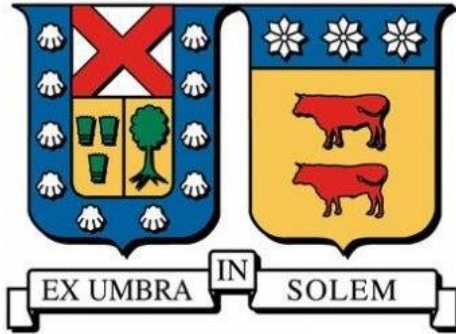




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA DEPARTAMENTO DE INDUSTRIAS

**FACTORES EXPLICATIVOS DE LA PRODUCTIVIDAD LEGISLATIVA EN LA CÁMARA DE
DIPUTADOS DE CHILE, 2022-2026: UN ANÁLISIS MEDIANTE MODELOS PREDICTIVOS**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERÍA CIVIL INDUSTRIAL

AUTOR

TOMÁS FELIPE ESTAY BONILLA

PROFESOR GUÍA

RAÚL STEGMAIER

SANTIAGO, 6 DE JULIO, 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título o Tesis de Postgrado

Título del trabajo: Factores explicativos de la productividad legislativa en la cámara de diputados de Chile, 2022-2026: un análisis mediante modelos predictivos

Nombre del candidato(a): Tomás Felipe Estay Bonilla

Carrera / Grado: Ingeniería Civil Industrial

Campus: Vitacura Departamento: Industrias

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Raúl Stegmaier Bravo, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente DEJO CONSTANCIA que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo NO contiene información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo CONTIENE información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por (marcar una opción):

o 6 meses o 12 meses o 2 años o 3 años o 5 años o 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 21/08/2026

Firma: _____

Estudiante o Candidato(a):

Fecha: 21/08/2026

Firma: _____

Agradecimientos

A lo largo de esta investigación he podido observar distintos aspectos de mi propia vida y en general me he dado cuenta de que he generado grandes lazos con mis compañeros y ahora buenos amigos de la universidad. Estos amigos han estado en cada parte del proceso de la universidad, comenzando desde el primer año hasta ahora. A esos amigos les agradezco mucho de todo corazón, siento que hemos crecido en conjunto y hemos estado en las buenas y en las malas, así que muchas gracias a ellos. Hablando de amigos no son los únicos que han estado, otros han estado desde el colegio, los cuales me han aconsejado más de una vez y me han hecho reflexionar muchos aspectos de mi vida logrando mejoras en ella también los tengo siempre presente y en alta estima.

A mis papás les agradezco también por siempre estar ahí, aunque ellos no lo crean a través de todas las dificultades y discusiones de este proceso los comprendía, pero me costo hacer un cambio para poder terminar la tesis. Sin embargo, a pesar de todo eso siempre me apoyaron y siempre se hacían notar, eso nunca lo olvidare y nunca dejare de agradecer todo lo que han hecho tanto por mí como mis hermanos.

A mis hermanos que, si bien no diría que me aportaron con la tesis, si me aportaron al hacerme sentir mejor y salir de los problemas de la rutina, además de que mi hermano siempre me daba buenos consejos.

Tambien me gustaría mencionar a mis abuelos que siempre se hacían presente con alguna llamada, para ir a verlos o almorzar juntos.

Y por último me gustaría dejar un agradecimiento a la universidad que me hizo crecer y ser una mejor persona en el ámbito laboral, nos enseñaron muchas cosas, pero por sobre todo le agradezco a mis profesores los cuales son los que dan la cara a la enseñanza y son los que realmente se preocupan por sus estudiantes y buscan lo mejor para ellos.

Resumen Ejecutivo

La productividad legislativa constituye un tema de creciente interés en Chile debido a la baja confianza ciudadana en el Congreso Nacional y a la necesidad de fortalecer los mecanismos de evaluación del desempeño parlamentario. Si bien existen indicadores tradicionales como la asistencia y la cantidad de proyectos presentados, estos no permiten explicar de manera integral las diferencias observadas en el desempeño de los diputados. En este contexto, la presente investigación tuvo como objetivo identificar los principales factores explicativos de la productividad legislativa de los diputados de la Cámara de Diputadas y Diputados de Chile durante el periodo 2022-2026 mediante la aplicación de modelos estadísticos y técnicas de aprendizaje automático.

El estudio adoptó un enfoque cuantitativo, no experimental y de carácter explicativo. Se trabajó con un censo correspondiente a los 157 diputados que ejercieron funciones durante las legislaturas 370 a 373. Como variable de estudio se construyó un índice de productividad legislativa que integró diversas dimensiones del trabajo parlamentario, incluyendo iniciativas legislativas, participación en comisiones, intervenciones y otras actividades propias del cargo. Paralelamente, se recopilieron variables individuales, académicas, políticas y de desempeño parlamentario, las cuales fueron sometidas a procesos de depuración, diagnóstico de multicolinealidad y selección mediante técnicas como VIF, LASSO, Elastic Net y Random Forest.

Posteriormente se estimaron y compararon distintos modelos predictivos, entre ellos regresión lineal, Ridge, LASSO, Elastic Net, Random Forest, Gradient Boosting, XGBoost y LightGBM. La evaluación se realizó utilizando métricas de desempeño como el coeficiente de determinación (R^2), el error cuadrático medio (RMSE) y análisis de robustez mediante múltiples particiones aleatorias de los datos.

Los resultados evidencian que las variables asociadas a la actividad legislativa presentan una capacidad explicativa considerablemente mayor que las características demográficas o académicas de los diputados. Asimismo, se observó que la asistencia parlamentaria, evaluada de manera aislada, posee una limitada contribución a la explicación sobre la productividad cuando se incorpora juntamente con otras variables. Entre los modelos evaluados, Ridge presentó el

mejor equilibrio entre capacidad predictiva, estabilidad e interpretabilidad, permitiendo identificar los factores con mayor afinidad con la productividad legislativa.

Los hallazgos obtenidos aportan evidencia empírica para comprender el desempeño parlamentario desde una perspectiva cuantitativa y ofrecen una metodología replicable para futuras investigaciones en gestión pública y análisis legislativo. Además, constituyen una base para el desarrollo de indicadores más robustos que permitan complementar las métricas tradicionales de evaluación parlamentaria y contribuir a una mayor transparencia, rendición de cuentas y fortalecimiento de la confianza ciudadana en las instituciones democráticas.

Índice General

Tabla de contenido

Tabla de contenido.....	6
1. Introducción.....	9
2. Objetivos.....	12
2.1. Objetivo General.....	12
2.2. Objetivos Específicos.....	12
3. Marco Teórico.....	13
3.1. Congreso de Chile.....	13
3.1.1. Cámaras del Congreso de Chile.....	13
3.1.2. Creación de una Ley.....	13
3.2. Marco Normativo y Regulatorio.....	14
3.2.1. Ley N° 21.233.....	14
3.2.2. Ley N° 21.603.....	15
3.2.3. Reglamentos internos de la Cámara de Diputados.....	15
3.2.4. Reglamentos internos del Senado.....	16
3.2.5. Comparaciones internacionales.....	17
3.3. Gestión Pública y Control de Desempeño.....	20
3.3.1. Enfoque de Nueva Gestión Pública (NGP):.....	20
3.3.2. Gobernanza y Responsabilidad:.....	21
3.3.3. Teoría de Incentivos:.....	21
3.4. Enfoques cuantitativos para la medición del desempeño legislativo.....	22
3.4.1. Medición del desempeño y modelización estadística.....	23
3.4.2. Modelos de regresión lineal.....	23
3.4.3. Regularización y selección de variables.....	24
3.4.4. Modelos no lineales y aprendizaje automático.....	25
3.4.5. Análisis de Robustez mediante Múltiples Semillas.....	27
3.4.6. Evaluación del desempeño de los modelos.....	28
3.5. Enfoques tradicionales de medición del desempeño organizacional.....	29
3.5.1. Indicadores de gestión y desempeño (KPIs).....	30
3.5.2. Aplicación de la ingeniería industrial en el sector público: control de procesos, eficiencia y mejora continua.....	30
3.5.3. Análisis de procesos: asistencia y registro parlamentario como flujo de trabajo.....	31
3.5.4. Balanced Scorecard (BSC): marco estratégico para la medición del desempeño.....	31
3.6. Percepción Ciudadana y Confianza Institucional.....	33
3.6.1. Estudios y encuestas sobre confianza en el Congreso Nacional (CEP, CADEM, Latinobarómetro). 33	

3.6.2.	Relación entre confianza ciudadana y desempeño legislativo percibido.	33
3.6.3.	Benchmarking institucional (comparar con parlamentos internacionales).	34
3.6.4.	Casos de éxito en el uso de métricas de gestión en organismos gubernamentales.	34
3.7.	Brecha entre asistencia formal y participación efectiva.....	35
3.7.1.	Concepto de asistencia formal vs. asistencia real (ej. llegada tardía, abandono de sesiones, ausencias parciales).....	35
3.7.2.	Limitaciones de los registros oficiales para reflejar la participación sustantiva.	35
3.7.3.	Impacto de esta brecha en la calidad del debate legislativo y en la percepción ciudadana.....	35
3.8.	Metodología Cross-Industry Standard Process for Data Mining (CRISP-DM).....	36
4	Metodología	39
4.1	Enfoque y Diseño de Investigación	40
4.2	Población y Muestra	40
4.2.1	Universo y Población.....	40
4.2.2	Tipo de Muestreo	41
4.2.3	Unidad de Análisis u Observación y Criterios de Selección.....	41
4.3	Definición y Operacionalización de Variables	41
4.3.1	Variables Dependientes	41
4.3.2	Variables Independientes.....	46
4.4	Fuentes y Recolección de Datos	51
4.5	Preparación y Limpieza de Datos	52
4.6	Estrategia de Modelado y Selección de Variables	52
4.7	Modelo Final y Validación	54
4.8	Herramientas y Software.....	54
4.9	Limitaciones Metodológicas	55
5	Resultados y Discusión	58
5.1	Análisis Descriptivo de los Datos	58
5.1.1	Composición Demográfica	58
5.1.2	Perfil Académico y Electoral	59
5.1.3	Distribución del Índice de Productividad y Afiliación Política	60
5.2	Diagnóstico Metodológico: Modelo con Datos Fallidos	61
5.3	Selección de Variables	62
5.3.1	Diagnóstico de Multicolinealidad (VIF)	62
5.3.2	Importancia de Variables según Random Forest	64
5.3.3	Selección por Consenso de Métodos Lineales	66
5.3.4	Variables Eliminadas por Otras Razones.....	67
5.4	Evaluación Comparativa de Modelos	68
5.5	Análisis de los Coeficientes del Modelo Ridge	70

5.6	Análisis Comparativo de los Diputados más Productivos	73
5.7	Perfil del Diputado Modelo: Características Individuales Asociadas a la Productividad Legislativa	75
6	Conclusiones	81
7	Recomendaciones	85
8	Referencias.....	88

Índice de Figuras

Ilustración 1: Balanced Scorecard.	32
Ilustración 2: Fases de la metodología CRISP-DM.	37
Ilustración 3. Importancia de variables obtenida mediante valores SHAP para el modelo Random Forest.	66

Índice de Tablas

Tabla 1. Puntajes asignados por el nivel educativo.....	47
Tabla 2. Indica todos los partidos políticos del periodo, junto con la coalición definida y si se le atribuye ser de la oposición u oficialismo.....	49
Tabla 3. Estadísticas descriptivas.....	58
Tabla 4. Variable académica y electoral.	60
Tabla 5. Valores del nivel de formación de los diputados con su cantidad de veces encontradas.....	60
Tabla 6. Estadísticas descriptivas de las variables de actividad parlamentaria.....	61
Tabla 7. Importancia de las variables en el modelo inicial de Random Forest.....	62
Tabla 8. Factor de inflación de la varianza (VIF) para todas las variables.	63
Tabla 9. Importancia de las variables por Random Forest.....	64
Tabla 10. Resumen de la selección de variables por consenso de métodos lineales.....	66
Tabla 11. Resultados y desempeño comparativo de los modelos estimados.	68
Tabla 12. Análisis de robustez con 30 semillas aleatorias.	69
Tabla 13. Coeficientes estandarizados del modelo Ridge, con la referencia de coalición centro para las coaliciones.....	70
Tabla 14. Diferencias porcentuales respecto al promedio general por cada grupo de productividad.	73
Tabla 15. Distribución de las coaliciones políticas por grupo de productividad y cantidades de diputados de dicha coalición dentro de los tops.	75

1. Introducción

Durante los años 2018 a 2022 en Chile se intensificó el debate sobre el funcionamiento de las instituciones políticas. Ese periodo, marcado por el denominado “estallido social”, no solo fue un momento de movilización ciudadana, sino también un punto de inflexión en la forma en que amplios sectores de la población perciben a los representantes públicos. Si bien una parte de las protestas se manifestó de manera pacífica, otra derivó en daños extendidos en infraestructura urbana. Este contexto generó una crítica sostenida hacia los actores políticos y, en particular, hacia los legisladores.

En el discurso público surgieron cuestionamientos respecto de aspectos concretos de la labor parlamentaria, entre ellos el nivel de asistencia, el ejercicio de licencias médicas bajo circunstancias controvertidas y el desempeño general de la función legislativa. El caso de las licencias médicas llamó la atención porque varios parlamentarios utilizaron este recurso para justificar ausencias prolongadas, incluso para realizar actividades distintas al descanso o tratamiento médico. Ese tipo de situaciones alimentaron la percepción de falta de compromiso con las responsabilidades propias del cargo.

Al mismo tiempo, el estallido social tensionó otras dimensiones del trabajo legislativo. Frente a un aumento de demandas ciudadanas en materia de seguridad pública, el Congreso tramitó diversas iniciativas relacionadas con ese ámbito. No obstante, varias de esas propuestas no lograron aprobación o quedaron sujetas a prórrogas extensas, lo que generó críticas adicionales desde distintos sectores de la opinión pública. Esta percepción negativa se refleja en las mediciones de opinión pública: según la Encuesta CEP N.º 89 (Centro de Estudios Públicos, 2023), el Congreso alcanzó solo un 8% de confianza ciudadana en 2023, siendo la institución política con peor evaluación junto a los partidos políticos. Este dato es consistente con los resultados de la Encuesta OCDE sobre Confianza en las Instituciones Públicas (OCDE, 2024), que sitúa a Chile entre los países con menor confianza parlamentaria dentro de la organización, con solo un 19% de la población declarando confianza alta o moderada en el parlamento nacional.

Ese conjunto de eventos plantea una cuestión central: cómo se comporta la productividad

legislativa en Chile y qué factores se asocian a ella. Para abordar ese problema es necesario, antes que nada, precisar el sentido de “productividad” en este contexto.

En términos generales, en el ámbito económico la productividad se entiende como la relación entre la cantidad de bienes o servicios producidos y los recursos utilizados para generarlos, es decir, como una medida de eficiencia en la utilización de insumos para obtener resultados (OCDE, 2001). En palabras de Krugman (1990), “la productividad no lo es todo, pero en el largo plazo es casi todo”, en el sentido de que la capacidad de una economía para mejorar el bienestar de su población depende en gran medida de su capacidad para producir más con los mismos recursos. Como referencia divulgativa del concepto, Banco Santander (2025) lo resume como “la cantidad de producto que obtiene una empresa en una hora de trabajo”. Si trasladamos estas ideas al Congreso de Chile, la productividad legislativa puede entenderse como la capacidad efectiva de los diputados para generar resultados normativos concretos, vale decir, leyes aprobadas, durante un período determinado. En este contexto, Volden y Wiseman (2014) proponen medir la productividad legislativa individual como la capacidad de cada legislador de hacer avanzar sus propias iniciativas hasta convertirlas en ley, distinguiendo entre quienes lideran el proceso y quienes simplemente lo apoyan, distinción que fundamenta la construcción del índice utilizado en esta investigación.

A partir de esta noción, se identifica una problemática concreta para la sociedad chilena: cuáles variables se asocian en mayor medida sobre la productividad legislativa de los diputados del congreso chileno. Por lo que este trabajo investigará diversas características individuales y de desempeño parlamentario que podrían explicar variaciones en productividad.

En torno a esta pregunta general surgen interrogantes más específicos que guiarán el desarrollo de esta investigación, los cuales se presentan a continuación:

- ¿Cómo se manifiesta la productividad legislativa de los diputados en Chile durante el periodo 2022-2026 cuando se mide a partir de indicadores cuantitativos de desempeño parlamentario?
- ¿Qué características individuales de los diputados, tales como sus estudios, experiencias previas y su comportamiento legislativo, se asocian con mayores niveles de productividad legislativa?

- ¿La asistencia parlamentaria contribuye en la explicación de la productividad legislativa cuando se considera de manera conjunta con otras variables explicativas?
- ¿Qué variables presentan mayor capacidad explicativa sobre la productividad legislativa al ser integradas en modelos predictivos?
- ¿Existen patrones diferenciados de productividad legislativa entre diputados con perfiles académicos y conductuales similares, pero con distinto nivel de asistencia?

2. Objetivos

2.1. Objetivo General

Analizar la productividad legislativa de los diputados en Chile durante el periodo 2022-2026 mediante la construcción de indicadores cuantitativos y el uso de modelos predictivos, con el fin de identificar qué características individuales se asocian de manera significativa en el desempeño legislativo.

2.2. Objetivos Específicos

Construir una medida sintética de productividad legislativa que permita evaluar de forma comparable el desempeño de los diputados, integrando distintas dimensiones del trabajo legislativo.

Sistematizar y cuantificar las variables de identificación, afiliación política y la actividad parlamentaria de los diputados, tales como antecedentes formativos, trayectoria y comportamiento en el ejercicio legislativo, con el fin de incorporarlas como variables explicativas del desempeño.

Evaluar la relevancia predictiva de variables de comportamiento parlamentario, incluyendo asistencia, puntualidad y participación, dentro del conjunto de factores explicativos del desempeño legislativo.

Aplicar modelos estadísticos y técnicas de aprendizaje automático para estimar la capacidad explicativa y predictiva de las variables seleccionadas sobre la productividad legislativa.

Comparar el desempeño de modelos lineales y no lineales con el objetivo de identificar qué enfoques permiten explicar de mejor manera la variabilidad observada en la productividad legislativa.

3. Marco Teórico

3.1. Congreso de Chile

El Congreso Nacional de Chile constituye el órgano legislativo del Estado y está compuesto por dos cámaras: la Cámara de Diputadas y Diputados y el Senado. Ambas corporaciones participan en la tramitación de las leyes y en el control político del Ejecutivo, cumpliendo funciones esenciales dentro del sistema democrático chileno.

3.1.1. Cámaras del Congreso de Chile

La cámara de diputados está compuesta de 155 representantes que provienen de 28 distritos electorales, los cuales tienen una duración en el cargo de 4 años. Dentro de sus responsabilidades están la de legislar, representar a los ciudadanos de su distrito y el de fiscalizar los actos del gobierno a través de la realización de acuerdos, la creación de comisiones investigadoras, entre otros.

La cámara de senadores está compuesta de 50 representantes, los que contemplan una duración de 8 años en el cargo. Sus responsabilidades son las de legislar y el de aprobar tratados internacionales, entre otras.

Estas funciones convierten a ambas cámaras en espacios centrales de producción normativa y deliberación política, lo que justifica la necesidad de evaluar el desempeño individual y colectivo de sus miembros más allá del cumplimiento formal de la asistencia.

3.1.2. Creación de una Ley

Para poder generar una ley en Chile se debe realizar ciertos pasos para que se pueda efectuar su creación.

En primer lugar, se debe realizar una propuesta de proyecto a ley, esta debe ser realizada a través de los parlamentarios o bien del presidente de la república.

En segundo lugar, se ingresa la moción a la cámara que haya iniciado el proceso o en caso de que sea uno de los siguientes casos se tramita inicialmente en determinada cámara, estos casos son los siguientes: Si se trata de una ley con relación a impuestos, presupuestos de Estado, contribuciones y reclutamiento de tropas el proceso inicia en la cámara de diputados. Y si es sobre amnistía e indulto en general se inicia el proceso en la cámara del senado. Luego de ingresar la propuesta, se discute si se aprueba la idea de legislar dicha propuesta, para

posteriormente una discusión particular donde se vota artículo por artículo. Si es aprobado pasa a la siguiente cámara. Si no es aprobado el proyecto de ley es rechazado y para una nueva instancia se debe realizar una nueva propuesta e iniciar el proceso desde cero.

En tercer lugar, se ingresa la propuesta de ley aprobada por la cámara anterior, y entra en el mismo proceso que el anterior punto, donde de ser aprobada pasa a manos del presidente de la república de Chile. En caso contrario puede haber dos caminos, el primero de estos es que en la revisión general de la propuesta se rechace ésta, por lo que genera un veto de la idea, en el segundo caso si es rechazada en la discusión particular esta se devuelve a la cámara anterior con diversos cambios presentados por la segunda cámara y se vuelve al punto anterior donde se vuelve a discutir y si es aprobada vuelve a la segunda cámara. Si en este proceso nuevamente se encuentran diferencia, se genera una comisión mixta de diputados y senadores, la cual propone un texto de consenso para una votación final.

En cuarto lugar, se genera la instancia del presidente, el cual tiene dos opciones, o aprueba dicho proyecto de ley o lo devuelve con modificaciones.

