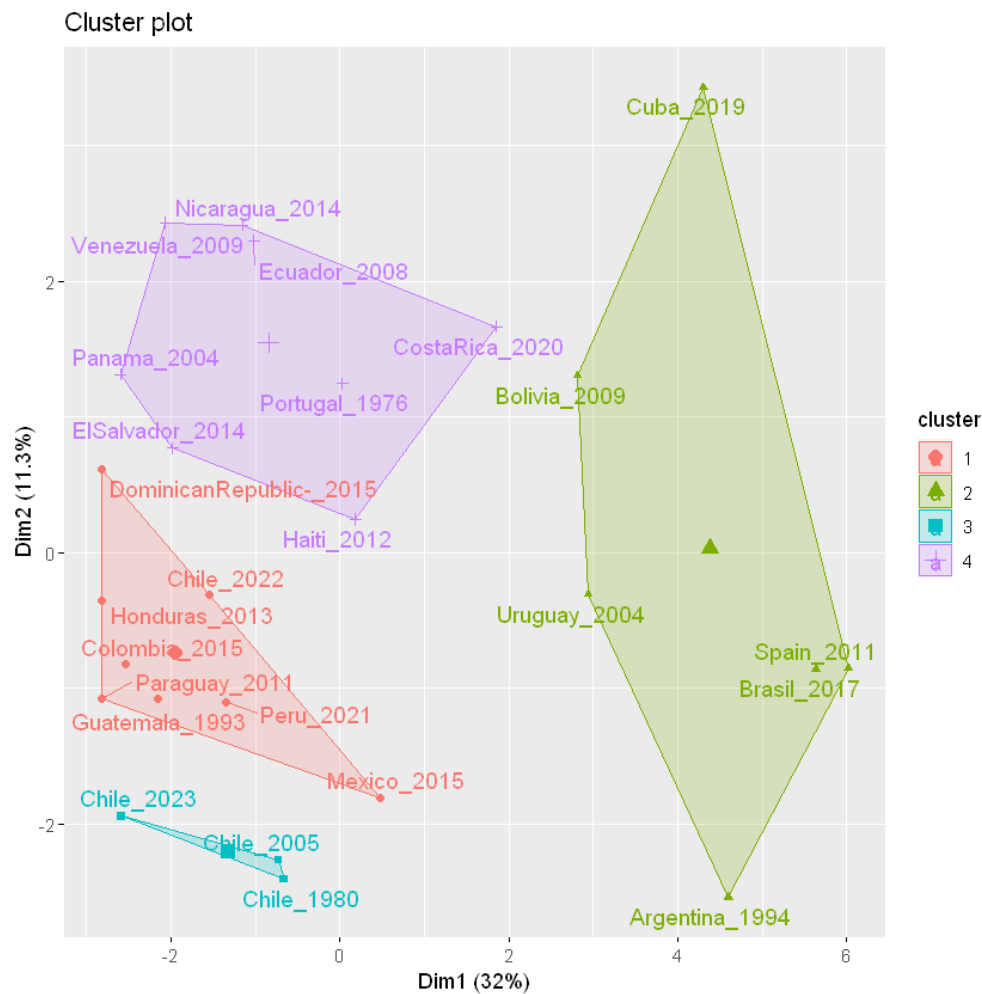


Figura 10: K-Means Clustering con $k = 4$. Elaboración propia.

