

TRABAJO DE INVESTIGACIÓN: MINERÍA DE DATOS EN ACCIDENTES VIALES 1

Estudio Exploratorio De Modelos De Minería De Datos Para Análisis Y Caracterización De  
Accidentes Viales Municipales

Juan Camilo González García

Trabajo de Grado para Optar por el título de Ingeniero Industrial

Director

Edgar Eduardo Córdoba Sarmiento

M. Sc. Ingeniería Industrial

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Estudios Industriales y Empresariales

Ingeniería Industrial Bucaramanga

2024

**Dedicatoria**

Dedicado a mis padres Adriana y Javier, quienes con sus enseñanzas me nutrieron como persona.

### **Agradecimientos**

A mi familia por su apoyo durante esta etapa de mi vida y el impulso por alcanzar nuevos horizontes.

A mis compañeros y amigos, resultantes de las coincidencias que me acompañaron y me siguen acompañando.

A mi director de proyecto Edgar Córdoba por su acompañamiento y apoyo durante este proceso.

**Tabla de Contenido**

	<b>Pág.</b>
Introducción .....	13
1. Definición del Proyecto .....	15
1.1. Planteamiento del Problema .....	15
1.2. Objetivos .....	17
1.2.1. Objetivo General.....	17
1.2.2. Objetivos Específicos.....	17
2. Marco de Referencia .....	18
2.1. Marco Teórico.....	18
2.1.1. Seguridad Vial .....	18
2.1.2. Accidente vial .....	18
2.1.3. Minería de Datos y su Aplicación a la Seguridad Vial .....	19
2.1.4. Evaluación del Desempeño Predictivo .....	19
2.1.5. Modelos Analíticos Evaluados para la Gestión de la Seguridad Vial .....	19
2.2. Marco de Antecedentes.....	21
2.2.1. Modelo de Predicción de Gravedad de Accidentes de Tránsito: un Análisis de los Siniestros en Bogotá, Colombia.....	22
2.2.2. Modelización Econométrica de los Accidentes de Tránsito en el Ecuador .....	22
2.2.3. Aplicación de Técnicas de Agrupamiento para Caracterizar Patrones de Siniestros Viales en Ecuador en el Año 2021.....	23

2.2.4. Un Análisis Bibliométrico y Referencial de Aprendizaje Automático y AutoML en la Predicción de la Severidad de Accidentes: El Estudio de Caso de Tres Ciudades.....	24
3. Metodología .....	24
3.1. Definición del Problema de Investigación.....	25
3.2. Búsqueda y Extracción de Información.....	26
3.2.1. Diseño de la Estrategia de Búsqueda. ....	26
3.2.2. Selección de Base de Datos .....	27
3.2.3. Extracción Preliminar de Información.....	27
3.3. Aplicación de los Criterios de Exclusión e Inclusión .....	27
3.4. Recopilación y Organización de la Información .....	28
3.5. Interpretación de Hallazgos .....	30
4. Revisión de Literatura.....	30
4.1. Protocolo de búsqueda .....	30
4.2. Análisis Bibliométrico .....	32
4.2.1. Publicaciones por Año .....	32
4.2.2. Conexiones Entre Autores .....	33
4.2.3. Palabras Claves .....	33
4.2.4. Principales Países.....	34
4.3. Análisis de Literatura.....	35
4.3.1. Factores Determinantes en la Ocurrencia y Severidad de Accidentes.....	36
4.3.2. Estrategias y Modelos Para la Gestión de la Seguridad Vial.....	38
4.4. Análisis de Literatura Complementario .....	44

4.4.1. Tipo de predicción abordado .....	45
4.4.2. Clasificación de Modelos Encontrados en la Literatura .....	46
4.4.3. Clasificación de Factores Encontrados en la Literatura.....	48
4.4.4. Uso de Tipo de Categorías de Factores Según el Tipo de Predicción .....	51
4.4.5. Análisis cruzado entre tipos de predicción, modelo y factores considerados.....	53
5. Recomendaciones para un Marco de Aplicación Efectivo .....	56
5.1. Implementación Según Tipo de Predicción .....	56
5.1.1. Estudio de Frecuencia .....	56
5.1.2. Estudio de Gravedad .....	57
5.1.3. Estudio simultaneo de frecuencia y gravedad.....	57
5.2. Implementación Según Tipo Modelo.....	58
5.2.1. Modelos Supervisados .....	58
5.2.2. Modelos No Supervisados .....	59
5.2.3. Modelos Híbridos.....	60
5.2.4. Modelos de Inteligencia Artificial Explicable .....	60
5.3. Implementación Según Tipo Categorías de Factores.....	61
6. Marco de Gestión para la Implementación de Modelos Predictivos en Seguridad Vial Municipal .....	63
6.1. Diagnóstico de Información Disponible. ....	63
6.1.1. Identificación y Recopilación .....	64
6.1.2. Clasificación Temática de Variables .....	64
6.1.3. Evaluación de la Calidad y Confiabilidad de los Datos.....	65
6.1.4. Identificación de Vacíos y Estrategias de Mitigación.....	66

