

**PROYECCIÓN DE ACCIDENTALIDAD LABORAL Y SUS CONSECUENCIAS,
ASOCIADAS A LA DINÁMICA DEL MERCADO LABORAL EN COLOMBIA**

**LUIS GUSTAVO MARTÍNEZ OSORIO
TUTOR: ESTEBAN LÓPEZ**

**UNIVERSIDAD DE LA SABANA
FACULTAD DE INGENIERÍA
MAESTRÍA EN ANALÍTICA APLICADA
TRABAJO DE GRADO
2023**



**Universidad
de La Sabana**

PÁGINA DE ACEPTACIÓN

Tutor del trabajo

Jurado 1

Jurado 2

Jurado 3

Tabla de contenido

1. INTRODUCCIÓN.....	1
2. PREGUNTA DE INVESTIGACIÓN.....	4
3. MARCO CONCEPTUAL.....	5
3.1. MARCO TEÓRICO.....	5
3.2. ESTADO DEL ARTE.....	11
4. OBJETIVOS.....	16
4.1. OBJETIVO GENERAL.....	16
4.2. OBJETIVOS ESPECÍFICOS.....	16
5. METODOLOGÍA.....	17
5.1. MUESTREO.....	17
5.2. EXPLORACIÓN.....	18
5.3. MODIFICACIÓN O TRANSFORMACIÓN.....	19
5.4. MODELAMIENTO.....	20
5.5. EVALUACIÓN.....	20
5.6. APLICACIÓN.....	21
6. DESARROLLO Y RESULTADOS.....	24
6.1. SELECCIÓN DE SECTORES ECONÓMICOS.....	24
6.2. EXPLORACIÓN DE DATOS.....	25
6.3. TRANSFORMACIÓN DE LOS DATOS.....	30
6.4. MODELAMIENTO.....	32
6.5. EVALUACIÓN.....	34
6.6. APLICACIÓN DE LOS MODELOS.....	37
6.7. ANÁLISIS DE RESULTADOS.....	40
7. CONCLUSIONES.....	44
8. TRABAJO FUTURO.....	47
9. LISTA DE REFERENCIAS.....	48
10. ANEXOS.....	50

LISTA DE TABLAS

Tabla 1. Matriz de estado del arte.....	12
Tabla 2. Escenarios de aplicación del modelo	21
Tabla 3. Total muertes por sector económico 2009 - 2022	24
Tabla 4. Variables estudiadas.....	25
Tabla 5. P-valores prueba Shapiro-Wilk por variable y sector económico.....	25
Tabla 6. Valores atípicos por sectores económicos y variables	31
Tabla 7. Componentes para la variable respuesta número de AT en el sector minas - clase de riesgo V	33
Tabla 8. Resultados de los modelos por variable y sector económico	35

LISTA DE FIGURAS

Figura 1. Metodología general	23
Figura 2. Total de Trabajadores – Sector Minas – Clase de riesgo V.....	26
Figura 3. Número de accidentes de trabajo – Sector construcción – Clase de riesgo V... 27	
Figura 4. Número de enfermedades laborales calificadas – Sector Inmobiliario	28
Figura 5. Interacción de variables objetivo – sector Inmobiliario	29
Figura 6. Mapa de calor, relacionamiento de variables – Sector Inmobiliario	30
Figura 7. Simulación variable Número AT en sector minas – clase de riesgo V.....	39
Figura 8. Simulación variable Número AT en sector transporte – clase de riesgo IV	39

LISTA DE ANEXOS

Anexo 1. Tablas de composición de componentes principales por sector económico.....	50
Anexo 2. Resultados escenarios de simulación por sector económico.	50
Anexo 3. Gráficas de aplicación del modelo en los diferentes sectores económicos.	50
Anexo 4. Gráficas de análisis multivariado.....	50

RESUMEN

La dinámica de los riesgos laborales es muy variable dependiendo los sectores económicos y las clases de riesgo en donde se desarrollen las actividades, ya que en estas se pueden presentar diversas cantidades de accidentes de trabajo y sus consecuencias asociadas como muertes, invalidez o incapacidades permanentes parciales. El análisis de la dinámica del mercado laboral permitiría identificar aquellos sectores económicos y sus respectivas clases de riesgo en las que se podrían presentar los mayores impactos sobre los aspectos asociados a la accidentalidad laboral, de manera que se pueda anticipar la ocurrencia de dichos eventos y fortalecer su respectiva gestión. El presente estudio tiene por objetivo desarrollar un modelo matemático y evaluar su ajuste con los datos históricos de los años 2009 al 2022 para los sectores económicos inmobiliario, construcción, minas, transporte, comercio y agricultura el cual permita proyectar la dinámica del mercado laboral colombiano durante los años 2023 y 2024 y su relación con el crecimiento en el número de trabajadores o empresas. Se utilizó la metodología SEMMA y se desarrolló un modelo matemático basado en el análisis de componentes principales (PCA) y la aplicación de una regresión lineal múltiple a partir de dichos componentes, el cual presentó diferentes ajustes según la variable analizada, el sector económico y la clase de riesgo asociada. Se proyectaron escenarios de crecimiento de los sectores económicos de acuerdo con las dinámicas del mercado laboral actual en Colombia y se evaluaron los cambios en las variables mencionadas, logrando proyectar una visión general de la accidentalidad laboral y sus consecuencias. Finalmente, los modelos mostraron ajustes muy variables, que van desde 0 para variables que nunca tuvieron ocurrencia hasta valores superiores a 0.9, evaluados con la medida del R^2 . Logrando generar predicciones con un buen nivel de ajuste para algunos sectores económicos y sus clases de riesgo, la cual serviría como insumo para toma de decisiones en materia de riesgos laborales.

Palabras clave: Accidente laboral, pensión, invalidez, indemnización, incapacidad parcial, riesgos laborales, mercado laboral, componentes principales, regresión lineal.

ABSTRACT

The dynamics of occupational risks vary greatly depending on economic sectors and the types of risks involved in the activities. This is because varying numbers of workplace accidents and their associated consequences—such as deaths, disabilities, or partial permanent incapacities—can occur in these sectors. Analyzing the dynamics of the labor market would allow for the identification of those economic sectors and their respective risk categories where the greatest impacts on aspects related to workplace accidents could occur. This would enable anticipation of such events and strengthen their respective management. This study aims to develop a mathematical model and evaluate its fit with historical data from 2009 to 2022 for the real estate, construction, mining, transportation, commerce, and agriculture sectors. This model would allow for projecting the dynamics of the Colombian labor market for the years 2023 and 2024 and its relationship with the growth in the number of workers or companies. The SEMMA methodology was used, and a mathematical model based on Principal Component Analysis (PCA) and the application of multiple linear regression from these components was developed. The model showed different levels of fit depending on the variable analyzed, the economic sector, and the associated risk category. Scenarios were projected for the growth of economic sectors based on the current labor market dynamics in Colombia, and changes in the mentioned variables were evaluated, successfully projecting an overall view of workplace accidents and their consequences. Finally, the models showed highly variable fits, ranging from 0 for variables that never occurred to values greater than 0.9, evaluated using the R^2 measure. Achieving to generate predictions with a good level of fit for some economic sectors and their risk categories, which would serve as input for decision-making in the area of occupational risks.

Keywords: Workplace accident, pension, disability, compensation, partial incapacity, occupational risks, labor market, principal components, linear regression.

1. INTRODUCCIÓN

Los trabajadores son una de las piezas más importantes en el desarrollo empresarial y económico del país, por lo cual, es de vital importancia que los gobiernos, empresas y trabajadores propiamente, unan esfuerzos para prevenir que estos sufran algún tipo de accidente o sus posibles consecuencias tales como incapacidades parciales permanentes, pensiones por invalidez o muertes, los cuales representan afectaciones directas leves o graves sobre la integridad de los trabajadores.

De acuerdo con las estadísticas de la Organización Internacional del Trabajo (OIT), al año mueren alrededor de 1.9 millones de trabajadores a causa de accidentes laborales en el mundo. De igual manera, la OIT informa que cada año se producen más de 360 millones de accidentes de origen laboral que presentan por lo menos 4 días de incapacidad, lo que representa una importante pérdida de fuerza laboral para las empresas (Organización Internacional del Trabajo, s.f.).

En línea con lo mencionado, se ha evidenciado un aumento en la conciencia pública y privada respecto a la necesidad de cuantificar los accidentes que surgen como producto de la dinámica laboral en los diferentes países, permitiendo tener un panorama más preciso sobre las afectaciones que pueden sufrir los trabajadores como producto de la prestación de sus servicios a los diferentes sectores económicos del país (Álvarez et al., 2019).

Uno de los factores que afectan la dinámica del mercado laboral es que para un mismo sector económico pueden existir diferentes niveles de riesgo, esto dependiendo de las actividades específicas que desarrollan las diferentes empresas, por lo cual, es importante tener claridad sobre esta clasificación y su aporte dentro del consolidado global de cada sector.

De acuerdo con lo establecido en el decreto 768 de 2022, las diferentes actividades económicas del país están clasificados en niveles de riesgo I, II, III, IV o V; clasificación que dependerá del nivel de complejidad de las actividades desarrolladas y la exposición de los trabajadores a los diferentes factores de riesgo.

En esta clasificación, los niveles IV y V son considerados de riesgo alto, y los niveles I, II y III son considerados niveles de riesgo bajo, de acuerdo con lo establecido en la resolución 0312 de 2019.

De igual manera, y para fines de la proyección que se busca realizar, se resalta que los accidentes de trabajo (AT) están definidos en el artículo 3 de la Ley 1562 de 2012 como “Es accidente de trabajo todo suceso repentino que sobrevenga por causa o con ocasión del trabajo, y que produzca en el trabajador una lesión orgánica, una perturbación funcional o psiquiátrica, una invalidez o la muerte”.

Algunos sectores económicos como el de la construcción y la agricultura, aportaron la mayor cantidad de accidentes laborales calificados para el año 2022, con un total de 59.228 y 40.000 respectivamente (Federación de Aseguradores Colombianos, 2023).

Por otra parte, las fatalidades por accidentes laborales permiten identificar otros sectores económicos de interés para la gestión de peligros y riesgos. En primera instancia, el sector económico de minas y canteras presentó la mayor cantidad de muertes por accidentes laborales para el año 2022, con un total de 113 fallecidos. En segundo lugar, se encontró el sector transporte, almacenamiento y comunicaciones con un total de 56 siniestros producto de accidentes laborales (Federación de Aseguradores Colombianos, 2023).

Las estadísticas del año 2022 demuestran que actualmente se siguen presentando altos números de accidentes laborales en los diferentes sectores económicos del país (aumentó un 5% respecto al año 2021, en el cual se presentaron un total de 319.948), así como también, una considerable cantidad de fatalidades que se originan de estos eventos, lo que deja en evidencia que aunque existen políticas públicas enfocadas a la prevención de riesgos laborales en general, como por ejemplo lo que establece la resolución 0312 de 2019 con la implementación del Sistema de Gestión en Seguridad y Salud en el Trabajo (SG-SST), el aumento de accidentes de origen laboral indica que dichas políticas no están siendo suficientes respecto a lo que demanda la dinámica del mercado laboral.

Lo anterior expuesto, presenta la necesidad de evaluar la dinámica actual del mercado laboral y proyectar el panorama futuro sobre las posibles incidencias o situaciones que se presentarán en el país en materia de accidentalidad laboral de acuerdo con la dinámica de los principales sectores económicos del mercado laboral, todo esto con el fin identificar patrones que permitan formular mecanismos de prevención o mitigación en los resultados de accidentes laborales en los próximos años, para lograr cuidar el recurso más valioso que tienen las empresas, el humano.

Para el desarrollo del presente estudio se usaron los datos que reporta de forma periódica la Federación de Aseguradores Colombianos – FASECOLDA, con relación a la accidentalidad laboral y sus consecuencias como muertes, invalidez o incapacidades permanentes, en los diferentes sectores económicos. Se realizará una proyección del comportamiento de las variables analizadas logrando identificar con esto los que podrían ser focos de interés para las autoridades laborales del país en los periodos futuros.

2. PREGUNTA DE INVESTIGACIÓN

¿Cuál es la proyección de crecimiento de la accidentalidad laboral y sus consecuencias en Colombia, en función de la dinámica del mercado laboral en los principales sectores económicos?

3. MARCO CONCEPTUAL

3.1. MARCO TEÓRICO

El presente estudio se desarrollará en torno a los conceptos de accidentes laborales en Colombia, así como los de indemnizaciones por pérdida de capacidad laboral y las pensiones por invalidez. La dinámica del mercado laboral, sus principales sectores económicos y la clasificación de los niveles de riesgo en las cuales desarrollan actividades cada uno de estos.