Si ambas cámaras aprueban las observaciones el proyecto de ley se aprueba con esas observaciones. En caso contrario se puede realizar una votación en la que, si 2/3 de los miembros de cada cámara están de acuerdo, el presidente debe aprobar el proyecto de ley como esta. Si no se cumplen los 2/3 no hay ley en los puntos en discrepancia.

3.2. Marco Normativo y Regulatorio

Inicialmente se necesita conocer bajo que conceptos el congreso nacional se rige, para su correcto funcionamiento de sus funciones y leyes que les afectan en el ámbito monetario.

3.2.1. Ley N° 21.233

La Ley 21.233 es una modificación a la constitución política de Chile donde se estipula que se va a crear una comisión que tiene como deberes el definir cuál va a ser el monto o sistema de remuneraciones para las autoridades a cargo del país mencionadas en el artículo 38 bis de la constitución. Además, se estipula que “los diputados y senadores percibirán como única renta una dieta equivalente a la remuneración de un ministro de Estado” (MINSEGPRES, 2020). Y se añade una disposición transitoria que indica que el Consejo de Alta Dirección Pública deberá fijar por única vez las remuneraciones de los diputados y senadores y ministros de Estado, hasta que la comisión defina éstas cuando tenga las facultades.

El artículo 38 bis de la constitución de Chile dice así “Las remuneraciones del

Presidente de la República, de los senadores y diputados, de los gobernadores regionales, de los funcionarios de exclusiva confianza del Jefe del Estado que señalan los números 7° y 10° del artículo 32 y de los contratados sobre la base de honorarios que asesoren directamente a las autoridades gubernativas ya indicadas, serán fijadas, cada cuatro años y con a lo menos dieciocho meses de anticipación al término de un período presidencial, por una comisión cuyo funcionamiento, organización, funciones y atribuciones establecerá una ley orgánica constitucional.” (MINSEGPRES, 2020)

La comisión estará conformada por

- a) Un exministro o exministra de Hacienda.
- b) Un exconsejero o exconsejera del Banco Central.
- c) Un excontralor o una excontralora o un ex subcontralor o una exsubcontralora de la Contraloría General de la República.
- d) Un expresidente o una expresidenta de una de las ramas que integran el Congreso Nacional.
- e) Un exdirector o una exdirectora Nacional del Servicio Civil.

3.2.2. Ley N° 21.603

La Ley 21.603 regula el funcionamiento, organización, funciones y atribuciones de la comisión para la fijación de remuneraciones que indica el artículo 38 bis de la constitución política de la república.

En esta ley se fija la creación y su objetivo de la comisión para la fijación de remuneraciones, también las funciones y atribuciones que tendrá y por último su organización y el funcionamiento que desarrollará.

3.2.3.Reglamentos internos de la Cámara de Diputados

El Reglamento de la Cámara de Diputadas y Diputados establece las normas que regulan el funcionamiento interno de la corporación, la organización de sus sesiones, las atribuciones de las comisiones y los deberes de los parlamentarios. Entre sus disposiciones más relevantes se encuentran aquellas que regulan la asistencia y participación de los diputados en las sesiones de sala y de comisiones, materias esenciales para el control de su desempeño legislativo.

De acuerdo con el artículo 7° del Reglamento, la asistencia a las sesiones de la Cámara es obligatoria para todos los diputados, salvo justificación debidamente acreditada. Asimismo, el artículo 8° establece que la presidencia de la Cámara puede aplicar sanciones a quienes no concurren o abandonen la sesión sin causa justificada. En particular, se

dispone que los diputados deben registrarse al momento del inicio y término de las sesiones, lo que da lugar a la elaboración de un registro oficial de asistencia, actualmente disponible al público en el sitio web institucional.

El reglamento también regula la disciplina parlamentaria, estableciendo que la reiteración de inasistencias o conductas que obstaculicen el funcionamiento de la Cámara pueden ser sancionadas con descuentos proporcionales en las dietas parlamentarias. Sin embargo, estas sanciones se aplican únicamente en casos de ausencia total, sin considerar la impuntualidad o la salida anticipada de los diputados, lo que constituye una limitación del sistema actual de control de asistencia.

Por último, el reglamento contempla la posibilidad de dar justificaciones de inasistencias por diversos motivos, estos pueden ser los siguientes: Licencia médica (problemas de salud), Impedimento grave (fallecimiento o accidente de algún familiar directo), Misión oficial (salida del país o dentro del territorio nacional), Permisos especiales (acordados por los comités parlamentarios), Permisos parentales(permisos postnatales, prenatales y parentales), Gestiones encomendadas por la comisiones (actividades relacionadas con el cargo, pero no siendo la sesión), Condiciones de riesgo (como por ejemplo el COVID-19), estas justificaciones deben ser acreditadas ante la Mesa de la Cámara. No obstante, la regulación no especifica mecanismos de control automatizado ni sanciones por atrasos, lo que ha generado críticas desde sectores que abogan por una mayor transparencia y rendición de cuentas en el desempeño parlamentario (Cámara de Diputadas y Diputados de Chile, 2023).

3.2.4.Reglamentos internos del Senado

El Reglamento del Senado de la República de Chile cumple funciones análogas, regulando el desarrollo de las sesiones, la organización de las comisiones y el comportamiento de los senadores. En relación con la asistencia, el artículo 23 del Reglamento dispone que los senadores tienen la obligación de asistir puntualmente a las sesiones de sala y de comisión, registrándose su presencia al inicio y término de cada una. Dichos registros conforman la base de los informes de asistencia publicados periódicamente en la página oficial del Senado.

En caso de inasistencia injustificada, el artículo 24 establece sanciones que pueden traducirse en multas pecuniarias, equivalentes a una fracción de la dieta parlamentaria, y en la pérdida del derecho a percibir remuneración por el día no trabajado. Sin embargo, de manera similar a la Cámara de Diputados, el reglamento no contempla sanciones por

retrasos o salidas anticipadas, ni un control detallado sobre la permanencia efectiva durante toda la sesión.

Además, el Reglamento del Senado permite justificar las ausencias por razones de salud, comisión de servicios (misiones oficiales) o fuerza mayor, con previa comunicación o un escrito a la mesa directiva del senado. La asistencia se considera cumplida con la mera inscripción en el registro electrónico, lo que significa que un senador puede ser contabilizado como presente incluso si su participación fue parcial o mínima.

El marco normativo del Senado, si bien promueve la transparencia mediante la publicación de los registros de asistencia, no incorpora herramientas de evaluación cualitativa del desempeño legislativo, tales como el nivel de participación en debates o la asistencia efectiva a las comisiones. Esta situación refleja una limitación estructural del sistema, centrado en un control de cumplimiento formal más que en la medición de la calidad del trabajo parlamentario (Senado de la República de Chile, 2024).

3.2.5. Comparaciones internacionales

Con el propósito de identificar prácticas relevantes para el control de asistencia y la evaluación del desempeño parlamentario, se analizan a continuación las reglas y prácticas de cinco países con tradiciones parlamentarias diversas, pero que contemplan los siguientes dos criterios: son bicamerales y tienen una población parecida o mayor a la de Chile y estos son: Alemania, Australia, Canadá, España y Japón. Para cada caso se menciona la normativa resumida sobre la asistencia, la disponibilidad pública de registros, y la existencia de mecanismos disciplinarios o definiciones ampliadas de “presencia” que vayan más allá de la mera firma en el registro.

En Alemania está el parlamento federal llamado Bundestag, donde la asistencia en los días de sesión está regulada por las normas internas y por la ley sobre el estatuto jurídico de los miembros del Bundestag, que establece la obligación de participar en el trabajo del parlamento y la existencia de un registro de asistencia que los miembros deben firmar en cada jornada de sesiones. La presidencia y la normativa legal contemplan consecuencias administrativas por no inscribirse o por no participar en votaciones nominales, sin embargo, el control suele centrarse en la constancia de la firma y la participación en votaciones, más que en el registro minuto a minuto de puntualidad o permanencia durante toda la sesión.

En Australia están las Standing Orders de la Cámara de Representantes y del Senado quienes precisan que el oficio del secretario mantiene las listas y los registros diarios de la

asistencia, los cuales se consignan en las Votes and Proceedings (votos y procedimientos de la sesión) y en los volúmenes de sesión que se publican oficialmente. Los nombres de los miembros presentes se toman en la cámara por el Serjeant-at-Arms (sargento de armas quien pone orden en la sesión) y se registran las ausencias en las actas oficiales, de este modo existe una publicación formal y periódica del registro de asistencia, si bien las reglas se orientan al registro de presencia y a la decencia del procedimiento más que a sanciones automáticas por tardanzas puntuales.

El Parlamento de Canadá (House of Commons) establece en sus Standing Orders (reglas de la casa) la obligación general de asistencia de los miembros y mantiene mecanismos formales de registro y de decoro durante la sesión. La normativa cubre, además, la correlación entre asistencia y otras actividades parlamentarias (por ejemplo, funciones en comisiones o misiones oficiales). Las reglas canadienses distinguen la asistencia a efectos de procedimiento y reconocen distintas formas de actividad parlamentaria que pueden considerarse “trabajo” más allá de la mera presencia en el Pleno, lo que permite un cómputo más amplio del desempeño del legislador.

En el caso de España el reglamento del Congreso de los Diputados dispone el deber de asistir a las sesiones del pleno y de las comisiones, y publica normas de registro y control de la actividad parlamentaria. En años recientes se han impulsado reformas reglamentarias para reforzar la transparencia y sancionar conductas que alteren el normal desarrollo de la actividad, así como para regular modalidades de votación remota en supuestos específicos. No obstante, la práctica normativa española continúa basándose, en lo esencial, en el registro formal de presencia y en justificativos reglamentarios para ausencias.

En el caso Japón están las normas del Diet (House of Representatives y House of Councillors) las cuales establecen horarios y procedimientos estrictos para las sesiones: la apertura de sesión se fija en horas concretas y los miembros deben estar presentes conforme a las disposiciones del reglamento de cada cámara. El sistema normativo japonés enfatiza la puntualidad y la asistencia, con reglas claras sobre la apertura de sesiones y la comprobación de la presencia de los miembros, aunque la información pública disponible tiende a centrarse en la convocatoria y el calendario más que en registros públicos minuto a minuto de la permanencia en sala.

Dentro de lo que se puede concatenar es que todos los parlamentos analizados disponen de registros oficiales de asistencia y los publican como parte de las actas o los votos y procedimientos (Australia, Canadá, España, Alemania, Japón). La existencia de

registros públicos constituye una práctica común que favorece la transparencia institucional.

Algunos sistemas como el de Canadá reconocen formalmente que la actividad parlamentaria puede incluir trabajo en comisiones u otras funciones que equivalen a “atender asuntos parlamentarios”, lo cual permite un cómputo más fiel del desempeño que la mera presencia en el pleno. La adopción de definiciones amplias reduce el sesgo de medir solo el tiempo físico en sala.

La mayoría de los reglamentos se concentra en la constancia de firma o del registro de asistencia y en la participación en votaciones, pocos ordenamientos automatizan el control de puntualidad o la permanencia continua sin controversias sobre privacidad o naturaleza representativa del cargo. En este punto, la experiencia comparada revela una tensión entre mayor granularidad en el control y preocupaciones operativas o políticas.

Las sanciones por inasistencia o conducta pueden existir (por ejemplo, descuentos o medidas disciplinarias), pero su alcance y frecuencia varían: algunos parlamentos priorizan sanciones administrativas y comités de conducta (Reino Unido o Canadá en su práctica), mientras que otros se limitan a descuentos o registros públicos que operan como control social. La adopción de sanciones por tardanza es poco frecuente entre los reglamentos de los países analizados.

Las prácticas internacionales confirman la viabilidad de publicar y supervisar registros de asistencia, pero muestran también que la mera publicación no garantiza la medición de la participación efectiva. Para mejorar la precisión del control en Chile, las lecciones más útiles son: (a) ampliar la definición de presencia para incorporar actividades parlamentarias relevantes (comisiones, votaciones, intervenciones), (b) valorar la implementación de registros que consideren puntualidad y permanencia cuando esto sea técnicamente viable y políticamente aceptable, y (c) diseñar mecanismos disciplinarios y de incentivos que equilibren la transparencia con garantías de debido proceso y representatividad.

El análisis de los sistemas parlamentarios descritos se basa en los reglamentos internos publicados por cada cámara, cuya sistematización comparada ha sido documentada por el Parlamento Europeo (1997) en su estudio sobre el control de asistencia en parlamentos nacionales, y por la base de datos Parline de la Unión Interparlamentaria (IPU, 2022), que registra el funcionamiento institucional de los parlamentos de más de 190 países. Dentro de lo que se puede concatenar, todos los parlamentos analizados disponen

de registros oficiales de asistencia y la publicación de estos registros públicos constituye una práctica común que favorece la transparencia institucional (Parlamento Europeo, 1997).

3.3. Gestión Pública y Control de Desempeño

En el siglo XIX, Wilson Woodrow, quien es considerado el padre de la administración pública, planteó en su publicación “*The Study of Administración*” en la revista *Political Science Quarterly* que debía existir una separación clara entre la política pública, liderada por representantes electos, y la administración pública, la cual debiese ser ejecutada por profesionales cualificados. Esta idea sentó las bases para entender a la administración pública como una disciplina propia.

Posteriormente, diversos autores y corrientes fueron nutriendo esta visión, hasta llegar a lo que se conoce como Nueva Gestión Pública (NGP), considerada una de las olas de reforma más relevantes en la historia contemporánea del sector público.

3.3.1. Enfoque de Nueva Gestión Pública (NGP):

Este nuevo modelo surgió entre los años 1980 y 1990 en los países anglosajones y fue rápidamente globalizado. Hood (1991) lo caracteriza como un conjunto de doctrinas administrativas orientadas a la introducción de métodos gerenciales del sector privado en el sector público, con énfasis en la medición de resultados, la definición de estándares explícitos de desempeño, la rendición de cuentas basada en outputs y la orientación hacia el usuario. En su formulación clásica, Hood (1991) identifica siete elementos doctrinales del modelo, entre los que destacan la gestión profesional activa, la medición explícita del desempeño mediante indicadores y la disciplina en el uso de recursos. Pollitt y Bouckaert (2004) amplían este marco señalando que la NGP no es un modelo uniforme sino un conjunto de tendencias reformistas que se implementaron de formas distintas según el contexto institucional de cada país, pero que comparten el objetivo central de hacer más eficiente y transparente la acción pública. En Chile, esta corriente inspiró reformas como el uso de instrumentos de evaluación con incentivos económicos desde 1998 y la creación del sistema de Alta Dirección Pública (DIPRES, 2019).

En Chile esta corriente inspiró reformas como el uso de instrumentos de evaluación con incentivos económicos desde 1998 y la creación del sistema de alta dirección pública en 2004, con el fin de profesionalizar la administración estatal y mejorar la eficiencia del gasto público

La principal innovación del modelo NGP es el uso de indicadores de desempeño, el realizar evaluaciones a los resultados, mostrar responsabilidad y transparencia a la ciudadanía, e ir metódicamente avanzando por medio de objetivos.

3.3.2. Gobernanza y Responsabilidad:

El concepto de gobernanza se refiere al modo en que las instituciones públicas, privadas y la sociedad civil interactúan para la toma de decisiones colectivas y la gestión de los recursos. En el ámbito legislativo, la gobernanza implica no solo la organización interna del Congreso, sino también los mecanismos que aseguran que su labor se realice con transparencia, legitimidad y apertura hacia la ciudadanía.

Asociado a la gobernanza está el concepto de responsabilidad pública o accountability, entendido como la obligación de los representantes de rendir cuentas por su desempeño y de someterse a mecanismos de evaluación y control. Schedler (1999) distingue dos dimensiones fundamentales del accountability: la answerability, que corresponde a la obligación de los representantes de informar y justificar sus decisiones ante quienes los mandatan, y la enforcement, que corresponde a la capacidad de imponer sanciones cuando esa obligación no se cumple. En el contexto parlamentario, la primera dimensión está parcialmente cubierta por los registros públicos de asistencia y votación, pero la segunda, vale decir, la capacidad de sancionar el desempeño deficiente mediante mecanismos institucionales formales es notoriamente débil en el sistema chileno. Por tanto, la aplicación de un enfoque de gobernanza más robusto permitiría avanzar hacia sistemas de control de desempeño que integren ambas dimensiones, con un impacto positivo en la confianza ciudadana y en la legitimidad institucional (OCDE, 2017).

Por tanto, la aplicación de un enfoque de gobernanza más robusto permitiría avanzar hacia sistemas de control de desempeño que integren nuevas métricas, generando un impacto positivo en la confianza ciudadana y en la legitimidad institucional.

3.3.3. Teoría de Incentivos:

La teoría de incentivos, desarrollada en su formulación moderna por Laffont y Tirole (1993), estudia cómo diseñar mecanismos de recompensas y sanciones que orienten el comportamiento de los agentes hacia los objetivos del principal, en contextos donde existe asimetría de información entre ambas partes. En el ámbito político, los ciudadanos actúan como principal y los representantes electos como agentes, generando una relación en la que el electorado delega poder, pero enfrenta dificultades para observar directamente el esfuerzo y las decisiones del legislador (Gailmard, 2012). Esta asimetría informacional es

precisamente la que justifica el diseño de indicadores de desempeño parlamentario: al hacer visible el comportamiento legislativo mediante métricas objetivas, se reduce la brecha de información y se habilitan mecanismos de accountability electoral más efectivos. La teoría de incentivos sostiene que las conductas de los individuos pueden orientarse a través de dichos mecanismos cuando estos están correctamente alineados con los resultados esperados (Laffont y Tirole, 1993). En el contexto parlamentario, la actual estructura de incentivos se centra en la asistencia formal como requisito básico, dejando fuera variables más sustantivas del desempeño real.

La teoría de incentivos sostiene que las conductas de los individuos pueden orientarse a través de mecanismos de recompensas y sanciones. En el sector público, esta teoría adquiere particular relevancia, ya que permite alinear los intereses individuales de los funcionarios con los objetivos colectivos de eficiencia y servicio a la ciudadanía.

En el contexto parlamentario, la actual estructura de incentivos se encuentra centrada en la asistencia formal como requisito básico, lo que puede inducir a comportamientos estratégicos, como llegar tarde o retirarse temprano sin que ello afecte la evaluación final. Desde una perspectiva profesional, este diseño es insuficiente, ya que premia el cumplimiento mínimo en lugar de promover la participación efectiva.

Aplicar la teoría de incentivos permitiría diseñar un sistema en el cual las remuneraciones o sanciones no se vinculen únicamente a la asistencia, sino también a factores como la puntualidad, la permanencia y la participación en los debates. Este rediseño podría mejorar significativamente la eficiencia legislativa y fortalecer la confianza hacia las personas.

Con esto en mente, la formulación de sistemas de incentivos o sanciones requiere previamente la construcción de métricas objetivas y comparables de desempeño legislativo, lo que justifica el uso de herramientas cuantitativas de análisis y modelización.

3.4. Enfoques cuantitativos para la medición del desempeño legislativo

La medición del desempeño legislativo presenta desafíos analíticos relevantes debido a la presencia simultánea de múltiples factores que contribuyen a explicar el resultado del trabajo parlamentario. Estos factores no responden a una única dimensión, sino que combinan elementos institucionales, individuales y contextuales, lo que dificulta su análisis mediante indicadores simples o aproximaciones de solo una variable. Frente a este escenario, el uso de enfoques cuantitativos multivariados permite capturar relaciones más complejas, evaluar el peso relativo de distintos factores y contrastar distintos supuestos

explicativos bajo un marco analítico consistente.

En este contexto, los modelos estadísticos y de aprendizaje automático se presentan como herramientas complementarias, donde los enfoques lineales funcionan como punto de partida interpretativo y los modelos no lineales permiten explorar estructuras más flexibles y potencialmente más cercanas al fenómeno observado.

3.4.1. Medición del desempeño y modelización estadística

La modelización estadística constituye una estrategia central para el análisis del desempeño en contextos institucionales complejos. En términos generales, este enfoque busca representar un fenómeno de interés mediante relaciones formales entre una variable que sintetiza el desempeño y un conjunto de factores explicativos, permitiendo evaluar asociaciones, magnitudes relativas y patrones generales en los datos.

En el ámbito del análisis legislativo, la modelización estadística permite abordar preguntas relacionadas con el desempeño de los parlamentarios desde una perspectiva empírica, superando enfoques basados exclusivamente en percepciones, juicios normativos o indicadores aislados. Al considerar simultáneamente múltiples dimensiones del trabajo parlamentario, estos modelos permiten controlar la influencia de distintos factores y evaluar su contribución relativa dentro de un mismo marco analítico.

Una ventaja adicional de este enfoque es su capacidad para identificar problemas habituales en el análisis de datos sociales, como la redundancia entre variables, la presencia de relaciones débiles o la concentración del efecto explicativo en un número reducido de factores. De este modo, la modelización estadística no solo cumple una función explicativa, sino también exploratoria, orientando decisiones metodológicas posteriores, como la selección de variables o la comparación entre distintos modelos.

En investigaciones aplicadas a la gestión pública y al estudio de instituciones políticas, este tipo de herramientas ha demostrado ser especialmente útil para analizar desempeño, eficiencia y resultados, siempre que se cuente con información estructurada y comparable. En contextos de análisis institucional y político, la modelización estadística se utiliza como base para la construcción y evaluación de índices de desempeño que permiten comparar unidades de análisis de forma sistemática.

3.4.2. Modelos de regresión lineal

Los modelos de regresión lineal constituyen el punto de partida clásico en el análisis cuantitativo del desempeño, debido a su claridad interpretativa y a la posibilidad de evaluar relaciones directas entre variables explicativas y una variable dependiente.