6.2.	Definición del Enfoque Predictivo.....	67
6.2.1.	Selección del Tipo de Predicción.....	67
6.2.2.	Criterios para Elegir un Enfoque Adecuado .....	67
6.2.3.	Enfoque Escalonado y Adaptativos .....	68
6.3.	Selección del Modelo de Minería de Datos .....	68
6.3.1.	Valoración de Condiciones Técnicas y Operativas .....	69
6.3.2.	Criterios de Selección del Modelo .....	69
6.4.	Aplicaciones Estratégicas de los Modelos Predictivos .....	70
6.4.1.	Priorización de Zonas y Momentos Críticos.....	71
6.4.2.	Intervención Focalizada de Factores de Riesgo.....	72
7.	Conclusiones .....	73
8.	Recomendaciones .....	74
	Referencias bibliográficas.....	76

**Lista de Tablas**

<b>Tabla 1</b>	Ecuación de búsqueda.....	31
<b>Tabla 2</b>	Criterios de Inclusión y Exclusión de Artículos. ....	31
<b>Tabla 3</b>	Segmentación de Artículos Basado en Tipo de Predicción .....	45
<b>Tabla 4</b>	Segmentación de Artículos Basado en Tipo de Modelo.....	47
<b>Tabla 5</b>	Relación de los modelos más empleados y su objetivo de predicción .....	48
<b>Tabla 6</b>	Frecuencia de categorías de factores en los artículos .....	50
<b>Tabla 7</b>	Tabla de comparación factores categorizados por tipo de predicción .....	51
<b>Tabla 8</b>	Tabla de comparación factores categorizados por tipo de modelo de predicción ....	52
<b>Tabla 9</b>	Factores propuestos para durante el proceso de valoración.....	65
<b>Tabla 10</b>	Consideraciones basadas en objetivo de predicción .....	67
<b>Tabla 11</b>	Preguntas para selección de enfoque de predicción.....	68
<b>Tabla 12</b>	Consideraciones para la selección de un modelo.....	70

### Lista de Figuras

<b>Figura 1.</b> Fases metodológicas desarrolladas.....	25
<b>Figura 2.</b> Publicaciones por año.....	32
<b>Figura 3.</b> Relación entre autores. ....	33
<b>Figura 4.</b> Relación entre Palabras Clave.....	34
<b>Figura 5.</b> Países principales de los artículos arrojados por la ecuación de búsqueda. ....	35
<b>Figura 6.</b> Caracterización de Variables de Siniestros Viales .....	37

**Lista de Apéndices**

**Apéndice A.** Bitácora de Investigación.

**Apéndice B.** Clasificación Cruzada de Modelos, Factores y Tipo de Predicción.

**Apéndice C.** Artículo de Carácter Publicable.

## Resumen

**Título:** Estudio Exploratorio De Modelos De Minería De Datos Para Análisis Y Caracterización De Accidentes Viales Municipales<sup>1</sup>

**Autor:** Juan Camilo González García<sup>2</sup>

**Palabras Clave:** Minería de Datos, Accidentes de Tránsito, Predicción, Frecuencia, Gravedad

**Descripción:** Esta investigación desarrolla un estudio exploratorio sobre el uso de modelos de minería de datos para el análisis, caracterización y predicción de accidentes de tránsito en entornos municipales.

A partir de una revisión sistemática de la literatura especializada, se identificaron los modelos predictivos más utilizados, así como los factores determinantes en su aplicación. Los modelos se clasifican en cuatro categorías: supervisados, no supervisados, híbridos y aquellos que incorporan técnicas de Inteligencia Artificial Explicable (XAI). Asimismo, se analiza su aplicación según el tipo de predicción abordada sea esta frecuencia, gravedad o predicción simultánea y los factores utilizados como predictores, lo que permite generar un análisis cruzado entre tipo de predicción, modelo y factores considerados.

Con base en estos hallazgos, se propone un marco metodológico de cinco fases para la implementación de modelos predictivos en la gestión de la seguridad vial, incluyendo el análisis de la información disponible, definición del tipo de predicción, selección del modelo y toma de decisiones basadas en los resultados. A través de este estudio se concluye que la minería de datos ofrece un potencial significativo para transformar la gestión de la seguridad vial, especialmente en el desarrollo de modelos escalables, integradores y alineados con los objetivos estratégicos locales.

---

<sup>1</sup> Trabajo de Grado

<sup>2</sup> Facultad de Ingenierías Fisicomécanicas. Escuela de Estudios Industriales y Empresariales. Director: Edgar Eduardo Córdoba Sarmiento M. Sc. Ingeniería Industrial

**Abstract**

**Title:** Exploratory Study of Data Mining Models for the Analysis and Characterization of Municipal Traffic Accidents<sup>3</sup>

**Author:** Juan Camilo González García<sup>4</sup>

**Key-Words:** Data mining, Road Accidents, Forecasting, Frequency, Severity

**Description:** This research presents an exploratory study on the application of data mining techniques for the analysis, characterization, and prediction of traffic accidents in municipal contexts.