3.1.1. Sectores Económicos

La clasificación de sectores económicos que se aplicará en el presente estudio será la realizada oficialmente por FASECOLDA, de acuerdo con su reporte mensual de estadísticas para el ramo de Riesgos Laborales. Partiendo de esto, se abordarán las dinámicas del mercado laboral para los sectores económicos inmobiliario, construcción, minas y canteras, transporte, almacenamiento y comunicaciones (transporte), comercio y agricultura, ganadería, caza y silvicultura (agricultura).

En primer lugar, se encuentra el sector inmobiliario, el cual consta de un conglomerado general de diferentes actividades económicas como lo son comercialización de bienes inmobiliarios, comercialización y alquiler de enseres domésticos, actividades relacionadas con bases de datos, alquiler de equipos de transporte, alquiler de maquinaria, entre otros. Como se puede observar, este sector presenta una alta variedad de actividades económicas, lo cual a su vez lo lleva a acumular una gran cantidad de población de trabajadores, siendo el que tiene la mayor cantidad de todos los sectores que se reportan, y, a su vez, es el que tiene la mayor cantidad de muertes de origen laboral para el periodo 2009 – 2022 (Federación de Aseguradores Colombianos, 2023).

De igual manera se encuentra el sector construcción, el cual abarca específicamente todas las actividades asociadas con trabajos eléctricos, pintura, instalaciones hidráulicas construcción de edificaciones y obras de ingeniería civil

en general, entre otras. Dada la complejidad de sus actividades y el nivel de riesgo de la mayoría de estas, el sector construcción es el segundo con la mayor cantidad de muertes de origen laboral en el periodo 2009 – 2022 (Federación de Aseguradores Colombianos, 2023).

Posteriormente, se encuentra el sector de minas y canteras, el cual compila todas aquellas actividades relacionadas con explotación minera, de hidrocarburos, extracción de minerales, gas, aglomerados, entre otros. Este sector económico compila la mayor parte de sus trabajadores en los niveles de riesgo laboral más altos (4 y 5) debido a la complejidad de sus actividades y es el tercero con mayor cantidad de muertes de origen laboral para el periodo 2009 – 2022.

De igual manera, se realizará el análisis del sector económico, transporte, el cual consolida actividades económicas como agencias de viajes, transporte terrestre intermunicipal, urbano colectivo, transporte fluvial, aéreo, entre otros. Este sector se presenta como el cuarto con la mayor cantidad de muertes de origen laboral en el periodo 2009 – 2022.

En quinto lugar, se realizará el análisis sobre el sector económico comercio, el cual también reúne una gran cantidad de trabajadores debido a la compilación de actividades económicas como comercio al por mayor y por menor de todo tipo de artículos alimenticios, calzado, equipos de oficina, textiles, entre otros. Asimismo, es el sector número cinco con la mayor cantidad de muertes de origen laboral durante el periodo 2009 – 2022.

Por último, se analizará el sector económico agricultura, el cual compila las actividades económicas asociadas a todos los servicios agrícolas y ganaderos, producción especializada de productos alimenticios, cría de ganado porcino, vacuno, entre otros.

3.1.2. Seguridad y Salud en el Trabajo

La seguridad y salud en el trabajo es un tema de amplio conocimiento y desarrollo en Colombia en la actualidad, teniendo sus orígenes hacia 1915 con la

Ley 57 de 1915 del Congreso de Colombia, referente a reparaciones por accidentes de trabajo, continuando a lo largo de los años con diferentes desarrollos técnicos y normativos, hasta la actualidad, en donde ya se cuentan con conceptos claramente definidos como el de accidentes de trabajo, SG-SST, entre otros.

En la actualidad, el gobierno de Colombia cuenta con un Plan Nacional de Seguridad y Salud en el Trabajo 2022 – 2031, en el cual se encuentran enmarcados todos los lineamientos asociados a la gestión de los peligros y riesgos en los lugares de trabajo, entre otras actividades. Este, complementado con la normatividad nacional vigente, permiten definir claramente aspectos asociados a la clasificación de los niveles de riesgo de los sectores económicos, la ocurrencia de accidentes o enfermedades laborales, entre otros (Ministerio del Trabajo, 2022)

Por otra parte, los esfuerzos del gobierno nacional en la gestión de los riesgos laborales, se han intensificado mediante el diseño de políticas públicas y aspectos como la obligatoriedad de implementación del SG-SST, el cual, actualmente está definido para los estándares mínimos enmarcados en la resolución 0312 de 2019 del Ministerio del Trabajo y otros aspectos mencionados en el Decreto 1072 de 2015. De esta manera, se han presentado impactos en la gestión de peligros y riesgos que se desarrolla en diferentes sectores económicos.

Dentro de los diferentes conceptos que se aplicarán en el estudio, se encuentra el de incapacidad permanente parcial, la cual corresponde a aquella situación que, producto de un accidente o una enfermedad laboral, genera una disminución definitiva de la capacidad laboral del trabajador de entre el 5% al 49%, lo cual representa que el trabajador que sufre esta situación se hace acreedor a una indemnización proporcional al daño sufrido (Ley 776 de 2002).

Finalmente, se tomará en cuenta el concepto de estado de invalidez, el cual a su vez hace que el trabajador sea acreedor a una pensión de invalidez, esta condición se presenta cuando, como producto de un accidente o enfermedad laboral, la persona pierde el 50% o más de su capacidad laboral.

3.1.3. Mercado Laboral

Es importante tener una visión referente al foco de interés del gobierno nacional actual en materia de crecimiento económico y mercado laboral, en el Plan Nacional de Desarrollo se presenta como prioridad el fomento de la formalización y crecimiento de algunos sectores económicos como el de la agricultura, mediante la formulación de criterios de equidad y transparencia en las unidades productivas agrícolas y pecuarias (Gobierno de Colombia, 2023).

Por otra parte, el gobierno nacional plantea en su plan cuatrienal de desarrollo para los años 2022 al 2026, políticas y estrategias de eficiencia energética, diseñando y promoviendo la implementación de sistemas de gestión eficiente de la energía, estimulando el consumo de energía eléctrica de forma racional y responsable, lo que se traduciría en un posible impacto negativo sobre el sector de minas y canteras (incluido hidrocarburos), esto unido a las iniciativas de fortalecimiento del marco normativo e incentivos para la descarbonización del sector transporte (Gobierno de Colombia, 2023).

De acuerdo con esto, se prevé que estos objetivos gubernamentales generen variaciones en las poblaciones de los diferentes sectores económicos, los cuales pueden sufrir cambios favorables o desfavorables en sus poblaciones de trabajadores, y de igual forma, estos cambios podrían tener algún grado de relación con los accidentes de origen laboral que se presenten en el país y sus diferentes consecuencias.

Adicionalmente, el gobierno nacional actual tiene propuestas para apoyar y fomentar la creación de nuevos sistemas de transporte público que estén alineados con las políticas energéticas del país y buscando la descarbonización del mismo, lo que podría impactar de forma positiva a la población de este sector económico. Sin embargo, igualmente el gobierno nacional propone desarrollar, entre otros, infraestructura férrea de carga a nivel nacional, lo cual podría, contrario a lo anterior, impactar de forma negativa la dinámica de fuerza laboral en este sector económico, por lo que la dinámica que se pueda presentar no es del todo

clara, teniendo en cuenta los objetivos del gobierno nacional (Gobierno de Colombia, 2023).

Por otra parte, de acuerdo con el reporte de la Superintendencia de sociedades con corte a diciembre de 2022, el número de procesos de insolvencia de empresas en el sector de la construcción aumentó un 8% con respecto al corte de septiembre del mismo año, pasando de 458 procesos a 495 en el cierre del año, conservando la tendencia al alza que presentó durante todo el año 2022, por lo que el panorama para este sector presenta un escenario de posible decrecimiento en los siguientes años (Superintendencia de Sociedades, 2022).

Finalmente, y basados en la primicia del gobierno nacional encaminada a una política pública de trabajo digno y decente, este buscará promover la formalización laboral, condición que en la actualidad afecta al 58.2% de la población nacional ocupada (Departamento Administrativo Nacional de Estadística, 2023); se espera que esta política de trabajo digno y decente, disminuya las proporciones de informalidad laboral, aumentando de esta forma la población de los diferentes sectores económicos en diferentes dinámicas.

Por lo anteriormente mencionado, el análisis de información y la proyección del crecimiento o disminución de los accidentes y enfermedades laborales en los diferentes sectores económicos del país, permitiría diseñar estrategias que mitiguen estos casos y permitan contrarrestar el crecimiento de estos aspectos y de esta forma atenuar su impacto en la salud de los trabajadores en Colombia.

3.1.4. Modelos matemáticos

3.1.4.1. Análisis de Componentes Principales (PCA)

El PCA es una de las técnicas estadísticas más comúnmente utilizadas en el análisis multivariado de datos, esta técnica tiene por objeto extraer la información más relevante de todo el conjunto de datos, enfocándose en la variable respuesta, y representando dicha información en nuevas variables ortogonales denominadas componentes principales. (Liu et al., 2022).

El PCA presenta importantes limitantes para grupos de datos con una gran cantidad de variables, sin embargo, se ajusta muy bien en aquellos con una cantidad reducida de las mismas, lo anterior se debe a la cantidad de componentes necesarios para explicar la variable respuesta, determinado en la cantidad de varianza que se logra explicar de dicha variable a partir de la formulación de n cantidad de componentes de los datos (sckit-learn.org, 2011).

3.1.4.2. Modelos de Regresión lineal

Los modelos de regresión lineal pueden desarrollarse de forma simple (una variable independiente y una variable dependiente) o de forma múltiple (dos o más variables independientes y una variable dependiente). Tienen por objetivo definir un trazado recto que permita la mejor predicción de valores, evaluando las distancias de los valores observados respecto a la línea definida por el modelo (Dascălu y Cozma, 2017).

Un modelo de regresión lineal múltiple podría expresarse como:

$$Y = A_0 + A_1X_1 + A_2X_2 + A_nX_n + \epsilon$$

- Y es la variable respuesta o dependiente
- X_n representa las variables independientes
- A_0 es el valor de Y cuando $X = 0$
- A_n Es la pendiente que representa el cambio en los valores de Y a partir de los cambios en X
- ϵ es el error aleatorio, correspondiente a la diferencia entre el valor observado y el predicho.

En el modelo de regresión lineal múltiple, es considerada como una extensión de la lineal simple, ya que en esta se incluye más de una variable como independiente o explicativa respecto a una única variable respuesta, por lo que su ajuste puede ser mejor en casos donde la variable respuesta tenga un nivel de

complejidad mayor, ya que esta permite obtener diversos coeficientes de regresión (positivos o negativos) que explican en mayor medida el comportamiento de la variable respuesta (Kutner et al., 2004).

El nivel de precisión del modelo dependerá principalmente de la relación lineal que exista entre las variables objeto de análisis; si dicha relación no es lineal, la precisión del modelo disminuiría y por ende sus predicciones serían menos acertadas (Kutner et al., 2004).

3.2. ESTADO DEL ARTE

En la actualidad, la aplicación de modelos matemáticos para la proyección de accidentes laborales y sus consecuencias está creciendo, y su aplicabilidad en diferentes sectores económicos es cada más frecuente y se convierte en un importante insumo para el análisis de comportamientos históricos en las empresas y para realizar la proyección de posibles escenarios futuros bajo los cuales se pueden tomar decisiones de carácter preventivo y de esta forma las empresas puedan hacer una mejor gestión en seguridad y salud en el trabajo y preservar la salud de sus colaboradores.

En el campo de riesgos laborales se ha evidenciado la aplicabilidad de diferentes modelos matemáticos con el fin de predecir el posible comportamiento de variables como la accidentalidad laboral de las empresas, clasificación de situaciones que puedan generar accidentes de trabajo, entre otras (tabla 1).