En su forma general, el modelo de regresión lineal múltiple puede expresarse como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

donde y_i representa la variable de desempeño, x_{ij} las variables explicativas, β_j los parámetros a estimar y ε_i el término de error.

Desde una perspectiva teórica, este modelo permite interpretar cada coeficiente como el efecto marginal promedio de una variable, manteniendo constantes las demás. Sin embargo, su validez descansa en supuestos clásicos como linealidad, independencia, homocedasticidad y ausencia de multicolinealidad, los cuales rara vez se cumplen plenamente en contextos institucionales complejos como el trabajo parlamentario.

Por esta razón, la regresión lineal se utiliza en este estudio principalmente como modelo inicial y de referencia, más que como herramienta final de explicación.

3.4.3. Regularización y selección de variables

Uno de los principales desafíos en modelos lineales multivariados es la multicolinealidad, es decir, la existencia de correlaciones elevadas entre variables explicativas. Este problema aumenta la inflación en las varianzas de los coeficientes estimados, reduce la estabilidad del modelo y dificulta la interpretación de los efectos individuales.

- Multicolinealidad y Factor de Inflación de la Varianza (VIF)

El Factor de Inflación de la Varianza (VIF) es una medida diagnóstica utilizada para detectar multicolinealidad. Se define como:

$$VIF_j = \frac{1}{1 - R_j^2}$$

donde R_j^2 corresponde al coeficiente de determinación obtenido al regredir la variable x_j sobre el resto de las variables explicativas.

Valores de VIF superiores a 5 o 10 suelen interpretarse como indicativos de multicolinealidad problemática.

En este estudio, el VIF se emplea como herramienta exploratoria inicial, permitiendo identificar redundancias antes de aplicar técnicas más avanzadas de regularización.

- LASSO y Ridge como control de complejidad

Para abordar de manera más robusta la multicolinealidad y el sobreajuste, se utilizan modelos de regresión regularizada.

El modelo Ridge introduce una penalización cuadrática sobre los coeficientes:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^k \beta_j^2 \right\}$$

Esta penalización reduce la magnitud de los coeficientes, pero no los elimina completamente.

El modelo LASSO, en cambio, utiliza una penalización absoluta:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^k |\beta_j| \right\}$$

La principal ventaja del LASSO es que puede llevar coeficientes exactamente a cero, funcionando simultáneamente como método de estimación y selección automática de variables. Por esta razón, en el estudio el LASSO tiene un peso metodológico mayor que el VIF, al permitir identificar subconjuntos relevantes de variables de forma empírica y controlada.

3.4.4. Modelos no lineales y aprendizaje automático

Dado que el trabajo parlamentario está asociado por múltiples factores que interactúan entre sí de manera no necesariamente aditiva ni proporcional, resulta razonable incorporar modelos capaces de capturar relaciones no lineales y efectos combinados entre variables. En este contexto, los modelos basados en árboles de decisión y sus extensiones constituyen una alternativa metodológica relevante.

Árboles de decisión

Un árbol de decisión es un modelo predictivo que segmenta el espacio de variables explicativas en regiones homogéneas a través de una secuencia de reglas binarias. En cada nodo del árbol se selecciona una variable y un punto de corte que maximizan la reducción de la varianza del error (en problemas de regresión) o la pureza de los nodos (en problemas de clasificación).

Formalmente, el modelo puede entenderse como una función por tramos:

$$\hat{y}(x) = \sum_{m=1}^M c_m \cdot I(x \in R_m)$$

donde R_m representa una región del espacio de variables definida por el árbol y c_m es el valor promedio de la variable dependiente en dicha región.

Los árboles de decisión presentan una alta interpretabilidad, ya que su estructura permite visualizar las reglas que conducen a cada predicción. Sin embargo,

un único árbol suele ser inestable y propenso al sobreajuste, lo que motiva el uso de modelos de ensamble.

Random Forest

Random Forest es un modelo de ensamble que combina un gran número de árboles de decisión construidos sobre muestras bootstrap del conjunto de datos y utilizando subconjuntos aleatorios de variables en cada división. La predicción final se obtiene como el promedio de las predicciones individuales de los árboles:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

Este enfoque reduce la varianza del modelo, mejora su capacidad de generalización y permite capturar interacciones complejas entre variables sin necesidad de especificarlas explícitamente.

Gradient Boosting

Gradient Boosting construye modelos de forma secuencial, donde cada nuevo árbol se ajusta para corregir los errores cometidos por el conjunto de árboles anteriores. El modelo puede representarse como:

$$F_m(x) = F_{m-1}(x) + v \cdot h_m(x)$$

donde $h_m(x)$ corresponde a un árbol ajustado al gradiente de la función de pérdida y v es la tasa de aprendizaje.

Este enfoque suele alcanzar un alto desempeño predictivo, aunque requiere un mayor control de hiperparámetros para evitar sobreajuste y presenta una menor interpretabilidad estructural que los modelos lineales.

XGBoost (Extreme Gradient Boosting)

XGBoost es una implementación optimizada del algoritmo Gradient Boosting desarrollada por Chen y Guestrin (2016). A diferencia del Gradient Boosting clásico, XGBoost incorpora regularización L1 y L2 directamente en la función objetivo, lo que reduce el riesgo de sobreajuste y mejora la capacidad de generalización del modelo. Adicionalmente, utiliza una aproximación de segundo orden del gradiente para el ajuste de los árboles, lo que acelera la convergencia y permite un entrenamiento eficiente incluso en conjuntos de datos de alta dimensionalidad.

LightGBM (Light Gradient Boosting Machine)

LightGBM es un framework de Gradient Boosting desarrollado por Ke et al. (2017) en Microsoft Research. Su principal innovación es la incorporación de dos

técnicas: el muestreo unilateral basado en gradiente (GOSS), que reduce el número de instancias utilizadas en el cálculo de la ganancia de información, y el agrupamiento exclusivo de características (EFB), que reduce la dimensionalidad del espacio de variables. Como resultado, LightGBM alcanza velocidades de entrenamiento significativamente mayores que XGBoost y Gradient Boosting clásico, manteniendo una precisión comparable.

Permutation Importance

La importancia por permutación (permutation importance) es una medida de la relevancia de cada variable en un modelo entrenado, introducida por Breiman (2001) en el contexto de los Random Forests. La idea central consiste en medir el incremento del error de predicción del modelo cuando los valores de una variable son aleatoriamente permutados, rompiendo así la relación entre dicha variable y la respuesta. Una variable es considerada importante si su permutación genera un aumento significativo del error, lo que indica que el modelo dependía de ella para realizar predicciones precisas. A diferencia de la importancia basada en la disminución de impureza, la permutation importance es independiente de la estructura interna del modelo y puede aplicarse a cualquier algoritmo de aprendizaje automático, lo que la convierte en una herramienta de interpretación agnóstica al modelo. Puede utilizarse en combinación con la importancia intrínseca del Random Forest y los valores SHAP para obtener una triangulación robusta de la relevancia de las variables explicativas.

3.4.5. Análisis de Robustez mediante Múltiples Semillas

Una limitación frecuente en la evaluación de modelos predictivos con muestras pequeñas es la dependencia del resultado respecto a la partición particular de los datos en conjuntos de entrenamiento y prueba. Para mitigar este problema, se puede implementar un análisis de robustez que consiste en repetir el entrenamiento y la evaluación de cada modelo con múltiples semillas aleatorias distintas, manteniendo constante la proporción de partición entre entrenamiento y prueba, y variando únicamente la asignación de observaciones a cada conjunto. Típicamente se reporta la media, la desviación estándar, el mínimo y el máximo del R^2 obtenido en cada semilla, lo que permite evaluar tanto el desempeño promedio de cada modelo como su estabilidad ante distintas configuraciones de los datos. Este enfoque complementa la validación cruzada y proporciona evidencia más sólida sobre qué modelos ofrecen resultados confiables más allá de una partición

específica.

3.4.6. Evaluación del desempeño de los modelos

La evaluación de los modelos cumple un doble objetivo: por una parte, permite comparar el desempeño predictivo de distintas especificaciones, por otra, contribuye a comprender en qué medida las variables incluidas aportan información relevante al análisis del desempeño legislativo. Para ello, se emplean métricas cuantitativas estándar y herramientas de interpretación de modelos complejos.

Antes de introducir las métricas específicas, es importante destacar que ninguna medida por sí sola es suficiente. En contextos aplicados, la evaluación debe considerar tanto la capacidad explicativa como la magnitud de los errores de predicción.

Coeficiente de determinación (R^2)

El coeficiente de determinación R^2 mide la proporción de la variabilidad total de la variable dependiente que es explicada por el modelo. Se define como:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

En términos simples, un valor alto de R^2 indica que el modelo logra reproducir adecuadamente los patrones generales observados en los datos. No obstante, esta métrica no informa directamente sobre el tamaño de los errores individuales ni garantiza una buena capacidad predictiva fuera de la muestra, especialmente en modelos complejos.

Error cuadrático medio (RMSE)

El Error Cuadrático Medio (RMSE) cuantifica el error promedio de las predicciones del modelo, penalizando de forma más severa los errores grandes. Se calcula como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Esta métrica se expresa en las mismas unidades que la variable dependiente, lo que facilita su interpretación sustantiva. Un RMSE menor indica que, en promedio, las predicciones del modelo se encuentran más cercanas a los valores observados.

El uso conjunto de R^2 y RMSE permite evaluar tanto la capacidad explicativa relativa como la precisión absoluta de los modelos estimados.

Interpretación de modelos no lineales mediante SHAP

Dado que los modelos no lineales sacrifican interpretabilidad estructural en favor de desempeño predictivo, se incorporan los valores SHAP (SHapley Additive exPlanations) como herramienta de interpretación.

SHAP se basa en la teoría de juegos cooperativos y asigna a cada variable una contribución marginal promedio sobre la predicción del modelo. El valor SHAP de una variable j se define como:

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [f(S \cup \{j\}) - f(S)]$$

donde:

- N representa el conjunto total de variables explicativas del modelo.
- S es un subconjunto de variables que no incluye a la variable j .
- $f(S)$ corresponde a la predicción del modelo utilizando únicamente las variables del subconjunto S .

En términos prácticos, SHAP permite identificar qué variables contribuyen positiva o negativamente a las predicciones, así como la magnitud de dicho aporte, tanto a nivel individual como agregado. De este modo, actúa como un puente interpretativo entre modelos complejos y el análisis sustantivo del fenómeno estudiado.

3.5. Enfoques tradicionales de medición del desempeño organizacional

El análisis del desempeño institucional ha dejado de ser un ámbito exclusivo del sector privado para convertirse en un componente central de la gestión pública moderna. Los gobiernos y organismos estatales enfrentan hoy la necesidad de optimizar recursos, estandarizar procesos, aumentar la eficiencia y fortalecer la rendición de cuentas. En este contexto, la incorporación de indicadores de gestión y desempeño, junto con metodologías de análisis provenientes de la ingeniería, permiten avanzar hacia una gestión pública más efectiva, transparente y basada en resultados.

La Nueva Gestión Pública (NGP) promueve precisamente esta lógica de modernización, impulsando el uso de herramientas cuantitativas que orienten las decisiones hacia la eficiencia y la satisfacción ciudadana. Dentro de este marco, los Key Performance Indicators (KPIs) y el Balanced Scorecard (BSC) constituyen instrumentos complementarios que posibilitan la medición, análisis y mejora continua del desempeño institucional, especialmente en contextos legislativos como el del Congreso Nacional de Chile.

3.5.1. Indicadores de gestión y desempeño (KPIs)

Los Key Performance Indicators (KPIs) son métricas cuantitativas diseñadas para medir el avance de una organización hacia objetivos estratégicos previamente definidos. Según Parmenter (2015), un KPI “es un indicador crítico de desempeño que proporciona evidencia sobre cuán exitosamente se está logrando un objetivo estratégico” (p. 4).

Estos indicadores permiten monitorear tanto los resultados como los procesos que los originan, entregando información oportuna para la toma de decisiones. En la gestión pública, los KPIs son esenciales para evaluar la eficiencia, la calidad del servicio, la transparencia y la satisfacción ciudadana, contribuyendo a fortalecer la legitimidad institucional.

Ejemplos de aplicación:

- En el sector privado: productividad laboral, cumplimiento de plazos, rotación de personal, satisfacción del cliente.
- En el sector público: tiempo promedio de atención, cobertura de vacunación, tasa de aprobación en evaluaciones educativas, nivel de satisfacción ciudadana con servicios municipales.

En Chile, la Dirección de Presupuestos (DIPRES) ha impulsado desde los años noventa programas de desempeño institucional basados en indicadores de cumplimiento y eficiencia (DIPRES, 2019). Asimismo, la OCDE (2017) destaca los KPIs como mecanismos clave para fortalecer la confianza pública y la rendición de cuentas gubernamental.

En el contexto del Congreso Nacional de Chile, estos indicadores permiten medir aspectos como la asistencia, participación efectiva, productividad legislativa y transparencia, facilitando una evaluación objetiva de la labor parlamentaria y promoviendo la gestión por resultados.

3.5.2. Aplicación de la ingeniería industrial en el sector público: control de procesos, eficiencia y mejora continua

La ingeniería industrial, tradicionalmente asociada al sector productivo, ha ampliado su campo de acción hacia la gestión pública, aportando metodologías que permiten analizar y optimizar procesos administrativos. Conceptos como control de procesos, eficiencia operativa y mejora continua son plenamente aplicables al ámbito estatal, donde los recursos son escasos y la demanda social por resultados concretos es creciente.

Mediante la aplicación de principios ingenieriles, es posible modelar procesos administrativos, identificar cuellos de botella, proponer indicadores de desempeño y diseñar sistemas de control que permitan una gestión más racional y transparente. Este enfoque técnico favorece la toma de decisiones basada en evidencia y fortalece la gestión de calidad dentro de las instituciones públicas.

3.5.3. Análisis de procesos: asistencia y registro parlamentario como flujo de trabajo

El análisis de procesos consiste en representar de manera ordenada las etapas que componen una actividad organizacional, identificando los puntos críticos que afectan su rendimiento. En el caso del Congreso Nacional, el proceso de asistencia y registro parlamentario puede entenderse como un flujo de trabajo compuesto por fases específicas: notificación de sesión, ingreso a sala, registro de asistencia, participación en debates, votaciones y registro final en actas.

Aplicar este enfoque al contexto legislativo permite detectar inconsistencias, como la posibilidad de que un parlamentario figure presente sin haber participado activamente, o se retire antes del término de la sesión. El tratamiento de estos procesos con herramientas de ingeniería posibilita identificar la brecha entre el cumplimiento formal y el cumplimiento efectivo de la función legislativa, lo que constituye una base esencial para diseñar indicadores más precisos y útiles.

3.5.4. Balanced Scorecard (BSC): marco estratégico para la medición del desempeño

El Balanced Scorecard (BSC), desarrollado por Kaplan y Norton (1992), surge como una herramienta que amplía la evaluación del desempeño organizacional más allá de los resultados financieros, integrando una visión estratégica que abarca diferentes dimensiones. Este modelo traduce la misión y visión institucional en un conjunto equilibrado de objetivos, indicadores y metas, organizados en cuatro perspectivas principales:

- Financiera, centrada en la administración eficiente y responsable de los recursos públicos.
- Clientes o ciudadanía, orientada a medir la satisfacción y confianza de los usuarios o beneficiarios de los servicios públicos.
- Procesos internos, enfocada en la calidad, consistencia y eficiencia de los procedimientos legislativos y administrativos.

- Aprendizaje y crecimiento, vinculada al desarrollo de capacidades, innovación y mejora organizacional.

En el sector público, el BSC se adapta para reflejar los objetivos de transparencia y eficiencia institucional. Su valor radica en que integra los KPIs en un marco estratégico, estableciendo vínculos entre los objetivos de alto nivel y las actividades operativas. En el caso del Congreso chileno, la aplicación del BSC permite organizar los indicadores de desempeño —como asistencia, productividad, eficiencia y percepción ciudadana— dentro de una estructura coherente que favorece la toma de decisiones basadas en resultados.

Para su implementación, el BSC requiere una serie de pasos: definición de misión y visión institucional, identificación de objetivos estratégicos, diseño del mapa estratégico, selección de indicadores, definición de metas y mecanismos de seguimiento. Este enfoque ha sido reconocido por organismos como el Balanced Scorecard Institute (2019) y Bernard Marr (2021) por su efectividad en el sector público, especialmente en el fortalecimiento de la rendición de cuentas y la comunicación estratégica.



Ilustración 1: Balanced Scorecard.

Si bien estas herramientas han sido ampliamente utilizadas en el ámbito de la gestión pública y privada, su aplicación directa al análisis del desempeño legislativo individual presenta limitaciones operativas y conceptuales, especialmente cuando se busca modelar relaciones estadísticas entre múltiples variables observables.

3.6. Percepción Ciudadana y Confianza Institucional

La confianza institucional constituye un elemento fundamental para la estabilidad democrática y la legitimidad de los poderes del Estado. En el caso del Congreso Nacional, la confianza ciudadana está estrechamente vinculada a la percepción del desempeño legislativo y a la transparencia en el uso de los recursos públicos. Aun cuando los registros formales de asistencia muestran cifras superiores al 95% en promedio, persiste una percepción social de desconexión e ineficiencia. Este fenómeno se explica, en parte, porque los datos oficiales no consideran aspectos cualitativos como la puntualidad, la permanencia en las sesiones o la participación en los debates parlamentarios.

Diversos estudios han demostrado que el descrédito en los parlamentos no se debe únicamente a su labor política, sino también a la forma en que se comunica y evalúa su desempeño frente a la ciudadanía. Por esta razón, es fundamental analizar la relación entre los indicadores objetivos de asistencia y productividad con la valoración subjetiva que la sociedad tiene de sus representantes.

3.6.1. Estudios y encuestas sobre confianza en el Congreso Nacional (CEP, CADEM, Latinobarómetro).

En Chile, diferentes encuestas de opinión reflejan un sostenido nivel de desconfianza hacia el Congreso. La Encuesta CEP (2024) muestra que solo un 8% de los encuestados confía en el Congreso Nacional, ubicándolo entre las instituciones peor evaluadas del país (CEP, 2024). De manera similar, la encuesta CADEM (2025) indica que el Congreso se sitúa en los últimos lugares en percepción de confianza, superando únicamente a los partidos políticos.

A nivel regional, el Latinobarómetro (2023) evidencia que el Congreso chileno se encuentra en la parte baja de los niveles de confianza en comparación con otros parlamentos latinoamericanos, con un 12% de confianza ciudadana, mientras que el promedio regional alcanza un 20%.

La OCDE (2024) también ha señalado que Chile presenta una de las brechas más grandes entre cumplimiento institucional formal y confianza percibida por la ciudadanía, destacando que la transparencia en la entrega de datos no ha logrado revertir el bajo prestigio social del Congreso.

3.6.2. Relación entre confianza ciudadana y desempeño legislativo percibido.

La literatura especializada indica que la confianza ciudadana no depende exclusivamente de la productividad legislativa objetiva, sino de cómo la ciudadanía percibe

la relevancia, transparencia y coherencia del trabajo parlamentario. Así, la asistencia superior al 95% en ambas cámaras no se traduce automáticamente en mayor confianza, porque no refleja factores como la puntualidad o la calidad del debate.

En este contexto, se genera una brecha entre desempeño formal y desempeño percibido. Mientras los registros muestran cumplimiento, la ciudadanía observa retrasos, sesiones vacías y falta de conexión con sus demandas. Este desajuste impacta directamente en la legitimidad institucional, ya que los ciudadanos tienden a evaluar el Congreso no solo por sus resultados normativos, sino también por señales visibles de compromiso y responsabilidad de sus integrantes.

Por ello, avanzar hacia sistemas de medición más integrales —que incluyan indicadores de gestión como puntualidad, permanencia y participación— podría contribuir no solo a mejorar la eficiencia legislativa, sino también a recuperar la confianza ciudadana en el Congreso Nacional.

3.6.3. Benchmarking institucional (comparar con parlamentos internacionales).

El benchmarking es una técnica de gestión que consiste en comparar el desempeño de una institución con referentes nacionales o internacionales, con el fin de identificar buenas prácticas transferibles.

En el caso del Congreso chileno, el benchmarking podría considerar parlamentos que han desarrollado sistemas más rigurosos de control de asistencia y participación, como el Parlamento Europeo, que publica reportes detallados de asistencia y participación en comisiones, o el Congreso de Estados Unidos, donde los registros incluyen votaciones individuales y porcentajes de participación por legislador. Estos ejemplos ofrecen insumos para evaluar la factibilidad de mejorar los mecanismos de control en Chile.

3.6.4. Casos de éxito en el uso de métricas de gestión en organismos gubernamentales.

Diversos organismos públicos han demostrado la utilidad de aplicar métricas de gestión en su trabajo:

En Chile, la Dirección de Presupuestos (DIPRES) implementa desde fines de los años noventa el Sistema de Evaluación y Control de Gestión, que utiliza indicadores de desempeño institucional para asignar recursos de manera más eficiente.

En Reino Unido, el modelo de Public Service Agreements estableció compromisos medibles de desempeño para cada ministerio, vinculando resultados con financiamiento.

En Canadá, la Auditoría General ha promovido la implementación de métricas de

desempeño en programas sociales y de salud, lo que ha permitido identificar falencias y optimizar recursos.

