r

Serie Proyectos de Investigación e Innovación

Superintendencia de Seguridad Social
Santiago - Chile

INFORME FINAL

Una estimación de la magnitud de accidentes que generan accidentes potencialmente graves y fatales en base al relato de la denuncia

Autor:
GDS Consultores

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 Social
Huérfanos 1376
Santiago, Chile.

INFORME FINAL

“Estimación de la magnitud de accidentes que generan accidentes potencialmente graves y fatales en base al relato de la denuncia”

Este trabajo fue seleccionado en la Convocatoria de Proyectos de Investigación e Innovación en Prevención de Accidentes y Enfermedades Profesionales 2023 de la Superintendencia de Seguridad Social (Chile), y fue financiado por Mutual de Seguridad CChC con recursos del Seguro Social de la Ley N°16.744 de Accidentes del Trabajo y Enfermedades Profesionales.

Presentado por GDS Consultores Ltda.

Christian Belmar, Investigador Principal
Catalina Araya, Investigadora Secundaria
Fernanda Ferrada, Analista de datos
Matias Philipp, Profesional de apoyo en IA
Alexis Montecinos, Asesor, Especialista en IA

2025

ÍNDICE

I. Resumen ejecutivo 6
II. Palabras claves 7
III. Introducción y antecedentes 7
IV. Objetivos 8
IV.1 Objetivo general 8
IV.2 Objetivos específicos 8
V. Definición del problema, pregunta de investigación o desafío de innovación 9
VI. Marco teórico y comparativa internacional en prevención de riesgos laborales 12
VI.1 Enfoques preventivos, normativas y marcos conceptuales 12
VI.2 Estándares internacionales y herramientas regulatorias 12
VI.3 Comparativa internacional y adaptación al contexto chileno 13
VI.4 Sustento teórico y análisis basado en datos 14
VI.5 Estrategias de prevención y gestión de riesgos 15
VI.6 Diferenciación normativa de incidentes laborales y clasificación CIE-10 16
VII. Revisión de literatura para el enfoque metodológico 18
VIII. Metodología del estudio 20
VIII.1 Diseño metodológico y estructura de datos 20
VIII.1.1 Características, cobertura y estructura de los datos 21
VIII.1.2 Universo elegible inicial y filtros aplicados 22
VIII.1.2 Tratamiento de datos nulos y limitaciones 25
VIII.1.3 Conexión con los objetivos 26
VIII.4 Etapas del proceso metodológico 26
VIII.4.1 Procesamiento de datos 27
VIII.4.2 Análisis de texto a través de Procesamiento de Lenguaje Natural 28
VIII.4.3 Variables de estudio 32
VIII.4.4 Estimación de la magnitud de los accidentes potencialmente graves y/o fatales 37
VIII.5 Herramientas metodológicas aplicadas en el estudio 37
VIII.5.1 Accidentes potencialmente graves y/o fatales 38
VIII.5.2 Justificación conceptual para estimación de la magnitud de accidentes potencialmente graves y/o fatales 39
VIII.5.3 Modelos de predicción 43
IX. Análisis de datos 46
IX.1 Descripción de variables y distribución general 46
IX.1.1 Edad 46
IX.1.2 Género 47
IX.1.3 Gravedad 48
IX.1.4 Calificación 50
IX.1.5 Días de reposo 52
IX.1.6 Accidentes Potencialmente Graves y Fatales 54
IX.1.7 Enfermedades Crónicas 56

X. Resultados **59**
X.1 Resultados analíticos para accidentes potencialmente graves y/o fatales (APG) 59
X.2 Análisis econométrico de accidentes potencialmente graves y severidad del daño. 69
X.2.1 Modelo de variable dependiente limitada: Logit 69
X.2.2 Modelos de aprendizaje automático: Random Forest 74
X.3 Resultados y conexión con los objetivos 76
XI. Validación de Resultados del Modelo de Accidentes Graves y Fatales **79**
XI.1 Reunión con Expertos de Mutual de Seguridad 79
XI.2 Taller de Expertos 80
XI.3 Mejoras y Ajustes Basados en la Validación 81
XII. Discusión **84**
XII.1 Aplicación de modelos generativos GPT 84
XII.2 Consideraciones y limitaciones metodológicas 89
XIII. Conclusiones **90**
XIV. Recomendaciones para Sistema de SST **93**
XIV.1 Tecnologías de monitoreo y enfoque multinivel 93
XIV.2 Programas educativos y adaptativos y acompañamiento técnico por perfil de riesgo 93
XIV.3 Implementación de tecnologías predictivas en la gestión de riesgos laborales 94
XV. Referencias **98**

I. Resumen ejecutivo

El presente estudio tuvo por objetivo identificar y caracterizar los accidentes laborales con alto potencial de gravedad y medir su magnitud en conformidad con los criterios establecidos en la Guía para la Identificación y Evaluación de Riesgos de Seguridad del ISP (2013), reconocida por la Superintendencia de Seguridad Social (SUSESO). A partir de un universo de 2.110.154 registros de denuncias aceptadas por la Mutual de Seguridad CChC (2011–2023), se construyó una base elegible de 1.934.760 observaciones que cumplían con criterios técnicos de completitud y consistencia.

La Guía del ISP establece que el riesgo se define como una combinación entre la probabilidad de ocurrencia de un evento y la magnitud de sus consecuencias. Bajo este enfoque, se clasificaron los accidentes conforme a su severidad esperada, considerando dimensiones clave como días de reposo, presencia de enfermedades crónicas y calificación del accidente (laboral, trayecto, enfermedad profesional). Estos criterios permitieron jerarquizar los eventos, fortaleciendo la trazabilidad del análisis en función de los estándares técnicos establecidos por la autoridad sanitaria.

Entre los principales resultados, se observó que los sectores económicos con mayor proporción de accidentes de alta severidad corresponden a actividades vinculadas a la construcción, transporte y manufactura. Asimismo, se detectaron diferencias relevantes por sexo, con una mayor proporción de casos graves entre hombres, particularmente en actividades de alto riesgo físico. En términos etarios, los trabajadores entre 30 y 49 años concentraron la mayor cantidad de accidentes potencialmente graves, mientras que los segmentos más jóvenes presentaron proporciones más altas de lesiones leves o con menor impacto funcional.

A nivel regional, se identificaron macrozonas con mayor incidencia de accidentes severos, lo que permite orientar estrategias de intervención territorial. Esta caracterización integral refuerza la necesidad de focalizar las acciones preventivas no solo en sectores específicos, sino también en perfiles de trabajadores y entornos laborales donde el riesgo presenta mayor intensidad.

En resumen, el estudio aporta evidencia empírica alineada con la Guía del ISP para apoyar la gestión de riesgos en entornos laborales complejos. Los resultados permiten fortalecer el diseño de estrategias de control, priorización de casos y focalización de políticas preventivas, desde una base técnica sólida y trazable.

II. Palabras claves

Palabras claves: Accidentes, Accidentes Trayecto, Accidentes Fatales, Prevención, Potencialmente Grave, Riesgo, Daño, Severidad

III. Introducción y antecedentes

En 2023, la Mutual de Seguridad registró 14.567 accidentes clasificados como graves, definidos por el nivel de daño conforme a la NT142 del Ministerio de Salud. Este sistema de clasificación se basa en el nivel de daño sufrido por el trabajador y permite enfocar los esfuerzos de prevención hacia las empresas con incidencias significativas, facilitando intervenciones puntuales. Aunque el modelo actual se ha basado tradicionalmente en un enfoque reactivo, que se centra en mitigar los daños una vez ocurridos, también existen esfuerzos preventivos implementados en el sistema. Sin embargo, estos esfuerzos se han limitado principalmente a la identificación y mitigación de riesgos ya existentes, lo que reduce la capacidad de anticipar y prevenir accidentes de alto impacto antes de que ocurran.

El presente estudio busca superar estas limitaciones mediante la identificación y caracterización de accidentes potencialmente graves y/o fatales (SIF Potential) en empresas adheridas a la Mutual. Utilizando técnicas de procesamiento de lenguaje natural (NLP) aplicadas a relatos históricos, se identifican patrones que permiten anticipar riesgos y fortalecer la gestión proactiva de la seguridad laboral. Este enfoque complementa la clasificación vigente, ofreciendo una visión anticipada y más precisa de los riesgos.

La gestión de riesgos de la Mutual de Seguridad ha evolucionado de un enfoque reactivo basado en datos retrospectivos, a incorporar también un enfoque preventivo basado en la detección temprana de riesgos mediante el uso de información en tiempo real. Este enfoque preventivo complementa los esfuerzos actuales, permitiendo la implementación de medidas proactivas antes de que los accidentes ocurran. Además, la Mutual de Seguridad cuenta con herramientas específicas para la identificación de riesgos de alto impacto, como el modelo de precursores y los sistemas de reportes, los cuales permiten una evaluación sistemática de los peligros y su probabilidad de ocurrencia, proporcionando un marco metodológico sólido para la gestión preventiva.

El análisis abarca el período 2011–2023, aplicando técnicas avanzadas de análisis de texto sobre relatos de accidentes para identificar factores recurrentes y mejorar la caracterización de estos eventos. Los hallazgos contribuirán a optimizar herramientas existentes y complementar los actuales enfoques de gestión de riesgos.

En cuanto a experiencias internacionales, el análisis comparativo realizado en el informe indica que países como Argentina, España, Finlandia y México han implementado estrategias preventivas exitosas basadas en la integración de tecnologías predictivas, análisis de datos históricos y modelos de gestión de riesgos. Estos enfoques han sido útiles para identificar patrones de siniestralidad y permitir intervenciones proactivas antes de que los accidentes graves se materialicen. Un estudio del Centro de Sistemas Públicos de la Universidad de Chile para el Instituto de Seguridad Laboral (ISL) destaca la relevancia de los sistemas de vigilancia epidemiológica y la integración de programas de formación como parte de las estrategias replicables para el contexto chileno (Centro de Sistemas Públicos, 2017).

El estudio se estructura en etapas: marco teórico, revisión de literatura, metodologías específicas (análisis de texto, modelos cuantitativos), parametrización de probabilidad y daño, y estimación econométrica de factores asociados a accidentes graves. Sus objetivos incluyen estimar la magnitud de estos eventos, validar modelos predictivos con criterio experto y formular recomendaciones preventivas, contribuyendo a un enfoque más informado y basado en datos para reducir la siniestralidad laboral.

IV. Objetivos

IV.1 Objetivo general

Estimar la magnitud de accidentes en el período 2011-2023, que fueron potencialmente graves y/o fatales a partir del relato de la denuncia, a través de procesamiento con analítica avanzada de texto.

IV.2 Objetivos específicos

- Definir teóricamente la clasificación de una denuncia como un accidente potencialmente grave y/o fatal. (O1)
- Determinar la magnitud de accidentes potencialmente graves y fatales en base al total de denuncias para el periodo 2011 - 2023. (O2)
- Validar con criterio experto los resultados del modelo de clasificación de accidentes potencialmente graves y fatales. (O3)
- Caracterizar los accidentes potencialmente graves y fatales según tipo de accidente, actividad económica y otras variables de interés. (O4)
- Levantar recomendaciones de intervención en base a los resultados, y difusión a nivel de actores claves resultados de estudio. (O5)

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

En Chile, la prevención de accidentes laborales y enfermedades profesionales sigue siendo un desafío crítico para la seguridad y la salud ocupacional. Durante el año 2022, la Mutual de Seguridad reportó 12.910 accidentes graves y accidentes con resultado de muerte, según lo establecido en la NT142 del Ministerio de Salud (MINSAL). Esta normativa define los accidentes graves en función del diagnóstico clínico y su implicancia en el ausentismo, la complejidad del tratamiento y la potencialidad de generar secuelas temporales o permanentes en el trabajador. Si bien este enfoque ha sido útil para la clasificación y gestión de casos, puede reforzar una aproximación predominantemente reactiva al análisis de la siniestralidad. No obstante, las mutualidades también desarrollan acciones preventivas relevantes, como asesorías en terreno, evaluaciones de riesgos y ambientes laborales, diseño y apoyo en programas de prevención, capacitaciones y gestión de comités paritarios, que reflejan un compromiso con la anticipación del riesgo, aunque sus resultados no siempre logran evitar los accidentes de alto impacto.

Aunque la gestión de riesgos de la Mutual de Seguridad se ha apoyado en datos retrospectivos para la prevención en el marco de este estudio, es importante señalar que la Mutual también utiliza información en tiempo real para la toma de decisiones en prevención. Esto permite complementar el uso de datos históricos con estrategias de seguridad laboral basadas en la detección temprana de riesgos.

Además, la Mutual de Seguridad cuenta con instrumentos específicos para la identificación de riesgos de alto impacto, como el modelo de precursores y los sistemas de reportes. En el presente estudio, se utilizó como marco metodológico la Guía para la Identificación y Evaluación de Riesgos de Seguridad en los Ambientes de Trabajo, elaborada por el Instituto de Salud Pública (ISP), la cual establece criterios técnicos estandarizados para la evaluación de peligros y su priorización. Esta guía permite una clasificación sistemática y replicable de los accidentes analizados, facilitando su integración en procesos automatizados de análisis de datos y asegurando consistencia técnica con el marco normativo vigente.

Esta investigación, basada en datos de accidentes del período 2011-2023, busca aplicar técnicas avanzadas de análisis de texto sobre relatos de accidentes, permitiendo identificar factores recurrentes de riesgo y mejorar la caracterización de estos eventos. Los hallazgos obtenidos contribuirán a fortalecer las herramientas existentes, optimizando su capacidad predictiva y complementando los enfoques actuales de gestión de riesgos en la Mutual de Seguridad.

A nivel de gestión en prevención de riesgos laborales, el enfoque tradicional se ha basado en tres estrategias clave: evaluaciones de riesgo, benchmarking y análisis retrospectivo de datos de

accidentes ocurridos. Estas herramientas, aunque ampliamente utilizadas, priorizan medidas correctivas, dejando rezagadas las intervenciones proactivas necesarias para abordar los riesgos desde una perspectiva de prevención y recuperación, como lo sugiere el modelo de gestión del Instituto por una Cultura de Seguridad Industrial (ICSI, 2019).

Un problema particular radica en que, si bien la Mutual de Seguridad cuenta con plataformas para la sistematización de situaciones con alto potencial de riesgo, como los sistemas de reportes, su alcance es limitado. Estos sistemas requieren que las empresas cuenten con un sistema de gestión adecuado, y su efectividad depende en gran medida de la relevancia que la gestión de seguridad tenga en las decisiones estratégicas de las empresas. Este es un aspecto crítico que será retomado al final del informe, a la luz de los resultados obtenidos, especialmente en relación con los sectores donde se detecta una menor capacidad de anticipación del riesgo.

Según los estudios de Krause (2012), un 21% de los accidentes laborales totales poseen características que podrían haber derivado en un accidente grave o fatal. Sin embargo, el desafío radica en otorgar escalabilidad a la detección de estos casos, considerando la masividad de texto disponible sobre circunstancias de accidentes. En otras palabras, el propósito de este estudio es analizar los relatos de accidentes como un insumo clave para identificar aquellos con alto potencial de daño, aunque no necesariamente hayan producido lesiones graves.

Sin embargo, el desafío radica en otorgar escalabilidad a la detección de estos casos, considerando la masividad de texto disponible sobre circunstancias de accidentes. En otras palabras, el propósito de este estudio es analizar los relatos de accidentes como un insumo clave para identificar aquellos con alto potencial de daño, aunque no necesariamente hayan producido lesiones graves. La ausencia de herramientas automatizadas que permitan realizar esta tarea de forma eficiente y replicable en grandes volúmenes de datos representa un vacío en los sistemas actuales de gestión del riesgo. Este vacío no solo dificulta la identificación de patrones o factores de riesgo asociados con los accidentes potencialmente graves y fatales, sino que también limita la capacidad de las empresas adherentes para implementar acciones correctivas y preventivas más efectivas y focalizadas. A partir del análisis anterior, surge la siguiente pregunta central de investigación:

¿Cuál es la magnitud de los accidentes laborales potencialmente graves y fatales respecto del total de accidentes ocurridos en las empresas adherentes a la Mutual de Seguridad CChC durante el período 2011-2023?

El desafío principal radica en desarrollar e implementar una metodología automatizada y escalable para identificar casos de accidentes laborales con potencial de gravedad y fatalidad en la Mutual de Seguridad. Esto implica:

1. Integrar técnicas avanzadas de análisis de datos que permitan clasificar automáticamente los registros de accidentes según su potencial de gravedad.
2. Validar los hallazgos mediante comparaciones con estudios previos y casos clasificados manualmente.
3. Proporcionar una herramienta práctica para las empresas adherentes que facilite la toma de decisiones preventivas basadas en datos, mejorando la eficacia de las intervenciones y reduciendo la ocurrencia de accidentes graves y fatales.

Este enfoque no solo contribuiría a mejorar la gestión de riesgos en las empresas adherentes, sino que también posicionaría a la Mutual de Seguridad como un referente en innovación preventiva en Chile, alineándose con las mejores prácticas internacionales en seguridad y salud ocupacional.

VI. Marco teórico y comparativa internacional en prevención de riesgos laborales

Esta sección establece el marco conceptual y normativo que fundamenta el estudio, integrando definiciones clave, enfoques teóricos, estándares internacionales y aprendizajes aplicables al contexto chileno. Se abordan los modelos proactivo y reactivo de gestión de riesgos, la clasificación legal de accidentes laborales y enfermedades profesionales, y una revisión comparativa de prácticas internacionales relevantes.

VI.1 Enfoques preventivos, normativas y marcos conceptuales

El enfoque reactivo se centra en responder a incidentes después de su ocurrencia, mientras que el enfoque proactivo busca anticiparse mediante la identificación de condiciones de riesgo. Este último resulta más eficaz al integrar tecnologías emergentes, datos históricos y análisis predictivo, como lo plantea Krause (2012) con el concepto de *Serious Injury and Fatality Potential* (SIF Potential), y el modelo del "diamante de la prevención" del ICSI (2019), que permiten detectar incidentes con alto potencial de gravedad, incluso cuando sus consecuencias fueron menores.

Por su parte, la Ley N° 16.744 y la Clasificación CIE-10 permiten distinguir entre accidentes de trabajo, trayecto y enfermedades profesionales, base para estrategias preventivas diferenciadas (SUSESO, 2023). La utilización de estas clasificaciones permite diseñar políticas focalizadas, monitorear tendencias y evaluar intervenciones.

Fomentar una cultura de seguridad laboral implica liderazgo visible, participación de trabajadores, sistemas de reporte y comunicación efectiva (Fernández-Muñiz et al., 2005; Muñoz Salazar et al., 2021). La mejora continua requiere auditorías, revisiones de incidentes y uso de KPI (Agudo Díaz et al., 2017), apoyados por tecnologías IoT para monitoreo en tiempo real (Acuña Acuña et al., 2024).

Prácticas como comités de seguridad, trazabilidad documental, y evaluación del clima organizacional (Oliver et al., 2005) fortalecen la gestión de riesgos. Burke et al. (2006) evidencian que programas de formación multifacética reducen la severidad de accidentes, en línea con el enfoque de la Mutual en competencias preventivas.

VI.2 Estándares internacionales y herramientas regulatorias

La OIT, mediante el Convenio N°155[1] y su guía "Crear una cultura de prevención", junto con el informe conjunto de la OMS/OIT (2021), proporcionan lineamientos clave sobre condiciones laborales seguras. A nivel nacional, los:

- **Anexo N°2: Autoevaluación del Cumplimiento Normativo**

Este documento permite a las empresas realizar una autoevaluación sobre la organización de la prevención de riesgos, condiciones de seguridad y protección personal. Permite detectar brechas en el cumplimiento de la normativa vigente, como la existencia de un Reglamento Interno de Higiene y Seguridad actualizado, el funcionamiento del Comité Paritario de Higiene y Seguridad, y la correcta implementación de protocolos para prevenir el acoso laboral y sexual, entre otros aspectos normativos. Además, el anexo incluye la evaluación de aspectos fundamentales como el saneamiento básico, la señalización de seguridad y la provisión de equipos de protección personal, asegurando que las empresas cumplan con los estándares establecidos.

- **Anexo N°3: Evaluación de Riesgos Críticos**

Este anexo facilita la identificación de las tareas laborales que implican riesgos críticos, tales como trabajos en espacios confinados, trabajos en alturas, exposición a temperaturas extremas y el manejo de maquinaria pesada. Las empresas pueden utilizar este anexo para detectar riesgos en sus actividades y establecer medidas de control preventivo, lo que es clave para evitar accidentes graves o fatales. Este anexo está alineado con las mejores prácticas internacionales en la gestión de riesgos laborales, ya que promueve la identificación temprana de los peligros y la implementación de protocolos de seguridad adecuados.

Ambos anexos, cuando se aplican correctamente, no solo permiten a las empresas cumplir con la normativa legal vigente, sino que también ayudan a fortalecer las estrategias preventivas, alineándose con las mejores prácticas internacionales en seguridad laboral.

VI.3 Comparativa internacional y adaptación al contexto chileno

En Europa, países como Alemania y el Reino Unido han reducido significativamente la siniestralidad laboral gracias al uso de tecnologías predictivas, programas de seguridad multinivel y dispositivos de monitoreo en tiempo real (OIT, 2019; EU-OSHA, 2020). En Latinoamérica, aunque se han implementado reformas en Chile, Brasil y México, persisten desafíos en el cumplimiento normativo y la cultura de seguridad (INSST, 2021; Fundacentro, 2020).

[1] Ver: https://normlex.ilo.org/dyn/normlex/es/f?p=NORMLEXPUB:12100:0::NO::P12100_ILO_CODE:C155

Estas experiencias ofrecen lecciones valiosas para el contexto chileno. La adopción de análisis predictivo, capacitación continua y participación organizacional son elementos replicables en sectores de alto riesgo como la minería y construcción.

VI.4 Sustento teórico y análisis basado en datos

La literatura reciente ha demostrado que la evaluación de riesgos mediante técnicas de procesamiento de lenguaje natural (NLP) puede facilitar la identificación de incidentes según su gravedad potencial, apoyando la priorización de acciones de mitigación y control. Este enfoque, aplicado en experiencias internacionales, fue retomado en este estudio para procesar los relatos de accidentes laborales y detectar patrones asociados a accidentes potencialmente graves (APG).

La prevención de riesgos laborales es fundamental para garantizar la seguridad y bienestar de los trabajadores en las empresas adheridas a Mutual de Seguridad CChC. Tradicionalmente, se ha adoptado un enfoque reactivo, como lo estipula la NT142 del Ministerio de Salud, que clasifica los accidentes graves en función del daño sufrido por el trabajador. Sin embargo, Krause (2012) propone un enfoque proactivo, enfocado en identificar situaciones de alto riesgo antes de que resulten en accidentes graves o fatales, lo que permite anticipar y gestionar los riesgos, aumentando la efectividad de las medidas de seguridad.

Este enfoque se complementa con el modelo del "diamante de la prevención" desarrollado por el Institut pour une Culture de Sécurité Industrielle (ICSI), que entrega un marco estructurado para integrar evaluación de riesgos, capacitación y cultura de seguridad. A diferencia de la pirámide tradicional de siniestralidad, el modelo del diamante plantea que muchos eventos leves o sin lesión pueden contener condiciones críticas que, bajo leves variaciones, podrían haber derivado en consecuencias graves o fatales (ICSI, 2019).

Krause (2016) refuerza este planteamiento indicando que aproximadamente el 21% de los accidentes laborales tienen el potencial de ser graves o fatales. Asimismo, el análisis de datos históricos permite identificar patrones y aplicar medidas preventivas, aunque el enfoque reactivo tiene limitaciones al centrarse únicamente en mitigar daños ya ocurridos. En este contexto, la utilización de la Clasificación Estadística Internacional de Enfermedades (CIE-10) resulta clave para formular políticas de prevención basadas en datos y fomentar un entorno laboral más seguro y saludable.

En este marco, se destaca también el rol del Instituto de Seguridad Laboral (ISL), organismo público encargado de administrar el Seguro Social contra Riesgos de Accidentes del Trabajo y Enfermedades Profesionales. Su función es clave dentro del sistema nacional de seguridad y salud en el trabajo, ya que no solo canaliza licencias médicas y prestaciones económicas, sino que también promueve la prevención de riesgos laborales en sus empresas adheridas. Junto con la SUSESO, el ISL refuerza la infraestructura necesaria para una gestión efectiva de riesgos,

abordando tanto accidentes como enfermedades profesionales desde un enfoque normativo y preventivo integral.

VI.5 Estrategias de prevención y gestión de riesgos

Las estrategias de prevención y gestión de riesgos laborales en Mutual de Seguridad CChC se fundamentan en herramientas como el benchmarking, el análisis de datos históricos y la implementación de sistemas de gestión. Para sustentar estas líneas de acción, se revisan tanto experiencias internacionales como enfoques desarrollados en el contexto chileno, con el objetivo de identificar buenas prácticas replicables y adaptar recomendaciones técnicas a la realidad local. Estas herramientas han sido reconocidas como esenciales para mejorar la efectividad y eficiencia de las políticas preventivas, especialmente en organizaciones complejas.

El estudio del Centro de Sistemas Públicos de la Universidad de Chile para el Instituto de Seguridad Laboral destaca buenas prácticas de Argentina, España, Finlandia y México, subrayando el valor de los programas de formación, la vigilancia epidemiológica y las estrategias integradas de prevención como ejes replicables para el contexto chileno (Centro de Sistemas Públicos, 2017). Complementariamente, el informe de Marsh Risk Consulting y RIMS (2016) documenta cómo organizaciones latinoamericanas han fortalecido sus políticas preventivas a través de modelos de continuidad operativa, gestión de crisis y control de riesgos emergentes. En esa línea, el análisis de datos históricos se reconoce como un insumo clave para construir modelos predictivos y definir intervenciones costo-efectivas basadas en patrones sectoriales y geográficos (OIT & Pontificia Universidad Católica de Chile, 2011)[2].

Desde una perspectiva nacional, el Centro de Políticas Públicas de la Pontificia Universidad Católica de Chile ha señalado la importancia de fortalecer la gestión de riesgos mediante mayor transparencia, focalización de recursos y mecanismos de evaluación de impacto. Estas recomendaciones apuntan hacia un sistema de salud ocupacional basado en evidencia e innovación en prevención.

En cuanto al diseño e implementación de sistemas de gestión, Fernández Muñiz, Montes Peón y Vázquez Ordás destacan que su éxito depende del tamaño organizacional, el compromiso de la alta dirección, la integración con otras áreas funcionales y la participación activa de los trabajadores. Estos factores estructurales determinan la sostenibilidad de las estrategias preventivas dentro de las organizaciones (Fernández-Muñiz et al., 2007).

A nivel operativo, Trillo Cabello (2022)[3] analiza la accidentalidad en obras de construcción y muestra cómo el análisis detallado de las fases de ejecución permite detectar puntos críticos,

[2] Ver documento: https://politicaspublicas.uc.cl/publicacion/analisis-del-sistema-de-seguridad-y-salud-ocupacional-en-chile-y-propuestas-de-mejoramiento/

[3] Ver documento: https://riuma.uma.es/xmlui/bitstream/handle/10630/24741/TD_TRILLO_CABELLO_Antonio_Cabello.pdf

facilitando la planificación de acciones preventivas más focalizadas. Por su parte, el Manual de Criterios de Buena Práctica Profesional del INSST (2011) subraya que la calidad técnica en prevención se potencia con formación especializada, integración transversal en la organización y el uso de estándares técnicos verificados.

En sectores de alto riesgo como la minería y el petróleo, la implementación de sistemas de control de riesgos críticos ha generado reducciones sustantivas en la siniestralidad, como evidencian Selleck y Cattani (2019)[4], al destacar la eficacia de prácticas como auditorías sistemáticas, controles verificables y protocolos de identificación de peligros. De forma complementaria, Kjellén (2023)[5] propone el uso de herramientas digitales que permitan verificar en tiempo real la efectividad de los controles críticos durante las operaciones, impulsando así una prevención más dinámica y basada en evidencia.

Finalmente, Burke et al. (2006)[6] demuestran que los programas de formación multifacética, especialmente aquellos enfocados en tareas de alto riesgo, reducen significativamente la severidad de los accidentes. Estos programas son más efectivos cuando incluyen componentes prácticos, simulaciones y retroalimentación directa, lo cual coincide con el enfoque de la Mutual de Seguridad en competencias preventivas y formación continua del personal.

En conjunto, estos hallazgos respaldan que la prevención de accidentes graves y fatales debe centrarse en la gestión efectiva de riesgos críticos, el uso de tecnologías predictivas, y el fortalecimiento de competencias mediante formación especializada. Su incorporación sistemática en el modelo preventivo de Mutual puede contribuir a una reducción sostenible de la siniestralidad, particularmente en sectores con mayor exposición al riesgo.

VI.6 Diferenciación normativa de incidentes laborales y clasificación CIE-10

La normativa vigente, en particular la Ley N° 16.744, establece la diferenciación entre tres tipos clave de incidentes laborales (SUSESO, 2023):

- **Accidentes de Trabajo Graves o Fatales:** Son aquellos de alta severidad que resultan en lesiones graves (amputaciones, reanimación) o en la muerte del trabajador, y requieren respuestas inmediatas (SUSESO, 2023).

