



Serie Proyectos de Investigación e Innovación

Superintendencia de Seguridad Social
Santiago - Chile

INFORME FINAL

Caracterización de la violencia laboral de la población trabajadora de Mutual de Seguridad, en base al relato del accidente.

Autores:

Carlos Flores Guzmán
Cristóbal Torres Rojas
Boris Garrido Arévalo
Camila Arlegui Méndez
Raimundo Mujica Sotomayor

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.

Índice

1	Introducción y Antecedentes	7
2	Definición del problema, pregunta de investigación y objetivos.....	11
2.1	<i>Definición del problema.....</i>	<i>11</i>
2.2	<i>Pregunta de investigación</i>	<i>12</i>
2.3	<i>Objetivos.....</i>	<i>12</i>
2.3.1	<i>Objetivo General.....</i>	<i>12</i>
2.3.2	<i>Objetivos Específicos</i>	<i>13</i>
3	Revisión de la literatura o experiencias relevantes	14
3.1	<i>Conceptualización de la violencia laboral: marco teórico y delimitaciones.....</i>	<i>14</i>
3.2	<i>Perspectiva internacional: evolución conceptual y metodológica</i>	<i>15</i>
3.3	<i>Enfoques tecnológicos: NLP y Machine Learning.....</i>	<i>16</i>
3.4	<i>Experiencias y evidencia en América Latina</i>	<i>17</i>
3.5	<i>Situación en Chile y relevancia institucional.....</i>	<i>18</i>
3.5.1	<i>Perspectiva de la legislación chilena sobre violencia laboral.....</i>	<i>18</i>
4	Descripción de la metodología de investigación.....	20
4.1	<i>Preparación y limpieza de datos.....</i>	<i>20</i>
4.2	<i>Justificación del tamaño muestral y confiabilidad.....</i>	<i>21</i>
4.3	<i>Fine-tuning de modelos LLM para clasificación binaria y por tipo de violencia.....</i>	<i>22</i>
4.3.1	<i>Clasificación binaria de violencia laboral</i>	<i>22</i>
4.3.2	<i>Clasificación según tipo de violencia (tipología Cal/OSHA).....</i>	<i>22</i>
4.4	<i>Validación experta.....</i>	<i>23</i>
4.4.1	<i>Validación experta de etiquetado.....</i>	<i>23</i>
4.4.2	<i>Validación experta de revisión.....</i>	<i>23</i>
4.5	<i>Análisis de caracterización mediante técnicas de NLP</i>	<i>24</i>
4.6	<i>Análisis cruzado con variables estructuradas.....</i>	<i>25</i>
4.7	<i>Interpretabilidad del modelo: SHAP Values.....</i>	<i>25</i>
4.8	<i>Generación de recomendaciones preventivas</i>	<i>25</i>
5	Resultados.....	26
5.1	<i>Procesamiento de Datos.....</i>	<i>26</i>
5.2	<i>Clasificación de Denuncias.....</i>	<i>27</i>
5.2.1	<i>Fine-tuning de modelos LLM para clasificación binaria</i>	<i>27</i>
5.2.2	<i>Justificación de la muestra y confiabilidad (Casos Totales)</i>	<i>29</i>
5.2.3	<i>Fine-tuning de modelos LLM para clasificación por tipo de violencia</i>	<i>30</i>
5.2.4	<i>Justificación de la muestra y confiabilidad (Tipo de Violencia)</i>	<i>31</i>
5.3	<i>Caracterización de Casos.....</i>	<i>32</i>
5.3.1	<i>Separación por Clústers</i>	<i>32</i>
5.3.2	<i>Dinámicas de escalamiento</i>	<i>34</i>

5.4	<i>Identificación de variables relevantes mediante red neuronal</i>	35
5.5	<i>Análisis descriptivo por Tipos de Violencia</i>	38
6	Discusión de Resultados	47
7	Recomendaciones para la prevención por tipo de violencia laboral	49
7.1	<i>Violencia externa</i>	49
7.2	<i>Violencia de servicio</i>	50
7.3	<i>Violencia interna</i>	51
8	Conclusiones	53
9	Referencias	55

Índice de Tablas

Tabla 1. Variables descriptoras de los relatos	27
Tabla 2. Matriz de Confusión - Modelo Violencia vs. No Violencia.....	28
Tabla 3. Indicadores - Modelo Violencia vs. No Violencia.....	28
Tabla 4. Matriz de Confusión - Modelo tipo de Violencia	30
Tabla 5. Indicadores - Modelo tipo de Violencia.....	31
Tabla 6. Clústers de relatos según similitudes semánticas y contextuales	32
Tabla 7. Distribución de Clústers por tipo de Violencia.....	32
Tabla 8. Términos recurrentes por Clúster.....	33
Tabla 9. Tabla Dinámicas de escalamiento	33

Índice de Figuras

Figura 1. Importancia de Variables - Método SHAP values en Modelo Violencia vs. No Violencia.....	36
Figura 2. Relación entre Tipo de Violencia, Profesión/Ocupación, Diagnóstico y Tipo de Accidente mediante gráfico Sankey	37
Figura 3. Número de casos por tipo de violencia y tipo de accidente.	38
Figura 4. Número de casos en violencia Externa y Accidentes de Trabajo por Actividad Económica	39
Figura 5. Proporción de Tipos de Violencia por Año.	40
Figura 6. Cantidad Total de Tipos de Violencia por Macro Segmento PYME.	41
Figura 7. Distribución de Tipos de Violencia por Actividad Económica.....	42
Figura 8. Distribución de Tipos de Violencia por Profesión/Ocupación más comunes.....	43
Figura 9. Cantidad total de Tipos de Violencia por Gravedad.	44
Figura 10. Cantidad total de Tipos de Violencia por Diagnóstico.	45
Figura 11. Cantidad total de Tipos de Violencia por Tipo Trastornos más comunes.	45

Resumen Ejecutivo

El presente proyecto busca abordar la creciente preocupación por la violencia laboral en Chile, caracterizándola en la población trabajadora afiliada a la Mutual de Seguridad, a partir del análisis de relatos de accidentes laborales. Para ello, se aplican técnicas avanzadas de Machine Learning como Procesamiento de Lenguaje Natural (NLP) y Modelo Grande de Lenguaje (LLM), permitiendo identificar y clasificar automáticamente los distintos tipos de violencia laboral definidos por la tipología Cal/OSHA: violencia tipo 1 (externa), tipo 2 (asociada a servicios) y tipo 3 (relaciones laborales, interna).

La investigación, de carácter retrospectivo y transversal, se apoya en el análisis de más de una década de registros históricos (2011-2023), que incluyen descripciones textuales de accidentes, datos administrativos y diagnósticos clínicos. Se comienza con la recolección, limpieza y estructuración de los datos, así como en el ajuste preliminar de modelos de lenguaje adaptados al contexto chileno. El análisis exploratorio ha permitido visibilizar patrones iniciales de violencia y establecer relaciones entre variables clave.

Este enfoque no solo permitió una caracterización más precisa y contextualizada de los hechos de violencia, sino que también proporcionó herramientas predictivas para anticipar denuncias, fortaleciendo la prevención y mejorando la gestión del riesgo en los entornos laborales. En última instancia, los hallazgos del proyecto buscan contribuir a la construcción de lugares de trabajo más seguros, saludables y equitativos, en línea con los desafíos actuales de la seguridad y salud ocupacional en Chile.

Palabras Claves

Violencia Laboral, trabajadores, Machine Learning, Modelos Grandes de Lenguaje (LLM)

1 Introducción y Antecedentes

La violencia laboral constituye un fenómeno complejo, multicausal y de creciente preocupación a nivel mundial. Reconocida por organismos internacionales como la Organización Internacional del Trabajo (OIT) y la Organización Mundial de la Salud (OMS), la violencia en el lugar de trabajo no solo compromete el bienestar físico y psicológico de los trabajadores, sino que también afecta negativamente la productividad, el clima organizacional y la sostenibilidad de las instituciones. Esta problemática ha adquirido una relevancia particular en contextos caracterizados por alta presión laboral, contacto frecuente con el público, desigualdades estructurales o deficiencias en la gestión de riesgos psicosociales.

En Chile, la preocupación por la violencia en el trabajo ha ido en aumento en las últimas décadas. Según datos de Mutual de Seguridad, entre los años 2016 y 2025 se observó un aumento del 62.83% en los casos reportados de violencia laboral, lo que representa una señal de alerta para los sistemas de prevención y protección de la salud ocupacional en el país, especialmente en sectores como Transporte, Servicios, Comercio y Educación que representan un 24%, 18%, 14% y 6% respectivamente, siendo casi el 60% de los casos, lo cual podría deberse a que los trabajadores están particularmente expuestos a situaciones de agresión, acoso o intimidación.

El abordaje de este fenómeno ha evolucionado desde una perspectiva exclusivamente jurídica o disciplinaria, hacia enfoques más integrales, centrados en la prevención, la detección temprana y la intervención basada en evidencia. En este marco, resulta clave contar con sistemas eficaces para identificar los distintos tipos de violencia que pueden manifestarse en el entorno laboral. Una de las tipologías más reconocidas internacionalmente es la desarrollada por la División de Seguridad y Salud Ocupacional de California (Cal/OSHA), que clasifica la violencia en tres tipos:

- **Tipo 1:** Violencia externa, donde el agresor no tiene vínculo con la víctima ni con la organización (por ejemplo, asaltos);
- **Tipo 2:** Violencia asociada a servicios, proveniente de clientes, pacientes o usuarios;
- **Tipo 3:** Violencia interna, que ocurre entre colegas, supervisores o subordinados.

Aunque esta clasificación ha sido adoptada en Chile como referente, su aplicación ha sido limitada debido a la escasez de estudios que caractericen detalladamente los hechos de violencia en los distintos sectores productivos. En particular, la violencia que se manifiesta en relatos espontáneos de accidentes laborales —a menudo redactados por los propios trabajadores o por profesionales de salud en atención primaria— ha sido poco explorada como fuente de información válida para comprender las dinámicas, contextos y consecuencias de estos hechos.

En el ámbito de seguridad laboral, se han comenzado a implementar tecnologías de vanguardia, como el Machine Learning, para analizar masivamente estos relatos y detectar

automáticamente denuncias relacionadas con violencia. Este avance marca un hito relevante en la incorporación de la inteligencia artificial en la gestión de la salud y seguridad ocupacional. No obstante, persiste una brecha significativa en la caracterización fina de los hechos de violencia: aún no se cuenta con una descripción sistemática de los contextos, actores involucrados, impactos físicos y psicológicos, y factores de riesgo asociados a cada tipología de violencia identificada.

En este contexto, surge el presente proyecto de investigación, cuyo propósito central es caracterizar los hechos de violencia laboral que afectan a la población trabajadora afiliada a la Mutual de Seguridad, a partir de los relatos de accidentes laborales registrados entre 2011 y 2025. Esta iniciativa se enmarca en un enfoque de análisis secundario de datos textuales, y busca aportar evidencia científica útil para la toma de decisiones, el diseño de políticas de prevención más efectivas, y la mejora de las condiciones de trabajo en el país.

El proyecto no solo responde a una necesidad técnica y epidemiológica, sino también a un mandato ético y normativo. La reciente promulgación de la Ley Karin en Chile, que establece nuevos estándares para la prevención, investigación y sanción del acoso laboral, refuerza la obligación de las organizaciones de desarrollar mecanismos proactivos y basados en datos para enfrentar la violencia en el trabajo. En este sentido, el desarrollo de herramientas automatizadas de análisis, junto con la validación experta de los hallazgos, representan un avance estratégico para Mutual de Seguridad y para el sistema de seguridad social en su conjunto.

Asimismo, este proyecto se nutre de una amplia base de antecedentes internacionales y nacionales que respaldan la utilización de modelos de lenguaje y técnicas de minería de texto como herramientas eficaces para analizar grandes volúmenes de datos no estructurados. Por ejemplo, *Recent Advances in Using Natural Language Processing to Address Public Health Research Questions Using Social Media and Consumer-Generated Data* (Conway, Hu & Chapman, 2019) destaca cómo el NLP ha sido aplicado con éxito en el análisis de datos textuales en salud pública. De manera complementaria, *Machine Learning in Occupational Safety and Health: A Systematic Review* (Maheronnaghsh & Zolfagharnasab, 2023) y *Named Entity Recognition in Electronic Health Records: A Methodological Review* (Durango, Torres-Silva & Orozco-Duque, 2023) muestran el potencial de estas tecnologías para identificar patrones de riesgo, predecir eventos críticos y fortalecer las capacidades institucionales en contextos complejos. En América Latina, el estudio *Metodología para el análisis de la violencia en el departamento de Bolívar mediante técnicas de Machine Learning* (Fernández Caraballo & Gómez Franco, 2018) evidencia la aplicabilidad de estos enfoques para abordar fenómenos sociales como la violencia laboral, utilizando bases de datos locales. Estas investigaciones respaldan la pertinencia de aplicar herramientas de inteligencia artificial, adaptadas al contexto chileno, para caracterizar hechos de violencia laboral a partir de relatos no estructurados.

A pesar del creciente interés institucional y académico por la violencia en el trabajo, persisten importantes brechas de conocimiento que limitan el desarrollo de intervenciones

eficaces. En primer lugar, existe una escasa caracterización por tipo y contexto de violencia laboral en Chile, especialmente en relación con la tipología internacionalmente reconocida (Cal/OSHA). Esta ausencia ha sido señalada por estudios como el de Ansoleaga (Ansoleaga, 2022), quienes advierten que la mayoría de las investigaciones en Chile se han concentrado en diagnósticos generales o en sectores específicos como salud, sin profundizar en las diferencias según tipo de agresor, vínculo con la víctima o contexto organizacional.

En segundo lugar, se evidencia una subutilización de fuentes textuales no estructuradas (como los relatos de accidentes laborales) en los sistemas de registro administrativos. Estos relatos, si bien están disponibles, no han sido sistemáticamente analizados ni integrados en los procesos de toma de decisiones, lo que representa una pérdida significativa de conocimiento situado. Como indica la evidencia internacional (Conway, 2019), el análisis de texto puede revelar patrones complejos de riesgo, emociones asociadas y relaciones entre eventos, lo cual es esencial para comprender fenómenos multifactoriales como la violencia laboral.

En tercer lugar, hay una falta de validación local de modelos predictivos o clasificatorios desarrollados en otros contextos, los cuales no siempre se ajustan a las dinámicas culturales, lingüísticas y normativas del país. Estudios como los de Maheronnaghsh y Zolfagharnasab (2023) destacan el potencial del aprendizaje automático para gestionar riesgos psicosociales, pero también insisten en la necesidad de adaptar los modelos a los datos específicos de cada región.

La incorporación de variables de género y condiciones estructurales sigue siendo deficiente en gran parte de las investigaciones. A pesar de que la violencia laboral puede tener expresiones diferenciadas según sexo, jerarquía o sector, estas dimensiones son frecuentemente tratadas de forma superficial o excluidas del análisis. Parra Osorio y Acosta Fernández (2010) subrayan la necesidad de incluir estas variables de manera explícita para evitar enfoques reduccionistas o normativos que invisibilicen las desigualdades estructurales en el mundo del trabajo.

El presente proyecto se hace cargo de estas brechas, proponiendo una estrategia metodológica robusta basada en técnicas de Machine Learning y modelos de lenguaje. Esta aproximación permitió avanzar en una caracterización más detallada de los hechos de violencia laboral, generar conocimiento contextualizado y fortalecer las capacidades institucionales para diseñar medidas preventivas más enfocadas.

Finalmente, el proyecto tiene un alto potencial de impacto en el campo de la Seguridad y Salud en el Trabajo (SST). Se estima que con las herramientas y metodologías generadas en la presente investigación, se podrá identificar correctamente sobre el 90% de los hechos de violencia según su tipo, lo que a nivel práctico se puede llegar a traducir en:

- Detectar y clasificar automáticamente hechos de violencia laboral y su respectivo tipo, incluso en grandes volúmenes de datos no estructurados.

