

ANEXO

N°45

INFORME

FINAL



Serie Proyectos de Investigación e Innovación
Superintendencia de Seguridad Social
Santiago - Chile

Análisis de factores determinantes, adicionales al diagnóstico, que permitan la detección temprana de casos de larga estadía mediante un modelo econométrico de predicción (MUSEG).

INFORME FINAL

Autor:
Sandra Alvear Vega
Héctor Vargas Garrido
Hugo Núñez Delafuente
Cristian Vidal Silva

Año publicación

2025



SUPERINTENDENCIA DE SEGURIDAD SOCIAL

SUPERINTENDENCE OF SOCIAL SECURITY

La serie Proyectos de Investigación e Innovación corresponde a una línea de publicaciones de la Superintendencia de Seguridad Social, que tiene por objetivo divulgar los trabajos de investigación e innovación en Prevención de Accidentes y Enfermedades del Trabajo financiados por los recursos del Seguro Social de la Ley 16.744.

Los trabajos aquí publicados son los informes finales y están disponibles para su conocimiento y uso. Los contenidos, análisis y conclusiones expresados son de exclusiva responsabilidad de su(s) autor(es), y no reflejan necesariamente la opinión de la Superintendencia de Seguridad Social.

Si requiere de mayor información, sobre el estudio o proyecto escriba a: investigaciones@suseso.cl.

Si desea conocer otras publicaciones, artículos de investigación y proyectos de la Superintendencia de Seguridad Social, visite nuestro sitio web: www.suseso.cl.

The Research and Innovation Projects series corresponds to a line of publications of the Superintendence of Social Security, which aims to disseminate the research and innovation work in the Prevention of Occupational Accidents and Illnesses financed by the resources of Law Insurance 16,744.

The papers published here are the final reports and are available for your knowledge and use. The content, analysis and conclusions are solely the responsibility of the author (s), and do not necessarily reflect the opinion of the Superintendence of Social Security.

For further information, please write to: investigaciones@suseso.cl.

For other publications, research papers and projects of the Superintendence of Social Security, please visit our website: www.suseso.cl.

Superintendencia de Seguridad
SocialHuérfanos 1376
Santiago,
Chile.



Análisis de factores determinantes, adicionales al diagnóstico, que permitan la detección temprana de casos de larga estadía mediante un modelo econométrico de predicción (MUSEG).

PREPARADO POR:
Sandra Alvear Vega
Héctor Vargas Garrido
Hugo Núñez Delafuente
Cristian Vidal Silva

Julio 2025

Índice

1. Introducción y Antecedentes	1
1.1. Accidentes en el trabajo	1
1.2. Accidentes en el trayecto	2
1.3. Enfermedades profesionales	3
2. Definición del problema, pregunta de investigación o desafío de innovación	4
2.1. Definición del problema	4
2.2. Pregunta de Investigación	5
2.3. Objetivos	5
2.3.1. Objetivo General	5
2.3.2. Objetivos Específicos	5
3. Revisión de la literatura o experiencias relevantes	6
3.1. Variables Personales y de la Salud del Trabajador	6
3.1.1. La edad del trabajador accidentado	6
3.1.2. Las condiciones de salud preexistentes en el trabajador accidentado	7
3.1.3. Los trabajadores hombres accidentados	7
3.1.4. La condición migrante del trabajador	8
3.1.5. El tipo de lesión generada por el accidente	8
3.1.6. La condición del accidente	8
3.2. Variables Situaciones y Temporales del Accidente	8
3.2.1. Día de la semana del accidente: hacia el fin de semana	8
3.2.2. Jornadas y hora del accidente: atardecer, anochecer y madrugadas	9
3.2.3. Accidentes de trayecto	9
3.2.4. Año de ocurrencia del accidente	10
3.2.5. Tiempos transcurridos en cualquier etapa de atención clínica o administrativa relacionada con el accidente	10
3.3. Variables de la Industria y la ubicación geográfica	10
3.3.1. Empresas de menor tamaño	10
3.3.2. Sectores con mayores proporciones de accidentes severos	10
3.3.3. Regiones o provincias donde ocurre el accidente	11
3.4. Sobre la estimación de tiempos de larga estadía o recuperación	11
4. Descripción de la metodología o etapas de la innovación	11
4.1. Metodología Machine Learning	12
4.1.1. Comprensión del negocio y los datos.	13
4.1.2. Preparación de los datos.	13
4.1.3. Ingeniería del Modelo.	13
4.1.4. Evaluación del Modelo	15
4.1.5. Interpretabilidad	17
4.2. Enfoque Econométrico	17
4.2.1. Variable dependiente	17
4.2.2. Variables independientes	17

4.2.3.	Análisis de los datos	17
4.2.3.1.	Accidentes del Trabajo	17
4.2.3.2.	Accidentes de Trayecto	18
4.2.3.3.	Enfermedad Profesional	18
5.	Resultados	19
5.1.	Enfoque Machine Learning	19
5.1.1.	Comprensión del negocio y los datos.	19
5.1.2.	Preparación de los datos.	23
5.1.2.1.	Limpieza del conjunto de datos.	23
5.1.2.2.	Ingeniería de características.	24
5.1.2.3.	Variable objetivo	27
5.1.3.	Ingeniería de Modelos	30
5.1.4.	Evaluación de Modelos	31
5.1.5.	Interpretabilidad	34
5.1.5.1.	Accidentes del Trabajo	34
5.1.5.2.	Accidentes del Trayecto	36
5.2.	Enfoque Econométrico	38
5.2.1.	Evaluación de Modelos	38
5.2.2.	Accidentes del Trabajo	39
5.2.2.1.	Estadísticos del modelo accidentes del trabajo.	39
5.2.2.2.	Interpretabilidad	39
5.2.3.	Accidentes del Trayecto	41
5.2.3.1.	Estadísticos del modelo accidentes del trayecto	41
5.2.3.2.	Interpretabilidad del modelo accidentes del trayecto	42
6.	DISCUSIÓN DE RESULTADOS	44
6.1.	Discusión de los resultados para los Accidentes del Trabajo	44
6.2.	Discusión de los resultados para los Accidentes del Trayecto	46
7.	Recomendaciones para Sistema de Seguridad y Salud en el Trabajo	46
8.	Conclusiones	47
9.	Referencias	49

Índice de figuras

1.	Diagrama de Metodología utilizada en estudio.	12
2.	Distribución de variables numéricas: edad, porcentaje de incapacidad y días de reposo.	22
3.	Días de reposo por categoría de variable CIE y tipo de accidente.	27
4.	Distribución de variable objetivo por tipo de accidente.	29
5.	Cross Validation en conjunto de test para accidentes de trabajo.	32
6.	Cross Validation en conjunto de test para accidentes de trayecto.	33
7.	Contribución de las 10 variables más importantes en la predicción del modelo para clasificar pacientes de larga estadía en accidentes de trabajo.	35
8.	Interpretabilidad local del modelo en accidentes de trabajo.	36
9.	Contribución de las 10 variables más importantes en la predicción del modelo para clasificar pacientes de larga estadía en accidentes de trayecto.	37
10.	Interpretabilidad local del modelo en accidentes de trayecto.	38

Índice de tablas

1.	Espacio de búsqueda de hiperparámetros por modelo. Los rangos del tipo $[a, b]$ indican un intervalo cerrado, es decir, se consideran todos los valores entre a y b , incluyendo los extremos.	16
2.	Descripción de variables del conjunto de datos	20
3.	Estadísticas descriptivas de variables numéricas	22
4.	Resumen de variables categóricas	23
5.	Categorías de variable CIE	24
6.	Categorías de variable CIUO. Cada categoría representa uno de los grandes grupos de la Clasificación Internacional Uniforme de Ocupaciones.	25
7.	Categorías de la variable <i>cod_forma_oit</i>	25
8.	Categorías de la variable <i>cod_agente_oit</i>	26
9.	Listado de variables del conjunto de datos	26
10.	Días promedio de reposo por categoría CIE para accidentes de trabajo y trayecto. .	28
11.	Categorías CIE cuyo número de casos de accidentes de trabajo no alcanza el umbral mínimo establecido.	29
12.	Categorías CIE cuyo número de casos de accidentes de trayecto no alcanza el umbral mínimo establecido.)	29
13.	Hiperparámetros óptimos determinados para cada modelo según tipo de accidente .	31
14.	Evaluación de calidad de los modelos según el criterio Akaike (AIC)	38
15.	Estadísticos de bondad de ajuste para el modelo de ausencia prolongada (Accidente del Trabajo)	39
16.	Efectos Marginales Promedio (EMP) de la Ausencia Prolongada en los Accidentes del Trabajo (70 % de los datos, Entrenamiento)	40
17.	Efectos Marginales Promedio (EMP) de la Ausencia Prolongada en los Accidentes del Trabajo (70 % de los datos, Entrenamiento)	41
18.	Estadísticos de bondad de ajuste para el modelo de ausencia prolongada (Accidente del Trayecto)	42
19.	Efectos Marginales Promedio (EMP) de la Ausencia Prolongada en los Accidentes del Trayecto (70 % de los datos, Entrenamiento)	43
20.	Efectos Marginales Promedio (EMP) de la Ausencia Prolongada en los Accidentes del Trayecto (30 % de los datos, Validación)	44

Resumen Ejecutivo

Al estar lejos del trabajo, debido a un accidente o enfermedad ocupacional, el trabajador puede deprimirse, sufrir consecuencias económicas y sociales, entre otras, todo lo cual impacta adversamente en su salud, calidad y esperanza de vida. Este proyecto tiene como objetivo determinar los factores adicionales al diagnóstico que permitan predecir anticipadamente si un paciente tendrá ausencia prolongada, considerando la información asociada al otorgamiento de reposo a pacientes de la Mutual de Seguridad, clasificados en tres grupos; accidentes del trabajo, del trayecto y enfermedades profesionales. Para ello se definió de un modelo Machine Learning (ML) + Explicación de Inteligencia Artificial (XAI), complementado con un enfoque econométrico. Para la selección del modelo ML, se emplea la técnica de benchmarking de modelos, el rendimiento de los algoritmos se evaluó en función de métricas como accuracy, precisión, recall y F1 score y para interpretar el modelo que presenta el mejor desempeño se usó Shapley Additive Explanations (SHAP). Desde el punto de vista econométrico se aplicó modelos logit y probit, cuyo rendimiento se evaluó según el método de máxima verosimilitud y el estadístico AIC (Akaike). Ambas metodologías aplicadas infieren que tanto para los accidentes del trabajo como del trayecto, gravedad, presencia de diagnósticos múltiples y porcentaje de incapacidad son variables propias del diagnóstico que influyen en la ausencia prolongada: En relación, con las variables distintas al diagnóstico destacan en (1) Accidentes del trabajo: sexo masculino, edad avanzada, accidente por caída y desempeño laboral en las PYMES (2) Accidentes del trayecto: edad avanzadas y desempeño en las PYMES, sexo masculino, seguridad vial y accidente ocurrido en zonas urbanas y para (3) enfermedades profesionales: los modelos no permiten realizar estimaciones estadísticamente significativas.

Palabras claves:

Ausencia Prolongada; Factores adicionales al diagnóstico; Accidentes del trabajo, Accidentes del trayecto, Enfermedades profesionales; Machine Learning; Modelos econométricos.

1. Introducción y Antecedentes

La Ley Chilena Nro. 16.744 distingue dos tipos de accidentes laborales: 1: En el lugar de trabajo y 2: En el trayecto. Adicionalmente, se consideran enfermedades profesionales. Según esta ley un accidente de trabajo es «toda lesión física sufrida por una persona causada o como consecuencia del trabajo». El accidente de trayecto es «el que se produce en el trayecto directo entre el lugar de trabajo y el lugar de residencia, y el que se produce en el trayecto entre dos lugares de trabajo, incluso de empleadores diferentes» (art.5°). Por su parte, una enfermedad profesional «es causada de una manera directa por el ejercicio de la profesión o el trabajo que realice una persona y que le produzca incapacidad o muerte» (art.7°). Los dos tipos de accidentes laborales y las enfermedades profesionales presentan notables diferencias. Según SUSESO, 2025, el año 2024 los accidentes del trabajo y del trayecto, con días perdidos, alcanzaron 143.199 (72 %) y 56.675 (28 %) casos. Las enfermedades profesionales para el mismo año alcanzaron 11.966 caso.

Según Ley N°16.744, el 78 % de la fuerza del trabajo ocupada se encuentran protegidas por el seguro de accidentes del trabajo y enfermedades profesionales, en 2024, equivalente a 7.252.538 persona, representando un incremento de 11 % en los últimos cinco años (SUSESO,2025).

1.1. Accidentes en el trabajo

En los accidentes del trabajo con días perdidos, para el año 2024, los hombres registraron un mayor número de accidentes (62 %), en relación con las mujeres (38 %). Se concentran en los tramos de edad de 25 a 34 años de (26 %), de 35 a 44 años (22 %) y de 45 a 54 años (20 %) (SUSESO,2025). Las actividades económicas que presentan mayor número de accidentes de trabajo son: Comercio reparación de vehículos y otros (14 %), Actividades inmobiliarias, empresariales y de alquiler (12 %), Industrias manufactureras (12 %) y Construcción (11 %) Los accidentes se concentran en empresas categorizadas como grandes, de más de 200 trabajadores (52 %), le siguen las empresas medianas de 50 a 199 trabajadores (21 %), luego las pequeñas empresas de 10 a 49 trabajadores (20 %) y finalmente las microempresas de 01 a 9 trabajadores (7 %) (SUSESO, 2025). El 57 % de los accidentes de trabajo ocurren en la región Metropolitana, le sigue las regiones de Valparaíso y del Biobío con un 9 % y 6 %, respectivamente (SUSESO,2025).

Tanto en hombres como en mujeres, cuando existe una lesión las extremidades son las zonas más afectadas. En los hombres, 42 % de los casos corresponde a miembros superiores (brazos y manos) y 24 % a inferiores (piernas y pies); en las mujeres, 39 % y 31 %, respectivamente. La forma de ocurrencia del accidente es causada mayoritariamente por caída de personas (42 %). Sin embargo, los hombres predominantemente sufren accidentes del trabajo por proyección de partículas (91 %) y atrapamiento o amputaciones (75 %). Por otro lado, en las mujeres dichos accidentes son causados por amenazas y agresiones de personas (52 %) y por tropiezo sin caída (46 %). En la misma línea, el agente material que causa el accidente del trabajo, a nivel global, es herramientas y aparatos (39 %), sin embargo, en los hombres son más representativas las máquinas (77 %) y en las mujeres personas y ambientes de trabajo (54 %) (SUSESO,2025).

El número de días perdidos, para la industria de las mutualidades, por accidentes del trabajo ha disminuido levemente, de 19 a 18,5 días promedio por cada accidente, entre los años 2015 y 2024. En particular, la Mutual de Seguridad presenta, al año 2024, un mayor número de días promedio perdido en comparación con la industria, 19,9 días. No obstante, en los últimos cinco años presenta una disminución de 3 días (SUSESO, 2025). Los accidentes del trabajo más graves los

enfrentan los hombres, mayoritariamente, por amputación traumática, caída de altura de más de 1,8 metros y maniobras de rescate (SUSESO, 2025). Por otro lado, la mayor cantidad de días promedios perdidos por accidente, se presentan en las actividades de explotación de minas y canteras (35,1 días) y transporte, almacenamiento y comunicaciones (23,9 días). Las empresas más pequeñas, de 1 a 9 trabajadores, presentan mayor número días promedio perdidos, 25,4 días, y las empresas más grandes, de 1.000 y más trabajadores, presentan menor número días promedio perdidos, 15,7 días. Los accidentes del trabajo presentan mayor número de días perdidos en las regiones de la Región de Ñuble (21 días promedio) y Región Del Libertador Gral. Bdo. O'Higgins (19,4 días promedio). La tasa de mortalidad por accidentes del trabajo al año 2024 representa 2,1 por 100.000 trabajadores protegidos, con una disminución de 0,6 por 100.000 trabajadores, en los últimos cinco años. La tasa de mortalidad se concentra en las actividades transporte y comunicaciones, minería, agricultura y pesca (SUSESO, 2025).

1.2. Accidentes en el trayecto

En los accidentes de trayecto, las mujeres presentan más accidentes que los hombres, 53 % y 47 % respectivamente, para el año 2024. Se concentran en los tramos de edad de 25 a 34 años de (30 %), de 35 a 44 años (22 %) y de 45 a 54 años (18 %) (SUSESO, 2025). Las actividades económicas que presentan mayor número de accidentes de trayecto son: Actividades inmobiliarias, empresariales y de alquiler (18 %), Comercio, reparación de vehículos y otros (10 %), Construcción (10 %) e industrias manufactureras (9 %). Los accidentes de trayecto se concentran en empresas categorizadas como grandes, de más de 200 trabajadores (64 %), le siguen las empresas medianas de 50 a 199 trabajadores (18 %), luego las pequeñas empresas de 10 a 49 trabajadores (14 %) y finalmente las microempresas de 01 a 9 trabajadores (4 %) (SUSESO, 2025). El 68 % de los accidentes de trayecto ocurren en la región Metropolitana, le sigue las regiones de Valparaíso y del Biobío con 9 % y 4 %, respectivamente (SUSESO, 2025).

En ambos sexos, las lesiones se concentran en las extremidades: en los hombres, un 39 % afecta a los miembros superiores (brazos y manos) y un 31 % a los inferiores (piernas y pies), mientras que en las mujeres las proporciones son 22 % y 44 %, respectivamente. La forma más representativa de ocurrencia de accidente de trayecto es por caída de personas, 41 % en hombres y 59 % en mujeres. En mujeres, las causas de accidente predominantes son: tropiezos sin caída (62 %), atrapamiento o amputaciones (60 %) y sobreesfuerzo y falsos movimientos (59 %). En los hombres, las formas más relevantes son: agresiones por animales (58 %), accidentes que involucran vehículos (56 %), amenazas o agresiones a personas (55 %) (SUSESO, 2025). El número de días perdidos por cada accidente de trayecto ha disminuido levemente entre los años 2015 y 2023, de 25,3 a 23,8 días promedio (SUSESO, 2025). En particular, la Mutual de Seguridad presenta, al año 2024, un mayor número de días promedio perdido en comparación con la industria, 26,2 días. No obstante, en los últimos cinco años presenta una disminución de 3,5 días (SUSESO, 2025). Las actividades de organizaciones y órganos extraterritoriales presentan en promedio 36,6 días perdidos por accidente de trayecto, le sigue explotación de minas y canteras con 32,3 días y agricultura, ganadería, caza y silvicultura con 31,8 días. Las empresas más pequeñas, de 1 a 9 trabajadores, presentan mayor número días promedio perdidos, 34,6 días, y las empresas más grandes, de 1.000 y más trabajadores, presentan menor número días promedio perdidos, 21,2 días. Los accidentes de trayecto presentan mayor número de días perdidos en las regiones de la zona central: Valparaíso, Maule y Del Libertador Gral. Bdo. O'Higgins (SUSESO, 2025).

La tasa de mortalidad por accidentes de trayecto al año 2024 representa 1,8 por 100.000 trabajadores protegidos, con una disminución de 0,9 entre los años 2015 y 2024 (SUSESO, 2025).