[4] Ver: Selleck, R. & Cattani, M. (2019). *Preventing fatalities in the construction industry—a review of critical risk management strategies*. Journal of Health, Safety and Environment, 35(3), 193–211. https://ro.ecu.edu.au/ecuworkspost2013/7176/?utm_source=chatgpt.com

[5] Ver**:** Kjellén, U. (2023). *Preventing fatal accidents in construction through the management of barriers*. Heliyon, 9(11), e21715. https://doi.org/10.1016/j.heliyon.2023.e21715

[6] Ver: Burke, M.J., Sarpy, S.A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R.O. & Islam, G. (2006). *Relative effectiveness of worker safety and health training methods*. American Journal of Public Health, 96(2), 315–324. https://doi.org/10.2105/AJPH.2004.059840

- **Accidentes de Trayecto:** Ocurren durante el traslado entre el hogar y el trabajo, o entre dos lugares de trabajo, y aunque se producen fuera del ambiente laboral, afectan la seguridad del trabajador (SUSESO, 2023).
- **Enfermedades Profesionales:** Estas enfermedades, son el resultado del ejercicio profesional y se desarrollan a lo largo del tiempo, requiriendo vigilancia y prevención continua (SUSESO, 2023).

Esta diferenciación es clave para definir estrategias de prevención específicas. La utilización de la Clasificación Internacional de Enfermedades (CIE-10) permite estandarizar y analizar los casos de enfermedades profesionales, identificar patrones asociados a sectores económicos, y evaluar la efectividad de intervenciones. Esta codificación homogénea facilita la mejora de protocolos y refuerza las capacidades preventivas de la Mutual de Seguridad CChC.

VII. Revisión de literatura para el enfoque metodológico

Esta sección tiene por objetivo fundamentar la estrategia metodológica del estudio mediante la revisión de antecedentes empíricos y técnicos sobre el uso de herramientas automatizadas para la clasificación de accidentes laborales. En particular, se analizan aplicaciones de procesamiento de lenguaje natural (NLP), minería de texto y aprendizaje automático en contextos de salud ocupacional y gestión de riesgos. La revisión se enfoca en experiencias que permitan mejorar la identificación temprana de accidentes potencialmente graves, a partir del análisis de datos narrativos.

La revisión fue de carácter exploratorio y se nutrió principalmente de literatura técnica, artículos científicos de acceso abierto y documentos institucionales especializados. Se priorizaron antecedentes con aplicaciones prácticas en contextos similares al de las mutualidades chilenas, particularmente en el uso de modelos automatizados de análisis de texto para mejorar la detección de riesgos laborales.

Diversos estudios han demostrado que la integración de NLP, algoritmos de clustering y modelación predictiva mejora la extracción de información y permite una prevención proactiva. Estas metodologías permiten que mutualidades como la Mutual de Seguridad CChC anticipen riesgos y diseñen estrategias más eficaces para reducir la siniestralidad laboral, en línea con las necesidades de entornos complejos.

Uno de los estudios destacados es el de Cardenas-Quispe, Pacheco, Manrique-Nugent y Quispe-Gonzales (2022)[7], quienes analizaron casos confirmados de COVID-19 en Cañete, Perú, utilizando clustering K-means para segmentar datos según la frecuencia de contagios. Esta metodología permitió identificar zonas de mayor riesgo y fue clave para formular políticas preventivas. Si bien se trata de un contexto epidemiológico, su enfoque predictivo basado en segmentación es aplicable a la gestión de riesgos laborales.

En el ámbito de la salud, también se incorporaron los aportes de Rodríguez-Mon Barrera, Guerrero Álvarez y Cano Cos (2021)[8], quienes emplearon representaciones semánticas como Word2Vec y Node2Vec para procesar textos clínicos. Estas técnicas permiten mapear conexiones entre términos técnicos, agrupar descripciones narrativas de accidentes y construir clústeres temáticos significativos, lo que es particularmente útil para el análisis de relatos de siniestros.

[7] Ver estudio: Análisis de Datos y Aplicación de Clustering en Registros de Casos Confirmados por COVID-19 en la Provincia de Cañete.

[8] Rodríguez-Mon Barrera, R., Guerrero Álvarez, M., & Cano Cos, A. (2021). Aplicación de técnicas de procesamiento de lenguaje natural y clustering en datos clínicos. *Revista de Ciencias de la Salud, 19*(3), 456–471. https://doi.org/10.1016/j.rcs.2021.03.005

Además, en estudios sobre licencias médicas y eventos graves se ha evidenciado que la minería de texto avanzada facilita la detección de patrones subyacentes, permitiendo estimar la probabilidad de ocurrencia de eventos críticos a partir de variables contenidas en los relatos. En sistemas de salud, estas técnicas han sido empleadas para priorizar casos en listas de espera, identificar comorbilidades y automatizar anotaciones clínicas. Su aplicación en el ámbito laboral permite enriquecer el análisis de datos estructurados con información semántica proveniente de las denuncias.

El uso combinado de NLP con algoritmos de aprendizaje supervisado y no supervisado posibilita caracterizar eventos pasados y anticipar situaciones de alto riesgo. Esta metodología ha sido adoptada en sistemas de seguridad laboral internacionales, fortaleciendo la prevención basada en evidencia.

A partir de los antecedentes revisados, se respalda una estrategia metodológica que integra análisis semántico, clustering temático y modelación predictiva. El análisis semántico, basado en técnicas de vectorización, permite representar el contenido de los relatos de accidentes de forma numérica, facilitando la identificación de similitudes lingüísticas entre casos. Estos vectores se emplearon en algoritmos como K-means y modelos supervisados como regresión logística, Probit y Random Forest, para estimar la probabilidad de que un accidente corresponda a un evento potencialmente grave (APG). Esta combinación metodológica ha demostrado complementar eficazmente el análisis tradicional, fortaleciendo la capacidad institucional de anticipar eventos críticos y diseñar medidas preventivas más focalizadas.

VIII. Metodología del estudio

Esta sección describe el enfoque metodológico empleado en el estudio, el cual se desarrolló en base a un diseño cuantitativo y aplicado, orientado a identificar y caracterizar los accidentes laborales con alto potencial de gravedad. El análisis se sustentó en el marco técnico de la *Guía para la identificación y evaluación de riesgos de seguridad en los ambientes de trabajo* del Instituto de Salud Pública de Chile (ISP, 2013), documento reconocido por la SUSESO y central en la estimación del riesgo de accidentes. Esta guía fue utilizada para operacionalizar la severidad de los accidentes laborales mediante el cálculo del Valor Esperado de la Pérdida (VEP), lo cual permitió jerarquizar los eventos según su riesgo potencial y establecer criterios objetivos para la priorización de casos y la formulación de estrategias preventivas.

Complementariamente, se emplearon herramientas de análisis de datos, tales como procesamiento de lenguaje natural (NLP) para transformar texto libre en representaciones vectoriales (TF-IDF y Word2Vec), agrupamiento temático mediante algoritmos K-medias, y modelos predictivos supervisados (regresión logística, Probit y Random Forest) para estimar la probabilidad de que un accidente sea potencialmente grave. Estas técnicas permitieron enriquecer el análisis normativo con evidencia empírica extraída desde los relatos.

Si bien el estudio no contempla recolección primaria sistemática, algunas decisiones metodológicas específicas fueron revisadas por expertos internos de la Mutual, lo que permitió reforzar la validez de ciertas definiciones operativas.

El enfoque metodológico adoptado permite operacionalizar los objetivos específicos del estudio en las distintas secciones del informe: la definición teórica y empírica de los accidentes potencialmente graves (O1), su medición a nivel nacional (O2), la validación de resultados mediante criterios expertos (O3), su caracterización sectorial y demográfica (O4), y la formulación de recomendaciones preventivas (O5), que se abordan en los capítulos VIII, IX, X, XI y XIII, respectivamente.

VIII.1 Diseño metodológico y estructura de datos

Esta sección presenta el diseño metodológico del estudio, detallando las fuentes de información utilizadas, la estructura y cobertura de la base de datos, los filtros de inclusión y exclusión aplicados, y las consideraciones técnicas para el tratamiento de datos. Además, se expone la forma en que este enfoque permite operacionalizar los objetivos específicos definidos en el estudio.

La metodología del estudio se fundamenta en el uso de información secundaria, complementada por un proceso de validación experta, con el objetivo de asegurar un análisis técnicamente riguroso y alineado con las prácticas vigentes en seguridad y salud en el trabajo. El acceso a la base de datos fue autorizado formalmente por la Mutual de Seguridad CChC, garantizando el uso ético y seguro de la información.

Se consideraron tres fuentes principales de información:

- **Revisión documental:** Incluyó normativas relevantes (como la NT142 del MINSAL y la Guía para la Identificación y Evaluación de Riesgos del ISP, 2013), literatura técnica nacional e internacional, y reportes institucionales relacionados con la siniestralidad laboral.
- **Validación técnica:** Se desarrolló mediante intercambios iterativos con profesionales expertos de la Mutual de Seguridad CChC, quienes aportaron criterios operativos para la depuración, interpretación y análisis de datos, así como la selección de variables más pertinentes para el enfoque preventivo.
- **Base de datos institucional:** Se accedió, mediante autorización formal, a una base de datos interna de la Mutual de Seguridad CChC que contiene más de 2,1 millones de registros de denuncias de accidentes laborales aceptadas entre los años 2011 y 2023. De este universo, 1,93 millones de casos fueron considerados técnicamente elegibles para el análisis, constituyendo el insumo principal del estudio.

VIII.1.1 Características, cobertura y estructura de los datos

La base de datos institucional incluyó más de 40 variables, tanto estructuradas como no estructuradas, correspondientes a registros de accidentes laborales entre 2011 y 2023. Estas variables abarcan información administrativa, clínica y descriptiva, tales como edad, género, diagnóstico, días de reposo, actividad económica, macrosegmento industrial, región, tipo de calificación del accidente (trabajo, trayecto o enfermedad profesional), resolución del caso, gastos médicos, centro de atención, reingresos, entre otras.

Para efectos del presente estudio, se seleccionaron únicamente aquellas variables con un nivel aceptable de completitud y relevancia técnica. En particular, se emplearon las variables textuales asociadas a la narrativa del accidente:

- **caso_que_hacia:** describe la actividad que realizaba el trabajador al momento del accidente.
- **caso_como_ocurrio:** indica cómo ocurrió el evento.
- **relato:** campo intermedio que combina caso_que_hacia y caso_como_ocurrio.

- **caso_anamnesis:** resume la evaluación clínica inicial.
- **relato_anamnesis:** variable principal para el análisis semántico, compuesta por la concatenación del relato (que ya incluye caso_que_hacia y caso_como_ocurrio) con el campo caso_anamnesis. Representa la narrativa más completa del caso.

Tal como fue validado por el equipo técnico de la Mutual de Seguridad CChC, se trabajó exclusivamente con la variable relato_anamnesis en los análisis de procesamiento de lenguaje natural (NLP), por ser la más informativa respecto a las circunstancias del accidente. En consecuencia, los campos individuales no fueron tratados de forma separada.

Alcance temporal, espacial y poblacional del análisis

- **Temporal:** El análisis se limitó al período 2011–2023, dado que la Mutual considera confiable la calidad y cobertura de los datos desde 2011 en adelante. Registros con fechas fuera de este rango o con errores cronológicos fueron excluidos. Para los casos sin fecha de ocurrencia (atdn_atd_fecha_ocurrencia), se utilizó la fecha de ingreso como proxy temporal.
- **Geográfico:** El estudio tiene cobertura nacional, con información proveniente de todas las regiones del país. Aunque no se realiza un análisis comparativo territorial detallado, la base permite futuras exploraciones a nivel regional o macrozonal.
- **Poblacional:** Se incluyeron únicamente trabajadores y trabajadoras afiliados a la Mutual de Seguridad cuyas denuncias de accidentes fueron aceptadas y contaban con resolución definitiva. Solo se consideraron registros con relato y anamnesis completos. Se excluyeron casos rechazados, pendientes o con datos críticos faltantes.

VIII.1.2 Universo elegible inicial y filtros aplicados

El universo elegible corresponde al subconjunto del total de denuncias aceptadas que cumplen con los requisitos técnicos mínimos para su inclusión en los análisis de texto, modelamiento predictivo y caracterización de accidentes potencialmente graves (APG). Para ello se aplicaron los siguientes criterios de exclusión:

- Registros con resolución “rechazada” o “pendiente”.
- Registros con fechas fuera del periodo 2011–2023.
- Casos con errores de digitación o sin fecha válida de ocurrencia (en estos casos, se utilizó la fecha de ingreso como aproximación).
- Registros sin contenido en los campos de texto ("relato", "caso_anamnesis" y “relato_como_ocurrio”).

- Casos duplicados o inconsistentes, identificados mediante verificación cruzada de identificadores únicos.

Además, se exigió que los registros contaran con contenido estructurado suficiente para aplicar técnicas de procesamiento de lenguaje natural (NLP), descartando aquellos con campos nulos o vacíos en las narrativas textuales.

La base incluye también variables estructuradas como edad, género, días de reposo, porcentaje de incapacidad, calificación del accidente (trabajo, trayecto o enfermedad profesional), región, actividad económica, y macrosegmento industrial. Estas fueron complementadas con variables semánticas derivadas del procesamiento de los relatos.

Para asegurar la consistencia analítica, se aplicaron los siguientes criterios de depuración:

- Se excluyeron registros calificados como “rechazados” o “pendientes”, dado que no corresponden a accidentes de origen laboral o no cuentan con resolución definitiva al momento del análisis.
- Se limitó el análisis al período 2011–2023, rango en el que la Mutual de Seguridad considera que la calidad y cobertura de los datos es confiable y completa.
- Para los casos sin fecha de ocurrencia registrada (atdn_atd_fecha_ocurrencia), se utilizó el campo de fecha de ingreso como proxy temporal, conforme a la validación de la contraparte.

Como resultado del proceso de depuración, se consideraron 1.934.760 registros iniciales que son elegibles para ser estudiados, los cuales representan el 91,7 % del universo aceptado (2.110.154 denuncias).

A continuación se presenta el detalle anual:

Tabla 1. Total inicial de denuncias aceptadas y casos elegibles para análisis de texto, por año

Año	Total universo aceptado	Total inicial elegible para análisis textual
2011	143.308	128.831
2012	154.560	140.990
2013	160.633	147.699
2014	164.474	150.479
2015	164.912	150.239
2016	171.483	154.872
2017	174.779	157.737
2018	173.091	156.467
2019	183.445	167.053
2020	131.682	120.290
2021	150.562	138.035
2022	184.176	174.118
2023	153.049	147.950
Total	2.110.154	1.934.760

Fuente: Elaboración propia con base en datos entregados por Mutual de Seguridad

La depuración de la base permitió conservar el 91,7 % del total de denuncias aceptadas entre 2011 y 2023, lo que asegura una muestra robusta y representativa del fenómeno en estudio. A lo largo del período, se observa una tasa de elegibilidad relativamente estable, con valores anuales cercanos al promedio general, lo que evidencia consistencia en la calidad del registro institucional.

Los años con mayor volumen de registros aceptados fueron 2019 (183.445), 2022 (184.176) y 2017 (174.779), reflejando una tendencia general de aumento previo a la pandemia. En contraste, el año 2020 muestra una disminución significativa en el número total de denuncias (131.682), atribuible a las restricciones sanitarias y reducción de actividades productivas durante ese período.

En términos de elegibilidad, la proporción de registros válidos se mantuvo por encima del 89 % en todos los años, con picos de cobertura en 2022 (94,5 %) y 2023 (96,7 %), lo que sugiere una mejora progresiva en la completitud y consistencia de los campos utilizados para el análisis textual.

Una vez aplicado este proceso de depuración, se obtiene una muestra inicial de 1.934.760 casos elegibles para el análisis, que incluye tanto accidentes leves como graves y/o fatales. Es importante señalar que en esta etapa aún no se ha realizado la clasificación de accidentes potencialmente graves (APG). Dicha clasificación se aplicará en fases posteriores del análisis, utilizando técnicas de procesamiento de lenguaje natural y modelamiento predictivo. En esa etapa, se excluirán expresamente los casos ya calificados como accidentes graves o fatales, con

el fin de asegurar que la variable APG represente únicamente eventos con *potencial* gravedad no reconocida oficialmente. Por lo tanto, los resultados que se presentan en esta sección corresponden al universo elegible inicial sobre el cual se aplicarán las metodologías de construcción de variables semánticas, incluida la variable APG.

VIII.1.2 Tratamiento de datos nulos y limitaciones

Una vez definido el universo elegible, se procedió a aplicar criterios complementarios de revisión de calidad y consistencia de los datos, los cuales se detallan a continuación.

Durante la fase de depuración, se aplicaron criterios rigurosos para asegurar la calidad, consistencia y trazabilidad de los datos utilizados en el análisis. En particular:

- Se excluyeron registros con datos faltantes en variables críticas como resolución del caso, fecha de ocurrencia, campos de relato o anamnesis.
- No se aplicaron imputaciones sobre variables críticas. Todos los análisis se basaron exclusivamente en información verificable y completa.
- En el caso de variables no críticas (por ejemplo, macrosegmento o región), se mantuvieron los registros aunque presentaran valores faltantes, siempre que su ausencia no comprometiera los objetivos del análisis. Estos valores fueron etiquetados de manera diferenciada.

Además, el análisis reveló limitaciones estructurales en algunas variables específicas del registro administrativo:

- rslc_porcentaje_incapacidad presenta un 99,2 % de datos faltantes, ya que solo aplica a casos con secuelas permanentes evaluadas por comités especializados. Su baja cobertura es esperada y no afecta la generalización de los resultados.
- caso_anamnesis y caso_como_ocurrio muestran omisiones principalmente en denuncias rechazadas, reingresos o enfermedades profesionales, en las que estos campos no siempre son requeridos.
- gasto y tramo_gasto presentan inconsistencias atribuibles a registros sin atención médica posterior (como altas inmediatas) o casos sin cobertura aplicable.
- Se identificaron errores de digitación en campos como edad (valores negativos o extremos), género (clasificación "N") y diagnósticos sin código CIE10, los cuales fueron corregidos mediante limpieza o exclusión de registros anómalos.

Este enfoque metodológico, validado por el equipo técnico de la Mutual de Seguridad CChC, garantiza que los modelos predictivos y análisis textuales se basen en un conjunto de datos

consistente, verificable y alineado con los estándares de calidad requeridos para este tipo de análisis.

VIII.1.3 Conexión con los objetivos

Este análisis inicial permite disponer de una base amplia, depurada y técnicamente consistente con los objetivos del estudio, lo que asegura la pertinencia de los datos empleados en los modelos y análisis posteriores. La alta cobertura, la validación de criterios de inclusión y la calidad de los registros seleccionados habilitan su uso como insumo confiable para la caracterización de accidentes y la identificación de patrones de riesgo.

Además, permite identificar patrones y posibles brechas en la clasificación original de accidentes, contribuyendo directamente al objetivo de prevenir casos potencialmente graves no observados.

La aplicación de los criterios de selección, depuración y estructuración metodológica se encuentra alineada con los objetivos específicos, del siguiente modo:

- Permite estimar la magnitud de los accidentes potencialmente graves y/o fatales desde un universo depurado y con información suficiente (Objetivo General y O2).
- Asegura la validez técnica del procesamiento textual mediante el uso de campos específicos, lo que habilita el desarrollo del diccionario de términos críticos y su integración en modelos clasificatorios (O1).
- Facilita la validación de resultados con criterio experto, ya que el conjunto depurado representa los casos efectivamente tramitados y clasificados institucionalmente (O3).
- Provee una base estructurada y segmentada que permite caracterizar los APG según variables clave como forma de ocurrencia, sector económico, edad y otros factores (O4).
- La trazabilidad de los filtros y fuentes garantiza que los resultados sean comunicables y útiles para formular recomendaciones dirigidas a los actores claves en la prevención de riesgos laborales (O5).

VIII.4 Etapas del proceso metodológico

Antes de detallar las etapas específicas del procesamiento de datos, a continuación se presentan los principales componentes metodológicos que estructuraron el análisis general del estudio:

- **Depuración y estructuración de datos:** Se aplicaron filtros para delimitar el universo elegible, excluir registros incompletos, corregir errores de digitación y eliminar duplicados. También se verificó la coherencia temporal de las fechas de ocurrencia.
- **Definición y construcción de variables analíticas:** A partir del campo relato_anamnesis, se generaron variables derivadas (como uso de herramientas y presencia de enfermedades crónicas), además de estructurar variables existentes (edad, género, días de reposo, etc.) para integrarlas al análisis cuantitativo.
- **Identificación de accidentes potencialmente graves (APG):** Se diseñó una clasificación binaria (APG/no APG) a través de técnicas de procesamiento de lenguaje natural y algoritmos de agrupamiento, reconociendo patrones comunes en accidentes con desenlace grave o fatal.
- **Aplicación de técnicas estadísticas y de aprendizaje automático:** Se emplearon modelos de regresión logística y probit para estimar la probabilidad de gravedad potencial. Asimismo, se aplicaron algoritmos no supervisados (K-medias) y Random Forest para caracterizar patrones y validar la capacidad predictiva.
- **Validación de resultados y análisis de resultados:** Se aplicaron métricas de ajuste y validación, y se revisaron los patrones textuales mediante talleres con expertos en prevención de riesgos laborales, ciencia de datos e inteligencia artificial. Esto permitió contrastar los hallazgos con criterios técnicos y normativos (NT142 y Guía ISP), asegurando su coherencia y pertinencia.

Este marco general orienta las etapas operativas que se detallan a continuación.

VIII.4.1 Procesamiento de datos

Las etapas de procesamiento parten de la base validada en la sección anterior, correspondiente al universo elegible de registros que cumplen criterios técnicos para su análisis. Dentro de las etapas del proceso metodológico, se emplean modelos predictivos, como Random Forest y regresión logística, que se integran de manera complementaria con la Guía para la Identificación y Evaluación de Riesgos de Seguridad en los Ambientes de Trabajo del ISP. Estas herramientas predictivas permiten anticipar los accidentes graves al analizar tanto datos históricos como en tiempo real. De este modo, se optimiza la gestión de riesgos al proporcionar una visión más precisa de la probabilidad de ocurrencia y severidad de los accidentes, reforzando el modelo normativo existente y facilitando la priorización de las medidas preventivas.

A continuación se presentan las etapas del procesamiento de datos

1. **Depuración del universo inicial**

Se aplicaron filtros secuenciales para garantizar la consistencia de los registros, incluyendo:

- Exclusión de denuncias rechazadas o con resolución pendiente.
- Eliminación de registros con errores de digitación o fechas fuera del periodo 2011–2023.
- Filtrado por presencia de campos clave como relato y anamnesis, esenciales para el procesamiento de texto.

2. **Tratamiento de datos faltantes**

- En variables críticas (fecha, resolución, relato): se excluyó completamente el registro.
- En variables no críticas (por ejemplo, macrosegmento o región): se retuvo el registro y se marcó como "faltante" sin imputar, siempre que su ausencia no afectará los objetivos analíticos.
- No se aplicaron imputaciones sobre campos utilizados en el modelamiento predictivo (APG, días de reposo, porcentaje de incapacidad).

3. **Construcción del set analítico**

- A partir de los filtros, se generó una base consolidada de 2.110.623 registros válidos para análisis inicial.
- De estos, 1.934.760 cumplían con los requisitos de completitud y calidad necesarios para aplicar técnicas de análisis textual y modelamiento de APG.
- Este conjunto representa el 91,7 % del total y constituye el universo efectivo de observaciones utilizadas en las siguientes etapas metodológicas.

VIII.4.2 Análisis de texto a través de Procesamiento de Lenguaje Natural

El análisis de texto se utiliza para extraer información significativa de relatos textuales relacionados con accidentes laborales. Dentro de las técnicas disponibles, el Procesamiento de Lenguaje Natural (NLP, por sus siglas en inglés), una rama de la inteligencia artificial, permite transformar datos no estructurados en información numérica y estructurada, mejorando la interpretación y facilitando la toma de decisiones. Uno de los mayores desafíos al trabajar con texto es convertir esta información no estructurada en una representación numérica que pueda ser utilizada en análisis estadísticos y modelos predictivos. Este proceso incluyó varias etapas, tales como limpieza, normalización, tokenización y vectorización de los textos, las que permitieron preparar el corpus para su análisis estructurado.

El procesamiento fue implementado en entorno Python 3.11, utilizando Jupyter Notebook como interfaz de desarrollo interactiva. Las herramientas empleadas incluyeron las bibliotecas pandas (2.2.2) para la manipulación de datos, nltk (3.8.1) para el tratamiento lingüístico básico (eliminación de palabras vacías en español, tokenización y normalización), re para expresiones regulares y limpieza de caracteres especiales, unicodedata para estandarización de acentos y

codificación, y scikit-learn (1.4.2) para el modelado vectorial mediante TF-IDF y exploración preliminar de agrupamientos y clasificación. Adicionalmente, se utilizaron matplotlib y seaborn para la visualización de patrones y frecuencias. Aunque se exploró la lematización con spaCy (3.7.2), se optó por una tokenización básica considerando criterios de eficiencia operativa dada la magnitud del corpus.

1. **Limpieza y normalización del texto:** Este proceso se aplicó directamente sobre los 1.934.760 registros válidos que contenían información en los campos relato y anamnesis, tras la depuración de la base de datos. El objetivo fue estandarizar el contenido textual para facilitar su análisis posterior mediante técnicas de procesamiento de lenguaje natural (NLP).

La limpieza incluyó:

- **Conversión a minúsculas para uniformar las entradas:** Todas las palabras fueron transformadas a minúsculas con el fin de evitar duplicidades léxicas artificiales derivadas del uso de mayúsculas o minúsculas. Por ejemplo, “Máquina” y “máquina” pasan a ser tratados como el mismo término. Esta transformación es esencial para una correcta tokenización y análisis de frecuencia.
- **Eliminación de signos de puntuación, números y caracteres especiales:** Se eliminaron todos los signos de puntuación (comas, puntos, paréntesis, etc.), así como dígitos y símbolos especiales (%, @, /, etc.), que no aportan valor semántico en este contexto. Esta operación reduce el “ruido” en el corpus y mejora la eficiencia de los algoritmos de vectorización y clasificación.

- **Eliminación de palabras vacías (stop-words) en español**: Se eliminaron palabras muy frecuentes del idioma español que no tienen carga semántica diferenciadora, como el, la, de, en, que, y, por, entre otras. Estas fueron identificadas mediante una lista predefinida estándar provista por la biblioteca **nltk**, ampliamente utilizada en procesamiento de lenguaje natural. Su exclusión permite centrar el análisis en términos con mayor capacidad informativa, como verbos de acción (ej. resbaló, cortó, atrapó) o sustantivos relevantes (ej. máquina, superficie, fractura).

Posteriormente, se aplicaron las siguientes técnicas de normalización semántica:

- Para reducir cada palabra a su forma canónica según su contexto gramatical se utiliza la técnica denominada Lematización. Por ejemplo, las formas verbales “trabajaba”, “trabajando” y “trabajó” fueron reducidas a la forma base “trabajar”. Esta técnica, basada en reglas lingüísticas del español, se utilizó para mejorar la precisión semántica al representar los relatos de accidentes de manera coherente y unificada.

- Se utilizó una reducción morfológica más agresiva sin considerar el contexto gramatical para identificar palabras en un contexto específico a través de una técnica

llamada Stemming con el último algoritmo denominado Snowball, permitiendo que, por ejemplo, las palabras “cortado”, “cortando” y “corte” fueron reducidas a la raíz común “cort”. Esta técnica fue útil para pruebas exploratorias de agrupamiento temático mediante K-means, con el objetivo de evaluar la sensibilidad del modelo a distintas formas de normalización semántica.

Ambos procedimientos fueron implementados utilizando librerías especializadas y aplicados a nivel de registro textual.

2. **Tokenización:** Una vez normalizado el contenido, se procedió a segmentar cada registro en unidades mínimas de análisis, lo que se denomina *tokens*, permitiendo transformar los relatos en vectores de palabras y preparar la información para la etapa de vectorización y modelado.

Este tratamiento se aplicó exclusivamente sobre el conjunto previamente filtrado y validado, garantizando la calidad semántica del insumo textual empleado en los modelos predictivos.

3. **Vectorización:**

Para transformar los relatos textuales en insumos numéricos analizables, se aplicaron dos métodos de vectorización complementarios:

- **TF-IDF (Term Frequency - Inverse Document Frequency):** Este método mide la relevancia de cada término dentro de un relato ajustándola según su frecuencia en todo el corpus. Así, términos frecuentes en un relato pero poco comunes en el conjunto total adquieren mayor peso, destacando palabras clave para la identificación de accidentes.

- **Word2Vec**: Este modelo representa cada palabra como un vector en un espacio multidimensional, capturando relaciones semánticas y sintácticas. De este modo, términos que comparten contexto aparecen cercanos entre sí, facilitando la detección de similitudes y patrones complejos en los relatos.

Aunque el Procesamiento de Lenguaje Natural (NLP) permite aplicaciones diversas como traducción automática o asistentes virtuales, en esta investigación se aplicaron exclusivamente técnicas orientadas a la detección de patrones léxicos asociados a accidentes potencialmente graves (APG). El objetivo fue generar variables clasificatorias y construir insumos analíticos coherentes con el diseño metodológico del estudio. La información estructurada resultante alimentó el modelo predictivo de APG y permitió realizar análisis estadísticos complementarios, como distribuciones de frecuencia, que reforzaron la validación empírica de los hallazgos.