- Identificar sectores económicos, perfiles ocupacionales y condiciones laborales asociadas con mayor riesgo de violencia.
- Generar alertas tempranas, medidas de prevención específicas y recomendaciones de política pública basadas en evidencia.

Además, al ser una solución escalable y replicable, esta iniciativa se puede extender a otras mutuales y sectores laborales del país, contribuyendo a reducir la exposición de los trabajadores a entornos hostiles, fortalecer la vigilancia de los factores psicosociales y mejorar la calidad del empleo en Chile.

En resumen, este proyecto se inscribe en un momento clave para la salud ocupacional en Chile. Aporta una mirada innovadora y profundamente contextualizada a una problemática urgente, con el potencial de generar transformaciones significativas tanto en la forma de comprender la violencia en el trabajo como en las estrategias para prevenirla y erradicarla.

2 Definición del problema, pregunta de investigación y objetivos

2.1 Definición del problema

La necesidad de caracterizar la violencia laboral con precisión en Chile surge porque, si bien las intervenciones han aumentado en cantidad, su eficacia se ve comprometida por la falta de diseño contextualizado, estandarización de protocolos y evaluación rigurosa, lo que limita su impacto real. Estudios nacionales destacan que muchas medidas se implementan de forma reactiva, sin un marco evaluativo claro y sin considerar las características específicas del entorno laboral chileno. Por ejemplo, Ansoleaga (p.17, 2022) señalan la *“escasa literatura de intervenciones sobre violencia laboral”* y advierten que la ausencia de estándares dificulta comparar resultados y valorar la efectividad de las acciones preventivas (no se consideran contextos, ni se evalúan resultados) (OIT, 2022; Ansoleaga, Ahumada & Henríquez, 2022). A nivel internacional, una revisión sistemática encontró que solo el 55% de las evaluaciones incluye seguimiento pre y post intervención, y que la mayoría de los estudios se concentra en países desarrollados y el sector salud, sin replicabilidad directa para Chile (Palma, Ansoleaga & Ahumada, 2018). En consecuencia, aunque se han implementado programas de capacitación, protocolos y políticas, las instituciones carecen de mecanismos robustos para medir su efectividad, lo que genera brechas entre la intención preventiva y los resultados reales. En este contexto, estas limitaciones no solo afectan la identificación y clasificación de denuncias, sino también la capacidad de generar políticas verdaderamente eficaces y sostenibles.

Como se señaló en la sección anterior, el aumento sostenido de los casos reportados y la adopción reciente de marcos regulatorios más exigentes, como la Ley Karin, han instalado con fuerza este fenómeno en la agenda pública y organizacional. Sin embargo, aún persiste una brecha crítica entre el reconocimiento del problema y la capacidad real de las instituciones para identificar, clasificar y analizar los hechos de violencia en sus múltiples dimensiones.

En particular, Mutual de Seguridad ha dado pasos importantes al comenzar a utilizar herramientas de análisis automático para detectar hechos de violencia en los relatos de accidentes laborales. No obstante, estos esfuerzos iniciales se han enfocado principalmente en la detección binaria (violento/no violento), sin alcanzar un nivel de detalle suficiente que permita diferenciar los tipos de violencia ni comprender a fondo sus características contextuales. Esto se traduce en una visión parcial del fenómeno y limita las posibilidades de diseñar medidas específicas para cada tipo de agresión, relación entre las partes involucradas o sector de actividad.

El problema central, por tanto, no es la ausencia total de datos, sino la falta de una estructura analítica adecuada para extraer conocimiento útil desde los datos ya disponibles,

particularmente aquellos en formato textual, como los relatos espontáneos de trabajadores y profesionales de salud. Estos textos, si bien contienen información clave sobre los hechos, relaciones laborales, consecuencias y entornos donde ocurre la violencia, no han sido sistemáticamente explotados con enfoques técnicos avanzados, ni los resultados sobre posibles clasificaciones de tipo de violencia validados bajo una tipología reconocida, como la propuesta por Cal/OSHA.

La consecuencia de esta situación es doble: por un lado, se pierde una oportunidad de generar conocimiento estratégico a partir de fuentes ya existentes, y por otro, se perpetúa una gestión de riesgos basada en intuiciones o registros fragmentarios, en lugar de evidencia robusta y contextualizada. Esta falta de caracterización impide establecer focos de prevención específicos, analizar tendencias temporales, o identificar sectores, cargos o condiciones laborales más vulnerables.

Frente a este diagnóstico, se hace necesaria una investigación que no solo detecte la violencia laboral en los registros de Mutual de Seguridad, sino que logre comprenderla en toda su complejidad, diferenciando tipos, relaciones, consecuencias y condiciones que la propician. Esta comprensión detallada es el insumo esencial para avanzar desde la mera vigilancia hacia una gestión preventiva proactiva, sensible al contexto y alineada con estándares internacionales.

2.2 Pregunta de investigación

En este marco, la pregunta que orienta esta investigación es la siguiente:

¿Cuáles son las características de los hechos de violencia que afectan a la población de trabajadores de Mutual de Seguridad?

Esta interrogante abarca tanto la tipificación de las denuncias como la descripción de los contextos donde ocurren, las relaciones entre víctimas y agresores, y los efectos reportados tras los eventos. Su abordaje permitirá no solo responder a una necesidad institucional concreta, sino también contribuir al desarrollo de un modelo replicable para otras organizaciones e industrias del país que enfrenten desafíos similares en la gestión de la violencia laboral.

2.3 Objetivos

2.3.1 Objetivo General

Caracterizar los hechos de violencia que afectan a los trabajadores de Mutual de Seguridad según la tipificación nacional vigente.

2.3.2 Objetivos Específicos

1. Identificar los tipos específicos de violencia que afectan a los trabajadores de Mutual de Seguridad, según la tipificación nacional vigente.
2. Analizar las características y contextos en los que se producen los actos de violencia contra los trabajadores de Mutual de Seguridad.
3. Describir las principales consecuencias físicas y psicológicas asociadas a los diferentes tipos de violencia que afectan a los trabajadores en el entorno laboral.
4. Proponer focos de prevención específicos para la prevención y mitigación de la violencia laboral, basadas en los resultados obtenidos en la caracterización.

3 Revisión de la literatura o experiencias relevantes

La violencia laboral es un fenómeno transversal que ha captado una atención creciente en la literatura científica, políticas públicas y prácticas de gestión organizacional, especialmente en contextos de alta interacción interpersonal. Diversos estudios coinciden en señalar que este tipo de violencia tiene un impacto negativo directo sobre la salud física y mental de los trabajadores, incluyendo trastornos psicológicos como ansiedad, depresión y estrés postraumático (Mento et al., 2020; Ávila Urdaneta, 2020). Asimismo, la evidencia muestra que genera consecuencias organizacionales significativas, expresadas en aumento del ausentismo, rotación de personal, deterioro del clima laboral y disminución de la productividad (Parra Osorio & Acosta Fernández, 2010; EU-OSHA, 2021). En este sentido, su abordaje efectivo requiere enfoques multidisciplinarios que integren conocimientos desde las ciencias sociales, la salud ocupacional, la gestión organizacional y, más recientemente, la ciencia de datos.

3.1 Conceptualización de la violencia laboral: marco teórico y delimitaciones

La violencia laboral es un concepto amplio y multifacético que ha sido definido desde distintas disciplinas, incluyendo la salud pública, el derecho laboral, la psicología organizacional y la sociología del trabajo. En 2003, la Organización Internacional del Trabajo (OIT), a través de un comité de expertos, elaboró un conjunto de recomendaciones centradas en el sector de servicios, que incluyeron también definiciones de carácter general útiles para comprender el fenómeno de la violencia en el trabajo. En este marco, la violencia se entiende como *“Toda acción, incidente o comportamiento que se aparta de lo razonable mediante el cual una persona es agredida, amenazada, humillada o lesionada por otra en el ejercicio de su actividad profesional o como consecuencia directa de la misma.”* (OIT, p.4, 2003). A su vez, se diferencia entre violencia interna —aquella que ocurre entre trabajadores dentro de la organización, incluyendo directivos y supervisores— y violencia externa, que involucra interacciones violentas entre trabajadores (y sus superiores) y terceras personas presentes en el entorno o situación laboral (OIT, 2003).

Cabe destacar que no toda tensión, conflicto o situación desagradable constituye violencia laboral. La literatura académica insiste en la importancia de diferenciar entre incidentes puntuales, conflictos funcionales propios del trabajo y actos sistemáticos de violencia o acoso. En este sentido, Parra Osorio y Acosta Fernández (2010) advierten que el acoso psicológico laboral debe entenderse como un proceso de conductas reiteradas y hostiles, sostenidas en el tiempo y capaces de generar un daño real en la víctima. No obstante, esta perspectiva resulta insuficiente para abarcar la totalidad del fenómeno, pues excluiría aquellas formas de violencia aisladas pero de gran impacto, como ocurre en la violencia tipo 1 (externa) o tipo 2 (proveniente de usuarios o clientes), que pueden darse tanto en el lugar de trabajo como en los trayectos hacia o desde él. Del mismo modo, la normativa chilena (Ley Karin, Ley N° 21.643) establece criterios para distinguir el acoso

laboral del ejercicio legítimo de la autoridad o el conflicto laboral ordinario. Para los fines de este proyecto, se excluyen de la definición de violencia laboral:

- Accidentes laborales sin componente relacional o social (por ejemplo, caídas, fallas técnicas, etc.), o sin intencionalidad.

Es decir, el análisis se centrará en aquellos relatos que describan comportamientos que implican daño intencionado, desequilibrio de poder, amenaza o agresión explícita o implícita, en línea con las definiciones establecidas en el campo de la salud ocupacional y los marcos normativos vigentes.

3.2 Perspectiva internacional: evolución conceptual y metodológica

En el plano internacional, *Workplace violence and mental health outcomes among healthcare workers: A systematic review* (Mento et al., 2020) documenta que la exposición a violencia en el lugar de trabajo está asociada con efectos negativos en la salud mental del personal sanitario, incluidos trastornos como ansiedad, depresión, estrés postraumático, agotamiento emocional, alteraciones del sueño y fatiga crónica. Por este motivo, este trabajo refuerza la importancia de implementar estrategias preventivas basadas en la evidencia —como intervenciones grupales, capacitación en manejo de agresiones y protocolos de contención psicológica— adaptadas a distintos entornos laborales. En Europa, el informe ESENER 3 de la Agencia Europea para la Seguridad y la Salud en el Trabajo (EU-OSHA, 2021) identifica la violencia laboral como un riesgo emergente que impacta de forma desproporcionada a sectores como salud, educación, transporte y comercio minorista, particularmente por la exposición frecuente al público, entornos estresantes o presiones temporales que exacerban la vulnerabilidad

Desde el punto de vista metodológico, se ha observado una evolución en el abordaje de este fenómeno: inicialmente dominado por encuestas y autoinformes, ha transitado hacia enfoques más complejos que incorporan datos administrativos, análisis de texto, redes sociales y minería de datos. En Estados Unidos, la Occupational Safety and Health Administration (OSHA) ha desarrollado guías específicas para abordar la violencia laboral, con énfasis en las políticas organizacionales y las intervenciones tempranas.

Uno de los aportes más relevantes en la clasificación conceptual es la tipología de la División de Seguridad y Salud Ocupacional de California (Cal/OSHA), que distingue entre tres tipos de violencia:

- **Tipo 1 (externa):** sin relación entre el atacante y la víctima, como en robos o asaltos.
- **Tipo 2 (servicio):** proveniente de usuarios, pacientes o clientes.
- **Tipo 3 (interna):** entre trabajadores o con vínculo laboral previo.

Esta clasificación ha sido ampliamente validada a nivel internacional y se ha convertido en una referencia clave para las organizaciones que buscan estructurar sus políticas de prevención.

Si bien la tipología de Cal/OSHA es la más utilizada a nivel internacional, existen otras clasificaciones relevantes en la literatura, como la propuesta por la OIT (OIT, 2003), que distingue entre violencia física, violencia psicológica, acoso moral y acoso sexual. Esta clasificación ha sido particularmente influyente en la formulación de políticas públicas en países latinoamericanos y en convenios internacionales como el Convenio 190 sobre la eliminación de la violencia y el acoso en el mundo del trabajo, ratificado por Chile en 2022.

En el caso chileno, la legislación vigente incorpora parcialmente estas perspectivas. La recientemente promulgada Ley Karin (Ley N° 21.643) establece una definición legal de acoso laboral, enfocada en conductas persistentes y reiteradas que resulten en menoscabo, maltrato o humillación. Sin embargo, no adopta explícitamente una tipología integral como la de Cal/OSHA o la OIT, lo que genera cierto vacío conceptual en la práctica institucional. Por esta razón, el presente proyecto considera pertinente complementar la normativa nacional con la clasificación de Cal/OSHA, ya que permite operacionalizar mejor el análisis a partir de los relatos y distinguir entre violencia externa, de servicio e interna, aspectos que no están diferenciados explícitamente en el marco legal chileno actual.

3.3 Enfoques tecnológicos: NLP y Machine Learning

En los últimos años, el uso de tecnologías emergentes ha abierto nuevas posibilidades para la detección, clasificación y predicción de hechos de violencia laboral. Maheronnaghsh y Zolfagharnasab (2023) realizaron una revisión sistemática de 42 estudios sobre la aplicación de técnicas de Machine Learning en seguridad y salud ocupacional, mostrando que estas herramientas han sido empleadas principalmente en la construcción, minería, manufactura y trabajos de oficina. Sus aplicaciones abarcan desde la clasificación de narrativas de accidentes hasta la predicción de riesgos ergonómicos, caídas, fatiga y exposición a ruido, con resultados que han permitido anticipar eventos críticos y diseñar indicadores de riesgo más precisos. Entre los aprendizajes destacan la necesidad de contar con bases de datos de calidad, la utilidad de algoritmos como Random Forest y Support Vector Machines para clasificar accidentes y predecir riesgos, y la portabilidad de los modelos a distintos contextos.

Por su parte, Tang et al. (2023) aplicaron algoritmos de aprendizaje automático a una muestra de 4.066 enfermeras en Columbia Británica (Canadá), identificando predictores psicosociales clave de violencia laboral. El estudio reveló que la falta de seguridad física, una deficiente gestión de cargas de trabajo y una baja protección psicológica eran los factores más asociados a la violencia tipo II (provocada por pacientes o visitantes), mientras que la falta de civismo y respeto, la cultura organizacional negativa y el escaso apoyo psicológico

se relacionaban con la violencia tipo III (entre trabajadores). Estos hallazgos muestran que el ML no solo permite detectar patrones de riesgo, sino también priorizar intervenciones concretas en ámbitos como el sector salud, contribuyendo a una mejor prevención y gestión de la violencia en el trabajo.

El Procesamiento de Lenguaje Natural (NLP) ha sido particularmente útil en el campo de análisis de datos no estructurados de textos que representan relatos de personas. Conway, Hu y Chapman (2019) muestran que, más allá de la simple clasificación de textos, estas técnicas han permitido detectar causas y dinámicas subyacentes de problemas de salud pública, como la desconfianza hacia las instituciones sanitarias o la vacilación frente a las vacunas, así como mapear narrativas emergentes en torno a enfermedades y consumo de sustancias. Un aporte clave ha sido la identificación de efectos psicológicos expresados en línea, tales como depresión, ansiedad, estrés postraumático o ideación suicida, evidenciando la capacidad del NLP para modelar cómo las personas comunican su malestar y sus trayectorias emocionales en entornos digitales. Estos aprendizajes no solo profundizan en la comprensión del fenómeno de la violencia laboral y sus impactos psicosociales, sino que también ofrecen herramientas para anticipar riesgos y diseñar intervenciones preventivas más sensibles y efectivas.