1.3. Enfermedades profesionales

En relación con las enfermedades profesionales, éstas son mayoritariamente femeninas (66 %), para el año 2024. Se ha producido una disminución, en los últimos cinco años, en el número de ellas (47 %). En particular, Mutual de Seguridad ha presentado una disminución, para el mismo periodo, superior a la industria (49 %). Las actividades económicas que concentran mayor número de enfermedades profesionales son: Administración pública y defensa; planes de seguridad social (23 %), Servicios sociales y de salud (23 %) y Enseñanza (11 %). Las enfermedades profesionales se concentran en empresas categorizadas como grandes, de más de 200 trabajadores (76 %), le siguen las empresas medianas de 50 a 199 trabajadores (12 %), luego las pequeñas empresas de 10 a 49 trabajadores (9 %) y finalmente las microempresas de 01 a 9 trabajadores (3 %) (SUSESO, 2025). El 49 % de ellas ocurren en la región Metropolitana, le siguen las regiones de Valparaíso y de los Lagos con 8 % y 6 %, respectivamente (SUSESO, 2025). Los diagnósticos más relevantes son: COVID-19 (24 %). Al excluir los casos de COVID-19, la patología de Salud Mental asume el primer lugar, alcanzando el 72 % del total de las enfermedades profesionales a nivel nacional y con 39,3 días promedio perdidos. En particular, este diagnóstico de enfermedad profesional cobra relevancia en las regiones de Tarapacá (82 %), Aysén (81 %) y Arica y Parinacota (80 %) (SUSESO, 2025). Es importante destacar, que los(as) trabajadores(as) que tienen un nivel alto en la escala de estrés de Setterlind, así como también, un nivel alto en salud mental (medido en forma independiente) presentan mayor probabilidad de tener un accidente laboral, manteniendo todo lo demás constante (SUSESO, 2020). En segundo lugar, en importancia se encuentran las patologías musculoesqueléticas (16 %), con 40,8 días promedio perdidos (SUSESO, 2025).

El presente informe responde a la siguiente pregunta de investigación:

¿cuáles son los factores adicionales al diagnóstico y de salud que permitan predecir anticipadamente la ausencia prolongada de un paciente que ha enfrentado (1) un accidente del trabajo, (2) un accidente del trayecto o (3) una enfermedad profesional.

El trabajo se organiza en las siguientes secciones:

1. Definición del problema: esta sección aborda el origen de la pregunta de investigación, objetivo general y objetivos específicos del proyecto.
2. Revisión de la literatura: se aborda la literatura relevante relacionada con variables personales y de salud del trabajador; variables situacionales y temporales del accidente; variables de la industria y situación geográfica y estimación de días de ausencia prolongada.
3. Metodología: se describe la metodología aplicada comprometida en la adjudicación del presente proyecto, a partir de las mismas dataset; los factores ajenos al diagnóstico y características propias del accidente: Machine Learning (ML) y Enfoque Econométrico. Los modelos de

ML, usando un enfoque supervisado, permiten un alto nivel de precisión predictiva. Paralelamente, los modelos econométricos, permiten determinar el efecto marginal o probabilidad de ocurrencia de ausencia prolongada de un paciente. De esta forma, ambos enfoques aportan perspectivas complementarias, permitiendo robustecer los resultados obtenidos al validar si los mismos patrones emergen de ambos enfoques. Los datos utilizados en este estudio son de fuentes secundarias, entregados por la Mutual de Seguridad, en el contexto del desarrollo del proyecto de investigación SUSESO P200-INNO297. La entrega de la base de datos garantiza el anonimato de las personas y el proyecto cuenta con la aprobación del Comité de Ética Científico para Proyectos de Investigación.

4. Resultados: se presentan en línea con los dos enfoques metodológicos aplicados; Machine Learning: comprensión del negocio y de los datos, preparación de los datos ingeniería y evaluación interpretabilidad de los modelos. Para los modelos econométricos se presenta la evaluación, estadísticos e interpretabilidad de ellos.
5. Discusión: en la presente sección se comparan los resultados obtenidos con los recursos teóricos adquiridos en la revisión sistemática de la literatura.
6. Recomendaciones: para sistema de seguridad y salud en el trabajo: A partir de los resultados obtenidos se plantean recomendaciones a Mutual de Seguridad.
7. Conclusión: Se muestran las conclusiones, en línea con los objetivos del trabajo y la metodología utilizada que permitió abordar adecuadamente el problema planteado.
8. Referencias: Finalmente se presenta la bibliografía aplicada en el desarrollo del presente proyecto.

2. Definición del problema, pregunta de investigación o desafío de innovación

2.1. Definición del problema

Los tiempos de ausencia prolongada necesarios para la recuperación de un trabajador/a después de un accidente o enfermedad ocupacional es un relevante problema de salud ocupacional, con creciente interés por parte de la investigación científica (Cancelliere et al., 2016; Comcare, 2023; Kang, 2022; Duong et al., 2022; Samoborec et al., 2018; Selander et al., 2002). Desde un punto de vista de la salud ocupacional, el retorno al trabajo es deseable para todos los actores involucrados, esto es, la sociedad, el gobierno, los empleadores y las familias (Poblete, 2022), pero en especial para el propio trabajador, ya que favorece su salud física, mental y el bienestar. Al estar lejos del trabajo durante un tiempo prolongado, el trabajador lesionado o enfermo puede deprimirse, aislarse, sufrir consecuencias socioeconómicas, perder autoestima y verse afectado en sus relaciones familiares y sociales. Todo lo anterior, genera consecuencias adversas para su calidad de vida, salud y esperanza de vida. De este modo, brindar apoyo temprano a los empleados lesionados para permitirles su recuperación y regreso al trabajo es fundamental; cuanto más tiempo esté alguien sin trabajar, es menos probable que retorne al trabajo, ya que la ausencia laboral tiende a perpetuarse. Por ejemplo,

si la persona está fuera del trabajo por 20 días, la posibilidad de volver al trabajo es de 70 por ciento (el documento original no informa qué ocurre con el 30 % restante); no obstante, si la persona está fuera por 70 días, las probabilidades de volver decrecen dramáticamente a solo un 35 % (Comcare, 2023).

La Organización Mundial de la Salud (OMS), sugiere un enfoque colaborativo para la formulación de políticas en salud que considere de manera sistemática las implicancias sanitarias de las decisiones que se toman en salud. Es así como es clave la acción intersectorial para la salud, en la que se debe “instar a adoptar una estrategia de salud integral en la que no solo se presten servicios de salud, sino que además, se aborden los determinantes sociales en salud, que en largo plazo se transformará en una disminución del gasto en salud (OMS, 2008; 2013).

De acuerdo con la información disponible en la base de datos de MUSEG, se orientará la búsqueda de variables directamente vinculadas o del tipo proxy que representen los factores adicionales al diagnóstico y las determinantes de salud, que inciden en una ausencia prolongada y que permitan diseñar los modelos econométricos. Por ello, en el presente proyecto se consideran los factores adicionales al diagnóstico, tales como: género del trabajador, edad, distinción entre accidentados de trayecto y en el trabajo, tipo de contrato (menos estables a más estables, parcial o jornada completa), tipo de trabajo (manuales/operativos/administrativos, otros), nivel educacional, ingresos económicos, actividad económica y tamaño empresa, tipo de apoyo familiar, de círculos sociales y empleadores, entre otras. También los determinantes de salud, tales como: diagnóstico principal y secundario de salud, días de licencia médica, tipo de accidente, indicaciones farmacológicas, dolor crónico, salud mental, días de permanencia hospitalaria, estándar mutual, entre otras.

2.2. Pregunta de Investigación

De este modo, conocer los factores determinantes, adicionales al diagnóstico médico, relacionadas con aquellos trabajadores que presentan ausencia prolongada pasa a ser un objetivo de investigación relevante. Por ello, el presente proyecto busca responder a la pregunta ¿cuáles son los factores adicionales al diagnóstico y de salud que permitan predecir anticipadamente la ausencia prolongada de un paciente que ha enfrentado (1) un accidente del trabajo, (2) un accidente del trayecto o (3) una enfermedad profesional? La información que se busca obtener será relevante no solo para efectos de predecir la probabilidad de ausencia prolongada, sino que, además, para establecer estrategias de mitigación y prevención focalizadas en aquellos aspectos modificables que ponen en posición de vulnerabilidad al trabajador en cuanto a padecer una recuperación de larga estadía.

2.3. Objetivos

2.3.1. Objetivo General

Determinar los factores adicionales al diagnóstico que permitan predecir anticipadamente si un paciente tendrá ausencia prolongada, considerando la información asociada al otorgamiento de reposo a pacientes de la Mutual de Seguridad.

2.3.2. Objetivos Específicos

1. Validar factores que explican los días de reposo otorgados con la literatura.

2. Desarrollar una metodología para determinar los principales factores que explican los días de reposo otorgados.
3. Desarrollar un modelo econométrico que permita predecir de forma anticipada si un paciente tendrá ausencia prolongada, tomando en consideración los principales factores identificados anteriormente.
4. Validar metodología y resultados de los modelos con criterio experto, testeando en casos fuera de muestra.
5. Proponer recomendaciones que contribuyan a la implementación de los resultados.
6. Difundir los resultados en congresos y revistas indexadas.

3. Revisión de la literatura o experiencias relevantes

Se realizó una revisión sistemática cualitativa de artículos publicados, con el propósito de responder a las siguientes tres preguntas: ¿a partir de cuántos días de reposo se considera como tiempo de ausencia prolongada en los trabajadores con accidente laboral?; ¿qué relación existe entre ausencia prolongada de los trabajadores con accidente laboral con los determinantes sociales de la salud? y ¿qué relación existe entre ausencia prolongada en trabajadores con accidente laboral con los determinantes de la salud? Para ello, se utilizó como referencia la declaración PRISMA, y el resultado de la revisión ha sido aceptado como artículo de revisión en la Revista Médica de Chile (Alvear et al., 2025).

A continuación, se indica un resumen de los principales hallazgos encontrados y que se relacionan con el presente proyecto, distinguiendo entre variables: a) Personales y de la salud del trabajador, b) Situaciones y temporales, y c) De la Industria y la ubicación geográfica. Además, se indicará un apartado final con antecedentes que sirvan de parámetro para definir un criterio temporal sobre la larga recuperación.

3.1. Variables Personales y de la Salud del Trabajador

3.1.1. La edad del trabajador accidentado

Se muestra invariablemente como un factor relevante y asociado con los mayores tiempos necesarios de tratamiento para los afectados. En especial, el rango etario desde el cual comienza a destacar o enfatizarse los largos tiempos de recuperación es a partir de los 45 a 50 años en adelante (Baraza et al., 2023; Baraza & Cugueró-Escofet, 2021; Berglund et al., 2021; González Fuentes et al., 2022; Yang et al., 2020). Cabe hacer presente que la cantidad de accidentes es mayor en los trabajadores más jóvenes, pero la severidad se incrementa en la medida que aumenta la edad.

El envejecimiento de la población trabajadora pasa a ser un tema relevante para la salud ocupacional, que requiere especial atención para la formulación de políticas públicas y políticas de seguridad ocupacional (Bravo et al., 2022; Fontaneda et al., 2022; López et al., 2018), así como en el diseño de los puestos de los trabajadores mayores, tales como reducción de sistemas de turnos y cambios en las tareas para proteger su seguridad laboral (Fuentes-Bargues et al., 2022). En esta misma línea, la edad promedio de los trabajadores españoles accidentados ha pasado de 35 años en

1996 a 41 años para el 2015 (López et al., 2018), en tanto que la edad promedio de los trabajadores en Chile ha pasado desde los 41,3 años en 2010 a los 43,8 años en 2018 (CIEDES, 2024). Entre las causas de la mayor severidad de los accidentes en los trabajadores mayores, y con ello, de mayores tiempos necesarios para la recuperación, podemos considerar la fragilidad o las condiciones de salud preexistentes en ellos (Boufous & Williamson, 2009), así como también podría deberse a que los trabajadores de mayor edad experimentan tiempos de reacción significativamente más lentos frente a situaciones peligrosas en general (Doroudgar et al., 2017).

3.1.2. Las condiciones de salud preexistentes en el trabajador accidentado

El hecho que los accidentes aumenten su gravedad en la medida que afecta a una persona de mayor edad, en especial a aquellos sobre 45 años, se ha relacionado con las posibles condiciones de salud preexistentes y los niveles de fatiga de los afectados (Rey-Merchan & Lopez-Arquillos, 2021, 2022; López et al., 2018).

De lo anterior se desprende la importancia de la salud física y mental de los trabajadores como un factor preventivo a considerar para amortiguar los eventuales efectos negativos de los accidentes ocupacionales (Street & Lacey, 2019). De hecho, se ha observado que la realización de actividad física está inversamente relacionada con la probabilidad de tener un accidente ocupacional (Bradley et al., 2023; Yao et al., 2019), así como fumar, el consumo de alcohol y las comorbilidades se asocian con mayor ausentismo de los trabajadores (Sampaio et al., 2023; Shockey & Esser, 2020). También, los problemas del sueño están relacionados con los accidentes laborales (Vargas-Garrido et al., 2021; Yao et al., 2019). Aunque los antecedentes de salud son un factor previo a los accidentes, también éstos afectan la salud física y mental de los trabajadores (Gu et al., 2020). Por ello, intervenciones psicosociales y de salud mental de los trabajadores afectados contribuyen a menores tiempos de reposo (Nicholas et al., 2020; Nieuwenhuijsen et al., 2020).

3.1.3. Los trabajadores hombres accidentados

Los trabajadores hombres tienen más tiempo perdido cuando se trata de accidentes severos o de mayor gravedad, y que por lo tanto requieren mayores tiempos totales de recuperación. Las mujeres tienen más días de ausentismo cuando se considera la totalidad de accidentes ocupacionales en su conjunto (que en su mayoría son leves). En relación con las enfermedades profesionales, en Chile, éstas son mayoritariamente femeninas (66 %), para el año 2024 (SUSESO, 2025). Como demuestra Fontaneda y colegas (2019), en cada edad las mujeres tienen más días de ausentismo que los hombres (excepto para las edades entre 45 a 55 donde no hay diferencias). No obstante, cuando se selecciona aquel subconjunto de casos más severos, entonces son los hombres quienes resultan más afectados, y entonces, requieren una mayor cantidad de días para su recuperación.

Cabe hacer presente que pudieran existir diferencias debido a realidades locales y culturales, pero que reforzarían el hallazgo principal. Por ejemplo, en España, las mujeres se ven afectadas por mayor proporción de accidentes y presentan más días de ausencia promedio que los hombres (Fontaneda et al., 2019). Sin embargo, la situación en Chile es diferente, los hombres se ven más afectados proporcionalmente en accidentes en el lugar de trabajo y las mujeres en los accidentes de trayecto, pero en ambos tipos de accidentes los hombres son quienes presentan más días de tiempo perdido en promedio considerando los datos de accidentes globales (SUSESO, 2019). Algunas explicaciones posibles serían la menor participación de las mujeres en el mercado laboral chileno,

que los hombres trabajadores tienen mayor edad que las mujeres trabajadoras (un 20 % de los hombres tienen 55 años o más, mientras que solo un 15 % de ellas está en ese rango etario), y que las mujeres jubilan a menor edad (60 años) que ellos (65) por lo que los trabajadores permanecen más tiempo en el mercado laboral en edades que implican mayor riesgo de fragilidad, y por tanto, accidentes con mayor severidad y ausentismo (SUSESO, 2024).

3.1.4. La condición migrante del trabajador

Cuando se toma el total de los accidentes, los locales tienden a presentar mayores días de ausentismo (González Fuentes et al., 2022); sin embargo, cuando se toma los casos más severos, entonces, son los inmigrantes quienes tienen más días perdidos (Baraza et al., 2023; Rey-Merchan & Lopez-Arquillos, 2021; Rey-Merchán et al., 2022). Algunas explicaciones de por qué los locales tienen más días de tiempo perdido obedece a que la población local es de mayor edad usualmente y las reclamaciones de los inmigrantes estén subrepresentadas por problemas de idioma, desconocimiento de la legislación o temor de informar una lesión laboral cuando se trata de accidentes leves o menos severos (Alacevich & Nicodemo, 2024; Baraza & Cugueró-Escofet, 2022). No obstante, en el caso de las lesiones con mayor severidad resulta más homogénea la tasa de denuncia por la evidencia y gravedad del evento, y en estos casos, son los inmigrantes quienes más días requieren para su recuperación.

3.1.5. El tipo de lesión generada por el accidente

Entre los tipos de lesiones más frecuentemente asociadas con mayor severidad y larga estadía se encuentran las fracturas, lesiones múltiples, amputaciones, con afectación principalmente a los miembros superiores e inferiores. Además, también se observan como consecuencias severas las contusiones de cráneo y lesiones internas (Baraza et al., 2023; Baraza & Cugueró-Escofet, 2021; Berglund et al., 2021; Fuentes-Bargues et al., 2022; González Fuentes et al., 2022).

3.1.6. La condición del accidente

Los tipos de accidentes que implican mayor severidad se relacionan con caídas, tropiezos y resbalones, con atrapamientos, donde interviene una máquina o equipo, o la pérdida de control de éste (Baraza et al., 2023; Baraza & Cugueró-Escofet, 2021; Berglund et al., 2021; Fontaneda et al., 2022; Fuentes-Bargues et al., 2022a; González Fuentes et al., 2022a; Höyükler et al., 2022).

3.2. Variables Situaciones y Temporales del Accidente

3.2.1. Día de la semana del accidente: hacia el fin de semana

El efecto “lunes” o el “efecto días de la semana” ha sido ampliamente descrito por la literatura (Fontaneda et al., 2024), esto es, la mayor cantidad de accidentes ocurre decrecientemente de lunes a viernes. Por ejemplo, desde los 21,82 % y 20,35 % los lunes a los 15,3 % y 15,78 % los viernes. También decrece los sábados y domingos, pero con porcentajes mucho menores, por ejemplo 3,4 % y 9,3 % los sábados a los 1,1 % y 3,1 % los domingos (Ahn, 2022; Baraza et al., 2023; Baraza & Cugueró-Escofet, 2021; Fontaneda et al., 2019, 2022; Fuentes-Bargues et al., 2022; López et al., 2018). No obstante, el incremento de los días perdidos de los accidentes ocurre de manera inversa,

esto es, los días perdidos promedio tienden a ser menores los lunes (22,8 días) comparados con los domingos (29,0) (López et al., 2018). Aunque esta dirección es la misma por sexo, algunos investigadores advierten que esto se replica en el caso de los hombres (de 23,4 a 27,5 de lunes a domingo), pero que en las mujeres el efecto se produce solo de lunes a viernes (de 25,2 a 27,1 de lunes a viernes) ya que el fin de semana decrece un poco (25,8 días promedio, pero todavía mayor que el lunes y martes en ellas) (Fontaneda et al., 2019).

Consistentemente, se ha observado que los accidentes con mayor severidad (y por tanto con mayor tiempo perdido) tienden a ocurrir en mayor proporción desde los jueves a domingo (Baraza & Cugueró-Escofet, 2021), los sábados y domingos (Fontaneda et al., 2022; Fuentes-Bargues et al., 2022), los viernes, sábados y jueves (Ahn, 2022), o bien los viernes (Baraza et al., 2023). Además, las personas mayores de 50 años tienden a accidentarse más hacia el final de la semana, los días jueves y viernes (Fontaneda et al., 2022). Estos hallazgos serían consistentes con la cada vez más creciente atención sobre los problemas de la fatiga o cansancio como variable interviniente en los accidentes ocupacionales (González Fuentes et al., 2022; López et al., 2018) y, al parecer, con aquellos que implican mayores días de tiempo perdido y severidad (Fischer et al., 2017; González Fuentes et al., 2022a; Rey-Merchan & Lopez-Arquillos, 2021).

3.2.2. Jornadas y hora del accidente: atardecer, anochecer y madrugadas

Las lesiones que implican mayor severidad y tiempo perdido tienden a ocurrir durante las jornadas de la noche, incluyendo el amanecer, y por las tardes y al atardecer (Ahn, 2022; Arocena & García-Carrizosa, 2023; López et al., 2018). Consistentemente también ocurren los accidentes más severos en los trabajadores que ingresan al turno por la noche y por la tarde (Yang et al., 2020).

En cuanto las horas de la jornada, se ha observado mayor riesgo de lesión severa a la tercera hora de la jornada (Fuentes-Bargues et al., 2022), entre la 5ta y la 7ma hora del turno, así como en la hora 13 (Yang et al., 2020). Si consideramos que los trabajadores pueden ingresar por la mañana, tarde o noche a su turno, esta información indica que los mayores riesgos serán en las tardes y noches, que coincide con el punto anterior.