En base a los relatos de accidentes laborales de 2023, se implementaron diccionarios temáticos construidos a partir de la Guía de Criterios de Notificación de Accidentes Graves SUSESO, lo

que permitió identificar palabras clave asociadas a cada tipo de accidente. A continuación se presenta la tabla completa de tipos de accidente y términos del diccionario utilizados para la clasificación automática:

Tabla 2. Total de cluster asociados a actividad económica y sus términos

Tipo de accidente	Palabra raíz	Términos del diccionario
Caídas	caid, altur	golp, nivel, altur, resbal, escaler, tropiez, mot, golpe, biciclet, torsion, desnivel, vehicul, motociclet, escalon, tropez, escal, andami
Atrapamiento	atrap, maquin	maquin, puert, camion, porton, bus, maquinari, microbus, aut, vehicul
Sustancias peligrosas	quimic, inhal	product, inhal, olor, limpi, clor, polv, gas, fug
Golpes / impacto	golp, vehicul	mot, fierr, vehicul, tablon, camionet, bus, conductor, ladron, frent, colect, chuz, micr
Incendio / explosión	fueg, explos	hum, inhal, amag, material, quimic, fueg, cort, combustion, producc, quemadur, estanqu, gas, electr, fall, llam, aceit, combuist, inflam
Cortes / amputaciones	cort, amput	ded, man, falang, anul, meniqu, pulg, vulcometal, serruch, sierr, motosierr, galleter, guillotin, minipim, soldador

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

Adicionalmente, se aplicó un enfoque no supervisado mediante el algoritmo K-medias, con el objetivo de descubrir agrupaciones naturales en los relatos vectorizados, sin depender de etiquetas predefinidas. A diferencia de una segmentación puramente léxica, en este caso se estableció una clasificación específica por actividad económica, utilizando diccionarios temáticos que asocian términos clave a contextos productivos concretos como construcción, comercio, transporte, agricultura, administración pública, alojamiento y alimentación, enseñanza y atención de la salud humana.

Cada clúster (K) se definió en función de la presencia de términos representativos de riesgos y accidentes característicos de cada sector económico. Por ejemplo, en Construcción se agruparon términos como desprendi, roc, altur y hormigon; en Comercio palabras vinculadas a asaltos, cortes o transporte de productos; y en Transporte y almacenamiento términos relacionados con colisiones, bencina o sustancias peligrosas. Este esquema permitió configurar clústeres coherentes con la realidad de cada sector, reflejando dinámicas y patrones de ocurrencia diferenciados.

El resultado de esta agrupación fue una clasificación de relatos por sectores productivos, revelando patrones léxicos comunes dentro de cada giro económico y facilitando la identificación de variables clave asociadas a riesgos específicos. Esta segmentación no supervisada refuerza la interpretación de los factores que inciden en la gravedad potencial de los accidentes, aportando evidencia para diseñar estrategias de prevención y control adaptadas a la naturaleza de cada actividad económica.

Tabla 3. Total de cluster asociados a actividad económica y sus términos

Actividad Económica	Clúster	Términos Asociados
Construcción	C1, C3, C7	desprendi, desprend, roc, golp, cay, estab, altur, golpe, mont, pasador, montaj, hormig, vald, hormigon, llen, bald
Comercio	C1, C5, C9	band, pas, mecher, transport, cajer, guardi, product, pick, up, camionet, ped, cubr, ded, man, cort, her, menisc, embalaj, embal, lca, rotur, asalt, amenaz, rob, individu, arma, angusti, delincuent, fueg, mied, agresion, diner, insomni, agred, verbal, emocional, carabiner, ansied, estres, denunci, pun, blanc, nervi, sujet, desconoc, dorm, suen, muert, ojo, extran, cuerp, salt, particul, esmeril, esqu
Actividades de servicios, administración y apoyo	C1, C2, C6, C7, C8, C10	fueg, amenaz, arma, asalt, angusti, emocional, armas, mied, labil, insomni, apunt, ansied, concili, intimid, suen, nervios, amenz, pistol, psicolog, eric, temor, rob, individu, encanon, verbal, muert, armad, nervi, ansios, martill, dolenci, rodamient, rro, golpe, golp, fij, estrab, aserrin, ojo, aserrader, podador, extran, cuerp, chisp, sold, salt, soldadur, soldador, quemadur, ojo, esmeril, arco, electr, prend, quem, enciend, antiparr, oid, masc, enchuf, epifor, angul, encend, cort
Transporte y almacenamiento	C7, C8	contien, sulfur, acid, aposit, cubet, ecn, bencin, inyect, salpic, aguj, maxi, median, energi, veloc, colision, taman, maxim, cinturon, vehicul, laseg
Agricultura y ganadería	C3, C4	descol, manguer, neg, ant, med, reconoc, dermatitis, tecnific, cam, app, maq, excav, retr, retroexcav, ra, atenc, chuz, maquinari, maquin, galleter
Administración pública	C2, C3, C4, C7	atasc, abrirl, impresor, fotocopi, tac, tropies, rodill, cae, tropiez, golp, jueg, activ, particip, cuerd, nin, jug, futbol, parvul, ojo, extran, salt, ocul, ojos, cuerp, gas, ardor, olor, gargant, particul, comienz, picazon, eman
Alojamiento y alimentación	C7, C9, C10	yod, mezon, meson, contr, golpe, vall, nev, ski, esqui, niev, pist, andarivel, sky, ticket, socorr, centr
Enseñanza	C1, C10	aguj, pinch, fuent, contamin, cortopunz, puncion, jering, hepatitis, vhb, examen, vacun, vhc, proced, muestr, truv, torsedur, manifiest, sufr, tobill, labor, pie, report, soldadur, estab
Atención de la salud humana	C1, C5, C10	cement, hipotiroidismoalergi, hrsrefier, coxal, resbaladiz, peldan, traser, mader, jefatur, glute, deport, paracetamol, column, cayend, nivel, segund, dos, oficin, espald, monitor, product, desd, contr, escalon, recint, mism, lumb, dentr, golp, suel, sent, haci, cuand, anosmi, ageusi, mialgi, caracteriz, antigen, cov, febril, tos, sinovac, coronavac, fiebr, estrech, cuantific, disne, cefale, dosis, palanc

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

VIII.4.3 Variables de estudio

Variables Incluidas en los datos originales

La base de datos proporcionada por la Mutual de Seguridad para el periodo 2011-2023 contiene información detallada sobre incidentes y accidentes laborales. La variable principal de los datos corresponde al relato anamnesis, el cual reúne información no estructurada de tres campos que son llenados al momento de crear el registro. A continuación, se describen las variables

incluidas en esta base de datos, independientemente de si fueron utilizadas en los análisis finales.

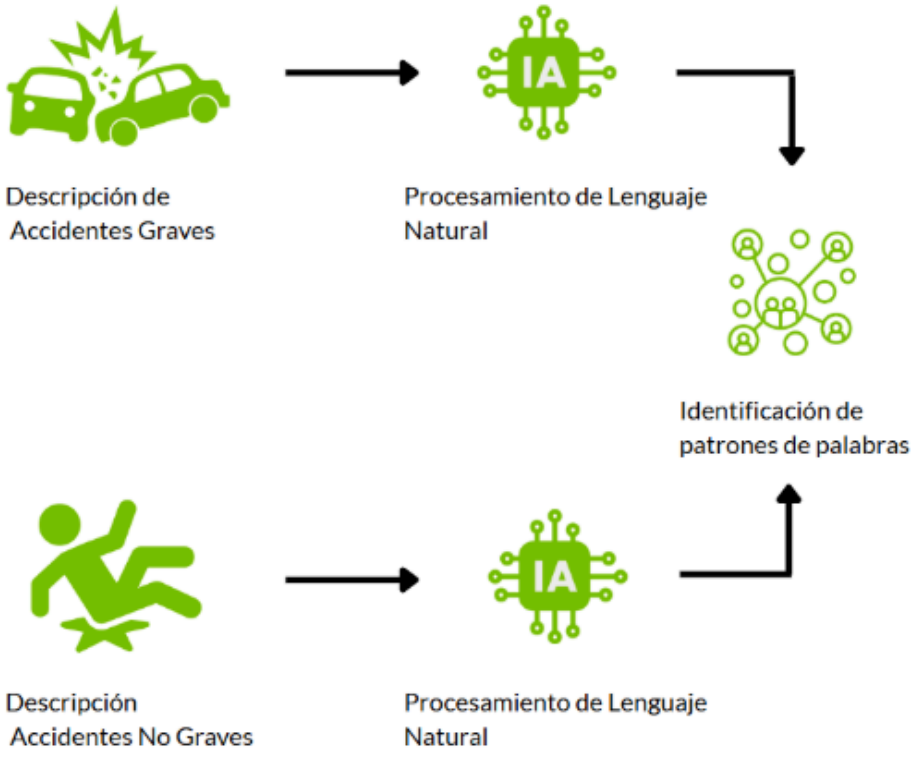
Estas variables permiten contextualizar y analizar los accidentes laborales desde diferentes dimensiones, como características demográficas, laborales, regionales y sectoriales, así como su gravedad e impacto. La información está estructurada en la siguiente tabla:

Tabla 4. Variables y descripción

Variable	Descripción
Relato anamnesis	Información no estructurada recopilada de tres campos al momento de registrar el accidente, se construye al juntar las variables "relato_como_ocurrio", "caso anamnesis", y "relato"
Edad	Factor relacionado con diferencias en resistencia física, experiencia laboral y capacidad de recuperación.
Género	Permite analizar diferencias en incidencia y gravedad de accidentes según género, influenciado por tipos de trabajo y condiciones laborales.
Días de reposo	Días de licencia generados por el accidente; mayor número de días se asocia con mayor gravedad y/o fatalidad.
Porcentaje de incapacidad	Nivel de incapacidad derivado del accidente; valores más altos indican mayor gravedad.
Gravedad	Clasificación de la gravedad del accidente según la Norma Técnica 142 del MINSAL.
Promedio de Trabajadores	Refleja la escala operativa de la empresa; influye en el nivel de riesgo y la implementación de medidas de seguridad.
Región	Considera variaciones en normativas de seguridad y prácticas laborales entre regiones.
Macrosegmento de la Industria	Clasificación del entorno industrial que refleja riesgos específicos y medidas de seguridad aplicables.
Actividad Económica	Clasificación de la actividad económica para contextualizar los riesgos asociados al sector.
Calificación	Estado del accidente o enfermedad profesional, con categorías: Sin Información, Accidentes de Trabajo, Accidentes de Trayecto y Enfermedades Profesionales.

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

Aplicación de métodos de NPL para la creación de nuevas variables



A partir de la base de datos proporcionada, se procedió a la creación de varias nuevas variables que permiten enriquecer el análisis de los accidentes laborales y facilitar la caracterización de los mismos. Las principales variables creadas son:

Creación de nuevas variables para el análisis

Variable Accidente Potencialmente Grave y/o Fatales

La variable "Accidente Potencialmente Grave" (APG) fue creada para clasificar los accidentes laborales como aquellos que por su contexto podrían haber sido graves, pero no tuvo consecuencias de este tipo. Para su construcción, se analizaron aquellos accidentes que tuvieron consecuencias graves, reconociendo a través de la técnica de Procesamiento de Lenguaje Natural (NPL por sus siglas en inglés) patrones de este tipo, los cuales se identifican y se buscan en accidentes que no tuvieron resultados graves.

El desarrollo de la nueva variable destinada a identificar este tipo de accidentes con características textuales similares a aquellos clasificados como graves y/o fatales, con el propósito de anticipar eventos de alta gravedad en función de patrones narrativos presentes en los relatos de los accidentes. Finalmente, se aplican técnicas de clustering (aprendizaje no supervisado) para la creación de las variables, el proceso se resume a continuación:

- **Segmentación de Accidentes Graves y/o Fatales:** Se seleccionó el subconjunto específico de la base de datos correspondiente a accidentes laborales graves y fatales registrados oficialmente como fatales. Este subconjunto fue utilizado como referencia para extraer características comunes en los relatos asociados a dichos eventos.

- **Análisis mediante Clustering:** A partir del conjunto de accidentes fatales, se aplicaron técnicas de clustering no supervisado para agrupar los relatos según similitud semántica. Para ello, se utilizaron representaciones vectoriales del texto como TF-IDF y Word2Vec. Esto permitió identificar patrones de lenguaje característicos de los accidentes con desenlace fatal, generando una base de conocimiento replicable.

- **Aplicación de Patrón a Base Completa:** Los patrones identificados fueron aplicados a toda la base de accidentes mediante técnicas de PLN. Se construyó un diccionario de términos relevantes como “caída de altura”, “atropello”, “golpe en la cabeza”, “pérdida de conciencia”, “aplastamiento”, “choque eléctrico”, entre otros. La detección de estas expresiones permitió identificar relatos compatibles con accidentes de alta gravedad.

- **Clasificación Dicotómica:** A partir de la detección total o parcial de ciertos patrones críticos en los relatos, se construyó una variable dicotómica que clasifica los casos como potencialmente graves y/o fatales (valor 1) o no (valor 0). Para el análisis de métricas, se consideraron únicamente aquellos casos con valor 1 en la variable APG que no hayan sido previamente registrados como accidentes graves. De este modo, se resguarda la calificación de “potencial”. Esta variable fue incorporada a la base de datos bajo el nombre **APG**.

- **Validación Metodológica:** La consistencia de esta clasificación fue evaluada mediante la revisión cruzada de casos efectivamente clasificados como graves y/o fatales. Asimismo, se contrastó con otras variables complementarias como gravedad, días de

reposo, uso de herramientas y presencia de enfermedades crónicas, lo que reforzó la robustez del enfoque adoptado.

Esta variable permite incorporar una dimensión adicional al análisis de riesgo, basada no solo en datos estructurados, sino en información narrativa contenida en los relatos de accidente, ampliando así las posibilidades de identificación temprana de eventos de alta severidad.

Variable de Enfermedades Crónicas

Se desarrolló una nueva variable destinada a identificar la presencia de enfermedades crónicas en los pacientes, basada en los relatos contenidos en la variable que une tanto el relato del accidente con la anamnesis, es decir, "relato_anamnesis". Esta variable fue construida utilizando el análisis de texto en los registros de accidentes, específicamente para el año 2022, con el objetivo de integrar información que pudiera identificar condiciones crónicas relevantes para el contexto del accidente.

La inclusión de esta variable se sustenta en evidencia empírica que demuestra que las enfermedades crónicas pueden aumentar el riesgo de accidentes laborales. Un estudio realizado por Kubo et al. (2013) en el sector manufacturero encontró que trabajadores con condiciones como hipertensión, diabetes, obesidad o trastornos musculoesqueléticos presentan una mayor probabilidad de sufrir accidentes en el trabajo, incluso al controlar por factores ocupacionales. Estos hallazgos respaldan la pertinencia de incorporar antecedentes de salud crónica al análisis, particularmente en investigaciones orientadas a identificar factores asociados a la severidad o recurrencia de los accidentes laborales.

Por tanto, la integración de esta variable busca fortalecer la capacidad predictiva del modelo para anticipar eventos con alta severidad, considerando factores de vulnerabilidad asociados al estado de salud del trabajador.

El proceso de creación de esta variable fue el siguiente:

- **Identificación de Patrones y Términos**: Se definieron las siglas AM (Antecedentes de Morbilidad) y AP (Antecedentes del Paciente) como los principales indicadores de enfermedades crónicas asociadas a los pacientes. Para asegurar la relevancia y alineación con los estándares de salud, se realizó una revisión de literatura médica y se utilizó el listado de enfermedades con cobertura del régimen de Garantías Explícitas en Salud (GES) como base para desarrollar la lista de enfermedades crónicas comunes. Algunas de las enfermedades crónicas incluidas en la lista son: diabetes, hipertensión, obesidad, asma, lupus, entre otras. Este enfoque asegura que las condiciones más prevalentes y reconocidas por el sistema de salud chileno sean consideradas en el análisis.

- **Análisis del Texto:** En cada registro de accidente, se realizaron búsquedas de condiciones crónicas dentro del campo relato_anamnesis. Para ello, se emplearon

expresiones regulares que permitieron detectar tanto enfermedades específicas como menciones explícitas del término “crónica”. De esta manera, se identificaron 24.185 observaciones que contenían al menos una de las enfermedades crónicas durante el año 2022, lo que representó el 13,13 % del total de accidentes registrados en ese año.

- **Validación y Negación**: Además de identificar la presencia de enfermedades crónicas, el análisis consideró las negaciones que aparecían en los relatos. En estos casos, se descartaron aquellos registros donde se especificaba que el paciente no presentaba antecedentes de enfermedades crónicas, garantizando así una clasificación más precisa y adecuada.

- **Vinculación con el GES:** El uso del GES como marco de referencia fue fundamental para validar las enfermedades crónicas identificadas en los relatos. Este sistema de salud chileno cubre una amplia gama de condiciones crónicas que, al ser incluidas en el listado de enfermedades, refuerzan la validez del análisis realizado. Al utilizar las enfermedades crónicas cubiertas por el GES, el proceso de clasificación se alinea con un estándar reconocido y formal en el país, asegurando que las condiciones crónicas identificadas en los registros tengan una correlación con las enfermedades de base que afectan a la población en general.

VIII.4.4 Estimación de la magnitud de los accidentes potencialmente graves y/o fatales

Una vez identificados los accidentes potencialmente graves (APG), se procedió a estimar su magnitud mediante la cuantificación del riesgo asociado a cada caso. Para ello, se calculó el nivel de riesgo individual y se clasificaron los accidentes según el sector económico correspondiente, considerando su frecuencia relativa en el total de eventos observados. El resultado de este procedimiento fue una matriz de distribución que relaciona sectores económicos con los distintos tipos de riesgo identificados, permitiendo establecer concentraciones de casos relevantes para orientar medidas preventivas específicas.

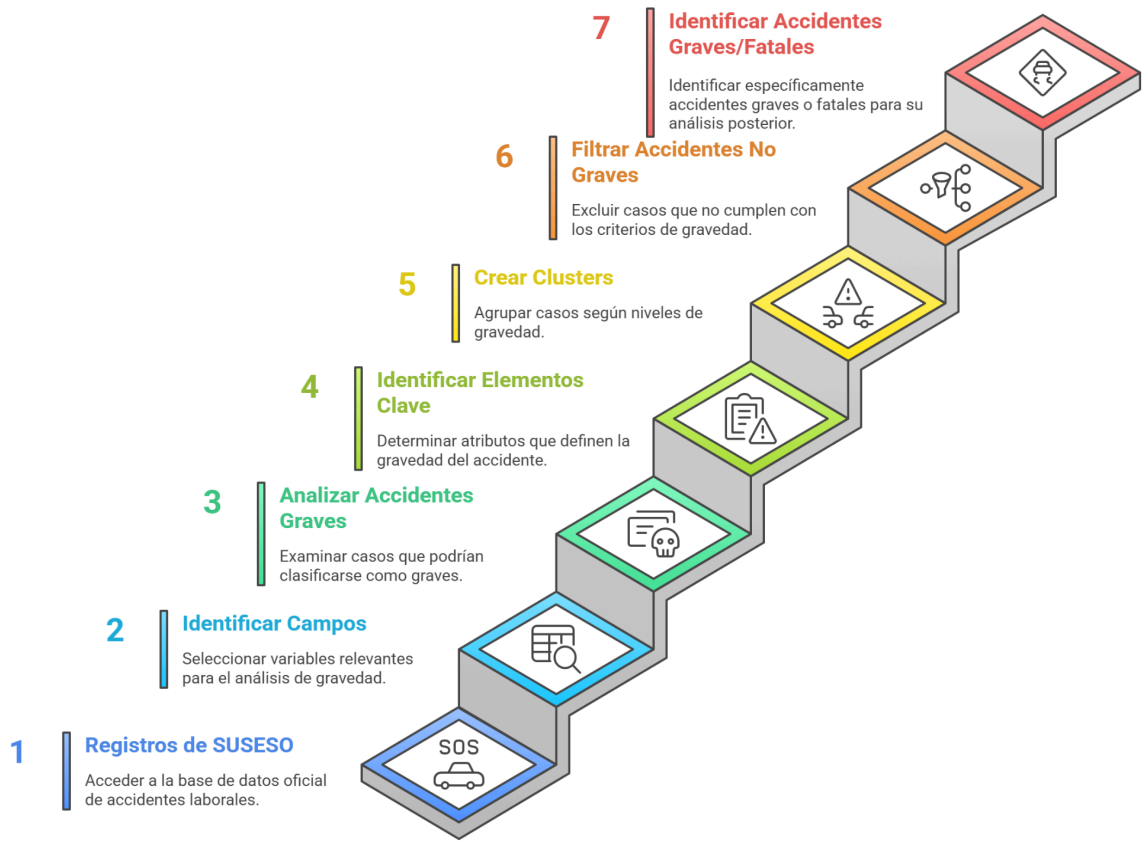
VIII.5 Herramientas metodológicas aplicadas en el estudio

Esta sección presenta las herramientas metodológicas aplicadas en el estudio, explicando su pertinencia en relación con los objetivos planteados. Cada técnica se seleccionó en función de su capacidad para abordar problemas específicos: identificación de condiciones de riesgo, caracterización de accidentes a partir de texto libre, detección de patrones ocultos y predicción de accidentes potencialmente graves. En conjunto, estas herramientas permiten desarrollar una aproximación robusta, replicable y alineada con estándares técnicos reconocidos en prevención de riesgos laborales.

VIII.5.1 Accidentes potencialmente graves y/o fatales

El objetivo de esta metodología es identificar accidentes que, aunque no hayan sido clasificados como graves, presentan condiciones que podrían haber resultado en consecuencias severas o fatales. Para ello, se emplea un análisis de texto sobre los antecedentes del accidente, permitiendo detectar aquellos eventos con un alto potencial de gravedad o fatalidad. Para lograr identificar este tipo de accidente, se procede a utilizar la siguiente estrategia de análisis de datos:

Figura 1. Estrategia identificación accidentes potencialmente graves y/o fatales



Fuente: Elaboración propia

La estrategia de identificación permite que el análisis de los accidentes se base en evidencia empírica, lo que facilita la detección de incidentes en contextos específicos de cada actividad económica. Esto garantiza que los casos identificados no se limiten únicamente a criterios teóricos, sino que reflejen la información registrada en los datos.

VIII.5.2 Justificación conceptual para estimación de la magnitud de accidentes potencialmente graves y/o fatales

La estimación de la magnitud del riesgo de accidentes potencialmente graves y/o fatales se basa en el marco metodológico establecido por la **Guía para la identificación y evaluación de riesgos de seguridad en los ambientes de trabajo del Instituto de Salud Pública (ISP, 2013)**[9] , instrumento técnico reconocido por la Superintendencia de Seguridad Social (SUSESO). Conforme a este marco, se emplea el método del Valor Esperado de la Pérdida (VEP), que permite cuantificar el riesgo mediante la fórmula:

$$VEP = Probabilidad \times Consecuencia$$

La estimación del VEP se realiza considerando exclusivamente aquellos riesgos no evitables, definidos como aquellos que no pueden eliminarse fácilmente sin incurrir en costos significativos o sin alterar los procesos productivos. Para cada riesgo identificado, se asignan valores a las siguientes variables:

- **Probabilidad (P):** Se clasifica como **baja (1), media (2) o alta (4)**, en función de la frecuencia de exposición, condiciones inseguras existentes y antecedentes de ocurrencia.

- **Consecuencia o severidad (C):** Se clasifica como **ligeramente dañina (1), dañina (2) o extremadamente dañina (4)**, considerando el impacto sobre la salud del trabajador (desde lesiones menores hasta accidentes fatales) y el nivel de daño material a la organización.

En este sentido, es fácil verificar que la VEP tomará valores determinados por la combinación de estas variables. En otras palabras, a cada accidente se le asigna un valor dado que representará el nivel de riesgo: 1 (trivial), 2 (tolerable), 4 (moderado), 8 (importante) y 16 (intolerable), los cuales se definen en la siguiente tabla:

[9] Ver: Instituto de Salud Pública de Chile (ISP), 2013. Guía para la identificación y evaluación de riesgos de seguridad en los ambientes de trabajo (D003-PR-500-02-001). Disponible en: https://www.ispch.cl/sites/default/files/D003-PR.500.02.001%20Gu%C3%ADa%20para%20la%20identificaci%C3%B3n%20y%20evaluaci%C3%B3n%20de%20riesgos%20de%20seguridad.pdf

Tabla 5. Magnitud del riesgo detectable

VEP	Nivel de Riesgo	Acción y Temporización
1	Trivial	No se requiere acción específica.
2	Tolerable	No se requiere mejorar la acción preventiva. Se sugiere revisar medidas más eficientes sin que impliquen carga económica. Se requieren verificaciones periódicas.
4	Moderado	Se deben implementar medidas correctivas en un plazo definido. Si las consecuencias son extremadamente graves, se deberá ajustar la probabilidad y medidas de control.
8	Importante	No debe iniciarse ni continuar el trabajo hasta reducir el riesgo. Si el trabajo está en curso, debe corregirse en un plazo menor al riesgo moderado.
16	Intolerable	Se debe suspender el trabajo hasta eliminar o reducir el riesgo. Si no es posible, el trabajo debe prohibirse.

Fuente: Guía para la identificación y evaluación de riesgos de seguridad en los ambientes de trabajo (D003-PR-500-02-001). Instituto de Salud Pública de Chile (ISP), 2013.

Esta metodología permite una cuantificación objetiva y replicable del riesgo, facilitando la priorización de medidas preventivas y el monitoreo de su efectividad.

Este enfoque, adaptado desde la Guía del ISP (2013), permite que el presente estudio utilice criterios técnicos ya validados institucionalmente para traducir el riesgo teórico en una medida operativa, alineando los hallazgos con las exigencias regulatorias vigentes.

A partir de estos componentes descritos, la magnitud de los accidentes potencialmente graves y/o fatales será determinada siguiendo esta metodología. Por lo consiguiente, se analizará cada uno de los accidentes conforme su probabilidad y su consecuencia de ocurrencia:

a) Cálculo de probabilidad:

Para las estimaciones de probabilidad es necesario analizar los componentes probabilisticos que se desprenden de las bases de datos. Como primera tarea en este estudio, se estudian los relatos de los accidentes y se encuentran aquellos que por sus condiciones son accidentes potencialmente graves (APG). Si se mide la frecuencias de estos, se está midiendo la probabilidad de que un accidente sea clasificado como potencialmente grave (APG) a través de un enfoque condicional, expresado como:

$$P(\text{APG}|\text{Accidente})$$

Es decir, al calcular la frecuencia relativa de los APG respecto del total de accidentes registrados, se obtiene una probabilidad condicional que representa el nivel de riesgo **dado que un accidente ha ocurrido**. Esta estimación constituye la base para el análisis posterior de

magnitud del riesgo, conforme a lo establecido en la Guía del ISP (2013), y se complementa más adelante con el cálculo de la severidad a partir de los días de reposo.

Por otro lado, para determinar la probabilidad de ocurrencia, se considera la frecuencia de los accidentes determinados por la actividad económica en que estos suceden, es decir que, $P(\text{Accidente})$ corresponde al cociente representa la frecuencia relativa de accidentes dentro de cada sector económico, en relación con el total de accidentes registrados. A esta métrica se le denomina **tasa de incidencia sectorial y** permite ponderar el riesgo de APG en función de la exposición estructural al accidente que presenta cada sector. Matemáticamente se define como:

$$P(\text{Accidente}) = \frac{\text{Total de accidentes en el sector economico}}{\text{Total de accidentes registrados en todos los sectores}}$$

Estas dos medidas definidas permiten estudiar la probabilidad total o riesgo absoluto de que ocurra un accidente potencialmente grave en el conjunto de la población trabajadora, se aplica la regla del producto de probabilidades:

$$P(\text{APG}) = P(\text{APG}|\text{Accidente}) \cdot P(\text{Accidente})$$

Donde:

- $P(\text{APG})$: probabilidad total de ocurrencia de un APG (riesgo absoluto),
- $P(\text{APG}|\text{Accidente})$: probabilidad condicional estimada a través de un modelo logit,
- $P(\text{Accidente})$: es la **tasa de incidencia sectoria**l, es decir, la probabilidad de que ocurra un accidente en la población trabajadora en determinado sector económico, lo que refleja la frecuencia relativa de accidentes dentro de cada sector económico en relación con el total de accidentes registrados.

Finalmente, el cálculo de $P(\text{APG})$ por sector económico se expresa como:

$$P(\text{APG en el sector } j) = P(\text{APG}|\text{Accidente})_j \cdot P(\text{Accidente})_j$$

Donde el subíndice j denota el sector económico específico. Este enfoque permite obtener una estimación más precisa del riesgo absoluto de ocurrencia de eventos graves, integrando tanto la probabilidad de severidad como la frecuencia relativa de accidentes por sector. Además, resulta especialmente útil para priorizar acciones preventivas en contextos de alta exposición combinada.