Clúster I: Democracia, Estado de derecho y derechos individuales

Este clúster de color rojo, se agrupan constituciones que enfatizan la consolidación de la democracia, el fortalecimiento del Estado y la protección de los derechos individuales. La propuesta de Chile 2022 se ubica preliminarmente en este clúster, reflejando su enfoque en estos temas centrales (Asamblea Constituyente de Chile, 2022). Algunos elementos característicos de las propuestas en este clúster incluyen:

- **Fortalecimiento del sistema judicial:** Se busca crear un sistema judicial independiente e imparcial, con un Tribunal Constitucional con funciones limitadas (Asamblea Constituyente de Chile, 2022).
- **Descentralización del poder:** Se propone una mayor descentralización del poder, con la creación de regiones y municipios con mayores atribuciones (Asamblea Constituyente de Chile, 2022).
- **Reconocimiento de los derechos individuales:** Se reconoce la primacía de los derechos individuales, incluyendo la libertad de expresión, la libertad de asociación y el derecho a la propiedad privada (Asamblea Constituyente de Chile, 2022).

Clúster III: Derechos sociales, participación ciudadana y transparencia

Agrupando textos constitucionales que se enfocan en la garantía de los derechos sociales, económicos y culturales, así como en el fomento de la participación ciudadana y la transparencia nos encontramos con el clúster tres de color celeste. La propuesta de Chile 2023 se ubica preliminarmente en este clúster, evidenciando su énfasis en estos aspectos (Convención Constitucional Constituyente de Chile, 2023). Logrando identificar las siguientes características:

- **Estado Social de Derecho:** Se define a Chile como un Estado Social de Derecho, con un rol activo en la garantía de los derechos sociales (Convención Constitucional Constituyente de Chile, 2023).
- **Derechos de la naturaleza:** Se reconocen los derechos de la naturaleza y se establece la obligación del Estado de protegerlos (Convención Constitucional Constituyente de Chile, 2023).
- **Participación ciudadana:** Se implementan mecanismos para fortalecer la participación ciudadana en la toma de decisiones públicas (Convención Constitucional Constituyente de Chile, 2023).

Los textos constitucionales que se encuentran en el Clúster III mantendrían un enfoque similar con la propuesta constitucional chilena del año 2023 en relación a los tópicos nombrados anteriormente. Incluyéndolos de la siguiente forma:

- **Constitución chilena de 1980:** Esta propuesta constitucional establece un Estado Social de

Derecho que garantiza los derechos fundamentales de las personas (Comisión de Estudios de la Constitución, 1980).

- **Reforma constitucional chilena del año 2005:** Esta reforma constitucional introduce reformas al sistema electoral, al sistema judicial y a la protección de los derechos fundamentales (Congreso Nacional de Chile, 2005).

5.4. Clustering K-Means con $k=5$

El análisis de K-means con $k = 5$ crea una estructura de clústeres más granular, dividiendo algunos clústeres de K-means = 4 en subgrupos más específicos. Las tendencias generales observadas en K-means = 4 se mantienen en K-means = 5, con una mayor diferenciación entre las propuestas dentro de cada clúster.