El desarrollo de modelos de lenguaje de gran escala (LLM) como GPT y BERT ha ampliado aún más el potencial analítico. Estos modelos permiten trabajar con grandes volúmenes de texto con altos niveles de precisión y contextualización, ofreciendo una alternativa poderosa a los métodos de codificación manual que suelen ser costosos y propensos a sesgos.

3.4 Experiencias y evidencia en América Latina

En América Latina, los estudios también han comenzado a explorar la aplicación de inteligencia artificial en el análisis de violencia. Fernández Caraballo y Gómez Franco (2018) aplicaron técnicas de Machine Learning para analizar patrones de violencia en el departamento de Bolívar (Colombia), evidenciando cómo estos enfoques permiten superar las limitaciones de la estadística tradicional al capturar interacciones complejas entre múltiples variables. Su estudio demostró que la combinación de datos históricos, contexto geográfico y variables sociodemográficas mejora significativamente la capacidad predictiva de los modelos.

En el ámbito específico del acoso laboral, Ávila Urdaneta (2020) analizó el “mobbing” como factor de riesgo psicosocial que puede incluso derivar en accidentes laborales. Este trabajo destaca cómo la violencia psicológica deteriora la salud mental de los trabajadores, favoreciendo estados de hipervigilancia, ansiedad y estrés extremo. Por su parte, Parra Osorio y Acosta Fernández (2010) realizaron una revisión sobre el acoso psicológico en

sectores como la educación y la salud, identificando una fuerte correlación entre ciertas dinámicas organizacionales y la persistencia del acoso.

3.5 Situación en Chile y relevancia institucional

En el contexto nacional, el Informe Anual de Seguridad y Salud en el Trabajo de SUSESO (2022) advierte que el 67 % de las licencias médicas por enfermedad profesional se deben a diagnósticos de salud mental, lo que sugiere que condiciones laborales adversas, como violencia o acoso, son factores relevantes en el incremento de estos casos. La reciente promulgación de la Ley Karin marca un punto de inflexión, al exigir a las organizaciones políticas preventivas más estrictas, protocolos formales y capacitaciones sobre violencia y acoso. Esta normativa impone nuevos estándares y al mismo tiempo habilita un marco para la innovación institucional, en el cual proyectos como este adquieren una relevancia estratégica.

No obstante, subsiste el desafío de avanzar hacia una caracterización más detallada y contextualizada de los distintos tipos de violencia, elemento esencial para orientar medidas preventivas específicas y efectivas.

3.5.1 Perspectiva de la legislación chilena sobre violencia laboral

La legislación chilena reconoce y aborda la violencia en el trabajo desde una perspectiva de salud y seguridad laboral, aunque no adopta explícitamente la tipología de Cal/OSHA (Tipo 1, 2 y 3), sí incorpora elementos de esta clasificación de manera implícita, especialmente al tratar la violencia como un riesgo psicosocial laboral.

i. Marco legal principal

- **Ley N° 16.744** sobre accidentes del trabajo y enfermedades profesionales:
Reconoce como accidente del trabajo cualquier hecho ocurrido por causa o con ocasión del trabajo, lo que incluye actos de violencia externa o interna si están vinculados a las funciones laborales.
- **Decreto Supremo N° 67** del Ministerio del Trabajo:
Establece que los actos de terceros que afecten a un trabajador en el ejercicio de sus funciones (como asaltos, agresiones, amenazas) pueden ser calificados como accidente del trabajo.
- **Ley 21.643:**
Establece un protocolo obligatorio de actuación para la prevención, investigación y sanción del acoso laboral, acoso sexual y otras formas de violencia en el trabajo.

Esta ley modifica el Código del Trabajo e incorpora nuevas obligaciones para empleadores, trabajadores y órganos fiscalizadores, buscando reforzar la protección frente a situaciones de violencia y acoso en el ámbito laboral.

- **Protocolo de Vigilancia de Riesgos Psicosociales en el Trabajo (MINSAL):**
Instrumento técnico obligatorio para ciertas empresas, que aborda el acoso, la violencia y el maltrato laboral como factores de riesgo psicosocial que deben ser evaluados, gestionados y prevenidos.

ii. **Enfoque predominante en la normativa**

- **Enfoque preventivo del riesgo psicosocial:**
La violencia (especialmente la interna y la del tipo 2) es tratada como parte de los factores organizacionales que pueden afectar la salud mental y física de los trabajadores.
- **Enfoque de accidente del trabajo:**
Cuando hay agresión o violencia externa (tipo 1), siempre que esté relacionada con el trabajo, se reconoce como accidente laboral, con derecho a atención por la mutual correspondiente.
- **Ley N° 21.643 (2024)** modifica el código del trabajo e introdujo el concepto de acoso laboral, acoso sexual y violencia en el trabajo, el cual se encuadra dentro de la violencia interna (tipo 3). Se establece la obligación del empleador de tomar medidas para prevenir y sancionar el acoso laboral, acoso sexual y violencia en el trabajo y permite que el trabajador afectado denuncie el hecho a través de los canales legales y administrativos.

4 Descripción de la metodología de investigación

Este proyecto adopta un enfoque metodológico cuantitativo, cualitativo, retrospectivo y transversal, basado en el análisis de relatos textuales de accidentes laborales registrados entre 2011 y 2025 en Mutual de Seguridad. Dado que el foco del estudio es comprender la naturaleza, contexto y consecuencias de los hechos de violencia laboral, se hace uso de técnicas avanzadas de análisis de datos, incluyendo modelos de lenguaje de gran escala (LLM), procesamiento de lenguaje natural (NLP) y algoritmos de Machine Learning.

La base de datos utilizada en este estudio fue entregada bajo un protocolo de anonimización establecido por la Mutual de Seguridad, asegurando la confidencialidad de las personas involucradas. Asimismo, el proyecto contó con la autorización del Comité de Ética Científico para Proyectos de Investigación, formalizada mediante la Resolución N.º 377, emitida el 20 de diciembre de 2024.

La investigación se desarrolló en varias etapas, permitiendo una transición progresiva desde la detección de hechos de violencia hasta su caracterización profunda, pasando por un proceso de clasificación supervisada y validada expertamente.

4.1 Preparación y limpieza de datos

El estudio utiliza el universo completo de relatos de accidentes laborales registrados en la Mutual de Seguridad entre 2011 y 2025, sin aplicar muestreo. El total disponible asciende a $N = 3.342.253$ registros. Cada registro contiene:

- Variables con el relato narrativo.
- Variables estructuradas con datos del trabajador.

Los datos fueron recopilados y proporcionados por la Mutual de Seguridad en formato CSV.

Los sectores económicos predominantes en la totalidad de los datos incluyen Construcción [20,5%], Comercio [13,8%], Servicios [10,7%], Industria manufacturera [10,4%] y Transporte [8,1%]. Las ocupaciones más representadas en la totalidad de los datos son Trabajos varios [33,8%], Servicios [5,5%], Conductores [5,0%], Sin oficio [4,5%] y Área enseñanza [3,7%].

Inicialmente, se realizó un proceso exhaustivo de limpieza y preprocesamiento de los datos. Esto incluyó:

- Eliminación de registros incompletos o duplicados:
 - Registros sin relato textual: 11,8%.
 - Registros con metadatos ausentes: < 0,01%.

- Registros con id duplicado: < 0,01%.
- Normalización y limpieza de textos
 - Corrección de caracteres especiales: Estandarización a la codificación UTF-8 (por ejemplo “Ã³” → “ó”).
 - Conversión a minúsculas.
 - Eliminación de puntuación y símbolos no informativos.
 - Eliminación de stopwords: Eliminar palabras vacías.
 - Lematización: Normalizar los verbos y sustantivos para que sean utilizados en su forma base (por ejemplo “trabajando” → “trabajar”).
 - Segmentación: Dividir en frases y tokens (palabras), manteniendo la coherencia de las estructuras gramaticales para un posterior análisis semántico. (por ejemplo “lleva 2 años como auxiliar” → “dos_años auxiliar”)
- Estandarización de campos claves:
 - Edad: Se validaron los rangos ingresados y se comprueba que coincidan con la variable tramo edad. Se eliminaron valores atípicos o mal registrados (<0.1%).
 - Sexo: Se transforma en categorías binarias (0: femenino, 1: masculino).
 - Ocupación: Se normalizó la ocupación utilizando un diccionario de sinónimos (por ejemplo “Maestro”, “Educador” → “Enseñanza”).
 - Diagnóstico: Se dividió la variable diagnóstico para incorporar diferentes subvariables (por ejemplo “Esguince tobillo grado 1” → “Esguince”, “Tobillo”, “Grado 1”).
 - Sector Económico: Se transforma a formato numérico utilizando la técnica de codificación one-hot encoding (crear una columna binaria para cada categoría única de la variable).

Esta etapa permitió preparar tanto los relatos limpios como sus respectivas variables estructuradas para etapas posteriores de entrenamiento y análisis.

4.2 Justificación del tamaño muestral y confiabilidad

Dado el volumen de relatos disponibles en la base de datos, resultaba metodológicamente inviable y poco eficiente evaluar los modelos sobre la totalidad del universo en las etapas iniciales. Por esta razón, se definió trabajar con una muestra probabilística aleatoria, que garantizara representatividad estadística suficiente para la validación preliminar del modelo.

Para determinar el tamaño muestral se consideraron criterios derivados de la teoría de muestreo clásico (Cochran, 1977), ampliamente utilizada en poblaciones grandes para la estimación de proporciones. En este marco, el tamaño muestral se obtiene a partir de la siguiente expresión:

$$n = \frac{Z^2 \cdot p \cdot (1 - p)}{E^2}$$

donde:

- n : tamaño de la muestra,
- Z : valor crítico asociado al nivel de confianza,
- p : proporción esperada de la característica de interés en la población,
- E : margen de error admisible.

La aplicación de este enfoque asegura que la muestra capture de manera adecuada la variabilidad de los relatos, incluyendo tanto casos de violencia como aquellos sin indicios de ella. Con ello se garantiza la confiabilidad del proceso de validación, a la vez que se optimizan los recursos y tiempos de cómputo, proporcionando un marco estadístico sólido para proyectar los resultados hacia el conjunto completo de datos en fases posteriores del proyecto.

4.3 Fine-tuning de modelos LLM para clasificación binaria y por tipo de violencia

Uno de los pilares metodológicos del proyecto fue la adaptación local (fine-tuning) de modelos de lenguaje preentrenados (LLM), como GPT o BERT, para el dominio específico de la violencia laboral en Chile. Este proceso se desarrolló en dos fases:

4.3.1 Clasificación binaria de violencia laboral

Se entrenó un primer modelo LLM para identificar automáticamente si un relato describe o no un hecho de violencia. Esta tarea se definió como una clasificación binaria supervisada, utilizando una muestra etiquetada previamente por expertos. La muestra se llevó a cabo según el método explicado en 4.2 Justificación del tamaño muestral y confiabilidad; y la etiquetación según lo detallado en 4.4.1 Validación experta de etiquetado. Se evaluó el rendimiento del modelo con métricas como precisión, recall y F1-score.

4.3.2 Clasificación según tipo de violencia (tipología Cal/OSHA)

A partir de los relatos detectados como violentos, se aplicó un segundo modelo LLM para clasificar los casos según la tipología de violencia laboral:

- Tipo 1: violencia externa.
- Tipo 2: violencia asociada a servicios.
- Tipo 3: violencia interna entre personas con vínculo laboral.

Ambos modelos fueron adaptados localmente utilizando técnicas de fine-tuning con datos reales, asegurando sensibilidad lingüística y contextual al lenguaje y particularidades del entorno laboral chileno. Para ello, se utilizaron recursos computacionales con capacidades de cómputo avanzado (GPUs) y modelos LLM preentrenado para realizar los ajustes correspondientes.

4.4 Validación experta

Dado el carácter sensible y complejo de los relatos de violencia, la validación del trabajo realizado con modelos de lenguaje requirió un proceso riguroso en dos etapas complementarias: una primera fase de validación experta de etiquetado realizada por el equipo investigador, y una segunda fase de validación experta de revisión llevada a cabo con especialistas de la Mutual de Seguridad en riesgos psicosociales laborales.

4.4.1 Validación experta de etiquetado

La primera validación consistió en un proceso de revisión manual de relatos clasificados por los modelos, a cargo del equipo investigador. Este estuvo integrado por un profesional de las ciencias sociales, un profesional experto en salud ocupacional y un profesional especializado en procesamiento de lenguaje natural (NLP). La diversidad de perfiles permitió analizar los relatos desde una perspectiva interdisciplinaria, garantizando un etiquetado consistente y ajustado tanto a criterios técnicos como al contexto de la salud y seguridad en el trabajo.

Este procedimiento tuvo como objetivos principales:

- Corregir errores de clasificación cometidos por el modelo.
- Generar un conjunto confiable y validado de relatos etiquetados por tipo de violencia.
- Ajustar y mejorar iterativamente el rendimiento de los modelos LLM.

La validación se realizó a partir de una muestra aleatoria y representativa de relatos, lo que permitió estimar el desempeño del modelo sobre el total de casos. Se aplicaron métricas estándar de evaluación, siendo la principal el F1 Score, que permite evaluar el equilibrio entre precisión y exhaustividad. Con este procedimiento se alcanzaron resultados de 93,67% en la detección de hechos de violencia y 92,23% en la clasificación por tipo de violencia, confirmando la solidez del enfoque metodológico y la confiabilidad de la base de relatos generada.

4.4.2 Validación experta de revisión

La segunda etapa correspondió a una validación con expertos en riesgos psicosociales laborales de la Mutual de Seguridad, quienes participaron en una sesión de análisis de resultados. En este espacio se discutieron los hallazgos obtenidos a partir de la clasificación

automática y la validación inicial, lo que permitió contrastar los resultados con la experiencia práctica de la institución.

Esta instancia tuvo como propósitos:

- Revisar y retroalimentar los resultados de los modelos.
- Contextualizar los hallazgos a la realidad organizacional y normativa vigente.
- Generar lineamientos preliminares para el diseño de recomendaciones preventivas.

Los insumos de esta sesión quedaron integrados en la sección 6. Discusión de Resultados, y a partir de ellos se definieron los lineamientos que posteriormente se consolidaron en la sección 7. Recomendaciones para la prevención por tipo de violencia laboral.

4.5 Análisis de caracterización mediante técnicas de NLP

Una vez establecida la base validada por tipo de violencia, se aplicaron técnicas de Procesamiento de Lenguaje Natural (NLP) para extraer patrones y características comunes dentro de cada grupo. Este análisis busca responder preguntas clave como:

- ¿Qué factores contextuales están presentes en los relatos violentos?
- ¿Cuáles son los principales gatillantes lingüísticos, semánticos o situacionales que aparecen en cada tipo de violencia?
- ¿Existen términos o frases recurrentes que indiquen causas estructurales o condiciones organizacionales de riesgo?

Para ello, se realizaron diversos procedimientos:

- Análisis de frecuencia de términos y coocurrencias, que permitió identificar palabras clave y asociaciones frecuentes específicas de cada tipo de violencia (por ejemplo, “jefatura”, “estrés” o “trato directo”)
- Modelado de tópicos, utilizando métodos de LDA y BERTopic, a partir del cual se extrajeron temas latentes con alta consistencia interna. Algunos tópicos recurrentes incluyeron situaciones en ambientes cotidianos, abuso de poder y agresiones de desconocidos.
- Clustering semántico de relatos, que agrupó automáticamente fragmentos similares, revelando perfiles comunes como “jefaturas”, “transporte público” o “guardias de seguridad”.
- Extracción de relaciones entre entidades, lo cual permitió mapear estructuras narrativas típicas dentro de cada categoría de violencia.