Una vez obtenida las probabilidades, se realiza un análisis de distribución de éstas y se identifican los diferentes rangos de probabilidad, donde la ocurrencia alta es si se encuentra por sobre el 75% de la distribución, baja si se encuentra por debajo del 50% y finalmente

ocurrencia media en los valores intermedios. Luego, se reparametriza según los valores identificados en la guía base del estudio según la siguiente distribución:

$$P(D) = \begin{cases} 1, & \text{si } P \leq P_{bajo} \\ 2, & \text{si } P_{bajo} < P \leq P_{medio} \\ 4, & \text{si } P > P_{medio} \end{cases}$$

b) Cálculo del daño probable:

Para encontrar el daño probable, variable que representa la consecuencia en términos de la guía que se utiliza como base en este estudio, se analizan los días de reposo que generó el accidente. Para esto, el total de la muestra se subdivide en deciles, y luego se califican los intervalos que indicarían daño en una escala de 1 a 3, siendo 1 el valor más bajo y 3 el más alto. Adicionalmente, se asigna el número mayor de daño probable si existe una herramienta o una enfermedad crónica asociada al APG. Se define D como daño probable, de la siguiente manera:

$$D = \begin{cases} 1, & \text{si } 0 \leq \text{días_reposo} < 13 \\ 2, & \text{si } 13 \leq \text{días_reposo} \leq 30 \\ 4, & \text{si días_reposo} > 30 \text{ o hay herramienta o enfermedad crónica involucrada} \end{cases}$$

Con las variables de daño y probabilidad, definimos el **Valor Esperado de Pérdida Probable**, en este sentido se le agrega el término último debido a que en el contexto, estamos examinando accidentes potenciales. Este es:

$$VEPP = D\ x\ Probabilidad\ APG\ reparametrizada$$

Este indicador permite ordenar los registros según su nivel de riesgo potencial, facilitando comparaciones entre sectores, calificaciones y períodos, y sirviendo como base para el desarrollo de estrategias preventivas más focalizadas.

Por lo tanto, el plan de trabajo para calcular la VEPP se resume cómo:

a) Cálculo de probabilidades: Se consideran la probabilidad de accidentes potencialmente graves condicional, la tasa de incidencia y se calcula la probabilidad de accidentes potencialmente graves. Se reparametriza conforme los indicadores de la **Guía para la identificación y evaluación de riesgos de seguridad en los ambientes de trabajo del Instituto de Salud Pública.**
b) Cálculo de daño: Se estima según los días de reposo o los elementos o enfermedades involucradas en el accidente. Se reparametriza, al igual que el caso anterior, conforme la **Guía para la identificación y evaluación de riesgos de seguridad en los ambientes de trabajo del Instituto de Salud Pública.**

c) Finalmente, se multiplican los valores para daño y probabilidades, obteniendo el Valor Esperado de Pérdida Probable (VEPP).

VIII.5.3 Modelos de predicción

El presente estudio incorpora modelos estadísticos y de Machine Learning con el objetivo de identificar factores predictivos asociados a accidentes potencialmente graves, mejorar la clasificación de eventos laborales y fortalecer las acciones preventivas. Se optó por una combinación de técnicas tradicionales (regresión logística y Probit) y avanzadas (Random Forest), dada su complementariedad metodológica y su aplicabilidad comprobada en contextos laborales complejos.

Regresión logística

La regresión logística permite modelar la probabilidad de que un accidente sea clasificado como potencialmente grave, utilizando como insumos variables explicativas cuantitativas y categóricas extraídas del relato, la anamnesis y otras características del evento. Dado que la variable dependiente es dicotómica (potencialmente grave / no potencialmente grave), esta técnica resulta idónea para problemas de clasificación binaria.

El modelo se expresa como:

$$log\left(\frac{P(y=1)}{P(y=0)}\right) = X\beta + \varepsilon,$$

Donde $P(y = 1)$ es la probabilidad de que un accidente sea potencialmente grave, X.es el vector de variables independientes, β el vector de coeficientes estimados y ε, el término de error.

Los coeficientes del modelo se interpretan mediante su exponenciación (e^{β}), lo que permite estimar la variación relativa en la probabilidad ante cambios unitarios en las variables explicativas. Para evaluar el ajuste del modelo y su capacidad predictiva, se utilizaron las siguientes métricas:

- Log-Likelihood y Pseudo-R² de McFadden: estimadores de ajuste global.
- Curva ROC y AUC: indican capacidad discriminativa del modelo.
- Significancia de coeficientes: determina la relevancia estadística de cada variable.

Esta técnica permitió identificar factores estadísticamente significativos asociados a la ocurrencia de accidentes potencialmente graves, facilitando la focalización de estrategias preventivas basadas en evidencia.

Regresión Probit

La regresión Probit es una técnica estadística similar a la regresión logística, pero utiliza la función de distribución acumulativa normal (CDF) para modelar la probabilidad de un evento. Este modelo es especialmente útil cuando se trabaja con variables dependientes categóricas y se busca asumir una distribución normal subyacente en los errores. Matemáticamente, se representa como:

$$P(y=1)=\Phi(X\beta)$$

donde:

- $P(y=1)$ es la probabilidad de que un accidente sea clasificado como potencialmente grave,
- Φ es la función de distribución acumulativa normal,
- X es el vector de variables independientes,
- β es el vector de coeficientes.

Este modelo es especialmente útil cuando se asume que los errores del modelo siguen una distribución normal, y permite evaluar los resultados bajo supuestos distintos a los del modelo logístico. En este estudio, su uso complementa el análisis, validando la consistencia de los factores explicativos identificados en la clasificación de accidentes potencialmente graves.

Modelo predictivo Machine Learning: Random Forest

Random Forest es un algoritmo de aprendizaje automático basado en árboles de decisión. Su funcionamiento se basa en la generación de múltiples árboles a partir de subconjuntos aleatorios del conjunto de datos (método bootstrap) y la combinación de sus predicciones mediante votación (clasificación) o promedio (regresión).

Este enfoque es especialmente robusto frente a problemas de sobreajuste y permite capturar relaciones no lineales complejas entre variables. En el presente estudio, Random Forest se empleó para modelar la probabilidad de ocurrencia de accidentes potencialmente graves, a partir de variables estructuradas y texto procesado. Su capacidad para manejar alta dimensionalidad e interacciones complejas lo convierte en una herramienta idónea para complementar los modelos tradicionales.

La combinación de técnicas estadísticas tradicionales (Logit y Probit) con modelos de Machine Learning avanzados (Random Forest) permitió validar patrones de riesgo mediante múltiples aproximaciones, reforzando la solidez de los resultados. Esta triangulación metodológica fortalece la capacidad del estudio para sustentar estrategias de prevención basadas en evidencia empírica y ajustadas a distintos perfiles y sectores de riesgo.

Además, la validación práctica de estos modelos con datos históricos y casos reales aporta beneficios operacionales directos para Mutual de Seguridad CChC. Por ejemplo, estudios previos han demostrado que modelos como Random Forest alcanzan precisiones superiores al 85 % en la predicción de accidentes laborales relacionados con factores como la falta de EPP, condiciones peligrosas y fatiga (Gutiérrez & Fernández, 2012). Este hallazgo es consistente con estudios recientes como el de Alassaf et al. (2023)[10], quienes aplicaron técnicas de Machine Learning en obras de construcción, obteniendo resultados altamente precisos en la predicción de accidentes mediante algoritmos como Random Forest y SVM.

Entre las principales ventajas de aplicar estos modelos destacan:

- Identificación de patrones ocultos de riesgo: permite detectar configuraciones de variables que anticipan condiciones peligrosas, incluso si no derivaron en un accidente grave.
- Optimización de recursos preventivos: prioriza sectores o actividades con mayor probabilidad de siniestralidad, mejorando la asignación de esfuerzos preventivos.
- Reducción de la siniestralidad grave: al intervenir sobre los factores que más influyen en la probabilidad de accidentes graves.
- Gestión basada en evidencia: promueve decisiones técnicas y adaptativas, fundamentadas en datos y patrones reales, más allá de las aproximaciones tradicionales normativas o reactivas.

En suma, esta estrategia metodológica no solo refuerza el rigor técnico del estudio, sino que permite transitar hacia un modelo de gestión preventiva basado en predicción, fortaleciendo las capacidades institucionales para anticipar y reducir accidentes laborales graves y fatales.

[10] Ver estudio: https://arxiv.org/pdf/2301.03567

IX. Análisis de datos

Esta sección presenta la caracterización descriptiva de las principales variables contenidas en la base de datos de accidentes laborales reportados entre los años 2011 y 2023, con el objetivo de identificar tendencias relevantes que sustenten el diseño y evaluación de modelos predictivos de severidad y riesgo. Para ello, se analiza la distribución anual de variables demográficas, clínicas y administrativas, incorporando desagregaciones por edad, género, gravedad del evento, calificación legal, duración del reposo médico, presencia de enfermedades crónicas y ocurrencia de accidentes potencialmente graves o fatales (APG).

IX.1 Descripción de variables y distribución general

IX.1.1 Edad

La variable "edad" agrupa a las personas afectadas por accidentes en distintos tramos etarios, lo que permite identificar la distribución de eventos según etapas del ciclo laboral. Este análisis permite observar tendencias demográficas que influyen en el perfil de riesgo y exposición.

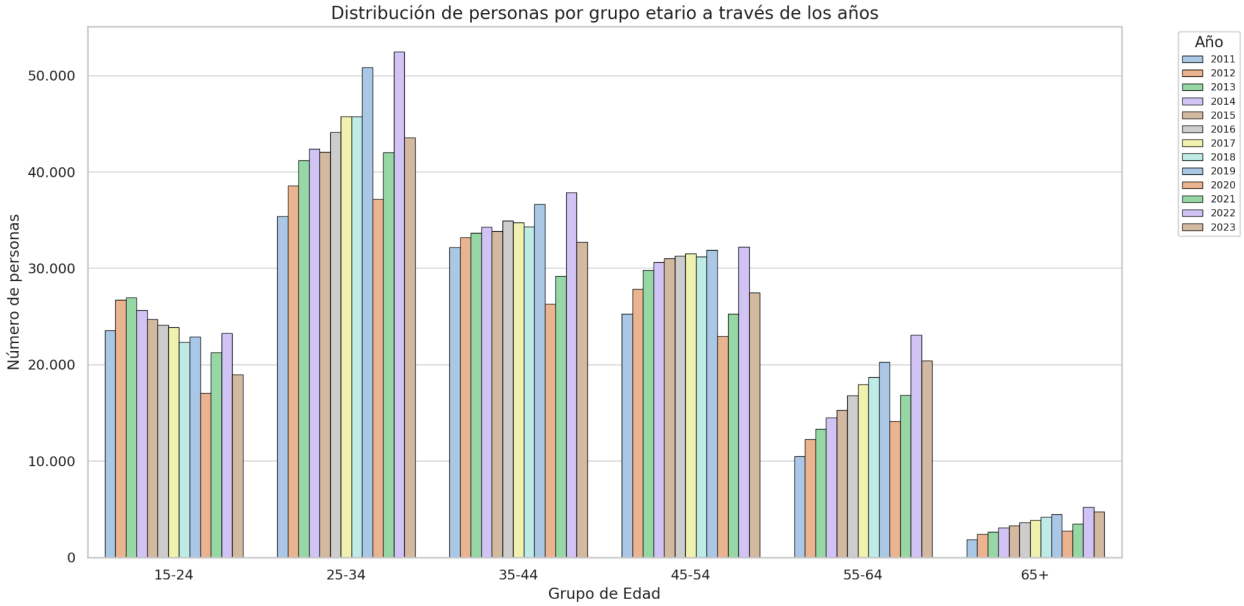
En general, los datos destacan el predominio de los grupos de edad intermedios (25-44 años), que representan la mayor proporción a lo largo del tiempo, y un notable aumento transversal en todos los segmentos etarios hacia los últimos años. Este cambio sugiere la necesidad de analizar posibles factores como políticas públicas, expansión de cobertura o cambios demográficos significativos.

Tabla 6. Distribución de accidentes por grupo etario (2011–2023)

Grupo de Edad	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
15-24	18.28%	18.96%	18.27%	17.03%	16.45%	15.58%	15.14%	14.27%	13.71%	14.16%	15.39%	13.36%	12.83%
25-34	27.52%	27.36%	27.91%	28.18%	28.02%	28.50%	29.01%	29.25%	30.44%	30.90%	30.44%	30.13%	29.46%
35-44	25.02%	23.55%	22.82%	22.77%	22.54%	22.56%	22.04%	21.94%	21.95%	21.85%	21.16%	21.74%	22.12%
45-54	19.62%	19.76%	20.19%	20.35%	20.63%	20.21%	19.97%	19.93%	19.08%	19.06%	18.30%	18.51%	18.58%
55-64	8.13%	8.68%	9.03%	9.63%	10.17%	10.82%	11.38%	11.95%	12.15%	11.74%	12.18%	13.26%	13.81%
65+	1.42%	1.69%	1.79%	2.04%	2.20%	2.33%	2.45%	2.67%	2.66%	2.28%	2.53%	3.00%	3.20%

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

Gráfico 1. Evolución del total de accidentes por grupos de edad años 2011-2023



Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

El gráfico 1 muestra la evolución anual del número total de accidentes registrados, desagregados por grupo etario. Se observa un predominio persistente de los tramos intermedios (25 a 44 años), así como un crecimiento sostenido en los grupos de mayor edad (55 años y más), especialmente desde 2016 en adelante. La caída observada en 2020 afecta transversalmente a todos los grupos, seguida de una recuperación parcial durante los años posteriores. Este comportamiento sugiere cambios estructurales en la distribución etaria de los trabajadores y posibles efectos vinculados a dinámicas del mercado laboral y políticas de prevención.

IX.1.2 Género

La variable pcnt_genero identifica el sexo biológico de la persona afectada, registrando dos categorías: F y M. Esta variable es categórica nominal dicotómica, y está completamente disponible en el conjunto de datos (sin valores nulos). Su análisis permite observar la distribución diferencial de accidentes y enfermedades profesionales según el género, lo cual resulta relevante para la evaluación de brechas estructurales, patrones ocupacionales diferenciados por sexo y potenciales sesgos en la exposición al riesgo o en la denuncia de casos.

En 2011, el 71,94 % de los casos registrados correspondían a hombres. Para 2023, esta proporción desciende al 59,37 %, evidenciando una disminución de 12,57 puntos porcentuales en un período de 13 años.

Simultáneamente, la proporción de mujeres aumenta desde un 28,06 % en 2011 hasta alcanzar el 40,63 % en 2023. Este crecimiento es progresivo, destacando repuntes en los periodos 2017–2018 (+0,84 pp) y 2021–2022 (+8,75 pp), posiblemente vinculados a cambios institucionales en el sistema de reporte o mayor visibilidad de los riesgos psicosociales.

Entre 2011 y 2014, la distribución se mantiene relativamente estable, con la proporción femenina fluctuando entre 28 % y 31 %. Una anomalía se observa en 2020, con una baja significativa en el total de casos (-51.759 respecto a 2019), y un leve aumento en la participación masculina (69,44 %), probablemente atribuible al contexto de pandemia y sus efectos sobre el acceso y registro de casos.

En 2023, la brecha de género se reduce a 18,74 puntos porcentuales, en comparación con los 43,88 pp registrados en 2011, acercándose progresivamente a un escenario más equilibrado en la distribución de casos por género.

Tabla 7. Distribución de accidentes por género (2011–2023)

Año	Masculino (M)	Femenino (F)	Total	% Masculino	% Femenino
2011	103.091	40.217	143.308	71,94	28,06
2012	109.074	45.486	154.560	70,57	29,43
2013	112.516	48.117	160.633	70,05	29,95
2014	112.670	51.804	164.474	68,5	31,5
2015	112.225	52.687	164.912	68,05	31,95
2016	114.701	56.782	171.483	66,89	33,11
2017	115.592	59.187	174.779	66,14	33,86
2018	113.035	60.056	173.091	65,3	34,7
2019	118.846	64.599	183.445	64,79	35,21
2020	91.442	40.244	131.686	69,44	30,56
2021	103.548	47.014	150.562	68,77	31,23
2022	110.544	73.632	184.176	60,02	39,98
2023	90.866	62.183	153.049	59,37	40,63

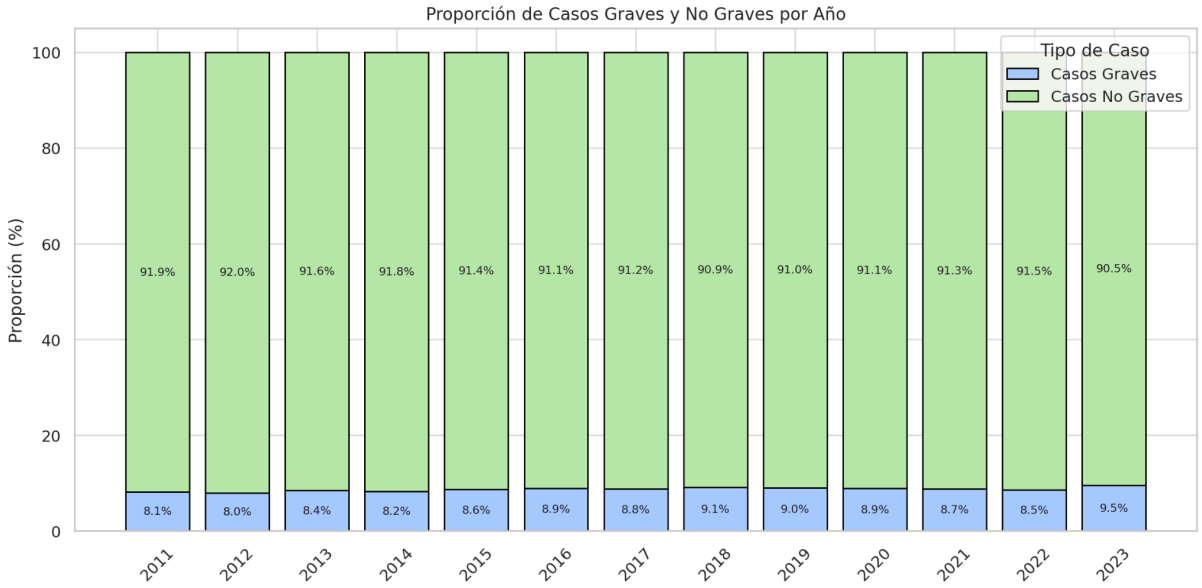
Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

IX.1.3 Gravedad

En este apartado se analiza la variable **Gravedad**, definida conforme a los criterios establecidos por el Ministerio de Salud y reportada por Mutual de Seguridad en las bases de datos originales. Esta variable permite clasificar los accidentes laborales como *graves* o *no graves*, según los lineamientos normativos vigentes. Para asegurar la trazabilidad con el universo completo de reportes institucionales, los indicadores presentados en esta sección se calculan sobre el **total de registros aceptados entre 2011 y 2023** (2.110.154 casos), previo a la aplicación de filtros o depuraciones posteriores empleadas en el análisis textual y el modelamiento predictivo.

Esta variable es clave para evaluar el impacto de los accidentes laborales tanto en términos de severidad como en la asignación de recursos para la prevención y mitigación de riesgos. En todos los años del período analizado, los casos no graves representan la mayoría de los registros, manteniéndose por sobre el 90 % del total anual.

Gráfico 2. Proporción de casos graves y no graves para los años 2011-2023



Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

El gráfico 2 presenta la proporción anual de casos graves y no graves registrados entre los años 2011 y 2023. Durante todo el período, los casos no graves concentran la mayoría de los registros, con porcentajes que se mantienen sobre el 90 %. La proporción de casos graves, en cambio, ha oscilado dentro de un rango estrecho, con valores mínimos en 2012 (7,97 %) y máximos en 2023 (9,52 %). A partir de 2016 se observa un leve aumento sostenido en la proporción de casos graves, que alcanza su mayor valor en el último año del período.

Tabla 8. Clasificación de accidentes por gravedad (2011–2023)

Año	Total accidentes	Total accidentes graves	Total accidentes no graves	Total accidentes graves(%)	Total accidentes no graves (%)
2011	143.308	11.648	131.660	8,13	91,87
2012	154.560	12.316	142.244	7,97	92,03
2013	160.633	13.522	147.111	8,42	91,58
2014	164.474	13.561	150.913	8,25	91,75
2015	164.912	14.240	150.672	8,63	91,37
2016	171.483	15.339	156.144	8,94	91,06
2017	174.779	15.438	159.341	8,83	91,17
2018	173.091	15.757	157.334	9,10	90,90
2019	183.445	16.507	166.938	9,00	91,00
2020	131.682	11.656	120.026	8,85	91,15
2021	150.562	13.171	137.391	8,75	91,25
2022	184.176	15.746	168.430	8,55	91,45
2023	153.049	14.567	138.482	9,52	90,48
Total	2.110.154	183.468	1.926.686	8.69	91.31

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

Cabe precisar que los valores presentados en este apartado corresponden al total de denuncias aceptadas entre 2011 y 2023 (N = 2.110.154), incluyendo tanto los accidentes calificados como graves (183.468 casos) como los no graves (1.926.686 casos). Por lo tanto, esta caracterización de la variable Gravedad se realiza previo a la aplicación de los filtros de elegibilidad y depuración descritos en secciones anteriores, y no debe confundirse con el universo reducido (N = 1.934.760) utilizado posteriormente para el análisis de texto y modelamiento de APG. Esto permite mantener la trazabilidad con el conjunto completo de registros institucionales y asegurar la comparabilidad con los criterios oficiales de clasificación.

IX.1.4 Calificación

La variable Calificación permite distinguir entre distintos tipos de eventos laborales según su origen y naturaleza: accidentes de trabajo (TB), accidentes de trayecto (TY) y enfermedades profesionales (EP). Esta clasificación es clave para comprender no solo el comportamiento del riesgo ocupacional, sino también las políticas institucionales de reconocimiento, fiscalización y prevención.

1. Predominio sostenido de los accidentes de trabajo (TB)

Los accidentes de trabajo (TB) representan la categoría mayoritaria en todo el período analizado. En 2011 se registraron 115.255 casos, cifra que alcanza un máximo de 129.855 en 2017 y luego se estabiliza en torno a los 103.000 a 118.000 casos en el trienio 2021–2023. Esto

refleja la persistencia de los riesgos operacionales directos en el lugar de trabajo como principal fuente de eventos calificados, a pesar de posibles mejoras en protocolos de seguridad.

Durante 2020 se observa una caída significativa en los registros TB (88.044 casos), coincidente con la pandemia por COVID-19 y sus efectos sobre la presencialidad y actividad económica. Posteriormente, la cantidad de eventos TB se recupera gradualmente.

2. Aumento progresivo de accidentes de trayecto (TY)

La categoría TY, que corresponde a accidentes ocurridos durante el trayecto entre el hogar y el trabajo, muestra una tendencia creciente desde 2011 (21.353 casos) hasta alcanzar su punto más alto en 2019 (43.950 casos), duplicando la cifra inicial en menos de una década.

Este incremento podría estar vinculado a:

- Mayor movilidad laboral y uso de transporte motorizado.
- Ampliación de los criterios de calificación.
- Mayor conocimiento por parte de los trabajadores sobre este tipo de cobertura.

Posteriormente, en 2020, se registra una contracción hasta 27.537 casos, también explicable por las medidas de confinamiento. Desde entonces, los valores vuelven a crecer, superando los 42.000 casos en 2022.

3. Crecimiento reciente en enfermedades profesionales (EP)

La categoría EP presenta cifras moderadas en la mayoría del período, con valores estables entre 2.200 y 3.200 casos anuales entre 2011 y 2019. Sin embargo, se observan dos rupturas significativas:

- En 2020, los casos EP se disparan a 10.891 registros, lo cual sugiere un efecto de la pandemia sobre el reconocimiento de enfermedades asociadas a condiciones laborales (como contagios o cuadros de salud mental).
- En 2022, se alcanza un máximo histórico de 17.260 casos, con un leve descenso posterior en 2023 (6.022 casos), lo cual puede reflejar cambios institucionales en la evaluación médica, mayor sensibilidad a los riesgos psicosociales o incluso ajustes en los criterios técnicos de calificación.

4. Casos sin calificación

La categoría “Sin Calificación” ha mantenido una proporción relativamente constante, fluctuando entre 4.200 y 8.100 casos anuales.

Tabla 9. Accidentes por tipo de calificación (2011–2023)

Año	Enfermedades Profesionales	Sin Calificación	Accidentes de Trabajo	Accidentes de Trayecto
2011	2.430	4.270	115.255	21.353
2012	2.249	5.052	123.267	23.992
2013	2.298	5.293	126.426	26.616
2014	2.360	6.006	127.979	28.129
2015	2.466	6.783	126.060	29.603
2016	2.555	7.862	128.923	32.143
2017	2.437	8.173	129.855	34.314
2018	3.183	8.192	127.798	33.918
2019	2.834	8.103	128.558	43.950
2020	10.891	5.210	88.044	27.537
2021	9.072	5.830	103.119	32.541
2022	17.260	6.033	118.269	42.614
2023	6.022	4.407	107.965	34.655
Total	66.057	81.214	1.551.518	411.365

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

IX.1.5 Días de reposo

La variable “rpss_dias_reposo” corresponde al número total de días de licencia médica otorgados como consecuencia de un accidente o enfermedad profesional. Esta variable constituye un indicador indirecto de la gravedad del caso, ya que un mayor número de días de reposo suele estar asociado a condiciones de mayor afectación funcional, recuperación prolongada o secuelas más complejas.

Al analizar los 2.110.154 registros disponibles entre los años 2011 y 2023, se construyó una distribución por rangos de duración del reposo, con los siguientes resultados

La duración del reposo médico asociado a accidentes o enfermedades profesionales constituye una medida indirecta de la gravedad funcional del evento, así como de su impacto en la capacidad laboral del trabajador. Para este análisis, los días de reposo fueron categorizados en cuatro tramos: 0 a 30 días, 31 a 180 días, 181 a 365 días y más de 365 días, y se analizó su distribución entre los años 2011 y 2023.

1. Predominio de eventos de corta duración

A lo largo de todo el período, la mayoría de los registros se concentra en el tramo 0 a 30 días, representando sistemáticamente más del 70 % de los casos anuales. Este patrón sugiere que la mayor parte de los eventos registrados corresponden a cuadros leves o moderados, con tiempos de recuperación breves. Por ejemplo, en 2011 se reportaron 93.846 eventos en este rango,

mientras que en 2022 la cifra aumentó a 99.395 casos, manteniendo su peso relativo como tramo dominante.

2. Estabilidad relativa en el rango medio (31 a 180 días)

El segundo tramo más frecuente corresponde a eventos que requieren entre 31 y 180 días de reposo, cuya participación anual se ha mantenido relativamente estable en torno a los 13.000 a 18.000 casos por año. Este grupo puede representar cuadros más complejos o con secuelas temporales, pero que en general permiten recuperación dentro del mismo año calendario.

3. Disminución progresiva de casos prolongados

Uno de los aspectos más destacables es la evolución de los casos de reposo prolongado, es decir, los tramos de 181 a 365 días y más de 365 días:

- En 2016 se registró el punto más alto para el tramo 181 a 365 días, con 3.158 casos.
- A partir de 2017 se observa una tendencia descendente, culminando en solo 893 casos en 2023, una reducción del 71,7 %.
- De forma aún más pronunciada, los casos con más de 365 días de reposo pasaron de 1.608 en 2016 a solo 13 en 2023, lo que representa una caída del 99,2 %.

Tabla 10. Distribución de accidentes por duración del reposo médico (2011–2023)

Año	0 a 30 días	31 a 180 días	181 a 365 días	Más de 365 días
2011	93.846	12.161	1.543	1.015
2012	98.998	12.736	1.722	1.050
2013	95.247	13.469	2.096	1.240
2014	96.391	13.945	2.153	1.144
2015	95.169	15.044	2.288	1.251
2016	99.259	17.363	3.158	1.608
2017	95.879	17.576	2.747	1.387
2018	89.982	17.757	2.792	1.212
2019	92.591	17.586	2.608	1.047
2020	72.254	13.857	2.135	580
2021	84.042	15.907	1.812	338
2022	99.395	18.286	1.941	369
2023	75.269	16.515	893	13
Total	**1.188.322**	**202.202**	**27.888**	**12.254**

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

Las variables construidas son las siguientes:

IX.1.6 Accidentes Potencialmente Graves y Fatales

La variable “APG” identifica aquellos accidentes que, sin haber sido clasificados formalmente como graves, presentan características que podrían haber derivado en consecuencias severas o fatales. Esta variable fue construida mediante técnicas de procesamiento de lenguaje natural (NLP), a partir del análisis semántico de los relatos y anamnesis registrados en las denuncias aceptadas entre 2011 y 2023.

La siguiente tabla muestra la distribución anual de accidentes laborales identificados como potencialmente graves (APG) entre los años 2011 y 2023, a partir del análisis semántico aplicado a relatos y anamnesis registrados en las denuncias aceptadas. La clasificación se realizó utilizando técnicas de procesamiento de lenguaje natural (NLP), que permitieron detectar configuraciones textuales asociadas a incidentes con alto riesgo, incluso cuando el accidente no fue clasificado formalmente como grave.

El criterio de clasificación se fundamenta en la identificación de patrones lingüísticos presentes en accidentes previamente catalogados como graves. Por ejemplo, si un accidente grave incluye términos como "choque" y "motocicleta", se pueden identificar otros registros con esas mismas características textuales como potencialmente graves, aun cuando no hayan producido consecuencias severas.