Coincidentemente, se observa que los accidentes de trayecto también tienden a ser más severos por las tardes y por las noches (Rey-Merchan & Lopez-Arquillos, 2021). Además, si consideramos que las horas del día con mayor severidad de accidentes, en orden de magnitud son: entre las 6 y 8 am, entre las 14 y 16 pm, y entre las 18 a 20 pm (Fontaneda et al., 2019), éstas parecen estar relacionadas con las horas en las cuales los trabajadores se desplazan desde y hacia su lugar de trabajo.

Toda la información anterior apunta a una mayor severidad de las lesiones, y con ello, mayor tiempo de días perdidos al atardecer, anochecer y en las madrugadas, lo que parece nuevamente estar relacionado con los problemas del cansancio y fatiga (Fischer et al., 2017; González Fuentes et al., 2022; Rey-Merchan & Lopez-Arquillos, 2021), así como los problemas del sueño (Pietilä et al., 2018), los cuales afectan principalmente a los trabajadores mientras se desplazan a sus trabajos (Vargas-Garrido et al., 2021).

3.2.3. Accidentes de trayecto

Los accidentes del trayecto son aquellos que implican mayores días de tiempo perdido y mayor severidad (Baraza et al., 2023; Baraza & Cugueró-Escofet, 2021; Fontaneda et al., 2019; Fuentes-

Bargues et al., 2022; Höyükler et al., 2022). Además, los accidentes de trayecto implican mayores costos, generan mayor discapacidad y muerte. Dada las diferencias en los patrones entre los accidentes en el trabajo y en el trayecto, algunos investigadores sugieren poner más atención en estos últimos y estudiarlos de manera diferenciada de los primeros (Vargas-Garrido et al., 2021).

3.2.4. Año de ocurrencia del accidente

La revisión de la literatura comparada muestra que algunos años resultan con mayores días necesarios para la recuperación (Fuentes-Bargues et al., 2022; Yang et al., 2020). Esto podría estar relacionados con períodos de expansión económica, que quizás aumente la demanda por los servicios médicos y con ello genere retrasos de atención y finalmente afecte la pronta atención. En Chile, por ejemplo, aunque el año 2020 las tasas de accidentes del trabajo y del trayecto son las más bajas del período 2015-2023, es el año con más días de tiempo perdido en el mismo período (SUSESO, 2024), probablemente debido a las altas demandas por el COVID-19.

3.2.5. Tiempos transcurridos en cualquier etapa de atención clínica o administrativa relacionada con el accidente

De los artículos revisados se desprende que mientras mayor el tiempo que pase el trabajador hospitalizado (Gray & Collie, 2019), mientras más tiempo demore desde el accidente a la atención del paciente, y más tiempo demore la calificación del evento como un accidente ocupacional (Besen et al., 2018) será mayor el tiempo de recuperación que requerirá el trabajador. En general, las tendencias parecen apuntar a que todo tiempo adicional que demore el proceso de provisión del servicio de salud y las gestiones de salud asociadas incidirán en más días de recuperación. Consistentemente, se ha observado que, a menor tiempo de atención del lesionado, los tiempos de alta hospitalaria son menores y el paciente obtiene mejores resultados de salud (Bharti et al., 2019; Iacobucci, 2021) debido a la prontitud de la atención. Igualmente, una de las variables que puede afectar la mayor cantidad de días de ausentismo en los trabajadores accidentados por las noches y los fines de semana se debería a que hay menos personal en los turnos para apoyar al afectado y entregar una asistencia oportuna.

3.3. Variables de la Industria y la ubicación geográfica

3.3.1. Empresas de menor tamaño

Los accidentes con más tiempo perdido y severidad se encuentran en las empresas más pequeñas (Ahn, 2022; Baraza et al., 2023; Baraza & Cugueró-Escofet, 2021; Berglund et al., 2021; Fontaneda et al., 2019; Rey-Merchan & Lopez-Arquillos, 2021). Las empresas pequeñas usualmente carecen de un adecuado sistema de seguridad laboral, muy necesarias para la prevención y gestión del riesgo ocupacional de los trabajadores (González Fuentes et al., 2022).

3.3.2. Sectores con mayores proporciones de accidentes severos

Construcción, empresas de transporte, y agricultura (Höyükler et al., 2022), más accidentes recurrentes (con mayor tiempo perdido) en producción que en servicios (Pietilä et al., 2018).

3.3.3. Regiones o provincias donde ocurre el accidente

Los accidentes que ocurren a trabajadores en algunas provincias o regiones presentan mayores días de recuperación. Por ejemplo, en el caso de España, frecuentemente aparece Extremadura y Galicia (más al norte y más al sur) (Baraza et al., 2023; Baraza & Cugueró-Escofet, 2021; Fuentes-Bargues et al., 2022). Como proxy, en Chile, las mayores tasas de accidentes se encuentran en las regiones de Arica (norte) y en las de Aysén y Magallanes (extremo sur) (SUSESO, 2024). Quizás esta información se relacione con la menor provisión de oferta de salud en las regiones más distantes a la zona central, o bien, a las mayores distancias a los centros de atención debido a la distancia geográfica, lo cual afecte los tiempos de respuesta oportuna en la atención de los pacientes más graves.

3.4. Sobre la estimación de tiempos de larga estadía o recuperación

Los trabajos con muestras españolas utilizan como criterio de clasificación de la gravedad de los accidentes una escala que hemos denominado “escala española”, que es una escala de gravedad de los accidentes, que son evaluadas por médicos, siguiendo los siguientes criterios de clasificación: Leve (sin implicancia para el trabajador), severo (con implicancias, pero recuperable), muy severo (puede causar invalidez o incapacidad parcial) y fatales (un 96 % de estos casos implican hasta dos días de ausentismo). Cabe hacer notar que la gran mayoría de los casos son leves (98 %). De este modo, si consideramos aquellos casos evaluados como severos, ellos muestran significativamente más días perdidos. López y colegas (2018) encuentran un promedio de recuperación de 24 días cuando los accidentes son tomados en conjunto, sin embargo, el promedio aumenta dramáticamente a 104 días cuando se trata de casos severos y fatales. Por ello, debido a que los casos severos aumentan a partir del tercer mes, la presente revisión sistemática permite sostener un punto de corte para definir una larga recuperación a partir de los 90 días.

Podemos concluir que los promedios de recuperación globales están entre los 24 a 31 días en España (dependiendo de la industria) considerando los accidentes en conjunto (predominando las lesiones leves). Coincidentemente, en Chile, el año 2023 los días de tiempo perdido promedios fueron de 18.3 en los accidentes del trabajo y 24 en los de trayecto. No obstante, cuando se consideran los accidentes evaluados como severos por médicos en España, los días de recuperación pasan a ser superiores a 90 días (con un promedio de 104 días). De este modo, existe fundamento para considerar como accidentes de larga estadía o recuperación algún punto entre esas cifras (90 a 104 días de recuperación). Por supuesto, este punto de corte requiere ser precisado y definido a futuro, dependiendo de variables tales como país, industria, tipo de lesión, criterios médicos, etc.

4. Descripción de la metodología o etapas de la innovación

El presente trabajo contempla dos metodologías complementarias, comprometidas en la adjudicación del proyecto, con el objetivo de determinar la ausencia prolongada de un trabajador que ha sufrido un accidente del trabajo, del trayecto o enfermedad profesional. A partir de las mismas variables; los factores ajenos al diagnóstico y características propias del accidente.

La primera de ellas, modelos Machine Learning (ML). Los modelos Machine Learning permiten un alto nivel de precisión predictiva, usando técnicas de aprendizaje supervisado permite el logro del objetivo del presente trabajo. La segunda de ellas, Modelos Económicos, aplicando el modelo

probit, se logra determinar el efecto marginal o probabilidad de ocurrencia de ausencia prolongada de un paciente.

De esta forma la presencia de ambos enfoques permite robustecer los resultados del proyecto, al validar si los mismos patrones emergen de ambos enfoques.

Los datos utilizados en este estudio son de fuentes secundarias, entregados por la Mutual de Seguridad, en el contexto del desarrollo del proyecto de investigación SUSESO P200-INNO297.

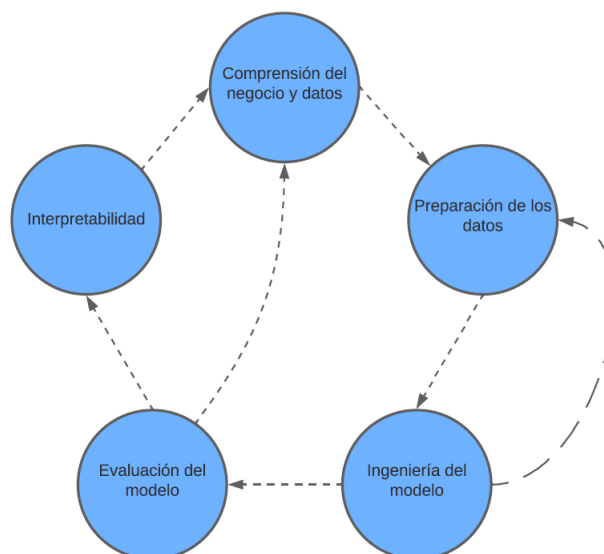
El conjunto de datos original consta de 3.327.747 observaciones y 33 variables que registran casos con fecha de admisión a la Mutual de Seguridad entre el 1 de enero de 2011 y el 31 de diciembre de 2024 en sus centros de atención a nivel nacional. Para asegurar la consistencia y relevancia analítica, se conservaron exclusivamente los siniestros clasificados como accidente aceptado, excluyéndose los reingresos y los casos rechazados. Se eliminaron además los registros con datos faltantes, salvo en la variable porcentaje de incapacidad, cuyos valores ausentes se imputaron con cero bajo la presunción de ausencia de incapacidad. Finalmente, se depuraron los registros que presentaban errores o inconsistencias en cualquier campo. Tras estas etapas de limpieza, la muestra final quedó constituida por 1.196.355 registros.

La entrega de la base de datos garantiza el anonimato de las personas de acuerdo con protocolo de anonimización de datos, definido por Mutual de Seguridad. El proyecto cuenta con la aprobación del Comité de Ética Científico para Proyectos de Investigación, según Resolución Nro. 352 del 22 de abril de 2024.

4.1. Metodología Machine Learning

Para asegurar la calidad del proyecto, utilizamos la metodología Cross-Industry Standard Process for Machine Learning con enfoque en la calidad (CRISP-ML(Q)). Esta metodología es ampliamente utilizada en proyectos de ML y publicaciones del sector salud (Kolyshkina and Simoff, 2021; Silva-Aravena et al., 2023). A esta metodología incorporamos una fase de interpretabilidad, que busca facilitar la comprensión de las predicciones del modelo utilizado. La Figura 1 describe las etapas de la metodología utilizada.

Figura 1: Diagrama de Metodología utilizada en estudio.



4.1.1. Comprensión del negocio y los datos.

En esta etapa se busca entender en profundidad el objetivo del proyecto desde una perspectiva organizacional y clínica. Para ello, se realizaron reuniones con la contraparte técnica, con el fin de comprender el negocio, la definición operativa de “ausencia prolongada” y el uso que podrían tener los resultados del modelo en la toma de decisiones.

Asimismo, se revisaron junto a la contraparte los conjuntos de datos disponibles, evaluando su estructura y descripción de variables incluidas. Esta interacción permitió identificar variables potencialmente relevantes, aclarar codificaciones internas y detectar posibles restricciones de uso o interpretación de la información entregada. Junto a lo anterior, se realiza una exploración inicial de los datos entregados por la Mutual.

4.1.2. Preparación de los datos.

En esta etapa se construye el conjunto de datos final que fue utilizado para en la fase siguiente. A partir de los datos provistos por la Mutual de Seguridad, se realiza una depuración para corregir errores, gestionar valores faltantes y eliminar registros inconsistentes. Además, se generan nuevas variables que podrían ser relevantes para predecir la ausencia prolongada. Los resultados de esta depuración y generación de variables se describen en las secciones 5.1.2.1 y 5.1.2.2 respectivamente.

Se realizó un análisis de posible desequilibrio de clases. Este desbalance de clases puede afectar el desempeño de los modelos predictivos, por lo que se evaluaron estrategias de tratamiento para mitigar este efecto. Finalmente, se transforma la información a un formato compatible con los algoritmos de aprendizaje automático, asegurando la calidad y coherencia de los datos.

4.1.3. Ingeniería del Modelo.

Para discriminar a pacientes de larga estadía, evaluamos 6 modelos de machine learning ampliamente utilizados en problemas de clasificación. Los modelos a evaluar son los siguientes:

1. **Classification and Regression Trees (CART).** El algoritmo CART utiliza árboles de decisión binarios para clasificar observaciones. Dado un conjunto de datos $D = \{(x_i, y_i)\}_{i=1}^n$, donde $x_i \in \mathbb{R}^m$ y $y_i \in \{0, 1\}$, cada división busca minimizar el *índice de Gini*:

$$\text{Gini}(R) = 1 - p_0^2 - p_1^2 \quad (1)$$

donde p_c es la proporción de muestras de clase c en la región R . La predicción final se realiza según la clase mayoritaria en la hoja correspondiente del árbol.

2. **Ensamble de Árboles de Decisión (Random Forest).** Es una técnica de aprendizaje por ensamblado que combina múltiples árboles CART entrenados sobre muestras bootstrap del conjunto de entrenamiento y subconjuntos aleatorios de variables (Breiman, 2001). Para un conjunto de datos $D = \{(x_i, y_i)\}_{i=1}^n$, la predicción del árbol j -ésimo es:

$$g_j(x_i) \in \{0, 1\} \quad (2)$$

La predicción final es el resultado de una votación mayoritaria entre los M árboles:

$$\hat{y}_i = \text{moda}\{g_1(x_i), g_2(x_i), \dots, g_M(x_i)\} \quad (3)$$

Random Forest puede también estimar la probabilidad de que un paciente tenga una larga estadía:

$$\mathbb{P}(y_i = 1|x_i) = \frac{1}{M} \sum_{j=1}^M g_j(x_i) \quad (4)$$

3. **Árboles Extremadamente Aleatorizados (ExtraTrees)**. Es similar a Random Forest pero introduce mayor aleatoriedad en el proceso de división de nodos (Geurts et al., 2006). En lugar de buscar el mejor punto de corte, selecciona puntos de corte aleatorios para cada variable. La predicción de cada árbol $g_j(x_i)$ y la final $\hat{y}_i$ siguen las mismas ecuaciones (2) y (3). Esta aleatoriedad adicional reduce la varianza del modelo.
4. **Light Gradient Boosting Machine (LightGBM)**. Es un modelo de *boosting* por gradiente que utiliza árboles de decisión como modelos base, entrenados de manera aditiva (Ke et al., 2017). Para una muestra x_i , la predicción en la iteración t es:

$$\hat{y}_i^{(t)} = \sigma \left(\sum_{k=1}^t f_k(x_i) \right) \quad (5)$$

donde $\sigma(z) = \frac{1}{1+e^{-z}}$ es la función sigmoide y f_k representa un árbol de regresión. La función objetivo aproximada es:

$$\text{obj}^{(t)} \approx \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \quad (6)$$

donde g_i y h_i son el gradiente y hessiano de la pérdida logarítmica respecto a $\hat{y}_i$.

5. **Categorical Boosting (CatBoost)**. Es un algoritmo de *boosting* optimizado para variables categóricas (Dorogush et al., 2018). Utiliza estadísticas ordenadas para codificar las variables y *ordered boosting* para evitar filtraciones de la variable objetivo. La predicción en la iteración t es:

$$\hat{y}_i^{(t)} = \sigma \left(\sum_{k=1}^t f_k(x_i) \right) \quad (7)$$

Y la función objetivo aproximada:

$$\text{obj}^{(t)} \approx \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \quad (8)$$

CatBoost transforma las variables categóricas mediante:

$$\text{TS}(x_{ij}) = \frac{\sum_{k < i} y_k \cdot \mathbb{I}[x_{kj} = x_{ij}] + a \cdot p}{\sum_{k < i} \mathbb{I}[x_{kj} = x_{ij}] + a} \quad (9)$$

donde a y p son parámetros de regularización.

6. **Extreme Gradient Boosting (XGBoost).** Es un algoritmo de *boosting* por gradiente que combina secuencialmente múltiples árboles de decisión (Chen y Guestrin, 2016). Para cada muestra, la predicción del árbol j es:

$$g_j(x_i) = w_{q(x_i)} \quad (10)$$

donde w_q es el peso de la hoja. La predicción total es la suma de las salidas de cada árbol:

$$\hat{y}_i = \sum_{j=1}^M g_j(x_i) \quad (11)$$

La función objetivo en la iteración t es:

$$\text{obj}^{(t)} = \sum_{i=1}^n \ell(y_i, \hat{y}_i^{(t-1)} + g_t(x_i)) + \Omega(f_t) \quad (12)$$

que se aproxima mediante expansión de Taylor de segundo orden:

$$\text{obj}^{(t)} \approx \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \quad (13)$$

4.1.4. Evaluación del Modelo

Para la selección del modelo, se emplea la técnica de benchmarking, la cual permite evaluar comparativamente diversos algoritmos con el objetivo de identificar aquel que presenta el mejor desempeño en la tarea de clasificación de pacientes. El modelo seleccionado será posteriormente utilizado para la interpretación de sus resultados. El rendimiento de los algoritmos se evalúa en función de métricas como accuracy, precision, F1-score y recall.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (15)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (16)$$

$$F_1 = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (17)$$

TP, FP, FN y TN corresponden, respectivamente, a verdaderos positivos, falsos positivos, falsos negativos y verdaderos negativos. Los verdaderos positivos (TP) representan el número de pacientes de larga estadía que el algoritmo clasifica correctamente como tales. Los falsos positivos (FP) corresponden al número de pacientes de baja estadía que el modelo identifica erróneamente como de larga estadía. Los falsos negativos (FN) son aquellos pacientes de larga estadía que son clasificados incorrectamente como de baja estadía. Por su parte, los verdaderos negativos (TN) representan el número de pacientes de baja estadía correctamente identificados.

La métrica accuracy, presentada en la ecuación (14), se define como el cociente entre el número total de pacientes correctamente clasificados y el total de pacientes analizados. La precisión, según la ecuación 15, corresponde a la proporción de pacientes de larga estadía correctamente identificados sobre el total de pacientes que el modelo clasifica como de larga estadía. El recall, descrito en la ecuación (16), se refiere a la proporción de pacientes de larga estadía correctamente identificados respecto del total real de pacientes pertenecientes a dicha categoría. Finalmente, el F1-score es la media armónica entre la precision y el recall, ecuación 17. Toma valores entre 0 y 1, y solo es alto cuando ambas métricas lo son.

En conjunto de datos se divide en 70 % para entrenamiento y 30 % para test. En el proceso de entrenamiento y selección del modelo, se utilizaron los hiperparámetros óptimos de cada modelo. Para el cálculo de los hiperparámetros óptimos se emplea el estimador de Parzen con estructura de árbol (Tree-Structured Parzen Estimator) como algoritmo de muestreo, y la técnica de poda basada en la mediana (Median Pruning) como método de descarte de configuraciones poco prometedoras. En la Tabla 1 se observa el espacio de búsqueda para cada modelo.

Tabla 1: Espacio de búsqueda de hiperparámetros por modelo. Los rangos del tipo $[a, b]$ indican un intervalo cerrado, es decir, se consideran todos los valores entre a y b , incluyendo los extremos.