Tabla 1. Matriz de estado del arte

Ítem	Año	Título	Autor(a)	Descripción	Modelo	Conclusiones
1	2023	Propuesta de diseño de un modelo matemático predictivo aplicado a la prevención de riesgos laborales en el sector de la construcción.	Villada J., Castaño L. y Pereira L.	Revisión de criterios de accidentalidad laboral, análisis de indicadores y principales modelos matemáticos para explicación del comportamiento de la accidentalidad laboral en el sector de la construcción	Regresión lineal	Se consideró que no existen relaciones entre las causas de accidentes laborales debido a que estos se presentan como eventos repentinos
2	2022	Generación de un modelo predictivo de la fuerza de agarre para trabajadores del sector panelero.	Delgado Y., Arroyo J., Monroy M. y Zea C.	Identificar la fuerza de agarre de los diferentes trabajadores para diseñar estrategias que eviten sobre esfuerzos y posibles lesiones osteomusculares	Regresión lineal múltiple	Debido a que el nivel de confianza en ninguno de los casos supera el 50% es correcto plantear que se debería ampliar la muestra poblacional, si se desea plantear un modelo para otro grupo poblacional, se debería discriminar para mano derecha e izquierda o dominante y no dominante, según intereses de los investigadores.
3	2021	Trend analysis and projection of work accidents cases based on work shifts, workers age, and accident types	Nai'em, M. Furqaan, Darwis, A. Muflihah y Sholihin Shiddiq	Análisis de los casos de accidentes laborales en el periodo 2009 - 2022 con base en los turnos de trabajo, edad de los trabajadores y tipos de accidentes.	Series de tiempo	El estudio permitió identificar que durante el periodo 2009-2015 los turnos de trabajo en la mañana representaron el 47% de los accidentes de trabajo, sin embargo, en la proyección que realizaron esta presentaría una disminución hasta un aproximado del 22% de los accidentes únicamente.
4	2019	Modelamiento predictivo de riesgos psicosociales en trabajadores de una empresa del sector educativo	Jesús P.	Uso de metodología CRISP-DM para desarrollar un modelo predictivo de la exposición al riesgo psicosocial en trabajadores de una institución educativa	Regresión logística binaria.	Se encontró que el modelo predictivo desarrollado tiene una predictibilidad del 73.9%, siendo aceptable para los requerimientos del negocio del sector educativo.
5	2021	Modelo de predicción de riesgos psicosociales en el transporte urbano de pasajeros usando técnicas de Inteligencia Artificial	Lara A.	Modelo de predicción de riesgos psicosociales en conductores de autobús	K-Nearest Neighbors - Árboles de decision Supported vector Machine- No equilibrado Supported vector Machine-	El estudio permitió identificar que el algoritmo de SVM tuvo el mejor rendimiento con un 93%, comparando con los otros que tuvieron un rendimiento alrededor del 87% para clasificar el estrés laboral. En consecuencia, en el trabajo realizado el modelo SVM con datos equilibrados se destaca en la predicción del estrés en la población objeto de estudio y subsecuentemente en la identificación de los riesgos psicosociales laborales que los originan, a través de las nueve dimensiones que se emplean en el cuestionario físico del INSHT de España.

					Equilibrado	
6	2017	The principal components analysis-method to reduce the collinearity in multiple linear regression model; application in medical studies	Dascălu, C., Cozma, C.	Uso del PCA para reducir colinealidad y optimizar la eficiencia de un modelo de regresión lineal múltiple para predecir las condiciones de salud de pacientes en función de su actividad cardiaca.	PCA – Regresión lineal múltiple	Los modelos de regresión lineal múltiples precedidos de un PCA son muy útiles para identificar la relación entre las variables objeto de análisis, facilitando la reducción de sets de datos en una determinada cantidad de componentes que explican de forma más precisa el comportamiento de la variable respuesta.
7	2012	Application of Principal Component Analysis & Multiple Regression Models in Surface Water Quality Assessment	Mustapha, A., Abdu, A.	Aplicación de PCA y regresión lineal múltiple para identificar las fuentes de contaminación y su contribución a la variación en la calidad del agua de un río.	PCA – Regresión lineal múltiple	El PCA permitió identificar los componentes que representaban el mayor impacto sobre la variación en la calidad del agua del río objeto de estudio y gracias a dicho análisis el modelo de regresión lineal múltiple tuvo un R ² de 0.942.
8	2021	Predicción de la accidentalidad laboral en la industria de pulpa y papel usando algoritmos de clasificación	Mosquera, R., Parra, L., Ledesma, A. y Bonilla, H.	Aplicación de modelos de clasificación para la identificación de situaciones de riesgo que se pueden convertir en accidentes laborales en la industria de pulpa y papel.	Árboles de decisión Redes Bayesianas	Se concluyó que el modelo de árboles de decisión presentó el mejor nivel de ajuste con una precisión del 90% en la identificación de situaciones con potencial de convertirse en accidentes laborales.

FUENTE: Construcción Propia

En el estudio desarrollado por (Villada et al., 2023), realizaron la aplicación de un modelo de regresión lineal para identificar la relación existente entre los factores y las condiciones de riesgo que ocasionan accidentes laborales en el sector de la construcción, en su estudio logrando determinar que el uso de modelos de regresión se presenta como una importante herramienta para identificación de posibles soluciones en problemas organizacionales asociados con la seguridad de los trabajadores y podrían utilizarse como un soporte para la toma de decisiones estratégicas a nivel corporativo.

Otro ejemplo de lo mencionado, en el estudio realizado por (Delgado et al., 2022), quienes utilizaron un modelo predictivo basado en regresión lineal múltiple para determinar la fuerza de agarre de una población de trabajadores del sector panelero, y de esta forma poder determinar las capacidades de carga que estos podrían tener, con el fin de diseñar estrategias de prevención de riesgos biomecánicos y promover la salud de dichos trabajadores. En este estudio utilizaron la regresión lineal múltiple, la cual les permitió identificar que los factores personales que más influyen en la fuerza de agarre de los trabajadores son el consumo de alcohol, la dominancia de la mano y la edad, a partir de lo cual, se pudieron generar medidas de mitigación del riesgo biomecánico.

Otra muestra del uso de modelos matemáticos en el campo de riesgos laborales es el estudio realizado por (Lara, 2021), en el cual se utilizaron los modelos de K-nearest neighbors (KNN), Decision Tree (DT) y Support Vector Machine (SVM) con el fin de realizar la clasificación e identificación de las diferentes condiciones que generan riesgos psicosociales, especialmente estrés laboral en conductores del sector público en la ciudad de Ambato, Ecuador. La aplicación de los modelos le permitió identificar que el modelo SVM presentó la mayor eficiencia con un score de 93% y una valoración Jaccard del 87%.

Por otra parte, el uso de PCA junto con modelos de regresión lineal se ha documentado en otros campos, como es el caso del estudio realizado por (Dascălu y Cozma, 2017), quienes utilizaron este modelo matemático para predecir cambios

en las condiciones de salud de un grupo de pacientes a partir del impacto de diferentes variables en la variable respuesta de actividad cardiaca de los pacientes. El PCA les permitió identificar un total de 11 componentes principales que más influencia presentaban sobre la variable respuesta, con los cuales lograron explicar un 75.41% de la varianza en la variable respuesta, y obtener un valor de R^2 de 0.417.

De igual manera, en el estudio realizado por (Mustapha y Abdu, 2012), se utilizó el PCA junto con un modelo de regresión lineal múltiple para determinar los componentes que presentaban mayor impacto en la variación de la calidad del agua en un río en Malasia. Lograron identificar un total de 5 componentes principales que explicaban un 83.1% de la varianza en la variable respuesta y con estos lograron un ajuste muy alto en su modelo de regresión lineal múltiple el cual presentó un valor de R^2 de 0.942.

Finalmente, podemos encontrar el estudio realizado por (Mosquera et al., 2021), asociado a la predicción de la accidentalidad en la industria de pulpa y papel mediante el uso de modelos matemáticos de clasificación basados en árboles de decisión y redes bayesianas, con el fin de determinar las condiciones de seguridad que presentaban mayor riesgo de materializarse en accidentes laborales en esta industria, concluyendo que para los datos recopilados el modelo de árboles de decisión tuvo un 90% de precisión en la determinación de las situaciones que se materializaron y en las que podrían materializarse a futuro.

4. OBJETIVOS

4.1. OBJETIVO GENERAL

Desarrollar un modelo matemático que permita proyectar la cantidad de accidentes laborales en Colombia y sus posibles consecuencias tales como muertes, invalidez o incapacidades permanentes parciales, en función de la dinámica de los principales sectores económicos del mercado laboral.

4.2. OBJETIVOS ESPECÍFICOS

- Consolidar la base de datos asociadas a las dinámicas del mercado laboral de los principales sectores económicos del país durante el periodo 2009 – 2022.
- Desarrollar, validar, calibrar y verificar un modelo que permita realizar la proyección del número de AT, muertes por AT, pensiones de invalidez e incapacidad permanente parcial por AT en el mercado laboral colombiano para los años 2023 y 2024.
- Analizar el efecto de diferentes escenarios de dinámica del mercado laboral sobre la ocurrencia de accidentes laborales en Colombia.

5. METODOLOGÍA

El presente estudio se desarrolla bajo la metodología SEMMA (muestreo, exploración, modificación, modelamiento y evaluación, por sus siglas en inglés) ya que esta contempla las diferentes etapas del tratamiento que se le dará a los datos con los que se desarrolló el mismo. Durante el desarrollo de la metodología se iniciará con la etapa de muestreo, en la cual se buscará tomar muestras de los datos para determinar aspectos como proporciones u otros y seleccionar los sectores económicos objeto de análisis (Barrios et al., 2019).

En la etapa de exploración se buscará tener una visión general de la información que reflejan los datos, se realizarán pruebas para identificar aspectos como la distribución de los mismos y otros. Posteriormente, en la tercera etapa, se realizará la modificación de los datos, de acuerdo con el desarrollo del análisis de estos, se evaluarán y ajustarán los valores atípicos que se puedan presentar en las series de datos, se aplicarán procesos de estandarización o normalización en caso de ser necesario, con el fin de tener la mejor calidad de los mismos.

En la cuarta etapa, se realizará el modelado, en esta, se evaluará la aplicación de diferentes modelos matemáticos para determinar cuál de estos permite obtener los mejores resultados sobre las proyecciones realizadas, a partir de la exploración y modificación que se realizó a los datos, para finalmente, realizar el proceso de evaluación, en el cual, se analizarán las interacciones que se presentaron y la información entregada por los modelos matemáticos aplicados, con el fin de concluir sobre los objetivos propuestos en el presente estudio.

5.1. MUESTREO

La presente investigación se plantea sobre la base general de estudio de los sectores económicos caracterizados en la base de datos de FASECOLDA, en su ramo de riesgos laborales. Para esto, se realizará una selección de sectores económicos de acuerdo con la cantidad de muertes que hayan presentado en el periodo de análisis (2009-2022).

La fuente de información corresponde a FASECOLDA, esta entidad, como agremiación de aseguradores en Colombia, consolida periódicamente la información reportada por las Asegurados de Riesgos Laborales - ARL sobre los trabajadores dependientes e independientes registrados en el Sistema General de Riesgos Laborales, permitiendo de esta forma obtener la información clara y confiable sobre la dinámica de la accidentalidad laboral en el país, sobre todos los trabajadores que tienen algún tipo de vinculación laboral con compañías y aquellos que realizan aportes de forma independiente. Se analizarán los datos correspondientes a los años 2009 – 2022.

Es importante aclarar que para las personas que realizan actividades laborales sin ninguna vinculación laboral, o que sean independientes, pero no realicen aportes al Sistema General de Seguridad Social, sus estadísticas no se encuentran registradas dentro de los datos de FASECOLDA, por lo que la información sobre posibles accidentes o enfermedades laboral que puedan ocurrir con estos colaboradores no se puede cuantificar dentro del presente estudio.

5.2. EXPLORACIÓN

En la exploración de los datos se realizará una revisión general de los mismos, identificando patrones que puedan interferir con los procesos de análisis (como por ejemplo caracteres especiales u otros). De igual manera, se realizará una revisión general de cada conjunto de datos, con el fin de entender su estructura, su alcance, la posible correlación entre las variables.

Dentro de las estrategias de exploración que se aplicarán, se desarrollará inicialmente una prueba Shapiro-Wilk con el fin de determinar si los sets de datos de cada sector económico tienen una distribución normal o no. Posteriormente, se realizará un análisis multivariado para conocer el comportamiento de las variables objeto de interés durante el periodo de estudio y la relación entre ellas para entender la influencia o dependencia que pueden tener entre ellas y a partir de esta interacción poder realizar la identificación de los posibles modelos matemáticos que se ajustarían mejor con los datos y posteriormente realizar la

comprobación de los mismos.

De igual manera, se realizará una identificación de datos atípicos, esto mediante la aplicación de la prueba ESD (extreme studentized deviate), esta permitió el análisis de las series de datos con las que se cuentan para cada variable y mediante la validación estadística de esta prueba se pudieron identificar aquellos valores atípicos que se encontraban en la misma.

5.3. MODIFICACIÓN O TRANSFORMACIÓN

En esta etapa se realizará la selección de las variables de acuerdo con el objeto de estudio, filtrando y seleccionando únicamente las que corresponden a accidentes, fatalidades, número de casos de pensiones por invalidez, número de casos de incapacidades parciales permanentes y crecimiento poblacional de trabajadores o empresas.

Lo anterior se realizará con el fin de poder formular los diferentes escenarios que se analizarán y sobre los cuales se realizará la proyección de los datos. Por lo cual, la identificación de la relación entre variables será de vital importancia para aumentar al máximo posible la precisión de las proyecciones que se obtendrán.