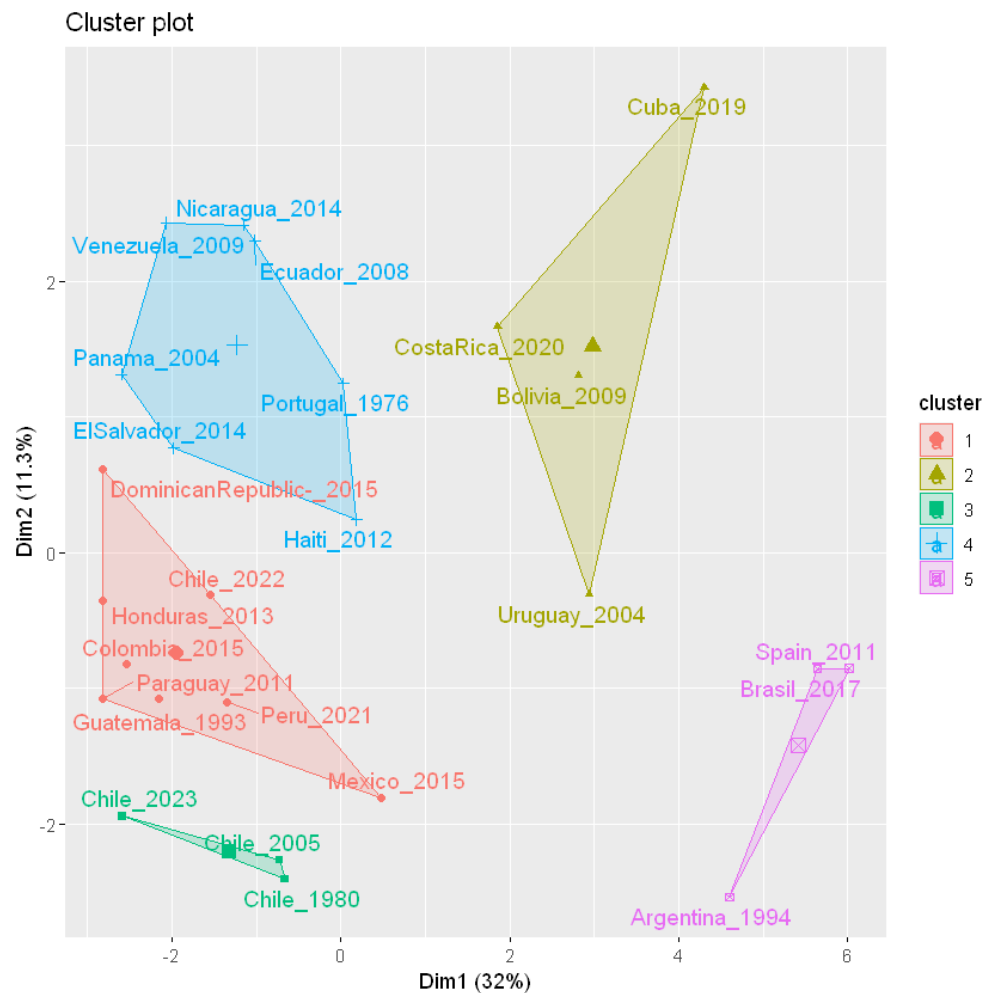


Figura 11: K-means Clustering con $k = 5$. Elaboración propia.

Al centrarnos en la ubicación de estas dos propuestas en sus respectivos clústeres, logramos identificar que la propuesta de Chile 2022 se ubica nuevamente con constituciones como Perú 2021 y Honduras 2013. De la misma forma, la propuesta de Chile 2023 se mantiene con la Constitución chilena del año 1980 y la reforma constitucional del año 2005.

5.5. Dendrograma Jerárquico

En la figura 12 se observa el dendrograma jerárquico generado a partir del clustering K-Means con $k=5$. En el eje horizontal se encuentran los nombres de las propuestas, y en el eje vertical se representa la distancia entre las propuestas. Cuanto más cerca estén dos propuestas en el eje vertical, más similares son entre sí. Donde se reafirma lo identificado anteriormente, se observa que la agrupación de las propuestas en los clústeres coincide en gran medida con los niveles de corte del dendrograma. Además, se mantiene la agrupación entre la propuesta constitucional de Chile 2022 con Brasil, Argentina, México, Honduras, Perú, Colombia, Guatemala y Paraguay. Del mismo modo, la agrupación de la Constitución de Chile 1980, la reforma constitucional chilena del 2005 y la propuesta constitucional del año 2023.