Modelo	Hiperparámetros	Espacio de Búsqueda
Random Forest	n_estimators	[50, 500]
	max_depth	[5, 50]
	min_samples_split	[2, 10]
	min_samples_leaf	[1, 10]
Extra Trees	n_estimators	[50, 300]
	max_depth	[5, 50]
	min_samples_split	[2, 10]
	min_samples_leaf	[1, 10]
LightGBM	n_estimators	[50, 500]
	max_depth	[3, 20]
	learning_rate	log-uniform [0.01, 0.3]
	num_leaves	[20, 150]
CatBoost	iterations	[50, 500]
	depth	[3, 10]
	learning_rate	log-uniform [0.01, 0.3]
XGBoost	n_estimators	[50, 500]
	max_depth	[3, 15]
	learning_rate	log-uniform [0.01, 0.3]
	subsample	[0.5, 1.0]
	colsample_bytree	[0.5, 1.0]
CART (Árbol de Decisión)	max_depth	[1, 50]
	min_samples_split	[2, 10]
	min_samples_leaf	[1, 10]
	criterion	{gini, entropy}

4.1.5. Interpretabilidad

Para interpretar el modelo que presenta el mejor desempeño utilizamos Shapley Additive Explanations (SHAP) debido a su robusta base teórica y amplio uso en la interpretación de modelos de aprendizaje automático (Keren Evangeline et al., 2022; Zhang et al., 2022).

SHAP es una metodología basada en la teoría de juegos que asigna a cada característica una contribución cuantitativa al resultado de una predicción, lo que permite comprender cómo y en qué medida cada variable influye en las decisiones del modelo.

La función de predicción se denota como $f(x)$, y F representa el conjunto de todos los parámetros de entrada. Los valores SHAP se calculan de la siguiente manera:

$$\Phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]; \quad (18)$$

Donde $|F|$ corresponde al número total de parámetros de entrada del modelo, S es un subconjunto de características que no incluye la i -ésima característica, $|S|$ representa la cantidad de elementos en ese subconjunto, y $f_S()$ denota la función de predicción del modelo restringida al conjunto S .

4.2. Enfoque Econométrico

La definición de un modelo econométrico que permita predecir si un paciente será o no de ausencia prolongada, en función de un conjunto de variables, distintas al diagnóstico principal y de salud.

4.2.1. Variable dependiente

La variable dependiente se construyó sobre la base del número de días de reposo (reposo). El número de días de lo que se considera el límite de una larga estadía o ausencia prolongada, se establece según lo indicado por la bibliografía (90 días o más) y lo recomendado por Mutual. Para efectos del modelo econométrico 1= con larga estadía y 0= sin larga estadía.

4.2.2. Variables independientes

Las variables independientes fueron definidas a partir de la base de datos entregada por Mutual de Seguridad y fueron agrupadas según lo recomendado por la literatura. Para efectos del modelo econométrico todas las variables independientes son transformadas en variables dummy. Las alternativas de respuesta de variables multicategóricas fueron convertidas en variables dummy independientes, es decir, con respuestas “0” y “1”.

4.2.3. Análisis de los datos

4.2.3.1. Accidentes del Trabajo La base de datos incluye 130.752 observaciones con accidente del trabajo. Del total, 6.228 (4,76 %) presentan ausencia prolongada y 124.524 (95,24 %) no presentan dicha situación. Se procedió a equilibrar la base de datos, tal como se definió para los modelos ML, para ello se seleccionó aleatoriamente 6.228 observaciones que no presentan ausencia prolongada, con el objetivo de presentar un número equilibrado entre observaciones que presentan y

no presentan ausencia prolongada. Finalmente, el estudio incluye 12.456 observaciones. De ellas, el 50 % (6.228) son trabajadores que han sufrido un accidente del trabajo, con ausencia prolongada y el 50 % restante son trabajadores sin larga estadía o ausencia prolongada. Se procedió a seleccionar al azar 70 % de los datos para entrenamiento y 30 % para efectos de validación del modelo.

4.2.3.2. Accidentes de Trayecto La base de datos incluye 59.956 observaciones con accidente de trayecto. Del total, 5.470 (9,12 %) presentan ausencia prolongada y 54.486 (90,88 %) no presentan. Se procedió a equilibrar la base de datos, tal como se definió en los ML, para ello se seleccionó aleatoriamente 5.470 observaciones que no presentan ausencia prolongada, con el objetivo de presentar un número equilibrado entre observaciones que presentan y no presentan ausencia prolongada. Finalmente, el estudio incluye 10.940 observaciones. De ellas, el 50 % (5.470) son trabajadores que han sufrido un accidente de trayecto, con ausencia prolongada y el 50 % restante son trabajadores sin ausencia prolongada. Se procedió a seleccionar al azar 70 % de los datos para entrenamiento y 30 % para efectos de validación del modelo.

4.2.3.3. Enfermedad Profesional La base de datos incluye 342 observaciones con enfermedad profesional. Del total, 29 (8,48 %) presentan ausencia prolongada y 313 (91,52 %) no presentan dicha condición. Se procedió a equilibrar la base de datos, tal como se definió en los ML, para ello se seleccionó aleatoriamente 29 observaciones que no presentan ausencia prolongada.

El análisis estadístico de los datos se realizó con el apoyo del Software Stata/SE 18. Se aplicó los modelos logit y probit, descritas en las ecuaciones 19 y 20 los cuales utilizan una función que asume únicamente valores entre cero y uno. El método para estimar ambos modelos es de Máxima Verosimilitud. Para evaluar y decidir entre ambos modelos se usó el estadístico AIC (Akaike) [3]. Las fórmulas de los modelos y del estimador de calidad se muestran a continuación:

$$\Pr(y = 1 \mid \mathbf{x}) = \Lambda(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k),$$

$$\Lambda(z) = \frac{e^z}{1 + e^z}. \quad (19)$$

Donde:

- $y = 1$: trabajador con ausencia prolongada;
- $\mathbf{X}$: vector de covariables distintas al diagnóstico y de salud;
- P : probabilidad de que $y = 1$;
- Λ : función logística;
- β : parámetros del modelo, estimados por máxima verosimilitud.

$$\Pr(y = 1 \mid \mathbf{x}) = \Phi(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k),$$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt. \quad (20)$$

Donde:

$y = 1$: trabajador con larga estadía;

P : probabilidad de que $y = 1$;

Φ : función de distribución acumulada (FDA) de la distribución normal estándar;

β : parámetros del modelo, estimados por máxima verosimilitud.

Para evaluar y decidir entre ambos modelos se usó el estadístico Akaike (AIC), que se describe en la ecuación 21.

$$AIC = 2k - 2\ln(\hat{L}) \quad (21)$$

Donde:

k = número de parámetros estimados en el modelo,

$\hat{L}$ = valor máximo de la función de verosimilitud del modelo.

A partir de dicho modelo se estiman los Efectos Marginales Promedio (AME). A través de esta técnica, es posible medir el impacto que, en promedio, cada variable independiente (controlando por los demás factores) tiene sobre la probabilidad de que una persona presente ausencia prolongada en Mutua de Seguridad. El cálculo de AME se obtiene a través de la ecuación 22.

$$AME = \frac{1}{N} \sum_{i=1}^N \frac{\partial E[y_i | \mathbf{x}_i, \mathbf{w}_i]}{\partial x} \beta_x \quad (22)$$

5. Resultados

5.1. Enfoque Machine Learning

5.1.1. Comprensión del negocio y los datos.

De reunión realizada con contraparte se recibe el conjunto de datos. Se compone de 3.327.747 instancias y 32 variables, con información del periodo 01/01/2011 hasta el 31/12/2024 para la fecha de admisión. La Tabla 2 describe las variables del conjunto de datos.

Tabla 2: Descripción de variables del conjunto de datos

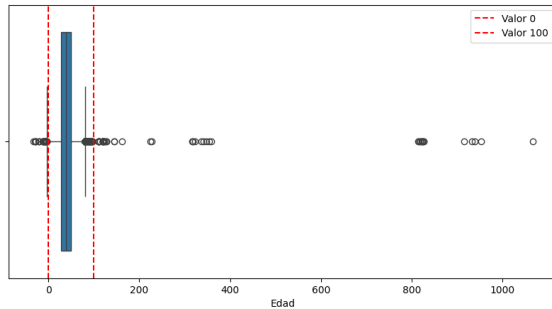
nº	variable	descripción
1	atdn_clasificacion_denuncia	Clasifica la denuncia en una combinación de 3 categorías, resolución_accidente (i.e. aceptado, rechazado o reingreso) , sin tiempo perdido (STP) o con tiempo perdido (CTP) y si es de trabajo (TB), trayecto (TY) o enfermedad profesional (EP). Por ejemplo, “Aceptado CTP TB” se interpreta como una denuncia por accidente de trabajo aceptada, con tiempo perdido.
2	atdn_atd_fecha_ocurrencia	Contiene la fecha en que ocurrió el accidente o siniestro. Es esencial para cualquier análisis cronológico.
3	atdn_fecha_admision	Indica la fecha en que se admitió la denuncia (puede haber un pequeño desfase con el ingreso del paciente). Es útil para medir tiempos de respuesta o el proceso administrativo posterior al siniestro.
4	edad	Es la edad del trabajador al momento en que fue emitida la denuncia.
5	pcnt_genero	Es el sexo del paciente (i.e Femenino (F); Masculino (M))
6	relato_anamnesis	Es la combinación del texto del “relato” con el texto de la anamnesis.
7	cmns_comuna	Identifica la comuna donde ocurre el accidente.
8	cmns_region	Es la región en la que se encuentra la comuna. Es útil para análisis geográficos a una escala más grande que la comuna.
9	porcentaje_incapacidad	Indica el grado de incapacidad que resultó del accidente. Es un cálculo para establecer grado de incapacidad del trabajador. Sobre 15 se considera como incapacidad leve (e.g. perdida de una falange) y más de 70 son casos crónicos (e.g. grandes amputados, postrados)
10	cnat_id_centro	Es un identificador único para el centro de atención médica donde se trató el paciente.
11	cnat_centro_atencion	Indica el nombre o descripción del centro de atención donde se trató el accidente. Complementa al ‘cnat_id_centro‘ que es respectivo ID. [HOSPITAL SANTIAGO]
12	caso_id_anonimo	Es un identificador único, específico para un caso en cuestión. Por ejemplo, un trabajador accidentado podría ingresar por una fractura y posteriormente por desarrollar una patología mental producto del accidente inicial. Dicho caso, se asocia a un solo ID de caso (i.e. accidente inicial), pero a dos ingresos por distintos diagnósticos (i.e. dos números de siniestros distintos). También considera los reingresos (un paciente puede reingresar por un caso particular, y tendrá un nuevo atdn_id_unico, pero el mismo caso_id_caso)
13	caso_id_reingreso	Indica si el caso es un reingreso (1) (está relacionado con un caso anterior) o es un nuevo caso (0). Es importante para entender patrones recurrentes o para filtrar reingresos si es necesario.
14	adhr_numero	Este es el identificador numérico único asignado a cada empresa afiliada a Mutual de Seguridad.
15	caso_fecha_ingreso	Indica la fecha en que se ingresó el caso en el sistema. Es útil para análisis administrativos y para medir tiempos de respuesta.

nº	variable	descripción
16	caso_nom_diag_1_may_est	Contiene el nombre del diagnóstico primario (de mayor estándar). Corresponde a una clasificación de las lesión o enfermedad.
17	caso_nom_diag_2_may_est	Es el nombre del segundo diagnóstico más relevante para el caso. Ayuda a entender casos que involucren múltiples lesiones o condiciones. Por ejemplo, una persona que es asaltada podría tener una fractura como diagnóstico de primer estándar, y en el de segundo estándar algún diagnóstico derivado de las consecuencias psicológicas.
18	caso_nom_diag_3_may_est	Es el nombre del tercer diagnóstico más relevante, siguiendo la misma lógica que los dos campos anteriores.
19	caso_profesion_ocupacion	Describe la profesión o la ocupación del individuo. Es relevante para identificar grupos de riesgo específicos.
20	caso_examen_fisico	Contiene los resultados del examen físico realizado. Proporciona información sobre el estado físico del individuo tras el accidente y el estado de daño.
21	grave	Es una variable binaria que indica si el accidente tuvo consecuencias graves (según la Norma Técnica 142 del MINSAL). Se utiliza para clasificar la gravedad del caso. (i.e. grave: 1; no grave: 0).
22	adhr_macrosegmento	Es una categorización del tipo de industria o sector al que pertenece el adherente. Útil para análisis sectoriales. [PYME, GRAN CORPORACIÓN, INDEPENDIENTE, ETC]
23	act_econ	Es el nombre de la actividad económica de la empresa adherente. Permite un análisis más detallado del tipo de industria en que ocurren los accidentes. Se extrae desde clcu_codigo.
24	diag_cod_cie10	Es el código del sistema de clasificación internacional de enfermedades (CIE-10).
25	rpss_dias_reposo	Indica el número de días de reposo definidos en licencia médica. Es fundamental para analizar la extensión del tratamiento y tiempo de recuperación.
26	reingreso	Es una variable binaria categórica que indica si el caso es un reingreso o no.
27	stp_ctp	Indica si el accidente tuvo tiempo perdido o no tuvo tiempo perdido. Se extrae desde atdn_clasificacion_denuncia.
28	resolucion_accidente	Es el resultado de la resolución del accidente. Indica si la denuncia se aceptó por el seguro. Se extrae desde atdn_clasificacion_denuncia.
29	calificacion	Esta variable indica si el accidente es de trabajo, trayecto, o enfermedad profesional. Los registros “sin calificación” son principalmente casos rechazados. Se extrae desde atdn_clasificacion_denuncia.
30	caso_area_apertura	Define el área de atención hospitalaria
31	ispr_nombre_isapre	Nombre de ISAPRE del paciente (si es que aplica) o si pertenece a FONASA.
32	pcnt_nacionalidad	Establece la nacionalidad del paciente (C: Chilena; E: Extranjera)

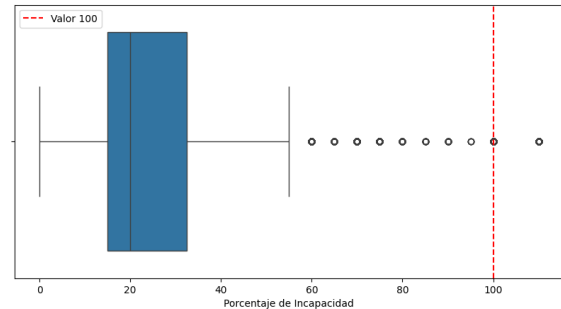
En la Tabla 3 se presentan las estadísticas descriptivas de las variables numéricas del conjunto de datos. Se evidencian algunas inconsistencias que sugieren la presencia de errores en los registros. Por ejemplo, en la variable *edad* se observan valores fuera del rango esperado, incluyendo edades negativas y un valor máximo de 1.066 años. Asimismo, en el *porcentaje de incapacidad* se identifican valores superiores al 100 %, y en los *días de reposo* se registran valores negativos, lo cual no resulta coherente con la naturaleza de la variable. Lo anterior también puede ser observado en la Figura 2.

Tabla 3: Estadísticas descriptivas de variables numéricas

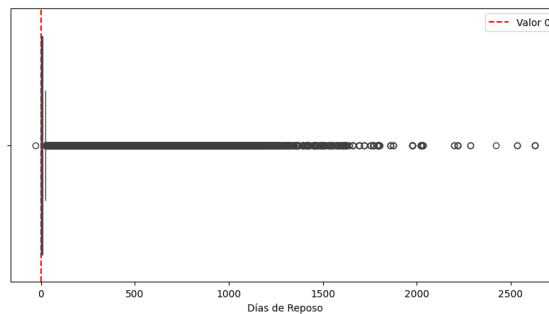
	edad	porcentaje_incapacidad	rpss_dias_reposo
promedio	38,99	27,91	15,78
desv estándar	13,05	21,61	57,10
min	-33	0	-29
25 %	28	15	0
50 %	38	20	2
75 %	49	32,5	10
max	1066	110	2627



(a) Distribución de la edad



(b) Porcentaje de incapacidad



(c) Días de reposo

Figura 2: Distribución de variables numéricas: edad, porcentaje de incapacidad y días de reposo.

En la Tabla 4 se observa que existen algunas variables categóricas que presentan un número elevado de clases. Por ejemplo, la variable *diag_cod_cie10* contiene 1.528 categorías distintas, mientras que *caso_profesion_ocupacion* registra 734. Esta alta cardinalidad puede generar problemas de dimensionalidad al momento de codificar dichas variables para su uso en modelos de

aprendizaje automático. Por tanto, será necesario aplicar técnicas de transformación que reduzcan la complejidad de estas variables, con el fin de evitar efectos negativos en el rendimiento del modelo.

Tabla 4: Resumen de variables categóricas

Variable	Clases	Moda	Freq	%
pcnt_genero	3	M	2.164.383	65,0 %
cmns_comuna	350	SANTIAGO	251.180	7,5 %
cmns_region	16	METROPOLITANA	1.771.210	53,2 %
cnat_centro_atencion	94	HOSPITAL SANTIAGO	805.190	24,2 %
caso_profesion_ocupacion	734	PERSONAS QUE REALIZAN TRABAJOS VARIOS	608.908	18,3 %
grave	2	0	3.066.462	92,1 %
adhr_macrosegmento	6	PYME	1.177.345	35,4 %
act_econ	20	CONSTRUCCIÓN	684.351	20,6 %
diag_cod_cie10	1528	S93.4	211.161	6,3 %
reingreso	2	No	3.049.923	91,7 %
stp_ctp	4	CTP	1.886.278	56,7 %
resolucion_accidente	3	Aceptado	2.356.807	70,8 %
calificacion	4	TB	1.716.278	51,6 %
pcnt_nacionalidad	2	C	3.201.848	96,2 %

5.1.2. Preparación de los datos.

El proceso de preparación de los datos contempló diversas etapas fundamentales para asegurar la calidad y pertinencia de la información utilizada en el modelado.

5.1.2.1. Limpieza del conjunto de datos. En primer lugar, se realizó una depuración del conjunto de datos, seleccionando exclusivamente los casos con calificación de accidente aceptada y excluyendo aquellos que correspondían a reingresos. A continuación, se eliminaron los registros con datos faltantes, con excepción de la variable porcentaje de incapacidad, cuyos valores ausentes fueron imputados con cero, bajo el supuesto de que la ausencia de este dato puede interpretarse como ausencia de incapacidad.

Asimismo, las variables que representaban fechas fueron transformadas al formato correspondiente, lo que permitió detectar inconsistencias cronológicas, como casos en los que la fecha de afectación del trabajador es posterior a la fecha de ingreso a la Mutual. Estos registros fueron eliminados por no cumplir con una secuencia temporal coherente.

Se identificaron casos en los que, a pesar de que el campo *relato_anamnesis* indica que se trata de un reingreso, la variable *reingreso* los clasifica erróneamente como "no reingreso". Estas inconsistencias fueron consideradas como errores de codificación y, por tanto, tales registros fueron excluidos del análisis.

Posteriormente, se aplicaron filtros adicionales para asegurar la validez de los valores en variables clave. Se conservaron únicamente los registros en los que la edad se encontraba en el intervalo [10–80], el porcentaje de incapacidad en el rango [0–100], y los días de reposo presentaban valores estrictamente mayores a cero.

5.1.2.2. Ingeniería de características. Se generaron nuevas variables derivadas de combinaciones o transformaciones de las originales, con el fin de enriquecer la representación de los casos. Las nuevas variables creadas son las siguientes:

1. **tiempo de espera:** diferencia entre las variables *atdn_fecha_admision* y *atdn_atd_fecha_ocurrencia* en horas.
2. **festivo:** variable binaria que indica si evento ocurre en día festivo.
3. **previo festivo:** variable binaria que indica si evento ocurre en día previo a festivo.
4. **posterior a festivo:** variable binaria que indica si evento ocurre en día posterior a festivo.
5. **día semana:** indica el día de la semana que ocurre evento.
6. **hora:** hora del día en que ocurre el evento.
7. **comuna:** la variable *cmms_comuna* fue transformada mediante una clasificación en tres categorías: rural, urbana y mixta, con el fin de facilitar su análisis dentro del modelo.
8. **CIE:** se construye desde la variable *diag_cod_cie10*. Esta nueva variable adopta valores en el rango [1–22], los cuales corresponden a los distintos capítulos de la clasificación CIE-10. En la Tabla 5 se presentan las categorías asociadas a cada valor.