En el conjunto de datos de FASECOLDA, asociado a trabajadores por sector económico, accidentes laborales y sus consecuencias por cada uno de estos, se realizará la selección de los sectores económicos que serán objeto de análisis, tomando como criterio de selección la cantidad de muertes que han presentado durante los años 2009 – 2022 y de esta forma, poder aplicar los diferentes modelos matemáticos que se desarrollarán.

Finalmente, se realizará el ajuste de los valores atípicos que se identificaron en la exploración, estos serán reemplazados con la medida de tendencia central que mejor se ajuste de acuerdo con el set de datos, logrando de esta forma eliminar el impacto que estos valores puedan tener en los modelos que se desarrollarán, permitiendo de esta forma aumentar su eficiencia y ajuste.

5.4. MODELAMIENTO

En la etapa del modelamiento se buscará desarrollar un modelo matemático cuya aplicación permita obtener diferentes valores de precisión en la predicción del comportamiento de las variables objeto de análisis en función cambios en la dinámica poblacional de los diferentes sectores económicos analizados.

Es importante aclarar que el desarrollo del presente estudio se realizará en el lenguaje de programación Python, por lo cual, se hará uso de diferentes librerías tales como Pandas, Numpy, Scikit-learn, entre otras, con el fin de estructurar de la mejor forma posible los datos, realizando las etapas de limpieza y procesamiento, para obtener las mayores precisiones sobre las proyecciones realizadas y de esta forma alcanzar el mejor análisis posible.

En esta etapa, se probarán modelos como series de tiempo, análisis de componentes principales (PCA, por sus siglas en inglés) combinado con la aplicación de regresiones lineales simples y múltiples a partir de la formulación de los componentes, buscando identificar los mejores niveles de ajuste de las predicciones con los conjuntos de datos.

5.5. EVALUACIÓN

Esta etapa se desarrollará mediante la definición de conjuntos de datos de prueba, para esto, se destinará el 80% de los datos de cada variable para el entrenamiento de los modelos y el 20% restante se destinará a un grupo de datos de muestreo, sobre el cual se probará la eficiencia de los diferentes modelos, esto con el fin de obtener una cantidad significativa de pruebas de validación sobre los valores predichos por los modelos y evaluar de mejor forma su nivel de ajuste.

El nivel de ajuste de los modelos se evaluará mediante el componente R^2 de los mismos, el cual se mide en nivel de 0-1, siendo 1 el mejor nivel de ajuste del modelo y 0 indicaría que el modelo no se ajusta al conjunto de datos de la variable de análisis.

5.6. APLICACIÓN

Una vez desarrollados y validados los modelos, estos se aplicarán para realizar una predicción del comportamiento de la accidentalidad laboral y sus consecuencias, junto con la dinámica poblacional del mercado laboral para los sectores económicos seleccionados. Esto se realizará mediante un análisis en el cual se formularán diferentes variabilidades en el comportamiento de los conjuntos de datos y evaluar de esta forma el posible comportamiento de cada variable en los periodos de tiempo futuros.

Teniendo en cuenta que el conjunto de datos cuenta con dos variables poblacionales (Número de empresas y Total de trabajadores), estas dos se formularán como variables exógenas en diferentes escenarios en los cuales puedan aumentar o disminuir durante los siguientes 24 meses contados a partir de enero de 2023. La variación que presentarán estas variables estará relacionada con los objetivos de desarrollo que presenta el actual gobierno nacional para los diferentes sectores económicos, esta se formulará mediante una función triangular con un escenario pesimista que indicaría un posible decrecimiento del sector, una moda que sería el comportamiento más probable para el mismo y un escenario optimista que sería aquel en el que se presente el crecimiento más alto. (tabla 2).

Tabla 2. Escenarios de aplicación del modelo

Sector Económico	Escenario Pesimista	Escenario más probable	Escenario Optimista
Inmobiliario	Decrecimiento del 12% anual (1% mensual).	Crecimiento del 12% anual (1% mensual).	Crecimiento del 36% anual (3% mensual).
Construcción	Decrecimiento del 24% anual (2% mensual).	No presenta crecimiento.	Crecimiento del 12% anual (1% mensual).
Minas	Decrecimiento del 24% anual (2% mensual)	No presenta crecimiento	Crecimiento del 12% anual (1% mensual).
Transporte	Decrecimiento del 12% anual (1% mensual).	Crecimiento del 12% anual (1% mensual).	Crecimiento del 36% anual (3% mensual).
Comercio	Decrecimiento del 12% anual (1% mensual).	Crecimiento del 12% anual (1% mensual).	Crecimiento del 36% anual (3% mensual).
Agricultura	Decrecimiento del 12% anual (1% mensual).	Crecimiento del 24% anual (2% mensual).	Crecimiento del 48% anual (4% mensual).

FUENTE: Construcción Propia

Es importante tener en cuenta que la variación en las variables poblaciones representaría aumento en el número de empresas y trabajadores en el Sistema General de Seguridad Social en Riesgos Laborales, por lo cual no indicarían necesariamente un aumento en el número de empleos en el país, sino que también podría darse en función de la formalización del empleo en dichos sectores.

El criterio de aplicación de los modelos será la precisión de los mismos, por lo cual, sólo se aplicarán para la simulación de escenarios en aquellas variables en donde los modelos presenten una precisión superior a 0.4.

Finalmente, en la figura 1 se observa el diagrama de la metodología general que se aplicará en el desarrollo del presente estudio.

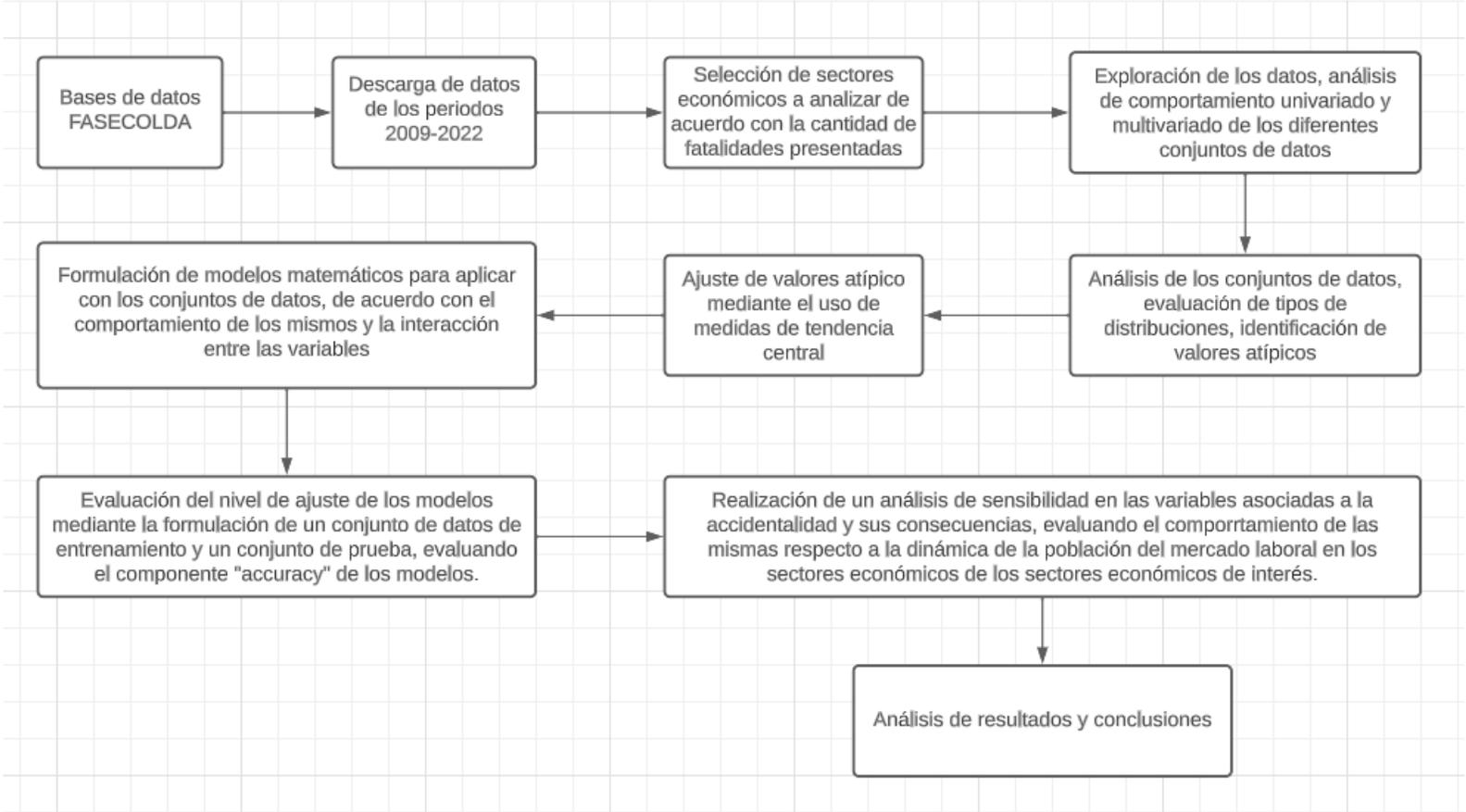


Figura 1. Metodología general

6. DESARROLLO Y RESULTADOS

6.1. SELECCIÓN DE SECTORES ECONÓMICOS.

De acuerdo con lo mencionado, los sectores económicos que se trabajarán en el presente estudio corresponden a aquellos que tuvieron la mayor cantidad de fatalidades durante los años 2009 – 2022, por lo cual, a continuación, se muestran las estadísticas para todos los sectores evaluados:

Tabla 3. Total muertes por sector económico 2009 - 2022

SECTOR ECONÓMICO	TOTAL MUERTES
Inmobiliario	1658
Construcción	1410
Minas y canteras	1230
Transporte, almacenamiento y comunicaciones	1145
Comercio	677
Industria manufacturera	643
Agricultura, ganadería, caza y silvicultura	478
Servicios comunitarios, sociales y personales	297
Administración pública y defensa	232
Servicios sociales y de salud	239
Eléctrico, gas y agua	94
Educación	64
Financiero	59
Hoteles y restaurantes	55
Servicio doméstico	29
Pesca	9
Órganos extraterritoriales	0

Fuente: Federación de Aseguradores Colombianos (2023).

De acuerdo con lo evidenciado en la tabla 3, el presente estudio se centrará en analizar los sectores económicos inmobiliario, construcción, minas y canteras (en adelante sector minas), transporte, almacenamiento y comunicaciones (en adelante sector transporte), comercio y agricultura, ganadería, caza y silvicultura (en adelante sector agricultura), los primeros cinco se eligen por ser los de mayor

cantidad de muertes materializadas durante los años 2009 al 2022 y el de agricultura se eligió por ser de especial interés para el gobierno nacional actual.

6.2. EXPLORACIÓN DE DATOS

6.2.1. Pruebas de normalidad

Los grupos de datos utilizados en el estudio corresponden a matrices de 8 variables por una cantidad de registros variables, puesto que esta depende de la cantidad de niveles de riesgos presenten en los diferentes sectores económicos, en donde, para los sectores inmobiliario, transporte, comercio y agricultura se presentaban todas las clases de riesgo (I, II, III, IV y V), el sector construcción no presentaba todas las clases de riesgo (I, II, III, IV y V), el sector minas no presentaba las clases de riesgo I y II. En la tabla 4 se pueden evidenciar las variables de los datos.

Tabla 4. Variables estudiadas

Número Empresas	Total Trabajadores	Número de AT	Muertes por AT	Pensiones de invalidez por AT	Indemnizaciones pagadas por AT

En el proceso de exploración de datos se realizaron diferentes mediciones que permitieron identificar los comportamientos individuales de las variables, relaciones entre ellas, dependencias, tendencias y otros. Inicialmente se realizó la prueba de normalidad para todos los sectores económicos mencionados, obteniendo los siguientes resultados en el test Shapiro-Wilk (tabla 5) para la hipótesis de que los datos son normales:

Tabla 5. P-valores prueba Shapiro-Wilk por variable y sector económico

Sector Económico	Número Empresas	Total Trabajadores	Número accidentes laborales	Número de Muertes por accidentes	Número de pensiones por invalidez	Número de indemnizaciones pagadas
Inmobiliario	0.00	2e-4	0.0036	5e-4	0.00	1e-3
Construcción	0.00	0.00	0.00	9e-3	0.00	0.0030
Minas	1e-3	0.00	0.00	0.00	0.00	0.00
Transporte	0.00	0.00	1e-3	1e-3	0.00	0.00
Comercio	0.00	0.00	0.038	0.00	0.00	0.00
Agricultura	0.00	0.00	0.00	0.00	0.00	0.00

FUENTE: Construcción propia

Como se evidencia en la tabla 5, todos los p-valores de las variables en los diferentes sectores económicos son menores al nivel de significancia 0.05, por lo cual, se determina que ninguna de las series de datos se distribuye normalmente.