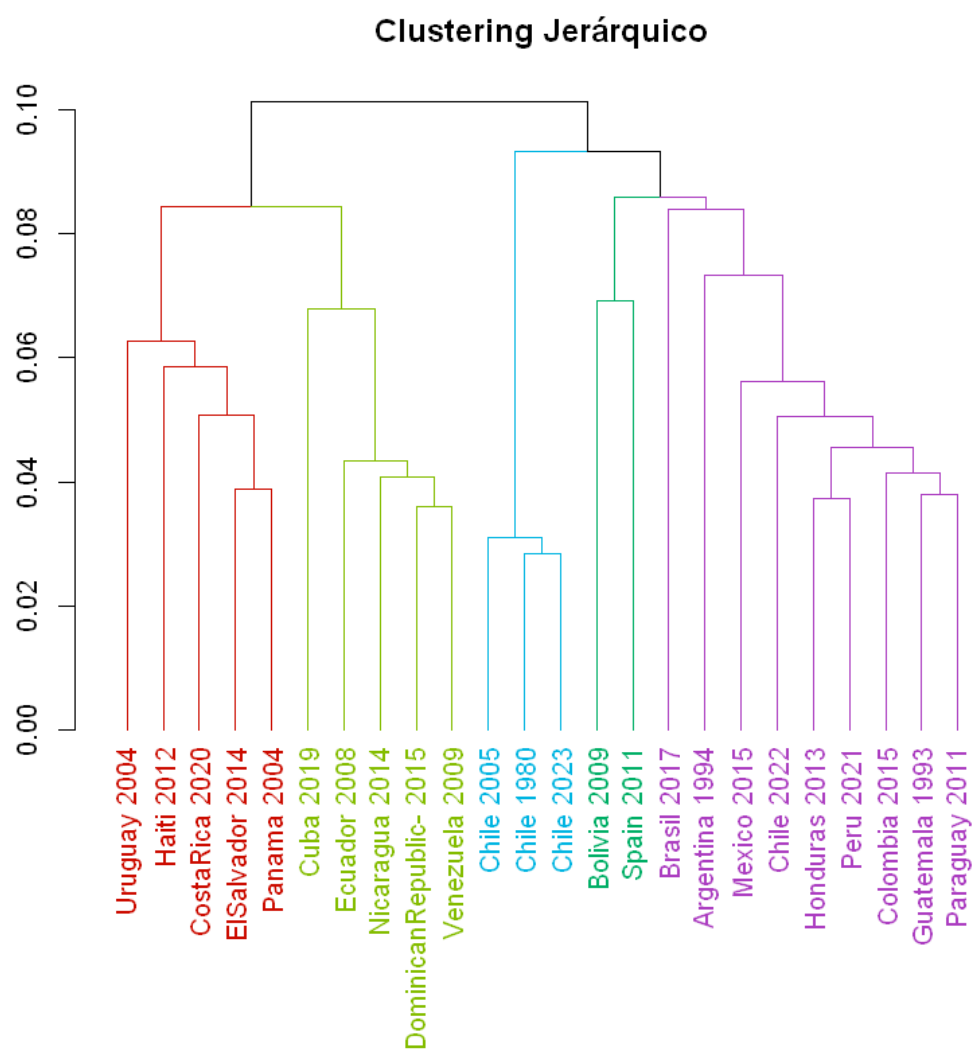


Figura 12: Dendrograma jerárquico con $k = 5$. Elaboración propia.

6. Conclusiones

El proceso constituyente en Chile ha sido un viaje complejo y dinámico, marcado por la elaboración de dos propuestas constitucionales significativas: la propuesta de 2022, rechazada en el plebiscito de salida, y la propuesta de 2023, también rechazada. Mediante el análisis de clustering y el dendrograma, se ha explorado en profundidad la ubicación de estas propuestas en los clusters identificados, revelando insights cruciales sobre su naturaleza y su relación con el discurso constitucional chileno.

6.1. Conclusiones del caso

Tras la extracción, emprendimos la tarea de limpiar los datos a analizar, eliminando elementos irrelevantes como signos de puntuación o caracteres especiales. Este proceso meticuloso fue crucial para garantizar la precisión del análisis posterior. Una vez limpios los datos, los transformamos en el formato Document Features Matrix (DFM), una representación que nos permitió capturar la esencia de cada texto constitucional.

Gracias al DFM y al poder del NLP, pudimos obtener una visualización clara de la distribución de los textos en los clústeres identificados. Esta visualización nos reveló patrones interesantes que antes permanecían ocultos, brindándonos insights valiosos sobre el discurso constitucional chileno.