Tabla 5: Categorías de variable CIE

Categoría	Glosa
1	Ciertas enfermedades infecciosas y parasitarias (A00-B99)
2	Neoplasias (C00-D49)
3	Enfermedades de la sangre y órganos hematopoyéticos y ciertos trastornos que afectan al mecanismo inmunológico (D50-D89)
4	Enfermedades endocrinas, nutricionales y metabólicas (E00-E89)
5	Trastornos mentales y de comportamiento (F01-F99)
6	Enfermedades del sistema nervioso (G00-G99)
7	Enfermedades del ojo y sus anexos (H00-H59)
8	Enfermedades del oído y de la apófisis mastoides (H60-H95)
9	Enfermedades del aparato circulatorio (I00-I99)
10	Enfermedades del aparato respiratorio (J00-J99)
11	Enfermedades del aparato digestivo (K00-K95)
12	Enfermedades de la piel y del tejido subcutáneo (L00-L99)
13	Enfermedades del aparato musculoesquelético y del tejido conectivo (M00-M99)
14	Enfermedades del aparato genitourinario (N00-N99)
15	Embarazo, parto y puerperio (O00-O9A)
16	Ciertas afecciones originadas en el período perinatal (P00-P96)
17	Malformaciones congénitas, deformidades y anomalías cromosómicas (Q00-Q99)
18	Síntomas, signos y resultados anormales de pruebas complementarias, no clasificados bajo otro concepto (R00-R99)
19	Lesiones traumáticas, envenenamientos y otras consecuencias de causas externas (S00-T88)
20	Causas externas de morbilidad (V00-Y99)
21	Factores que influyen en el estado de salud y contacto con los servicios sanitarios (Z00-Z99)
22	Códigos para situaciones especiales (U00-U99)

9. **CIUO**: se construye desde la variable *caso_profesion_ocupacion*. Esta nueva variable adopta valores en el rango [0–9], los cuales corresponden a los grandes grupos de la Clasificación Internacional Uniforme de Ocupaciones (CIUO). En la Tabla 6 se presentan las categorías asociadas a cada valor.

Tabla 6: Categorías de variable CIUO. Cada categoría representa uno de los grandes grupos de la Clasificación Internacional Uniforme de Ocupaciones.

Grandes grupos	Glosa
1	Directores, gerentes y administradores
2	Técnicos y profesionales de nivel medio
3	Técnicos y profesionales de nivel medio
4	Personal de apoyo administrativo
5	Trabajadores de los servicios y vendedores de comercios y mercados
6	Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros
7	Artesanos y operarios de oficios
8	Operadores de instalaciones, máquinas y ensambladores
9	Ocupaciones elementales
0	Ocupaciones de las Fuerzas Armadas

10. **Multidiagnóstico**: variable binaria que indica para una instancia que un paciente recibe diagnósticos adicionales al diagnóstico primario. Esta variable se construye desde las variables *caso_nom_diag_1_may_est*, *caso_nom_diag_2_may_est* y *caso_nom_diag_3_may_est*.

A sugerencia de contraparte de la Mutual, se incorporaron variables adicionales que no estaban presentes en el conjunto de datos original. Estas nuevas variables fueron entregadas con posterioridad a la recepción del conjunto inicial. A continuación, se detallan:

11. **cod_forma_oit**: representa las características del acontecimiento que dio lugar directamente a la lesión, es decir, la forma en que el objeto o sustancia involucrada entró en contacto con la persona afectada. Esta variable categórica puede tomar valores en el rango [1–9]. En la Tabla 7 se describen las distintas categorías que puede asumir esta variable.

Tabla 7: Categorías de la variable *cod_forma_oit*

Categoría	Glosa
1	Caídas de personas
2	Caídas de objetos
3	Pisadas sobre, choques contra, o golpes por objetos, a excepción de caídas de objetos
4	Atrapada por un objeto o entre objetos
5	Esfuerzos excesivos o falsos movimientos
6	Exposición a, o contacto con, temperaturas extremas
7	Exposición a, o contacto con, la corriente eléctrica
8	Exposición a, o contacto con, sustancias nocivas o radiaciones
9	Otras formas de accidente, no clasificadas bajo otros epígrafes, incluidos aquellos accidentes no especificados

12. ***cod_agente_oit***: representa el agente material relacionado con la lesión o con el accidente que la provocó. Esta variable categórica puede tomar valores en el rango [1–7]. En la Tabla 8 se describen las distintas categorías que puede asumir.

Tabla 8: Categorías de la variable *cod_agente_oit*

Categoría	Glosa
1	Máquinas
2	Medios de transporte y de manutención
3	Otros aparatos
4	Materiales, sustancias y radiaciones
5	Ambiente del trabajo
6	Animales
7	Agentes no clasificados por falta de datos suficientes

13. ***tipo_marca***: clasificación de denuncias según marca, esta pueden ser seguridad vial (SV), agresiones (MA) y seguridad industrial (SI).

Al considerar las nuevas variables, el conjunto de características a considerar en el estudio son las observadas en la Tabla 9.

Tabla 9: Listado de variables del conjunto de datos

nº	variable
1	CIE
2	CIUO
3	festivo
4	prev_festivo
5	post_festivo
6	mult_diag
7	urbano
8	rural
9	comuna_mixta
10	grave
11	tiempo_espera
12	dia_semana
13	hora
14	rslc_porcentaje_incapacidad
15	adhr_macrosegmento
16	act_econ
17	tipo_marca
18	edad
19	rpss_dias_reposo
20	fonasa
21	genero
22	nacionalidad
23	forma_oit
24	agente_oit

5.1.2.3. Variable objetivo La variable objetivo construida para el presente estudio permite identificar a los pacientes de larga estadía, entendidos como aquellos cuya duración del reposo médico excede lo esperable según su diagnóstico. Dado que la cantidad de días de reposo varía significativamente según la causa médica del accidente, se definió una regla específica por categoría de diagnóstico, utilizando la clasificación de la variable CIE. La variabilidad en los días de reposo por cada categoría CIE para accidentes del trabajo y trayecto se puede observar en las Figuras 3a y 3b respectivamente.

(a) Accidentes del Trabajo

(b) Accidentes del Trayecto

En particular, se consideró como paciente con ausencia prolongada aquel cuya cantidad de días de reposo supera el promedio de días de reposo de todos los pacientes que presentan la misma categoría *CIE*. Esta definición relativa permite capturar desviaciones dentro de cada grupo diagnóstico, ajustando así la identificación de casos prolongados al contexto clínico particular de cada tipo de lesión o enfermedad. Este criterio utilizado recoge además la indicación de la contraparte de Mutua, que sugería incorporar en la definición la variabilidad de los días perdidos por tipo de diagnóstico, de modo que la construcción de la variable refleje las particularidades de cada lesión o enfermedad. El promedio de cada categoría por tipo de accidente que se utilizó como umbral se observe en la Tabla 10.

27

Tabla 10: Días promedio de reposo por categoría *CIE* para accidentes de trabajo y trayecto.

CIE	Trabajo	Trayecto
0	5,08	6,75
1	5,12	2,00
2	25,00	–
3	19,00	–
4	44,71	56,86
5	46,51	50,55
6	178,66	293,56
7	8,03	44,68
8	24,12	34,71
9	60,11	54,45
10	46,31	71,16
11	14,95	37,88
12	11,60	75,88
13	35,34	72,16
14	20,75	12,90
15	3,00	7,00
16	10,40	11,15
17	94,50	80,00
18	5,15	3,95
19	20,96	25,94
20	200,86	200,00
21	11,66	9,80
22	6,22	–

Con el objetivo de asegurar la robustez del análisis y evitar sesgos derivados de tamaños muestrales reducidos, se decidió considerar únicamente aquellas categorías *CIE* que presentan al menos 100 observaciones en el conjunto de datos. Esta restricción permite obtener estimaciones más estables y representativas del comportamiento de los días de reposo dentro de cada categoría. Incluir categorías con muy pocos registros podría distorsionar los resultados, incrementar la varianza de las métricas de evaluación y dificultar la generalización de los hallazgos. Por tanto, el umbral de 100 casos se establece como un criterio mínimo para garantizar validez interna en los análisis posteriores. En las Tablas 11 y 12 se muestran las categorías que no superan el umbral y se excluyen del estudio para cada tipo de accidente.

Tabla 11: Categorías CIE cuyo número de casos de accidentes de trabajo no alcanza el umbral mínimo establecido.

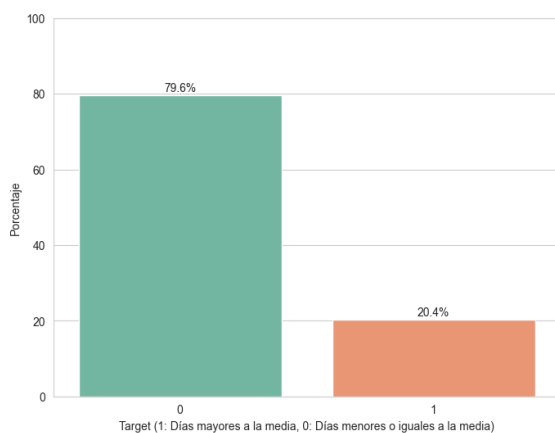
CIE	Casos
14	68
16	49
4	42
22	18
20	7
2	5
3	2
15	2
17	2

Tabla 12: Categorías CIE cuyo número de casos de accidentes de trayecto no alcanza el umbral mínimo establecido.)

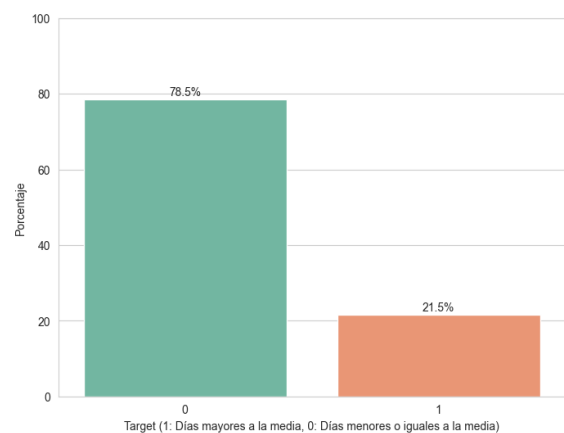
CIE	Casos
10	47
11	46
16	39
0	36
12	31
9	29
14	29
4	14
17	2
15	2
1	1
20	1

Para las variables categóricas seleccionadas, se implementó la creación de variables dummy, permitiendo su incorporación en algoritmos que requieren entradas numéricas.

Al observar las Figuras 4a y 4b, se evidencia una distribución desigual entre los casos positivos y negativos para la variable objetivo. Esta desproporción sugiere la presencia de un desbalance de clases, lo cual puede afectar negativamente el desempeño de los modelos predictivos. Por lo tanto, es necesario aplicar técnicas de procesamiento adecuadas con el fin de mitigar este problema y mejorar la capacidad del modelo para generalizar adecuadamente.



(a) Accidentes del Trabajo



(b) Accidentes del Trayecto

Figura 4: Distribución de variable objetivo por tipo de accidente.

Cada subconjunto fue dividido en un set de entrenamiento (70 %) y prueba (30 %), con el propósito de evaluar el desempeño de los modelos de forma robusta.

Finalmente, para mitigar el desequilibrio de clases, se aplicó la técnica SMOTE (Synthetic

Minority Over-sampling Technique), con el fin de generar ejemplos sintéticos de la clase minoritaria y así mejorar la capacidad predictiva del modelo sin inducir sesgo por subrepresentación.

5.1.3. Ingeniería de Modelos

Para cada modelo y para ambos tipos de accidentes (“Trabajo” y “Trayecto”) se realizó un proceso sistemático de optimización de hiperparámetros mediante Optuna, una biblioteca basada en búsqueda bayesiana y “pruning” de ensayos ineficientes. Esta estrategia exploró exhaustivamente el espacio definido para cada hiperparámetro en la Tabla 1 y, empleando validación cruzada, se identificó la combinación que maximizó el desempeño de cada modelo. La determinación de estos valores óptimos es crucial por dos razones: primero, eleva el rendimiento predictivo de cada modelo al adaptar su complejidad a las características específicas del conjunto de datos; segundo, proporciona un marco justo de comparación, ya que todos los algoritmos son evaluados en su mejor configuración posible en lugar de con parámetros arbitrarios o por defecto. De este modo, los resultados obtenidos permiten afirmar de forma objetiva que modelo ofrece la mayor capacidad explicativa y la mejor generalización para cada escenario, permitiendo de manera robusta seleccionar el modelo final. Los valores de hiperparámetros óptimos de cada modelo y tipo de accidente se observan en la Tabla 13.

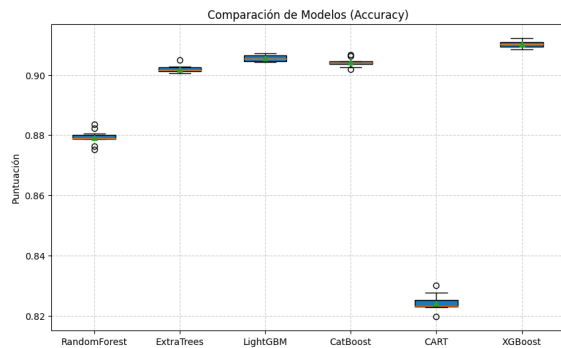
Tabla 13: Hiperparámetros óptimos determinados para cada modelo según tipo de accidente

Modelo	Hiperparámetro	Trabajo	Trayecto
Random Forest	n_estimators	45	49
	max_depth	29	30
	min_samples_split	5	9
	min_samples_leaf	1	1
Extra Trees	n_estimators	260	260
	max_depth	48	48
	min_samples_split	2	2
	min_samples_leaf	1	1
LightGBM	n_estimators	437	417
	max_depth	20	20
	learning_rate	0.0806	0.0769
	num_leaves	134	134
CatBoost	iterations	480	442
	depth	10	10
	learning_rate	0.1982	0.1512
XGBoost	n_estimators	457	480
	max_depth	15	15
	learning_rate	0.0777	0.0483
	subsample	0.9962	0.8914
	colsample_bytree	0.5019	0.5167
CART	max_depth	34	37
	min_samples_split	9	8
	min_samples_leaf	10	10
	criterion	gini	entropy

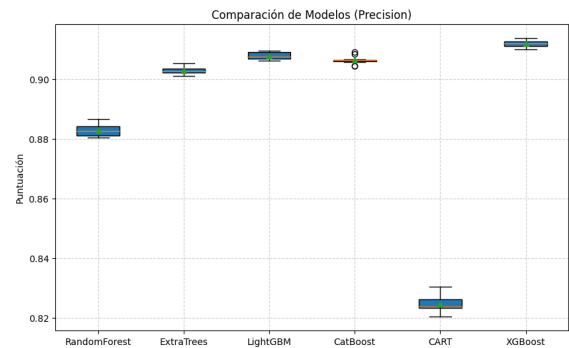
5.1.4. Evaluación de Modelos

En las Figuras 5 y 6 se observa el desempeño de los modelos ya optimizados —Random Forest, Extra Trees, LightGBM, CatBoost, CART y XGBoost en la tarea de clasificar a los pacientes con ausencia prolongada según el tipo de accidente. Conviene aclarar que, al tratarse de modelos supervisados, durante el proceso de entrenamiento se suministra tanto el conjunto de variables independientes como el valor conocido de la variable dependiente categórica (1=ausencia prolongada, 0= sin ausencia prolongada); es decir, el algoritmo “aprende” la relación entre ambas.

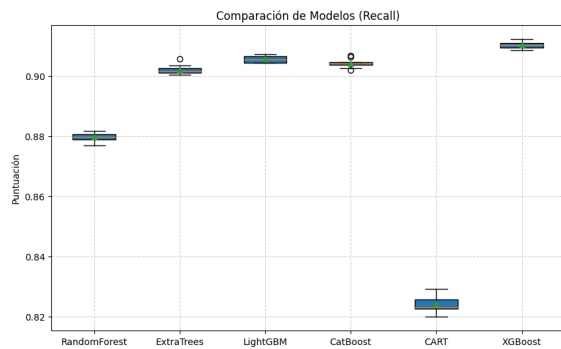
Para garantizar un análisis integral se emplearon métricas ampliamente aceptadas en tareas de clasificación binaria: accuracy, precision, recall y F1-score. Con este conjunto de métricas se busca no solo determinar cuál algoritmo ofrece el mayor rendimiento nominal, sino también comprender la solidez y la relevancia de sus predicciones en distintos tipos de accidentes.



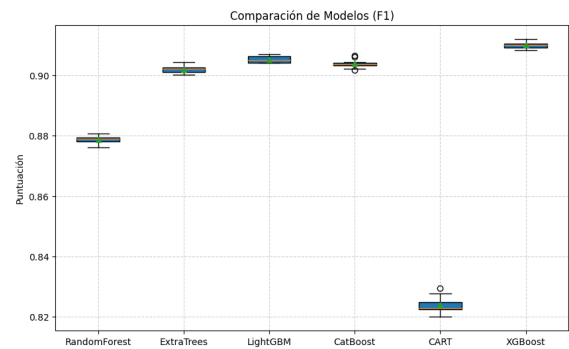
(a) accuracy



(b) precision

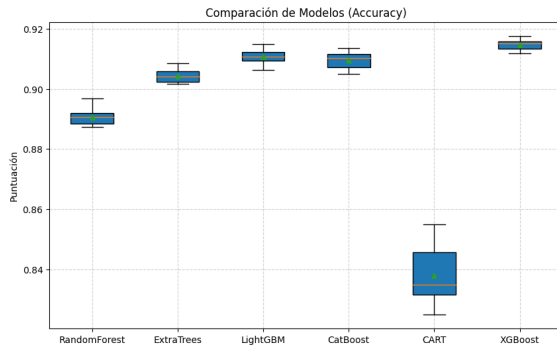


(c) recall

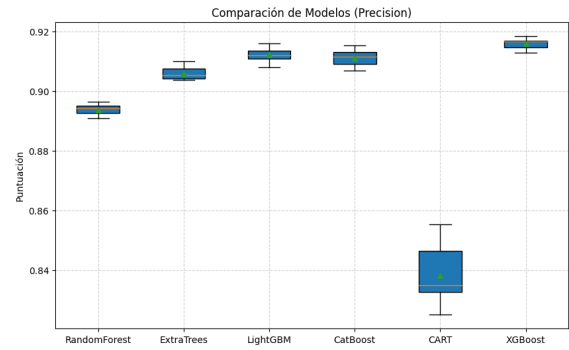


(d) F1

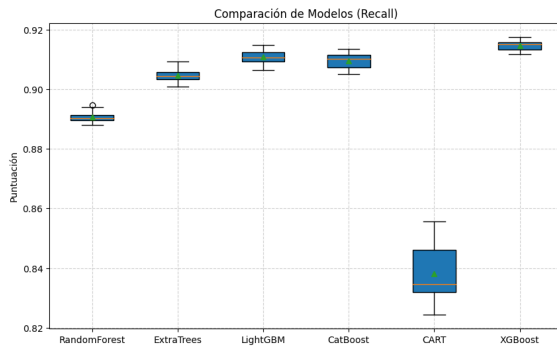
Figura 5: Cross Validation en conjunto de test para accidentes de trabajo.