4.6 Análisis cruzado con variables estructuradas

De manera complementaria, se integraron las variables estructuradas del dataset (como sector económico, ocupación, edad, género, macrosegmento de empresa, días de reposo, etc.) con los relatos clasificados, para identificar correlaciones significativas. Este cruce permitió entender:

- Qué grupos ocupacionales o sectores presentan mayores niveles de violencia y de qué tipo.
- Qué condiciones organizacionales o contextuales son más recurrentes en cada categoría.
- Si existen diferencias relevantes según sexo, edad o región.

4.7 Interpretabilidad del modelo: SHAP Values

Para comprender cómo cada variable influye en las predicciones del modelo, se utilizaron los valores SHAP (Shapley Additive Explanations), una técnica de interpretación basada en la teoría de juegos cooperativos. Estos valores permiten asignar a cada característica un peso que refleja su contribución al resultado del modelo para una observación específica, lo que permite interpretar decisiones individuales y tendencias generales.

En el contexto de este estudio, los SHAP values resultan particularmente útiles para identificar qué variables estructuradas tienen mayor influencia al clasificar un hecho de violencia. El uso de SHAP permite además respaldar con evidencia empírica interpretaciones que, de otro modo, podrían ser meramente especulativas. Esto amplía los alcances del modelo desde una herramienta predictiva a un instrumento explicativo, útil para apoyar decisiones preventivas o diseñar intervenciones focalizadas en función de las variables que más inciden en los distintos tipos de hechos de violencia.

4.8 Generación de recomendaciones preventivas

A partir de la caracterización cualitativa y cuantitativa, se elaboraron recomendaciones prácticas orientadas a la prevención, detección y mitigación de los distintos tipos de violencia laboral. Estas recomendaciones son contextualizadas según los hallazgos del proyecto y alineadas con las obligaciones normativas actuales, incluyendo las exigencias de la Ley Karin.

5 Resultados

5.1 Procesamiento de Datos

El análisis comenzó con 3.342.253 denuncias totales y 84 variables, incluyendo relatos textuales. Después de la depuración se eliminaron 10 denuncias, quedando en total 3.342.243 denuncias y 30 variables finales. Pasos:

- Remoción de duplicados: Se encontraron solo 10 registros duplicados, asociados a un id incorrecto.
- Normalización y limpieza de textos: Consistió en limpiar las variables *caso_que_hacia*, *caso_como_ocurrio* y *caso_anamnesis*, eliminando los stopwords, signos de puntuación, espacios y caracteres especiales.

Estandarización de campos clave: Unificación de categorías en variables asociadas al diagnóstico y partes del cuerpo lesionada: *caso_nom_diag_1_may_est*, *caso_nom_diag_2_may_est* y *caso_nom_diag_3_may_est*, a la actividad económica: *act_econ*. Validación de rangos en variables numéricas (edad o días de reposo) corrigiendo y eliminando valores fuera de rango.

Se identificaron las variables de interés para caracterizar y clasificar los relatos de violencia laboral:

Variable	Tipo	Descripción
Año	Fecha	Año del evento
Mes	Fecha	Mes del evento
Día	Fecha	Día del evento
Aceptado/Rechazado	Booleana	Indica si fue aceptado/rechazado por Mutua
Grave	Booleana	Indica si el caso fue grave
Días reposo	Numérica	Cantidad total de días de reposo indicados
Días consumidos	Numérica	Días de reposo efectivamente tomados
Días otorgados	Numérica	Días de reposo autorizados
Porcentaje incapacidad	Numérica	Porcentaje de incapacidad otorgado
Gasto	Numérica	Gasto asociado al caso
Número trabajadores	Numérica	Total de trabajadores durante el período
Promedio trabajadores	Numérica	Promedio de trabajadores durante el período
Día de la semana	Categoría	Día de la semana del evento
Tramo edad	Categoría	Rango de edad del afectado

Género paciente	Categoría	Género de la persona afectado
Comuna	Categoría	Comuna donde ocurrió el evento
Región	Categoría	Región donde ocurrió el evento
Clasificación tipo accidente	Categoría	Tipo de accidente laboral
Clasificación tipo siniestro	Categoría	Tipo de siniestro registrado
Fin de semana	Booleana	Si ocurrió sábado o domingo
Horario	Categoría	Rango horario del evento
Feriado	Booleana	Indica si fue un día feriado
Diagnóstico	Categoría	Diagnóstico principal
Parte del cuerpo	Categoría	Parte del cuerpo afectada
Subdiagnóstico	Categoría	Diagnóstico secundario
Tramo gasto	Categoría	Rango de gasto asociado
Actividad económica	Categoría	Rubro económico en el que se desempeña el afectado
Macrosegmento	Categoría	Categoría del sector económico
Profesión/Ocupación	Categoría	Rol o cargo del trabajador afectado
Descripción caso	Texto	Relato en texto libre del evento ocurrido

Tabla 1. Variables descriptoras de los relatos

Esta estructura facilita el análisis y entrenamiento de los modelos.

5.2 Clasificación de Denuncias

Una de las primeras tareas del proyecto consiste en adaptar modelos de lenguaje de gran escala (LLM) al dominio específico de los relatos de accidentes laborales registrados por Mutual de Seguridad, con el objetivo de identificar automáticamente si un relato contiene o no hechos asociados a violencia laboral. Esta etapa es clave, ya que permite filtrar los millones de registros existentes y focalizar el análisis posterior únicamente en aquellos casos potencialmente relevantes para la caracterización.

5.2.1 Fine-tuning de modelos LLM para clasificación binaria

Se realizó el fine-tuning de un modelo LLM local, entrenado localmente con datos de Mutual, para realizar una clasificación binaria sobre los relatos: Violencia vs. No Violencia. Para validar el desempeño del modelo, se trabajó con una muestra aleatoria de 1.000 relatos seleccionada a partir del universo total de 3.342.253 relatos únicos.

La muestra fue rotulada manualmente, estableciendo como verdad de referencia que:

- 959 casos correspondían a No Violencia

- 41 casos correspondían a Violencia

El modelo, tras su entrenamiento, fue evaluado sobre esta muestra, obteniendo los siguientes resultados:

- De los 959 casos de No Violencia, 958 fueron correctamente clasificados y 1 fue clasificado erróneamente como Violencia.
- De los 41 casos de Violencia, 37 fueron correctamente clasificados y 4 fueron erróneamente clasificados como No Violencia.

Se obtiene la siguiente Matriz de Confusión:

		Realidad	
		Si	No
Predicción	Si	37	1
	No	4	958

Tabla 2. Matriz de Confusión - Modelo Violencia vs. No Violencia

A partir de esta matriz de confusión, se calcularon los indicadores estándar para modelos de clasificación binaria:

Métrica	Valor
Sensibilidad	90,24%
Especificidad	99,90%
Valor Predictivo Positivo	97,37%
Valor Predictivo Negativo	99,58%
Razón de Verosimilitud Positiva	867,7
Razón de Verosimilitud Negativa	0,097
F1-Score	93,67%

Tabla 3. Indicadores - Modelo Violencia vs. No Violencia

- Sensibilidad: La capacidad para identificar correctamente los casos positivos.
- Especificidad: La capacidad para identificar correctamente los casos negativos.
- Valor Predictivo Positivo: La probabilidad de que un caso clasificado como positivo realmente lo sea.
- Valor Predictivo Negativo: La probabilidad de que un caso clasificado como negativo realmente lo sea.
- Razón de Verosimilitud Positiva: Cuánto más probable es que un resultado positivo provenga de un caso realmente positivo, en comparación con uno negativo.
- Razón de Verosimilitud Negativa: Cuánto más probable es que un resultado negativo provenga de un caso realmente negativo, en comparación uno positivo.
- F1-score: Mide el equilibrio entre el Valor Predictivo Positivo y la Sensibilidad.

Estos resultados pueden considerarse altamente satisfactorios, dado que el modelo evidenció una elevada capacidad para identificar correctamente tanto los casos positivos como los negativos. Este nivel de desempeño es coherente con los criterios discutidos por Rainio, Teuho y Klén (2024), quienes señalan que la precisión por sí sola puede ser engañosa en contextos complejos y que resulta necesario complementarla con métricas como el F1 Score, la sensibilidad y la especificidad, dado que aportan una visión más equilibrada sobre la calidad de un clasificador. Aunque no se encontraron estudios específicos en el ámbito de la violencia laboral que establezcan umbrales de referencia para estas métricas, la evidencia aportada por Rainio, Teuho y Klén (2024) en aplicaciones transversales, particularmente en el área médica donde se evalúan modelos con datos sensibles y desbalanceados, sugiere que valores de F1 superiores al 80%, acompañados de alta especificidad y sensibilidad, son interpretados como indicadores de rendimiento confiable y efectivo. Esto refuerza que los resultados obtenidos en este estudio, en un dominio igualmente complejo y sensible como el análisis de relatos de violencia laboral, pueden ser considerados robustos y metodológicamente sólidos.

5.2.2 Justificación de la muestra y confiabilidad (Casos Totales)

Dado que el total de relatos únicos disponibles supera los tres millones (3.342.243) y siguiendo el criterio expuesto por Cochran (1977), se optó por evaluar el modelo sobre una muestra aleatoria de 1.000 casos. Para llegar a ese número, se consideró un nivel de confianza de 95% ($Z=1,96$), una proporción máxima de variabilidad de $p=0,5$ y un margen de error deseado de 3,1% ($E=0,031$), valor inferior al umbral de 5% que suele considerarse aceptable en estudios exploratorios y de validación (Cochran, 1977).

Si bien la prevalencia de relatos que contienen indicios de violencia en la base completa es cercana al 4%, la distribución observada de la muestra (4,1%) resulta consistente con ese valor, lo cual refuerza la validez del procedimiento de muestreo aleatorio empleado.

Los altos niveles de desempeño alcanzados en la muestra permiten proyectar una aplicación efectiva del modelo sobre el universo completo, especialmente en tareas de filtrado inicial de relatos, donde el objetivo es reducir el universo de trabajo a aquellos casos más relevantes para su posterior análisis cualitativo.

Este desempeño se considera robusto en comparación con métodos tradicionales de clasificación textual supervisados, como aquellos basados en árboles aleatorios, que suelen enfrentar mayores dificultades al trabajar con lenguaje ambiguo o narrativas no estructuradas. En este contexto, el modelo LLM fine-tuneado demuestra una capacidad superior para distinguir con alta precisión los relatos que describen violencia laboral, contribuyendo de forma efectiva a la automatización del filtrado inicial de grandes volúmenes de texto y fortaleciendo las siguientes etapas del proyecto.

Una segunda tarea consiste en adaptar nuevamente los modelos de lenguaje de gran escala (LLM) para identificar los tres tipos de violencia, en los relatos previamente asociados a hechos de violencia laboral.

5.2.3 Fine-tuning de modelos LLM para clasificación por tipo de violencia

Se parte de la premisa de que una visión global o única del fenómeno de la violencia en el trabajo, tiende a invisibilizar diferencias clave entre sus distintos tipos, dificultando así la formulación de respuestas específicas y efectivas. En este contexto, se consideró fundamental contar con un modelo capaz de distinguir entre las distintas manifestaciones de violencia laboral, permitiendo una gestión pertinente del riesgo.

Se realizó otro fine-tuning de un modelo LLM local, entrenado un subconjunto de los datos de Mutual, aquellos asociados a hechos de violencia, para realizar una clasificación ternaria: Violencia externa, Violencia asociada a servicios y Violencia Interna. Para validar el desempeño del modelo, se trabajó con una muestra aleatoria de 600 relatos seleccionados a partir del universo total de casos asociados a hechos de violencia de 159.715 relatos únicos.

La muestra fue rotulada manualmente, estableciendo como verdad de referencia que:

- 307 casos correspondían a Violencia externa
- 183 casos correspondían a Violencia asociada a servicios
- 110 casos correspondían a Violencia interna

El modelo, tras su entrenamiento, fue evaluado sobre esta muestra, la siguiente Matriz de Confusión:

		Realidad		
		Externa	Servicio	Interna
Predicción	Externa	283	7	5
	Servicio	9	174	4
	Interna	15	2	101

Tabla 4. Matriz de Confusión - Modelo tipo de Violencia

A partir de esta matriz de confusión, se calcularon los indicadores estándar para modelos de clasificación en su versión macro, lo que implica que se evaluaron de forma independiente para cada una de las clases, y luego se promediaron sin ponderar por cantidad los casos en cada clase:

Métrica	Valor
Sensibilidad	92,75%
Especificidad	96,98%

Valor Predictivo Positivo	93,60%
Valor Predictivo Negativo	95,97%
Razón de Verosimilitud Positiva	34,14
Razón de Verosimilitud Negativa	0,0746
F1-Score	93,16%

Tabla 5. Indicadores - Modelo tipo de Violencia

- Sensibilidad: La capacidad para identificar correctamente los casos positivos.
- Especificidad: La capacidad para identificar correctamente los casos negativos.
- Valor Predictivo Positivo: La probabilidad de que un caso clasificado como positivo realmente lo sea.
- Valor Predictivo Negativo: La probabilidad de que un caso clasificado como negativo realmente lo sea.
- Razón de Verosimilitud Positiva: Cuánto más probable es que un resultado positivo provenga de un caso realmente positivo, en comparación con uno negativo.
- Razón de Verosimilitud Negativa: Cuánto más probable es que un resultado negativo provenga de un caso realmente negativo, en comparación uno positivo.
- F1-score: Mide el equilibrio entre el Valor Predictivo Positivo y la Sensibilidad.

Los resultados obtenidos pueden considerarse satisfactorios por diversos motivos. En primer lugar, las métricas de evaluación reflejan un desempeño robusto en la clasificación de las distintas categorías, lo cual es especialmente relevante considerando la naturaleza compleja y ambigua de los relatos analizados. Estos relatos no siempre presentan los hechos de manera directa ni con una estructura uniforme, lo que representa un desafío considerable para la interpretación automática.

Asimismo, el modelo logra este nivel de rendimiento sin recurrir a estructuras excesivamente complejas, lo que se enmarca en el principio de parsimonia: alcanzar un buen desempeño predictivo con una arquitectura relativamente sencilla favorece tanto la interpretabilidad como la generalización del modelo. En conjunto, estos factores permiten clasificar los resultados como adecuados en función del contexto y los objetivos de estudio.

5.2.4 Justificación de la muestra y confiabilidad (Tipo de Violencia)

Para el subconjunto de casos identificados como violencia, en total 159.715 relatos, siguiendo el criterio de Cochran (1977), se seleccionó una muestra aleatoria de 600 relatos. Este tamaño permite trabajar con un margen de error aproximado de 4% bajo un nivel de confianza del 95%, lo que resulta suficiente para validar el rendimiento del modelo en este grupo específico de interés.

5.3 Caracterización de Casos

5.3.1 Separación por Clústers

A partir de la base validada, se aplicaron distintas técnicas de Procesamiento de Lenguaje Natural (NLP) para caracterizar los relatos de violencia según cada tipo identificado.

- **Factores contextuales presentes en los relatos violentos**

Como parte del análisis de caracterización, se aplicó una técnica de clusterización para agrupar los relatos en seis grupos temáticos distintos, basados en similitudes semánticas y contextuales.

Los grupos temáticos no fueron definidos a priori ni forzados a ajustarse a una taxonomía preestablecida, sino que emergieron inductivamente desde los propios datos. Las denominaciones asignadas a cada clúster son etiquetas orientativas, pensadas como mnemónicos de patrones narrativos recurrentes que, por diseño, no implican una correspondencia con las categorías de Tipo de Violencia.