Tabla 11. Accidentes potencialmente graves (2011–2023)

Accidente Potencialmente Grave			
Año	**No**	**Si**	**Total**
2011	111.733	6.627	118.360
2012	123.314	6.441	129.755
2013	128.601	6.665	135.266
2014	130.967	7.105	138.072
2015	129.761	7.505	137.266
2016	132.055	8.964	141.019
2017	135.055	8.749	143.804
2018	133.506	8.717	142.223
2019	143.696	8.325	152.021
2020	102.849	6.793	109.642
2021	118.124	7.836	125.960
2022	148.558	10.674	159.232
2023	124.756	9.112	133.868
Total	**1.662.975**	**103.513**	**1.766.488**

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

Cabe señalar que el análisis de accidentes potencialmente graves (APG) se realizó sobre un subconjunto de 1.766.488 registros, correspondiente a los casos que, habiendo superado los

filtros iniciales de elegibilidad general, contaban con información textual suficiente y estructurada para ser procesada mediante técnicas de procesamiento de lenguaje natural (NLP). Esto implica que algunos registros fueron excluidos en esta etapa específica por carecer de relatos o anamnesis válidas, presentar inconsistencias lingüísticas o no cumplir con los criterios técnicos mínimos necesarios para aplicar el modelo de clasificación semántica. Esta diferencia respecto al universo elegible inicial (1.934.760 casos) responde, por tanto, a una depuración adicional orientada a garantizar la calidad del análisis textual y la validez de la variable APG construida.

En total, se identificaron 103.513 accidentes potencialmente graves entre 2011 y 2023, lo que representa una incidencia acumulada de 5,86 % sobre los 1.766.488 casos analizados. El número absoluto de APG presenta una tendencia creciente entre 2011 y 2022, con un máximo en 2022 (10.674 casos), seguido por una leve baja en 2023 (9.112 casos), pese a una disminución general en el total de accidentes reportados ese año.

La proporción relativa de APG respecto del total anual también muestra un crecimiento progresivo durante el período analizado. Este aumento podría deberse a una mejora en la sensibilidad del modelo de clasificación, a una mayor exposición estructural a riesgos complejos o a cambios en las condiciones laborales, especialmente en el contexto posterior a la pandemia. En particular, el año 2020 muestra una fuerte caída en el total de accidentes (109.642), atribuible a las restricciones sanitarias derivadas del COVID-19. Sin embargo, la proporción de APG se mantuvo en niveles similares a años anteriores (6.793 casos, equivalente al 6,20 %), lo que sugiere una persistencia de situaciones de alto riesgo a pesar de la menor actividad general.

La distribución de accidentes potencialmente graves según su calificación revela patrones consistentes con la estructura general de los accidentes laborales reportados en el período 2011–2023. Del total de 103.513 APG identificados mediante el análisis semántico, la mayoría corresponde a accidentes de trabajo, que representan el 61,7 % del total (63.842 casos). Le siguen los accidentes de trayecto, con un 22,4 % (23.157 casos), mientras que las enfermedades profesionales constituyen el 15,6 % (16.141 casos). Finalmente, se observan 373 registros sin calificación explícita (0,4 %), los cuales fueron detectados por el modelo a partir de patrones léxicos coincidentes, a pesar de la ausencia de codificación formal.

En términos anuales, la evolución muestra una tendencia de crecimiento sostenido de los APG en todas las categorías, con un aumento especialmente marcado en accidentes de trayecto y enfermedades profesionales a partir de 2020. Este fenómeno podría estar relacionado con cambios en la organización del trabajo, la movilidad y la exposición a riesgos post-pandemia. Además, destaca el año 2022 como el periodo con mayor número de APG en todas las categorías, especialmente en enfermedades profesionales (2.291 casos), lo que refuerza la hipótesis de que la identificación automática está capturando situaciones de riesgo que no siempre quedan reflejadas en la clasificación original.

En conjunto, esta clasificación por tipo de accidente permite dimensionar con mayor precisión qué categorías concentran los mayores riesgos potenciales y facilita la definición de estrategias de intervención focalizadas, tanto en prevención de accidentes de trayecto (por ejemplo, seguridad vial laboral), como en entornos de trabajo con alta exposición a enfermedades de origen ocupacional.

Tabla 12. Accidentes potencialmente graves según calificiación (2011–2023)

Año	Accidente Trayecto	Accidente Trabajo	Enfermedad profesional	Sin calificación	Total
2011	1.263	4.656	701	7	6.627
2012	1.280	4.562	589	10	6.441
2013	1.407	4.667	583	8	6.665
2014	1.477	4.954	665	9	7.105
2015	1.647	5.064	787	7	7.505
2016	1.892	5.693	1.370	9	8.964
2017	1.993	5.462	1.284	10	8.749
2018	1.956	5.267	1.481	13	8.717
2019	2.070	4.876	1.370	9	8.325
2020	1.374	3.632	1.773	14	6.793
2021	1.691	4.216	1.914	15	7.836
2022	2.673	5.627	2.291	83	10.674
2023	2.434	5.166	1.333	179	9.112
Total	**23.157**	**63.842**	**16.141**	**373**	**103.513**

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

Adicionalmente, dentro de la categoría APG se incluyen casos identificados como potencialmente fatales, a partir del análisis de bases de datos específicas de accidentes con desenlace mortal. En estos casos, se asigna un valor de 1 a la variable APG cuando el evento presenta condiciones que, bajo otras circunstancias, podrían haber resultado en una muerte efectiva. Esta identificación se basa en criterios técnicos definidos por las mutualidades, y no implica que el accidente haya provocado efectivamente una fatalidad, sino que evidencia configuraciones de riesgo extremo que deben ser consideradas en el diseño de acciones preventivas prioritarias.

IX.1.7 Enfermedades Crónicas

La variable denominada en la base como “CRONICA_FINAL” identifica si el paciente presenta una enfermedad crónica, según lo registrado en la anamnesis clínica. Se codifica como 1 cuando se detecta una condición crónica relevante en el relato, lo que permite reconocer casos de mayor complejidad médica y potencial impacto en la recuperación.

A continuación, se muestra la distribución anual de accidentes laborales que presentan condición crónica entre los años 2011 y 2023.

Tabla 13. Accidentes con enfermedades crónicas (2011–2023)

Año	No Crónica	Crónica	Total	% Crónica
2011	134.306	9.002	143.308	6,28 %
2012	143.817	10.743	154.560	6,95 %
2013	146.552	14.081	160.633	8,77 %
2014	148.664	15.810	164.474	9,61 %
2015	148.057	16.855	164.912	10,22 %
2016	152.477	19.006	171.483	11,08 %
2017	154.759	20.020	174.779	11,45 %
2018	152.312	20.779	173.091	12,00 %
2019	160.714	22.731	183.445	12,39 %
2020	116.127	15.555	131.682	11,81 %
2021	131.392	19.170	150.562	12,73 %
2022	159.991	24.185	184.176	13,13 %
2023	135.949	17.100	153.049	11,17 %

Fuente: Elaboración propia a partir de la información proporcionada por Mutual de seguridad.

1. Tendencia creciente sostenida (2011–2022)

Entre los años 2011 y 2022, se observa un incremento continuo en el número y proporción de casos con condición crónica. En 2011, estos representaban el 6,28 % del total de accidentes (9.002 casos), mientras que en 2022 alcanzan el 13,13 % (24.185 casos), más que duplicando su peso relativo en el conjunto de eventos registrados.

Este aumento puede deberse a múltiples factores, entre los que destacan:

- Mayor capacidad diagnóstica y seguimiento longitudinal de casos.
- Reconocimiento de patologías complejas y de evolución prolongada, especialmente en el ámbito psicosocial y musculoesquelético.
- Cambios normativos o institucionales que incentivan la trazabilidad de los casos más allá del episodio agudo inicial.

2. Punto de inflexión posterior a 2015

A partir de 2015, el porcentaje de casos crónicos supera por primera vez el 10 %, consolidando una nueva fase de comportamiento estructural del sistema. Esta ruptura puede estar vinculada a:

3. Comportamiento durante y después de la pandemia (2020–2023)

En 2020, producto de la pandemia por COVID-19, se observa una disminución general del total de eventos (131.682), pero los casos crónicos mantuvieron un nivel elevado (11,81 %), lo que sugiere una persistencia estructural de este tipo de casos, incluso en contextos de menor movilidad o exposición laboral.

En los años 2021 y 2022, el indicador sigue creciendo y alcanza su punto más alto en 2022 (13,13 %). Sin embargo, en 2023 se aprecia una ligera disminución en términos absolutos (17.100 casos) y relativos (11,17 %), lo que podría indicar una corrección posterior al efecto acumulado post-pandemia, cambios en criterios de calificación, o una mejora en el manejo preventivo temprano de casos.

Como conclusión el análisis exploratorio de las variables disponibles ha permitido identificar patrones relevantes sobre la evolución de los accidentes laborales en Chile entre 2011 y 2023. Se observan tendencias estructurales, como:

- La predominancia de grupos etarios intermedios y el aumento sostenido de trabajadores mayores de 55 años.
- La reducción de la brecha de género, junto con una creciente participación femenina.
- La persistencia de accidentes no graves como mayoría, aunque con un aumento relativo de los casos graves en los últimos años.
- La identificación automatizada de accidentes potencialmente graves o fatales, con una proporción acumulada de 5,86 % y una tendencia creciente hasta 2022.
- El alza progresiva de casos con enfermedades crónicas, consolidando nuevos desafíos en la recuperación y prevención de largo plazo.

Estos patrones sugieren que el sistema de registro ha mejorado en su capacidad de capturar datos clínicos y laborales más completos, lo que permite alimentar modelos predictivos más precisos. Dichos modelos pueden focalizarse en los sectores económicos, perfiles sociodemográficos y condiciones de riesgo más relevantes, fortaleciendo las estrategias de prevención y fiscalización.

X. Resultados

A continuación, se presentan los resultados del análisis realizado sobre los accidentes potencialmente graves (APG) ocurridos entre 2011 y 2023. El análisis se enfoca en la identificación de estos accidentes según el sector económico y la probabilidad de que un accidente registrado sea clasificado como grave. Este estudio utiliza un modelo predictivo basado en datos históricos para evaluar las tendencias y riesgos en diferentes sectores, lo que permite priorizar las intervenciones y estrategias de prevención. Los resultados proporcionan información detallada sobre la distribución y características de los accidentes potencialmente graves, con el objetivo de guiar las futuras políticas de seguridad y salud laboral.

X.1 Resultados analíticos para accidentes potencialmente graves y/o fatales (APG)

Para la identificación de los accidentes potencialmente graves y/o fatales (APG), los datos fueron tratados conforme al procedimiento revisado en puntos anteriores. Utilizando técnicas de procesamiento de lenguaje natural (NLP) para clasificar los relatos de accidentes según criterios definidos en la Guía del ISP. Esta metodología permitió procesar grandes volúmenes de información no estructurada y transformarla en variables categorizadas con validez técnica, asegurando la trazabilidad y consistencia en la aplicación de los criterios de gravedad.

En promedio, un 5,85% de los accidentes reportados entre 2011 y 2023 fueron clasificados como potencialmente graves. El año con mayor proporción fue 2023, alcanzando un 6,81%, mientras que el valor más bajo se registró en 2013, con un 4,93%. Estos valores reflejan una tendencia relativamente estable en la proporción de APG durante el período analizado, con una leve alza en los últimos años, lo que sugiere un cambio estructural en las condiciones de riesgo, posiblemente asociado a factores externos como la pandemia o transformaciones en los procesos laborales.

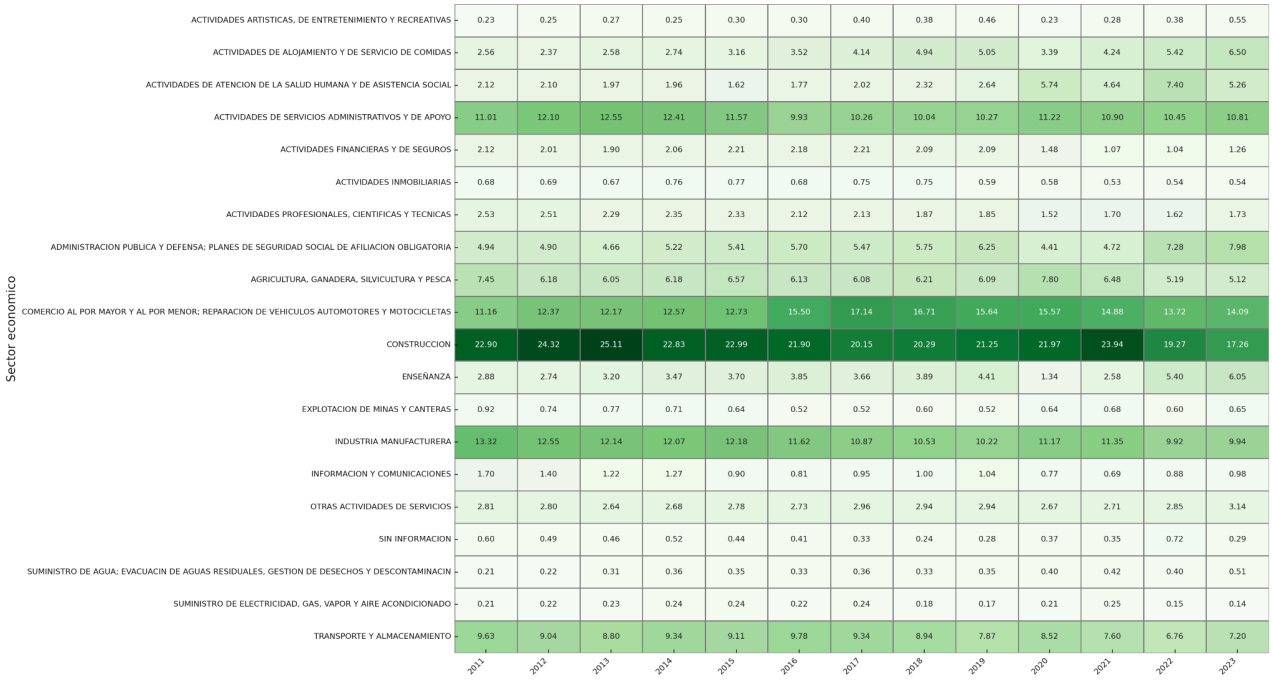
Para evaluar la probabilidad de ocurrencia de APG según sector económico, se calcularon **tasas de incidencia** anuales, entendidas como la proporción de APG sobre el total de accidentes registrados en cada sector y año. Tal como se mencionó anteriormente, este cálculo corresponde a:

$$P(\text{Accidente}) = \frac{\text{Total de accidentes en el sector economico}}{\text{Total de accidentes registrados en todos los sectores}}$$

El análisis muestra que el sector Construcción concentró sistemáticamente la mayor proporción de APG a lo largo del período, con un promedio calculado entre los años 2011 y 2023 de 17,13%, seguido por Comercio al por mayor y al por menor; Reparación de vehículos automotores y motocicletas (13,65%), y la Industria Manufacturera (12,51%). En contraste, los

sectores con menor proporción promedio de APG fueron Suministro de electricidad, gas, vapor y aire acondicionado (0,20%) y Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación (0,40%). Estos resultados se resumen a continuación:

Fig. 3. Tasa de incidencia promedio por sector (promedios años 2011-2023)



Fuente: Elaboración propia.

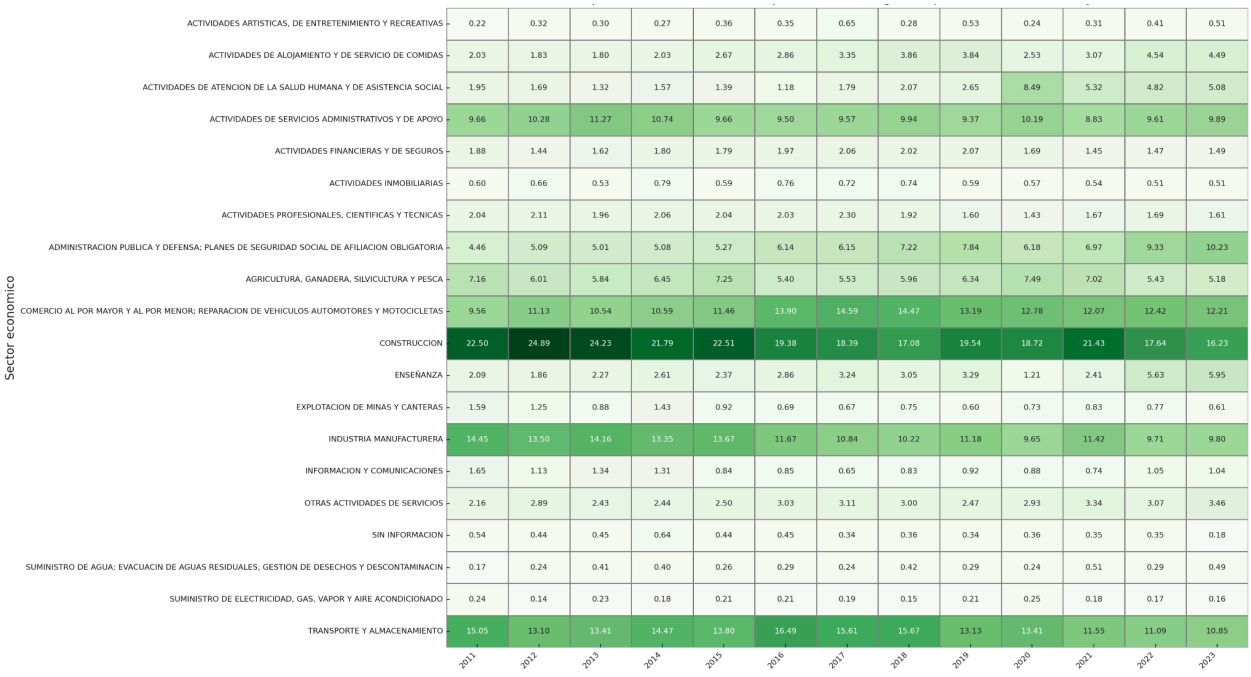
Por otro lado, se analiza la probabilidad condicional asociada a que, dada la ocurrencia de un accidente, este sea potencialmente grave según el sector económico. Es decir, se calcula:

$$P(\text{APG}|\text{Accidente})$$

Para ello, se sigue el esquema presentado en la Figura 3, identificando los Accidentes Potencialmente Graves (APG) y calculando sus respectivas probabilidades con respecto a su frecuencia en cada sector económico.

Es importante interpretar esta probabilidad como la frecuencia relativa de APG dentro del total de accidentes ocurridos en cada sector, es decir, dado que ocurre un accidente, cuál es la probabilidad de que sea potencialmente grave.

Fig. 4. Probabilidad de que dado que ocurre un accidente, este sea un APG por sector económico (promedios años 2011-2023)



Fuente: Elaboración propia.

Al realizar el análisis agregado para el período 2011-2023, se encuentra que los sectores con la mayor probabilidad promedio entre esos años de que un accidente sea potencialmente grave, dado que ocurre un accidente, son:

- **Construcción (20.33%)**
- **Transporte y almacenamiento (13.66 %)**
- **comercio al por mayor y por menor, reparación de vehículos automotores y motocicletas (12.22 %)**

Es relevante aclarar que estos sectores no necesariamente tienen más accidentes ; sino que, cuando ocurre uno, este tiene una mayor probabilidad de ser grave.

Esta interpretación es clave para diferenciar entre frecuencia de accidentes y gravedad de los eventos, proporcionando una perspectiva más precisa sobre los riesgos asociados a cada sector económico. Los resultados se muestran a continuación:

Tabla 14. Probabilidad de que ocurra un accidente sea un APG por sector económico (promedios años 2011-2023)

Sector económico	Probabilidad (%) promedio de que al ocurrir un accidente este sea APG (2011–2023)
Actividades artísticas, de entretenimiento y recreativas	0,37
Actividades de alojamiento y de servicio de comidas	2,99
Actividades de atención de la salud humana y de asistencia social	3,02
Actividades de servicios administrativos y de apoyo	9,89
Actividades financieras y de seguros	1,75
Actividades inmobiliarias	0,62
Actividades profesionales, científicas y técnicas	1,88
Administración pública y defensa; planes de seguridad social de afiliación obligatoria	6,54
Agricultura, ganadería, silvicultura y pesca	6,24
Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	12,22
Construcción	20,33
Enseñanza	2,99
Explotación de minas y canteras	0,9
Industria manufacturera	11,82
Información y comunicaciones	1,02
Otras actividades de servicios	2,83
Sin información	0,4
Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación	0,33
Suministro de electricidad, gas, vapor y aire acondicionado	0,19
Transporte y almacenamiento	13,66

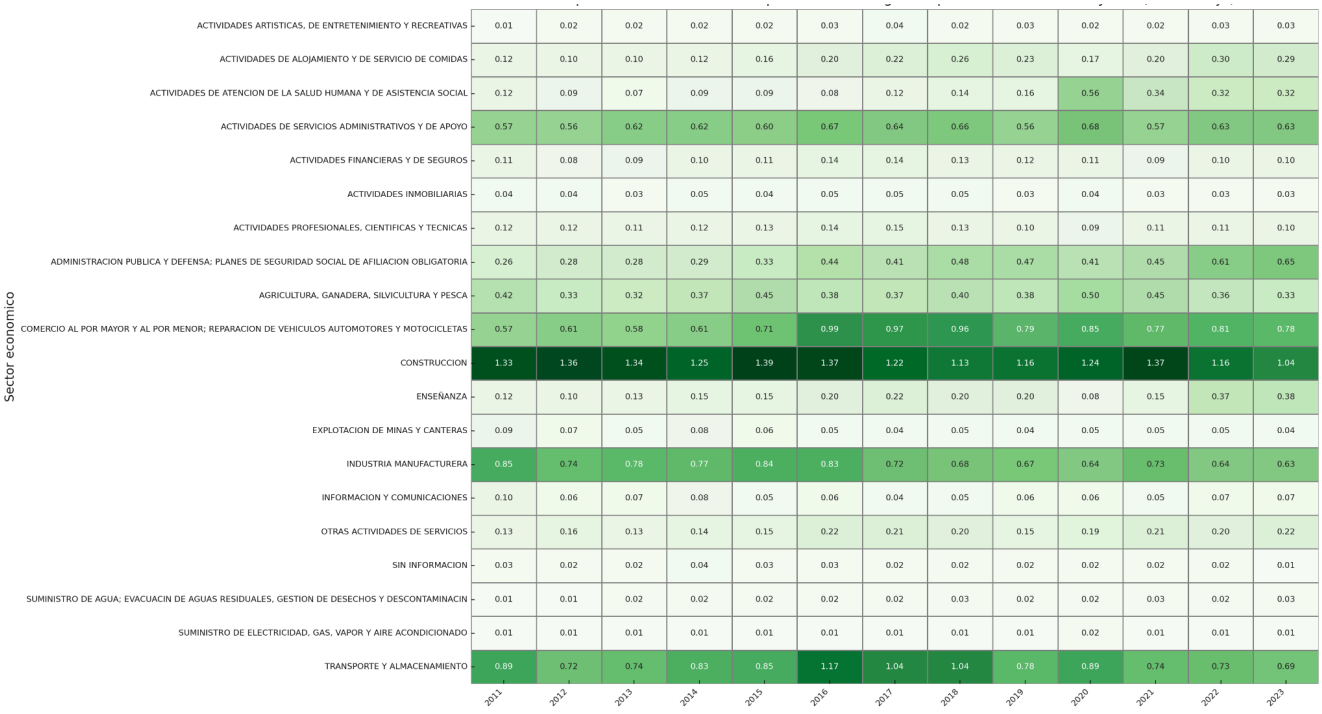
Fuente: Elaboración propia

Para calcular las probabilidades de que un accidente sea potencialmente grave, se multiplica la probabilidad de ocurrencia por la probabilidad de que, dado que ocurre un accidente, este sea un APG para cada sector "j". Es decir:

$$P(\text{APG en el sector } j) = P(\text{APG}|\text{Accidente})_j \cdot P(\text{Accidente})_j$$

El resultado se muestra en el siguiente mapa de calor, donde el sector de la construcción, transporte y comercio concentran las probabilidades más altas:

Fig.5. Probabilidad de que ocurra un accidente potencialmente grave (APG) por sector económico



Fuente: Elaboración propia.

Una vez obtenidas las probabilidades, se distribuyen para cada actividad económica, estas se asignan individualmente a cada accidente registrado según el sector económico correspondiente. Luego de re-parametrizar dichas probabilidades según la Guía del ISP, el cual define las probabilidades según los siguientes criterios:

$$P(D) = \begin{cases} 1, & \text{si } P \leq P_{bajo} \\ 2, & \text{si } P_{bajo} < P \leq P_{medio} \\ 4, & \text{si } P > P_{medio} \end{cases}$$

Para llevar a cabo la parametrización, se analizó la distribución de probabilidades de ocurrencia por sector económico, bajo la siguiente segmentación (en tres percentiles):

- Nivel Bajo (valor 1): Corresponde a actividades económicas donde la tasa de incidencia se ubica por debajo del percentil 50 de la distribución anual
- Nivel Medio (Valor 2): Corresponde actividades donde la tasa de incidencia se encuentra en el percentil 50 y 75
- Nivel Alto (Valor 4): Corresponde aquellas actividades con tasas de incidencia superiores al percentil 75, donde están la mayor probabilidad de accidentes.

Este enfoque permite aplicar la escala de riesgo definida en la Guía del ISP (2013), transformando las frecuencias observadas en niveles estandarizados de riesgo operativo, facilitando su interpretación y priorización dentro de los sistemas de gestión de seguridad laboral, obteniendo la siguiente distribución de probabilidades reparametrizadas:

Tabla 15. Distribución de los registros obtenidos según diferentes valores de probabilidad reparametrizada de daño.

Año	Totales probabilidad reparametrizada			Total
	1	2	4	
2011	2.506	3.706	1.408	7.620
2012	4.152	2.822	88	7.063
2013	4.232	2.796	95	7.124
2014	5.931	1.231	251	7.412
2015	3.033	1.458	3.448	7.939
2016	248	1.636	7.251	9.135
2017	3.817	1.388	3.351	8.556
2018	2.537	1.204	4.813	8.555
2019	5.095	2.541	41	7.677
2020	3.855	2.431	2.257	8.543
2021	4.508	644	3.108	8.260
2022	4.086	2.718	1.648	8.452
2023	4.843	900	2.487	8.229
Total general	**48.844**	**25.474**	**30.246**	**104.563**

Fuente: Elaboración propia.

Estos valores se relacionan con la severidad del daño, el cual se ha calculado según la distribución de los días de reposo. En este sentido, se presenta que un accidente puede causar más daño que otro en función a que necesita una mayor cantidad de días para recuperarse. Esta variable se calcula según la siguiente ecuación:

$$D = \begin{cases} 1, & \text{si } 0 \leq \text{días_reposo} < 13 \\ 2, & \text{si } 13 \leq \text{días_reposo} \leq 30 \\ 4, & \text{si días_reposo} > 30 \text{ o hay herramienta o enfermedad crónica involucrada} \end{cases}$$

Se añaden las variables de herramientas o enfermedad crónica como un factor que aumenta el daño probable. El resultado se muestra en la siguiente tabla:

Tabla 16. Distribución de los registros obtenidos según severidad del daño.

Año	Severidad del daño			Total
	1	2	4	
2011	104.334	13.236	612	118.182
2012	115.300	13.963	697	129.960
2013	120.846	14.294	753	135.893
2014	123.877	13.756	736	138.369
2015	123.434	13.384	832	137.650
2016	124.515	14.643	779	139.937
2017	127.635	14.274	880	142.789
2018	126.813	13.513	879	141.205
2019	137.922	13.177	890	151.989
2020	92.432	15.219	719	108.370
2021	110.708	12.992	765	124.465
2022	141.789	13.982	832	156.603
2023	118.224	13.104	758	132.086
Total	**1.567.829**	**179.537**	**10.132**	**1.757.498**

Fuente: Elaboración propia.