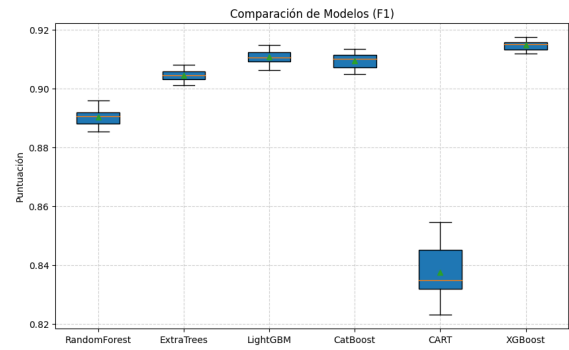
(a) accuracy



(b) precision



(c) recall



(d) F1

Figura 6: Cross Validation en conjunto de test para accidentes de trayecto.

El análisis comparativo revela que, para ambos tipos de accidentes y con independencia de la métrica utilizada (accuracy, precision, recall y F1 score), el modelo XGBoost alcanza de manera consistente el mejor desempeño, superando el 90 % en todos los indicadores. Estas evidencias sitúan a XGBoost como la opción más robusta para la clasificación de pacientes de ausencia prolongada y justifican su elección para la etapa siguiente del estudio, correspondiente a la interpretabilidad del modelo y la determinación de los factores que ejercen mayor influencia sobre sus predicciones.

5.1.5. Interpretabilidad

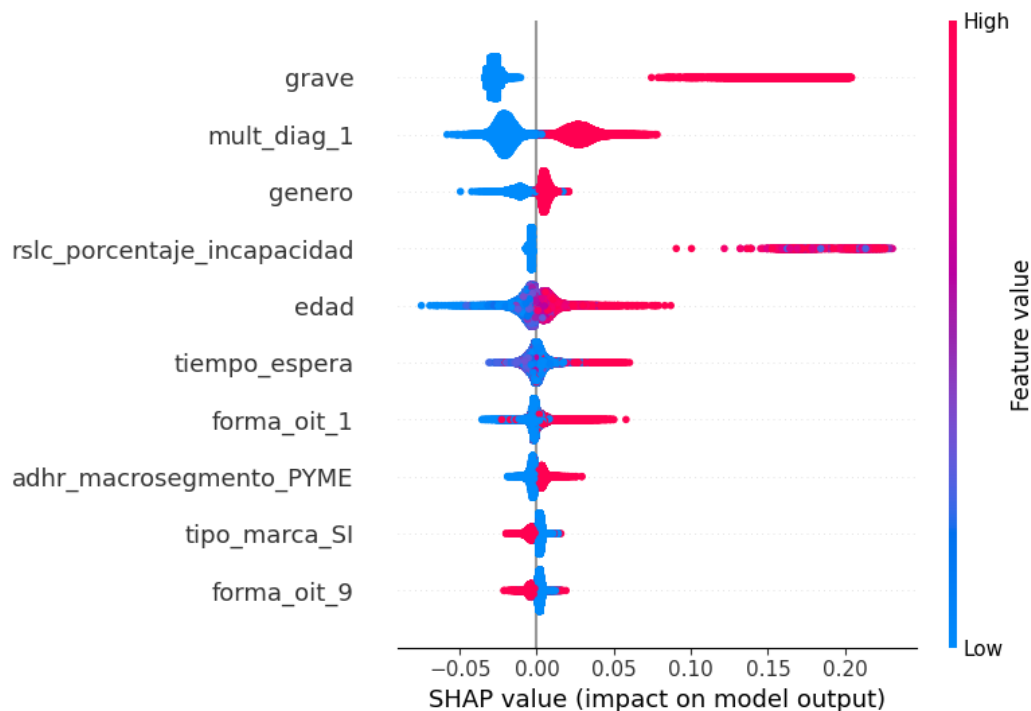
En esta sección abordamos la interpretabilidad del modelo de clasificación de pacientes con ausencia prolongada por accidentes de trabajo y de trayecto. Tras el proceso de benchmarking, el algoritmo XGBoost, ajustado con hiperparámetros óptimos, demostró el desempeño más alto y, por tanto, constituye la base del análisis de interpretabilidad. Para comprender el razonamiento subyacente a sus predicciones y ofrecer transparencia tanto a nivel agregado como individual, aplicamos el método SHAP (SHapley Additive exPlanations). Este enfoque, fundamentado en la teoría de juegos, cuantifica la contribución marginal de cada variable al resultado del modelo, permitiendo así una interpretación global, identificando las características que impulsan el comportamiento general del modelo, y una interpretación local, que explica de manera específica las decisiones asignadas a cada paciente. Para la interpretabilidad global del modelo se utilizan los valores SHAP, los cuales permiten entender la contribución de cada variable en las predicciones. Un impacto positivo en el modelo indica que el valor de esa variable tiende a aumentar la probabilidad de que se clasifique un caso como positivo (ausencia prolongada), mientras que un impacto negativo implica que su valor contribuye a disminuir dicha probabilidad. El orden en que aparecen las variables refleja la importancia de estas para el modelo. En la interpretabilidad local del modelo, se analiza el impacto de cada predictor en un caso individual. En ella se representan las aportaciones de cada variable a la predicción específica de ese registro. Es normal que la importancia y aun las variables destacadas difieran de las observadas en la interpretabilidad global, ya que esta última promedia los efectos absolutos de cada predictor sobre todo el conjunto de datos, resaltando los factores relevantes en términos generales. Por el contrario, la interpretación local considera únicamente cómo los valores concretos de ese caso desplazan la predicción respecto del valor base, lo que puede resaltar variables cuya influencia media es modesta, pero que resultan determinantes para el individuo analizado. El punto de partida es el valor base (promedio del modelo); a partir de él, las barras rojas (o hacia la derecha) muestran aportes positivos que empujan la predicción hacia el resultado de interés (ausencia prolongada), mientras que las barras azules (o hacia la izquierda) reflejan aportes negativos que la alejan de dicho resultado. La suma de todas las contribuciones, más el valor base, coincide exactamente con la predicción final para ese caso, permitiendo ver de manera transparente qué variables y en qué magnitud determinaron el resultado particular analizado. De este modo, los resultados de SHAP proporcionan evidencia clara a los profesionales de salud, facilitando la priorización y la asignación eficiente de recursos en función de los factores que más influyen en la permanencia prolongada de los pacientes.

5.1.5.1. Accidentes del Trabajo En la Figura 7 se presentan las diez variables con mayor peso en la clasificación de pacientes de larga estadía en accidentes de trabajo.

La más importante es la variable *grave* (1= grave, 0= no grave). Cuando el caso se registra como grave (puntos rojos), los valores SHAP se ubican notablemente hacia la derecha, incrementando la probabilidad de ausencia prolongada; por el contrario, los casos no graves (puntos azules) producen el efecto contrario. En segundo lugar, aparece *mult_diag 1*: la presencia de diagnósticos múltiples refuerza la probabilidad de una ausencia prolongada frente a cuadros simples. La tercera variable, *genero* (1 = masculino, 0 = femenino), muestra un sesgo claro, los pacientes masculinos aportan contribuciones positivas al riesgo de ausencia prolongada, mientras que los femeninos lo reducen. A continuación, valores elevados en *rslc_porcentaje_incapacidad* y una *edad* avanzada añaden peso positivo, lo que refleja mayor severidad clínica y menor capacidad de recuperación. La

variable tiempo espera refleja que demoras extensas (rojo) aumentan la probabilidad de ausencias prolongadas, mientras que esperas breves (azul) lo atenúan. Cuando la característica que dio lugar a la lesión es “Caídas de personas” (*forma_oit_1*=1) aumenta la probabilidad de ausencia prolongada. Un efecto similar se presenta cuando el trabajador pertenece a una PYME (*adhr_macrosegmento_PYME* =1) desplaza los valores SHAP hacia la derecha. Cuando una denuncia se marca distinta a seguridad industrial (*tipo_marca_SI*=0) aumenta la probabilidad de ser de larga estadía. Finalmente, los casos cuya característica que dio lugar a la lesión es clasificada como “otras formas de accidente” (*forma_oit_9*=1) disminuyen la probabilidad de ser de ausencia prolongada.

Figura 7: Contribución de las 10 variables más importantes en la predicción del modelo para clasificar pacientes de larga estadía en accidentes de trabajo.



La Figura 8a muestra la interpretabilidad local del modelo, para esto se analizan dos casos particulares, un caso que el modelo predice correctamente como de ausencia prolongada y un segundo caso que el modelo predijo exitosamente sin ausencia prolongada. En la Figura 8a se presenta el caso de un paciente que el modelo pronostica correctamente como ausencia prolongada. En ella, puede observarse que características individuales del paciente considera el modelo para realizar la predicción y cómo los valores de esta modifican la probabilidad base (0,2021) hasta alcanzar un valor de predicción de 0,53. El factor con mayor impacto positivo es el indicador de gravedad (“*grave* = 1”), que desplaza los valores SHAP hacia la derecha (área en rojo) y eleva sustancialmente la probabilidad de una estancia prolongada. Le siguen la presencia de múltiples diagnósticos (“*mult_diag_1* = 1”), la afiliación al sector Construcción (“*act_econ_CONSTRUCCION* = 1”), la edad avanzada (“*edad* = 53”), el horario de ocurrencia nocturno (“*hora_22* = 1”), característica que dio lugar a la lesión es clasificada como distinta a “otras formas de accidente” (“*forma_oit_9* = 0”) y empresa es distinta de una PYME (“*adhr_macrosegmento_PYME* = 0”); cada uno de estos atributos

aporta una contribución positiva adicional que refuerza la predicción de larga ausencia prolongada. Por el contrario, que la característica que dio lugar a la lesión es distinta a “Caídas de personas” (“*forma_oit_1* = 0”) muestra la única contribución en azul de mayor magnitud, reduciendo parcialmente la probabilidad de ausencia prolongada. La suma de todas estas contribuciones individuales permite explicar cómo el modelo aumenta la probabilidad desde el valor base del 20,21 % hasta aproximadamente el 53 %, clasificando correctamente al paciente con ausencia prolongada.



(a) Ejemplo de paciente de larga estadía correctamente clasificado



(b) Ejemplo de paciente de corta estadía correctamente clasificado

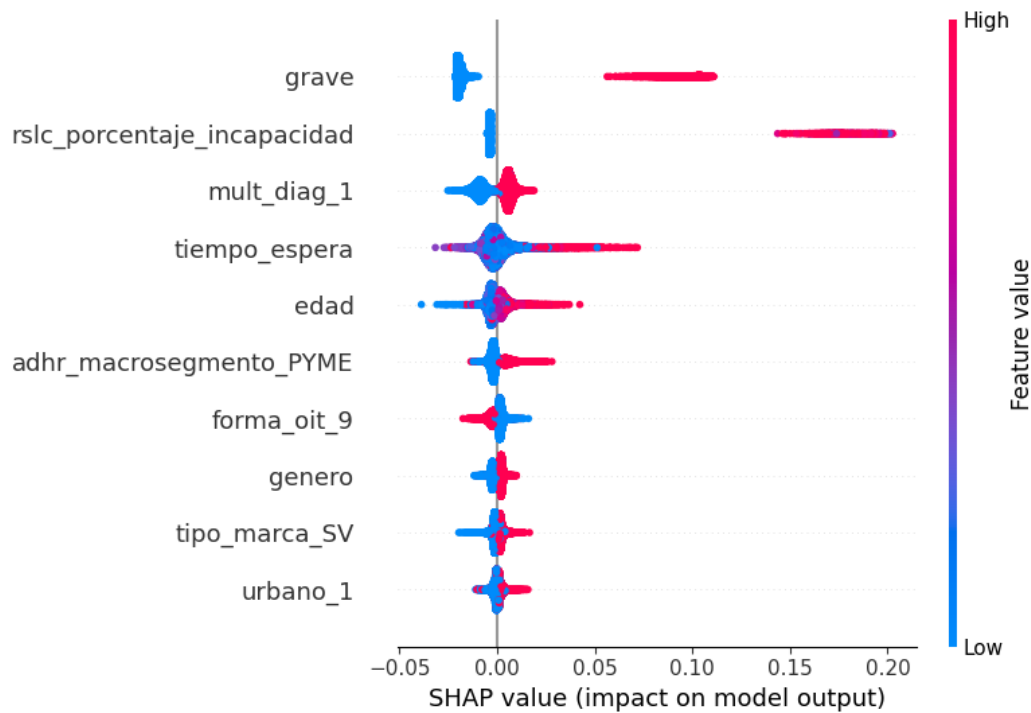
Figura 8: Interpretabilidad local del modelo en accidentes de trabajo.

La Figura 8b muestra el caso de un paciente de accidente trabajo que el modelo clasifica correctamente como sin ausencia prolongada. En ella, puede observarse cómo las características individuales del paciente modifican la probabilidad base (0,2021) hasta alcanzar un valor de predicción de 0,15. El primer conjunto de atributos con impacto positivo en la probabilidad de ausencia prolongada incluye presencia de un caso con múltiples diagnósticos (“*mult_diag_1* = 1”), que ocurre a las 13:00 hrs. (“*hora_13* = 1”) en un día viernes (“*día_semana_4* = 1”), cada uno de los cuales desplaza ligeramente los valores SHAP hacia la derecha (área en rojo) y tiende a incrementar la probabilidad de una ausencia prolongada. Sin embargo, este efecto positivo es superado por las contribuciones negativas (áreas en azul) de los siguientes factores: la ausencia de gravedad (“*grave* = 0”), la pertenencia al macrosegmento Grandes Contribuyentes (“*adhr_macrosegmento_GC* = 1”), ser del género femenino (“*genero* = 0”), trabajar en el sector de alojamiento y servicios de comidas (“*act_econ_ACTIVIDADES DE ALOJAMIENTO Y DE SERVICIO DE COMIDAS* = 1”), un tiempo de espera breve (“*tiempo_espera* = 0,1454”) y la inexistencia de porcentaje de incapacidad (“*rslc_porcentaje_incapacidad* = 0”). Estas características desplazan los valores SHAP hacia la izquierda, reduciendo la probabilidad de una ausencia prolongada desde el valor base de 0.2021 hasta aproximadamente el 0,15, lo que explica por qué el modelo clasifica correctamente a este paciente como de baja estadía o sin ausencia prolongada.

5.1.5.2. Accidentes del Trayecto La Figura 9 presenta la contribución global de las diez variables con mayor influencia en el modelo que predice ausencias prolongadas posteriores a accidentes de trayecto. La característica de mayor peso es la gravedad del caso (*grave*; 1 = grave, 0 = no grave): los eventos clasificados como graves (puntos rojos) desplazan la probabilidad estimada hacia valores más altos, mientras que los no graves (puntos azules) la reducen. Le sigue el porcentaje de incapacidad (*rslc_porcentaje_incapacidad*); un grado elevado de incapacidad incrementa la probabilidad de ausencia prolongada, en tanto que valores bajos o nulos la mitigan. La presencia de diagnósticos múltiples (*mult_diag_1*), los tiempos de espera extensos (*tiempo_espera*) y la

mayor edad del paciente también se asocian con una mayor propensión a ausentarse por periodos prolongados. Asimismo, los trabajadores pertenecientes a empresas clasificadas como PYME (*adhr_macrosegmento_PYME*) muestran una probabilidad superior de registrar ausencias de larga duración. En contraste, los casos cuya forma de accidente se cataloga como “otras formas” (*forma_oit_9* = 1) reducen la probabilidad de ausencia prolongada. El modelo también asigna relevancia al género: los pacientes de sexo masculino exhiben una mayor probabilidad de ausentarse por plazos extensos. Finalmente, las denuncias marcadas como seguridad vial (*tipo_marca_SV* = 1) y los accidentes ocurridos en zonas urbanas (*urbano_1* = 1) se asocian con probabilidades superiores de ausencia prolongada.

Figura 9: Contribución de las 10 variables más importantes en la predicción del modelo para clasificar pacientes de larga estadía en accidentes de trayecto.



La Figura 10a muestra cómo las características clínicas y contextuales de un paciente elevan la probabilidad base de ausencia prolongada (0,2135) hasta 0,52, lo que conduce al modelo a clasificarlo correctamente como caso de ausencia prolongada. El impulso más significativo corresponde al porcentaje de incapacidad (*rslc_porcentaje_incapacidad* = 60), cuyo marcado efecto positivo desplaza los valores SHAP hacia la derecha y domina la explicación de la predicción. Le siguen, en orden de relevancia, la gravedad del accidente (*grave* = 1), un tiempo de espera considerable, la edad de 58 años y la ocurrencia del evento en una zona urbana (*urbano_1* = 1), todas ellas reforzando la probabilidad de estancia prolongada. El único contrapeso apreciable proviene de la causa de la lesión atrapamiento por un objeto o entre objetos (*forma_oit_4* = 1), que atenúa parcialmente la probabilidad. En conjunto, estas variables incrementan la probabilidad hasta el 0,52 respaldando la decisión del modelo de clasificar al paciente como ausencia prolongada tras su accidente de trayecto.

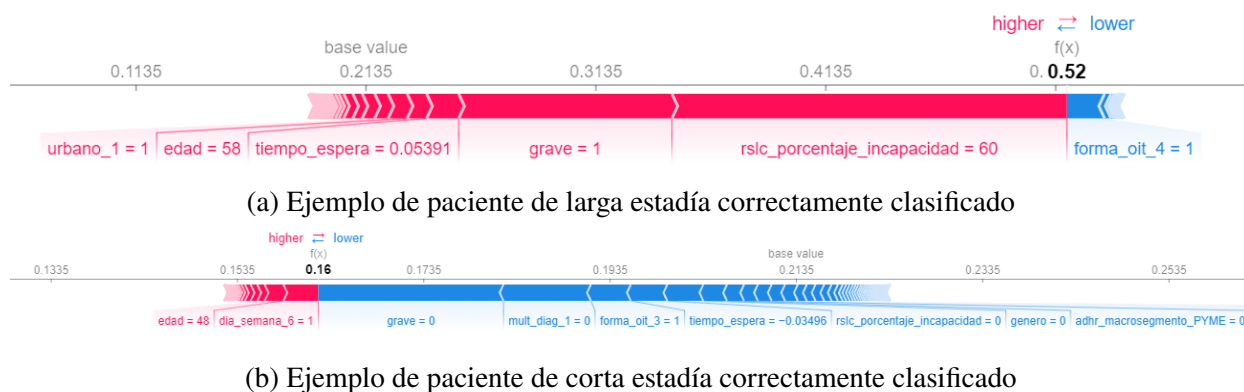


Figura 10: Interpretabilidad local del modelo en accidentes de trayecto.

La Figura 10b corresponde a un caso correctamente clasificado como sin ausencia prolongada, esto es, sin ausencia prolongada. El gráfico evidencia cómo las variables individuales ajustan la probabilidad base (0,2135) hasta el valor de predicción final (0,16). La edad del paciente (48 años) y el hecho de que el accidente ocurriera un domingo ($día_semana_6 = 1$) desplazan levemente los valores SHAP hacia la derecha, aumentando de forma marginal la probabilidad de ausencia prolongada. No obstante, dichos incrementos se ven ampliamente contrarrestados por los efectos negativos de la ausencia de gravedad y de diagnósticos múltiples ($grave = 0$; $mult_diag_1 = 0$), la causa del accidente es pisadas sobre, choques contra o golpes por objetos ($forma_oit_3 = 1$), un tiempo de espera breve ($tiempo_espera = -0,03494$), la falta de incapacidad ($rslc_porcentaje_incapacidad = 0$), el género femenino ($género = 0$) y la pertenencia a un segmento empresarial distinto de PYME ($adhr_macrosegmento_PYME = 0$). Estas variables desplazan los valores SHAP hacia la izquierda, reduciendo desde el valor base 0,2135 inicial hasta aproximadamente el 0,16 lo que explica la correcta clasificación del paciente como caso de baja ausencia prolongada.

5.2. Enfoque Econométrico

5.2.1. Evaluación de Modelos

Teniendo presente la medida de calidad de los modelos Logit y Probit, basada en el criterio Akaike (AIC), se infiere que ellos son bastante similares y los resultados no dependen de la elección de un modelo Logit o Probit. Se toma la decisión de continuar con el modelo Probit, dado que los errores estándar son calculados automáticamente y es más recomendable en muestras grandes, Tabla 14.