Estos casos demuestran que las herramientas de análisis y control, al ser adaptadas al sector público, pueden generar mejoras significativas en transparencia, eficiencia y legitimidad institucional.

3.7. Brecha entre asistencia formal y participación efectiva

El análisis de la asistencia parlamentaria en Chile refleja porcentajes elevados en ambas cámaras, con registros oficiales superiores al 95% en promedio. No obstante, estas cifras esconden una brecha significativa entre la asistencia formal, entendida como la mera presencia registrada en sala, y la participación efectiva, que implica puntualidad, permanencia y contribución activa a las sesiones.

3.7.1. Concepto de asistencia formal vs. asistencia real (ej. llegada tardía, abandono de sesiones, ausencias parciales).

La asistencia formal corresponde al registro oficial que constata la presencia de un parlamentario en sala, independientemente de si este llega tarde o se retira anticipadamente. En contraste, la asistencia real o efectiva considera no solo la presencia física, sino también la puntualidad en el inicio de las sesiones, la permanencia durante toda la jornada legislativa y la participación en los debates y votaciones. Casos documentados de legisladores que se registran como presentes pese a ingresar con más de una hora de retraso o que permanecen escasos minutos en sala muestran que ambos conceptos pueden divergir de manera significativa.

3.7.2. Limitaciones de los registros oficiales para reflejar la participación sustantiva.

Los registros oficiales del Congreso no distinguen entre asistencia puntual y tardía, ni tampoco detallan la permanencia o la calidad de la participación de los parlamentarios. Esta limitación genera un sesgo en la información disponible, pues un alto porcentaje de asistencia no garantiza que los legisladores hayan estado presentes durante los debates centrales o que hayan contribuido activamente al proceso legislativo. En consecuencia, los datos actuales tienden a sobrestimar el nivel real de compromiso parlamentario.

3.7.3. Impacto de esta brecha en la calidad del debate legislativo y en la percepción ciudadana.

La brecha entre asistencia formal y participación efectiva afecta directamente la

calidad del debate legislativo. La ausencia parcial o la falta de intervención de parlamentarios disminuye la diversidad de puntos de vista, limita la deliberación democrática y puede retrasar la tramitación de proyectos de ley.

Al mismo tiempo, esta situación incide en la percepción ciudadana. Aunque los datos oficiales muestran un cumplimiento formal elevado, la ciudadanía observa escenas de salas con escasa presencia, retrasos recurrentes y discusiones con baja participación, lo que alimenta la desconfianza hacia el Congreso. De este modo, la brecha identificada contribuye a explicar por qué, a pesar de los altos niveles de asistencia registrados, los índices de confianza institucional permanecen entre los más bajos del sistema político.

3.8. Metodología Cross-Industry Standard Process for Data Mining (CRISP-DM)

La CRISP-DM (Cross Industry Standard Process for Data Mining) es un modelo metodológico desarrollado a fines de la década de 1990 con el objetivo de estandarizar el desarrollo de proyectos de minería de datos y análisis predictivo. Este enfoque propone un marco estructurado para organizar las etapas involucradas en la extracción de conocimiento a partir de datos.

El modelo describe el ciclo de vida de un proyecto de análisis de datos mediante seis fases principales: comprensión del problema, comprensión de los datos, preparación de los datos, modelamiento, evaluación y despliegue. Estas etapas forman un proceso iterativo en el que es posible retroceder entre fases a medida que se obtiene nueva información o se detectan problemas en los datos o en los modelos desarrollados.

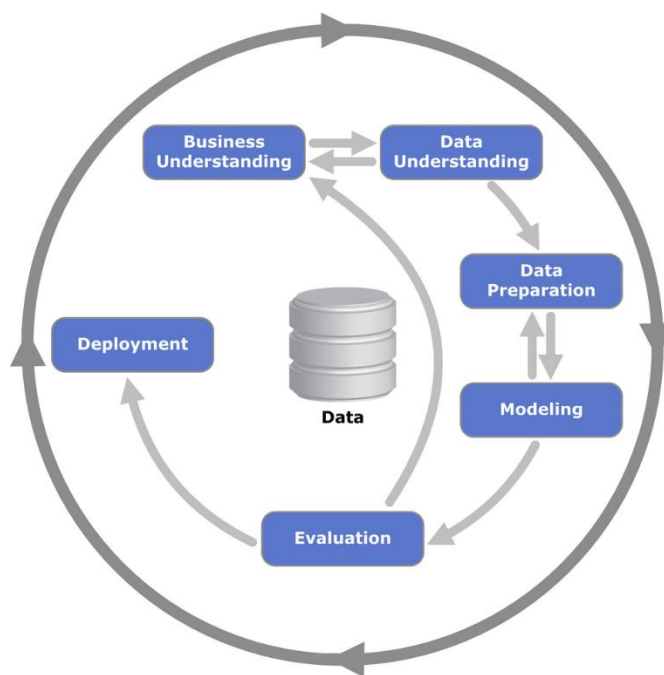


Ilustración 2: Fases de la metodología CRISP-DM.

El modelo CRISP-DM organiza el proceso analítico en las siguientes fases:

1. Comprensión del problema (Business Understanding)

Corresponde a la etapa inicial del proceso. En esta fase se identifican los objetivos del proyecto, el problema que se busca analizar y los criterios que permitirán evaluar los resultados obtenidos. También se definen las preguntas de investigación y el alcance del análisis.

2. Comprensión de los datos (Data Understanding)

En esta etapa se realiza una exploración inicial del conjunto de datos disponible. Incluye la recopilación de los datos, la descripción de sus características, la identificación de variables relevantes y la detección de posibles problemas como valores faltantes, inconsistencias o errores en la información.

3. Preparación de los datos (Data Preparation)

Esta fase comprende el conjunto de procedimientos necesarios para construir el dataset que será utilizado en el análisis. Entre las tareas más comunes se encuentran la selección de variables, limpieza de datos, transformación de variables, integración de diferentes fuentes de información y generación de nuevas variables derivadas.

4. Modelamiento (Modeling)

Durante esta etapa se aplican técnicas estadísticas o algoritmos de aprendizaje automático con el objetivo de identificar patrones en los datos o construir modelos capaces de explicar o predecir un fenómeno determinado. Esta fase puede incluir la estimación de

múltiples modelos y el ajuste de sus parámetros.

5. Evaluación (Evaluation)

En esta fase se analizan los resultados obtenidos por los modelos desarrollados. El objetivo es verificar si el modelo responde adecuadamente al problema planteado inicialmente y si los resultados son consistentes con los objetivos del análisis. También se comparan distintos modelos para identificar el que presenta mejor desempeño.

6. Despliegue o implementación (Deployment)

Corresponde a la etapa final del proceso. En esta fase los resultados del análisis son interpretados y utilizados para generar conocimiento o apoyar procesos de toma de decisiones. En investigaciones académicas esta etapa suele materializarse en la interpretación de los resultados y su discusión en relación con la literatura existente.

Una característica central de CRISP-DM es su estructura cíclica. Las fases del proceso no necesariamente se ejecutan de forma estrictamente secuencial, sino que el análisis puede requerir volver a etapas anteriores para ajustar los datos, modificar las variables utilizadas o mejorar los modelos desarrollados. Este enfoque permite adaptar el proceso analítico a distintos tipos de proyectos de minería de datos y análisis estadístico.

4 Metodología

El objetivo principal de esta investigación consiste en analizar los factores asociados a la productividad legislativa de los diputados, mediante la construcción de un índice de productividad y el desarrollo de modelos analíticos que permitan identificar las variables que explican dicho fenómeno.

Considerando la naturaleza del problema de investigación y el tipo de análisis requerido, se optó por estructurar el proceso metodológico utilizando la metodología CRISP-DM. Este enfoque resulta particularmente adecuado para investigaciones que combinan la construcción de indicadores compuestos, el análisis exploratorio de datos y la aplicación de modelos predictivos.

La utilización de CRISP-DM en esta investigación responde a varias razones metodológicas. En primer lugar, el estudio requiere integrar distintas fuentes de información y variables de diversa naturaleza, incluyendo indicadores de actividad legislativa, características individuales de los diputados y variables institucionales. La estructura propuesta por CRISP-DM permite organizar de manera sistemática las etapas de selección, preparación y transformación de estos datos.

En segundo lugar, la investigación contempla la aplicación y comparación de múltiples modelos analíticos con el objetivo de identificar aquellos que presentan mejor capacidad explicativa respecto de la productividad legislativa. La fase de modelamiento y evaluación propuesta por CRISP-DM facilita la comparación sistemática entre distintos enfoques analíticos, permitiendo seleccionar el modelo más adecuado para el conjunto de datos analizado.

En tercer lugar, el enfoque iterativo de CRISP-DM resulta particularmente útil en investigaciones exploratorias donde la relación entre variables no es completamente conocida. En este estudio, la exploración inicial de los datos permitió identificar relaciones preliminares entre variables, detectar problemas de multicolinealidad y ajustar el conjunto de variables utilizadas en los modelos analíticos.

Finalmente, el uso de esta metodología contribuye a garantizar la transparencia y replicabilidad del proceso de investigación. Al estructurar el análisis en fases claramente definidas, estas son: comprensión del problema, comprensión de los datos, preparación de los datos, modelamiento, evaluación e interpretación. Esto facilita la comprensión del procedimiento seguido y permite que futuros estudios puedan replicar o ampliar los

resultados obtenidos.

En consecuencia, la metodología CRISP-DM proporciona un marco estructurado que permite integrar de manera coherente las distintas etapas del proceso analítico desarrollado en esta investigación, desde la construcción del índice de productividad legislativa hasta la estimación y evaluación de los modelos explicativos.

4.1 Enfoque y Diseño de Investigación

La presente investigación se fundamenta en un enfoque cuantitativo, de carácter no experimental y explicativo, esto quiere decir que se evaluará de manera numérica los datos y variables, no se obtendrán los datos por medio de experimentos, solo se recolectarán y se analizarán y será de carácter explicativo dado que se buscará encontrar las variables explicativas al fenómeno buscado. Se aborda el análisis del Congreso Nacional de Chile desde la perspectiva de la gestión pública moderna, tratándolo como una organización pública que transforma recursos institucionales, tiempo y capital humano en resultados legislativos observables. Bajo este enfoque, la productividad se entiende como la capacidad efectiva de los diputados para generar productos legislativos concretos durante un periodo determinado.

El diseño es longitudinal retrospectivo, ya que se observa el fenómeno en su contexto natural sin manipular variables, analizando el comportamiento acumulado durante las legislaturas 370 a 373 (periodo 2022-2026). Adicionalmente, los procesos de la investigación se estructuran en fases sucesivas comenzando con la comprensión del negocio, la comprensión de los datos, la preparación, el modelado, la evaluación e implementación. Dentro de este marco, se aplican técnicas de minería de datos con un enfoque predictivo-explicativo, combinando análisis descriptivo de la población en estudio con modelos de regresión regularizada y algoritmos de aprendizaje automático, con el objetivo de identificar las variables individuales asociadas a la productividad legislativa.

4.2 Población y Muestra

4.2.1 Universo y Población

La población objetivo de esta investigación abarca a la totalidad de los integrantes de la Cámara de Diputadas y Diputados de Chile durante el periodo legislativo correspondiente a las legislaturas 370, 371, 372 y 373. El horizonte temporal del estudio se extiende desde el inicio del periodo constitucional, el 11 de marzo de 2022, hasta la fecha de corte de la actualización de los datos utilizados, el 3 de noviembre de 2025.

La composición institucional de la Cámara se basa en 155 escaños, representativos

de los 28 distritos del país. Sin embargo, debido a la dinámica parlamentaria que incluye renunciaciones, fallecimientos, inhabilitaciones legales y reemplazos por cupos vacantes, el universo de individuos que han ejercido el cargo durante el periodo analizado asciende a un total de 157 diputados.

La información que caracteriza a esta población fue recopilada mediante la revisión sistemática de fuentes oficiales primarias, específicamente el portal web de la Cámara de Diputadas y Diputados de Chile y la Biblioteca del Congreso Nacional (BCN), garantizando así la veracidad y validez pública de los registros.

4.2.2 Tipo de Muestreo

Dado el objetivo de analizar la productividad total del periodo y la disponibilidad de datos integrales, se trabajó bajo la modalidad de Censo. Esto implica que no se seleccionó una muestra estadística, sino que se incorporó en el estudio a la totalidad de los sujetos que conforman la población histórica del periodo, o sea los 157 diputados.

Al incluir tanto a los diputados que iniciaron el mandato en 2022 como a quienes asumieron posteriormente por reemplazo, se elimina el sesgo de selección y se obtiene una visión completa de la producción legislativa generada. Este enfoque permite que los modelos predictivos posteriores se entrenen con la totalidad de la variabilidad de actores y situaciones ocurridas en el periodo.

4.2.3 Unidad de Análisis u Observación y Criterios de Selección

La unidad de análisis primaria son los diversos diputados entendidos como un individuo único que acumula toda su respectiva información. Cada uno de los 157 registros en la base de datos representa a un legislador caracterizado por un conjunto de atributos sociodemográficos, académicos, conductuales y políticos. Para garantizar la integridad de la base de datos final, se consideró a cualquier persona que haya ejercido el cargo de diputado en alguna de las sesiones comprendidas entre el 11 de marzo de 2022 y el 3 de noviembre de 2025, sin aplicar criterios de exclusión por afiliación política o tiempo en el cargo, dando el total de 157 parlamentarios.

4.3 Definición y Operacionalización de Variables

Para el análisis cuantitativo, se definieron dos grandes grupos de variables: las variables independientes (predictoras o características) y las variables dependientes (variables a estudiar). A continuación, se detalla su operacionalización.

4.3.1 Variables Dependientes

Como anteriormente fue explicado la productividad legislativa se define como la

capacidad efectiva de los diputados para poder generar productos legislativos concretos (resultados “visibles” o “tangibles”) durante un periodo establecido (periodo de estudio). Teniendo eso en cuenta para esta investigación se definieron cuatro métricas como la base de la variable dependiente, incluyendo cinco variables que se utilizan en estas, las cuales son las siguientes: artículos tramitados y leyes publicadas, puntaje de complejidad, peso de autoría, nivel de importancia para generar el índice de productividad normalizado.

Índice de Productividad (IP):

$$IP_i = \frac{\sum_{j \in L_i} N_j \cdot A_{i,j}}{T_i}$$

El índice de productividad de cada diputado se obtiene sumando los productos entre el nivel de importancia y el peso de autoría de cada ley en que participó, y dividiendo ese total por el tiempo efectivo en el cargo medido en años.

El numerador captura la productividad acumulada bruta del diputado: a mayor número de leyes, mayor nivel de importancia de esas leyes y mayor protagonismo como autor, mayor es el valor del numerador. Un diputado que fue autor exclusivo de tres leyes de nivel 3 obtiene un numerador de 9, mientras que uno que participó como coautor en seis leyes de nivel 1 obtiene un numerador de 3, reflejando que no solo importa la cantidad sino también la calidad y el liderazgo en la actividad legislativa.

La división por $T(i)$, el tiempo activo en años, es necesaria porque el período de observación no es uniforme para todos los diputados. Algunos ejercieron el cargo durante los cuatro años completos de las legislaturas 370 a 373, mientras que otros ingresaron por reemplazo en fechas posteriores, acumulando inevitablemente menos oportunidades de participar en leyes aprobadas. Sin esta normalización, los diputados con mayor tiempo en el cargo tendrían sistemáticamente índices más altos, independientemente de su intensidad de trabajo real, lo que sesgaría cualquier comparación entre legisladores. La normalización transforma el índice en una medida de productividad por unidad de tiempo, haciendo comparables a diputados con distintas duraciones de mandato dentro del período analizado.

Nivel de Importancia:

El nivel de importancia se calcula de dos maneras simultaneas, la primera es la cuantitativa obtenida a través de un proceso de clasificación por medio del puntaje de complejidad y la segunda es por medio del criterio del investigador. Esto a su vez tiene

una diferenciación extra, para evitar el sesgo haciendo más relevante el criterio cuantitativo por sobre el cualitativo.

$$N_j = \frac{1}{3} \cdot N_{j,cual} + \frac{2}{3} \cdot N_{j,cuant}$$

Nivel de Importancia Cualitativo:

Con el objetivo de diferenciar el impacto sustantivo de la actividad legislativa, se construyó una clasificación de tres niveles que subdivide las leyes según su relevancia para la ciudadanía y el funcionamiento del sistema político.

El primer nivel agrupa iniciativas de bajo impacto estructural. Tienen un valor de 1,0. Se trata, en general, de leyes de carácter administrativo, simbólico o particular. Dentro de esta categoría se incluyen, por ejemplo, aquellas destinadas a otorgar nacionalidad por gracia, prorrogar la vigencia de normativas existentes, declarar feriados, asignar nombres a infraestructuras públicas, autorizar la construcción de monumentos o permitir actividades específicas como eventos deportivos. Estas iniciativas, si bien cumplen funciones institucionales, no generan cambios significativos en las condiciones sociales o económicas de la población.

El segundo nivel corresponde a modificaciones legales de alcance acotado o sectorial. Tienen un valor de 2. Aquí se ubican reformas que introducen ajustes menores en normativas existentes o que afectan ámbitos específicos sin alterar de manera sustantiva su funcionamiento general. Ejemplos de este tipo son modificaciones puntuales al Código de Comercio, regulaciones administrativas particulares o ajustes vinculados a permisos de construcción y otras materias similares. Su impacto es limitado y tiende a concentrarse en grupos o sectores definidos.

El tercer nivel incluye leyes de alta relevancia, entendidas como aquellas que generan efectos significativos en la sociedad. Se trata de iniciativas que crean o modifican normas en áreas clave como la economía, la seguridad, las políticas sociales o los servicios básicos. Tienen un valor de 3.

Si bien la clasificación de las leyes por nivel de importancia introduce un componente de juicio del investigador, se fundamenta en criterios observables y replicables basados en el contenido normativo de cada ley. Adicionalmente, dicho criterio fue complementado con una medida objetiva de complejidad normativa basada en el conteo ponderado de artículos por ley, explicado anteriormente, lo que reduce la dependencia del juicio personal. De igual manera se reconoce como una limitación metodológica que futuras investigaciones podrían mejorar mediante clasificaciones institucionales o validación por

expertos.

Nivel de Importancia Cuantitativo:

$$N_{j\text{ cuant}} = \begin{cases} 1, & \text{sí } C_j \leq 1,6 \\ 2, & \text{sí } 1,6 < C_j \leq 3,45 \\ 3, & \text{sí } C_j > 3,45 \end{cases}$$

El nivel de importancia transforma el puntaje de complejidad continuo $C(j)$ en una escala ordinal de tres categorías, con el fin de distinguir entre leyes de distinto impacto normativo sin introducir diferencias entre leyes con complejidades similares.

Los puntos de corte se determinan combinando un criterio de salto relativo (ratio) sobre la distribución empírica ordenada de los 53 puntajes de complejidad, en lugar de un criterio de salto absoluto. Este ajuste se justifica porque, en una distribución de valores crecientes, un mismo salto absoluto representa una discontinuidad proporcionalmente mayor cuanto menor es la base sobre la que ocurre. El análisis de razones entre valores consecutivos identificó dos discontinuidades relativas relevantes: la primera entre 1,4 y 1,8 (razón=1,29), y la segunda, la de mayor magnitud relativa antes del ingreso a la cola de valores atípicos, entre 2,9 y 4,0 (razón=1,38). Los cortes se fijaron en los puntos medios de ambos intervalos (1,6 y 3,45), generando tres niveles: un nivel bajo ($C_j \leq 1,6$) que concentra el 69,8% de las observaciones ($n=37$) y corresponde al grueso de la actividad legislativa de menor densidad normativa; un nivel medio ($1,6 < C_j \leq 3,45$) con 20,8% de las observaciones ($n=11$); y un nivel alto ($C_j > 3,45$) con el 9,4% restante ($n=5$), que agrupa tanto leyes de impacto moderado-alto como los casos extremos de la distribución. Estos últimos no se excluyen del análisis pese a su carácter atípico, dado que corresponden a leyes sustantivamente relevantes para el universo de estudio, su agrupación en un único nivel superior evita que su magnitud distorsione la escala ordinal sin eliminar su contribución al índice.

La elección de tres niveles, en lugar de cuatro o más, responde al tamaño reducido del universo de leyes analizado, donde una mayor granularidad produciría categorías con muy pocas observaciones y menor robustez estadística. La escala ordinal 1-2-3 es además intuitiva para la interpretación del índice final: una ley de nivel 3 tiene tres veces más peso que una de nivel 1 en la construcción de la productividad del diputado.

Puntaje de Complejidad:

$$C_j = A_{prin} \cdot 0,45 + A_{sub} \cdot 0,33 + A_{trans} \cdot 0,22$$

El puntaje de complejidad se determina en base a la cantidad de artículos principales, subartículos y de artículos transitorios. El ponderador que acompaña a cada uno de ellos se detalla de la siguiente manera:

Los artículos principales reciben un peso de 0,45 por ser el núcleo autónomo de la ley. Cada artículo principal establece una norma completa e independiente, es decir, puede existir y tener efecto jurídico por sí solo sin depender de otro artículo.

Los subartículos reciben un peso de 0,33 porque, si bien constituyen trabajo legislativo real y reflejan la densidad técnica de la ley, son estructuralmente subordinados al artículo que los contiene, lo que significa que su autonomía normativa es parcial.