Uno de los hallazgos más significativos es la afinidad semántica que presenta la propuesta constitucional chilena del año 2023 con la Constitución de 1980 y la reforma de 2005. Es interesante observar cómo, independientemente de los métodos de agrupamiento utilizados, ya sea k-means con $k=4$ y $k=5$, o el clustering jerárquico, esta propuesta sigue ubicándose en clústeres compartidos con estas cartas magnas. Este fenómeno sugiere una conexión conceptual entre la propuesta de 2023 y las constituciones anteriores, destacando una continuidad en ciertos elementos fundamentales del discurso constitucional chileno a lo largo del tiempo. Por el contrario, la propuesta de 2022 muestra una agrupación en clústeres distintos, lo que resalta aún más la evolución del discurso constitucional en Chile y evidencia un cambio de enfoque entre las propuestas analizadas.

La ubicación de las propuestas de Chile 2022 y 2023 en clústeres separados no solo revela una evolución del discurso constitucional, sino que también sugiere diferencias sustanciales en sus enfoques y contenidos. Esta distribución espacial de los textos podría reflejar las prioridades y preocupaciones políticas predominantes en cada período específico, capturando el pulso cambiante de la sociedad chilena.

Además, la agrupación de las propuestas en clústeres separados pone de manifiesto la flexibilidad y adaptabilidad del proceso constituyente chileno. Esta característica crucial permite que el proceso

se adapte y evolucione para responder a las necesidades y aspiraciones cambiantes de la sociedad a lo largo del tiempo. El proceso constituyente no es estático, sino que se transforma dinámicamente para reflejar las prioridades y preocupaciones de cada era.

Los resultados del análisis del discurso constitucional proporcionan una base sólida para el diálogo constructivo entre los diversos actores involucrados en el proceso. Al comprender las diferentes perspectivas y ubicaciones de las propuestas, los actores pueden comenzar a identificar puntos de convergencia y divergencia, allanando el camino para la construcción de consensos y la toma de decisiones informadas. El diálogo constructivo es fundamental para avanzar hacia un nuevo marco constitucional que responda a las necesidades y aspiraciones de la sociedad chilena.

6.2. Conclusiones finales

El uso de métodos agrupamiento, ha sido fundamental para la comprensión del proceso constituyente en Chile. El análisis de clustering ha permitido identificar agrupamientos de propuestas constitucionales con características similares. Esto ha facilitado la comprensión de las diferentes tendencias y enfoques presentes en el debate constitucional. La identificación de estos agrupamientos ha tenido un impacto significativo en la comprensión del proceso constituyente chileno, facilitando la comparación de propuestas, revelando patrones y enriqueciendo el debate. Sin embargo, aún hay un camino por recorrer para profundizar en el análisis y extraer conclusiones más robustas.

En este sentido, se hace hincapié en la importancia de futuras investigaciones colaborativas con investigadores con conocimiento en el ámbito legislativo. Esta colaboración interdisciplinaria permitiría realizar un análisis más en profundidad, en relación al tipo de contenido que abarcan los textos constitucionales. Al incorporar la experiencia de expertos en el ámbito legislativo, se podrían identificar patrones y tendencias más específicas, permitiendo extraer conclusiones más precisas y relevantes de los clústeres.