Tabla 14: Evaluación de calidad de los modelos según el criterio Akaike (AIC)

Modelo	AIC Logit	AIC Probit
Accidentes del Trabajo	0.807	0.807
Accidentes del Trayecto	0.895	0.897
Enfermedades Profesionales	0.614	0.616

5.2.2. Accidentes del Trabajo

5.2.2.1. Estadísticos del modelo accidentes del trabajo. Se ejecutó dos modelos econométricos con una primera ronda de entrenamiento (70 % de los datos) y otra de validación (30 % restante). Ambas rondas presentan una buena bondad del ajuste (Count $R^2 = 0.81$), con varianzas explicadas entre 44 % a 59 % en el entrenamiento y entre 45 % a 60 % en la validación, para los estadísticos Cox-Snell y Nagelkerke, respectivamente (ver Tabla 15).

Tabla 15: Estadísticos de bondad de ajuste para el modelo de ausencia prolongada (Accidente del Trabajo)

Estadístico	Training	Validación
Número de observaciones	8 332	3 588
LR χ^2 (9)	4 817.815	2 129.300
Pr > χ^2	0.000	0.000
Count R^2	0.807	0.808
Sensibilidad	74.1 %	74.1 %
Especificidad	86.6 %	87.4 %
Clasificados correctamente	80.7 %	80.8 %
Cox-Snell R^2	0.44	0.45
Nagelkerke R^2	0.59	0.60

5.2.2.2. Interpretabilidad Los efectos marginales promedios de modelo de entrenamiento se indican en la Tabla 14. Como se observa, las variables que incrementan la probabilidad de tener una ausencia prolongada son ser hombre (11.6 %), tener un accidente tipo caída (8.9 %), presentar más de un diagnóstico (35 %), y que el accidente sea de gravedad (50 %). Por otro lado, las variables que disminuyen la probabilidad de ausencia prolongada son: tener edad menor de 45 años (-8.4 %), ser artesano u operario de oficio (-5.7 %), trabajar con materiales, sustancias y radiaciones (-26.0 %), tener marca seguridad industrial (-17.7 %), y ser evaluado con un porcentaje de incapacidad menor por el accidente (-59.5 %).

Tabla 16: Efectos Marginales Promedio (EMP) de la Ausencia Prolongada en los Accidentes del Trabajo (70 % de los datos, Entrenamiento)

Variable	Valor	Predic (dy/dx) (a,b)	[95 % Conf. Interv.]
Hombre Mujer (base)	Si = 1, No = 0 —	0.116*** —	[0.077; 0.154] —
Tramo edad: Edad <45 años Edad ≥ 45 (base)	Si = 1, No = 0 —	-0.084*** —	[-0.111; -0.057] —
Caída de persona Otra forma OIT (base)	Si = 1, No = 0 —	0.089*** —	[0.055; 0.123] —
Artesanos y operarios de oficio Otra CIUO (base)	Si = 1, No = 0 —	-0.057** —	[-0.097; -0.018] —
Materiales, sustancias y radiaciones Otros Agentes OIT (base)	Si = 1, No = 0 —	-0.260*** —	[-0.348; -0.172] —
Marca: Seguridad Industrial Otro tipo marca (base)	Si = 1, No = 0 —	-0.177*** —	[-0.207; -0.147] —
Multidiagnóstico	Si = 1, No = 0	0.350***	[0.324; 0.376]
Gravedad	Si = 1, No = 0	0.501***	[0.470; 0.524]
Porcentaje de incapacidad (0–14 %)	Si = 1, No = 0	-0.595***	[-0.609; -0.581]

^a dy/dx representa el efecto marginal, que es la tasa de cambio de la variable dependiente (y) con respecto a la variable independiente (x).

^b p-valor: ***p<0.001; **p<0.01; *p<0.05.

Los efectos marginales promedios del modelo de validación confirman los resultados anteriores en la misma dirección y similares magnitudes (ver Tabla 17). Las variables que incrementan la probabilidad de ausencia prolongada son ser hombre (11.8 %), tener un accidente tipo caída (2.2 %), presentar más de un diagnóstico (37.1 %), y que el accidente sea de gravedad (50 %), en tanto que las variables que disminuyen la probabilidad son: tener edad menor de 45 años (-12.4 %), ser artesano u operario de oficio (-5.7 %), trabajar con materiales, sustancias y radiaciones (-14.0 %), tener marca seguridad industrial (-14.2 %), y ser evaluado con un porcentaje de incapacidad menor por el accidente (-54.9 %).

Tabla 17: Efectos Marginales Promedio (EMP) de la Ausencia Prolongada en los Accidentes del Trabajo (70 % de los datos, Entrenamiento)

Variable	Valor	Predic (dy/dx) (a,b)	[95 % Conf. Interv.]
Hombre	Si = 1, No = 0	0.118***	[0.060; 0.177]
Mujer (base)	–	–	–
Tramo edad: Edad < 45 años	Si = 1, No = 0	–0,124***	[–0,166; –0,083]
Edad ≥ 45 (base)	–	–	–
Caída de persona	Si = 1, No = 0	0.022	[–0,030; 0.076]
Otra forma OIT (base)	–	–	–
Artesanos y operarios de oficio	Si = 1, No = 0	–0,057**	[–0,097; –0,018]
Otra CIUO (base)	–	–	–
Materiales, sustancias y radiaciones	Si = 1, No = 0	–0,140*	[–0,278; –0,003]
Otros Agentes OIT (base)	–	–	–
Marca: Seguridad Industrial	Si = 1, No = 0	–0,142***	[–0,188; –0,096]
Otro tipo marca (base)	–	–	–
Multidiagnóstico	Si = 1, No = 0	0.371***	[0.333; 0.411]
Gravedad	Si = 1, No = 0	0.504***	[0.473; 0.537]
Porcentaje de incapacidad (0–14 %)	Si = 1, No = 0	–0,549***	[–0,575; –0,524]

^a dy/dx representa el efecto marginal, que es la tasa de cambio de la variable dependiente (y) con respecto a la variable independiente (x).

^b p-valor: *** p<0.001; ** p<0.01; * p<0.05.

5.2.3. Accidentes del Trayecto

5.2.3.1. Estadísticos del modelo accidentes del trayecto Igualmente, se ejecutó dos modelos econométricos. Una primera ronda de entrenamiento con 70 % de los datos y otra de validación con el 30 % restante. La primera obtuvo un Count R2 = 0.82 y la segunda de 0.83, con varianzas explicadas entre 39 % a 51 % (entrenamiento) y entre 41 % a 54 % (validación), para los estadísticos Cox-Snell y Nagelkerke, respectivamente (ver Tabla 18).

Tabla 18: Estadísticos de bondad de ajuste para el modelo de ausencia prolongada (Accidente del Trayecto)

Estadístico	Training	Validación
Número de observaciones	7 657	3 283
LR χ^2 (9)	3 727.407	1 710.520
Pr > χ^2	0.000	0.000
Count R^2	0.823	0.827
Sensibilidad	81.9 %	82.3 %
Especificidad	82.6 %	83.1 %
Clasificados correctamente	82.3 %	82.7 %
Cox–Snell R^2	0.385	0.406
Nagelkerke R^2	0.513	0.541

5.2.3.2. Interpretabilidad del modelo accidentes del trayecto En la Tabla 19 se muestra los efectos marginales promedios del modelo de entrenamiento, donde se observa que las variables que incrementan la probabilidad de tener una ausencia prolongada son: sexo hombre (8.6 %), presentar más de un diagnósticos (16.3 %), y que el accidente sea de gravedad (64 %); en tanto que las variables que disminuyen la probabilidad de ausencia prolongada son: tener edad menor de 45 años (-9.0 %), tener una lesión traumática, envenenamiento y otras (-51.3 %), ser consecuencia de causa externa de movilidad (-45.9 %), resultar como consecuencia de una agresión (-14.6 %), ocurrir el día lunes (-4.4 %) y martes (-5.8 %).

Tabla 19: Efectos Marginales Promedio (EMP) de la Ausencia Prolongada en los Accidentes del Trayecto (70 % de los datos, Entrenamiento)

Variable	Valor	Predic (dy/dx) (a,b)	[95 % Conf. Interv.]
Hombre	Si = 1, No = 0	0.086***	[0.056; 0.116]
Mujer (base)	—	—	—
Tramo edad: Edad < 45 años	Si = 1, No = 0	-0,090***	[-0,118; -0,062]
Edad ≥ 45 (base)	—	—	—
Lesiones traumáticas, envenenamientos y otras consecuencias de causas externas	Si = 1, No = 0	-0,513***	[-0,532; -0,493]
Otra forma OIT (base)	—	—	—
Causas externas de morbilidad	Si = 1, No = 0	-0,459***	[-0,485; -0,433]
Otra forma OIT (base)	—	—	—
Marca: Agresiones	Si = 1, No = 0	-0,146***	[-0,206; -0,089]
Otro tipo marca (base)	—	—	—
Lunes	Si = 1, No = 0	-0,044**	[-0,079; -0,008]
Otro día (base)	—	—	—
Martes	Si = 1, No = 0	-0,058***	[-0,094; -0,023]
Otro día (base)	—	—	—
Multidiagnóstico	Si = 1, No = 0	0.163***	[0.132; 0.195]
Gravedad	Si = 1, No = 0	0.638***	[0.620; 0.657]

^a dy/dx representa el efecto marginal, que es la tasa de cambio de la variable dependiente (y) con respecto a la variable independiente (x).

^b p-valor: *** p<0.001; ** p<0.01; * p<0.05.

Los efectos marginales promedios del modelo de validación (ver Tabla 20) arrojan la misma dirección y con similares magnitudes que el modelo de entrenamiento. Las variables que incrementan la probabilidad de ausencia prolongada son ser hombre (13.3 %), presentar más de un diagnóstico (17.6 %), y la gravedad o severidad del accidente (65.7 %), en tanto que las que disminuyen la probabilidad son tener edad menor de 45 años (-11.1 %), las lesiones traumáticas (-51.5 %), por otras causas externas (-46.3 %), ser resultado de una agresión (-14.3 %).

Tabla 20: Efectos Marginales Promedio (EMP) de la Ausencia Prolongada en los Accidentes del Trayecto (30 % de los datos, Validación)

Variable	Valor	Predic (dy/dx) (a,b)	[95 % Conf. Interv.]
Hombre	Si = 1, No = 0	0.133***	[0.087; 0.179]
Mujer (base)	—	—	—
Tramo edad: Edad < 45 años	Si = 1, No = 0	-0,111***	[-0,156; -0,067]
Edad ≥ 45 (base)	—	—	—
Lesiones traumáticas, envenenamientos y otras consecuencias de causas externas	Si = 1, No = 0	-0,515***	[-0,539; -0,491]
Otra forma OIT (base)	—	—	—
Causas externas de morbilidad	Si = 1, No = 0	-0,463***	[-0,504; -0,423]
Otra forma OIT (base)	—	—	—
Marca: Agresiones	Si = 1, No = 0	-0,143***	[-0,232; -0,054]
Otro tipo marca (base)	—	—	—
Lunes	Si = 1, No = 0	-0,010	[-0,065; 0.043]
Otro día (base)	—	—	—
Martes	Si = 1, No = 0	0.007	[-0,050; 0.066]
Otro día (base)	—	—	—
Multidiagnóstico	Si = 1, No = 0	0.176***	[0.129; 0.223]
Gravedad	Si = 1, No = 0	0.657**	[0.629; 0.685]

^a dy/dx representa el efecto marginal, que es la tasa de cambio de la variable dependiente (y) con respecto a la variable independiente (x).

^b p-valor: *** p < 0.001; ** p < 0.01; * p < 0.05.

6. DISCUSIÓN DE RESULTADOS

6.1. Discusión de los resultados para los Accidentes del Trabajo

Consistente con hallazgos previos, aquellos accidentes que implican mayor tiempo de recuperación ocurren principalmente en los hombres. Por ejemplo, en el trabajo de Fontaneda y colegas (2019), cuando se selecciona el subconjunto de accidentes más severos, los hombres son quienes resultan más afectados. En esta misma línea, las tasas de accidentes con muertes son significativamente mayores entre los hombres debido a múltiples variables, tales como la mayor toma de riesgo por parte de ellos, los distintos puestos de trabajos ejecutados por ellos, el mayor consumo de alcohol y más las actividades recreativas que afectan sus descansos, entre otras (Eevers & Gee, 1986).

En la revisión sistemática de Alvear y colegas (2025) se observa que a mayor edad del trabajador accidentado hay más días necesarios para su recuperación, información que resulta consistente con nuestro hallazgo de que los trabajadores menores de 45 años tienen menos probabilidad de una ausencia prolongada. En este sentido, el envejecimiento de la población trabajadora pasa a ser un tema relevante para la salud ocupacional, por ejemplo, la edad promedio de los trabajadores

españoles accidentados ha pasado de 35 años en 1996 a 41 años para el 2015 (López et al., 2018), en tanto que la edad promedio de los trabajadores en Chile ha pasado desde los 41,3 años en 2010 a los 43,8 años en 2018 (CIEDES, 2024). Entre las causas de la mayor severidad de los accidentes en los trabajadores mayores, y con ello, de mayores tiempos necesarios para la recuperación, podemos considerar la fragilidad, las condiciones de salud preexistentes en ellos (Boufous & Williamson, 2009), variable que también ha aparecido como resultado en este trabajo. Además, los trabajadores de mayor edad experimentan tiempos de reacción más lentos, lo cual finalmente podría repercutir en la mayor severidad de los accidentes que sufren (Doroudgar et al., 2017). Respecto a la relación de la gravedad y frecuencia de accidentes con las condiciones de salud preexistentes de los trabajadores mayores, se han relacionado los accidentes más severos con condiciones de salud preexistentes y los niveles de fatiga de los afectados (Rey-Merchan & Lopez-Arquillos, 2021; López et al., 2018), así como también hay mayor frecuencia de accidentes laborales cuando hay antecedentes médicos familiares (diabetes) o el trabajador tiende a ser más sedentario (Pietilä et al., 2018).

Cabe hacer presente que las dos variables anteriores de riesgo (esto es, ser hombre y ser trabajador de edad mayor a 45 años) se refuerzan debido a la legislación de nuestro país, donde las mujeres jubilan a los 60 años. Los hombres permanecen más tiempos expuestos a este doble riesgo, producto de sus trayectorias laborales más extensas por su jubilación a partir de los 65 años.

En cuanto a las variables del diagnóstico, tales como la gravedad del accidente y el porcentaje de incapacidad menor, asociados con mayor y menor probabilidad de un accidente de larga recuperación, respectivamente, existe evidencia consistente que relaciona severidad o gravedad con la duración del tratamiento hospitalario y la duración del tiempo de recuperación final. Así, por ejemplo, en España se clasifica la gravedad de los accidentes mediante la “escala española”, que es una escala de gravedad de los accidentes, que son evaluadas por médicos, siguiendo los siguientes criterios de clasificación: Leve (sin implicancia para el trabajador), severo (con implicancias, pero recuperable), muy severo (puede causar invalidez o incapacidad parcial) y fatales (un 96 % de estos casos implican hasta dos días de ausentismo). Cabe hacer notar que la gran mayoría de los casos son leves (98 %). De este modo, si consideramos aquellos casos evaluados como severos, ellos muestran significativamente más días perdidos. (Rey-Merchan y López-Arquillos, 2021; López y colegas, 2018). Igualmente, se ha observado que aquellos trabajadores accidentados que requieren más días de hospitalización producto de la gravedad de sus lesiones requerirán mayor tiempo de atención total para su recuperación (Gray & Collie, 2019).

Asimismo, la evidencia previa indica que los tipos de accidentes que implican mayor severidad están asociados con caídas, tropiezos y resbalones (Baraza et al., 2023; Baraza & Cugueró-Escofet, 2021; Berglund et al., 2021; Fontaneda et al., 2022; Fuentes-Bargues et al., 2022; González Fuentes et al., 2022; Höyükler et al., 2022). Esta información es consistente utilizando las técnicas de inteligencia artificial que encuentran entre las primeras siete palabras relacionadas con tiempos de hospitalización “atrapó” (segunda), “resbaló” (tercera), “cayó” (sexta), y máquina (séptima) (Khairuddin et al., 2023). Nuestros resultados también apuntan a que las caídas se relacionan con un 9 % más de probabilidad de terminar en un accidente con ausencia prolongada.

Finalmente, la última variable, extraída del método de inteligencia artificial, indica que mientras mayor tiempo transcurra entre el accidente y la atención médica, mayor será el tiempo de ausencia prolongada. Por un lado, es posible inferir una relación entre la mayor gravedad del accidente y la necesidad de traslado profesional para el trabajador accidentado hasta el centro asistencial respectivo, lo que requiere espera en vez que el trabajador llegue por sus medios o en otro tipo de transporte. Del mismo modo, el trabajador accidentado de gravedad podría ser atendido en

una primera instancia en otro centro y luego derivado al centro del organismo administrador. En cualquier caso, la evidencia previa ha establecido esta relación, al observar que mientras mayor el tiempo que demore desde el accidente a la atención del paciente, así como más tiempo demore la calificación del evento como un accidente ocupacional será mayor el tiempo de recuperación que requerirá el trabajador (Besen et al., 2018; Bharti et al., 2019; Iacobucci, 2021). Asimismo, como la mayor frecuencia de accidentes más graves ocurren por las noches y los fines de semana, es posible que se deba a que hay menos personal en los turnos para apoyar al afectado y entregar una asistencia oportuna.

6.2. Discusión de los resultados para los Accidentes del Trayecto

Tal como fue revisado en el caso de los accidentes del trabajo, las variables encontradas y la discusión se mantiene para los accidentes del trayecto. Esto es, hay relaciones entre la ausencia prolongada con la mayor gravedad de la lesión, que es la variable que más se relaciona con la variable resultado (64 %). Asimismo, contribuye a la ausencia prolongada por accidente del trayecto el ser hombre, tener multidiagnóstico, y ser mayor de 45 años.

Además, se observa que disminuye la probabilidad si el accidente corresponde a una lesión traumática o consecuencia de causa externa a la movilidad, y ser resultado de una agresión.

Finalmente, nuestros resultados arrojan que es menos probable que los accidentes de trayecto con ausencia prolongada ocurran un lunes o martes, lo cual es completamente consistente con la literatura previa. Se ha observado que los accidentes con mayor severidad (y por tanto con mayor tiempo perdido) tienden a ocurrir en mayor proporción desde los jueves a domingo (Baraza & Cugueró-Escofet, 2021), los sábados y domingos (Fontaneda et al., 2022; Fuentes-Bargues et al., 2022), los viernes, sábados y jueves (Ahn, 2022), o bien los viernes (Baraza et al., 2023). Consistentemente, investigaciones reportan que los días perdidos promedio por accidente tienden a ser menores los lunes (22,8 días) comparados con los domingos (29,0) (López et al., 2018).

7. Recomendaciones para Sistema de Seguridad y Salud en el Trabajo

De los resultados obtenidos se plantean las siguientes recomendaciones a Mutual de Seguridad.

- Revisar los protocolos administrativos de atención a pacientes, en particular aquellos relacionados con pacientes considerados graves y con más de un diagnóstico. Lo anterior, con el objetivo de disminuir los tiempos de espera y evitar de esta forma la ausencia prolongada.
- Reforzar las políticas de supervisión en el sector de la construcción, en particular factores presentes en el ambiente de trabajo y características físicas y fisiológicas de los trabajadores, que puedan generar accidentes tipo caídas, con ausencia prolongada.
- Supervisar la realización de capacitación laboral sobre contenidos en seguridad para el trabajo por parte de los empleadores, en particular en la industria de la construcción.
- Realizar un estudio que identifique los factores sociales y de salud que inciden en los accidentes del trabajo y trayecto de las personas mayores, cuyos resultados promuevan en las empresas condiciones laborales especiales para dicho grupo etario.