Los artículos transitorios reciben el peso más bajo, un valor de 0,22, porque su vigencia es limitada en el tiempo. No forman parte del cuerpo normativo permanente de la ley, sino que regulan situaciones de transición, plazos de implementación o excepciones acotadas que dejan de tener efecto una vez completado el período.

Peso de Autoría:

$$A_{i,j} = \begin{cases} 0,67 & \text{si el diputado } i \text{ es autor de la ley } j. \\ 0,33 & \text{si participa como coautor de la ley.} \end{cases}$$

El peso de autoría diferencia entre dos roles legislativos cualitativamente distintos. En primer lugar, un diputado que figura como autor de una moción es quien tuvo la idea inicial, la redactó, la presentó formalmente ante la Cámara o Comisión en la que participe y asumió la responsabilidad política de impulsarla durante su tramitación. Este rol implica un nivel de involucramiento activo, iniciativa y liderazgo que justifica asignarle el valor máximo de la escala, fijado en 1,0.

Y, en segundo lugar, un diputado que figura como coautor, en cambio, manifestó su respaldo a una iniciativa ajena sin necesariamente haber participado en su redacción ni en la conducción de su proceso de tramitación. Su participación es relevante para el avance de la ley, pues suma apoyo político a la moción, pero su contribución es de menor intensidad que la del autor principal, es por esto por lo que se le asigna un valor de 0,5.

Artículos Tramitados: Es la cantidad total de artículos tramitados en las leyes en que participó cada diputado.

Leyes Publicadas: Es el número de leyes publicadas en el Diario Oficial en que participó cada diputado.

4.3.2 Variables Independientes

Este grupo de variables intentan explicar las diversas características a estudiar o analizar para poder entender las variables dependientes u objetivos. Las cuales se dividieron de la siguiente manera, para lograr un mejor entendimiento:

- Variables de Identificación

Diputado: Apellidos y nombres de los individuos, es el identificador de la base de datos.

Edad: Edad actual de la persona.

Género: Variable dummy, 1 si es de género Masculino y 0 si es de género femenino.

Nivel de Votación: La variable nivel de votación, mide el respaldo electoral relativo de cada diputado en su respectivo distrito calculado de la siguiente manera:

$$v_{nc} = \frac{N_v \cdot N_c}{T_d}$$

Siendo v_{nc} el nivel de votación, N_v es número de votos por el candidato, N_c es el número de candidatos elegidos por distrito y T_d es el total de votos válidos del distrito.

Esto con el fin de ponderar la variable de una manera que se normalice según la cantidad de personas del distrito y la cantidad de candidatos seleccionados por dicho distrito.

Nivel de Formación: Es una variable cuantitativa mediante un sistema de puntajes. La escala se construyó sobre un rango entre 0 y 100, donde el valor mínimo corresponde a los diputados que solo cuenten con el nivel de la enseñanza media completa, ya que es el requisito mínimo establecido por la legislación chilena para ejercer el cargo de diputado y el puntaje máximo es cuando el diputado en cuestión tiene todos los niveles académicos cursados y aprobados.

Los demás puntajes se asignan de acuerdo con el nivel alcanzado, siendo una suma de cada nivel académico alcanzado, ósea que si realizo un doctorado y tiene un título universitario se le asigna el puntaje dado por la tabla 2.

Los valores base definidos para cada nivel educativo son los siguientes:

Tabla 1. Puntajes asignados por el nivel educativo.

Nivel educativo	Puntaje base
Doctorado	43
Magíster	32
Título profesional universitario	20
Título técnico de nivel superior	10
Egresado de carrera universitaria	10
Licenciatura	5
Diplomado o postítulo	5
Enseñanza media completa	0

La lógica del puntaje asignado para cada nivel educativo está basada en la progresión en términos de duración, especialización y profundidad académica, es así como programas de posgrado, como magísteres o doctorados, implican un nivel de especialización superior al pregrado, como también mayores exigencias académicas y mayor inversión de tiempo en formación avanzada.

En relación con las categorías de licenciatura y egresado universitario, se estableció una regla de bloqueo metodológico. Si un diputado posee título profesional, magíster o doctorado, los valores correspondientes a licenciatura se fijan en 0 puntos. Y para el caso de los egresados su puntaje se fijará en cero si un diputado posee un título universitario profesional. La razón de esta decisión es que estos niveles constituyen etapas intermedias del mismo proceso formativo y, por tanto, su inclusión simultánea produciría una doble contabilización del mismo ciclo educativo. Sin embargo, cuando la licenciatura o la condición de egresado representan el nivel máximo de formación alcanzado, se utilizan sus valores base correspondientes.

Experiencia: Años de experiencia acumulados en cámaras del congreso siendo

diputado o senador.

Experiencia laboral: Años de experiencia previa al periodo 2022-2026 en el mundo privado.

Sindicato: Variable dummy, 1 si participó o dirigió sindicato, 0 si no.

- Variables de Afiliación Política

Partido: Son todos los partidos políticos del periodo 2022 a 2026.

Coalición: Es una variable categórica construida mediante la agrupación de los partidos políticos en bloques ideológicos según los siguientes tres criterios:

El primer criterio corresponde a la ubicación ideológica según el programa de cada partido. Para ello se revisaron sus principios doctrinarios, programas políticos y posicionamientos públicos en torno a temas relevantes como el rol del Estado, políticas sociales, reformas institucionales y derechos civiles. Este análisis permitió identificar afinidades programáticas que agrupan a los partidos en grandes familias ideológicas.

El segundo criterio considera la trayectoria política e institucional de los partidos dentro del sistema político chileno. Este elemento resulta especialmente relevante para distinguir entre partidos que forman parte de tradiciones políticas consolidadas como los partidos que integraron los gobiernos de la Concertación y la Nueva Mayoría y aquellos que emergen en ciclos políticos más recientes, particularmente a partir de las movilizaciones sociales de la década de 2010 y del surgimiento de nuevas coaliciones políticas.

El tercer criterio corresponde al comportamiento político observable durante el período 2022-2026. Se consideraron especialmente los patrones de coordinación parlamentaria, la participación en bloques legislativos, las alianzas políticas formales e informales y la cercanía programática con determinadas agendas legislativas. Este criterio permite complementar la clasificación ideológica con evidencia empírica sobre la conducta política efectiva de los partidos dentro del Congreso.

A partir de la combinación de estos tres criterios se definieron cinco categorías ideológicas: izquierda progresista, izquierda tradicional, centro, derecha

tradicional y derecha radical, las que se explicaran a continuación y la distribución que tiene cada grupo está en la columna 2 de la Tabla 1

La categoría izquierda progresista agrupa a partidos vinculados al ciclo político más reciente de la izquierda chilena, caracterizados por agendas programáticas orientadas a transformaciones estructurales en materias sociales, ambientales e institucionales.

La categoría izquierda tradicional corresponde a partidos de centroizquierda con una trayectoria política consolidada en el sistema democrático chileno desde el retorno a la democracia en 1990.

La categoría centro agrupa a partidos que se sitúan en una posición intermedia dentro del eje ideológico izquierda-derecha y que no mantienen una identificación orgánica con los bloques tradicionales de la izquierda ni de la derecha. Estos partidos suelen enfatizar posiciones moderadas, orientadas hacia la estabilidad institucional y la búsqueda de acuerdos transversales.

La categoría derecha tradicional agrupa a los partidos que han constituido históricamente el principal bloque político de la derecha chilena desde el retorno a la democracia, caracterizados por posiciones programáticas favorables a la economía de mercado, un rol más limitado del Estado en la provisión de bienes públicos y una orientación política institucionalista.

Finalmente, la categoría derecha radical corresponde a partidos que se ubican en una posición más ideológicamente definida dentro del espectro conservador, caracterizados por discursos políticos más confrontacionales y por una mayor distancia programática respecto del gobierno durante el período analizado.

Tabla 2. Indica todos los partidos políticos del periodo, junto con la coalición definida y si se le atribuye ser de la oposición u oficialismo.

Partido político	Coalición ideológica	Alineación
Independientes	Centro	Mixto
Movimiento Amarillos por Chile	Centro	Oposición
Partido Demócrata Cristiano	Centro	Oposición
Partido Demócratas Chile	Centro	Oposición
Partido Liberal de Chile	Centro	Oficialismo

Partido Nacional Libertario	Derecha Radical	Oposición
Partido Republicano	Derecha Radical	Oposición
Partido Social Cristiano	Derecha Radical	Oposición
Evolución Política	Derecha Tradicional	Oposición
Renovación Nacional	Derecha Tradicional	Oposición
Unión Demócrata Independiente	Derecha Tradicional	Oposición
Federación Regionalista Verde Social	Izquierda Progresista	Oficialismo
Frente Amplio	Izquierda Progresista	Oficialismo
Partido Acción Humanista	Izquierda Progresista	Oficialismo
Partido Comunista	Izquierda Progresista	Oficialismo
Partido Ecologista Verde	Izquierda Progresista	Oficialismo
Partido Humanista	Izquierda Progresista	Oficialismo
Partido Por la Democracia	Izquierda Tradicional	Oficialismo
Partido Radical de Chile	Izquierda Tradicional	Oficialismo
Partido Socialista	Izquierda Tradicional	Oficialismo

Oficialismo: Variable dummy, 1 si es partido oficialista al gobierno del periodo 2022-2026, 0 si no. Primero el termino oficialista u oficialismo se refiere al grupo de personas que están alineados hacia el gobierno en mandato, y la forma de clasificar se realiza a partir de la combinación de tres criterios complementarios. El primer criterio corresponde a la pertenencia formal a la coalición de gobierno. Se consideraron como oficialistas aquellos partidos que integraron las coaliciones políticas que apoyaron al gobierno durante el período analizado, particularmente las agrupaciones políticas que participaron en la coordinación programática y legislativa del Ejecutivo.

El segundo criterio corresponde a la participación directa en el Poder Ejecutivo. Se considera como evidencia de alineación oficialista la ocupación de cargos ministeriales, subsecretarías u otras posiciones de responsabilidad política dentro del gobierno.

Y el tercer criterio corresponde al comportamiento legislativo observable de los partidos durante el período de estudio. Se analizaron patrones generales de votación, coordinación parlamentaria y posicionamientos públicos frente a las principales iniciativas impulsadas por el Ejecutivo.

En base a estos tres criterios se identificó que partidos estuvieron alineados al gobierno, donde se puede ver en la columna 3 de la tabla N°1.

- Variables de Actividad Parlamentaria

Intervenciones: Número total de intervenciones en la cámara en el periodo por analizar.

Intervenciones en Comisión: Número de intervenciones dentro de las comisiones que participa cada diputado.

Proyectos de Ley Presentados: Número de proyectos ingresados a la cámara como propuestas de ley.

Asistencia: Valor normalizado de la asistencia con valores entre 0 y 1 del periodo total.

Puntualidad: Valor normalizado de la puntualidad con valores entre 0 y 1 del periodo total.

Presidencia de Comisión: Número de veces elegido presidente de alguna comisión.

Cantidad de Comisiones: Cantidad de comisiones en las que participa.

Preside Cámara: Cantidad de veces elegido presidente de la cámara de diputados.

4.4 Fuentes y Recolección de Datos

La recolección se realizó mediante un análisis de documentos y extracción de datos oficiales. Las fuentes primarias fueron el sitio web de la Cámara de Diputadas y Diputados de Chile, para registros de actividad siendo la fecha, el número de sesión, el número de leyes, las comisiones, si es o no oficialista, la asistencia, la hora de llegada del diputado y la hora de inicio y termino de la sesión, etc. Por otro lado, la Biblioteca del Congreso Nacional de Chile (BCN) fue para la recolección de datos como la edad, el género, la

experiencia legislativa y previa en el mercado laboral privado, sus estudios, etc. La información se consolidó en una base de datos tipo Excel estructurada para su procesamiento.

4.5 Preparación y Limpieza de Datos

Una vez consolidada la base de datos, se procede a ejecutar un proceso riguroso que incluye la limpieza, la transformación de las variables independientes.

Tal y como se ha mencionado anteriormente esta investigación se desarrolla por medio del codificado de las variables para luego ser medidas por diversos modelos lineales como no lineales, para ello se debe hacer un preprocesamiento de las variables para que estén listas para utilizar en dichos modelos.

Para comenzar las variables categóricas o simplemente las variables que permiten clasificar en diversos grupos o categorías como, por ejemplo: género, partidos políticos, etc. Se transforman a través de un One-Hot Encoding, método que genera variables binarias para cada categoría y al mismo tiempo se elimina la primera columna para evitar la multicolinealidad, más conocida como la trampa dummy.

Para el caso de outliers o valores extremos se analizan todas las variables numéricas como es el caso de las intervenciones, el número de artículos, la edad, etc. Para todos estos casos se decide mantener estos valores ya que reflejan las diferencias reales de desempeño legislativo entre cada diputado o diputada, lo que permitirá realizar análisis más reales al final de la investigación, estos valores no son errores de medición, por lo cual su eliminación generaría un sesgo en el modelo.

También se realiza el escalado para las variables, donde se aplica la función StandardScaler la cual estandariza los valores de todas las variables independientes para asegurar que los modelos de regularización (Ridge, LASSO) funcionaran correctamente y que todas las variables tuvieran la misma escala numérica.

4.6 Estrategia de Modelado y Selección de Variables

La selección de las variables finales se realiza mediante una estrategia combinada que integra el diagnóstico de multicolinealidad y el análisis de importancia de variables. Se

descarta el uso de métodos de selección por regularización lineal como criterio definitivo de inclusión o exclusión, dado que estos métodos penalizan variables desde una perspectiva que no aplica a modelos no lineales como Random Forest, XGBoost o LightGBM. En consecuencia, para la selección de variables finales se consideran los siguientes criterios:

- a) Diagnóstico de Multicolinealidad: Se realiza un análisis preliminar utilizando el Factor de Inflación de la Varianza (VIF) con la función `statsmodels` para detectar relaciones lineales fuertes entre covariables, las cuales pueden aumentar la varianza de los estimadores en modelos lineales y distorsionar la interpretación de los coeficientes.
- b) Selección por Consenso de Métodos Lineales: Se aplican tres métodos de selección de forma independiente: LASSO (regularización L1), ElasticNet (combinación de regularización L1 y L2) y `SelectFromModel` basado en importancia de coeficientes. Se genera una tabla de frecuencia cruzando los resultados de los tres métodos y se conservan las variables seleccionadas por al menos dos de los tres criterios. Este enfoque garantiza que la inclusión de cada variable no dependa de las particularidades de un único método, sino que cuente con respaldo estadístico consistente. Cabe destacar que el criterio de selección considera la magnitud del coeficiente asignado por cada método, independientemente de su signo, dado que una variable con una asociación negativa y significativa aporta información predictiva equivalente a una con su contraparte, y excluirla por su dirección constituiría un error metodológico.
- c) Validación mediante Valores SHAP y Feature Importance: Como un paso más de comprobación, se utilizan los valores SHAP (del inglés *SHapley Additive exPlanations*, o Explicaciones Aditivas de Shapley) para verificar la coherencia entre las variables seleccionadas y su importancia real dentro del modelo final. SHAP es un método de interpretabilidad de modelos de aprendizaje automático fundamentado en la teoría de juegos cooperativos, que asigna a cada variable una contribución individual al valor predicho para cada observación. A diferencia de las métricas de importancia globales del Random Forest, que miden la reducción promedio de impureza, SHAP permite identificar tanto la magnitud como la

dirección del efecto de cada variable sobre la predicción, ofreciendo una medida más robusta y desagregada de la relevancia de cada predictor. Su inclusión en esta etapa tiene como objetivo confirmar que el conjunto de variables seleccionadas por los métodos lineales sea también coherente con la estructura interna del modelo no lineal principal, y que las conclusiones interpretativas no dependan de un único método de evaluación de importancia.

4.7 Modelo Final y Validación

Una vez reducido el set de variables, se procede a entrenar los modelos con una configuración y división del conjunto de datos de la siguiente forma: Se divide el conjunto de datos en un 70% para entrenamiento y un 30% para prueba, con semilla aleatoria fija para garantizar la reproducibilidad de los resultados.

Para identificar el modelo con mejor desempeño bajo las condiciones y variables del estudio, se evaluaron nueve algoritmos distintos: cuatro modelos lineales regularizados (Ridge, LASSO, ElasticNet y Regresión Lineal) y cinco modelos no lineales basados en árboles (Random Forest, Gradient Boosting, XGBoost y LightGBM). Esta comparación amplia permite establecer con evidencia cuál familia de modelos se ajusta mejor a la estructura de los datos, sin asumir la superioridad de ninguno de los métodos.

Cada modelo se evalúa mediante tres métricas complementarias. Primero, validación cruzada de cinco pliegues sobre el conjunto de entrenamiento, que permite estimar el desempeño del modelo en datos no vistos sin depender de una única partición. Segundo, R^2 en el conjunto de prueba, que mide la proporción de variabilidad de la variable dependiente explicada por el modelo sobre el 30% de datos reservados. Tercero, un análisis de robustez con 30 semillas aleatorias distintas, que evalúa la estabilidad de cada modelo ante diferentes particiones de los datos, reportando la media, desviación estándar, mínimo y máximo del R^2 obtenido. Los resultados completos de esta comparación se presentan en el capítulo de Resultados y Discusión.

4.8 Herramientas y Software

El desarrollo computacional de la investigación se lleva a cabo en un entorno de ciencia de datos utilizando el lenguaje de programación Python. A continuación, se

detallan las librerías utilizadas en cada etapa del proceso:

Manipulación y procesamiento de datos: Se utiliza la librería pandas para la gestión de DataFrames, limpieza y transformación de datos, NumPy se utiliza para operaciones vectoriales y álgebra lineal y SimpleImputer de scikit-learn para el tratamiento de valores faltantes.

Machine learning y modelado: Scikit-learn constituye el núcleo del análisis, utilizándose sus módulos para preprocesamiento (StandardScaler), división de datos (división de entrenamiento y testeo, KFold, puntaje de la validación cruzada), modelos lineales (Regresión Lineal, Ridge, Lasso, Elastic Net, Lasso CV, Elastic Net CV), modelos no lineales basados en árboles (Random Forest Regressor, Gradient Boosting Regressor, Decision Tree Regressor) y selección de variables (Select From Model). Adicionalmente, se utilizan XGBoost (XGB Regressor) y Light GBM (LGBM Regressor) como algoritmos de ensamble de alto rendimiento, y pyGAM (Linear GAM) para la estimación de modelos aditivos generalizados como análisis complementario.

Diagnóstico estadístico: statsmodels para el ajuste de modelos OLS y el cálculo del Factor de Inflación de la Varianza (VIF) mediante variance inflation factor.

Interpretabilidad: SHAP para el análisis de importancia de variables mediante valores de Shapley, y permutation importance junto a Partial Dependence Display de scikit-learn para la evaluación de la relevancia de variables y el análisis de dependencia parcial.

Visualización: Matplotlib y Seaborn para la generación de gráficos de diagnóstico y visualización de resultados.

Entorno de desarrollo: Google Colaboratory (Colab), que permite la ejecución reproducible del código en la nube con acceso a recursos computacionales sin configuración local.

4.9 Limitaciones Metodológicas

Si bien los modelos estimados permiten explicar una proporción relevante de la variabilidad, la investigación presenta limitaciones inherentes a la naturaleza de los datos:

Ausencia de variables cualitativas: El análisis se basa exclusivamente en resultados cuantificables (artículos, leyes, asistencia), dejando fuera aspectos cruciales como la calidad técnica de las iniciativas, su impacto normativo o la relevancia política de las leyes.

Roles informales: No es posible capturar mediante variables públicas el liderazgo político informal, la capacidad de negociación en “pasillos” o la influencia en redes de poder no registradas en actas, factores que afectan significativamente la productividad real.

Factores contextuales: El modelo no incorpora eventos externos (coyunturas políticas, crisis sanitarias o económicas) que pueden condicionar la agenda legislativa en periodos específicos, independientemente de la voluntad del diputado.

Datos disponibles: La investigación está limitada a la información pública y estructurada que reportan las instituciones oficiales, datos privados o no reportados no pueden ser considerados en el análisis.

Período único de análisis: el estudio cubre solo las legislaturas 370 a 373 (2022-2026), que corresponden a un gobierno de izquierda con una derecha en oposición. Las relaciones de las coaliciones encontradas pueden ser específicos de este contexto político y no generalizables a otros períodos donde la composición del gobierno sea distinta.

Corte temporal de los datos: la última medición es el 3 de noviembre de 2025, vale decir, el período no está completo. Diputados que ingresaron más recientemente tienen menos tiempo de observación, y aunque el índice normaliza por tiempo activo, los últimos meses del período quedan fuera del análisis.

Subjetividad en la clasificación de leyes: el criterio de tres niveles de importancia (1, 2, 3) basado en los cortes de complejidad normativa introduce un componente de decisión del investigador. Si bien se fundamenta en criterios observables, otra clasificación podría arrojar resultados distintos.

Solo mociones parlamentarias: el estudio considera únicamente las leyes originadas en mociones de diputados, excluyendo los mensajes del Ejecutivo. Esto implica que leyes de alto impacto impulsadas por el gobierno no están capturadas en el índice de productividad, lo que puede sesgar la comparación entre bloques con distinta relación con el poder ejecutivo.

Tamaño muestral reducido: con 157 observaciones, el análisis de robustez con 30 semillas es una medida necesaria pero no suficiente para garantizar la estabilidad de los resultados en muestras de mayor tamaño. Esto limita la generalización estadística de los hallazgos.