7. Referencias bibliográficas

- **Asamblea Constituyente de Chile** (2022). *Propuesta de nueva Constitución para Chile*. Disponible en <https://www.chileconvencion.cl/wp-content/uploads/2022/07/Texto-Definitivo-CPR-2022-Tapas.pdf>
- **Avendaño, O.** (2019). *Estallido social en Chile: Los dilemas políticos desde octubre del 2019*. Revista De Ciencia Política, 57(2), 105–119. Disponible <https://doi.org/10.5354/0719-5338.2019.61558>
- **Avendaño, O. Osorio, N.** (2021). *Estallido Social en Chile: Desafíos y proyecciones del período entre octubre de 2019 y mayo de 2021*. Diponible: https://rua.ua.es/dspace/bitstream/10045/119477/1/Ambos-Mundos-n02_01.pdf
- **BCN.** (2024). Biblioteca del Congreso Nacional. *Guía de Formación Cívica - La Constitución*. Disponible: https://www.bcn.cl/formacioncivica/detalle_guia?h=10221.3/45675#:~:text=En%20t%C3%A9rminos%20simples%20una%20Constituci%C3%B3n,%E2%80%9D%20o%20%E2%80%9C%C3%B3digo%20Pol%C3%ADtico%E2%80%9D.
- **BCN.** (2024). Biblioteca del Congreso Nacional. *Ley Fácil - Proceso Constitucional 2023*. Disponible: <https://www.bcn.cl/portal/leyfacil/recurso/proceso-constitucional-2023>
- **Benoit, K.** (2023). Cran R Project. *Quanteda: Quantitative Analysis of Textual Data*. 18 Mayo Disponible: <https://cran.r-project.org/web/packages/quanteda/index.html>
- **ChileAtiende.** (2022). *Proceso constituyente*. Disponible: <https://www.chileatiende.gob.cl/fichas/72504-proceso-constituyente#:~:text=La%20Convenci%C3%B3n%20Constituyente%20tendr%C3%A1%20por,tarea%20que%20le%20fue%20encargada.>
- **Comisión de Estudios de la Constitución** (1980). *Propuesta de nueva Constitución para la República de Chile*. Disponible en <https://www.bcn.cl/leychile/constituciones>
- **Congreso Nacional de Chile** (2005). *Reforma Constitucional de 2005*. Disponible en: <https://uchile.cl/noticias/192165/mitos-y-verdades-sobre-las-reformas-constitucionales-de-2005>
- **Convención Constitucional Constituyente de Chile** (2023). *Borrador de propuesta de nueva Constitución para Chile*. Disponible: <https://www.biobiochile.cl/especial/nuevo-proceso-constituyente/noticias/2023/10/30/borrador-de-nueva-propuesta-de-constitucion-descargalo-aqui.shtml>
- **Furbush, J.** (2018). *Machine learning: a quick and simple definition*. Disponible: <https://www.oreilly.com/ideas/machine-learning-a-quick-and-simple-definition>
- **Gallardo, F.** (2024). Diario Constitucional. *Métodos Jerárquicos de Análisis - Capítulo 3*. Universidad de Granada. Disponible: <https://www.ugr.es/~gallardo/pdf/cluster-3.pdf>