- Solicitar a los empleadores planes de prevención específicos para el colectivo de trabajadores considerados mayores, basadas en la formación y la identificación continua de los riesgos de accidentabilidad, que cambian según las circunstancias de cada industria y contextos culturales y laborales.
- Sensibilizar mediante capacitación y talleres a encargados de seguridad ocupacional y a los propios trabajadores sobre la asociación entre fatiga, somnolencia y la severidad de los accidentes laborales. Hemos revisado en este trabajo que los accidentes que requieren mayor tiempo de recuperación (mayor gravedad) tienden a ocurrir al final de la semana (jueves a sábados) y en los horarios de madrugada, noche y al finalizar los turnos de trabajo.
- Reforzar políticas de capacitación en relación con sensibilizar a los trabajadores de sexo masculino que se encuentran más expuestos a presentar accidentes del trabajo y del trayecto que implican una ausencia prolongada después del evento. Igualmente, continuar con investigaciones que busquen explicar los factores asociados con estas consecuencias adversas para la salud y seguridad ocupacional de ellos.
- En nuestro trabajo hemos encontrado que hay una asociación entre accidentes del trayecto y mayor tiempo de recuperación (mayor severidad) en zonas urbanas. Se requiere explorar los factores asociados con esta relación, por ejemplo, medio de transporte más prevalente y accidente (tipo de vehículo, transporte público, traslado a pie, etc.), efectos de la seguridad vial (intersecciones, letreros, convivencia vial), mayor presencia de ciclistas y peatones, entre otros.
- Solicitar a los empleadores de empresas PYMES fortalecer los planes de formación a trabajadores en prevención de accidentes, tendientes a tomar conciencia de la protección personal y respeto a las reglas establecidas en los ambientes de trabajo.
- Implementar un proyecto piloto que permita aplicar en un grupo controlado de equipo sanitario las variables de predicción de ausencia prolongada generadas en el presente estudio.

8. Conclusiones

En primer lugar, es importante destacar que el objetivo planteado en este trabajo ha sido cumplido satisfactoriamente. La metodología utilizada permitió abordar adecuadamente el problema planteado, lo que ha facilitado la obtención de resultados sólidos y conclusiones pertinentes.

Luego, el presente trabajo permite concluir que existen características propias del diagnóstico de los pacientes que permiten clasificarlos correctamente con ausencia prolongada, tanto en accidentes del trabajo, como del trayecto. Estás son: un paciente con un alto indicador de gravedad, con presencia de más de un diagnóstico y con mayor porcentaje de incapacidad.

En línea con el objetivo de nuestro trabajo, se infiere que los factores adicionales al diagnóstico con mayor peso en la clasificación de pacientes con ausencia prolongada son:

- a) **Accidentes del trabajo:** Las variables que predicen la ausencia prolongada son: género masculino, edad avanzada, ha presentado un tiempo de espera prolongado para ser atendido, *forma_oit_1* (“caídas de personas”), con desempeño laboral en macrosegmentos PYMES

y denuncia distinta a seguridad industrial (*tipo_marca_SI*=0). Por otro lado, los casos cuya característica que dio lugar a la lesión es clasificada como “otras formas de accidente” (*forma_oit_9*=1) disminuyen la probabilidad de ser de ausencia prolongada. El modelo econométrico, adicionalmente, muestra que los trabajadores que desempeñan el oficio de “Artesano y Operarios de Oficios”, tienen menor probabilidad de tener ausencia prolongada.

- b) **Accidentes del trayecto:** Las variables que predicen la ausencia prolongada son: el incremento de tiempos de espera para ser atendido, presentar edad avanzada, desempeño laboral en macrosegmentos PYMES, pacientes de sexo masculino, denuncias marcadas como seguridad vial (*tipo_marca_SV* = 1) y accidentes ocurridos en zonas urbanas (*urbano_1* = 1). En contraste, los casos cuya forma de accidente se cataloga como “otras formas” (*forma_oit_9* = 1) reducen la probabilidad de ausencia prolongada. Adicionalmente, el modelo econométrico muestra que la ausencia prolongada disminuye si el accidente ocurre los lunes o martes, en relación con otros días de la semana.

Los resultados del estudio proporcionan evidencia clara a MUTUAL para implementar políticas adecuadas, que podrían evitar que los accidentes laborales o del trayecto generen ausencia prolongada, facilitando la priorización y la asignación eficiente de recursos.

Es destacable la variable edad avanzada de los pacientes, en tal sentido, el envejecimiento de la población trabajadora pasa a ser un tema relevante para la salud ocupacional, teniendo presente el perfil demográfico de la población chilena. En este trabajo hemos observado que aumentan los tiempos de recuperación cuando el trabajador/a tiene una edad mayor a 45 años, lo cual es un dato relevante si consideramos que la edad promedio de los trabajadores en Chile ha pasado desde los 41,3 años en 2010 a los 43,8 años en 2018 (CIEDES, 2024).

La evidencia también muestra que los hombres, sumado a edad avanzada, podrían presentar mayor ausencia prolongada después de un accidente del trabajo, lo anterior, en línea con la mayor edad de jubilación en relación con las mujeres (65 años).

Es relevante la presencia de ausencia prolongada en el macrosegmento PYMES. Dicho sector presenta el mayor porcentaje de empleabilidad a nivel país, en tanto, el impacto no sólo es para el trabajador, también para las empresas que conforman dicho sector, en particular, el aumento de los costos de producción y la consecuente pérdida de competitividad.

Para las enfermedades profesionales, ambas metodologías no permiten realizar estimaciones estadísticamente significativas.

9. Referencias

- Ahn, S. (2022). Derivation of Risk Factors to Quantify the Risk of Safety Accidents for Small and Medium-Sized Enterprises in Construction Industry. *Sustainability (Switzerland)*, 14(12). <https://doi.org/10.3390/su14127306>
- Alacevich, C., & Nicodemo, C. (2024). The Effect of Immigration on Occupational Injuries: Evidence from Administrative Data*. *Oxford Bulletin of Economics and Statistics*, 86(2), 209–235. <https://doi.org/10.1111/obes.12572>
- Alvear-Vega, S., Vargas-Garrido, H., Núñez-Delafuente, H., & Vidal-Silva, C. (2025, en publicación). Factores sociales del tiempo de recuperación de un trabajador, después de un accidente laboral: Una revisión sistemática. *Rev Med Chile* 2025; 153(6).
- Arocena, P., & García-Carrizosa, L. M. (2023). Determinants of the duration of sick leave due to occupational injuries: Evidence from Spanish manufacturing. *Journal of Safety Research*, 86, 311–317. <https://doi.org/10.1016/j.jsr.2023.07.009>
- Baraza, X., & Cugueró-Escofet, N. (2021). Severity of occupational agricultural accidents in Spain, 2013–2018. *Safety Science*, 143. <https://doi.org/10.1016/j.ssci.2021.105422>
- Baraza, X., & Cugueró-Escofet, N. (2022). Immigration and occupational accidents: A comparative study of accident severity among foreign and Spanish citizens in the agricultural sector. *Spanish Journal of Agricultural Research*, 20(3). <https://doi.org/10.5424/sjar/2022203-18792>
- Baraza, X., Cugueró-Escofet, N., & Rodríguez-Elizalde, R. (2023). Statistical analysis of the severity of occupational accidents in the mining sector. *Journal of Safety Research*, 86, 364–375. <https://doi.org/10.1016/j.jsr.2023.07.015>
- Berglund, L., Johansson, M., Nygren, M., Samuelson, B., Stenberg, M., & Johansson, J. (2021). Occupational accidents in Swedish construction trades. *International Journal of Occupational Safety and Ergonomics*, 27(2), 552–561. <https://doi.org/10.1080/10803548.2019.1598123>
- Besen, E., Young, A., Gaines, B., & Pransky, G. (2018). Lag times in the work disability process: Differences across diagnoses in the length of disability following work-related injury. *Work*, 60(4), 635–648. <https://doi.org/10.3233/WOR-182771>
- Bharti, A., Mohan, K., Singh, R., Vajpai, I., Midha, T., & Diwedi, A. (2019). A clinico-epidemiological study on trauma cases attending a tertiary care hospital in North India. *Journal of Orthopaedics and Allied Sciences*, 7(1), 17. https://doi.org/10.4103/joas.joas_42_18
- Boufous, S., & Williamson, A. (2009). Factors affecting the severity of work related traffic crashes in drivers receiving a worker's compensation claim. *Accident Analysis and Prevention*, 41(3), 467–473.

<https://doi.org/10.1016/j.aap.2009.01.015>

- Bradley, E., Fayaz, S., Farish, C., & Coulson, M. (2023). Evaluation of physical health, mental wellbeing, and injury in a UK Police Firearms unit. *Police Practice and Research*, 24(2), 232–244. <https://doi.org/10.1080/15614263.2022.2080066>
- Bravo, G., Castellucci, H. I., Lavallière, M., Arezes, P. M., Martínez, M., & Duarte, G. (2022). The influence of age on fatal work accidents and lost days in Chile between 2015 and 2019. *Safety Science*, <https://doi.org/10.1016/j.ssci.2021.105599>
- Breiman, L. (2001). Random Forests. *Machine learning*, 45, 5-32.
- CIEDES (2024). Edad promedio de ocupados en Chile sube a 44 años y se prevé que siga aumentando. Recuperado de <https://www.ciedess.cl/601/w3-article-2664.html>
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- COMCARE (2023). Return to work. Australian Government. Recuperado de <https://www.comcare.gov.au/about/forms-pubs/docs/pubs/rehabilitation/return-to-work-information-sheet.pdf>
- Doroudgar, S., Chuang, H. M., Perry, P. J., Thomas, K., Bohnert, K., & Canedo, J. (2017). Driving performance comparing older versus younger drivers. *Traffic Injury Prevention*, 18(1), 41–46. <https://doi.org/10.1080/15389588.2016.1194980>
- Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.
- Eevers, J. E., & Gee, E. M. (1986). Playing It Safe: Accident Mortality and Gender Roles. *Sociological Focus*, 19(4), 349–360. <http://www.jstor.org/stable/20831413>
- Fischer, D., Lombardi, D. A., Folkard, S., Willetts, J., & Christiani, D. C. (2017). Updating the “Risk Index”: A systematic review and meta-analysis of occupational injuries and work schedule characteristics. In *Chronobiology International* (Vol. 34, Issue 10, pp. 1423–1438). Taylor and Francis Ltd. <https://doi.org/10.1080/07420528.2017.1367305>
- Fontaneda, I., Camino López, M. A., González Alcántara, O. J., & Greiner, B. A. (2022). Construction Accidents in Spain: Implications for an Aging Workforce. *BioMed Research International*, 2022. <https://doi.org/10.1155/2022/9952118>
- Fontaneda, I., Camino López, M. A., González Alcántara, O. J., & Greiner, B. A. (2024). The “weekday Effect”: A Decrease in Occupational Accidents from Monday

- to Friday - An Extension of the “monday Effect.” *BioMed Research International*, 2024. <https://doi.org/10.1155/2024/4792081>
- Fontaneda, I., Camino López, M. A., González Alcántara, O. J., & Ritzel, D. O. (2019). Gender differences in lost work days due to occupational accidents. *Safety Science*, 114, 23–29. <https://doi.org/10.1016/j.ssci.2018.12.027>
- Fuentes-Bargues, J. L., Sánchez-Lite, A., González-Gaya, C., Rosales-Prieto, V. F., & Reniers, G. (2022). A study of situational circumstances related to Spain’s occupational accident rates in the metal sector from 2009 to 2019. *Safety Science*, <https://doi.org/10.1016/j.ssci.2022.105700>
- González Fuentes, A., Busto Serrano, N. M., Sánchez Lasheras, F., Fidalgo Valverde, G., & Suárez Sánchez, A. (2022). Work-related overexertion injuries in cleaning occupations: An exploration of the factors to predict the days of absence by means of machine learning methodologies. *Applied Ergonomics*, <https://doi.org/10.1016/j.apergo.2022.103847>
- Gray, S. E., & Collie, A. (2019). Work absence following road traffic crash in Victoria, Australia: A population-based study. *Injury*, 50(7), 1293–1299. <https://doi.org/10.1016/j.injury.2019.05.032>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63, 3-42.
- Gu, J. K., Charles, L. E., Fekedulegn, D., Ma, C. C., Violanti, J. M., & Andrew, M. E. (2020). Occupational injury and psychological distress among U.S. workers: The National Health Interview Survey, 2004–2016. *Journal of Safety Research*, 74, 207–217. <https://doi.org/10.1016/j.jsr.2020.06.002>
- Hösükler, E., Turan, T., & Erkol, Z. Z. (2022). Analysis of injuries and deaths by trauma scores due to occupational accidents. *Ulusal Travma ve Acil Cerrahi Dergisi*, 28(9), 1258–1269. <https://doi.org/10.14744/tjtes.2022.22796>
- Iacobucci, G. (2021). Overcrowding and long delays in A&E caused over 4000 deaths last year in England, analysis shows. *BMJ (Clinical Research Ed.)*, 375, n2835. <https://doi.org/10.1136/bmj.n2835>
- Khairuddin, M. Z. F., Hasikin, K., Razak, N. A. A., Mohshim, S. A., & Ibrahim, S. S. (2023). Harnessing the Multimodal Data Integration and Deep Learning for Occupational Injury Severity Prediction. *IEEE Access*, 11, 85284–85302. <https://doi.org/10.1109/ACCESS.2023.3304328>
- López, M. A. C., Alcántara, O. J. G., Fontaneda, I., & Mañanes, M. (2018). The Risk Factor of Age in Construction Accidents: Important at Present and Fundamental in the Future. *BioMed Research International*, 2018. <https://doi.org/10.1155/2018/2451313>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T-Y. (2017) ‘LightGBM: A Highly Efficient Gradient Boosting Decision Tree’,

- Keren Evangeline, I.; Angeline Kirubha, S.; Glory Precious, J. Prediction of Breast Cancer Recurrence in Five Years using Machine Learning Techniques and SHAP. In *Intelligent Computing Techniques for SmartEnergy Systems*; Springer, 2022; pp. 441–453.
- Kolyshkina, I., & Simoff, S. (2021). Interpretability of Machine Learning Solutions in Public Healthcare: The CRISP-ML Approach. *Frontiers in big data*, <https://doi.org/10.3389/fdata.2021.660206>, <https://www.frontiersin.org/articles/10.3389/fdata.2021.660206/full>.
- Nicholas, M. K., Costa, D. S. J., Linton, S. J., Main, C. J., Shaw, W. S., Pearce, G., Gleeson, M., Pinto, R. Z., Blyth, F. M., McAuley, J. H., Smeets, R. J. E. M., & McGarity, A. (2020). Implementation of Early Intervention Protocol in Australia for ‘High Risk’ Injured Workers is Associated with Fewer Lost Work Days Over 2 Years Than Usual (Stepped) Care. *Journal of Occupational Rehabilitation*, 30(1), 93–104. <https://doi.org/10.1007/s10926-019-09849-y>
- Nieuwenhuijsen, K., Verbeek, J. H., Neumeyer-Gromen, A., Verhoeven, A. C., Bültmann, U., & Faber, B. (2020). Interventions to improve return to work in depressed people In *Cochrane Database of Systematic Reviews* (Vol. 2020, Issue 12). John Wiley and Sons Ltd. <https://doi.org/10.1002/14651858.CD006237.pub4>
- Pietilä, J., Räsänen, T., Reiman, A., Ratilainen, H., & Helander, E. (2018). Characteristics and determinants of recurrent occupational accidents. *Safety Science*, 108, 269–277. <https://doi.org/10.1016/j.ssci.2017.12.020>
- Rey-Merchan, M., & Lopez-Arquillos, A. L. (2021). Organizational and personal factors in occupational traffic injuries at work in Spain. *Traffic Injury Prevention*, 22(7), 519-523. <https://doi.org/10.1080/15389588.2021.1964075>
- Rey-Merchán, M. del C., López-Arquillos, A., & Rey-Merchán, A. M. (2022). Characteristics of Occupational Injuries among Spanish Nursing Workers. *Healthcare (Switzerland)*, 10(2). <https://doi.org/10.3390/healthcare10020220>
- Sampaio, F. E., Oliveira, M. J. S., Areosa, J., & Facas, E. (2023). Prevalence of Comorbidity and Its Effects on Sickness-Absenteeism among Brazilian Legislative Civil Servants. *International Journal of Environmental Research and Public Health*, 20(6). <https://doi.org/10.3390/ijerph20065036>
- Shockey, T. M., & Esser, M. B. (2020). Binge Drinking by Occupation Groups among Currently Employed U.S. Adults in 32 States, 2013–2016. *Substance Use and Misuse*, 55(12), 1968–1979. <https://doi.org/10.1080/10826084.2020.1784947>
- Silva-Aravena, F., Núñez Delafuente, H., Gutiérrez-Bahamondes, J. H., & Morales, J. (2023). A Hybrid Algorithm of ML and XAI to Prevent Breast Cancer: A

- Strategy to Support Decision Making. *Cancers*,15(9), 2443. <https://doi.org/10.3390/cancers15092443>
- Street, T. D., & Lacey, S. J. (2019). Accounting for employee health: The productivity cost of leading health risks. *Health Promotion Journal of Australia*, 30(2), 228–237. <https://doi.org/10.1002/hpja.200>
- SUSESO (2025). Boletín Anual. Estadísticas sobre Seguridad Social 2024. Recuperado de <https://www.suseso.cl/608/w3-propertyvalue-10364.html>
- SUSESO (2025). Informe Regional 2024. Estadísticas de Accidentes del Trabajo y Enfermedades Profesionales. Recuperado de https://www.suseso.cl/607/articles-757454_archivo_01.pdf.
- SUSESO (2025). Estadísticas de Accidentabilidad 2024. Recuperado de https://www.suseso.cl/607/articles-755198_archivo_01.pdf
- SUSESO (2024). Informe Anual de Seguridad y Salud en el Trabajo 2023. Recuperado de <https://www.suseso.cl/607/w3-article-729454.html>
- SUSESO (2023). Estadística de la Seguridad Social 2023. Recuperado de <https://www.suseso.cl/607/w3-article-729383.html>.
- SUSESO (2020). Serie Documentos de Trabajo. Documento de Trabajo N°20. Modelos predictivos por aprendizaje automatizado de accidentes laborales y factores psicosociales del trabajo Recuperado de: https://www.suseso.cl/607/articles-617781_archivo_01.pdf
- SUSESO (2019). Estadísticas de la Seguridad Social 2018. Informe Anual. Recuperado de https://www.suseso.cl/605/articles-578297_recurso_2.pdf
- Vargas-Garrido, H., Moyano-Díaz, E., & Andrades, K. (2021). Sleep problems are related to commuting accidents rather than to workplace accidents. *BMC Public Health*, 21(1). <https://doi.org/10.1186/s12889-021-10737-5>
- Yang, L., Branscum, A., Smit, E., Dreher, D., Howard, K., & Kincl, L. (2020). Work-related injuries and illnesses and their association with hour of work: Analysis of the Oregon construction industry in the US using workers' compensation accepted disabling claims, 2007-2013. *Journal of Occupational Health*, 62(1). <https://doi.org/10.1002/1348-9585.12118>
- Yao, Y., Zhao, S., An, Z., Wang, S., Li, H., Lu, L., & Yao, S. (2019). the associations of work style and physical exercise with the risk of work-related musculoskeletal disorders in nurses. *International Journal of Occupational Medicine and Environmental Health*, 32(1), 15–24. <https://doi.org/10.13075/ijomeh.1896.01331>
- Zhang, G.; Shi, Y.; Yin, P.; Liu, F.; Fang, Y.; Li, X.; Zhang, Q.; Zhang, Z. A machine learning model based on ultrasound image features to assess the risk of sentinel lymph node metastasis in breast cancer patients: Applications of scikit-learn and SHAP. *Frontiers in Oncology* **2022**, 12.