5 Resultados y Discusión

5.1 Análisis Descriptivo de los Datos

La muestra final para el análisis de productividad se compone de 157 diputados que ejercieron el cargo durante el período comprendido entre el 11 de marzo de 2022 y el 3 de noviembre de 2025. Este apartado caracteriza a la población en estudio, explorando sus atributos demográficos, académicos y la distribución inicial de la variable objetivo antes de la aplicación de los modelos predictivos.

5.1.1 Composición Demográfica

El parlamento chileno analizado presenta una estructura demográfica marcada por una predominancia masculina y una edad madura. De los 157 diputados, 101 son hombres (64,3%) y 56 son mujeres (35,7%), lo que refleja que aún existe un desequilibrio en la representación parlamentaria, por temas de brechas de género. La edad promedio de los legisladores es de 51,3 años con una mediana de 50,6 años, y el 50% de los diputados se concentra entre los 40,7 y los 59,3 años, dando a entender que la mayoría está en una etapa adulta y más experimentada de la vida, pero aun así existen diputados de todo el rango etario proporcionando así distintas visiones de los temas a tratar.

En el caso de la experiencia legislativa o mejor dicho en el congreso, se tiene una media de 6,0 años, lo que da a entender que en promedio cada diputado lleva más de un periodo completo, pero la realidad es que esto está considerado casi al final del periodo analizado, por ende, casi todos llevan cerca de 4 años cumplidos, exceptuando los diputados que se unieron posteriormente.

En el caso de la experiencia laboral en empresas privadas se obtiene que existe una media de 12 años trabajados en la industria privada por parte de los diputados, lo que quiere decir que la mayoría a obtenido experiencia fuera del mundo público logrando así tener una visión más amplia del país.

Tabla 3. Estadísticas descriptivas.

Variable	Media	Mediana	Desv. Est.	Mín.	Máx.
Edad [años]	51,3	50,6	12,2	28,9	82,8

Género masculino	0,64	1,0	0,48	0,0	1,0
Experiencia legislativa [años]	6,0	3,7	3,8	0,5	27,7
Experiencia laboral privada [años]	12,0	10,0	7,6	0,0	30,0

5.1.2 Perfil Académico y Electoral

El nivel de formación es construido mediante una escala de valores asignados a diversos niveles educativos quedando valores de 0 a 1. Esta variable revela una heterogeneidad considerable en el capital educativo de los legisladores, ya que se aprecian distintos tramos educativos de los diputados. El primer tramo encontrado esta entre los puntajes de 0 a 0,39, en el cual se puede notar que están las personas que terminaron sus estudios desde la enseñanza media hasta un título profesional con algún diplomado. El segundo tramo es de 0,4 a 0,59, donde los legisladores ya poseen el grado de un magister en sus estudios, obteniendo un caso donde solo egreso de la universidad, pero los otros 36 casos tienen por lo menos un título profesional y en algunos casos diplomados. Y el tercer tramo de 0,6 a 1,0 son personas que realizan estudios académicos de alto calibre como lo es un doctorado, además de obtener el grado de magister y título profesional sumando tres candidatos y un último candidato que alcanzo todos los niveles académicos propuestos en esta investigación. El puntaje promedio es de 0,27, con una desviación estándar de 0,2, lo que indica que el promedio de los diputados alcanza el nivel educativo de título profesional con algún diplomado y si consideramos una desviación estándar se llega al segundo tramo encontrado, pero llega al punto de tener magister con un egreso o un técnico con algún diplomado, no con un título profesional.

Básicamente se hizo una ponderación de la cantidad de personas electas por distrito por la proporción de votos que tuvo el candidato, es por esto por lo que hay diputados con valores mayor a 1, donde el caso es que tuvieron una gran proporción de los votos del distrito aun con mucha competencia elegida.

En el caso de la variable nivel de votación presenta una media de 0,42 y una desviación estándar de 0,26. Lo que significa que la media de diputados tiene cerca de un 42% ponderado de sus respectivos distritos, esto ya que la variable se normaliza multiplicando por la cantidad de candidatos elegidos en dicho distrito, y esto a su vez está

calculado por la cantidad de gente que está en ese distrito. Además, se puede notar que la gran mayoría de los datos están cerca de la media, pero aun así existen datos que se escapan totalmente de los demás. Estos son mayores a 1 y esto se da porque existen candidatos que obtuvieron grandes cantidades de votos en distritos con mucha población, lo que a su vez aumenta el número de candidatos elegidos en dicho distrito.

Otro análisis por mencionar es que la mayoría de los candidatos con una cantidad de votos sobresalientes o mayores que 1 son de la coalición de izquierda tanto tradicional como progresista, esto es destacable, ya que se podría deducir que los partidos de izquierda logran llegar a consensos mayores que los partidos de centro o de derecha.

Tabla 4. Variable académica y electoral.

Variable	Media	Mediana	Desv. Est.	Mín.	Máx.
Nivel de Formación	0,27	0,20	0,21	0,00	1,00
Nivel de Votación	0,42	0,36	0,26	0,0	1,65

Tabla 5. Valores del nivel de formación de los diputados con su cantidad de veces encontradas.

Nivel de Formación	Cantidad
0	14
0,05	6
0,1	16
0,15	1
0,2	63
0,25	16
0,42	1
0,52	26
0,57	10
0,95	3
1	1

5.1.3 Distribución del Índice de Productividad y Afiliación Política

El índice de productividad esta ponderado a través del tiempo en el cargo que ha disputado el diputado en el periodo de estudio y esté presenta una distribución con asimetría positiva y cola derecha extendida, lo que significa que la productividad legislativa está altamente concentrada en un pequeño grupo de diputados que destaca ampliamente frente a la mayoría de los legisladores que se sitúan en valores bajos del

índice.

En el caso del resto de variables que se enfocan en datos obtenidos de la cámara se puede notar que en todas existe una media bastante menor a los valores máximos respectivos, lo que confirma lo mismo apreciado con el índice de productividad. No es así con los valores mínimos donde estos o pueden ser pequeños por bajo desempeño o es por los nuevos diputados que se unieron posteriormente al cargo.

Y en el caso de la variable asistencia se puede ver que todos los diputados asisten a sus sesiones, lo cual no indica que estén presentes durante toda la sesión, ya que la asistencia dentro de la cámara se mide si llegan, aunque sea los últimos cinco minutos. Esto no es así con la variable de puntualidad donde se considera si llega dentro de la hora estipulada de la sesión, es por eso por lo que se aprecia que decae mucho los valores de esta variable en comparación con la anterior, aunque aun así hay casos excepcionales que llegan casi todas las sesiones del periodo a la hora asignada.

Tabla 6. Estadísticas descriptivas de las variables de actividad parlamentaria.

Variable	Media	Mediana	Desv. Est.	Mín.	Máx.
Leyes Publicadas	2,66	2,0	2,1	0,0	9,0
Proyectos de Ley Presentados	78,6	71,0	47,3	4,0	234,0
Intervenciones	127,9	113,0	79,4	2,0	412,0
Intervención en Comisión	89,2	76,0	63,1	0,0	318,0
Cantidad de Comisiones	3,60	4,0	1,4	0,0	7,0
Asistencia	0,971	0,978	0,039	0,81	1,0
Puntualidad	0,584	0,591	0,191	0,06	0,97
Índice de Productividad	1,08	0,85	0,93	0,0	5,64

5.2 Diagnóstico Metodológico: Modelo con Datos Fallidos

Antes de presentar el modelo explicativo principal, se busca mencionar que hubo un modelo inicial que incluía la variable artículos tramitados como predictor independiente, con el objetivo de documentar un problema metodológico relevante detectado durante el análisis.

Este modelo alcanzó un R^2 de 0,631 en el conjunto de prueba con Random Forest,

implicando que este modelo se puede predecir con un altísimo nivel de precisión para ser una investigación de este tipo. Sin embargo, el análisis de importancia de variables reveló que la variable artículos tramitados concentraba el 74,9% del poder predictivo del modelo, siendo su permutation importance de 1,36 frente a solo un 0,25 para la segunda variable más importante, lo cual implicaba que existía un sesgo absoluto de una variable para explicar el modelo, por lo que se toma la decisión de incluir esta variable como parte de la creación de la variable dependiente, siendo excluida del modelo explicativo final.

Tabla 7. Importancia de las variables en el modelo inicial de Random Forest.

Variable	Feature Importance	Permutation Importance	Importancia acumulada
Artículos Tramitados	0,750	1,361	74,95%
Proyectos de Ley Presentados	0,115	0,247	86,48%
Intervenciones	0,035	0,029	89,97%
Intervención en Comisión	0,024	0,020	92,34%
Nivel de Votación	0,018	0,015	94,18%
Resto de variables	0,058	0,087	100,00%

5.3 Selección de Variables

Antes de presentar el desempeño comparativo de los modelos seleccionados, en esta sección se describen qué variables fueron consideradas como predictoras y bajo qué criterios se determinó su inclusión en el modelo final, añadiendo al final las variables descartadas y sus razones. La importancia de variables se evalúa mediante tres métodos complementarios: diagnóstico de multicolinealidad mediante el Factor de Inflación de la Varianza (VIF), feature importance y permutation importance del Random Forest, y consenso entre tres métodos lineales de selección (LASSO, ElasticNet y SelectFromModel).

5.3.1 Diagnóstico de Multicolinealidad (VIF)

El Factor de Inflación de la Varianza permite detectar relaciones lineales fuertes entre las variables predictoras, las cuales pueden inflar la varianza de los estimadores en modelos lineales. La Tabla 9 presenta los valores de VIF obtenidos para el conjunto final de

variables, una vez excluidas aquellas que generaban data leakage, para esto se consideran variables que apuntaban a lo mismo, como por ejemplo la asistencia y la puntualidad o la edad con las experiencias, dado que ambas tienen un comportamiento muy similar dado a que mientras más edad tengas, más años tienes de experiencia en el promedio.

Tabla 8. Factor de inflación de la varianza (VIF) para todas las variables.

Variable	VIF
Puntualidad	11,23
Cantidad de Comisiones	9,74
Proyectos de Ley Presentados	7,41
Intervenciones	4,83
Nivel de Votación	4,42
Experiencia	4,40
Experiencia Laboral	3,69
Nivel de Formación	3,67
Coalición Derecha Tradicional	3,58
Género Masculino	3,46
Coalición Izquierda Progresista	3,05
Intervención en Comisión	2,91
Coalición Izquierda Tradicional	2,42
Coalición Derecha Radical	2,32
Preside Comisión	1,65

En la Tabla 9 se puede ver que hay un valor que supera 10, lo que indica que existe multicolinealidad aun así se busca tener en cuenta otros métodos de análisis de variables dado que se trabaja con modelos tanto lineales como no lineales, de los cuales los últimos dentro de su programación no les genera complicaciones las multicolinealidades, además de que parte de la investigación está centrada en analizar si la asistencia o puntualidad tienen un impacto positivo o negativo en la productividad legislativa. También variables

como puntualidad y cantidad de comisiones, que presentan los VIF más elevados del conjunto, son al mismo tiempo dos de las variables con mayor poder explicativo según los métodos de importancia descritos más adelante. Tal como se discutió en la metodología, dado que el modelo principal corresponde a Ridge, una técnica de regularización que penaliza explícitamente la magnitud de los coeficientes, la presencia de multicolinealidad moderada no compromete la estabilidad de las estimaciones, ya que la regularización L2 reduce precisamente la varianza que la colinealidad tendería a inflar en una regresión lineal sin penalización.

5.3.2 Importancia de Variables según Random Forest

Como análisis complementario al modelo lineal, se entrena un Random Forest sobre el mismo conjunto de variables con el fin de contrastar el ordenamiento de importancia obtenido por un método no lineal. La Tabla 10 presenta la importancia intrínseca del modelo (feature importance) junto con la importancia por permutación, calculada de forma independiente a la estructura del árbol. Además, con el propósito de complementar el análisis de los modelos predictivos e identificar las contribuciones explicativas de cada variable sobre la productividad legislativa, se estimó un modelo Random Forest y posteriormente se calcularon los valores SHAP (SHapley Additive exPlanations). Esta técnica permite interpretar modelos de aprendizaje automático cuantificando el aporte individual de cada variable a las predicciones realizadas y queda plasmada en la Figura 3.

Tabla 9. Importancia de las variables por Random Forest.

Variable	Feature Importance	% Acumulado	Perm. Importance
Proyectos de Ley Presentados	0,3750	37,50%	0,8834
Intervención en comisión	0,1371	51,21%	0,2010
Puntualidad	0,0988	61,10%	0,1172
Intervenciones	0,0881	69,90%	0,1069
Nivel de Votación	0,0753	77,43%	0,0825
Nivel de Formación	0,0467	82,11%	0,0498

Coalición Izquierda Tradicional	0,0376	85,87%	0,0689
Experiencia Laboral	0,0405	89,92%	0,0332
Cantidad de Comisiones	0,0354	93,45%	0,0405
Resto de variables	0,0655	100,00%	0,0915

Las primeras dos variables, proyectos de ley presentados e intervención comisión, concentran en conjunto el 51,2% de la importancia total del modelo, lo que confirma que la iniciativa legislativa y el trabajo técnico en comisiones son los dos pilares dominantes de la productividad. Resulta especialmente relevante que la permutation importance de proyectos de ley presentados (0,8834) sea casi cinco veces mayor que la de la segunda variable (0,2010), una brecha considerablemente más amplia que la observada en feature importance, lo que proporciona evidencia adicional sobre la relevancia predictiva de esta variable, dado que la permutation importance cuantifica el cambio en el error de predicción cuando se rompe la relación estadística entre la variable y el resultado, y no únicamente su contribución estructural dentro de los árboles.

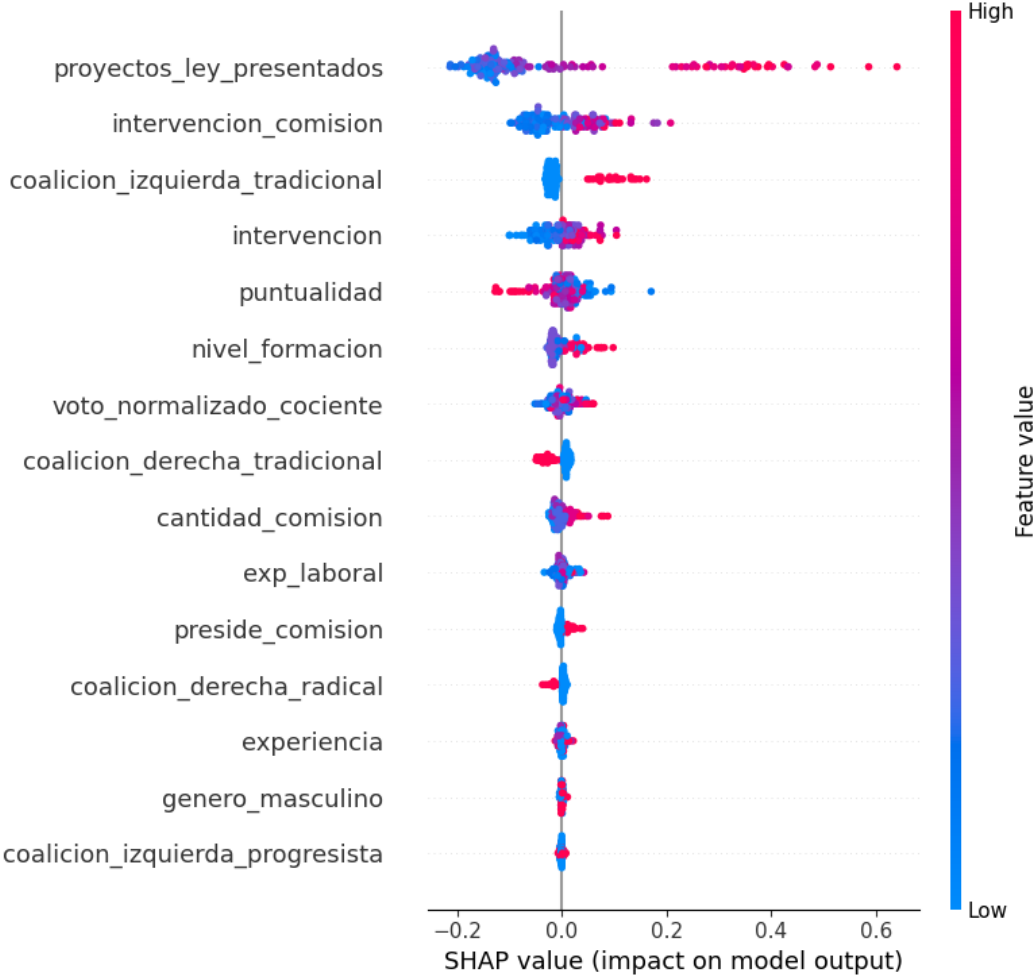


Ilustración 3. Importancia de variables obtenida mediante valores SHAP para el modelo Random Forest.

5.3.3 Selección por Consenso de Métodos Lineales

Siguiendo el criterio metodológico descrito en el capítulo anterior, se aplican los tres métodos de selección lineal (LASSO, ElasticNet y SelectFromModel) y se conservan como variables finales aquellas seleccionadas por al menos dos de los tres métodos. La Tabla 11 resume este proceso de consenso.

Tabla 10. Resumen de la selección de variables por consenso de métodos lineales.

Variable	LASSO	ElasticNet	SelectFromModel	Suma Total
Proyectos de Ley Presentados	Sí	Sí	Sí	3
Nivel Formación	Sí	Sí	Sí	3
Intervención en Comisión	Sí	Sí	Sí	3
Cantidad de Comisiones	Sí	Sí	Sí	3

Coalición Izquierda Tradicional	Sí	Sí	Sí	3
Coalición Derecha Radical	Sí	Sí	Sí	3
Coalición Derecha Tradicional	Sí	Sí	Sí	3
Coalición Izquierda Progresista	Sí	Sí	Sí	3
Puntualidad	Sí	Sí	No	2
Nivel de Votación	Sí	Sí	No	2
Experiencia	No	No	No	0
Preside Comisión	No	No	No	0
Intervención	No	No	No	0
Experiencia Laboral	No	No	No	0
Género Masculino	No	No	No	0

Ocho variables son seleccionadas de forma unánime por los tres métodos, lo que les otorga el respaldo más robusto dentro del conjunto: proyectos de ley presentados, nivel de formación, intervención en comisión, cantidad comisión y las cuatro variables de coalición política con mayor diferenciación respecto al centro. Dos variables adicionales, puntualidad y nivel de votación, son seleccionadas por dos de los tres métodos, quedando igualmente incorporadas al modelo final por cumplir el umbral de consenso definido. En contraste, cinco variables no son seleccionadas por ningún método: experiencia, preside comisión, intervención, experiencia laboral y género masculino. Esto no significa necesariamente que dichas variables carezcan de toda relación con la productividad, sino que su el grado de explicación al que llegan, una vez controlado por las demás variables del modelo, no resulta suficientemente distinto de cero para ser considerado un predictor relevante bajo los tres criterios lineales aplicados. Es importante notar que algunas de estas variables sí mostraron coeficientes pequeños y no nulos en el análisis de sensibilidad de Ridge, lo que se profundiza en la sección 5.5.

5.3.4 Variables Eliminadas por Otras Razones

En esta subsección se consideran tres variables que previamente fueron eliminadas al análisis anterior, estas son: Oficialismo, Partido y Sindicato.

En el caso de oficialismo y partido fueron eliminadas del modelo porque generaban

mucha multicolinealidad con la variable de las coaliciones, por lo que afectaba mucho antes de los resultados por ende fue eliminada.

Y el caso de la variable sindicato fue ya que no generaba ningún valor en el modelo por lo tanto solo generaba ruido, es por esto por lo que tambien fue eliminada previamente a realizar los resultados finales.

5.4 Evaluación Comparativa de Modelos

Una vez definido el conjunto final de variables predictoras, se entrenan ocho algoritmos para identificar cuál ofrece el mejor desempeño bajo las condiciones del estudio. La Tabla 12 presenta los resultados de la evaluación mediante validación cruzada de cinco pliegues (R^2 CV), R^2 en el conjunto de prueba y RMSE. La Tabla 13 complementa este análisis con los resultados del análisis de robustez con 30 semillas aleatorias.

Tabla 11. Resultados y desempeño comparativo de los modelos estimados.

Modelo	R^2 CV	R^2 Test	RMSE
Ridge	0,057	0,164	0,313
ElasticNet	0,049	0,167	0,471
LASSO	0,046	0,145	0,317
Regresión Lineal	0,023	0,146	0,317
Random Forest	0,009	0,084	0,328
Gradient Boosting	-0,150	0,066	0,331
XGBoost	-0,146	-0,097	0,359
LightGBM	-0,283	-0,039	0,349
Árbol de Decisión	-0,476	-0,009	0,344

Los cuatro modelos lineales regularizados (Ridge, ElasticNet, LASSO y Regresión Lineal) ocupan las primeras cuatro posiciones en R^2 test, todos por encima de 0,148, mientras que los cinco modelos no lineales presentan un desempeño notoriamente más bajo e incluso negativo en el caso de Gradient Boosting, XGBoost, LightGBM y el Árbol de Decisión en la validación cruzada R^2 . Un R^2 de validación cruzada negativo indica que, en promedio, el modelo se desempeña peor que simplemente predecir la media de la variable

dependiente, lo que sugiere sobreajuste severo dado el tamaño muestral reducido (157 observaciones) en relación con la complejidad de estos algoritmos. El modelo Ridge obtiene el mejor R^2 test (0,164) y RMSE más bajo (0,313) dentro de su categoría, consolidándose como el candidato más sólido para la interpretación de coeficientes. Y para comprobar esto se procede a realizar un análisis con 30 simulaciones diferentes y realizando promedios de estas.