6.2.2. Análisis Univariado

Posteriormente, se realizó la revisión individual de las variables objeto de estudio en el periodo de análisis (2009 - 2022), pudiendo de esta forma realizar el análisis del comportamiento histórico de las variables y posibles tendencias que se puedan presentar en dicho comportamiento.

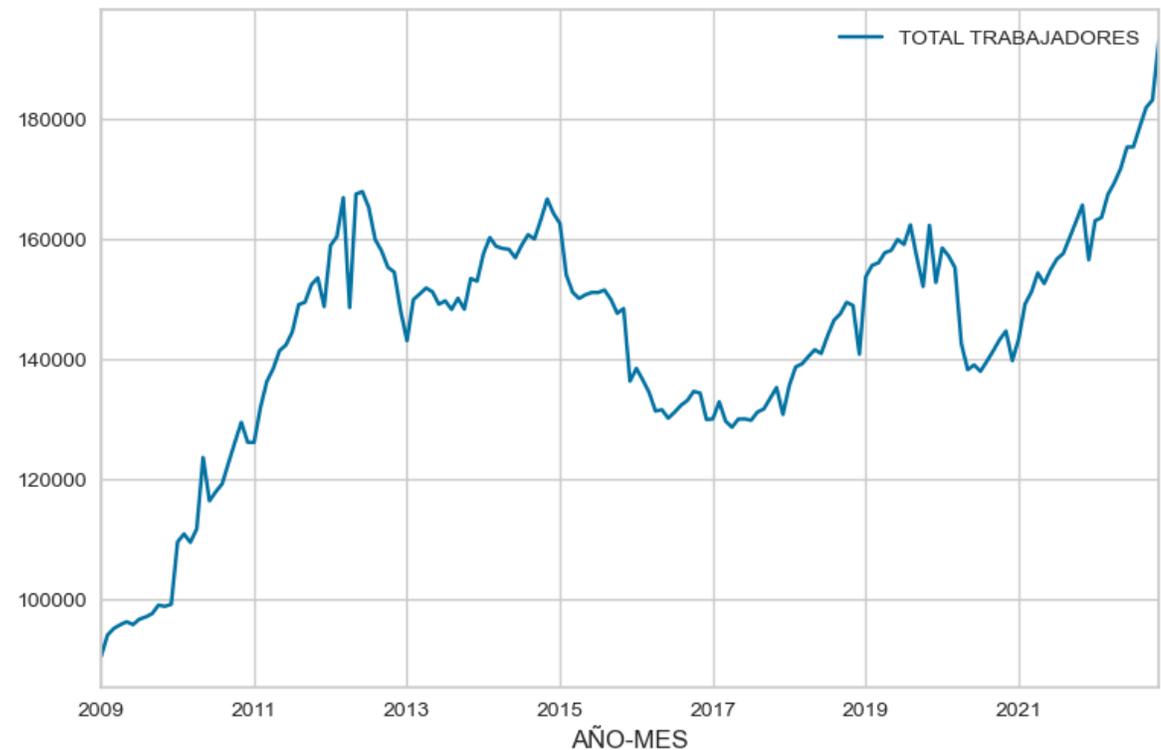


Figura 2. Total de Trabajadores - Sector Minas – Clase de riesgo V

En la figura 2 se evidencia el comportamiento de la variable total trabajadores para el sector minas en la clase de riesgo V, en la cual se puede observar un crecimiento durante el periodo de análisis en el que la población de trabajadores aumentó un 60% entre enero de 2009 y diciembre de 2022.

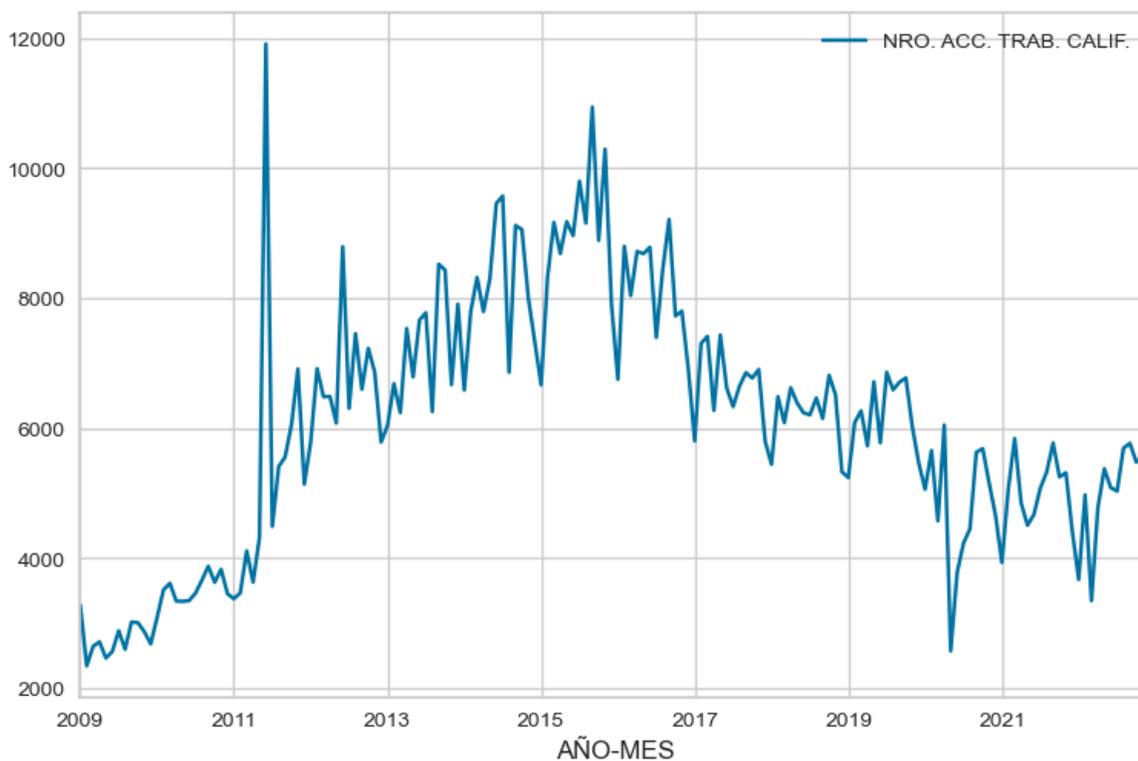


Figura 3. Número de accidentes de trabajo - Sector construcción - Clase de riesgo V

En la figura 3 se observa el comportamiento de la variable número de accidentes de trabajo calificados para el sector económico construcción en la clase de riesgo V, como se evidencia, esta presenta una variación en su tendencia, la cual venía siendo creciente hasta finales del 2015, momento en el cual empezó a descender gradualmente, sin embargo, el valor final de la serie sigue siendo mayor (aumentó un 37%) que el valor inicial de la misma. De igual manera, se evidencia un leve pico de accidentalidad en junio de 2011, presentando su valor más alto con un total de 11913 accidentes para este mes.

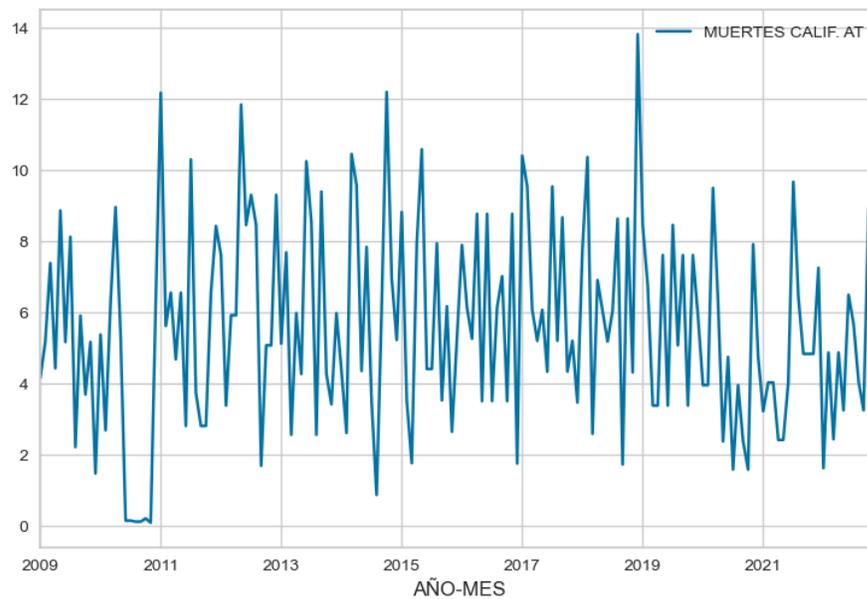


Figura 4. Número de enfermedades laborales calificadas - Sector Inmobiliario

En la figura 4 se observa el comportamiento de la variable número de muertes por AT para el sector transporte en su clase de riesgo IV, esta presentó una variación constante entre aumento y disminución de la cantidad de muertes, presentando su pico máximo de 14 muertes en un mes y presentando un aumento del 25% entre el valor inicial y el final de la serie.

6.2.3. Análisis Multivariado

El desarrollo del análisis multivariado se centró principalmente en dos aspectos, el primero consistente en la revisión general de las variables y el grado de correlación que presentaban entre ellas, con esto, se buscó identificar cuáles tendrían un mayor nivel de relación y cuáles no presentaban ninguno. De igual manera, a partir de la exploración general, se realizó una exploración específica entre las variables que presentaron mayores relaciones y entender de mejor forma su interacción.