Based on a systematic review of specialized literature, the study identifies the most frequently used predictive models and their contributing factors. The models are classified into four categories supervised, unsupervised, hybrid and those incorporating Explainable Artificial Intelligence (XAI) techniques. Additionally, their application scope according to the type of prediction addressed frequency, severity or simultaneous prediction, as well as the factors used as predictors, including road, environmental, human and vehicle-related factors, capturing relationship between prediction type with factors and models used.

As result of these findings, the research proposes a five-phase methodological framework for the application of predictive models in road safety management, including diagnosis of data availability, defining the predictive objective, model selection, and data-driven decision making. The study concludes that data mining has a significant potential to improve road safety management, especially through the development of scalable and integrative models aligned with local strategic objectives.

---

<sup>3</sup> Bachelor's Thesis

<sup>4</sup> Facultad de Ingenierías Fisicomécanicas. Escuela de Estudios Industriales y Empresariales. Director: Edgar Eduardo Córdoba Sarmiento M. Sc. Ingeniería Industrial

## **Introducción**

Durante las últimas décadas, los accidentes de tránsito han representado un problema de salud pública a nivel global, con consecuencias significativas en términos de pérdidas humanas y costos. Según la Organización Mundial de la Salud (2023), los accidentes de tránsito representan una de las principales causas de muerte entre personas jóvenes, siendo agravado en países de ingresos medio y bajos, donde la gestión de la seguridad vial enfrenta desafíos estructurales. En este contexto, el desarrollo de herramientas que permitan la mitigación de los efectos, así como sus ocurrencias, se vuelven de inmenso valor para los entes gubernamentales y ciudadanos.

Ante este panorama, el presente trabajo de investigación propone un marco exploratorio enfocado en la aplicabilidad de diversos modelos de minería para la predicción, caracterización y análisis de los accidentes viales, mediante la identificación de patrones críticos, variables incidentes y zonas de riesgo. Este enfoque se articula dentro de una analítica predictiva que busca superar las limitaciones de los modelos estadísticos tradicionales al permitir un manejo flexible y adaptativo de los datos disponibles.

En el ámbito nacional, a pesar de las normativas e instrumentos regulatorios establecidos para la reducción de los accidentes de tránsito, todavía hay desafíos para acortar las brechas entre el desarrollo efectivo de estrategias preventivas creadas a partir de datos históricos y su efectividad, los métodos estadísticos tradicionales han sido útiles para entender el fenómeno de manera descriptiva, pero tienen limitaciones cuando se trata de identificar patrones ocultos, evaluar múltiples factores y prever comportamientos en entornos cambiantes (Rodríguez et al., 2003).

En respuesta a esta problemática, la integración de técnicas avanzadas de minería de datos y el aprendizaje automático se presenta como una gran oportunidad para mejorar la gestión de la seguridad vial. Estas técnicas permiten analizar grandes cantidades de

información, descubrir relaciones no evidentes entre variables y crear modelos predictivos que se pueden aplicar tanto a nivel territorial como institucional.

El objetivo de esta investigación es abordar patrones relevantes dentro de la aplicación de modelos predictivos y factores determinantes estudiados dentro de la literatura, que pueden ser aplicables en contextos similares. Para esto, se generó una revisión narrativa de la literatura especializada abarcando 105 artículos, seguido de una sugerencia metodológica que conecta los resultados teóricos con una estructura de gestión adaptada a situaciones locales.

Este documento está organizado en seis secciones: la parte inicial describe el tema de la investigación, su importancia y el método sugerido; La segunda parte especifica los objetivos y explica la importancia del estudio; La tercera parte discute lo teórico y situacional, el cuarto describe la revisión de la literatura; El quinto revela los hallazgos y análisis; y el sexto propone un plan para administrar y ejecutar el marco basado en los resultados encontrados.

## **1. Definición del Proyecto**

### **1.1. Planteamiento del Problema**

Los accidentes de tránsito son un desafío significativo para la seguridad vial y el desarrollo urbano, generando impactos sociales, económicos y ambientales. A nivel global, la siniestralidad vial es una de las principales causas de muerte y lesiones, afectando no solo a conductores y pasajeros, sino también a peatones y otros actores viales (Ahmed et al., 2023; Jeong et al., 2022). La identificación de patrones y factores de riesgo presentes en el transcurso de estos eventos es fundamental para la implementación de estrategias de mitigación efectivas por parte de los organismos de control y las agencias de seguridad vial.