Con estos valores, se puede calcular el Valor Esperado de la Pérdida (VEP) que en este caso será un valor probable (VEPP), dado que se basa en el cálculo de APG, por lo que se define su cálculo cómo:

$$VEPP = \text{Probabilidad de APG} \times \text{Severidad del daño (D)}$$

Donde el valor de probabilidad de APG se encuentra reparametrizado según se explicó anteriormente. Esto implica que el valor de VEPP tomará cuatro valores que definirán el nivel de riesgo. El resultado se muestra en la siguiente tabla, donde el grueso de los accidentes se concentra en riesgos de tipo trivial, tolerable y moderado:

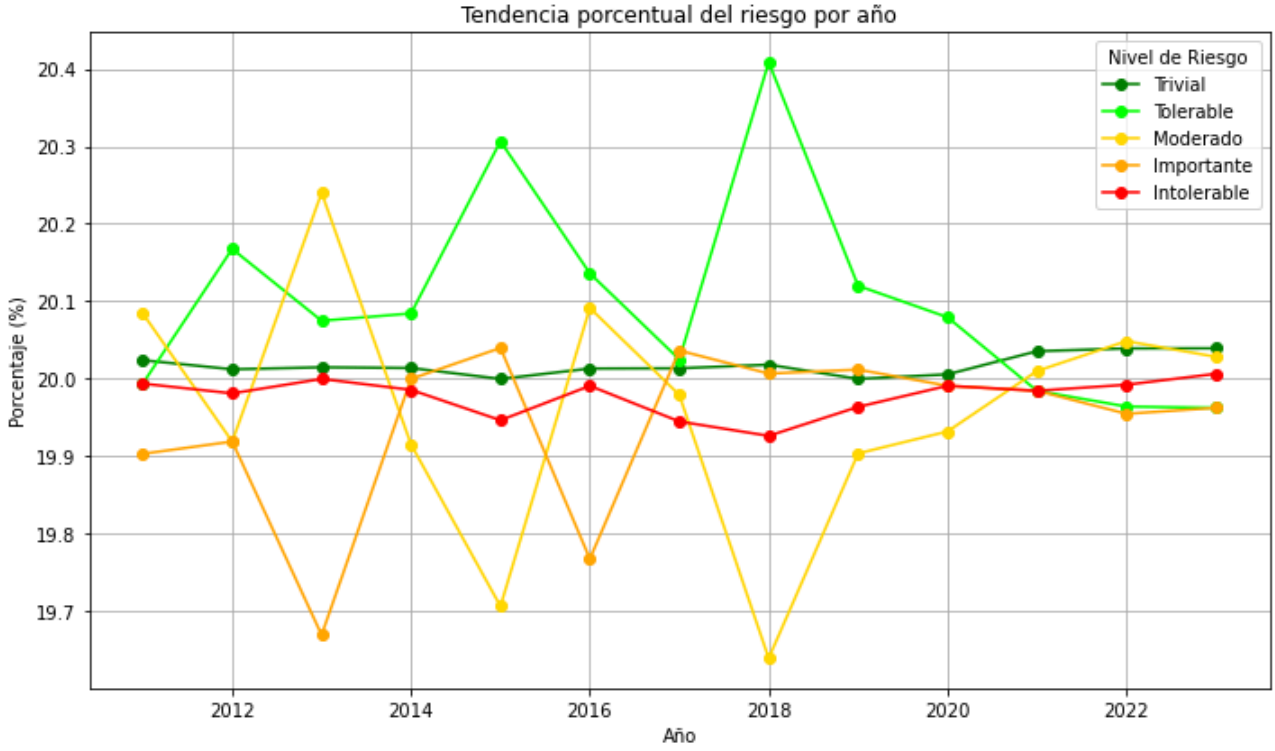
Tabla 17. Distribución de los registros obtenidos según riesgo.

Año	Riesgo					Total
	Trivial	Tolerable	Moderado	Importante	Intolerable	
2011	1.327	1.325	1.331	1.319	1.325	6.627
2012	1.289	1.299	1.283	1.283	1.287	6.441
2013	1.334	1.338	1.349	1.311	1.333	6.665
2014	1.422	1.427	1.415	1.421	1.420	7.105
2015	1.501	1.524	1.479	1.504	1.497	7.505
2016	1.794	1.805	1.801	1.772	1.792	8.964
2017	1.751	1.752	1.748	1.753	1.745	8.749
2018	1.745	1.779	1.712	1.744	1.737	8.717
2019	1.665	1.675	1.657	1.666	1.662	8.325
2020	1.359	1.364	1.354	1.358	1.358	6.793
2021	1.570	1.566	1.568	1.566	1.566	7.836
2022	2.139	2.131	2.140	2.130	2.134	10.674
2023	1.826	1.819	1.825	1.819	1.823	9.112
Total	**20.722**	**20.804**	**20.662**	**20.646**	**20.679**	**103.513**

Fuente: Elaboración propia.

El número total es menor debido a completitud de datos (es decir, no todos pueden ser reparametrizados debido a que no todos poseen la información disponible al realizar el cruce entre probabilidad y tipo de actividad económica) Al analizar la tendencia porcentual, resalta la tendencia positiva de crecimiento de los accidentes potencialmente graves con riesgo moderado, mientras que los de riesgo importante se mantienen con una tendencia horizontal hasta el 2019 que es donde empiezan a bajar, aumentando los de tipo tolerable. Sin embargo, cabe destacar que existe una tendencia promedio creciente para los accidentes con riesgo intolerable, pasando de un 3.9% a un 7.7% entre 2012 y 2019, retornando a la tendencia a la baja en años posteriores. El 2020 actúa como un año crítico en los cambios de tendencias para todos los tipos de riesgo.

Gráfico 3. Tendencia porcentual del riesgo por año.



Fuente: Elaboración propia.

Al analizar por sector se observa que el sector de la construcción presenta el mayor número de accidentes potencialmente intolerables (8.737), es decir, eventos con alta probabilidad de ocurrencia y una alta probabilidad de generar daño considerable. En contraste, el sector de industria manufacturera, aunque tiene el tercer mayor volumen de accidentes totales (11.819), estos se concentran principalmente en el nivel de riesgo moderado (3.243 casos), junto con niveles relevantes de riesgo importante e intolerable. Además, manufactura es el segundo sector con mayor cantidad de registros de riesgo moderado, después de "Transporte y almacenamiento". Un punto relevante es que en el sector Transporte, aproximadamente el 28 % del total de accidentes se ubica en la categoría de riesgo importante, lo que refuerza su criticidad. Los detalles por sector se muestran en la siguiente tabla.

Tabla 18. Distribución de los registros obtenidos según riesgo para los diferentes sectores económicos

Sector económico	Riesgo					Total
	Trivial	Tolerable	Moderado	Importante	Intolerable	
Actividades artísticas, de entretenimiento y recreativas	400	15	0	0	0	415
Actividades de alojamiento y de servicio de comidas	1.417	1.382	281	169	75	3.324
Actividades de atención de la salud humana y de asistencia social	1.043	1.389	442	200	97	3.171
Actividades de servicios administrativos y de apoyo	1.176	3.010	2.605	1.844	1.624	10.259
Actividades financieras y de seguros	1.253	432	58	33	6	1.782
Actividades inmobiliarias	584	45	2	0	0	631
Actividades profesionales, científicas y técnicas	1.311	475	84	41	15	1.926
Administración pública y defensa; planes de seguridad social de afiliación obligatoria	632	3.034	1.577	1.117	636	6.996
Agricultura, ganadería, silvicultura y pesca	1.402	2.941	980	665	543	6.531
Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas	1.429	1.606	3.953	3.030	2.747	12.765
Construcción	2.648	608	2.192	6.728	8.737	20.913
Enseñanza	1.251	1.471	347	185	60	3.314
Explotación de minas y canteras	747	101	25	3	0	876
Industria manufacturera	1.462	1.953	3.243	2.546	2.615	11.819
Información y comunicaciones	873	139	23	15	0	1.050
Otras actividades de servicios	1.406	1.119	242	117	61	2.945
Sin información	370	6	2	0	0	378
Suministro de agua; evacuación de aguas residuales, gestión de desechos y descontaminación	331	10	0	0	0	341
Suministro de electricidad, gas, vapor y aire acondicionado	193	1	0	0	0	194
Transporte y almacenamiento	794	1.067	4.606	3.953	3.463	13.883
Total General	**20.722**	**20.804**	**20.662**	**20.646**	**20.679**	**103.513**

Fuente: Elaboración propia.

Los resultados se alinean con los criterios definidos en la Guía del ISP (2013), que establece niveles de probabilidad y severidad como herramientas clave para la evaluación del riesgo. Al

aplicar una clasificación basada en percentiles para reparametrizar las probabilidades de accidentes potencialmente graves (APG), se logra una diferenciación más precisa entre sectores económicos y períodos específicos. Esto permite priorizar aquellos sectores con mayores niveles de riesgo y orientar estrategias preventivas hacia actividades con alta probabilidad y severidad de accidentes, como la construcción y la industria manufacturera.

El análisis evidencia que sectores como construcción, transporte, manufactura y comercio presentan riesgos significativos tanto en frecuencia como en severidad de accidentes. En este contexto, se recomienda implementar políticas públicas focalizadas en estos sectores, reforzando la supervisión de condiciones laborales y ampliando los programas de capacitación en seguridad ocupacional.

El año 2020 destaca como un punto de inflexión, con cambios notables en las tendencias de riesgo, lo que sugiere la necesidad de fortalecer los protocolos de seguridad ante contingencias globales, como la pandemia, que modifican las dinámicas laborales y exponen a los trabajadores a nuevos escenarios de riesgo.

Desde una perspectiva de mitigación, los esfuerzos deben centrarse en los tipos de accidentes con mayor severidad y en las actividades económicas con alta incidencia de accidentes graves. La construcción y el comercio sobresalen tanto por la cantidad como por la gravedad de los eventos reportados, lo que requiere intervenciones preventivas específicas y de alta prioridad. En contraste, sectores como atención de la salud o actividades artísticas, aunque con menor incidencia general, podrían beneficiarse de estrategias dirigidas a reducir riesgos particulares que afectan su desempeño en materia de seguridad.

X.2 Análisis econométrico de accidentes potencialmente graves y severidad del daño.

X.2.1 Modelo de variable dependiente limitada: Logit

Para identificar los factores que incrementan la probabilidad de que un accidente laboral sea clasificado como potencialmente grave (APG), se estimó un modelo econométrico de tipo logit. Esta decisión metodológica se fundamenta en que la variable dependiente es binaria y el interés radica tanto en la identificación de variables explicativas significativas como en la evaluación del riesgo relativo entre grupos de trabajadores.

El modelo incorpora variables sociodemográficas (edad, género), clínicas (presencia de enfermedad crónica), y características del evento (tipo de calificación). La estimación se realizó sobre una base nacional de accidentes laborales ocurridos entre 2011 y 2023, compuesta por más de 1,4 millones de registros.

Con el objetivo de identificar los factores asociados a la ocurrencia de accidentes potencialmente graves (APG), se estimó un modelo econométrico tipo probit, el cual resulta adecuado para modelar una variable dependiente dicotómica (1: APG, 0: no APG), asumiendo una distribución normal acumulada en los errores.

Este tipo de especificación permite interpretar los efectos marginales como probabilidades y facilita una evaluación estadística de las variables explicativas en términos de riesgo relativo.

La especificación estimada fue la siguiente:

$$\Pr(APG_i = 1) = \Phi(\beta_0 + \beta_1 \cdot \text{genero}_i + \beta_2 \cdot \text{crónica}_i + \beta_3 \cdot \text{edad_grupo}_i + \beta_4 \cdot \text{calificación}_i)$$

Donde

- Φ es la función de distribución normal acumulada,
- genero_i indica si el trabajador es mujer,
- cronica_i representa presencia de enfermedad crónica,
- edad_grupo_i agrupa en tramos etarios (<30, 30–44, 45–59, 60+),
- calificacion_i corresponde al tipo de evento reportado (Enfermedad Profesional, Accidente de Trabajo, Accidente de Trayecto, Sin Calificación).

La siguiente tabla muestra los coeficientes estimados mediante regresión *logit*, que en esta etapa se considera equivalente funcional al *probit* en términos de dirección e interpretación de los efectos, y que permite una mayor estabilidad numérica para esta base de datos.

Tabla 19. Resultados del modelo logit para la probabilidad de accidente potencialmente grave (APG)

Variable	Coeficiente	Error estándar	z-valor	p-valor	OR	IC 95% inferior	IC 95% superior
Constante	-0.9103	0.011	-86.16	0,000	0.4	0,394	0,411
Género femenino	-0.0656	0.004	-18.27	0,000	0.94	0,929	0,944
Enfermedad crónica	0.1464	0.006	26.09	0,000	1.16	1,144	1,171
Edad 30-44 años	0.0808	0.004	19.74	0,000	1.08	1,076	1,093
Edad 45-59 años	0.2943	0.004	65.82	0,000	1.34	1,332	1,353
Edad 60+ años	0.5601	0.007	76.13	0,000	1.75	1,734	1,774
Calificación: Sin info	0.5232	0.013	39.75	0,000	1.69	1,650	1,729
Calificación: TB	0.6875	0.01	67.06	0,000	1.99	1,969	2,011
Calificación: TY	1.012	0.011	94.33	0,000	2.75	2,710	2,789

Fuente: Elaboración propia,

Todos los efectos reportados son estadísticamente significativos al 1%, y los intervalos de confianza son estrechos, lo que indica una alta precisión de las estimaciones.

El OR (Odds Ratio) o razón de probabilidades es una medida que se obtiene al exponenciar el coeficiente de un modelo logit (logístico). Representa la proporción de probabilidades de que ocurra un evento (en este caso, que un accidente sea potencialmente grave) frente a que no ocurra, dado un cambio unitario en la variable explicativa, manteniendo constantes las demás variables.Por ejemplo:

- Un OR > 1 indica que la variable aumenta la probabilidad del evento.
- Un OR < 1 indica que la variable reduce la probabilidad del evento.
- Un OR = 1 implica que no tiene efecto.

En el caso del análisis presente, la constante del modelo (OR = 0.40) refleja que, en ausencia de todos los factores de riesgo, la probabilidad base de que un accidente sea clasificado como APG es baja, lo que se condice con el hecho de que la mayoría de los registros corresponden a eventos no graves.

Los resultados muestran que los accidentes de trayecto (calificacion_TY) presentan una probabilidad 2,75 veces (OR=2,75) mayor de ser considerados APG en comparación con enfermedades profesionales (categoría base), mientras que los accidentes de trabajo (calificacion_TB) presentan un riesgo 1,99 veces mayor. Los registros sin información de calificación muestran un riesgo 69% mayor de ser clasificados como APG en comparación con enfermedades profesionales.

La edad también presenta una relación creciente con el riesgo: personas entre 45 y 59 años tienen un 34% más de probabilidad de sufrir un APG respecto a menores de 30 años, y los mayores de 60 años presentan un incremento de 75% .En el tramo de 30 a 44 años, la probabilidad de APG se incrementa en torno a 8% respecto a menores de 30 años.

La presencia de una enfermedad crónica (CRONICA_FINAL) se asocia con un incremento de 16% en el riesgo de APG. En contraste, ser mujer (genero_f = 1) se relaciona con una reducción del 6,4%[11] en la probabilidad de ocurrencia de este tipo de eventos, manteniendo constantes las demás variables.

Todos los efectos reportados son estadísticamente significativos al 1% y los intervalos de confianza del 95% se presentan en la tabla de resultados. El modelo converge correctamente y muestra un ajuste razonable, dada la distribución altamente desbalanceada de la variable dependiente.

[11] En variables con coeficiente negativo, el OR se interpreta como una reducción relativa de la probabilidad del evento. Por ejemplo, para **Género femenino**, el OR = exp(-0,0656) ≈ 0,936 indica que la probabilidad es un 6,4% menor respecto al grupo base. Matemáticamente, esto se calcula como **1 − OR = reducción porcentual** cuando OR < 1.

Magnitud de efecto y jerarquía de riesgos

Los resultados revelan patrones sistemáticos y estadísticamente robustos en la determinación del riesgo de APG. El factor de mayor peso es el tipo de calificación del accidente. En comparación con los eventos calificados como enfermedades profesionales (categoría base):

- Un accidente de trayecto incrementa en 175% la probabilidad de APG (OR = 2,75).
- Un accidente de trabajo eleva el riesgo en 99% (OR = 1.99).

Estos hallazgos reflejan que la forma en que se tipifica el accidente es un determinante estructural del riesgo, y probablemente captura aspectos relacionados con el contexto del suceso, la severidad del daño y los procedimientos de notificación.

La edad también emerge como un factor significativo. En relación con los trabajadores menores de 30 años:

- Quienes tienen entre 45 y 59 años presentan un 34% más de probabilidad de sufrir un APG.
- En el grupo de 60 años y más, el riesgo es 75% superior (OR = 1,75).

Este patrón sugiere que los trabajadores mayores, además de estar potencialmente más expuestos a riesgos acumulativos, podrían presentar mayor fragilidad fisiológica ante eventos similares.

Otro hallazgo relevante es el efecto de la presencia de enfermedades crónicas. Quienes declaran condiciones crónicas presentan un 15,8% más de probabilidad de clasificar el accidente como APG (OR = 1,16), lo que podría deberse a una menor capacidad de recuperación o mayor vulnerabilidad al daño.

Finalmente, el modelo muestra que ser mujer se asocia con una probabilidad 6,4% menor de sufrir un APG (OR = 0,936), lo cual podría estar vinculado a diferencias en los tipos de funciones desempeñadas, exposición a tareas de alto riesgo o factores de reporte, aunque este punto requiere mayor indagación cualitativa.
La tabla permite establecer una jerarquía de variables según su impacto relativo sobre la probabilidad de sufrir un accidente potencialmente grave (APG).

Tabla 20. Razones de probabilidad (Odds Ratio) del modelo logit para accidentes potencialmente graves (APG)

Variable	OR	IC 95% inferior	IC 95% superior	Interpretación breve
calificación_TY	2.75	2,71	2,789	Máximo riesgo: >175% vs. Enfermedad Profesional
calificación_TB	1.99	1,969	2,011	>99% más riesgo que enfermedad profesional
edad_grupo_60+	1.75	1,734	1,774	>75% más probabilidad que menores de 30 años
edad_grupo_45-59	1.34	1,332	1,353	>34% más riesgo
CRONICA_FINAL	1.16	1,144	1,171	>15.8% más riesgo si hay enfermedad crónica
edad_grupo_30-44	1.08	1,076	1,093	>8.4% más probabilidad respecto a menores de 30
genero_f	0.94	0,929	0,944	Ser mujer → 6.4% menor riesgo de APG

Fuente: Elaboración propia.

Para evaluar la capacidad del modelo logit en términos de clasificación individual, se aplicó una validación cruzada de 5 pliegues, utilizando como métrica de desempeño el área bajo la curva ROC (AUC). Esta medida permite estimar la capacidad del modelo para distinguir correctamente entre eventos de tipo APG (positivos) y no APG (negativos), independientemente del umbral de clasificación elegido.

Los resultados obtenidos se presentan a continuación:

Tabla 21. Resultados de validación cruzada (5-fold) del modelo logit

	AUC
1	0,55
2	0,558
3	0,55
4	0,571
5	0,591
Promedio	**0,564**

Fuente: Elaboración propia.

El AUC promedio obtenido es de 0,564, lo cual indica una capacidad discriminativa modesta pero superior al azar (AUC = 0,5). Este resultado es consistente con lo esperado en contextos con las siguientes características:

- La variable dependiente (APG) es poco frecuente en la base de datos.
- Existe un fuerte desbalance de clases, donde los casos positivos representan un bajo porcentaje del total.
- El modelo logit impone restricciones lineales y aditivas, lo que puede limitar su capacidad para capturar relaciones complejas o interacciones no lineales entre variables.

Pese a esta limitación, el modelo logra identificar patrones sistemáticos de riesgo, y sus coeficientes son estadísticamente robustos, por lo que su validez explicativa es alta, aunque su utilidad como clasificador predictivo es limitada.

La validación cruzada indica que el modelo logit ofrece un punto de partida valioso para comprender los factores de riesgo, pero que su rendimiento como clasificador predictivo es insuficiente para aplicaciones de priorización o monitoreo automatizado. En consecuencia, se justifica avanzar hacia modelos basados en aprendizaje automático, complementados con técnicas de balanceo de clases, con el fin de:

- Aumentar la precisión de la clasificación individual.
- Mejorar la detección temprana de eventos de alta severidad.
- Reducir falsos negativos (accidentes graves no identificados).

X.2.2 Modelos de aprendizaje automático: Random Forest

El modelo predictivo Random Forest fue aplicado para la clasificación de accidentes potencialmente graves (APG), obteniendo un desempeño robusto a nivel general, con una precisión global del 94%. Este valor evidencia una alta capacidad del algoritmo para discriminar entre eventos potencialmente graves y no graves, en un contexto de desbalance de clases.

No obstante, al desagregar por clase, el modelo presentó un recall de 63% y un F1-score de 59% para los casos positivos (APG), lo cual indica que si bien el modelo logra identificar una parte significativa de los accidentes potencialmente graves, aún persisten casos que no son correctamente clasificados. Esto es esperable dada la baja proporción de eventos APG respecto del total (5,86%), lo que condiciona el rendimiento en la clase minoritaria.

En cuanto a la importancia de las variables, la variable “rpss_dias_reposo” concentra el mayor peso en el modelo, con una importancia relativa de 0,937, lo que sugiere una alta capacidad discriminativa. Le siguen, con valores considerablemente menores, las variables edad (0,038), calificación (0,018), género (0,004) y crónica_final (0,002). Esta jerarquía de importancia respalda la noción de que los días de reposo constituyen un proxy efectivo de la severidad potencial del evento.

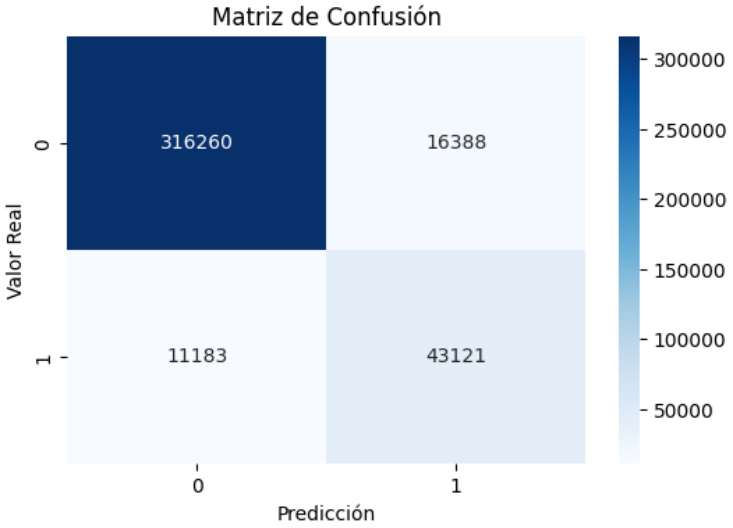
Tabla 22. Principales resultados del Rondom Forest

Variable	Importancia
rpss_dias_reposo	0,937385
edad	0,037974
calificacion	0,018154
pcnt_genero	0,004064
CRONICA_FINAL	0,002423

Fuente: Elaboración propia.

La matriz de confusión del modelo (ver gráfica) muestra un alto rendimiento en la clasificación de la clase negativa, con una precisión del 97% (316.260 verdaderos negativos). No obstante, se identificaron 16.388 falsos positivos (casos clasificados erróneamente como APG) y 11.183 falsos negativos (casos APG no detectados). Este último valor representa un 5% del total real de eventos potencialmente graves, lo cual es relevante desde una perspectiva preventiva, dado que implica la omisión de casos que podrían derivar en consecuencias severas. Estos resultados evidencian la necesidad de continuar ajustando los umbrales de decisión y/o aplicar estrategias de balanceo de clases, con el fin de mejorar la sensibilidad del modelo hacia los casos críticos sin comprometer significativamente su precisión general.

Gráfico 4. Matriz de confusión.

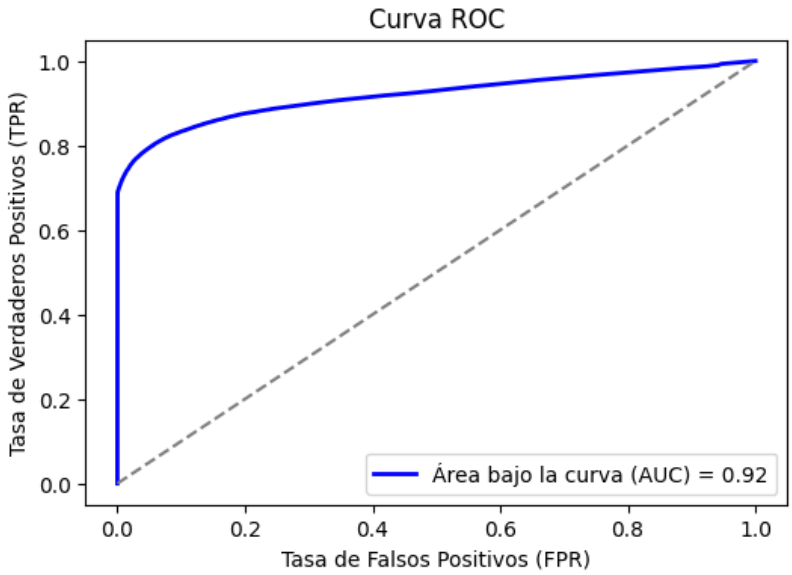


Fuente: Elaboración propia.

La curva ROC generada (ver gráfico 5) refuerza este análisis, mostrando un área bajo la curva (AUC) de 0,92, lo cual indica un desempeño sobresaliente del modelo en la discriminación entre las clases. Este valor confirma que el modelo es altamente eficaz para distinguir entre accidentes potencialmente graves y no graves. No obstante, se observa que la separación más efectiva ocurre en rangos donde se equilibran los falsos positivos y falsos negativos, lo que

sugiere que el ajuste del umbral de decisión podría ser una estrategia adecuada para mejorar la sensibilidad hacia los casos críticos sin deteriorar significativamente el rendimiento general del modelo.

Gráfico 5. Matriz de confusión.



Fuente: Elaboración propia.

En conclusión, el modelo Random Forest aplicado en este estudio demuestra ser una herramienta efectiva para la clasificación de accidentes potencialmente graves (APG), especialmente en contextos con grandes volúmenes de datos y alta desproporción de clases. Su precisión global del 94% y un AUC de 0,92 en la curva ROC reflejan una alta capacidad discriminativa en términos generales. Sin embargo, el desempeño sobre la clase positiva (APG) aún puede ser optimizado, dado que persisten falsos negativos con implicancias preventivas relevantes. La alta dependencia del modelo en la variable "días de reposo" sugiere la necesidad de revisar en profundidad el rol de este indicador en los sistemas de vigilancia. En tanto, variables con bajo aporte relativo podrían ser descartadas para mejorar la eficiencia del modelo sin comprometer su poder predictivo. En conjunto, estos resultados permiten orientar futuras mejoras en el diseño de sistemas de alerta temprana y focalización de intervenciones preventivas, especialmente en sectores y perfiles con mayor exposición a eventos de alta gravedad potencial.

X.3 Resultados y conexión con los objetivos

El análisis de datos de accidentes laborales registrados entre 2011 y 2023 ha permitido identificar patrones en la ocurrencia de accidentes potencialmente graves (APG), así como evaluar los factores que influyen en su severidad y distribución en distintos sectores económicos. Utilizando técnicas de procesamiento de lenguaje natural (PLN), modelos

econométricos y aprendizaje automático, se han obtenido resultados que contribuyen a la comprensión y prevención de estos accidentes. A continuación, se presentan los principales hallazgos del estudio.

1. **Identificación de Accidentes Potencialmente Graves (APG):** Entre 2011 y 2023, se identificaron un total de 103.513 accidentes potencialmente graves a partir del análisis de datos de la Mutual de Seguridad. La proporción de APG respecto del total de accidentes ha variado a lo largo del tiempo, con un aumento significativo en 2022 y 2023, lo que se asocia con cambios en la metodología de registro y análisis. Sectores como construcción, comercio y manufactura presentan las mayores tasas de APG en términos absolutos, mientras que sectores como actividades financieras y científicas muestran una mayor proporción relativa de accidentes potencialmente graves en comparación con su total de accidentes registrados.

2. **Riesgo de APG por Sector Económico:** Con base en los valores promedio de la probabilidad de un accidente potencialente grave para el período 2011-2023, se observa que el sector de la construcción presenta la mayor probabilidad total de accidentes potencialmente graves, alcanzando un 3,94%. Esto lo posiciona como el sector de mayor riesgo, debido tanto a su alta incidencia de accidentes como a la gravedad asociada a estos eventos. Otros sectores con probabilidades significativas son el comercio al por mayor y al por menor (2,39%), actividades de servicios administrativos y de apoyo (1,93%), y agricultura, ganadería, silvicultura y pesca (1,32%). Estos resultados resaltan la necesidad de priorizar la implementación de medidas preventivas en sectores con mayor actividad económica y riesgo inherente. Por el contrario, sectores como suministro de agua, evacuación de aguas residuales, gestión de desechos y descontaminación (0,02%), suministro de electricidad, gas, vapor y aire acondicionado (0,03%), y actividades inmobiliarias (0,13%) presentan las menores probabilidades totales de APG. Aunque estos sectores muestran menores niveles de riesgo en términos de frecuencia, es fundamental mantener controles para prevenir accidentes graves en actividades críticas. Este análisis refuerza la importancia de utilizar $P(\mathrm{APG})$ como herramienta clave para orientar políticas públicas y estrategias de mitigación específicas, adaptadas a las características de cada sector económico.

3. **Severidad del Daño y Días de Reposo:** La severidad del daño en los accidentes laborales ha sido analizada utilizando el número de días de reposo como indicador clave. Se observa una tendencia creciente en la cantidad de días de reposo hasta 2020, con una mediana que varía entre 2 y 3 días en la mayoría de los años, pero con valores extremos que elevan el promedio hasta 41,28 días en 2020. Posteriormente, en 2022 y 2023, se registró una disminución en el promedio de días de reposo, posiblemente asociada a cambios en criterios de registro o en la severidad de los accidentes reportados.

4. **Factores Asociados a los APG según Modelos Econométricos:** Los modelos logit y probit estimados indican que los principales factores que incrementan la probabilidad de un APG incluyen:

    - **Edad:** A medida que aumenta la edad, la probabilidad de que un accidente sea clasificado como APG también se incrementa.
    - **Sexo masculino:** Los hombres presentan una mayor probabilidad de sufrir accidentes potencialmente graves en comparación con las mujeres.
    - **Días de reposo:** A mayor cantidad de días de reposo, mayor es la probabilidad de que un accidente sea clasificado como APG.
    - **Calificación del accidente:** Ciertas categorías de calificación de accidentes (como enfermedades profesionales o accidentes de trayecto) tienen menor probabilidad de ser considerados APG en comparación con accidentes de trabajo.