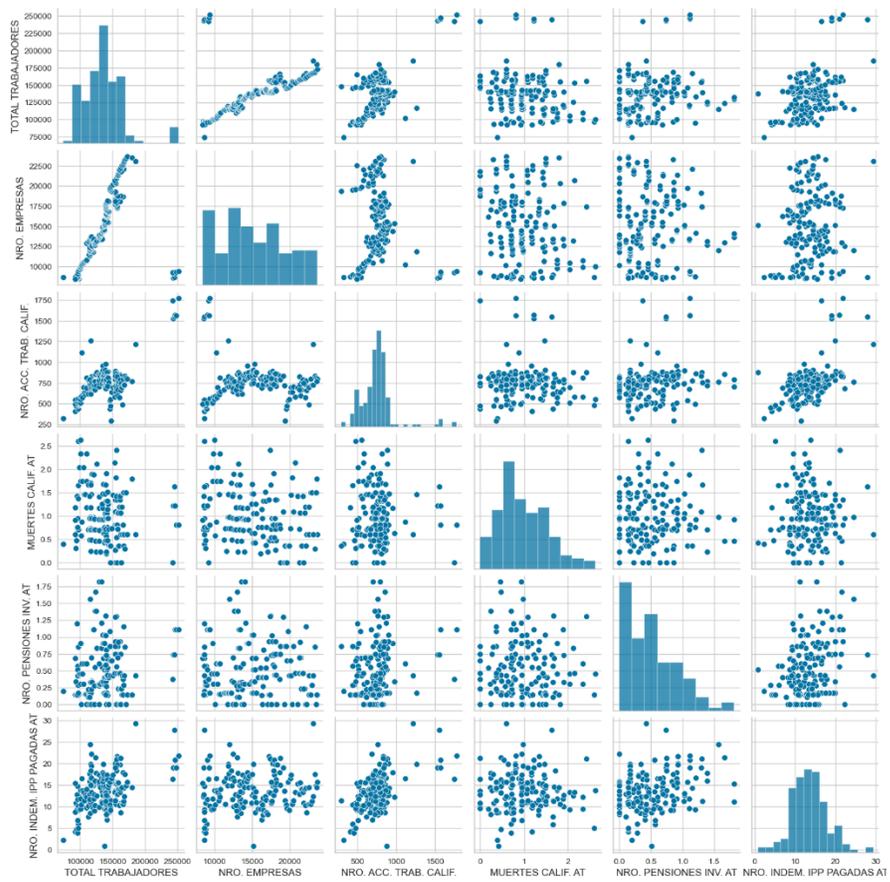


Figura 5. Interacción de variables objetivo - sector Inmobiliario

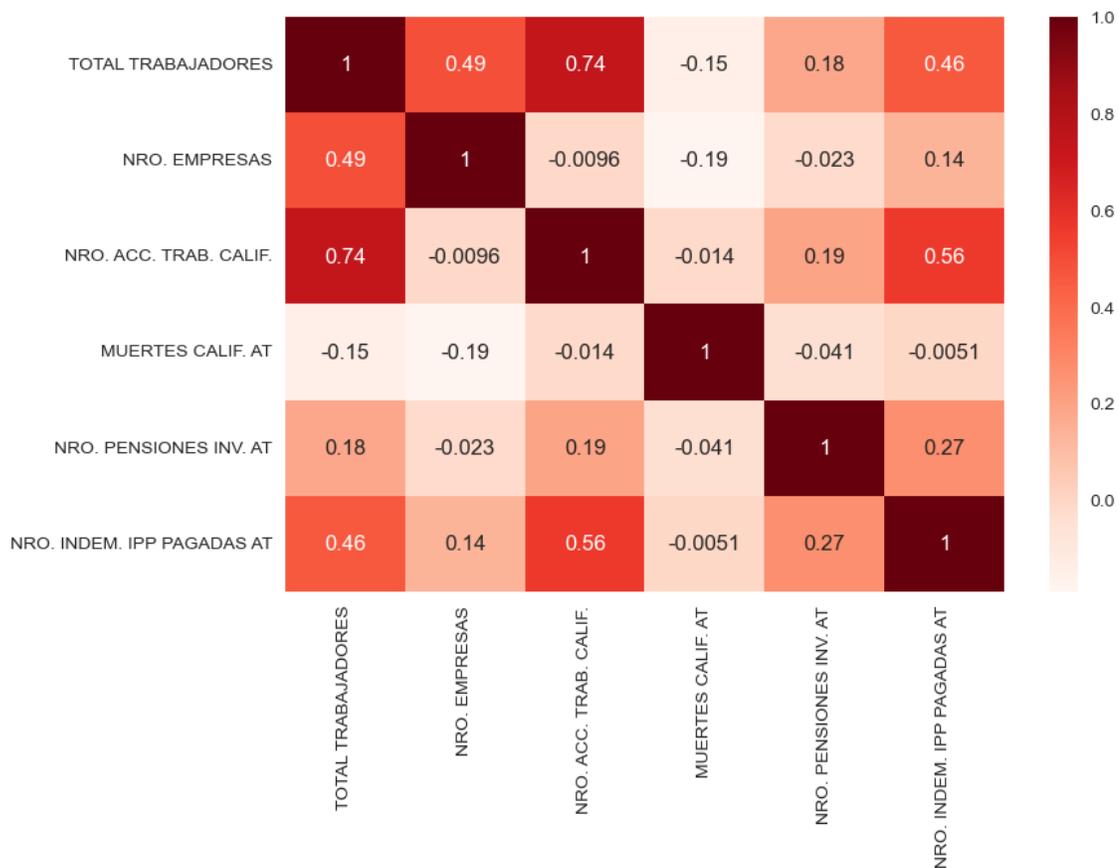


Figura 6. Mapa de calor, relacionamiento de variables - Sector Inmobiliario

En las figuras 5 y 6 se evidencia el análisis multivariado que se realizó, el mismo se aplicó para cada uno de los sectores económicos en todas sus clases de riesgo, logrando identificar las correlaciones entre variables y las tendencias que presentaban las mismas. Estos resultados se pueden evidenciar en el anexo 4.

6.3. TRANSFORMACIÓN DE LOS DATOS

Para la definición de los sets de datos de cada sector económico se realizó una proyección de composición para las diferentes clases de riesgo, esto debido a que los registros mensuales estaban consolidados unificando todas las clases de riesgo y no las discriminaba de forma individual, mientras que dicha segregación por clase de riesgo sólo se encontraba disponible en el reporte anual, por lo cual, se tomó la composición por clases del reporte anual y se le aplicó a los registros mensuales, para que de esta forma se pudiera obtener el registro mensual por clase de riesgo y poder hacer el análisis objetivo para las combinaciones sector económico – clase

de riesgo – variables objetivo.

Posteriormente, se procedió a realizar un análisis de datos atípicos para las variables objeto de interés, para esto, se desarrolló un Generalized Extreme Studentized Deviate (ESD) test con el cual se realizó el análisis de los datos de cada variable con el fin de identificar si estas presentaban uno o más valores atípicos, y posteriormente, una vez identificados estos valores, se procedió con la imputación de los mismos mediante el uso de la mediana (tabla 6).

Tabla 6. Valores atípicos por sectores económicos y variables

Sector económico	Variable	Cantidad de valores atípicos imputados por variable
Inmobiliario	Total trabajadores	0
	Número de empresas	0
	Número accidentes laborales	0
	Muertes por accidentes laborales	0
	Número pensiones por invalidez	0
	Número de indemnizaciones pagadas	0
Construcción	Total trabajadores	0
	Número de empresas	0
	Número accidentes laborales	2
	Muertes por accidentes laborales	0
	Número pensiones por invalidez	0
	Número de indemnizaciones pagadas	0
Minas y Canteras	Total trabajadores	1
	Número de empresas	0
	Número accidentes laborales	1
	Muertes por accidentes laborales	6
	Número pensiones por invalidez	0
	Número de indemnizaciones pagadas	0
Transporte	Total trabajadores	0
	Número de empresas	0
	Número accidentes laborales	1
	Muertes por accidentes laborales	0
	Número pensiones por invalidez	0
	Número de indemnizaciones pagadas	0
Comercio	Total trabajadores	0
	Número de empresas	0
	Número accidentes laborales	0

	Muertes por accidentes laborales	0
	Número pensiones por invalidez	0
	Número de indemnizaciones pagadas	0
Agricultura	Total trabajadores	1
	Número de empresas	1
	Número accidentes laborales	1
	Muertes por accidentes laborales	1
	Número pensiones por invalidez	0
	Número de indemnizaciones pagadas	0

Fuente: Construcción propia

6.4. MODELAMIENTO

Una vez realizada la exploración de los datos y tomando como referencia la revisión bibliográfica, se definió la implementación de un modelo de regresión lineal combinado con PCA, en el cual, el PCA permitirá definir los componentes más relevantes del set de datos para explicar la varianza de la variable respuesta y posteriormente se evaluará el nivel de precisión del modelo de regresión lineal mediante una implementación múltiple, tomando como variables independientes los componentes generados y evaluando cada una de las variables respuesta.

El nivel de precisión del modelo se evaluará mediante el parámetro R^2 . El modelo se considerará aplicable al proceso de simulación cuando el mismo tenga una precisión superior a 0.4, a partir de la cual se realizarán las posteriores predicciones de acuerdo con los escenarios formulados.

6.4.1. Proyección de modelos

A continuación, se describirá la aplicación del modelo para el sector minas en su clase de riesgo V, esta misma aplicación se realizó para todos los sectores económicos en cada una de sus clases de riesgo.

El modelo se aplicó sobre el set de datos con los valores atípicos ajustados. El análisis se realizó de forma individual por clase de riesgo, teniendo en cuenta que las variaciones entre clases como la I y la V podrían ser muy amplias y afectarían la precisión del modelo. Por lo anterior, se realizó el filtro del set de datos en cada una de las 5 clases y se aplicó el PCA, dado esto, se obtuvieron un total de 168 registros (12 meses por 14 años) para cada una de las variables de análisis en cada una de las clases.

En la aplicación del PCA se realizó un preprocesamiento del set de datos mediante el módulo StandardScaler de la librería sklearn, permitiendo ajustar los componentes a una media de 0 y desviación estándar de 1, facilitando con esto el proceso de entrenamiento del modelo y previniendo que se presenten sesgos en función de alguna variable o componente que tenga mayor nivel de influencia en la variable respuesta.

Posteriormente se procedió a determinar el número de componentes más adecuado para el entrenamiento del modelo, el criterio para esta selección fue realizar el entrenamiento con el número de componentes necesario para explicar al menos el 80% de la varianza de la variable respuesta, esto se determinó mediante el módulo `pca.explained_variance_ratio` de la misma librería.

A continuación, se realizó la medición del aporte que presentó cada variable a la creación de los diferentes componentes que explican la varianza de la variable respuesta. En la tabla 7 se evidencia que para el sector económico minas en su clase de riesgo V, el número adecuado de componentes para explicar al menos el 80% de la varianza de la variable respuesta número de AT fue 3. Así mismo, se evidencia la contribución que realizaron las diferentes variables a la construcción de cada uno de los componentes, teniendo presente que los valores más altos en las variables representan una alta correlación entre dicha variable y la variable respuesta.

Tabla 7. Componentes para la variable respuesta número de AT en el sector minas - clase de riesgo V

Componente	Número empresas	Total Trabajadores	Número muertes AT	Número pensiones AT	Número de indemnizaciones AT
1	-0.520	-0.558	-0.349	-0.300	-0.453
2	0.113	0.190	-0.358	-0.793	0.439
3	-0.109	-0.011	0.841	-0.497	-0.180

FUENTE: Construcción propia

En el anexo 1 se evidencian las tablas de composición de los componentes para cada uno de los modelos por sector económico, clase de riesgo y variable.

6.5. EVALUACIÓN

Finalmente, se realizó la formulación de un set de entrenamiento y un set de pruebas del modelo, conformados por el 80% y 20% de los datos respectivamente. Con el 80% de los datos se realizó el entrenamiento del modelo de regresión lineal y obtener predicciones que posteriormente fueron evaluadas con el set de pruebas, obteniendo de esta forma la precisión del modelo mediante el R^2 .

6.5.1. Resultados generales

A continuación, se evidencian los resultados obtenidos de la aplicación del modelo en las diferentes variables objeto de análisis para cada sector económico (tabla 8).

Tabla 8. Resultados de los modelos por variable y sector económico

Sector económico	Clase de riesgo	Número de accidentes de trabajo		Número de muertes por accidente de trabajo		Número de pensiones por invalidez		Número de indemnizaciones pagadas	
		Cantidad Componentes	R ²	Cantidad Componentes	R ²	Cantidad Componentes	R ²	Cantidad Componentes	R ²
Inmobiliario	1	2	0.690	2	0.435	3	0.469	3	0.413
	2	4	0.479	3	0.224	3	0.218	4	0.542
	3	3	0.773	3	0.394	3	0.54	3	0.59
	4	3	0.48	3	0.17	3	0.562	3	0.777
	5	3	0.68	2	0.04	3	0.501	3	0.534
Construcción	2	4	0.473	3	0.06	3	0.096	3	0.049
	3	3	0.815	3	0.412	3	0.237	3	0.562
	4	3	0.554	3	0.034	3	0.135	3	0.341
	5	3	0.603	3	0.148	3	0.105	3	0.321
Minas	3	3	0.36	4	0	4	0	3	0.07
	4	3	0.921	3	0.33	2	0.201	2	0.066
	5	3	0.851	2	0.047	3	0.141	3	0.246
Transporte	1	4	0.203	3	0.002	4	0.092	3	0.246
	2	3	0.803	3	0.089	3	0.036	3	0.547
	3	3	0.649	3	0.034	3	0.005	3	0.34
	4	3	0.795	2	0.012	2	0.223	3	0.656
	5	2	0.946	2	0.945	2	0.89	2	0.897
Comercio	1	3	0.025	3	0.292	3	0.086	3	0.004
	2	4	0.117	4	0.074	3	0.024	4	0.204
	3	4	0.494	3	0.09	3	0.088	4	0.42
	4	3	0.831	2	0.129	2	0.02	3	0.155
	5	3	0.716	3	0.402	3	0.163	3	0.79
Agricultura	1	3	0.15	4	0.1	3	0.012	3	0.036
	2	4	0.37	3	0.036	3	0.035	3	0.303

	3	3	0.529	2	0.125	3	0.746	3	0.642
	4	4	0.431	3	0.138	3	0.11	3	0.357
	5	3	0.865	3	0.217	3	0.469	3	0.581

FUENTE: Construcción propia

6.6. APLICACIÓN DE LOS MODELOS

Una vez desarrollados los modelos y obtenidos los resultados de componentes y precisión de los mismos, se realizó la selección de aquellos cuya precisión fue superior a 0.4 para ser aplicados en los escenarios de simulación. Después de seleccionar las combinaciones de variable-clase de riesgo-sector económico que cumplen con el criterio, se procedió a la aplicación del modelo en los escenarios definidos en el numeral 5.6.