En la literatura científica, se han estudiado diversos modelos de minería de datos enfocados en la predicción y caracterización de accidentes viales (Jeong et al., 2022; Ahmed et al., 2023). Estos modelos permiten analizar múltiples variables con incidencia en la accidentalidad, tales como condiciones meteorológicas, infraestructura vial, comportamiento del conductor y características del vehículo, entre otras. La capacidad de estos modelos para procesar grandes volúmenes de información y detectar relaciones complejas entre variables ha abierto nuevas posibilidades para la intervención y acción de las entidades responsables de la seguridad vial.

A pesar de los avances en la aplicación de modelos predictivos y analíticos en la seguridad vial, sigue existiendo un vacío en la adopción y aplicación práctica de estos enfoques por parte de los organismos reguladores. Muchas agencias de seguridad vial aún dependen de modelos estadísticos convencionales enfocados a análisis descriptivos, los cuales pueden ser insuficientes para generar respuestas oportunas y personalizadas según las particularidades de cada entorno urbano o carretera. La falta de integración de técnicas avanzadas de minería de datos en la gestión de la seguridad vial limita la capacidad de las entidades para tomar

decisiones basadas en evidencia y diseñar políticas más efectivas, en la búsqueda de mitigar la frecuencia y severidad de los accidentes.

En este contexto, el propósito de este estudio es analizar y presentar modelos de minería de datos aplicados a la caracterización y predicción de accidentes de tránsito, con el objetivo de proporcionar herramientas que permitan a las entidades y organismos de control mejorar sus estrategias de intervención y acción. Para ello, se explorarán las principales variables con incidencia en la accidentalidad, identificadas a partir de la literatura existente, y se evaluará la efectividad de diferentes enfoques analíticos en la detección de frecuencia y severidad.

A través de este estudio, se busca generar insumos que permitan a los actores involucrados en la gestión de la seguridad vial optimizar la toma de decisiones y diseñar estrategias de mitigación basadas en datos reales y modelos predictivos avanzados, logrando así una reducción efectiva en la ocurrencia y gravedad de los siniestros viales.

## **1.2. Objetivos**

### ***1.2.1. Objetivo General***

Realizar un estudio exploratorio enfocado en modelos de minería de datos para el análisis y caracterización de accidentes viales, con el fin de generar insumos que favorezcan la gestión de la seguridad vial a nivel municipal.

### ***1.2.2. Objetivos Específicos***

1. Realizar una revisión de la literatura sobre los algoritmos de minería de datos aplicados al análisis y caracterización de accidentes viales, con especial énfasis en aquellos utilizados para predecir la gravedad y la frecuencia de accidentes.
2. Identificar los factores comunes y determinantes en los modelos que predicen la gravedad y la frecuencia de accidentes viales, proporcionando un análisis integral de estos factores.
3. Explorar y analizar la relación entre los diferentes modelos de minería de datos encontrados, evaluando cómo pueden integrarse en un marco de gestión eficaz para la seguridad vial municipal.
4. Desarrollar un artículo científico con base en los resultados obtenidos, adecuado para ser publicado en una revista académica.

## 2. Marco de Referencia

### 2.1. Marco Teórico

La seguridad vial es un problema global al que ha cobrado importancia dentro de la investigación científica debido a su impacto social y económico. Debido al avance en la inteligencia artificial y minería de datos, se ha desarrollado herramientas que permiten analizar patrones de los accidentes, además de predecir la severidad y frecuencia de siniestros en los entornos urbanos y carreteras interurbanas.

Este marco teórico representa los fundamentos de la seguridad vial, la minería de datos aplicada a la predicción de accidentes y los principales modelos de datos utilizados en la literatura para determinar y analizar los factores de riesgo implícitos.

#### 2.1.1. Seguridad Vial

La seguridad vial se define como el conjunto de acciones y prácticas implementadas para prevenir los accidentes de tránsito y/o reducir sus consecuencias. En América Latina, países como Colombia han implementado planes nacionales de seguridad vial, con iniciativas enfocadas en mejorar la infraestructura y aplicar tecnologías de análisis predictivo

#### 2.1.2. Accidente vial

Los accidentes de tránsito pueden ser clasificados según su causa de acuerdo con el marco de clasificación propuesto por Jeong et al., (2022), dentro del cual se describen los siguientes segmentos:

- **Error humano:** Exceso de velocidad, conducción bajo efectos del alcohol, distracción del conductor.
- **Condiciones viales:** Estado del pavimento, señalización inadecuada, fallas estructurales de las vías.
- **Factores ambientales:** Lluvia, neblina, iluminación insuficiente, entre otras.

Es importante recalcar que las consecuencias de los siniestros viales afectan tanto a las víctimas como al sistema de salud y economía del país, consecuencia del costo implícito en la rehabilitación, asistencia médica, pérdidas laborales y arreglos de infraestructura.

### ***2.1.3. Minería de Datos y su Aplicación a la Seguridad Vial***

La minería de datos es una técnica que permite el análisis de grandes volúmenes de información con el propósito de identificar patrones, tendencias y relaciones implícitas entre variables. Su aplicación dentro de la seguridad vial ha permitido el desarrollo de modelos predictivos que ayudan a mejorar la planificación y gestión del tránsito (Ghandour et al., 2020; Hadjidimitriou et al., 2020; Santos et al., 2021).