Clúster	Descripción Violencia	% Casos Totales
1	Jerárquica	23,7%
2	De labores internas	17,0%
3	En trayecto	20,1%
4	En transporte público	14,0%
5	A guardias y comercio	15,1%
6	En educación y SENAME	10,1%

Tabla 6. Clústers de relatos según similitudes semánticas y contextuales

Posteriormente, se detalla la distribución de estos clústeres según los tipos de violencia definidos en el estudio:

Clúster	Descripción Violencia	% Casos Violencia Externa	% Casos Violencia de Servicios	% Casos Violencia Interna
1	Jerárquica	4,3%	9,2%	86,4%
2	De labores internas	34,2%	23,9%	41,9%
3	En trayecto	90,2%	8,0%	1,7%
4	En transporte público	52,3%	45,0%	2,6%
5	A guardias y comercio	67,7%	18,5%	13,8%
6	En educación y SENAME	3,1%	61,7%	35,2%

Tabla 7. Distribución de Clústers por tipo de Violencia

Tal como se observa, clústers como "En trayecto" y "Jerárquica" no alcanzan un 100% en los tipos de violencia, que intuitivamente cabría esperar (Externa e Interna, respectivamente). Estas derivaciones se explican por dos factores:

1. Casos Borde/Ambigüedad Contextual:

En Trayecto, aparecen relatos donde el incidente ocurre dentro de instalaciones o en tránsito asociado a una prestación. En Jerárquica, algunos relatos enfatizan la asimetría de poder, aunque la interacción primaria esté situada en escenarios de servicio o frente a terceros. En este marco, en ambos clústers coexisten los tres tipos de violencia.

2. Ruido en la clasificación:

En el proceso de clasificación según tipo de violencia, existe un margen de error que puede estar generando la presencia de porcentajes minoritarios de otros tipos de violencia en cada clúster.

- **Términos recurrentes en los relatos**

El análisis de n-gramas (conjunto de palabras que se repiten), permitió identificar los principales gatillantes presentes en cada clúster:

Clúster	Términos recurrentes en los relatos
1	Jefatura directa, Acoso laboral jefatura, Estrés/sobrecarga laboral, Malos tratos, Licencia médica.
2	Acoso laboral, Realizando labores habituales, Lugar de trabajo, Sufre agresión.
3	Sufre asalto, Trayecto trabajo, Arma blanca, Arma de fuego, Trayecto domicilio-trabajo, Caminando vía pública.
4	Conductor Transantiago, Sufre agresión, Conduciendo bus, Arma blanca, Sufre agresión pasajero, Ruta habitual.
5	Guardia de seguridad, Sufre asalto, Mano, Arma de fuego, Recibe golpe.
6	Trato directo, Sala de clases, Malos tratos, SENAME, Licencia médica, Antecedentes mórbidos, Sufre golpe, Crisis de pánico.

Tabla 8. Términos recurrentes por Clúster

Se observa una mayor incidencia de violencia interna en ocupaciones caracterizadas por relaciones jerárquicas marcadas y escasa autonomía, como las que se dan en contextos de subordinación directa (Clúster 1), así como en espacios educativos y residenciales con alta carga emocional (Clúster 6: educación y SENAME). Además, el Clúster 2 (De labores internas) sugiere que buena parte de la violencia interna ocurre en el marco de las tareas diarias: son episodios de hostigamiento y agresión vinculados a la rutina laboral y a dinámicas organizacionales que no siempre se explicitan como "jerárquica". Por otro lado,

los trabajadores de transporte público (Clúster 4) y los guardias de seguridad o personal de comercio (Clúster 5) presentan altos niveles de exposición a violencia externa, particularmente asociada al contacto frecuente con usuarios o público general, muchas veces en condiciones de inseguridad física. Finalmente, el Clúster 3 (En trayecto) evidencia un riesgo transversal asociado a los trayectos laborales, donde la violencia ocurre fuera del entorno laboral directo, pero impacta significativamente. Esta caracterización permite inferir que, tanto la naturaleza de la ocupación, como las condiciones estructurales del entorno inciden en la probabilidad de exposición a diferentes formas de violencia.

5.3.2 Dinámicas de escalamiento

El análisis de los relatos mediante técnicas de NLP no solo permite identificar factores contextuales y agrupamientos temáticos, sino también evidenciar dinámicas de escalamiento y consecuencias psicológicas asociadas a los hechos de violencia. En diversos clústeres, particularmente aquellos vinculados a violencia jerárquica (Clúster 1) y en contextos institucionales como educación y SENAME (Clúster 6), emergen patrones narrativos que reflejan repeticiones sostenidas de malos tratos, ausencia de respuesta institucional y deterioro progresivo de la salud mental de las personas afectadas.

Términos como estrés, licencia médica, crisis de pánico, ansiedad y malos tratos constantes aparecen reiteradamente en los relatos, lo que sugiere que los eventos de violencia no son hechos aislados, sino que forman parte de procesos acumulativos con efectos sostenidos en el bienestar emocional. Asimismo, la falta de mecanismos efectivos de contención o intervención por parte de las jefaturas o instituciones parece contribuir a la cronificación del malestar y a una percepción de desprotección estructural. Esta evidencia refuerza la necesidad de abordar la violencia laboral no solo desde una perspectiva normativa o disciplinaria, sino también desde un enfoque preventivo y de salud mental.

A continuación, se sintetizan los términos recurrentes identificados por tipo de violencia, que funcionan como gatillantes lingüísticos y permiten comprender mejor sus contextos de aparición:

Tipo de Violencia	Términos Recurrentes	Caracterización
Interna	Acoso, sobrecarga, malos tratos, jefatura directa	Relacionada con jerarquías, condiciones adversas y dinámicas organizacionales rígidas
Externa	Arma blanca, asalto, golpes	Asociada a agresiones físicas o amenazas ocurridas fuera del entorno institucional
Asociada a Servicios	Trato directo, pasajero, atención	Derivada de interacciones conflictivas en contextos de atención o servicio

Tabla 9. Tabla Dinámicas de escalamiento

Estos términos sugieren que la violencia no solo responde a hechos puntuales, sino que se inserta en contextos laborales estructuralmente expuestos, donde la ausencia de mecanismos de contención y la naturalización del maltrato contribuyen al deterioro sostenido del bienestar psicológico.

5.4 Identificación de variables relevantes mediante red neuronal

Como complemento al análisis cualitativo de relatos presentado en la sección anterior, se implementó una red neuronal para identificar las variables estructuradas más relevantes al momento de determinar si un relato corresponde o no a un hecho de violencia. Esta estrategia buscó fortalecer la interpretación de los hallazgos mediante el cruce de enfoques: por un lado, el análisis semántico de los relatos permite captar las dinámicas y contextos narrados por los trabajadores; por otro, el análisis de variables estructuradas permite identificar patrones objetivos y cuantificables que pueden estar actuando como factores de riesgo.

En este análisis se consideraron exclusivamente las variables provenientes de la parte estructurada de los datos, las cuales fueron procesadas una vez que los hechos de violencia habían sido previamente identificados. Esto permitió centrar el análisis en los factores objetivos y cuantificables registrados en la base de datos, complementando la interpretación cualitativa derivada del procesamiento textual.

Para calcular la importancia de cada variable, se utilizó el método SHAP values (Shapley Additive Explanations), una técnica de interpretabilidad basada en teoría de juegos, que permite cuantificar la contribución individual de cada variable a la predicción del modelo. Esta metodología entrega explicaciones consistentes y globales (a nivel general del modelo), facilitando la comprensión de los factores que influyen en la clasificación y aumentando la transparencia del proceso.

Para estimar la relevancia de cada variable en las decisiones del modelo, se utilizó el método SHAP values (Shapley Additive Explanations), una técnica de interpretabilidad basada en la teoría de juegos. Esta metodología permite cuantificar la contribución individual de cada variable a las predicciones, entregando explicaciones globales sobre el comportamiento del modelo.

El gráfico que se presenta a continuación resume visualmente la importancia relativa de cada variable estructurada en el proceso de clasificación.

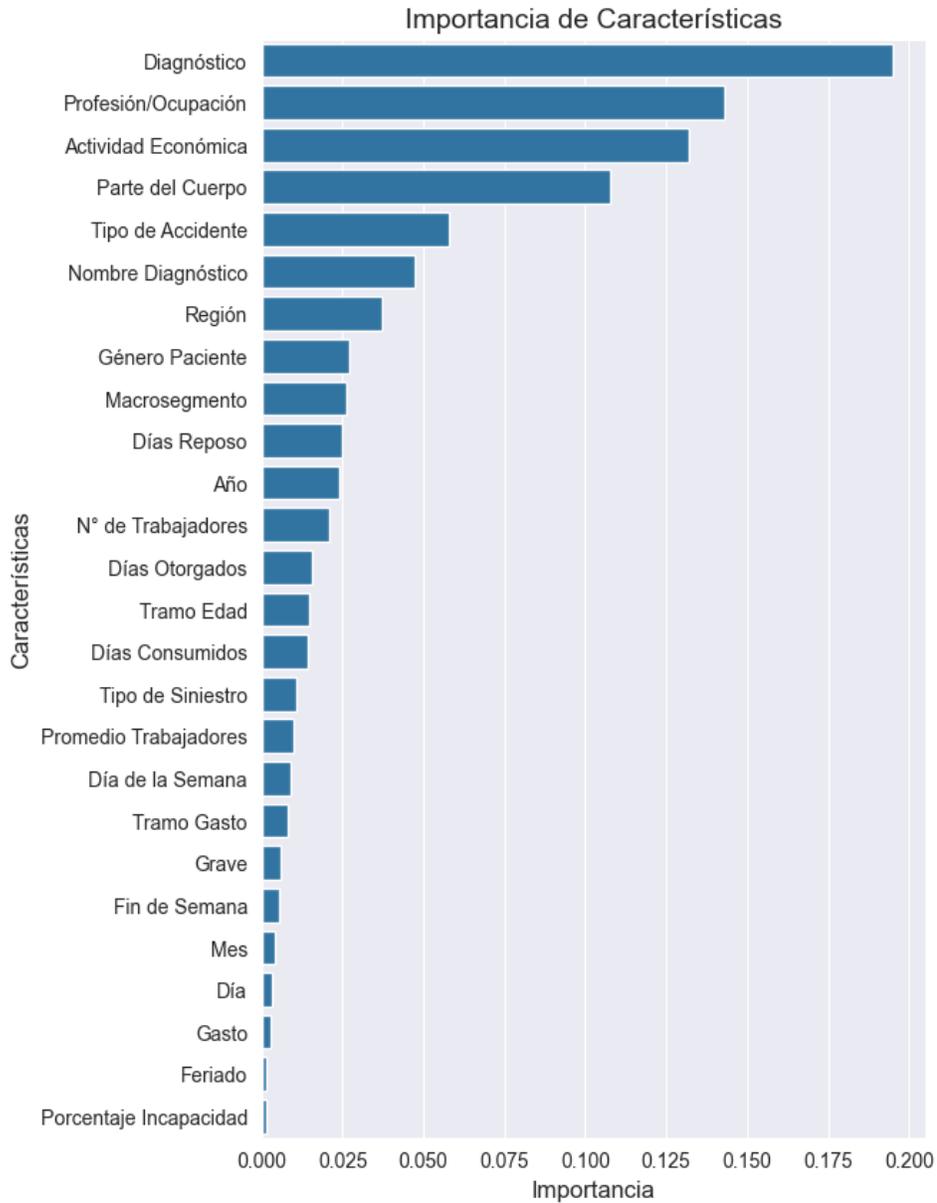


Figura 1. Importancia de Variables - Método SHAP values en Modelo Violencia vs. No Violencia

Las primeras cinco variables con mayor impacto son:

- Diagnóstico
- Profesión/Ocupación
- Actividad Económica
- Parte del Cuerpo
- Tipo de Accidente

Estas variables no deben interpretarse de manera aislada, sino como elementos interdependientes cuya combinación es aprovechada por el modelo para detectar patrones complejos. Su identificación permite ampliar la visión de una lógica puramente descriptiva a una comprensión más estructural de los factores que la acompañan.

Comprender qué variables estructuradas están más asociadas a relatos violentos contribuye a responder preguntas clave sobre las causas subyacentes de este fenómeno. Por ejemplo, la alta relevancia del diagnóstico o de ciertas ocupaciones puede estar revelando entornos laborales con mayor exposición a estrés o al trato directo con usuarios. El tipo de accidente y las partes del cuerpo afectadas podrían estar actuando como proxies de situaciones críticas en contextos de mayor vulnerabilidad.

En el siguiente gráfico de tipo Sankey, se muestra la relación entre Tipo de Violencia, Profesión/Ocupación, Diagnóstico y Tipo de Accidente, para los casos más comunes. El gráfico refleja la frecuencia en la que aparece cada grupo, a través de los enlaces, lo que permite identificar patrones de concentración.

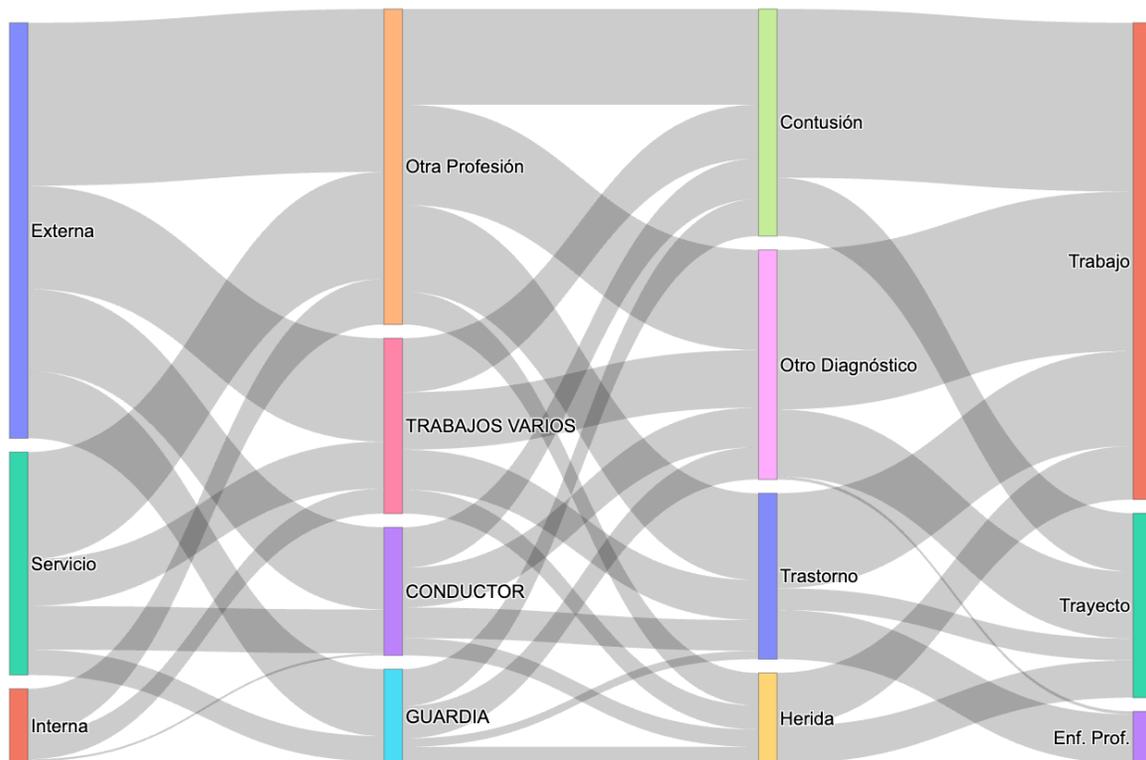


Figura 2. Relación entre Tipo de Violencia, Profesión/Ocupación, Diagnóstico y Tipo de Accidente mediante gráfico Sankey

5.5 Análisis descriptivo por Tipos de Violencia

Se realizó un análisis descriptivo de las ocurrencias registradas, segmentando la información según los tipos de violencia identificados por el modelo. Esta desagregación permitió examinar la distribución y concentración de los casos en función de las variables de mayor interés.