Dada la configuración del set de datos para cada sector económico, se contaba con un total de 168 registros para cada clase en cada variable (12 meses por 14 años). Para la aplicación del modelo, se definieron como variables exógenas las de número de empresas y total de trabajadores, por lo cual, se procedió a realizar formulación aleatoria mediante una función triangular entre los escenarios pesimista, más probable y optimista. A continuación, se muestran las ecuaciones aplicadas para realizar la predicción de los nuevos valores en cada variable:

$$C_{n+1} = A_{n+1} + B_{n+1} + D_n + E_n + F_n$$

$$C_{n+24} = A_{n+24} + B_{n+24} + D_{n+23} + E_{n+23} + F_{n+23}$$

$$D_{n+1} = A_{n+1} + B_{n+1} + C_{n+1} + E_n + F_n$$

$$D_{n+24} = A_{n+24} + B_{n+24} + C_{n+24} + E_{n+23} + F_{n+23}$$

$$E_{n+1} = A_{n+1} + B_{n+1} + C_{n+1} + D_{n+1} + F_n$$

$$E_{n+24} = A_{n+24} + B_{n+24} + C_{n+24} + D_{n+24} + F_{n+23}$$

$$F_{n+1} = A_{n+1} + B_{n+1} + C_{n+1} + D_{n+1} + E_{n+1}$$

$$F_{n+24} = A_{n+24} + B_{n+24} + C_{n+24} + D_{n+24} + E_{n+24}$$

- Donde A y B son las variables exógenas
- Donde C, D, E y F son las variables respuestas analizadas (número de AT, muertes por AT, pensiones por invalidez, incapacidad permanente parcial).
- Donde n es el último valor real del set datos.
- Donde n+1 es el primer valor predicho para cada una de las variables.
- Donde n+24 representa el último valor predicho para cada una de las variables objetivo (que presentaría el valor de cada variables en el mes de diciembre de 2024).

Se evidencia que se parte inicialmente con los 168 registros reales en cada variable, después, se formula de forma aleatoria mediante función triangular los registros 169 para las variables exógenas número de empresas y total de trabajadores. A partir de dicha formulación aleatoria, se aplica el modelo tomando los registros 169 de las variables exógenas y los registros 168 de las otras variables (diferentes a la variable objetivo 1) y se aplica el modelo para predecir el registro 169 de la variable objetivo 1, posteriormente, se vuelven a tomar los registros 169 de las variables exógenas, el registro 169 de la variable objetivo 1 (predicho en el paso anterior) y los registros 168 de las otras variables (diferentes a la variable objetivo 2) y se aplica el modelo para predecir el registro 169 de la variable objetivo 2, esto mismo se realiza para las variables objetivo 3 y 4.

Posteriormente, se formulan nuevamente de forma aleatoria mediante función triangular los registros 170 de las variables exógenas y se repite el proceso para las 4 variables objetivo (número de AT, muertes AT, pensiones invalidez AT, indemnizaciones pagadas AT), hasta lograr la predicción de los 24 registros siguientes al 168 (completando un total de 192 registros para cada variables) los cuales representarían el comportamiento de las variables durante los 24 meses siguientes al periodo de análisis (enero de 2023 a diciembre de 2024).

Una vez obtenidos los 192 registros, se itera y se repite el proceso realizando 100 replicaciones, para posteriormente calcular el promedio del comportamiento de cada variable durante los siguientes 24 meses al periodo de análisis y calcular su intervalo de confianza con un nivel de confianza del 95%. A continuación, se observa la aplicación del modelo en la simulación del comportamiento para las variables número de AT y número indemnizaciones pagadas AT en los sectores minas y transporte respectivamente para las clases de riesgo V y IV respectivamente (figuras 7 y 8).

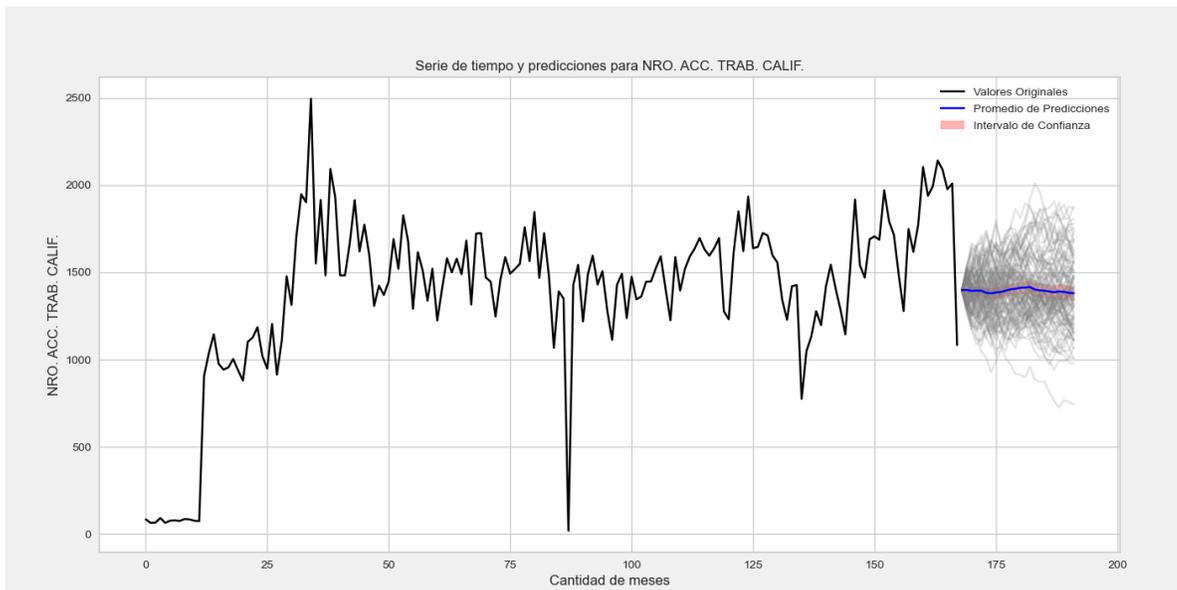


Figura 7. Simulación variable Número AT en sector minas - clase de riesgo V

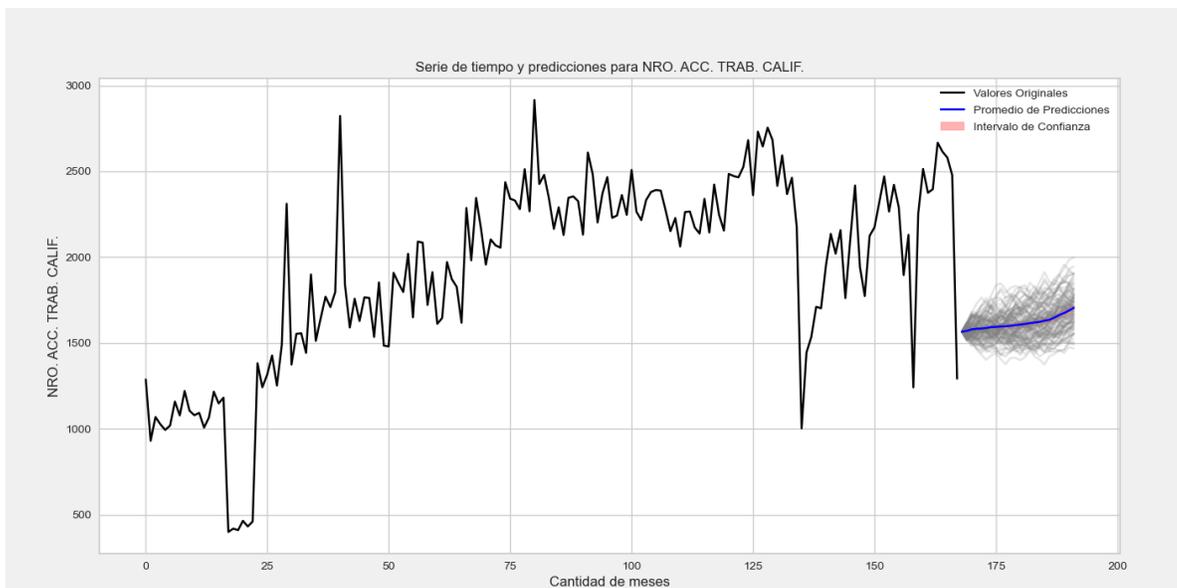


Figura 8. Simulación variable Número AT en sector transporte - clase de riesgo IV

En el anexo 2 se encuentran los resultados de las simulaciones ejecutadas para aquellas combinaciones sector económico – clase de riesgo – variables en las cuales el modelo presentó una precisión superior a 0.4, aquellas que se marcan como “NA” corresponden a las combinaciones en las cuales el modelo tuvo una precisión inferior al criterio por lo cual no se ejecutaron las simulaciones. Igualmente, en el anexo 3 se encuentran las gráficas que muestran el

comportamiento de las variables en los diferentes sectores económicos y clases de riesgo para los cuales se aplicó el modelo y se realizó la simulación de los escenarios.

6.7. ANÁLISIS DE RESULTADOS

El modelo desarrollado presentó diferentes niveles de precisión en función del comportamiento de las diferentes variables y sus comportamientos, aunque en todos los casos el PCA lograba explicar un valor superior al 80% de la varianza en la variable respuesta, el modelo de regresión lineal múltiple no siempre lograba obtener un buen nivel de precisión con la variable respuesta, especialmente para aquellas variables que, por sus características, no presentaban muchas ocurrencias como lo eran el número de muertes AT y el número de pensiones de invalidez por AT.

La simulación de los escenarios en los diferentes sectores permitió identificar los comportamientos de las variables objeto de interés en los diferentes sectores económicos y clases, sin embargo, en muchos casos esas variables no presentaron una clara relación con las variables exógenas generadas, por lo que su comportamiento no estaba directamente vinculado con la dinámica poblacional del sector económico analizado, como por ejemplo el sector inmobiliario en su clase de riesgo I, la cual presentó crecimientos y decrecimientos en los 24 meses de predicción, pero la variable número de accidentes de trabajo no siempre presentó el mismo comportamiento (aumento o disminución) que las variables exógenas.

Sin embargo, para el mismo sector inmobiliario en su clase de riesgo I, se presentó una relación proporcional entre la variable muertes por AT y el número de AT ocurridos, en donde la primera presentó el mismo comportamiento que la segunda en función de los aumentos o disminuciones evidenciados. Teniendo en cuenta que el número de muertes por AT presentada para este sector en esta clase es bajo (en promedio 1 mensual) mientras que el del número de accidentes es más representativo (aproximadamente 2200 accidentes mensuales), se podría interpretar que una buena gestión de riesgos laborales que permita disminuir en determinada cantidad los accidentes de trabajo podría representar alcanzar el 0 en

el número de fatalidades para este sector económico en esta clase de riesgo de forma más frecuente.

Para la variable muertes por AT, el modelo sólo presentó una precisión superior a 0.4 en el 18.5% de las clases sobre las que se aplicó (5 de 27), esto debido principalmente a que la ocurrencia de este acontecimiento era muy poco frecuente en todos los sectores económicos en relación con la población de los mismos y la ocurrencia del número de AT, los cuales eran significativamente mayores a la presencia de fatalidades productos de los AT materializados en cada sector económico.

Un comportamiento similar se presentó para la variable número de pensiones de invalidez por AT, para la cual el modelo sólo presentó una precisión superior a 0.4 en el 26% de los casos en que se aplicó (7 de 27 clases analizadas), debido a una situación similar relacionada con la baja ocurrencia de esta situación en relación con la población de empresas, colaboradores y ocurrencia de AT.

El sector económico inmobiliario en su clase de riesgo III presenta un comportamiento desfavorable para la variable indemnizaciones permanentes parciales ya que en el año 2020 se presentaba en promedio una indemnización por cada 103 accidentes, para el año 2022 esta cifra se situó en 1 indemnización por cada 71 accidentes y para el año 2024 se espera que se encuentre en 1 indemnización por cada 68 accidentes presentados, por lo que este deberá ser un foco de especial interés en este sector económico y clase de riesgo.

Para el sector económico construcción, en su clase de riesgo III, se espera que el comportamiento del número de AT sea bastante estable, con variaciones +5% a -5%, no se espera que pueda presentar disminuciones significativas en función de una variación poblacional leve. Por lo cual, el posible impacto negativo que pueda tener este sector económico en los próximos meses debería representar una oportunidad para que las empresas puedan optimizar su gestión de riesgos laborales y disminuir significativamente su accidentalidad laboral. En el campo de las muertes por AT no se espera que se materialicen muertes para esta clase de riesgo en este sector económico, por lo cual su foco debería estar en los accidentes

leves o graves para de esta forma conservar sus bajos registros de mortalidad.

Para el sector económico Minas, en su clase de riesgo V, no se esperan grandes variaciones en sus registros de accidentes de trabajo, la variación esperada es del 1% en forma creciente o decreciente sin que esta tenga una relación directa con la variación poblacional. Este comportamiento estable presenta una relación con los rigurosos estándares de seguridad y salud en el trabajo que se manejan en este sector económico.

La clase de riesgo IV del sector transporte presenta un comportamiento favorable en el número de indemnizaciones permanentes parciales, ya que estas pasaron de presentarse 1 por cada 37 accidentes de trabajo en 2020 a 1 por cada 47 en accidentes en 2022 y se espera que para el año 2024 se presente 1 indemnización por cada 64 accidentes de trabajo, alcanzando un aumento del 97% en el número de accidentes necesarios para que se materialice un accidente de trabajo que conlleve una pérdida de capacidad laboral superior al 5% e inferior al 50% en un trabajador de este sector.