### ***2.1.4. Evaluación del Desempeño Predictivo***

Evaluar el rendimiento de los modelos es fundamental para garantizar la eficacia, especialmente en contextos en los que la falla pueden ser críticos como la seguridad vial. Dentro de esto, se destaca la medición de la precisión de los modelos. La precisión responde a la proporción de predicciones realizadas correctamente por un modelo en relación con el total de predicciones (Google Inc., 2024).

### ***2.1.5. Modelos Analíticos Evaluados para la Gestión de la Seguridad Vial***

A continuación, se presentan y describen de manera breve algunos de los modelos analíticos analizados durante el proceso de revisión, agrupados según su enfoque metodológico dentro de la taxonomía de Faisal et al., (2025) dentro de la cual el aprendizaje automático (*Machine Learning*) puede ser descompuesto en dos grandes clases de técnicas las de aprendizaje supervisado y no supervisado, basado en su tipo de entrenamiento.

**Aprendizaje Automático Supervisado.** La característica principal de estos modelos es el entrenamiento a través de la ingesta de datos etiquetados previamente para entrenar el modelo, donde cada conjunto de entrenamiento contiene características y su valor esperado,

permitiendo al modelo generar relaciones entre ambos grupos de datos y posteriormente de un nuevo conjunto de datos.

Estos modelos pueden ser diseñados con dos objetivos principales: clasificación, o regresión según al diseño haya sido utilizado, siendo la principal diversidad que los modelos que participan en clasificación puedan pronosticar variable tipo discreta (Ejemplo: Nivel gravedad de un accidente: Fatal, grave u leve (Pérez-Sala et al., 2023)), , mientras que los que participan en regresión han sido dirigidos a variable número continuo (Ejemplo: Número de siniestros en un tramo (Singh et al., 2020)).

Dentro de los modelos de clasificación podemos encontrar Bayesiano Ingenuo (*Naive Bayes*) y regresión logística, en los modelos de regresión dedicados principalmente se encuentran modelos como regresión lineal y *Polynomial Regression*, incluyendo modelos avanzados de regresión. También existen modelos no especializados como pueden ser *Decision Tree*, *Random Forest*, *K-Nearest Neighbor* y *Support Vector Machine*.

**Aprendizaje Automático no Supervisado.** A diferencia de los modelos de aprendizaje supervisados, su fase de entrenamiento no depende de datos etiquetados previamente, siendo particularmente útiles en fases exploratorias de datos, ya que su objetivo principal es identificar patrones, estructurar o relaciones subyacentes de la información.

Dentro de este grupo se pueden clasificar en dos divisiones agrupamiento de datos (*Clustering*) o reducción de dimensionalidad, consistiendo la primera en separar la información en grupos con cierto grado de similitud, mientras la segunda busca simplificar las características principales de la data, conservando la mayor cantidad de información con el propósito de simplificar la interpretación.

Los principales modelos agrupamiento están *K-Means Clustering*, *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* y Clustering jerárquico, abordados en la literatura se encontraron reglas de asociación como *Apriori Algorithm*, además del uso de

*LigthGMB* para agrupación (Cai, 2020; Kuyumcu et al., 2023; Rezashoar et al., 2024), mientras el relacionado con la reducción de dimensionalidad se encuentra el análisis de componente principal (*Principal Component Analysis*).

**Inteligencia Artificial Explicable (XAI).** Dentro del creciente mundo de la inteligencia artificial se necesita la creación de herramientas con la capacidad de explicar la manera en la que modelos complejos como las redes neuronales realizan la toma de decisiones. Autores como Arrieta et al., (2019) dentro de su revisión a define estas herramientas como elementos con la capacidad de analizar y detallar el funcionamiento de modelos complejos, teniendo en cuenta la audiencia a la que se dirige dicha información. Esta característica es fundamental para aumentar la confianza y efectividad en el uso de Inteligencia Artificial (IA) en contextos sensibles como la seguridad vial. Entre las técnicas XAI más destacadas se encuentra métodos como SHAP, LIME y la visualización de relevancia de características.

## **2.2. Marco de Antecedentes**

Dentro del marco de la seguridad vial, el creciente interés por forma de reducir el número de accidentes ha aumentado en interés en los últimos años debido al impacto que conlleva social y económicamente. Según la Organización mundial de la Salud (OMS) se estima que anualmente, fallecen aproximadamente 1,19 millones de personas a causa de siniestros viales (World Health Organization, 2023).

En Colombia, se establece el Plan Nacional de Seguridad Vial (PNSV) 2022-2031 como estrategia para disminuir las fatalidades consecuencia de los accidentes de tránsito. A pesar de esto, análisis de la accidentalidad en municipalidades como Bucaramanga, muestran que, a pesar de la disminución en la tasa de accidentes, la tasa de mortalidad en grupos como motociclista y peatones ha aumentado (Observatorio Nacional de Seguridad Vial, 2023).