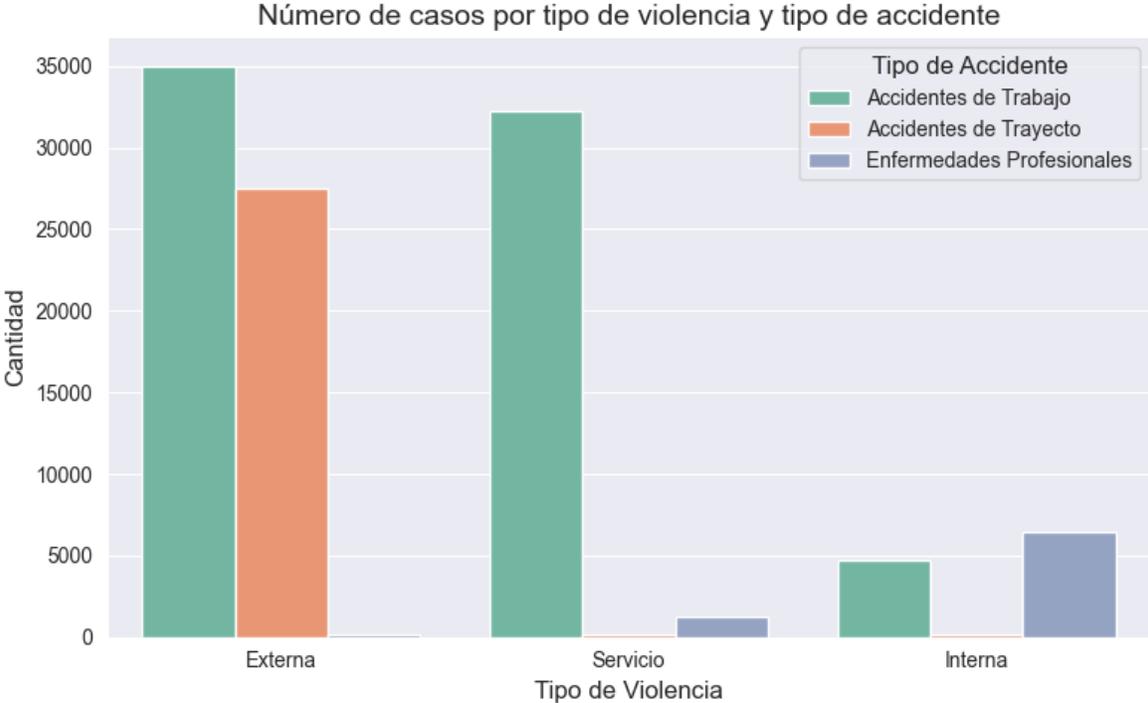


Figura 3. Número de casos por tipo de violencia y tipo de accidente.

Fuente: Denuncias de violencia ingresados y calificados como aceptados por Mutual de Seguridad, periodo 2011-2025.

Se observa una mayor frecuencia de ingresos asociados a violencia Tipo 1 (externa), con predominancia de accidentes de trabajo y trayecto. La violencia Tipo 3 (interna) presenta una alta proporción de enfermedades profesionales, lo que podría indicar una exposición prolongada a factores psicosociales.

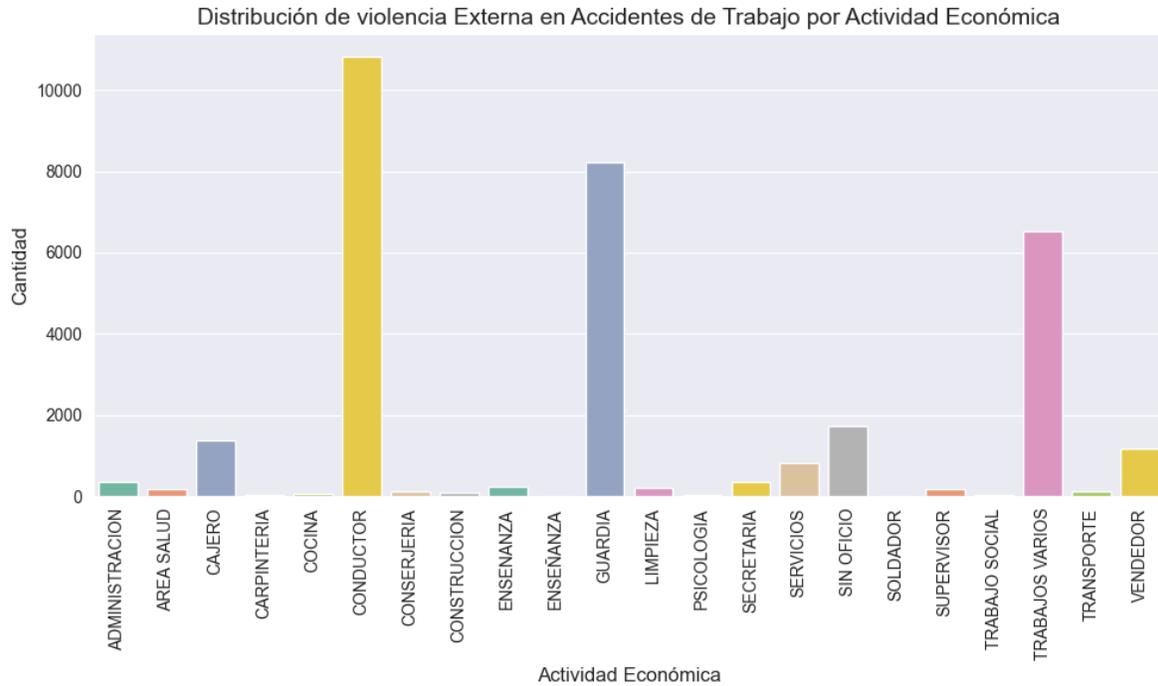


Figura 4. Número de casos en violencia Externa y Accidentes de Trabajo por Actividad Económica .

El gráfico muestra que la violencia externa en accidentes de trabajo se concentra principalmente en ocupaciones con alta exposición al público, como conductores, guardias de seguridad y trabajos varios. Estos resultados sugieren que los casos de violencia externa no ocurren de manera aleatoria, sino en actividades donde existe contacto frecuente con terceros.

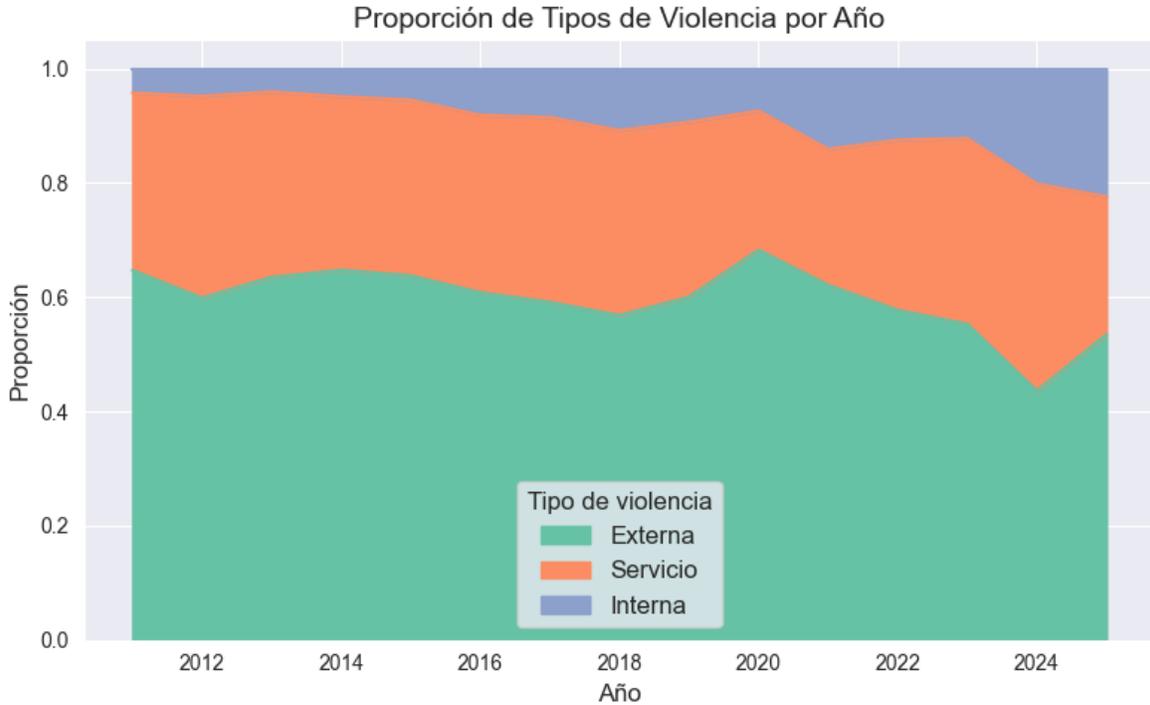


Figura 5. Proporción de Tipos de Violencia por Año.

Fuente: Denuncias de violencia laboral ingresados y calificados como aceptados por Mutuo de Seguridad entre 2011-2025.

Evolución proporcional de casos aceptados por violencia laboral según tipo, evidenciando una disminución relativa de la violencia externa y un aumento progresivo de la violencia interna en el periodo analizado.

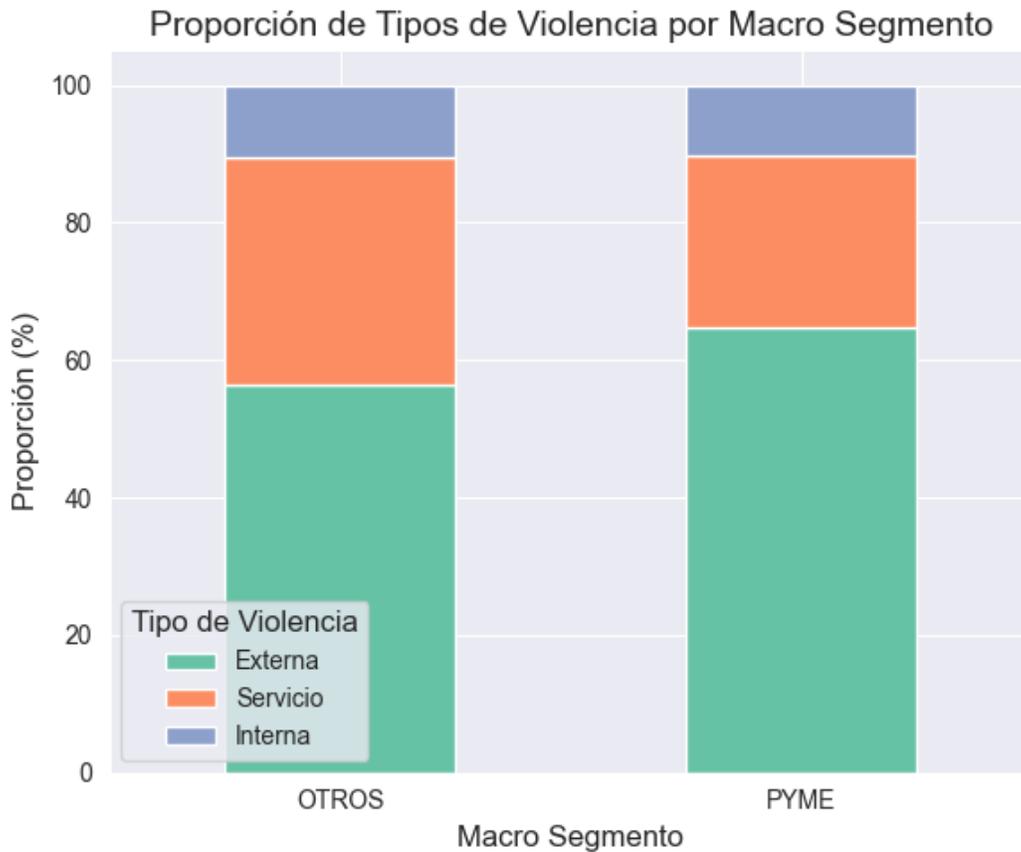


Figura 6. Cantidad Total de Tipos de Violencia por Macro Segmento PYME.

Fuente: Denuncias de violencia laboral aceptados por Mutual de Seguridad, 2011–2025.

La distribución de los casos evidencia una predominancia de violencia externa como principal modalidad reportada en ambos macrosegmentos. Les siguen los eventos asociados a funciones de servicio, mientras que la violencia interna mantiene una proporción significativamente menor. Estos datos sugieren una mayor exposición operacional y contextual al riesgo externo, especialmente en organizaciones fuera del segmento PYME, los que involucran GC, CORP, EMPR, TI y TCP.

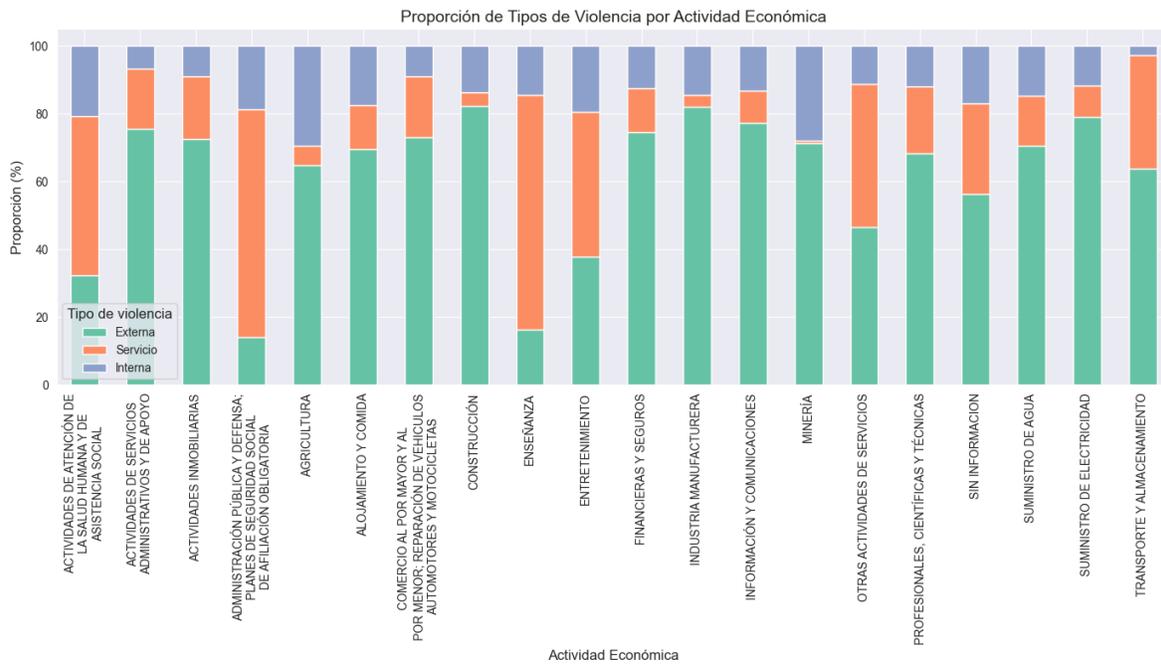


Figura 7. Distribución de Tipos de Violencia por Actividad Económica.

Fuente: Mutual de Seguridad, denuncias aceptadas 2011–2025.

La distribución de tipos de violencia varía según la actividad económica: la violencia externa predomina en sectores de contacto con público o exposición a riesgos (comercio, transporte, construcción, minería), la violencia de servicio se concentra en áreas de atención directa como salud y educación, mientras que la violencia interna es más visible en contextos institucionales y jerárquicos como la administración pública.