El sector transporte, en su clase de riesgo V, es el sector que mayor crecimiento presenta en sus registros de accidentes de trabajo para el periodo simulado, aumentado un 13% su promedio de accidentes mientras que su población presentaría un decrecimiento de únicamente el 2%, el aumento de estos accidentes de trabajo se vería reflejado principalmente en accidentes leves ya que no se evidencian variaciones significativas el número de muertes por AT, el número de pensiones por invalidez (el promedio continúa en un valor cercano a 0 para los 24 meses simulados en las dos variables) y tampoco en las indemnizaciones pagadas por AT, las cuales mantienen su promedio alrededor de las 4 mensuales.

Es importante tener en cuenta el comportamiento del número de muertes por AT en la clase de riesgo I del sector inmobiliario, ya que estas venían presentando un comportamiento muy favorable al pasar de 1 muerte por cada 1493 accidentes en 2021 a 1 muerte por cada 2481 accidentes en promedio del 2022. Sin embargo, se espera que para el los años 2023 y 2024 se materialice una muerte por cada 1601 y 1667 accidentes respectivamente.

Para el sector económico comercio, en su clase de riesgo V, la variable número de accidentes presentó un comportamiento inversamente proporcional a la del total de trabajadores, en el cual, la población del sector aumentó en torno al 5% durante los 24 meses simulados, mientras que el número de AT presentó una disminución el 7%. Este comportamiento podría estar asociado al fenómeno de formalización laboral, en donde se espera un crecimiento poblacional que correspondería a personas que ya laboran actualmente en dicho sector, pero en el ámbito informal, y su formalización no representaría necesariamente un aumento en los registros de accidentalidad laboral.

Finalmente, para el sector agricultura, en su clase de riesgo III, se espera un importante crecimiento poblacional asociado principalmente a las políticas de formalización laboral del gobierno nacional, y teniendo en cuenta que este a su vez tiene el índice de informalidad laboral más alto (85% de acuerdo con el informe del DANE), esperando un crecimiento en número de trabajadores de alrededor del 15% y un 5% en número de empresas. Sin embargo, se espera que el comportamiento de la accidentalidad laboral tenga un comportamiento inverso al de las variables poblacionales, en donde el número de AT presentaría una disminución del 10; disminución que no se reflejaría en el número de pensiones por AT o indemnizaciones pagadas por AT, las cuales mantendrían sus promedios en alrededor de 1% y 10% respectivamente.

7. CONCLUSIONES

- El comportamiento del mercado laboral en los diferentes sectores económicos analizados es muy variable entre sus diferentes clases de riesgo y en función de las políticas públicas del actual gobierno nacional. El estudio permitió desarrollar diferentes escenarios que simulaban el comportamiento individual de cada sector y el comportamiento de la accidentalidad laboral y sus consecuencias en función de dicha variabilidad.
- Los mayores crecimientos poblacionales que se presentarán de acuerdo con las replicaciones realizadas, corresponden igualmente a los sectores con mayores índices de informalidad laboral, por lo que se considera que las políticas de formalización laboral del actual gobierno jugarán un papel fundamental en la dinámica laboral de los años 2023 y 2024.
- Se realizó la consolidación de un total de 6 sets de datos (1 por cada sector económico de estudio), los cuales estaban conformados por una línea temporal (cada uno de los meses desde enero de 2009 hasta diciembre de 2022), dos variables poblacionales y 4 variables respuesta que correspondían al panorama de la accidentalidad laboral y sus respectivas consecuencias como lo eran las muertes, pensiones por invalidez e indemnizaciones permanentes parciales.
- El modelo desarrollado corresponde a un PCA con número variable de componentes en función de la explicación de la varianza para cada variable respuesta, posteriormente aplicar una regresión lineal múltiple con los componentes principales y las diferentes variables respuesta. Permitiendo predecir el comportamiento de dichas variables en los 24 meses de los años 2023 y 2024 con diferentes escenarios de dinámica poblacional.
- El modelo de regresión lineal precedido por un PCA permitió obtener ajustes significativos (alrededor del 90%) en los sectores minas, transporte y agricultura en sus clases de riesgo V para la variable número de AT, por lo que esta presenta un comportamiento lineal que se ve impactado por los

cambios en las variables poblacionales (número de empresas y total de trabajadores).

- Se evidenció que para la variable número de AT, los componentes se conformaron principalmente por las variables número de empresas y total trabajadores, evidenciando que dicha variable guarda relación con las poblaciones en cada sector económico y sus respectivas clases de riesgo.
- Para las variables muertes por AT, pensiones por invalidez e indemnizaciones permanente parciales, se evidenció que sus componentes fueron conformados principalmente por la variable número de accidentes en todos los sectores económicos y clases de riesgo, lo cual está completamente alineado con la lógica de ocurrencia de los accidentes y la posterior consecuencia que puede materializarse en alguna de las variables mencionadas.
- La formulación de sets de entrenamiento y validación permitió realizar la correcta calibración y validación del modelo desarrollado, lo cual, junto con el análisis individual de las clases de riesgo por sector económico permitió obtener altos niveles de precisión para diferentes sectores económicos, puesto que el comportamiento de las diferentes clases dentro de un mismo sector económico era muy variable en cuanto a población de trabajadores y ocurrencia de accidentes y las diferentes consecuencias.
- El desarrollo de escenarios individuales por sector económico y ajustados con las tendencias actuales del mercado laboral para cada uno, permitió simular comportamientos más ajustados a la realidad individual de cada sector económico y de la misma manera obtener predicciones más precisas del comportamiento de las variables respuesta en cada sector.
- El ajuste individual de los sets de datos para cada sector y las predicciones realizadas se podrían utilizar como insumo para el análisis de tendencias en cada uno, determinación de focos de interés en materia de riesgos laborales desde el punto de vista de las empresas, las ARL, o las autoridades

nacionales en materia de trabajo y riesgos laborales.

- De acuerdo con los resultados obtenidos, no se esperan cambios abruptos en el comportamiento de la accidentalidad laboral y sus consecuencias asociadas para los años 2023 y 2024 en ninguno de los sectores económicos analizados, puesto que ninguno presentó un crecimiento o decrecimiento superior al 20% en ninguna de las variables objeto de interés.
- Se debe prestar especial atención a los sectores económicos minas en su clase de riesgo V, construcción en su clase de riesgo V y transporte en su clase de riesgo IV, junto con sus riesgos asociados, ya que presentaron las estadísticas más altas de fatalidades para los periodos analizados.
- Los resultados permiten identificar que el mayor crecimiento en los accidentes de trabajo estaría en las categorías de leves teniendo en cuenta que para ninguna clase de ningún sector económico se evidenciaron variaciones significativas en las variables de muertes, pensiones e indemnizaciones.

8. TRABAJO FUTURO

El presente estudio permitió identificar nuevas líneas sobre las cuales podrían desarrollarse futuros estudios complementarios. Dentro de las líneas identificadas se encuentran las siguientes:

- Estudios complementarios que prueben la implementación de otros modelos matemáticos diferentes a la regresión lineal múltiple, buscando identificar mejores ajustes con las variables objeto de interés en los sectores económicos analizados y lograr predicciones más precisas que permitan realizar nuevas evaluaciones sobre el comportamiento de la accidentalidad de acuerdo con la dinámica del mercado laboral.
- Estudios complementarios que incluyan nuevas variables que permitan explicar de forma complementaria el comportamiento de la accidentalidad laboral en la dinámica del mercado laboral. Podría considerarse la inclusión de variables cualitativas que permitan explicar con un mayor nivel de profundidad los conceptos asociados a los comportamientos de las variables cuantitativas.
- Se propone el desarrollo de un estudio sobre algún sector específico de interés, integrando dinámicas específicas del sector como los peligros asociados a las actividades desarrolladas por los trabajadores con el fin de evaluar la incidencia de estas en los índices generales de accidentalidad laboral.

9. LISTA DE REFERENCIAS

- Álvarez, S., Palencia, F., & Riaño, M. (2019). Comportamiento de la accidentalidad y enfermedad laboral en Colombia 1994 – 2016. *Revista de la Asociación Española de Especialistas en Medicina del Trabajo*, 28(1), 10-19. https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1132-62552019000100002&lng=es&nrm=iso&tlng=es
- Barrios, M., Jimeno, M., Villalba, P., y Navarro, E. (2019). Novel data mining methodology for healthcare applied to a new model to diagnose metabolic syndrome without a blood test. *Diagnostics*, 9(4). <https://doi.org/10.3390/diagnostics9040192>
- Dascălu, C. G., y Cozma, C. D. (2017). The principal components analysis-method to reduce the collinearity in multiple linear regression model; application in medical studies. <https://www.researchgate.net/publication/268265047>
- Delgado, Y., Arroyo, J., Monroy, M., y Zea, C. (2022). Generación de un modelo predictivo de la fuerza de agarre para trabajadores del sector panelero. *Ingeniería y Desarrollo en la Nueva Era*, 460-469. <https://doi.org/10.5281/zenodo.7381846>
- Departamento Administrativo Nacional de Estadística. (2023). Boletín Técnico Gran Encuesta Integrada de Hogares (GEIH) Enero-marzo 2023 - Ocupación informal. https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/Nueva_medicion_informalidad.pdf
- Federación de Aseguradores Colombianos (2023). Riesgos Laborales – Estadísticas del ramo. Federación de Aseguradores Colombianos – FASECOLDA. Recuperado el día 25 de marzo de 2023 de <https://sistemas.fasecolda.com/rldatos/>.
- Gobierno de Colombia (2023). Plan Nacional de Desarrollo 2022 - 2026. Bogotá D.C., Colombia. Departamento nacional de Planeación – DNP. Recuperado de <https://www.dnp.gov.co/plan-nacional-desarrollo/pnd-2022-2026>.
- Kutner, M. H., Nachtsheim, C., Neter, J., y Li, W. (2004). *Applied linear statistical models*.
- Lara, A. (2021). Modelo de predicción de riesgos psicosociales en el transporte urbano de pasajeros usando técnicas de Inteligencia Artificial. Universidad Técnica de Ambato.
- Ley 776 de 2002. Por la cual se dictan normas sobre la organización, administración y prestaciones del Sistema General de Riesgos Profesionales. 17 de diciembre de 2002. D.O. No. 45037.
- Ley 1562 de 2012. Por la cual se modifica el Sistema de Riesgos Laborales y se dictan otras disposiciones en materia de Salud Ocupacional. 11 de julio de 2012. D.O. No. 48488.
- Liu, Z., Cheng, S., y Liu, P. (2022). Prediction model of BOF end-point P and O contents based on PCA-GA-BP neural network. *High Temperature Materials and Processes*, 41(1), 505-513. <https://doi.org/10.1515/htmp-2022-0050>
- Ministerio del Trabajo. (2022). Plan nacional de Seguridad y Salud en el Trabajo 2022 - 2031.
- Mosquera, R., Parra, L., Ledesma, A. J., & Bonilla, H. F. (2021). Predicción de la accidentalidad laboral en la industria de pulpa y papel usando algoritmos de clasificación. *Informacion Tecnologica*, 32(1), 133-142. <https://doi.org/10.4067/S0718-07642021000100133>
- Mustapha, A., y Abdu, A. (2012). Application of Principal Component Analysis & Multiple Regression Models in Surface Water Quality Assessment. *Journal of Environment and Earth Science*, 2(2). www.iiste.org
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanhderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. y Duchesnay, E. (2011). *Sklearn.decomposition.PCA*. Scikit learn. Recuperado el día 26 de julio de 2023 de <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>.

- Resolución 0312 de 2019 [Ministerio del Trabajo]. Por la cual se definen los Estándares Mínimos del Sistema de Gestión de la Seguridad y Salud en el Trabajo SG-SST. 13 de febrero de 2019.
- Superintendencia de Sociedades. (2022). Atlas de Insolvencia en Colombia.
- Organización Internacional del Trabajo (Sin fecha). Seguridad y Salud en el Trabajo. Organización Internacional del Trabajo. Recuperado el día 10 de febrero de 2023 de <http://www.ilo.org/global/topics/safety-and-health-atwork/lang--es/index.html>.
- Villada, J., Castaño, L., y Pereira, L. (2023). Propuesta de diseño de un modelo matemático predictivo aplicado a la prevención de riesgos laborales en el sector de la construcción. Universidad ECCI.

10.ANEXOS

Anexo 1. Tablas de composición de componentes principales por sector económico

Anexo 2. Resultados escenarios de simulación por sector económico.

Anexo 3. Gráficas de aplicación del modelo en los diferentes sectores económicos.

Anexo 4. Gráficas de análisis multivariado.