Ante esta problemática, la disponibilidad de datos abiertos y el uso de herramientas de Big Data, minería de datos y Machine Learning han permitido desarrollar modelos predictivos y de caracterización de siniestros viales. Estos enfoques han sido utilizados con éxito en diferentes regiones del mundo para mejorar la comprensión de los factores que influyen en la ocurrencia y severidad de los accidentes.

Dentro de los trabajos encontrados relacionados al tópico, se destacan los siguientes por su relación geográfica similar a Colombia, además de su marco de aplicación dentro de la toma de decisiones de los organismos de gestión de movilidad.

### ***2.2.1. Modelo de Predicción de Gravedad de Accidentes de Tránsito: un Análisis de los Siniestros en Bogotá, Colombia.***

En el caso de Monroy & Díaz (2018) desarrollaron un modelo estructurado con el propósito de predecir la gravedad y analizar siniestros viales basado en registros históricos de accidentes en la ciudad de Bogotá, Colombia. Este proceso se desarrolló a partir de un proceso de “clustering” de data considerando variables mixtas a partir de técnicas *K-prototypes*.

Adicionalmente se usó un modelo de regresión logística para predecir la gravedad de estos eventos a partir de diferentes variables como actores viales involucrados, lugar y hora del siniestro vial a través de esto se permitió determinar localidades con alto nivel de accidentalidad.

### ***2.2.2. Modelización Econométrica de los Accidentes de Tránsito en el Ecuador***

En el caso de Castillo et al. (2020) se aborda la problemática de los accidentes de tránsito en Ecuador desde enfoques analíticos como predicción de accidentes a través de modelos ARIMA y la identificación de los factores determinantes en la ocurrencia de accidentes a través de un modelo *legit multinominal*.

Basado en datos históricos semanales de accidentes de tránsito registrados entre el 2015 y 2018 proporcionados por la Dirección Nacional de Control del Tránsito y Seguridad Vial de

Ecuador, además de la predicción de la cantidad de accidentes, el estudio exploró los factores humanos y físicos que influyen en la ocurrencia de accidentes, diferenciando los tipos de vehículos involucrados.

A través de esto se determinó que los motociclistas y conductores de vehículos pesados tienden a ser más propensos a sufrir accidentes de tránsito con heridas graves, asimismo, el exceso de velocidad y la embriaguez tienen alto impacto en los siniestros viales, además, factores externos como las condiciones climatológicas y el estado de las carreteras tienden a afectar significativamente el riesgo de accidentes según el tipo de vehículo involucrado.

### ***2.2.3. Aplicación de Técnicas de Agrupamiento para Caracterizar Patrones de Siniestros Viales en Ecuador en el Año 2021.***

En el estudio realizado por Maza (2023) enfocado en la caracterización de patrones en los siniestros viales en Ecuador durante el 2021, a través de modelos de clustering de datos no supervisados. Esta investigación se fundamentó en la metodología CRISP-DM, que permitió estructurar el proceso de minería de datos desde la comprensión del problema hasta la generación de modelos analíticos.

La metodología utilizada constó de identificación de conglomerados con mayor incidencia de accidentes viales en Ecuador y proporcionar información relevante a las entidades de control para la toma de decisiones, mediante la aplicación de 3 algoritmos de agrupamiento, *K-means* para optimización del número de clústeres mediante el método de “Elbow”, agrupamiento jerárquico y *DBSCAN* basado en la densidad adecuado para la identificación de agrupaciones de diferente tamaño y forma.

La investigación permitió segmentar los accidentes de tránsito en diferentes regiones del Ecuador, asimismo, se encontró que las provincias con mayor cantidad de siniestros fueron Guayas, Pichincha y Manabí. Para cada una de estas regiones se establecieron patrones de comportamiento y causas principales que tenían impacto dentro de los siniestros, meses de

mayor incidencia y horarios críticos, esto permitió proveer a los entes de control la información para facilitar la implementación de políticas públicas basadas en datos.