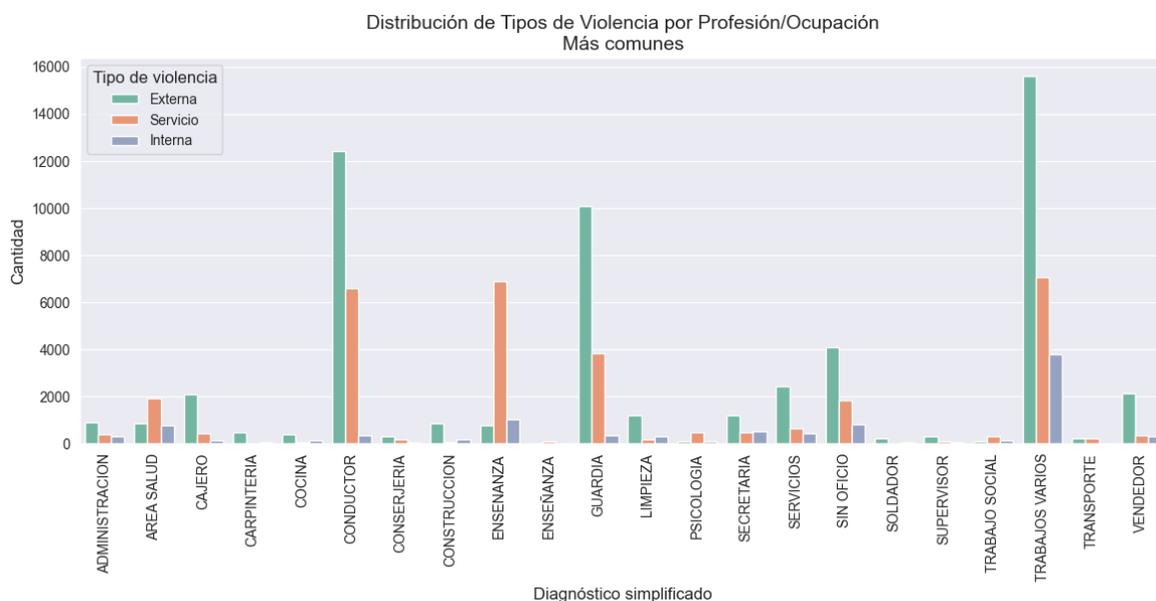


Figura 8. Distribución de Tipos de Violencia por Profesión/Ocupación más comunes.

Fuente: Mutual de Seguridad, denuncias aceptadas 2011–2025

Las ocupaciones con mayor interacción con usuarios externos, como conductores y guardias, concentran la mayoría de los casos de violencia externa, mientras que áreas como salud y educación presentan una alta proporción de violencia tipo servicio. Aunque ambas categorías pueden parecer cercanas, el modelo las diferencia según el origen de la agresión: la violencia externa corresponde a situaciones de robo, asalto o agresiones físicas o verbales provenientes de público general, en contextos donde el trabajador no presta un servicio específico (por ejemplo, transporte, seguridad o comercio).

Mientras que la violencia de servicios se asocia a contextos con relación directa y prolongada entre trabajador y usuario, donde el conflicto surge dentro del marco del servicio prestado (por ejemplo, salud o educación). De este modo, el modelo clasifica como externa las agresiones ligadas a inseguridad física en el entorno laboral y como de servicios aquellas propias de la interacción en funciones de cuidado, formación o atención especializada.

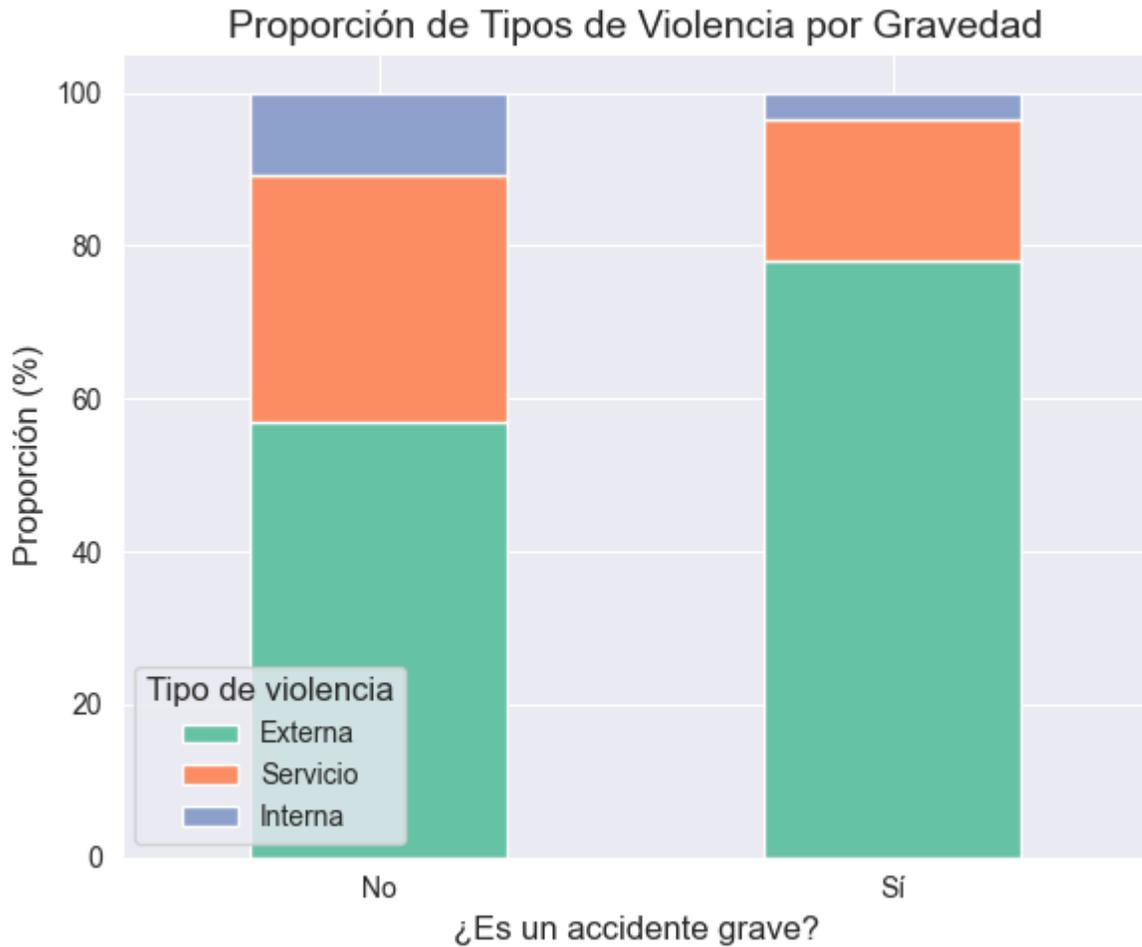


Figura 9. Cantidad total de Tipos de Violencia por Gravedad.

Fuente: Mutua de Seguridad, casos aceptados 2011–2025

Predominan los casos no graves, concentrados en violencia externa. No obstante, este tipo también presenta mayor proporción de accidentes graves, evidenciando mayor riesgo de daño físico. Violencias internas y de servicio se asocian principalmente a afectación no traumática.

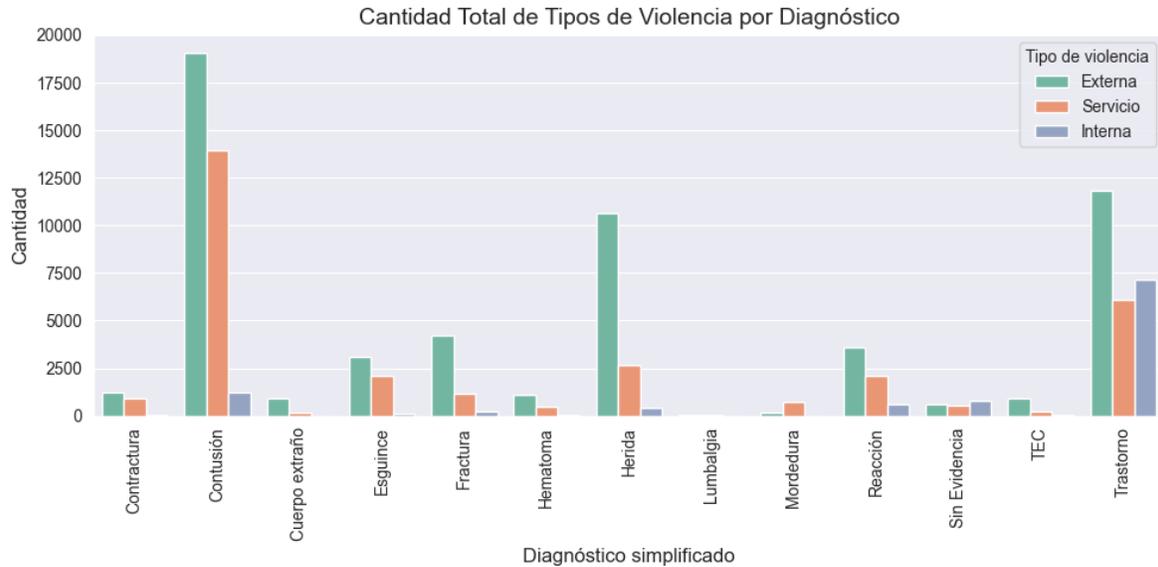


Figura 10. Cantidad total de Tipos de Violencia por Diagnóstico.

Fuente: Mutua de Seguridad, casos aceptados 2011–2025.

El diagnóstico permite inferir la naturaleza de la violencia: la violencia externa tiende a generar lesiones físicas visibles, mientras que la violencia interna y de servicio se manifiestan más comúnmente como trastornos psicosociales o de salud mental.

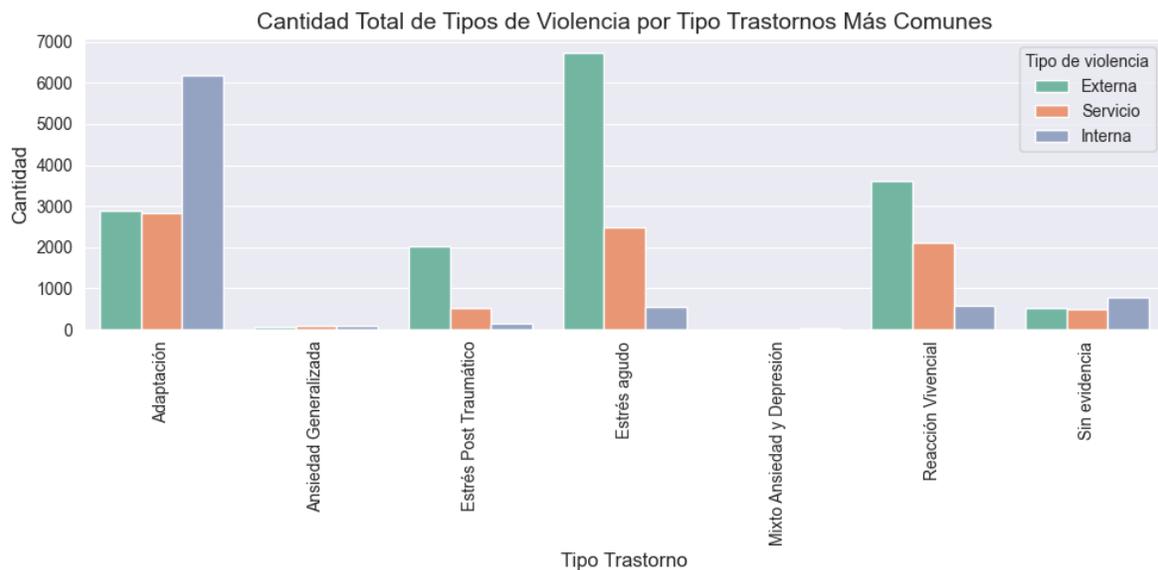


Figura 11. Cantidad total de Tipos de Violencia por Tipo Trastornos más comunes.

Fuente: Mutua de Seguridad, casos aceptados 2011–2025.

La violencia interna se relaciona más fuertemente con trastornos de adaptación, lo que sugiere un daño sostenido en el tiempo vinculado al entorno organizacional. En cambio, la violencia externa se vincula más a reacciones agudas de estrés, con alta carga emocional y aparición rápida de síntomas. La violencia de servicio muestra un patrón mixto, pero con claro predominio de afectación emocional.

6 Discusión de Resultados

A partir de los resultados y de las discusiones generadas con los expertos en riesgos psicosociales laborales de la Mutual de Seguridad según la metodología descrita en **4.4.2 Valdiación experta de revisión**, se acordó enfatizar en que las medidas preventivas no deben limitarse a entregar herramientas individuales a los trabajadores, sino que deben estar orientadas a una gestión organizacional más amplia, incorporando acciones de protección física (como guardias, mamparas o protocolos frente a contingencias), así como articulación con el entorno mediante mapeo de riesgos y coordinación con planes de seguridad pública. Asimismo, se destacó que la Ley Karin (Ley N° 21.643) y el Protocolo de Riesgos Psicosociales constituyen el piso mínimo que toda organización debe cumplir, pero que deben complementarse con protocolos internos específicos de actuación frente a agresiones y situaciones de emergencia. Finalmente, los expertos señalaron la necesidad de analizar factores estructurales de la organización —como la sobrecarga laboral, el diseño de puestos de trabajo, el liderazgo y la justicia organizacional— dado que muchas veces son estas condiciones las que generan o facilitan la aparición de violencia de usuarios o violencia interna.

El análisis integral de la información evidencia que la violencia laboral presenta patrones claramente diferenciados según su tipología, gravedad, categoría de ingreso, actividad económica y ocupación, lo que tiene implicancias directas en la gestión preventiva.

En primer lugar, la violencia externa destaca como el tipo de mayor incidencia, especialmente en accidentes de trabajo y trayecto, superando los 35.000 y 27.500 casos respectivamente. Esta tendencia confirma que la exposición a terceros ajenos a la organización constituye uno de los principales focos de riesgo, particularmente en actividades que implican desplazamiento, contacto directo con usuarios o trabajo en espacios abiertos. La presencia de un porcentaje considerable de accidentes graves dentro de esta tipología refuerza la necesidad de fortalecer las medidas de control en terreno, implementar protocolos de actuación ante incidentes y evaluar la capacitación del personal en manejo de situaciones agresivas.

En contraste, la violencia interna se concentra principalmente en enfermedades profesionales (más de 6.000 casos), lo que sugiere que los conflictos interpersonales y los riesgos psicosociales prolongados tienen un peso significativo en la generación de daño. Al no presentar predominio de lesiones traumáticas graves, la intervención debe orientarse a estrategias de gestión del clima laboral, resolución temprana de conflictos, vigilancia de la salud mental y fortalecimiento de los canales de denuncia interna.

Por su parte, la violencia de servicio presenta su mayor carga en accidentes laborales (más de 32.000 casos) y está asociada de manera directa a la naturaleza de las funciones ejecutadas, especialmente en actividades de atención de público o prestación de servicios en entornos de alta presión operativa. La ausencia de casos en trayecto y la baja incidencia en enfermedades profesionales evidencian que se trata de incidentes de impacto

inmediato, que requieren protocolos específicos de atención segura, formación en comunicación efectiva y mecanismos de contención post-incidente.

El análisis por actividad económica permite identificar focos prioritarios de intervención:

- La violencia externa presenta mayor incidencia en transporte y almacenamiento, atención de la salud humana y asistencia social y comercio al por mayor y menor, sectores que combinan alta movilidad y contacto con terceros.
- La violencia de servicio se concentra en atención de la salud humana y asistencia social, administración pública y defensa y transporte y almacenamiento, donde la presión laboral y el contacto frecuente con usuarios incrementan el riesgo de incidentes.
- La violencia interna destaca en administración pública y defensa, enseñanza y transporte y almacenamiento, lo que evidencia la necesidad de reforzar la gestión psicosocial en estos sectores.

Por ocupación, se identifican como grupos más expuestos Trabajos Varios, Conductores y Guardias, seguidos por Enseñanza y Sin Oficio. Estos perfiles combinan condiciones de alta exposición, interacción constante con terceros y entornos operativos complejos, lo que incrementa la probabilidad de incidentes.

En conjunto, los resultados respaldan la necesidad de estrategias preventivas diferenciadas y específicas según tipo de violencia, sector productivo y ocupación, priorizando:

1. Violencia externa: control de riesgos en terreno, protocolos de seguridad y capacitación en manejo de incidentes.
2. Violencia interna: intervenciones psicosociales, gestión de convivencia y mecanismos de resolución de conflictos.
3. Violencia de servicio: adaptación de procesos, mejora en la atención segura y formación en comunicación efectiva.