- **Garrote, E.** (2024). Diario Constitucional. *¿Qué es una Constitución y para qué sirve?*. Disponible: <https://www.diarioconstitucional.cl/articulos/que-es-una-constitucion-y-para-que-sirve/>
- **Gill, T.** (2019). Universidad Nacional de Lomas de Zamora. *Aplicación de Data Analytics para contribuir al desarrollo de nuevo métodos de estimación de la demanda en los modelos de negocio con sistema de venta directa*. Disponible: <http://repositorio.unlz.edu.ar:8080/handle/123456789/559>
- **Hirscheberg, J.** (2015). Science. *Advances in natural language processing - Special Issue Review*. 17 Julio Disponible: <https://www.science.org/doi/abs/10.1126/science.aaa8685>
- **IBM** (2024). IBM Cloud. *¿Qué es el aprendizaje no supervisado?*. Disponible: <https://www.ibm.com/mx-es/topics/unsupervised-learning>
- **Jurafsky, D., & Martin, J. H.** (2020). Stanford University. *Speech and Language Processing (3er ed.)*. Disponible: <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>
- **Larrechi, M.** (2004). Universitat Rovira i Virgili. *Exploración de resultados multidimensionales: Análisis de agrupaciones mediante métodos jerárquicos*. Disponible: <http://rodi.urv.es/quimio/general/agrup.pdf>
- **López, V.** (2019). *Aplicación y comparativa de cuatro modelos de clustering para datos GTE*. Universitat Oberta de Catalunya. Disponible: <https://openaccess.uoc.edu/bitstream/10609/90626/6/vlopezsanchTFM0119memoria.pdf>
- **Lu, P** (2021). *Component: K-Means Clustering*. 11 Marzo. Disponible: <https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/k-means-clustering?view=azureml-api-2>
- **MacQueen, J..** (2024). *K-Means Clustering Analysis - Mailman School of Public Health*. Disponible: <https://www.publichealth.columbia.edu/research/population-health-methods/k-means-cluster-analysis>
- **Maldonado, M.** (2023). Universidad de Chile *¿Cuáles son los próximos pasos del proceso constitucional 2023?*. 22 Noviembre. Disponible: <https://uchile.cl/noticias/211499/cuales-son-los-proximos-pasos-del-proceso-constitucional-2023>
- **Manrique, E.** (2020). Revista Ibérica de Sistemas y Tecnologías de Información *Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo*. 06 Enero. Disponible: <https://www.proquest.com/openview/c7e24c997199215aa26a39107dd2fe98/1?pq-origsite=gscholar&cbl=1006393>
- **Marr, B.** (2016). *A short history of machine learning. - Every Manager Should Read- Forbes*. 19 Febrero. Disponible: <https://www.forbes.com/sites/bernardmarr/2016/02/19/>

a-short-history-of-machine-learning-every-manager-should-read/?sh=38e474b915e7

- **Mora, M.** (2023). *Cluster jerárquico - Descubre los métodos y técnicas avanzadas para una segmentación empresarial más precisa*. 28 Julio. Disponible: <https://es.linkedin.com/pulse/cluster-jer%C3%A1rquico-descubre-los-m%C3%A9todos-y-t%C3%A9nicas-mora-caballero>
- **Murphy, K.P.** (2012). *Machine learning: a probabilistic perspective*. Boston, Massachusetts: MIT Press.
- **Ramírez, F.** (2020). *¿Qué es y cómo funciona la Inteligencia Artificial?* Disponible: <https://uchile.cl/noticias/164233/que-es-y-como-funciona-la-inteligencia-artificial>
- **Romero, J.** (2023). *Estadística Multivalente - Análisis cluster (AC)*. Disponible: <https://digibug.ugr.es/bitstream/handle/10481/85861/AC.pdf?sequence=1&isAllowed=y>
- **Sánchez, R.** (2018). *Ciencia de Datos con R*. Disponible: <https://rsanchezs.gitbooks.io/ciencia-de-datos-con-r/content/paquetes/paquetes.html>
- **Sharma, P.** (2019). *The Ultimate Guide to K-Means Clustering: Definition, Methods and Applications*. 03 Noviembre. Disponible: <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/>
- **Urrejola, J.** (2019). *La cronología del estallido social de Chile*. 25 Noviembre. Disponible: <https://www.dw.com/es/la-cronolog%C3%ADa-del-estallido-social-de-chile/a-51407726>
- **Vallalta, J.** (2024). *Data Mining: Aprendizaje supervisado y no supervisado*. Disponible: <https://healthdataminer.com/data-mining/aprendizaje-supervisado-y-no-supervisado/>
- **Watanabe, K.** (2023). *quanteda: Quantitative Analysis of Textual Data*. 18 Mayo. Disponible: <https://cran.r-project.org/web/packages/quanteda/index.html>
- **Zúñiga, F.** (2022). *El rechazo en el plebiscito chileno de 4 de septiembre de 2022: perspectivas constitucionales*. 25 Octubre. Disponible: <https://revistas.uv.es/index.php/cuadernosconstitucionales/article/view/25721/21890>