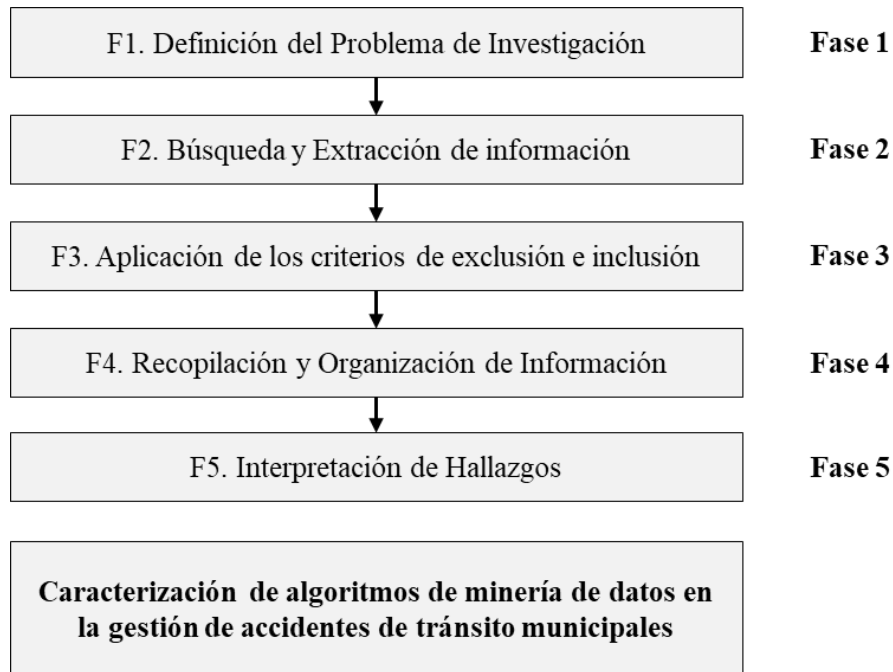
#### ***2.2.4. Un Análisis Bibliométrico y Referencial de Aprendizaje Automático y AutoML en la Predicción de la Severidad de Accidentes: El Estudio de Caso de Tres Ciudades***

El estudio realizado por Angarita-Zapata et al. (2021) es de utilidad como un marco de implementación y comprensión de las limitaciones asociadas al enfoque del Machine Learning dentro del contexto nacional, dentro de su proceso se evaluaron diferentes bases de datos y modelos, así como los resultados obtenidos en relación con diferentes ciudades.

Particularmente el estudio aborda una comparación del rendimiento de diferentes algoritmos para minería de datos con conjuntos gubernamentales de Medellín, Bogotá y Bucaramanga. Uno de los hallazgos importantes es relacionado con el desbalance de la data para el data set de Bucaramanga, en general los resultados muestran la posibilidad de aplicar este tipo de técnicas con una precisión razonable con la información disponible gubernamentalmente.

### **3. Metodología**

La metodología tomada como base para esta investigación fue una adaptación de la metodología propuesta por Khan et al. (2003) para revisiones sistemáticas de literatura, ajustándola a un enfoque exploratorio y narrativo, teniendo en cuenta el enfoque y alcance del proyecto. A continuación, se exponen las fases metodológicas desarrolladas para la ejecución del proyecto (Ver Figura 1).

**Figura 1***Fases metodológicas desarrolladas.*

*Nota.* Metodología adaptada de *Five steps to conducting a systematic review*, Por Khan et al., 2003, Journal of the Royal Society of Medicine.

### 3.1. Definición del Problema de Investigación

En esta primera fase, se contemplan los accidentes de tránsito como un problema significativo para la seguridad vial y el desarrollo urbano. La siniestralidad vial a nivel global es una de las principales causas de muerte y lesiones que afectan no únicamente a conductores, sino también a pasajeros, peatones y demás actores viales, por esto, la identificación de patrones, datos o factores de riesgo en la ocurrencia de este tipo de accidentes es fundamental para el diseño e implementación de nuevas estrategias de mitigación o prevención, por parte de organismos de control y agencias de seguridad vial.

Algunas de las causas que llegan a provocar accidentes de este tipo son el error humano, el exceso de velocidad, la conducción bajo efectos de alcohol u otras sustancias psicoactivas, la no utilización de elementos de seguridad, las distracciones durante la conducción y la falta de seguridad de la infraestructura vial entre otras (Organización Mundial de la Salud, 2023).

Lo anterior surgió como motivación para realizar un estudio exploratorio, que se enfoque en modelos de minería de datos propuestos en la literatura con el propósito de caracterizar los factores asociados a los accidentes viales, y proporcionar herramientas que favorezcan a entidades u organismos de control la gestión de la seguridad vial. Para ello, se evaluaron variables relacionadas a la predicción de frecuencia y gravedad con incidencia en la accidentalidad previamente identificadas en la literatura, y se evalúa la efectividad de los diferentes modelos de minería de datos para dicho fin.

### **3.2. Búsqueda y Extracción de Información**

Con base en lo propuesto en la primera etapa metodológica, se dispone la búsqueda y extracción de la información para alcanzar dicho planteamiento. Por tanto, se debe realizó una contextualización o revisión preliminar de literatura asociada para familiarizarse con la terminología relacionada en estudios de este tipo. La estrategia de búsqueda estará diseñada basada en los principios de trazabilidad y transparencia metodológica de Khan et al. (2003).

#### **3.2.1. *Diseño de la Estrategia de Búsqueda.***

Se definieron los términos de búsqueda basados en tesauros y palabras clave específicas del área de interés, basados en la revisión de literatura preliminar. Entre los términos considerados estarán: “Data Mining”, “Predictive Models”, “Traffic Accidents”, “Road Accidents”, “Severity”, “Risk Factors” y “Frequency”; estos términos fueron utilizados para la construcción de ecuaciones de búsqueda para la extracción de la mayor cantidad de artículos pertinentes para la investigación.