Tabla 12. Análisis de robustez con 30 semillas aleatorias.

Modelo	Media	Desv. Est.	Mínimo	Máximo
Ridge	0,250	0,129	-0,150	0,461
ElasticNet	0,237	0,105	-0,062	0,410
Linear	0,233	0,144	-0,195	0,427
LASSO	0,206	0,116	-0,147	0,379
Random Forest	0,205	0,162	-0,300	0,399
LightGBM	0,103	0,178	-0,392	0,349
Gradient Boosting	0,057	0,220	-0,504	0,407
XGBoost	-0,004	0,276	-0,717	0,374
Árbol de Decisión	-0,275	0,487	-1,791	0,415

El análisis de robustez confirma y refuerza el ordenamiento observado con una sola partición de datos. Ridge no solo obtiene la media de R^2 más alta (0,250) entre las 30 semillas evaluadas, sino que además presenta una de las desviaciones estándar más bajas del conjunto (0,129), lo que indica que su desempeño es consistente independientemente de qué observaciones caigan en el conjunto de entrenamiento o de prueba. En el extremo opuesto, el Árbol de Decisión exhibe una desviación estándar de 0,487, casi de cuatro veces superior a la de Ridge, con un mínimo de -1,791, evidenciando una inestabilidad extrema ante distintas particiones de los datos. Esta diferencia de estabilidad es tan importante como la diferencia de desempeño promedio: un modelo cuyo R^2 puede variar entre -1,791 y 0,415 según la muestra no ofrece bases confiables para la interpretación de resultados, mientras que Ridge mantiene un comportamiento predecible y reproducible, razón adicional para seleccionarlo como modelo principal de esta investigación.

5.5 Análisis de los Coeficientes del Modelo Ridge

Los coeficientes estandarizados del modelo Ridge permiten cuantificar la asociación parcial entre cada variable y el índice de productividad, manteniendo constantes las demás variables del modelo. Al estar estandarizados, son directamente comparables entre sí independientemente de las unidades originales de cada variable. La Tabla 14 presenta el ordenamiento completo de predictores, tomando como categoría de referencia la coalición de centro.

Tabla 13. Coeficientes estandarizados del modelo Ridge, con la referencia de coalición centro para las coaliciones.

Variable	Coef. estandarizado	Dirección
Proyectos de Ley Presentados	+0,285	↑ Positivo
Coalición Derecha Tradicional	-0,184	↓ Negativo
Coalición Derecha Radical	-0,180	↓ Negativo
Coalición Izquierda Progresista	-0,113	↓ Negativo
Cantidad de Comisiones	+0,078	↑ Positivo
Intervención en Comisión	+0,050	↑ Positivo
Nivel de Formación	+0,050	↑ Positivo
Nivel de Votación	+0,037	↑ Positivo
Coalición Izquierda Tradicional	-0,031	↓ Negativo
Puntualidad	-0,022	↓ Negativo
Intervención	+0,019	↑ Positivo
Experiencia	-0,008	≈ Neutro
Género Masculino	+0,006	≈ Neutro
Experiencia Laboral	-0,003	≈ Neutro
Preside Comisión	+0,001	≈ Neutro

Antes de analizar cada variable, es importante precisar cómo deben interpretarse los coeficientes con signo negativo, ya que un error común es asumir que un coeficiente negativo implica que la variable carece de relevancia en el modelo. Ocurre exactamente lo contrario: un coeficiente negativo indica que la variable sí es relevante y que su relación

con la variable dependiente es estadísticamente significativa, solo que dicha relación es inversamente proporcional. Esto significa que, a medida que el valor de la variable aumenta, el índice de productividad predicho disminuye, mientras todas las demás variables permanecen constantes. La magnitud del coeficiente, independientemente de su signo, indica la intensidad de la asociación estadística entre la variable y el índice de productividad. Así, un coeficiente de $-0,184$ representa una asociación de mayor magnitud que uno de $+0,050$, aunque el primero sea negativo y el segundo positivo. En otras palabras, mientras mayor es el valor absoluto de un coeficiente negativo, mayor es la caída esperada en la variable dependiente ante un aumento de esa variable independiente. Por esta razón, en la Tabla 14 las variables se ordenan por magnitud absoluta del coeficiente y no por su signo, de modo que coalición derecha tradicional ($-0,184$) aparece como el segundo predictor más pronunciado del modelo, justo después de proyectos de ley presentados, pese a tener signo negativo.

En el caso de la iniciativa legislativa la variable de proyectos de ley presentados es el predictor de mayor magnitud con un coeficiente de $+0,285$, siendo más del doble que el siguiente predictor relevante. Un diputado que presenta proyectos de ley significativamente por encima del promedio tiene, en términos esperados, un índice de productividad $0,285$ unidades mayor que uno promedio, con todas las demás características constantes. Este resultado establece que la iniciativa legislativa activa es el factor individual que más explica estadísticamente la productividad en la investigación en el período analizado.

En el caso de las variables de coalición se presentan los valores más heterogéneos del modelo y deben leerse con especial atención a su magnitud y signo. Tomando como referencia la coalición de centro, todos los demás bloques muestran coeficientes negativos, lo que significa que, en comparación con un diputado de centro con características idénticas en las demás variables, pertenecer a cualquier otro bloque político reduce el índice de productividad esperado. La derecha tradicional muestra el coeficiente más negativo del modelo completo ($-0,184$), seguida muy de cerca por la derecha radical ($-0,180$), lo que indica que ambos bloques son los más fuertemente asociados a una menor productividad relativa, aun así, superan incluso a variables de comportamiento parlamentario como cantidad de comisiones o intervención en comisión en magnitud

absoluta. La izquierda progresista presenta un coeficiente de $-0,113$, una caída esperada considerablemente menor que la de los bloques de derecha, pero igualmente negativa. La izquierda tradicional, con $-0,031$, es la que más se aproxima al rendimiento del centro, siendo la coalición más débil en términos de magnitud del conjunto. Además, cabe destacar que como el periodo en estudio es de un sector de izquierda los resultados encontrados tengan relación directa, esto es solo una teoría o conjetura que para probarla habría que realizar el mismo estudio, pero con diversos periodos de estudio considerando los diferentes sectores políticos.

En el caso de la participación institucional la cantidad de comisiones presenta un coeficiente de $+0,078$, siendo el tercer predictor en magnitud dentro de las variables de comportamiento parlamentario. Lo que indica que participar en un mayor número de comisiones se asocia positivamente con la productividad. La variable intervención en comisión presenta un coeficiente de $+0,050$, lo que indica una asociación positiva con el índice de productividad, aunque de menor magnitud que la observada para la iniciativa legislativa bruta. Es interesante notar que la brecha entre intervención en comisión ($+0,050$) e intervención en sala ($+0,019$) es de más del doble, lo que es coherente con el hallazgo de que el trabajo en comisiones tiene una mayor conexión con la productividad que con la visibilidad pública en el pleno.

En el caso del capital humano y de los perfiles de los diputados, el nivel de formación presenta un coeficiente de $+0,050$, indicando que una mayor formación académica se relaciona positivamente con el índice de productividad, con una magnitud equivalente a la observada para la intervención en comisión. El nivel de votación presenta un coeficiente de $+0,037$, lo que indica una relación positiva de menor magnitud con el índice de productividad. Esta asociación puede interpretarse como un reflejo del capital político disponible para impulsar y negociar iniciativas legislativas, sin que ello implique una relación causal.

Para el caso de la puntualidad la que presenta un coeficiente de $-0,022$, significa que una mayor puntualidad de los diputados se asocia con una caída pequeña en el índice de productividad esperado. Este resultado contraintuitivo con el inicio de la investigación es coherente con la hipótesis de que los diputados con mayor carga de trabajo legislativo y

múltiples responsabilidades simultáneas presentan naturalmente menor puntualidad formal en algunas sesiones, sin que ello comprometa su productividad real, esto no quiere decir que la relación inversa implica que la impuntualidad cause productividad, sino que simplemente ambas podrían ser consecuencia de una agenda parlamentaria más intensa.

Finalmente, las variables como experiencia (-0,008), género masculino (+0,006), experiencia laboral (-0,003) y preside la comisión (+0,001) presentan coeficientes cuya magnitud absoluta es inferior a 0,01, es decir, prácticamente nula en términos prácticos independientemente de su signo. A diferencia de los casos anteriores, donde el signo negativo señalaba una relación inversa relevante, en estas cuatro variables la cercanía a cero indica una ausencia de relación discernible sobre la productividad una vez controladas las demás dimensiones del modelo. Esto es consistente con que ninguna de estas cuatro variables fue seleccionada por los métodos de consenso lineal descritos en la Tabla 11, reforzando la coherencia entre ambos análisis.

5.6 Análisis Comparativo de los Diputados más Productivos

Con el objetivo de complementar los hallazgos del modelo, se realiza un análisis descriptivo comparativo entre los grupos de mayor productividad (top 10, top 15 y top 20) según el índice y el promedio general de la cámara. La Tabla 15 presenta las diferencias porcentuales en las principales variables cuantitativas y la Tabla 16 muestra la distribución de coaliciones en cada grupo.

Tabla 14. Diferencias porcentuales respecto al promedio general por cada grupo de productividad.

Variable	Media general	Top 10	Top 15	Top 20	Δ% T10	Δ% T20
Leyes Publicadas	2,66	7,00	6,60	6,30	+163%	+137%
Proyectos de Ley Presentados	78,6	116,7	118,1	111,7	+48%	+42%
Intervención en Comisión	89,2	138,9	160,9	133,7	+56%	+50%
Intervenciones (sala)	127,9	152,4	165,3	153,7	+19%	+20%
Cantidad de Comisiones	3,60	4,60	4,47	4,30	+28%	+19%
Nivel de Formación	0,355	0,580	0,505	0,457	+63%	+29%
Nivel de Votación	0,425	0,454	0,518	0,497	+7%	+17%

Asistencia	0,971	0,985	0,978	0,978	+1,5%	+0,8%
Puntualidad	0,584	0,483	0,521	0,551	-17%	-6%
Experiencia Laboral	11,95	10,40	8,53	9,40	-13%	-21%

Las variables de comportamiento parlamentario son las que presentan las brechas más amplias y consistentes entre el top y el promedio general, lo que confirma a nivel descriptivo el mismo patrón identificado por los coeficientes del modelo Ridge. Los proyectos de ley presentados en la cámara se mantienen entre 42% y 48% por encima del promedio en los tres grupos analizados, una diferencia notablemente estable que respalda su rol como el predictor más robusto de la productividad. En el caso de las intervenciones en alguna comisión presenta la brecha relativa más amplia de todas las variables de actividad, alcanzando +56% en el top 10 y +50% en el top 20, muy por encima de la brecha observada en intervención de sala (+19% a +20%), lo que confirma nuevamente que el trabajo técnico en comisiones discrimina con mayor fuerza la productividad que la visibilidad en el pleno.

Entre las variables de perfil individual, nivel de formación exhibe la diferencia más marcada del conjunto: los diputados del top 10 presentan un nivel de formación 63% superior al promedio general, brecha que se reduce a 29% en el top 20, lo que sugiere que la formación académica diferencia con especial fuerza a los diputados de más alto desempeño relativo, aunque su relación se diluye levemente al ampliar el grupo de comparación. El nivel de votación muestra el patrón inverso, aumentando de +7% en el top 10 a +17% en el top 20, lo que indica que el respaldo electoral tiene mayor peso diferenciador en grupos más amplios que en el núcleo de los diputados más productivos.

Dos variables presentan un comportamiento que confirma directamente los hallazgos de coeficientes negativos descritos en la sección 5.5. La asistencia prácticamente no varía entre el top y el promedio general (entre +0,8% y +1,5%), lo cual es coherente con que esta variable nunca fue seleccionada como predictor relevante por ningún método de selección, dada su escasa varianza en el conjunto completo de diputados. Y la puntualidad, en cambio, es sistemáticamente inferior en los grupos de mayor productividad respecto al promedio general, con una caída de 17% en el top 10 que se atenúa a 6% en el top 20, replicando exactamente la dirección negativa que mostró su coeficiente en el modelo Ridge

(-0,022) y reforzando la interpretación de que los diputados más activos legislativamente presentan, en promedio, menor puntualidad formal lo que puede indicar que los mejores diputados tienen más responsabilidades provocando que no sean tan puntuales como sus colegas que no son tan productivos. La experiencia laboral exhibe igualmente un comportamiento negativo (-13% a -21%), consistente con su coeficiente cercano a cero, pero de signo negativo en el modelo, lo que sugiere que una mayor trayectoria en el sector privado no se asocia con mayor productividad legislativa, e incluso muestra una asociación levemente inversa en este análisis descriptivo.

Tabla 15. Distribución de las coaliciones políticas por grupo de productividad y cantidades de diputados de dicha coalición dentro de los tops.

Coalición	% Congreso general	Top 10 (n)	Top 15 (n)	Top 20 (n)
Centro	12,1%	3	5	5
Derecha tradicional	28,7%	4	6	7
Izquierda tradicional	16,6%	2	3	6
Izquierda progresista	27,4%	1	1	2
Derecha radical	15,3%	0	0	0

El centro está fuertemente sobrerrepresentado en todos los grupos: con solo el 12,1% de representación en el congreso general, concentra 3 diputados en el top 10, 5 en el top 15 y en el top 20. La derecha radical está completamente ausente en los tres grupos pese a representar el 15,3% del congreso. La izquierda progresista, siendo el bloque más grande con 27,4%, tiene solo 1 representante dentro del top 10 y el top 15, y 2 en el caso del top 20. Esto evidencia que las coaliciones que están más cercanas al centro tienen mejores resultados en la productividad legislativa del país. Además, estos patrones son directamente consistentes con los coeficientes del modelo Ridge.

5.7 Perfil del Diputado Modelo: Características Individuales Asociadas a la Productividad Legislativa

Una pregunta alterna al centro de esta investigación es la capacidad de lograr relacionar qué características individuales, vale decir, aquellas atribuibles a la persona antes de asumir el cargo, se asocian con un mayor desempeño legislativo. Este análisis

tiene una aplicación práctica concreta: orientar la selección de candidatos a diputado priorizando aquellos atributos que, según la evidencia empírica recopilada, se asocian con una mayor productividad esperada en términos de leyes aprobadas e importantes.

Para ello, se integran dos fuentes de información complementarias: los coeficientes estandarizados del modelo Ridge, que cuantifican la asociación parcial de cada variable manteniendo constantes las demás, y el análisis comparativo de los grupos de mayor productividad (top 10, top 15 y top 20), que permite observar los patrones descriptivos que distinguen a los diputados más productivos del resto de la cámara. Ambas fuentes convergen en un perfil coherente que se detalla a continuación por medio de las variables.

Género: No se aprecia una diferencia estadísticamente relevante para poder definir si es mejor el femenino o el masculino.

El modelo Ridge asigna un coeficiente de +0,000 a la variable género masculino, una magnitud prácticamente nula que indica que el género no tiene una capacidad explicativa sobre la productividad legislativa una vez controladas las demás variables. Este resultado se refuerza con el análisis descriptivo del top 10, donde la distribución es de 6 hombres y 4 mujeres, una proporción notablemente más paritaria que la del congreso general, donde los hombres representan el 64,3% de los diputados. La evidencia apunta, por tanto, a que el género no constituye un predictor de productividad, y que las brechas observables en datos brutos entre hombres y mujeres se explican por diferencias en el acceso a comisiones, en la cantidad de proyectos presentados y en otras variables de comportamiento parlamentario, y no por el género en sí mismo. Por lo tanto, el diputado modelo puede ser hombre o mujer indistintamente, dado que, en condiciones equivalentes de actividad parlamentaria, la productividad esperada es la misma.

Edad: levemente inferior al promedio de la cámara

La variable edad fue excluida del modelo final por su alta multicolinealidad con experiencia legislativa (VIF de 41,9 en el análisis inicial), por lo que no se cuenta con un coeficiente directo que cuantifique su asociación parcial. Sin embargo, el análisis descriptivo comparativo revela que los diputados del top 10 son en promedio entre 3 y 4 años más jóvenes que el promedio general de la cámara, que es de 51,3 años. Esto sugiere que el rango etario asociado a mayor productividad se ubica aproximadamente entre los 45 y los 50 años, es decir, en la etapa media de la vida profesional y política de un legislador. Una interpretación plausible es que diputados en ese rango han acumulado suficiente experiencia y capital político como para negociar y hacer avanzar proyectos, pero aún se

encuentran en una fase activa de construcción de su carrera legislativa que los incentiva a generar resultados concretos. Los extremos del espectro etario, tanto los diputados más jóvenes como los de mayor edad, muestran niveles de productividad en promedio inferiores al grupo central, aunque esta diferencia debe interpretarse con cautela dado que la edad no fue incluida como predictor formal del modelo, debido a la posible multicolinealidad con otras variables.

Nivel de formación académica: Según los valores obtenidos el nivel académico ronda el magíster como umbral óptimo.

El nivel de formación es uno de los predictores de perfil individual con mayor respaldo en el modelo, con un coeficiente Ridge de +0,038 y una diferencia de +63% en el top 10 respecto al promedio general. Traducido a la escala construida para esta variable, que va de 0 a 1 y está dividida en tres tramos, el diputado modelo presenta un puntaje situado entre 0,40 y 0,59, correspondiente al segundo tramo analizado. Este tramo agrupa a quienes poseen algún grado de magíster con algún título universitario o técnico o algún diplomado lo que indica un nivel de formación académica avanzada pero no siendo el más alto posible. Los diputados con puntajes en el tercer tramo, vale decir, aquellos con el nivel máximo de aprendizaje, o sea un doctorado, también presentan productividad elevada, pero el segundo tramo es el que concentra la mayor densidad de diputados del top, lo que sugiere que la formación de magíster es el umbral a partir del cual la formación académica comienza a aportar una capacidad predictiva positiva sobre la productividad.

La interpretación sustantiva de este resultado es que una formación avanzada facilita la redacción de proyectos normativos técnicamente más viables, la comprensión de materias legislativas complejas y la capacidad de argumentar con solidez en sesiones de comisión, donde se negocia el contenido específico de las leyes. No implica que un candidato sin magíster no pueda ser productivo, pero la evidencia sugiere que la formación académica es un activo relevante en el contexto legislativo chileno.

Experiencia legislativa previa: No se llega a apreciar una capacidad predictiva cuando se descarta.

El coeficiente de experiencia legislativa en el modelo Ridge es de -0,001, un valor prácticamente nulo que indica que haber sido diputado anteriormente no garantiza mayor productividad en el período analizado. El análisis descriptivo del top muestra una media de 6,45 años de experiencia, levemente superior al promedio general de 6,01 años, pero la diferencia es mínima y no estadísticamente relevante. Este resultado desafía la intuición

de que la trayectoria parlamentaria previa es una condición necesaria para la productividad, y sugiere que el conocimiento acumulado en períodos anteriores no se transfiere automáticamente en términos de resultados legislativos, posiblemente porque cada período legislativo tiene una agenda, una correlación de fuerzas y una dinámica política propias que requieren adaptación independientemente de la experiencia previa. Un candidato nuevo puede, por tanto, rendir igual que uno veterano si reúne las demás características del perfil descrito.

Experiencia laboral privada: Baja o moderada, con sesgo hacia el sector público

El coeficiente de experiencia laboral en el modelo es de -0,006, negativo y pequeño, lo que indica una asociación levemente inversa entre años de trayectoria en el sector privado y la productividad legislativa. El análisis del top 10 confirma este patrón: los diputados más productivos tienen en promedio 10,4 años de experiencia privada frente a los 11,95 años del promedio general, una diferencia de -13% que se amplía a -21% en el top 20. El diputado modelo tiene, por tanto, una trayectoria más orientada al servicio público, la academia o la política desde temprano, que a una carrera extensa en empresa privada. Una explicación plausible es que la experiencia en el sector público desarrolla habilidades más directamente aplicables al trabajo legislativo, como la comprensión del funcionamiento institucional del Estado, la gestión de normas y regulaciones, y la negociación en entornos con múltiples actores con intereses distintos. Esto no significa que la experiencia privada sea un obstáculo, sino que no constituye una ventaja comparativa relevante para la productividad legislativa específica, esto dado que el promedio de años trabajados en el sector privado es de casi 12.

Respaldo electoral: Superior al promedio del distrito, sin llegar a ser el candidato estrella.

El coeficiente de nivel de votación en el modelo es de +0,017, positivo y moderado, lo que indica que un mayor respaldo electoral relativo se asocia con mayores niveles de productividad, aunque la magnitud de esta asociación es inferior a la observada para variables como los proyectos de ley presentados o la participación en comisiones. El análisis del top muestra que los diputados más productivos tienen un nivel de votación un 7% superior al promedio del conjunto en el top 10, diferencia que aumenta al 17% en el top 15, lo que sugiere que el respaldo electoral tiene mayor peso diferenciador en grupos de productividad media-alta que en el núcleo de los más productivos. El diputado modelo no necesita ser el candidato más votado de su lista ni de su distrito, pero sí debe contar con

una base electoral suficiente como para tener capital político que le permita negociar iniciativas con otros bloques y construir las mayorías necesarias para aprobar proyectos. Un respaldo electoral muy bajo, en cambio, puede limitar la capacidad de influencia interparlamentaria que es clave para hacer avanzar la agenda legislativa propia.