5. **Resultados del Modelo Predictivo Random Forest:** La aplicación de modelos de Machine Learning permitió construir un modelo Random Forest con una precisión del 94% para la clasificación de APG. Sin embargo, el modelo muestra un recall del 63% para la clase positiva (APG), lo que sugiere dificultades en la detección de algunos casos de APG. La variable más influyente en la clasificación de APG es el número de días de reposo, seguido por la edad y la calificación del accidente. Aunque el modelo logra discriminar entre accidentes graves y no graves, se identificaron 2507 falsos positivos y 1789 falsos negativos, lo que indica áreas de mejora en la optimización del balance entre precisión y recall.

Estos hallazgos permiten orientar estrategias preventivas en sectores con alta incidencia de APG, mejorar la identificación de riesgos y reforzar medidas de seguridad laboral para reducir la ocurrencia de estos eventos. De este modo, los resultados obtenidos y validados permiten cumplir los objetivos planteados y orientan intervenciones preventivas alineadas con la evidencia generada.

XI. Validación de Resultados del Modelo de Accidentes Graves y Fatales

El proceso de validación de los resultados del modelo de clasificación de accidentes potencialmente graves y fatales se realizó mediante la participación de los expertos de la Mutual de Seguridad, así como expertos externos en áreas clave como inteligencia artificial, ciencia de datos y seguridad laboral. La validación fue llevada a cabo en dos instancias: una reunión interna con Oscar Muñoz y Luis Alarcón de la Mutual de Seguridad, y un Taller de Expertos que incluyó un grupo multidisciplinario.

XI.1 Reunión con Expertos de Mutual de Seguridad

Para validar los resultados preliminares del modelo de clasificación, se llevó a cabo una sesión técnica con dos profesionales de la Mutual de Seguridad, ambos con experiencia reconocida en prevención de riesgos, salud ocupacional, innovación organizacional y ciencia de datos aplicada. La validación se estructuró como una sesión presencial de revisión guiada, basada en una pauta temática diseñada por el equipo de investigación, que consideró cinco dimensiones: (1) consistencia normativa del modelo con la NT142 y el Compendio SUSESO, (2) pertinencia del diccionario de términos críticos, (3) validez de los patrones semánticos detectados, (4) incorporación de variables estructuradas y contextuales, y (5) recomendaciones de mejora aplicativa.

La dinámica combinó la exposición de resultados con visualizaciones y el análisis conjunto de ejemplos concretos de accidentes clasificados como potencialmente graves (APG), con énfasis en la interpretación de los criterios utilizados por el modelo. Las observaciones realizadas fueron registradas en una minuta y sistematizadas por el equipo técnico, utilizándose posteriormente para ajustar el modelo predictivo y reforzar su coherencia normativa.

En esta reunión, el equipo de investigación se centró en evaluar los resultados preliminares del modelo de clasificación basado en técnicas de Procesamiento de Lenguaje Natural (NLP) y minería de texto. Se discutieron los criterios de clasificación de los accidentes graves y fatales, y se evaluó si el modelo desarrollado alineaba adecuadamente con las normativas vigentes, como la NT142 y el Compendio Normativo 744 de la SUSESO. En ese sentido, se recomendó integrar como referencia la Guía para la Identificación y Evaluación de Riesgos de Seguridad en los Ambientes de Trabajo del Instituto de Salud Pública (ISP), para asegurar que los criterios de clasificación se alinearan con las normas y estándares regulatorios establecidos. La guía fue destacada como herramienta útil para asegurar coherencia en la identificación de riesgos y la evaluación de daños, especialmente en relación con accidentes graves y fatales.

Se validaron patrones identificados por el modelo, como la relación entre atrapamientos y edad avanzada, los cuales fueron considerados consistentes con la literatura especializada y los datos históricos de la Mutual. Este hallazgo fue reconocido como relevante para la clasificación de accidentes graves, y se señaló que el modelo permite incorporar factores adicionales que no suelen estar disponibles en los registros tradicionales. Se destacó, entre otros, la inclusión de enfermedades crónicas y el sector económico como variables adicionales que permiten una caracterización más completa de los eventos, lo que mejora la precisión del análisis y aporta valor preventivo.

Sin embargo, también se identificaron limitaciones. Los expertos observaron que el modelo se focaliza principalmente en las consecuencias del accidente (por ejemplo, su gravedad), sin incorporar de forma explícita las condiciones previas al evento. Esta carencia fue considerada una debilidad metodológica relevante, ya que limita la capacidad del modelo para capturar factores de riesgo subyacentes. Como resultado, se recomendó incorporar nuevas variables o diccionarios que reflejen estos elementos previos, con el fin de avanzar hacia una caracterización más integral del riesgo y robustecer las capacidades predictivas del modelo.

XI.2 Taller de Expertos

El Taller de Expertos reunió a un grupo multidisciplinario de profesionales con experiencia en prevención de riesgos laborales, inteligencia artificial, análisis de datos y políticas públicas. Su propósito fue revisar los resultados preliminares del modelo de clasificación de accidentes graves y fatales, y aportar observaciones técnicas que permitieran fortalecer su diseño metodológico y su aplicabilidad operativa.

La actividad se desarrolló en formato presencial y adoptó una metodología participativa de carácter técnico. Durante la sesión, el equipo de investigación presentó los objetivos, las técnicas utilizadas y los principales hallazgos del estudio, generando un espacio de diálogo estructurado en torno a los componentes centrales del modelo. Los expertos invitados, provenientes de ámbitos académico, institucional y de consultoría, realizaron una revisión crítica de los insumos presentados, discutiendo tanto su validez técnica como su utilidad práctica para la gestión del riesgo. El proceso de validación se apoyó en una pauta temática diseñada previamente, que organizó la discusión en torno a aspectos como la cobertura sectorial del modelo, la representatividad de las variables utilizadas, la alineación con marcos normativos y la capacidad predictiva del enfoque implementado. Las observaciones surgidas durante el taller fueron sistematizadas por el equipo técnico y utilizadas como insumo para ajustar tanto el modelo como sus herramientas complementarias, incluyendo el diccionario semántico.

Uno de los principales aportes del taller fue la recomendación de complementar el enfoque actual, centrado en la gravedad del accidente como resultado observable, con una mirada orientada a las causas subyacentes o condiciones previas que puedan haber influido en la

ocurrencia del evento. Esta sugerencia coincidió con lo planteado en validaciones previas y permitió reforzar la importancia de incorporar variables que permitan anticipar riesgos antes de que se materialicen. Asimismo, se propuso ampliar los diccionarios utilizados en el modelo para reflejar con mayor precisión las particularidades de sectores productivos críticos como la construcción y el transporte, que concentran una alta incidencia de accidentes graves pero no estaban suficientemente representados en los primeros resultados. Esta expansión semántica permitiría aumentar la sensibilidad del modelo ante patrones específicos de siniestralidad.

Otra recomendación importante fue explorar el uso de modelos predictivos más avanzados, como aquellos basados en machine learning o deep learning, que integren distintos tipos de datos, textuales, estructurados y visuales, en un enfoque multimodal. Esta integración podría mejorar la capacidad anticipatoria del modelo, haciendo posible detectar configuraciones de riesgo con mayor precisión y en etapas más tempranas. Finalmente, se abordó la calidad de los datos utilizados, en particular los relatos clínicos y narrativas textuales de los accidentes, advirtiéndose que estos campos pueden estar sujetos a sesgos derivados del estilo de registro o de la discrecionalidad de los profesionales clínicos. En este sentido, se recomendó estandarizar los procedimientos de documentación y establecer criterios mínimos para la calidad narrativa, con el objetivo de garantizar la consistencia y confiabilidad de los datos empleados en futuros modelos de análisis.

Las observaciones entregadas por los expertos durante el taller no solo enriquecieron la discusión técnica, sino que fueron consideradas activamente como parte del proceso de validación metodológica del estudio. Varias de las recomendaciones planteadas, como la necesidad de ampliar los diccionarios por sector económico, incorporar variables asociadas a condiciones previas al accidente y avanzar hacia modelos predictivos más sofisticados, fueron integradas en el rediseño del modelo, en el ajuste de herramientas analíticas y en la formulación de recomendaciones estratégicas. Este taller permitió, por tanto, consolidar la coherencia técnica del estudio y fortalecer su aplicabilidad en la gestión preventiva de la siniestralidad laboral.

XI.3 Mejoras y Ajustes Basados en la Validación

A partir de las sugerencias y recomendaciones recibidas durante las reuniones con los expertos de la Mutual de Seguridad y el Taller de Expertos, se llevaron a cabo diversas mejoras y ajustes en el modelo de clasificación de accidentes potencialmente graves y fatales. Estos cambios fueron diseñados para mejorar la precisión, representatividad y efectividad del modelo en la identificación de riesgos laborales, con un enfoque preventivo más robusto.

- **Incorporación de causas previas:** Una de las principales recomendaciones fue integrar un enfoque preventivo que no solo evaluará las consecuencias de los accidentes, sino también las condiciones previas que podrían predisponer a un accidente grave. En respuesta a esta sugerencia, se incorporó un análisis más profundo de factores de riesgo

ocupacionales, como las condiciones laborales, el equipo utilizado y otros aspectos relacionados con el entorno de trabajo. Además, se incluyó un enfoque sobre factores organizacionales, como la carga de trabajo, los procedimientos de seguridad y la cultura organizacional, que pueden influir en la ocurrencia de accidentes. Este ajuste permite al modelo identificar áreas de intervención proactiva antes de que los accidentes ocurran, mejorando la prevención.

- **Expansión de cobertura sectorial:** En respuesta a la recomendación de los expertos sobre la representación de sectores críticos, se amplió la base de datos del modelo para incluir información sobre sectores laborales de alto riesgo, como construcción, transporte y salmonicultura. Estos sectores, que presentan una alta incidencia de accidentes graves, no estaban adecuadamente reflejados en los resultados iniciales del modelo. Al incorporar más datos específicos de estos sectores, se mejoró la representatividad del modelo, permitiendo que los resultados sean más precisos y aplicables a las realidades de estos sectores clave. De este modo, el modelo se vuelve más inclusivo y adecuado para las empresas de estos sectores, que a menudo enfrentan riesgos laborales únicos.

- **Avance hacia modelos predictivos:** Se está trabajando en la implementación de modelos de machine learning para llevar el enfoque del modelo de clasificación a uno más predictivo. Este avance permitirá anticipar con mayor precisión los accidentes antes de que ocurran, basándose no solo en los datos históricos de accidentes, sino también en nuevas fuentes de datos, como sensores e imágenes, que proporcionarán información más detallada y actualizada sobre las condiciones laborales en tiempo real. La integración de estos modelos permitirá una gestión de riesgos más proactiva, capaz de identificar patrones y factores de riesgo emergentes, lo que mejorará las intervenciones preventivas y permitirá acciones correctivas más tempranas.

- **Mejoras en los registros de datos:** Se identificó que los sesgos en los relatos de accidentes podrían afectar la precisión del análisis. En respuesta a esta preocupación, se implementaron estrategias para reducir estos sesgos, como la revisión y estandarización de los registros clínicos y de accidentes. Además, se establecieron nuevos protocolos de recolección de datos que aseguran una captura uniforme y coherente de la información, minimizando las discrepancias en la documentación de los incidentes. Estos nuevos protocolos se basan en estándares homogéneos y en un enfoque más sistemático, lo que facilita el análisis de los datos y garantiza que la información utilizada para las predicciones sea más confiable y consistente. Como parte de este avance, se planea incorporar modelos transformador generativo previamente entrenado (GPT por sus siglas en inglés) para procesar y analizar los relatos de los accidentes en lenguaje natural, lo que permitirá identificar patrones de riesgo adicionales y mejorar la clasificación y segmentación de los eventos. Los modelos GPT podrían ayudar a mejorar la precisión de las predicciones al analizar grandes volúmenes de datos

textuales y ofrecer una interpretación más precisa de los relatos de accidentes, contribuyendo así a una gestión de riesgos más proactiva.

Como conclusión, el proceso de validación con criterio experto fue fundamental para ajustar y mejorar los resultados del modelo de clasificación de accidentes potencialmente graves y fatales. A través de las reuniones con Oscar Muñoz y Luis Alarcón, y el Taller de Expertos, se validaron los hallazgos, se identificaron fortalezas y limitaciones, y se generaron recomendaciones clave que han permitido fortalecer la capacidad predictiva del modelo y asegurar su aplicabilidad a la gestión de riesgos laborales.

XII. Discusión

Esta sección analiza los principales hallazgos y aportes metodológicos del estudio, situándolos en el contexto de la evolución de las herramientas de procesamiento de lenguaje natural y su aplicación en la gestión de la seguridad laboral. Se destaca la comparación entre enfoques tradicionales (NLP estadístico) y la incorporación exploratoria de inteligencia artificial generativa (GPT) para ampliar las capacidades interpretativas. A partir de esta revisión, se discuten oportunidades, desafíos y consideraciones clave para fortalecer la trazabilidad, la automatización y la calidad de los sistemas de clasificación y prevención de accidentes. Finalmente, se presentan las principales limitaciones metodológicas identificadas.

XII.1 Aplicación de modelos generativos GPT

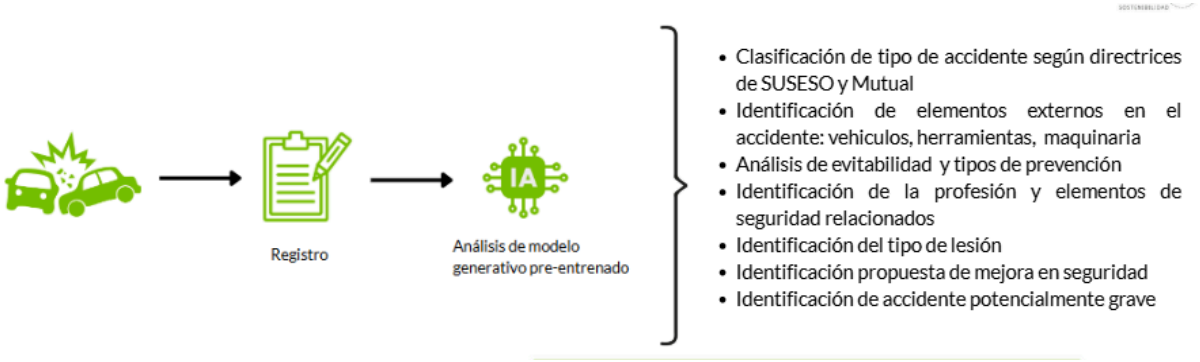
A modo de discusión, se dispone de una medida complementaria a los análisis basados en técnicas tradicionales de procesamiento de lenguaje natural tratadas en este informe. Consiste en la utilización de un componente exploratorio mediante el uso de inteligencia artificial generativa (IAG), específicamente un modelo tipo GPT (Generative Pre-trained Transformer) para el análisis de los relatos de los accidentes. Esta etapa tuvo por objetivo ampliar la capacidad interpretativa del sistema y evaluar la factibilidad de aplicar modelos de lenguaje de última generación a la caracterización de accidentes laborales. El siguiente cuadro muestra las diferencias de las aplicaciones metodológicas:

NLP Tradicional (Informe original SUSESO)	NLP Generativo con GPT-4 (Análisis actual)
• Requiere preprocesamiento exhaustivo: limpieza, normalización y etiquetado manual de datos.	• El relato libre del accidente es procesado directamente como entrada (“lectura comprensiva” automatizada).
• Aplica técnicas estadísticas como conteo de tokens, frecuencia de términos y análisis de *embeddings* semánticos.	• Se le consulta al modelo si el accidente es potencialmente grave, y este responde con clasificación y justificación inmediata.
• El desarrollo de modelos de clasificación exige entrenamiento supervisado, etiquetado previo y validación cruzada.	• No se requiere entrenamiento local ni codificación previa: el modelo aprende de insumos de SUSESO y razona en contexto.
• Los criterios de clasificación son rígidos: cambios metodológicos implican reentrenar modelos completos.	• Permite adaptar criterios de análisis (por ejemplo, definir gravedad según nuevas normativas) sin modificar el modelo base.
	• Acelera la transformación de texto en variables estructuradas, habilitando análisis masivo de forma escalable.

A diferencia del procesamiento léxico convencional, los modelos generativos como GPT se caracterizan por capturar relaciones semánticas de largo alcance y comprender el contexto global del relato. Esto permite no solo clasificar el tipo de accidente de forma más precisa, sino también inferir causas implícitas, evaluar si el evento era evitable, asignar niveles de severidad y sugerir recomendaciones preventivas consistentes. Esta capacidad de razonamiento en contexto habilitó la extracción de variables que permiten identificar características específicas del accidente, tales como:

- Clasificación del tipo de accidente según directrices de SUSESO y Mutual.
- Identificación de elementos externos involucrados (vehículos, herramientas, maquinaria).
- Evaluación de evitabilidad y tipos de prevención posibles.
- Inferencia de la profesión o rol del trabajador y elementos de seguridad presentes o ausentes.
- Tipo de lesión asociada al evento.
- Propuesta de mejora en seguridad expresada por el modelo.
- Determinación de si el accidente corresponde a un evento potencialmente grave.

La incorporación de estas variables permitió enriquecer el análisis cuantitativo posterior y estructurar un conjunto de atributos adicionales que no estaban disponibles en las bases originales. En términos operativos, la capacidad del modelo para transformar texto libre en variables estructuradas representa un avance significativo en la automatización del análisis de accidentes laborales, particularmente por su adaptabilidad a nuevas normativas sin necesidad de mayor reentrenamiento. Esto se resume en la siguiente figura:



Estos resultados se alinean con la literatura sobre inteligencia artificial generativa, que destaca su utilidad en tareas de interpretación de lenguaje natural en dominios técnicos, su capacidad de adaptación a distintos marcos conceptuales y su aplicabilidad en contextos donde los datos estructurados son limitados o incompletos (Bommasani et al., 2022; Kaplan et al., 2020).

Si bien los modelos de lenguaje generativo ofrecen ventajas considerables en la comprensión contextual y estructuración semántica de textos complejos, también presentan riesgos

metodológicos que deben ser considerados. Uno de los más relevantes es el fenómeno de las alucinaciones, entendidas como la generación de contenido factualmente incorrecto o inventado con alta coherencia gramatical. Este comportamiento, ampliamente documentado en la literatura especializada, puede comprometer la validez de los resultados si no se implementan mecanismos de control adecuados. Las alucinaciones pueden adoptar múltiples formas: desde clasificaciones incorrectas hasta recomendaciones sin fundamento en el texto de origen, particularmente cuando el modelo debe inferir información ausente de forma explícita.

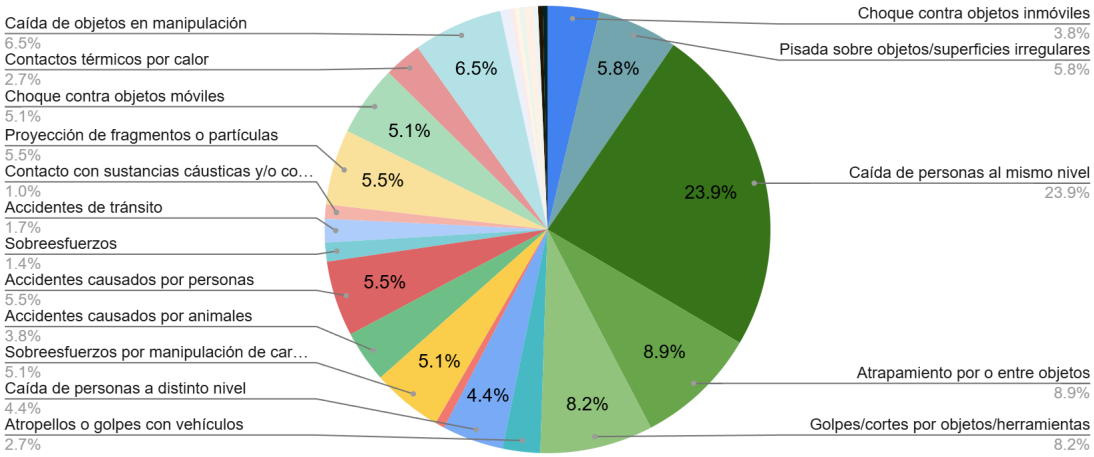
Este fenómeno ha sido descrito detalladamente por Ji et al. (2022)[12], quienes definen la alucinación como "la generación de contenido inconsistente con las entradas o con hechos del mundo real". Mündler et al. (2023)[13] analizan cómo los modelos pueden incurrir en alucinaciones autocontradictorias, generando afirmaciones que se niegan entre sí en una misma salida. Asimismo, Islam et al. (2024)[14] ofrecen un panorama actualizado de las técnicas de mitigación más efectivas, destacando el rol de mecanismos de validación y re-pregunta como los aplicados en este estudio).

Para mitigar este riesgo, el estudio implementó un procedimiento de verificación antagonista. Este consistió en reenviar las respuestas generadas por el modelo GPT a una segunda iteración del mismo modelo, solicitando que evaluará si la clasificación realizada y los atributos extraídos se encontraban efectivamente presentes y justificables en el relato original. Este mecanismo de autocontrol permitió identificar y corregir inconsistencias, reduciendo así la incidencia de errores semánticos y asegurando que las salidas estuvieran respaldadas por el contenido textual real. Este enfoque, si bien exploratorio, demostró ser una estrategia efectiva para contener las alucinaciones inherentes al modelo, y constituye un precedente valioso para futuras aplicaciones de inteligencia artificial generativa en contextos institucionales con requerimientos de trazabilidad y calidad verificable.

A modo de prueba, se analizaron 300 correspondientes al 2023 aplicando esta metodología de análisis con modelos GPT. Basando el análisis en el documento “Guía para la identificación y evaluación de riesgos de seguridad en los ambientes de trabajo” esta muestra un total de 46 diferentes riesgos laborales asociados y su respectiva calificación y definición. Al crear el conjunto de instrucciones sistemático (prompt), se le entrega el contexto necesario para el análisis y posterior identificación, logrando definir el relato dentro del marco de los casos de estudio . En total se obtuvo los siguientes accidentes identificados y su distribución:

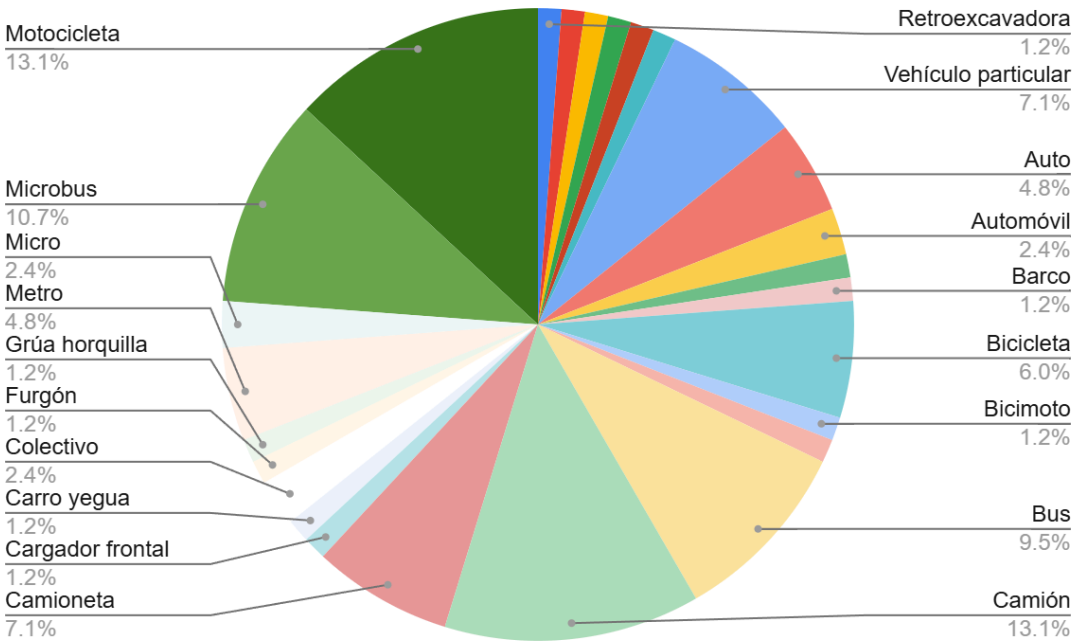
[12] Ver: https://arxiv.org/abs/2202.03629
[13] Ver: https://arxiv.org/abs/2305.15852
[14] Ver: https://arxiv.org/abs/2401.01313

Gráfico 6. Distribución de accidentes identificados



Por otro lado, se observa que el 63,4% de los accidentes analizados no presentan un vehículo asociado. Dentro del 36,6% restante, los tipos de vehículo más frecuentemente identificados son el camión (13,1%) y la motocicleta. No obstante, al agrupar los registros por tipo de transporte, se estima que el 29,8% de los accidentes corresponden a incidentes vinculados al transporte público, incluyendo metro, microbuses, buses y colectivos. En contraste, la participación de vehículos especializados como retro excavadoras o grúas horquilla es marginal, representando un porcentaje muy bajo del total de casos.

Gráfico 7. Distribución de vehículos asociados a accidentes.



En relación con el reconocimiento de contextos, el análisis asistido por modelos GPT demostró una capacidad superior no solo para identificar accidentes potencialmente graves, sino también para descartarlos adecuadamente, incluso en casos donde la técnica tradicional de NLP, basada en coincidencia de términos, sugería lo contrario. Un ejemplo ilustrativo corresponde a un relato inicialmente clasificado como "potencialmente grave" por el modelo tradicional, pero que fue reclasificado por el modelo generativo, argumentando: "Aunque el accidente involucra una caída y golpes, no se menciona la presencia de lesiones graves, ni involucra altura, maquinaria, electricidad o químicos." Este tipo de evaluación evidencia que el modelo generativo no se limita a patrones léxicos, sino que interpreta el contexto semántico completo, permitiendo una clasificación más precisa.

En términos cuantitativos, el modelo GPT identificó 129 accidentes como potencialmente graves, mientras que la técnica tradicional clasificó 17 casos adicionales dentro de esta categoría, presumiblemente debido a su sensibilidad frente a palabras clave. Esta diferencia refuerza la importancia de integrar enfoques basados en comprensión contextual, especialmente en dominios donde las decisiones analíticas tienen implicancias en gestión del riesgo y prevención institucional.

Otro aspecto particularmente valioso del uso de modelos generativos fue su capacidad para identificar causales del accidente a partir del contexto narrativo, incluso cuando estas no estaban explícitamente formuladas en el relato. A diferencia de los modelos tradicionales, que se limitan a reconocer patrones léxicos o secuencias de términos, GPT fue capaz de inferir relaciones causales indirectas, como la interacción entre condiciones ambientales, fallas operacionales y omisiones en el uso de elementos de protección personal. Asimismo, el modelo generó recomendaciones de mejora y medidas preventivas contextualizadas, que incluyeron desde modificaciones en procedimientos hasta ajustes en la capacitación y supervisión. Esta funcionalidad refuerza el valor del modelo no solo como herramienta clasificadora, sino también como apoyo en la generación de conocimiento aplicado para la gestión preventiva.

En síntesis, la incorporación de modelos de inteligencia artificial generativa en el análisis de accidentes laborales representa un avance metodológico relevante, al permitir una interpretación más rica, contextualizada y operativamente útil de la información contenida en relatos no estructurados. Si bien existen desafíos asociados a su implementación, como el control de alucinaciones y la necesidad de validación cruzada, los resultados obtenidos evidencian su potencial para complementar y robustecer los enfoques tradicionales. La capacidad del modelo para extraer variables latentes, descartar falsos positivos y generar recomendaciones preventivas en lenguaje natural ofrece nuevas oportunidades para el desarrollo de sistemas de análisis predictivo y de gestión de riesgos laborales más adaptativos, precisos y escalables.

XII.2 Consideraciones y limitaciones metodológicas

Si bien los resultados obtenidos evidencian un avance significativo en la aplicación de modelos de procesamiento del lenguaje natural y herramientas de inteligencia artificial generativa, es necesario considerar ciertas limitaciones inherentes al diseño metodológico, los datos disponibles y la naturaleza del fenómeno analizado:

1. **Dependencia de la calidad del texto narrativo:** El análisis se basa en campos de texto libre, cuya calidad depende del registro inicial. Errores tipográficos o descripciones escuetas pueden afectar la precisión del modelamiento.

2. **Sesgo por exclusión de registros incompletos:** Se excluyeron más de 175 mil casos sin relatos textuales, lo que podría introducir sesgos si hay patrones sistemáticos en esos registros.

3. **Desbalance de clases en la variable dependiente:** La baja proporción de accidentes potencialmente graves dificultó el desempeño de modelos como logit. Se aplicaron estrategias como ponderación y Random Forest, pero se reconocen desafíos en sensibilidad. Se recomienda considerar sobremuestreo (e.g., SMOTE) en futuras aplicaciones.

4. **Naturaleza no experimental del estudio:** Las relaciones observadas son asociativas y no implican causalidad. Los hallazgos deben interpretarse como correlaciones útiles para orientar la prevención, no como inferencias causales.

5. **Heterogeneidad no observada entre sectores y años:** Aunque se controló por actividad económica y año, pueden existir factores no observados (cultura organizacional, cobertura de mutualidades, cambios normativos) que afecten la ocurrencia o el registro.

6. **Limitaciones computacionales:** Dado el volumen de datos, algunas técnicas avanzadas (como validación cruzada o uso masivo de modelos LLM) se aplicaron solo sobre muestras representativas. Estas decisiones buscaron asegurar trazabilidad y eficiencia sin comprometer la validez general de los resultados.