Este enfoque permitirá optimizar la asignación de recursos, aumentar la eficacia de las medidas de prevención y reducir la ocurrencia e impacto de eventos violentos en el ámbito laboral.

7 Recomendaciones para la prevención por tipo de violencia laboral

En base a los resultados obtenidos y considerando el marco normativo vigente, en particular la Ley 21.643 (Ley Karin) y el Protocolo de Vigilancia de Riesgos Psicosociales (CEAL/SUSESO), se proponen medidas técnicas, administrativas e ingenieriles diferenciadas para cada tipo de violencia, complementadas con recomendaciones específicas en función de los diagnósticos y trastornos identificados en la población afectada.

7.1 Violencia externa

- Evaluación periódica de riesgos psicosociales y de seguridad: identificar funciones y puestos con mayor exposición (atención de público, transporte, vigilancia, trabajo nocturno, visitas en terreno, etc.), siguiendo los lineamientos del Protocolo CEAL para riesgos psicosociales y considerando la Ley Karin, que refuerza la obligación de prevenir y sancionar conductas de acoso laboral y violencia en el trabajo.
- Establecer procedimientos claros de actuación ante agresiones, donde como mínimo se indique como reportar, derivar y registrar los incidentes, asegurando su coherencia con las exigencias de la Ley Karin en cuanto a plazos, investigación y protección de la víctima.
- Implementar una bitácora digital o física para documentar hechos violentos, identificar patrones y establecer medidas correctivas, garantizando resguardo de datos y trazabilidad.
- Generación de programa de capacitaciones al personal, donde como mínimo se considere entrenar al personal en comunicación efectiva, manejo de clientes y resolución de conflictos, incorporando módulos sobre derechos, obligaciones y canales de denuncia establecidos por la Ley Karin.
- Refuerzo del liderazgo y supervisión preventiva, por medio de la formación a líderes en el reconocimiento temprano de señales de cualquier tipo violencia y en la aplicación de protocolos, reforzando su rol como garantes de un ambiente seguro.
- Disponer de un plan de contingencia por medio de procedimientos alternativos de operación en situaciones de alto riesgo, incluyendo suspensión temporal de atención presencial cuando corresponda.
- Para la actividad económica de transporte se recomienda la implementación de un entorno de trabajo seguro para conductores mediante iluminación estratégica en paraderos y terminales, eliminación de puntos ciegos y optimización de la visibilidad en cabinas. Incorporar televigilancia en tiempo real desde centrales de monitoreo, cámaras a bordo, sistemas GPS y botones de pánico conectados directamente a seguridad o Carabineros. Complementar con sistemas de vocería que faciliten la comunicación inmediata ante situaciones de riesgo, protocolos claros de actuación

frente a incidentes y coordinación directa con fuerzas de seguridad, asegurando tanto la prevención como una respuesta rápida y efectiva.

- Generar un diseño seguro del entorno laboral a través de accesos controlados, iluminación estratégica, eliminación de puntos ciegos, optimización de la visibilidad. Fomentando el uso de herramientas tecnológicas como cámaras de vigilancia, alarmas, detectores de movimiento y botones de pánico conectados a una central de monitoreo.
- En puntos de atención a público, contar con alguna barrera física protectora para prevenir agresiones directas.
- Se recomienda implementar un seguimiento integral de los trabajadores afectados, incluyendo evaluación médica y psicológica inmediata, contención post-incidente y programas de rehabilitación temprana. Esta violencia, caracterizada por agresiones o amenazas de terceros ajenos a la organización, se asocia principalmente con lesiones físicas de diversa gravedad, estrés agudo y síntomas de estrés postraumático. La aplicación de estas medidas busca prevenir secuelas físicas y psicosociales, favoreciendo la recuperación completa de los trabajadores.

7.2 Violencia de servicio

- Realizar diagnósticos periódicos de riesgos psicosociales aplicando herramientas como CEAL/SM-SUSESO, complementados con el levantamiento de incidentes y situaciones de riesgo recurrentes, para diseñar planes de acción específicos.
- Implementar políticas institucionales de tolerancia cero a la violencia de usuarios/clientes, difundiendo el Protocolo de Actuación establecido en la Ley 21.643 y asegurando que todo el personal conozca los pasos de reacción y denuncia.
- Capacitar al personal en manejo de situaciones complejas, contención emocional, comunicación asertiva y primeros auxilios psicológicos, con énfasis en sectores de mayor exposición.
- Diseñar y adaptar los espacios de atención para garantizar seguridad, visibilidad y vías de escape, incorporando control de accesos, televigilancia, botones de pánico y, cuando sea necesario, barreras físicas de protección.
- Establecer protocolos claros de actuación y denuncia que aseguren confidencialidad, ausencia de represalias y acompañamiento psicológico y legal, incluyendo derivación a programas de apoyo del organismo administrador.
- Desarrollar campañas de sensibilización dirigidas a usuarios y coordinar acciones preventivas con carabineros, seguridad privada u otros actores locales para reforzar la protección del personal.
- Para el sector educativo, se recomienda la generación de un diagnóstico periódico de riesgos psicosociales con herramientas como CEAL/SM-SUSESO, la recopilación y análisis de incidentes con terceros (apoderados, estudiantes u otros actores externo), la adopción y difusión de políticas de tolerancia cero conforme a la Ley 21.643, la capacitación continua del personal en manejo de conflictos, contención

emocional y primeros auxilios psicológicos, el diseño seguro de espacios mediante control de accesos, televigilancia, botones de pánico y rutas de evacuación, así como la ejecución de protocolos de respuesta temprana coordinados con redes externas y campañas de sensibilización dirigidas a estudiantes y familias para promover el respeto y la resolución pacífica de conflictos.

- Implementar monitoreo psicológico y programas de apoyo individual y grupal para los trabajadores expuestos a alta carga emocional, incluyendo evaluaciones periódicas del estado emocional y estrategias de prevención de trastornos por estrés crónico. La violencia de servicio, presente durante la interacción con usuarios, clientes o beneficiarios, se relaciona principalmente con estrés agudo, crisis de pánico, ansiedad y lesiones físicas leves, y estas medidas buscan proteger la salud y seguridad del personal durante la prestación de servicios.

7.3 Violencia interna

- Fomentar una cultura organizacional basada en el respeto, la inclusión y la colaboración, como campañas de sensibilización sobre convivencia, comunicación asertiva y resolución de conflictos.
- Implementar y difundir políticas internas de prevención de acoso laboral, maltrato y violencia, incluyendo el Protocolo de Actuación de la Ley 21.643, junto con canales de denuncia confidenciales.
- Capacitar periódicamente en gestión de conflictos, liderazgo, normativa legal vigente y primeros auxilios psicológicos, fortaleciendo las capacidades preventivas y de contención de líderes y supervisores.
- Aplicar instrumentos como el Cuestionario CEAL/SM-SUSESO para identificar riesgos psicosociales y desarrollar planes de acción correctiva con seguimiento continuo.
- Actuar de forma inmediata y objetiva ante denuncias, asegurando apoyo psicológico, acompañamiento legal si corresponde, y medidas correctivas o disciplinarias según la gravedad del caso.
- Fortalecer el compromiso institucional mediante la participación de la alta dirección y el funcionamiento efectivo de comités de convivencia laboral con representación de trabajadores.
- Optimizar el diseño y distribución de los espacios de trabajo para favorecer la visibilidad, la comunicación y la supervisión, evitando zonas aisladas y asegurando iluminación adecuada; incluir áreas comunes que promuevan interacciones positivas.
- Incorporar sistemas de control de acceso y televigilancia en áreas críticas, junto con señalética preventiva que refuerce el respeto y la convivencia, cuidando siempre la confidencialidad y la normativa de datos.
- Habilitar espacios seguros para entrevistas, mediaciones o contención emocional, ofreciendo privacidad y resguardo físico en situaciones de conflicto.

- Establecer programas de intervención psicológica y psicosocial, evaluación periódica de la salud mental, seguimiento clínico y derivación a especialistas en salud mental cuando corresponda. La violencia interna, derivada de conflictos, acoso o agresiones entre trabajadores, se vincula con trastornos de adaptación, ansiedad, depresión, alteraciones del sueño y síntomas psicosomáticos. Estas acciones permitirán prevenir la cronificación de los trastornos, promover la resiliencia del trabajador y mejorar el clima laboral.

8 Conclusiones

Los resultados del análisis cuantitativo y del entrenamiento de modelos de *Machine Learning* y LLM permiten concluir que el enfoque adoptado ofrece un desempeño altamente robusto para la identificación y clasificación de hechos de violencia laboral. En la tarea de clasificación binaria, el modelo alcanzó un F1-Score de 93,67%, con una sensibilidad de 90,2% y una especificidad de 99,9%, mientras que la clasificación según la tipología de Cal/OSHA (violencia externa, de servicio e interna) obtuvo un F1-Score de 93,16%.

Estos resultados confirman que los modelos fine-tuneados son capaces de procesar con alta precisión un volumen masivo de relatos (más de 3,3 millones), lo que constituye un avance significativo en la capacidad de detección automatizada de violencia laboral.

Adicionalmente, el uso de técnicas de *NLP* permitió identificar clústeres temáticos y factores contextuales relevantes, como trayectos laborales, transporte público, jerarquías internas o entornos educativos, y mediante métodos de interpretabilidad como los SHAP values se destacó la importancia de variables como diagnóstico, ocupación, actividad económica y tipo de accidente.

Todo ello demuestra que la combinación de modelos de lenguaje, validación experta y análisis interpretativo no solo es eficaz para la clasificación, sino que también entrega información valiosa para comprender los determinantes y contextos de la violencia laboral.

El análisis realizado permitió establecer que la violencia laboral presenta patrones diferenciados según su origen, gravedad, actividad económica y ocupación, con la violencia externa como la de mayor frecuencia e impacto físico, la violencia interna asociada principalmente a daños psicológicos por exposición prolongada a conflictos interpersonales, y la violencia de servicio vinculada a incidentes en la interacción directa con usuarios o beneficiarios.

Los diagnósticos más frecuentes derivados de estos eventos incluyen lesiones traumáticas de diversa gravedad (predominantes en violencia externa), así como trastornos de adaptación, ansiedad, depresión, estrés agudo y en algunos casos, síntomas compatibles con estrés postraumático. Este perfil clínico evidencia que, además del riesgo físico, la violencia laboral constituye una amenaza significativa para la salud mental de los trabajadores, con potencial de generar secuelas crónicas si no se actúa oportunamente.

En este contexto, se confirma la necesidad de implementar estrategias preventivas específicas por tipo de violencia, integrando medidas técnicas, administrativas e ingenieriles alineadas con la Ley 21.643 (*Ley Karin*) y el Protocolo de Vigilancia de Riesgos Psicosociales CEAL/SUSESO. Dichas medidas deben contemplar la detección temprana de lesiones físicas y trastornos psicológicos, protocolos de atención y derivación a

especialistas, así como programas de seguimiento que aseguren la rehabilitación y reintegro seguro al trabajo.

En definitiva, la gestión efectiva de la violencia laboral requiere una planificación preventiva sostenida, basada en evidencia, que priorice la protección integral de la salud física y mental de los trabajadores, optimice el uso de recursos y contribuya a la consolidación de entornos laborales más seguros, saludables y resilientes.

9 Referencias

- Ansoleaga, E., Ahumada, M., & Henríquez, S. (2022). *¿Son efectivas las intervenciones para prevenir y/o intervenir sobre la violencia laboral? Un análisis de la evidencia científica*. *Praxis Psy*. 22. 6-22. 10.32995/praxispsy.v22i36.170.
- Ávila Urdaneta, J. (2020). El acoso laboral o *mobbing* como elemento causante de accidentes laborales. *Revista de Seguridad y Salud en el Trabajo*, 18(2), 55–67. http://scielo.senescyt.gob.ec/scielo.php?pid=S2631-27862020000300123&script=sci_arttext
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.
- Conway, M., Hu, M., & Chapman, W. W. (2019). Recent advances in using natural language processing to address public health research questions using social media and consumer-generated data. *Yearbook of Medical Informatics*, 28(1), 208–217. <https://www.thieme-connect.de/products/ejournals/abstract/10.1055/s-0039-1677918>
- Durango, M. C., Torres-Silva, E. A., & Orozco-Duque, A. (2023). Named entity recognition in electronic health records: A methodological review. *Healthcare Informatics Research*, 29(4), 286–300. <https://ehir.org/journal/view.php?doi=10.4258/hir.2023.29.4.286>
- European Agency for Safety and Health at Work [EU-OSHA]. (2021). *Third European Survey of Enterprises on New and Emerging Risks (ESENER 3)*. <https://osha.europa.eu/en/publications/esener-2019-overview-report-how-european-workplaces-manage-safety-and-health>
- Fernández Caraballo, E., & Gómez Franco, Y. (2018). *Metodología para el análisis de la violencia en el departamento de Bolívar mediante técnicas de machine learning*. Universidad Tecnológica de Bolívar. <https://hdl.handle.net/20.500.12585/1118>
- Maheronnaghsh, S., & Zolfagharnasab, H. (2023). Machine learning in occupational safety and health: A systematic review. *International Journal of Occupational and Environmental Safety*, 7(1), 14-32. https://ijooes.fe.up.pt/index.php/ijooes/article/view/2184-0954_007-001_001586
- Mento, C., Silvestri, M., Bruno, A., Muscatello, M. R. A., Cedro, C., Pandolfo, G., & Zoccali, R. (2020). Workplace violence against healthcare professionals: A systematic review: A systematic review. *Aggress. Violent Behav.* 2020;51:10138. <https://www.sciencedirect.com/science/article/abs/pii/S1359178919301181>
- Organización Internacional del Trabajo [OIT]. (2003). *Repertorio de recomendaciones prácticas sobre la violencia en el lugar de trabajo en el sector de los servicios y medidas para combatirla*. Reunión de expertos, Ginebra (8–15 de octubre de 2003). International Labour Office.

https://www.ilo.org/wcmsp5/groups/public/@ed_protect/@protrav/@safework/documents/normativeinstrument/wcms_112578.pdf

- Organización Internacional del Trabajo [OIT]. (2022). *La violencia laboral en Chile en el marco del Convenio 190 de la OIT*. International Labour Office.
- Occupational Safety and Health Administration (OSHA). *Guidelines for Preventing Workplace Violence for Healthcare and Social Service Workers*. U.S. Department of Labor. <https://www.osha.gov/workplace-violence>
- Palma, A., Ansoleaga, E., & Ahumada, M. (2018). Violencia laboral en trabajadores del sector salud: Revisión sistemática. *Revista Médica de Chile*, 146(7), 836–844. https://www.scielo.cl/scielo.php?pid=S0034-98872018000200213&script=sci_arttext
- Parra Osorio, L., & Acosta Fernández, M. (2010). La investigación cuantitativa del acoso psicológico laboral en los sectores de la educación superior y de la salud: Una revisión sistemática. *Revista Latinoamericana de Psicología*, 42(1), 159–172. <https://www.redalyc.org/articulo.oa?id=265419646010>
- Rainio, O., Teuho, J. & Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci Rep* 14, 6086 (2024). <https://www.nature.com/articles/s41598-024-56706-x>
- Superintendencia de Seguridad Social (SUSESO). (2022). Informe anual de seguridad y salud en el trabajo 2022. Santiago, Chile: SUSESO. https://www.suseso.cl/607/articles-707000_archivo_01.pdf
- Tang, X., Boamah, S. A., & Kaulius, M. (2023). Workplace Predictors of Violence against Nurses Using Machine Learning Techniques: A Cross-Sectional Study Utilizing the National Standard of Psychological Workplace Health and Safety. *Healthcare*, 11(7), 1008. <https://www.mdpi.com/2227-9032/11/7/1008>