Coalición política: Centro o izquierda tradicional en el período analizado

Dentro del grupo de variables políticas, la coalición presenta la asociación de mayor magnitud con el índice de productividad, solo por detrás de las variables de actividad parlamentaria. Tomando como referencia la coalición de centro, todos los bloques de izquierda y derecha muestran coeficientes negativos o levemente positivos en el modelo. El bloque de derecha radical presenta el coeficiente más negativo con -0,076, seguido de la derecha tradicional con -0,068, mientras que la izquierda progresista tiene -0,036 y la izquierda tradicional muestra el coeficiente más próximo al centro con +0,013. Esto significa que, con características individuales idénticas en todas las demás variables, un candidato de centro tiene la mayor productividad esperada del período, seguido de la izquierda tradicional, mientras que un candidato de derecha radical tendría la menor productividad esperada.

El análisis del top refuerza este patrón: el centro, representando solo el 12,1% del congreso, concentra 3 diputados en el top 10, 5 en el top 15 y 5 en el top 20, siendo el bloque más sobrerrepresentado en proporción a su tamaño. La derecha radical, con 15,3% de representación, está completamente ausente en los tres grupos.

Es fundamental, sin embargo, destacar que este resultado es específico del período 2022-2026, que corresponde a un gobierno de izquierda. Una hipótesis razonable es que la productividad de un bloque político está parcialmente condicionada por su posición respecto al oficialismo en cada período, dado que los bloques más cercanos al gobierno tienen mayor facilidad para hacer avanzar sus iniciativas en la agenda legislativa. Esta hipótesis requeriría ser contrastada con datos de otros períodos con distintas composiciones de gobierno para asociar si el patrón es estructural o coyuntural. Por lo tanto, la recomendación de priorizar candidatos de centro o izquierda tradicional como criterio de selección debe interpretarse con cautela y en el contexto político específico de cada elección.

Por lo tanto, considerando las diversas variables analizadas el perfil del candidato a diputado con mayor productividad legislativa esperada según la evidencia de esta investigación es el siguiente. Se trata de una persona de cualquier género, de entre 45 y 50

años aproximadamente, con formación académica de nivel magíster, con una trayectoria orientada principalmente al servicio público, pero habiendo obtenido experiencia laboral dentro del mundo privado, con un respaldo electoral superior al promedio de su distrito, aunque no necesariamente el candidato más votado de su lista, y con afiliación a una coalición de centro o izquierda tradicional en el contexto de un período de gobierno de izquierda.

Es fundamental, sin embargo, interpretar este perfil con la perspectiva adecuada respecto a su poder predictivo real. Las características individuales descritas explican solo una fracción de la productividad legislativa observada, dado que el modelo Ridge, que es el mejor estimador disponible en este estudio, explica en promedio el 24,7% de la variabilidad del índice de productividad. El 75,3% restante depende de factores no observables mediante registros públicos, siendo los que en mayor medida ayudan a explicar estadísticamente la iniciativa legislativa activa medida por la cantidad de proyectos de ley que el diputado efectivamente presenta, el trabajo técnico sostenido en comisiones, y la capacidad de negociación informal que no queda registrada en ninguna fuente de datos institucional.

Esto implica que las características del perfil individual son condiciones facilitadoras, no garantías de productividad. Un candidato con el perfil óptimo tiene una ventaja de partida respecto a uno sin ese perfil, pero su desempeño real dependerá fundamentalmente de las decisiones que tome dentro del congreso: cuántos proyectos presenta, en cuántas comisiones trabaja activamente, y cuán eficaz es en construir las alianzas políticas necesarias para convertir sus iniciativas en leyes aprobadas. La selección de candidatos basada en el perfil descrito es, por tanto, una condición necesaria pero no suficiente para asegurar productividad legislativa.

6 Conclusiones

Esta investigación buscó responder una pregunta concreta: ¿qué características individuales de un diputado se asocian con una mayor productividad legislativa en la Cámara de Diputados de Chile durante el período 2022-2026? Para abordarla, se construyó un índice de productividad original que combina la cantidad de leyes aprobadas, la complejidad normativa de cada ley y el rol de autoría del diputado, normalizado por el tiempo efectivo en el cargo. Sobre este índice se entrenaron y compararon ocho algoritmos de modelado, y se interpretaron sus resultados a través de múltiples métodos de análisis de importancia de variables. El conjunto de hallazgos obtenidos permite formular conclusiones sustantivas sobre la naturaleza de la productividad legislativa y las variables que contribuyen a explicarla estadísticamente.

El primer hallazgo, y quizás el más importante para encuadrar todos los demás, es que el modelo que mejor explica la productividad legislativa es Ridge, un modelo lineal regularizado, con un R^2 promedio de 0,250 en el análisis de robustez con 30 semillas aleatorias y un R^2 de 0,164 en el conjunto de prueba. La superioridad de Ridge sobre todos los algoritmos no lineales evaluados, incluyendo Random Forest (0,190), Gradient Boosting (0,041), XGBoost (-0,047) y LightGBM (0,088), no es un resultado trivial. Indica que la relación estadística entre las características individuales de los diputados y su productividad es fundamentalmente lineal y aditiva: cada variable contribuye de forma independiente y proporcional al resultado, sin interacciones complejas ni umbrales no lineales. Esto simplifica la interpretación de los resultados y los hace más robustos desde el punto de vista metodológico.

El segundo hallazgo es el ordenamiento de los predictores. La variable con mayor capacidad predictiva sobre el índice de productividad es la cantidad de proyectos de ley presentados, con un coeficiente estandarizado de +0,161 y una importancia de Random Forest de 0,375, siendo casi tres veces más relevante estadísticamente que la segunda variable. Este resultado establece que la iniciativa legislativa activa presenta la asociación estadística más fuerte con la productividad en el conjunto de variables analizado. Un diputado que presenta muchos proyectos muestra sistemáticamente un índice de productividad más alto, independientemente de su coalición, su formación o su experiencia. En segundo lugar, aparece la intervención en comisiones (0,137 en RF, coeficiente +0,028), lo que confirma que el trabajo técnico en los espacios donde se

negocia el contenido específico de las leyes se asocia más fuertemente con la productividad que la visibilidad pública en el pleno, donde la intervención en sala presenta una capacidad predictiva menor, con solo 0,088 de importancia. Esta jerarquía, comisión por encima de sala, es uno de los hallazgos más relevantes para la comprensión del trabajo parlamentario real en Chile.

El tercer hallazgo concierne las variables de perfil individual. Nivel de formación es el predictor de características personales con mayor respaldo en el modelo, con un coeficiente de +0,038 y una importancia RF de 0,047, confirmando que una formación académica avanzada, en el rango de magíster o doble titulación, se asocia positivamente con la productividad. El nivel de votación presenta un coeficiente de +0,017, indicando que el respaldo electoral se asocia de forma positiva y moderada con la productividad, lo que puede interpretarse como reflejo del capital político disponible para negociar, aunque sin que esta relación permita establecer causalidad. La experiencia laboral privada muestra un coeficiente levemente negativo de -0,006, lo que sugiere que trayectorias más orientadas al servicio público se asocian estadísticamente con valores más altos de productividad que las carreras extensas en el sector privado. El género presenta un coeficiente de prácticamente cero (+0,000), confirmando que el género no presenta capacidad predictiva significativa sobre la productividad legislativa cuando se controlan las demás variables del modelo.

El cuarto hallazgo es la asociación de la coalición política con la productividad. Tomando como referencia la coalición de centro, todos los bloques muestran coeficientes que difieren en magnitud y signo. La izquierda tradicional es el único bloque con coeficiente positivo (+0,013), mientras que la izquierda progresista presenta -0,036, la derecha tradicional -0,068 y la derecha radical -0,076. Este patrón tiene una explicación contextual plausible: el período 2022-2026 corresponde a un gobierno de izquierda, por lo que los bloques de derecha, en posición de oposición durante todo el período, muestran estadísticamente valores de productividad más bajos. Sin embargo, el hecho de que la izquierda progresista, siendo el bloque de gobierno, presente un índice promedio inferior al centro y a la izquierda tradicional sugiere que la experiencia institucional acumulada y la capacidad de negociación transversal se asocian más fuertemente con la productividad que la sola alineación ideológica con el ejecutivo. Este hallazgo debe interpretarse con cautela: es probable que sea específico del período analizado y no generalizable a otros contextos políticos.

El quinto hallazgo es que la puntualidad presenta una asociación negativa con la productividad, con un coeficiente de $-0,019$, resultado que se confirma en el análisis descriptivo donde los diputados del top 10 muestran en promedio un 17% menos de puntualidad que el promedio general. La interpretación estadísticamente más consistente con el conjunto de variables del modelo es que los diputados con mayor carga de trabajo legislativo presentan naturalmente menor puntualidad formal en sesiones específicas. La relación observada no permite establecer que la impuntualidad contribuya causalmente a la productividad, sino únicamente que ambas variables se asocian en la dirección descrita una vez controladas las demás variables del modelo. Esta conclusión es relevante porque cuestiona el uso de la puntualidad como indicador de desempeño parlamentario.

El sexto hallazgo, quizás el más importante para la política pública, es que el modelo explica solo el 25,0% de la variabilidad en la productividad legislativa de forma robusta. El 75% restante corresponde a factores no capturados por los datos públicos institucionales disponibles: la negociación política informal, las redes de influencia interparlamentaria, el capital social acumulado, la coyuntura política específica de cada período y la capacidad de articular coaliciones ad hoc para proyectos concretos. Esta brecha entre lo observable y lo explicado estadísticamente tiene una implicancia directa: las características del perfil individual son variables con capacidad predictiva, pero no determinantes en el sentido causal del término. El candidato con el perfil estadísticamente asociado a mayor productividad presenta una ventaja de partida observable en los datos, pero su desempeño real dependerá de factores que este modelo no puede anticipar.

Esta investigación no fue diseñada para establecer relaciones de causalidad entre las variables analizadas y la productividad legislativa, sino para identificar patrones de asociación que permitan comprender qué características presentan una mayor relación estadística con el desempeño parlamentario durante el período 2022-2026. Esta distinción resulta especialmente relevante porque varias de las variables explicativas corresponden a actividades desarrolladas por los propios diputados durante el mismo período en que se mide la productividad, como la presentación de proyectos de ley, las intervenciones en comisiones o las intervenciones en sala. En consecuencia, la importancia estadística de estas variables debe interpretarse como evidencia de que coexisten con mayores niveles de productividad dentro del período analizado y no como demostración de que constituyen su causa. Los modelos desarrollados permiten identificar qué factores presentan una mayor capacidad para explicar y predecir la variabilidad observada en la productividad legislativa,

pero no establecer relaciones causales independientes entre dichas variables. Por ello, los resultados deben entenderse como una caracterización empírica de la productividad legislativa chilena durante las legislaturas estudiadas y como un punto de partida para futuras investigaciones que, mediante metodologías específicamente orientadas al análisis causal, profundicen en los mecanismos que explican estas asociaciones.

En suma, la productividad legislativa en Chile durante el período 2022-2026 se asocia estadísticamente, en primer lugar, con el comportamiento activo del diputado dentro del congreso (iniciativa legislativa y trabajo en comisiones), en segundo lugar, con el posicionamiento político (coalición y capital electoral) y, en tercer lugar, con características del perfil previo (formación académica y orientación de carrera). Variables como el género, la experiencia previa y la puntualidad muestran una asociación marginal o estadísticamente no significativa con la productividad en el contexto de este modelo. Esta arquitectura de asociaciones estadísticas constituye una aproximación empírica al estudio de la productividad legislativa en Chile y aporta evidencia útil para comprender los factores que se relacionan con ella, así como una base metodológica para futuras investigaciones orientadas a profundizar en los mecanismos que explican el desempeño parlamentario.

7 Recomendaciones

A partir de los hallazgos de esta investigación, se formulan las siguientes recomendaciones dirigidas a distintos actores del sistema político y académico chileno.

Primera recomendación: rediseñar los indicadores públicos de evaluación parlamentaria. La asistencia formal, que promedia un 97% con mínima varianza entre diputados, y la puntualidad, que muestra una asociación negativa con la productividad en el modelo estimado, no deberían utilizarse como indicadores centrales de desempeño legislativo. Se recomienda avanzar hacia sistemas de evaluación que incorporen la cantidad de proyectos de ley presentados, el número de intervenciones en comisiones y la cantidad de leyes aprobadas ponderadas por complejidad normativa, que son precisamente las variables con mayor capacidad predictiva sobre la productividad según los resultados de esta investigación. Estos indicadores capturan la actividad legislativa real con mayor precisión estadística que la asistencia formal.

Segunda recomendación: diseñar un sistema de KPIs parlamentarios basado en la evidencia empírica. Los resultados de esta investigación identifican con precisión las variables con mayor capacidad predictiva sobre la productividad, lo que proporciona una base estadística sólida para construir un sistema de indicadores clave de desempeño para los diputados. Se recomienda que la Cámara de Diputados, en colaboración con la Biblioteca del Congreso Nacional y la DIPRES, desarrolle un sistema de KPIs que incluya al menos tres dimensiones: iniciativa legislativa medida por proyectos de ley presentados y su porcentaje de avance en tramitación, actividad en comisiones medida por intervenciones efectivas y asistencia a sesiones técnicas y productividad normativa medida por leyes aprobadas ponderadas por su complejidad. Este sistema reemplazaría el uso de la asistencia formal como indicador principal, dado que los datos muestran que dicha variable no presenta capacidad discriminante entre diputados con distintos niveles de productividad.

Tercera recomendación: implementar un Balanced Scorecard parlamentario como marco integrador de evaluación. Un sistema de KPIs aislados puede generar incentivos no previstos si las dimensiones no están articuladas entre sí. Se recomienda estructurar los indicadores dentro de un Balanced Scorecard adaptado al contexto legislativo, organizando el desempeño del diputado en cuatro perspectivas equilibradas: productividad normativa

(leyes aprobadas y complejidad), actividad parlamentaria (comisiones e iniciativa legislativa), representación ciudadana (vínculo con electores y capital electoral) y desarrollo institucional (formación académica y capacitación). Este marco permite evaluar el desempeño desde una perspectiva integral, superando la reducción del trabajo parlamentario a una sola métrica.

Cuarta recomendación: considerar la formación académica como criterio en los procesos de selección de candidatos. Los resultados muestran que el nivel de formación se asocia positivamente con la productividad legislativa, aunque esta relación no puede interpretarse como causal a partir del diseño metodológico de esta investigación. Se recomienda que los partidos políticos incorporen criterios de preparación técnica en la evaluación de candidatos, no como requisito excluyente sino como factor que el análisis estadístico sugiere relacionado con el desempeño. El umbral estadísticamente asociado a mayor productividad corresponde a formación de magíster o doble titulación universitaria.

Quinta recomendación: replicar el estudio en períodos distintos para evaluar la estabilidad de las asociaciones identificadas. La asociación entre coalición política y productividad identificada en esta investigación es específica del período 2022-2026. Para determinar si los patrones observados son robustos o propios de este contexto político particular, se recomienda replicar el análisis en períodos con distintas composiciones de gobierno. Si el centro resulta estadísticamente más asociado a mayor productividad independientemente del período, ello apuntaría a una ventaja estructural de los partidos de centro vinculada a su capacidad de negociación transversal. Si el patrón cambia con el gobierno, confirmaría que la posición relativa respecto al oficialismo es el factor estadísticamente relevante, más que la identidad de la coalición en sí misma.

Sexta recomendación: mejorar la transparencia y estructuración de los datos parlamentarios. Actualmente, la información sobre actividad en comisiones, autoría de proyectos de ley y seguimiento de mociones existe, pero no está consolidada ni es fácilmente accesible para investigadores o ciudadanos. Se recomienda que la Cámara de Diputados avance hacia un sistema de datos abiertos que publique de forma estructurada y actualizada estos registros, lo que permitiría tanto investigaciones más robustas como una ciudadanía más informada respecto al desempeño de sus representantes.

Séptima recomendación: incorporar variables adicionales en investigaciones futuras. El 75% de la variabilidad no explicada estadísticamente por el modelo apunta a factores que no están disponibles en registros públicos. Se recomienda que futuras investigaciones

exploren la incorporación de redes de copatrocinio entre diputados, análisis de discurso en intervenciones parlamentarias, medidas de alineación temática con la agenda del ejecutivo y métricas de centralidad en redes políticas derivadas de votaciones nominales. Estas variables podrían aumentar significativamente la capacidad explicativa del modelo y ofrecer una comprensión más completa de los patrones estadísticos que se asocian con la productividad legislativa, sin que ello implique por sí solo establecer relaciones causales.

8 Referencias

- Banco Santander. (2025). Productividad. Glosario Banco Santander. <https://www.bancosantander.es/glosario/productividad>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cámara de Diputadas y Diputados de Chile. (2023). Reglamento de la Cámara de Diputadas y Diputados. Biblioteca del Congreso Nacional de Chile. <https://www.bcn.cl/leychile/navegar?idNorma=82415>
- Centro de Estudios Públicos. (2025). Encuesta Nacional CEP N.º 93, abril-mayo 2025. https://static.cepchile.cl/uploads/cepchile/2025/05/07-144852_hgoj_CEP93_VF_anexos.pdf
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., y Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc. <https://ceeivalencia.emprenemjunts.es/?op=8&n=27469>
- Chen, T., y Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- Dirección de Presupuestos (DIPRES). (2019). Sistema de Evaluación y Control de Gestión: Informe de Evaluaciones 2019. Ministerio de Hacienda, Gobierno de Chile. <https://www.dipres.gob.cl/597/w3-propertyvalue-23076.html>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Gailmard, S. (2012). Accountability and principal-agent models. University of California, Berkeley. [https://www.law.berkeley.edu/archive/files/csls/Gailmard_-_Accountability_and_Principal-Agent_Models\(2\).pdf](https://www.law.berkeley.edu/archive/files/csls/Gailmard_-_Accountability_and_Principal-Agent_Models(2).pdf)
- Hastie, T., Tibshirani, R., y Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2.a ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19. <https://doi.org/10.1111/j.1467-9299.1991.tb00779.x>
- Inter-Parliamentary Union. (2022). Parline database on national parliaments. <https://data.ipu.org>
- Irrarrazaval, I. (2023). Confianza institucional en Chile: un desafío para el desarrollo. *Puntos de*

- Referencia, 682. Centro de Estudios Públicos.
https://static.cepchile.cl/uploads/cepchile/2023/12/pder682_-irarrazaval.pdf
- James, G., Witten, D., Hastie, T., y Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2.a ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Kellermann, M. (2012). Legislative productivity of the U.S. Congress, 1789–2004. *Legislative Studies Quarterly*, 37(1), 105–130.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., y Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154. <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>
- Krugman, P. (1990). The age of diminished expectations: U.S. economic policy in the 1980s. MIT Press.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., y Li, W. (2005). Applied linear statistical models (5.a ed.). McGraw-Hill.
- Laffont, J.-J., y Tirole, J. (1993). A theory of incentives in procurement and regulation. MIT Press.
- Latinobarómetro. (2023). Informe 2023. <https://www.latinobarometro.org>
- Lundberg, S. M., y Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- Ministerio Secretaría General de la Presidencia. (2020, 20 de mayo). Ley N° 21.233: Modifica la Constitución Política para reducir las remuneraciones parlamentarias y de otras autoridades del Estado. Diario Oficial de la República de Chile. <https://www.bcn.cl/leychile/navegar?idNorma=1145781>
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2001). Measuring productivity: Measurement of aggregate and industry-level productivity growth. OECD Publishing. <https://doi.org/10.1787/9789264194519-en>
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2017). Government at a glance 2017. https://www.oecd.org/content/dam/oecd/es/publications/reports/2017/07/government-at-a-glance-2017_g1g74dbf/9789264304543-es.pdf
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2024). OECD survey on

drivers of trust in public institutions - 2024 results: Country notes - Chile.
https://www.oecd.org/es/publications/2024/06/oecd-survey-on-drivers-of-trust-in-public-institutions-2024-results-country-notes_33192204/chile_c3d29801.html

Parlamento Europeo. (1997). Checking the attendance of members in the national parliaments. European Centre for Parliamentary Research and Documentation.
[https://www.europarl.europa.eu/RegData/etudes/etudes/join/1997/166944/IPOL-AFCO_ET\(1997\)166944_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/etudes/join/1997/166944/IPOL-AFCO_ET(1997)166944_EN.pdf)

Parmenter, D. (2015). Key performance indicators: Developing, implementing, and using winning KPIs (3.a ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119019855>

Pollitt, C., y Bouckaert, G. (2004). Public management reform: A comparative analysis (2.a ed.). Oxford University Press.

Senado de la República de Chile. (2024). Reglamento del Senado. Biblioteca del Congreso Nacional de Chile. <https://www.bcn.cl/leychile/navegar?idNorma=82539>

Schedler, A. (1999). Conceptualizing accountability. En A. Schedler, L. Diamond y M. F. Plattner (Eds.), *The self-restraining state: Power and accountability in new democracies* (pp. 13–28). Lynne Rienner Publishers.

Volden, C., y Wiseman, A. E. (2014). *Legislative effectiveness in the United States Congress: The lawmakers*. Cambridge University Press.

Wilson, W. (1887). The study of administration. *Political Science Quarterly*, 2(2), 197–222. <https://doi.org/10.2307/2139277>

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6.a ed.). Cengage Learning.