Estas limitaciones no invalidan los hallazgos, pero refuerzan la necesidad de considerar el enfoque propuesto como una herramienta complementaria dentro de una estrategia más amplia de gestión del riesgo. La interpretación de los resultados debe realizarse considerando estos factores y orientarse a la mejora continua de los sistemas de información, recolección de datos y capacidades analíticas institucionales.

XIII. Conclusiones

El presente estudio permitió identificar patrones y tendencias en los accidentes laborales reportados a la Mutual de Seguridad CChC, con especial atención a los accidentes potencialmente graves (APG). A través de modelos predictivos y análisis estadístico, se caracterizaron factores asociados a su ocurrencia, como edad, sexo, días de reposo y tipo de calificación, lo que refuerza la necesidad de medidas focalizadas para la mitigación del riesgo. Los hallazgos muestran diferencias significativas por perfil y segmento laboral, lo que valida la importancia de un enfoque diferenciado en la gestión de la seguridad y salud en el trabajo.

Uno de los principales aportes metodológicos del estudio fue la aplicación de técnicas de procesamiento de lenguaje natural (NLP) para clasificar y caracterizar los APG a partir de relatos textuales. Esta estrategia permitió sistematizar grandes volúmenes de información no estructurada, identificar patrones de riesgo no observables con técnicas tradicionales y construir variables analíticas robustas. La coherencia entre los resultados obtenidos y los patrones empíricamente conocidos valida esta metodología como una herramienta efectiva para la detección temprana y segmentación de riesgos en contextos operativos reales.

Desde una perspectiva sectorial, se confirma una mayor concentración de APG en sectores como construcción y manufactura, lo que sugiere la necesidad de implementar protocolos específicos y estrategias preventivas intensivas en estos ámbitos. Paralelamente, sectores con menor exposición, como electricidad, gas y agua, pueden utilizarse como referencia para optimizar la asignación de recursos, en función del riesgo estimado. Esta diferenciación sectorial permite diseñar planes de acción más eficientes y alineados con las condiciones reales del entorno productivo.

El análisis temporal reveló un cambio estructural en la siniestralidad a partir de 2020, posiblemente asociado a los efectos de la pandemia, lo que refuerza la necesidad de adaptar continuamente las estrategias preventivas frente a eventos disruptivos externos.

En términos prácticos, los resultados obtenidos refuerzan la importancia de integrar herramientas de análisis de datos en la gestión preventiva. La incorporación de modelos predictivos permite anticipar escenarios de riesgo y diseñar intervenciones antes de que los accidentes ocurran, lo que fortalece la eficacia de las decisiones preventivas. Este enfoque, sustentado en datos históricos y procesamiento automatizado de información textual, contribuye a la generación de políticas de seguridad más adaptativas, con mayor capacidad de respuesta y ajustadas a las condiciones reales del trabajo.

Se destaca el uso de la Guía para la Identificación y Evaluación de Riesgos de Seguridad en los Ambientes de Trabajo del Instituto de Salud Pública como referencia normativa clave para la clasificación de los eventos analizados, especialmente en relación con la identificación de

riesgos y la evaluación de daños en los accidentes laborales. Esta guía permitió asegurar coherencia regulatoria y estandarización en la identificación de accidentes graves, alineando el estudio con los criterios normativos vigentes para la medición de riesgo y daño en los entornos laborales. Además, se conecta directamente con las normas internacionales en la gestión de la seguridad laboral, como los estándares establecidos por la Organización Internacional del Trabajo (OIT) y las Normas ISO 45001, que proporcionan un marco normativo global para reforzar la identificación estandarizada de riesgos en el lugar de trabajo. A partir de esta experiencia, se recomienda avanzar en el desarrollo de instrumentos analíticos actualizables que permitan fortalecer los sistemas de vigilancia de riesgos, basados en evidencia empírica y criterios normativos vigentes.

Asimismo, los modelos logit y probit aplicados permitieron confirmar con evidencia estadística que factores como la edad, el género, la presencia de enfermedad crónica y el tipo de accidente, particularmente los accidentes de trabajo y de trayecto en comparación con las enfermedades profesionales, tienen una relación significativa con la probabilidad de que un evento sea clasificado como potencialmente grave. Aunque estos modelos mostraron alta validez para explicar patrones de riesgo, sus limitaciones como clasificadores predictivos refuerzan la necesidad de complementarlos con modelos de aprendizaje automático como Random Forest y técnicas de balanceo de clases para mejorar la identificación temprana de accidentes graves.

Finalmente, este estudio subraya la necesidad de consolidar una cultura de seguridad basada en el uso de evidencia analítica para la toma de decisiones en prevención de riesgos. La implementación de estrategias proactivas, combinadas con la optimización de modelos predictivos y el uso de tecnologías de análisis semántico, puede contribuir significativamente a la reducción de accidentes graves en los entornos laborales. A partir de estos hallazgos, se recomienda continuar con investigaciones que integren nuevas fuentes de datos, variables contextuales y metodologías de inteligencia artificial avanzadas, como modelos generativos, con el fin de fortalecer la capacidad institucional de anticipación frente a riesgos emergentes y promover una gestión preventiva más continua, sectorializada y efectiva.

Los hallazgos de este estudio se traducen en una herramienta de gestión para la prevención de accidentes laborales: un sistema analítico automatizado que permite identificar casos con alta probabilidad de severidad a partir de sus características observables y relatos textuales. Esta herramienta puede ser implementada como un sistema de alerta temprana que priorice casos para investigación, refuerce la trazabilidad institucional y focalice recursos preventivos en sectores, funciones o perfiles con mayor exposición crítica.

El análisis por forma de ocurrencia, diagnóstico, días de reposo y edad entrega criterios objetivos que pueden ser integrados en matrices de riesgo, planes de intervención diferenciados y estrategias de capacitación. Esta orientación está en línea con la literatura internacional, que destaca la necesidad de adaptar las estrategias preventivas a los riesgos específicos de cada sector. Estudios previos han demostrado que sectores con mayores riesgos laborales, como la

construcción y la manufactura, requieren enfoques preventivos intensivos, mientras que sectores con menor exposición, como electricidad y gas, deben centrarse en protocolos más específicos (Raineri & Figueroa, 2018; González & Sánchez, 2019). El presente estudio refuerza estos enfoques mediante la aplicación de modelos predictivos y análisis sectoriales que permiten definir acciones preventivas más focalizadas.

Asimismo, uno de los hallazgos relevantes del estudio es la dependencia de los sistemas de reporte respecto a la existencia de una cultura organizacional que priorice la seguridad como un eje estratégico. Tal como se señaló en la sección de definición del problema, la efectividad de estos sistemas se ve limitada cuando las empresas no disponen de mecanismos consolidados para la gestión preventiva o no asignan suficiente prioridad a la prevención. Los resultados obtenidos refuerzan este diagnóstico, al mostrar que sectores con alta concentración de accidentes potencialmente graves (como la construcción) presentan una mayor recurrencia de registros con relatos incompletos, baja especificidad técnica y escasa mención de medidas preventivas, lo que podría reflejar deficiencias en los procesos de reporte y seguimiento. Este punto resulta clave para orientar futuras estrategias institucionales, tanto en términos de fiscalización como de apoyo técnico. Este hallazgo da cuenta de lo anticipado en la definición del problema, en cuanto a que la capacidad de anticipación del riesgo se ve afectada por la debilidad en los sistemas de gestión preventiva y la cultura organizacional de las empresas, especialmente en sectores con alta siniestralidad. Por tanto, los resultados empíricos obtenidos confirman esta limitación estructural y refuerzan la necesidad de intervenciones diferenciadas.

Finalmente, se destaca la importancia de abordar los riesgos ocultos: situaciones que no se manifiestan como eventos críticos en los registros tradicionales, pero que podrían haber derivado en consecuencias graves bajo condiciones distintas. Este tipo de riesgo requiere una combinación de análisis técnico y lectura contextual, junto con programas de formación que fortalezcan la capacidad de detección temprana en los entornos de trabajo. Profundizar en esta línea permitiría a la Mutual de Seguridad CChC reforzar su capacidad de anticipación ante escenarios de alto impacto, consolidando un enfoque más integral de gestión preventiva.

XIV. Recomendaciones para Sistema de SST

Las recomendaciones que se presentan a continuación integran tanto los hallazgos del estudio como las observaciones recogidas en el proceso de validación técnica realizado con expertos en prevención de riesgos.

XIV.1 Tecnologías de monitoreo y enfoque multinivel

Para mejorar la prevención en sectores de alto riesgo como la construcción y la minería, es fundamental no solo adoptar tecnologías de monitoreo en tiempo real, sino también abordar los desafíos que dificultan su implementación efectiva.

Uno de los principales desafíos es la inversión inicial requerida para estas tecnologías, lo que puede ser una barrera significativa para empresas de menor tamaño. En este sentido, se recomienda explorar modelos de financiamiento progresivo o esquemas de incentivos que fomenten la adopción de estas herramientas.

Otro aspecto clave es la capacitación del personal. La introducción de nuevas tecnologías requiere competencias específicas, por lo que es fundamental diseñar estrategias de formación continua para asegurar que los trabajadores puedan interpretar y reaccionar correctamente ante la información proporcionada por los sistemas de monitoreo.

Además, la integración de estas tecnologías en la cultura organizacional de las empresas representa un desafío importante. En muchos casos, la resistencia al cambio puede dificultar la adopción de enfoques multinivel de seguridad. Para superar esta barrera, se recomienda promover la seguridad como un valor central dentro de la organización, fomentando la participación activa de los trabajadores en el diseño y evaluación de medidas preventivas.

Más allá de la simple adopción de tecnologías y enfoques de gestión, las estrategias de prevención deben ajustarse a la realidad estructural de cada empresa, considerando sus capacidades económicas, el nivel de capacitación del personal y la cultura organizacional existente. Solo mediante un enfoque integral y adaptado se podrá garantizar una implementación efectiva y sostenible en el tiempo.

XIV.2 Programas educativos y adaptativos y acompañamiento técnico por perfil de riesgo

A partir del análisis realizado, se propone fortalecer las capacidades formativas y de asesoría técnica mediante un modelo de acompañamiento progresivo, orientado a sectores y empresas con menor madurez preventiva y mayor exposición a eventos potencialmente graves. Esta estrategia debe considerar:

- **Diagnóstico inicial del nivel de madurez preventiva por sector económico,** tamaño de empresa y comportamiento histórico de siniestralidad, como base para una asignación diferenciada de recursos técnicos y formativos.

- **Diseño de planes de apoyo sectoriales,** que integren asesoría técnica, formación continua y herramientas específicas para empresas con baja capacidad estructural de gestión del riesgo. Se recomienda que los contenidos formativos estén alineados con los hallazgos derivados del análisis histórico de accidentes, identificando segmentos críticos de trabajadores o condiciones de exposición recurrentes.

- **Esquemas de certificación o reconocimiento,** que motiven a las empresas a incorporar buenas prácticas preventivas y formación continua alineada con estándares técnicos, generando incentivos positivos a la mejora sostenida.

- **Integración de la formación en seguridad a los planes de desarrollo profesional,** de forma que los contenidos no se perciban como obligaciones normativas aisladas, sino como parte de una estrategia de mejora del desempeño, condiciones de trabajo y productividad.

- **Incorporación de protocolos diferenciados para sectores con alta incidencia de accidentes potencialmente graves (APG),** como construcción o manufactura, incluyendo rutinas específicas para la gestión del riesgo operativo y la respuesta ante emergencias.

- **Mecanismos de seguimiento y retroalimentación continua,** que permitan monitorear avances en cultura preventiva, aplicación de protocolos y reducción de eventos críticos, facilitando ajustes oportunos a las estrategias formativas.

Esta recomendación apunta a fortalecer el vínculo entre evidencia empírica, capacidades institucionales y cultura organizacional, permitiendo una intervención más efectiva y contextualizada en la reducción de la siniestralidad laboral, y contribuyendo a superar barreras organizacionales que limitan la apropiación y sostenibilidad de las prácticas preventivas.

XIV.3 Implementación de tecnologías predictivas en la gestión de riesgos laborales

La adopción de tecnologías predictivas en la gestión de riesgos laborales es fundamental para mejorar la anticipación y prevención de accidentes graves, especialmente en sectores de alto riesgo como la minería, la construcción y la manufactura. No obstante, su implementación efectiva enfrenta varios desafíos, sobre todo en empresas de menor tamaño que carecen de los recursos necesarios. Para abordar estos desafíos y facilitar su implementación en el contexto de la Mutual de Seguridad CChC, se pueden considerar las siguientes acciones:

- **Desarrollo de soluciones escalables y accesibles:** Uno de los principales obstáculos para las empresas pequeñas adheridas es la inversión inicial requerida para las tecnologías predictivas. En este sentido, se recomienda que la Mutual ofrezca soluciones escalables y accesibles, mediante el uso de plataformas de análisis de datos en la nube o software de código abierto, que puedan ser implementadas sin grandes inversiones iniciales. Proyectos piloto o versiones de bajo costo de estas soluciones permitirían a las empresas comenzar a utilizar las tecnologías predictivas sin comprometer recursos importantes.

- **Fortalecimiento del apoyo institucional y modelos de colaboración:** Para superar las barreras económicas, la Mutual podría explorar modelos de financiamiento progresivo o subsidios dirigidos a empresas más pequeñas. Además, se recomienda fomentar la colaboración interempresarial entre empresas del mismo sector dentro de la Mutual, para compartir recursos, infraestructura tecnológica y conocimientos sobre las mejores prácticas de seguridad y prevención. Esta colaboración permitiría que las empresas más pequeñas accedan a las herramientas predictivas necesarias sin incurrir en grandes costos adicionales.

- **Fortalecimiento de la capacitación técnica:** La falta de capacitación técnica en las empresas adheridas es otro reto importante. Se recomienda que la Mutual de Seguridad desarrolle programas de formación continua enfocados específicamente en el uso de tecnologías predictivas y análisis de datos. Estos programas deben ser diseñados no solo para capacitar a los empleados de las empresas adheridas, sino también para fortalecer la capacidad técnica de los profesionales que implementarán estas tecnologías dentro de las organizaciones, garantizando que puedan interpretar los datos y actuar preventivamente sobre los riesgos detectados.

- **Cambio cultural organizacional:** La integración de tecnologías predictivas no solo requiere de herramientas tecnológicas, sino también de un cambio cultural organizacional. Para superar la resistencia al cambio en las empresas, es esencial que la Mutual promueva la seguridad como un valor central en las empresas adheridas, involucrando a los empleados en el diseño y evaluación de medidas preventivas. Además, la alta dirección de las empresas debe liderar este proceso, asegurando que la adopción de nuevas tecnologías predictivas sea vista como una oportunidad para mejorar la seguridad laboral.

- **Adaptación de los modelos predictivos a las necesidades del sector:** Cada sector industrial presenta características y riesgos específicos. En este sentido, se recomienda que la Mutual de Seguridad ayude a las empresas a adaptar los modelos predictivos a las particularidades de su sector. Las soluciones personalizadas para sectores como la minería, la construcción o la manufactura permitirán una identificación más precisa de los riesgos y una priorización efectiva de las medidas preventivas. Se recomienda que la

Mutual trabaje directamente con expertos en cada sector para adaptar las tecnologías a las realidades específicas de cada industria.

- **Evaluación continua y ajuste de estrategias:** Para asegurar la efectividad de las tecnologías predictivas, se sugiere que la Mutual implemente sistemas de retroalimentación continua que permitan ajustar las estrategias de seguridad a medida que se obtienen nuevos datos. Estos sistemas permitirían optimizar la gestión de riesgos dentro de las empresas adheridas, asegurando que las intervenciones preventivas sean siempre actualizadas y eficaces.

- **Colaboración con centros de investigación y expertos en IA:** Se recomienda que la Mutual busque colaboraciones estratégicas con universidades y centros especializados en inteligencia artificial (IA). Esto permitiría acceder a modelos predictivos avanzados que mejoren la precisión de las predicciones de accidentes y optimicen las intervenciones preventivas. La Mutual podría facilitar el acceso de sus empresas adheridas a estos modelos de IA generativa y análisis de datos para mejorar tanto la capacidad predictiva como la eficacia de las intervenciones.

Proyección tecnológica y líneas de innovación

Además de las acciones priorizadas para la adopción actual de tecnologías predictivas, se propone una línea de desarrollo futuro orientada a incorporar herramientas más avanzadas de inteligencia artificial. Estas innovaciones no constituyen medidas inmediatas, pero representan oportunidades clave para robustecer la gestión preventiva y avanzar hacia un sistema más adaptativo y contextualizado.

Entre estas proyecciones se destacan:

- **Integración de inteligencia artificial generativa (IAG):** Los modelos de IAG permiten reconocer patrones semánticos complejos en textos incompletos o ambiguos, facilitando una clasificación más precisa de accidentes potencialmente graves y fatales. Esta capacidad mejora el entendimiento del contexto narrativo de los accidentes, incluyendo factores como herramientas, condiciones físicas, fatiga laboral o exposición a agentes nocivos.

- **Aplicación de IA analítica para predicción de concentraciones de riesgo:** Se sugiere explorar el uso de algoritmos que detecten agrupaciones de actividades críticas o condiciones recurrentes de alto riesgo a partir del análisis histórico. Esta tecnología podría anticipar escenarios críticos y orientar la focalización preventiva con mayor precisión.

- **Expansión del análisis contextual:** Incorporar herramientas de procesamiento de lenguaje natural para analizar el entorno organizacional previo al accidente, incluyendo

rutinas laborales, cargas de trabajo y protocolos vigentes. Esto permitiría diseñar intervenciones focalizadas más efectivas.

- **Fortalecimiento de la calidad y actualización de datos:** Se recomienda avanzar en la integración de sensores y dispositivos inteligentes que provean datos en tiempo real sobre las condiciones laborales. Esto aumentaría la capacidad de respuesta ante riesgos emergentes y enriquecería los modelos predictivos existentes.

- **IA en la gestión preventiva adaptativa:** Finalmente, se propone utilizar IA no solo para predecir, sino también para ajustar automáticamente las estrategias preventivas en función de los datos observados, optimizando así la asignación de recursos y el monitoreo de intervenciones en tiempo real.

Estas líneas complementan los mecanismos ya sugeridos y refuerzan el objetivo de una gestión preventiva basada en evidencia, capaz de adaptarse a contextos dinámicos y anticipar escenarios de alto impacto mediante el uso de tecnologías avanzadas.

XV. Referencias

- Acuña Acuña, E. G., Huertas Rosales, Á., & Vásquez Espinoza, S. (2024). Sistemas de monitoreo IoT para la seguridad laboral en Costa Rica. *New Trends in Qualitative Research, 20*(4), e1054. https://doi.org/10.36367/ntqr.20.4.2024.e1054
- Agudo Díaz, F. J., Rubio Collar, M. Á., & Seisdedos Rodríguez, I. (2017). La mejora continua en la gestión de la prevención de riesgos laborales en la empresa desde la vigilancia colectiva de la salud. *Revista de la Asociación Española de Especialistas en Medicina del Trabajo, 26*(1).
- Bhanot, N., Singh, H., Sharma, D., Jain, H., & Jain, S. (2019). Python vs. R: A text mining approach for analyzing the research trends in Scopus database. *arXiv preprint* arXiv:1911.08271. Disponible en: https://arxiv.org/abs/1911.08271
- Boix, P., & Rodríguez de Prada, A. (s.f.). *Manual CBP: Criterios de Buena Práctica Profesional en Actividades Preventivas*. Instituto Nacional de Seguridad y Salud en el Trabajo (INSST). https://www.insst.es/documents/94886/96076/CBP+TEXTO.pdf
- Cardenas-Quispe, G., Pacheco, A., Manrique-Nugent, R., & Quispe-Gonzales, D. (2022). *Análisis de casos confirmados de COVID-19 en la provincia de Cañete mediante clustering K-means.* Revista Peruana de Ciencia de Datos, 4(1), 23–35.
- Cano Cos, A. (2020). *Aprendiendo vectores de palabras a partir de normas de asociación*. Repositorio Unican.
- Centro de Políticas Públicas UC. (2020). *Análisis del sistema de seguridad y salud ocupacional en Chile y propuestas de mejoramiento*. Pontificia Universidad Católica de Chile. https://politicaspublicas.uc.cl/publicacion/analisis-del-sistema-de-seguridad-y-salud-ocupacional-en-chile-y-propuestas-de-mejoramiento/
- Centro de Sistemas Públicos de la Universidad de Chile. (2017). *Evaluación de diseño, procesos, resultados e impacto del Plan de Prevención del Instituto de Seguridad Laboral: Informe final*. Universidad de Chile. https://www.isl.gob.cl/wp-content/uploads/2024/08/2016-2017-Investigacion-Evaluacion-del-Plan-de-Prevencion-del-Instituto-de-Seguridad-laboral.pdf
- EU-OSHA (European Agency for Safety and Health at Work). (2020). *Panorama Europeo de la Seguridad y Salud en el Trabajo*. EU-OSHA.
- Fernández Muñiz, B., Montes Peón, J. M., & Vázquez Ordás, C. J. (2007). Factores precursores de la implantación del sistema de gestión de la seguridad laboral en las empresas españolas. *Revista de Economía Industrial*. https://dialnet.unirioja.es/descarga/articulo/2486895.pdf
- Fernández-Muñiz, B., Montes-Peón, J. M., & Vázquez-Ordás, C. J. (2005). Antecedentes del comportamiento del trabajador ante el riesgo laboral: Un modelo de

cultura positiva hacia la seguridad. *Revista de Psicología del Trabajo y de las Organizaciones, 21*(3), 207–234.

- Fundacentro (Fundação Jorge Duprat Figueiredo de Segurança e Medicina do Trabalho, Brasil). (2020). *Estudios sobre la Seguridad en el Trabajo en Brasil y Latinoamérica*. Fundacentro.
- García, A., & López, B. (2017). Evaluación de riesgos laborales y su relación con los accidentes en la industria manufacturera. *Revista de Riesgos Laborales y Salud Ocupacional, 24*(4), 87–101.
- Guerrero Álvarez, R. (2019). *Aplicación del modelo word2vec para el análisis de sentimientos en tweets en idioma inglés*. DocPlayer.
- Gutiérrez, A., & Fernández, J. (2012). Estrategias de prevención de riesgos laborales en diferentes industrias. *Revista de Prevención de Riesgos Laborales, 18*(2), 45–58.
- Haddad, G., & Anton, M. (2019). Análisis de datos en prevención de riesgos laborales. *Revista de Prevención Analítica, 21*(3), 105–118.
- ICSI. (2019). *Instituto por una Cultura de Seguridad Industrial*. Disponible en: https://www.icsi-eu.org/
- ICSI. (2019). *Los riesgos más importantes*. Disponible en: https://www.icsi-eu.org/
- ILO. (2009). *Estudio general sobre seguridad y salud de los trabajadores*. Organización Internacional del Trabajo.
- ILO. (2017). *Estudio general relativo a los instrumentos de seguridad y salud en el trabajo sobre el marco promocional, la construcción, las minas y la agricultura*. Organización Internacional del Trabajo.
- INSST (Instituto Nacional de Seguridad y Salud en el Trabajo, España). (2021). *Guía sobre la Gestión de la Seguridad en el Trabajo*. Disponible en: https://www.insst.es/
- Institut pour une Culture de Sécurité Industrielle (ICSI). (2019). *Lo esencial de la prevención de los accidentes graves, mortales y tecnológicos mayores*. https://www.icsi-eu.org/sites/default/files/2020-07/Icsi_essentiel_ES_prevencion-accidentes-graves-mortales-technol%C3%B3gicos-mayores_2019.pdf
- Johnson, S., & Brown, D. (2016). Implementación de medidas preventivas efectivas en el sector minero. *Revista de Seguridad en la Minería, 28*(2), 78–92.
- Jones, P., & Smith, R. (2018). Gestión de riesgos laborales en empresas industriales. *Revista de Seguridad en la Industria, 27*(1), 53–68.
- Krause, T., & Russell, L. (1994). The behavior-based approach to proactive accident. *Professional Safety, 39*(3), 22–26.
- Krause, T. (2012). New perspectives in fatality and serious injury prevention. *Presentation at Fatality Prevention Forum 2012*, Coraopolis, PA, USA.
- Kubo, J., Goldstein, B. A., Cantley, L. F., Tessier-Sherman, B., Galusha, D., Slade, M. D., Chu, I. M., & Cullen, M. R. (2013). *Contribution of health status and prevalent chronic disease to individual risk for workplace injury in the manufacturing environment*. Occupational and Environmental Medicine, 70(9), 555–560. https://doi.org/10.1136/oemed-2013-101653

- Marsh Risk Consulting & RIMS. (2016). *II Benchmark de Gestión de Riesgos en Latinoamérica: Evolución, tendencias y oportunidades*. Marsh McLennan. https://www.marsh.com/content/dam/marsh/Documents/PDF/LAC/LAC-es/benchmark-v15.pdf
- Martínez, E., & González, M. (2011). Análisis comparativo de estrategias de prevención de riesgos laborales en la industria. *Revista de Seguridad Industrial, 25*(4), 112–127.
- Martínez-Oropesa, C., & Cremades, L. V. (2012). Liderazgo y cultura en seguridad: su influencia en los comportamientos de trabajo seguros de los trabajadores. *Salud de los Trabajadores, 20*(2), 179–192.
- MINSAL (Ministerio de Salud, Chile). (2020). *Norma Técnica N° 142 para la Notificación de Accidentes de Trabajo y Enfermedades Profesionales*. Disponible en: https://www.minsal.cl
- Ministerio de Salud de Chile. (2018). *Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud, Décima Revisión (CIE-10)*. Disponible en: https://www.minsal.cl
- Muñoz Salazar, P., Calavacero Balza, N., & Díaz Tramolao, P. (2021). Artefactos preventivos y cultura de seguridad en empresas subcontratadas de minería. *Revista Venezolana de Gerencia, 26*(Esp.5), 490–504.
- Mutual de Seguridad. (2022). *Guía DS 67* [PDF]. Disponible en: https://www.mutual.cl
- Oliver, A., Tomás, J. M., & Cheyne, A. (2005). Clima de seguridad laboral: naturaleza y poder predictivo. *Revista de Psicología del Trabajo y de las Organizaciones, 21*(3), 253–268.
- Organización Internacional del Trabajo (OIT). (2011). *Evaluación del sistema de salud y seguridad ocupacional en América Latina*. OIT. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@americas/@ro-lima/@sro-santiago/documents/publication/wcms_178071.pdf
- Organización Internacional del Trabajo (OIT). (2019). *Informe Mundial sobre la Seguridad y la Salud en el Trabajo*. Disponible en: https://www.ilo.org
- Organización Internacional del Trabajo (OIT). (2019). *Safety and health at the heart of the future of work: Building on 100 years of experience*. Ginebra: OIT. https://www.ilo.org/global/topics/safety-and-health-at-work/lang--en/index.htm
- Organización Internacional del Trabajo (OIT) & Pontificia Universidad Católica de Chile. (2011). *Comparación internacional de sistemas de salud y seguridad laboral*. OIT. https://www.ilo.org/wcmsp5/groups/public/@americas/@ro-lima/@sro-santiago/documents/publication/wcms_178071.pdf
- Pérez, J., & Gómez, E. (2021). Prevención de riesgos laborales en el sector químico. *Revista de Prevención en la Industria Química, 29*(2), 40–55.
- Rodríguez, L., & Gómez, F. (2015). Experiencias de éxito en la gestión de riesgos laborales en la industria química. *Revista de Seguridad en la Industria Química, 22*(3), 92–105.

- Rodríguez-Mon Barrera, B. (2021). *Análisis de informes médicos de pacientes mediante técnicas de NLP y consulta de Metatesauros*. Repositorio Comillas.
- Rodríguez-Mon Barrera, R., Guerrero Álvarez, M., & Cano Cos, A. (2021). *Aplicación de técnicas de procesamiento de lenguaje natural y clustering en datos clínicos.* Revista de Ciencias de la Salud, 19(3), 456–471. https://doi.org/10.1016/j.rcs.2021.03.005
- Sánchez, M., & López, N. (2022). Gestión de riesgos laborales en el sector de la construcción. *Revista de Seguridad en la Construcción, 32*(3), 112–125.
- SUSESO (Superintendencia de Seguridad Social). (2023). Disponible en: https://www.suseso.cl
- Trillo Cabello, A. F. (2022). *Accidentalidad en obras de construcción. Análisis con enfoque en las fases de obra*. Universidad de Málaga. https://riuma.uma.es/xmlui/bitstream/handle/10630/24741/TD_TRILLO_CABELLO_Antonio_Cabello.pdf
- Universidad Internacional de Valencia. (2016). *Siniestralidad laboral en Europa y Latinoamérica*. Disponible en: https://www.universidadviu.com
- Valenzuela, C., & Torres, A. (2014). Benchmarking en prevención de riesgos laborales. *Revista de Prevención Comparada, 16*(1), 25–37.
- Vargas, R., & Ramírez, P. (2013). Experiencias de éxito en la gestión de riesgos laborales en el sector de la construcción. *Revista de Seguridad en la Construcción, 30*(1), 67–81.
- World Health Organization (WHO) & International Labour Organization (ILO). (2021). *WHO/ILO joint estimates of the work-related burden of disease and injury, 2000–2016: Global monitoring report*. World Health Organization. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed_dialogue/@lab_admin/documents/publication/wcms_819788.pdf