Asimismo, se consideraron la inclusión de sinónimos y términos alternativos ampliar el rango de la búsqueda, dando la oportunidad de integrar estudios relevantes publicados con terminologías no asociadas directamente con los términos planteados inicialmente.

### ***3.2.2. Selección de Base de Datos***

Se seleccionaron bases de datos científicas reconocidas en campos interdisciplinarios, específicamente Scopus y Web of Science (WOS). Estas bases de datos ofrecen características específicas de los artículos, como el filtrado por tipo de documento, área temática y fecha de publicación, lo que simplificó el proceso de extracción al centrar los resultados logrados con el objetivo de la investigación.

En el procedimiento de búsqueda, se utilizó filtros preliminares para definir los resultados a artículos publicados con fechas entre el 2018 y 2024 y categorizados como artículos de investigación o revisiones de literatura vinculadas al tema de estudio.

### ***3.2.3. Extracción Preliminar de Información***

Los artículos seleccionados fueron descargados y organizados para realizar una revisión inicial. Se realizó una lectura preliminar de del título y resumen para confirmar la pertinencia temática después de proceder pasar a la aplicación de los criterios de inclusión y exclusión en la siguiente fase.

## **3.3. Aplicación de los Criterios de Exclusión e Inclusión**

Una vez realizada la búsqueda de documentos el proceso de su selección busca identificar aquellos que cumplen los criterios de inclusión y exclusión establecidos para la investigación. En este caso, se decide incluir aquellos artículos publicados entre el año 2018 y 2024, en los que el área de investigación y temática estuviesen directamente relacionada a accidentes de tránsito, ingeniería y análisis de datos.

Por otro lado, se requieren criterios de exclusión tales como aquellos documentos que sean de fuentes como artículos con accesos anticipado, conferencias y cualquier otro formato diferente a artículo publicado.

### **3.4. Recopilación y Organización de la Información**

Posterior a la aplicación de la depuración de los documentos a través de los criterios de inclusión y exclusión, se realizó de manera sistemática la consolidación y organización de la información de los documentos que cumplen con los criterios mencionados. Esta fase tuvo como propósito organizar la información en un formato homogéneo con los datos relevantes de cada estudio, favoreciendo el análisis posterior y síntesis.

Para ordenar la información, se utilizó una bitácora de revisión creada en una hoja de cálculo (Excel), la cual permitió sistematizar de manera ordenada los atributos clave de cada artículo examinado.

Esta bitácora contiene los siguientes campos: título del artículo, principales autores, año de publicación, el objetivo del estudio, la metodología utilizada, los modelos de minería de datos utilizados, las variables analizadas, el tipo de predicción realizada (frecuencia, severidad o ambos elementos), los principales resultados obtenidos y las limitaciones identificadas en cada investigación.

Una vez finalizada la recopilación y sistematización de la información en la bitácora, se desarrolló la fase de análisis e interpretación de los hallazgos, para etapa tiene como finalidad sintetizar los resultados obtenidos de manera narrativa, permitiendo identificar patrones, tendencias y país de estudio.

El estudio de los datos se realizó de manera inductiva, lo que permitirá centrar la exploración temática en las investigaciones examinadas. Especialmente, se buscó:

- Determinar los métodos de minería de datos más utilizadas en el estudio y caracterización de accidentes viales municipales.

- Establecer las variables principales que suelen ser consideradas dentro de los modelos predictivos (como tipo de vía, condiciones meteorológicas, comportamiento del conductor, entre otras).
- Identificar patrones y tendencias en desarrollo en la aplicación de algoritmos para predicción de gravedad o frecuencia de accidentes.
- Evaluar los enfoques metodológicos utilizados en las diversas investigaciones, enfocando las similitudes y diferencias en los supuestos usados para la creación de modelos predictivos.
- Detectar limitaciones encontradas por los autores, como la disponibilidad de datos, sesgos en los registros de accidentes, o restricciones metodológicas asociadas al lugar donde de donde se obtuvieron los datos.

Además, se analizaron los entornos geográficos de aplicación de los modelos (urbano, rural, municipal) para valorar su habilidad para adaptarse en los diferentes contextos, permitiendo explorar y analizar la relación entre los diferentes modelos encontrados, proponiendo posibles lineamientos para su integración en un marco eficaz para la seguridad vial a nivel municipal, en coherencia con el objetivo 3.

El análisis se enriqueció empleando herramientas de síntesis como tablas de resumen y matrices comparativas, las cuales apoyan la exposición ordenada de los resultados más importantes. Finalmente, los descubrimientos fueron debatidos de manera crítica, incorporarlos en una narración consistente que no solo sintetice la situación actual del conocimiento del tema, sino que también detecte vacíos para futuros estudios y sugiera posibilidades de optimización en la aplicación de métodos de minería de datos aplicados a la administración de la seguridad vial a escala municipal.

### **3.5. Interpretación de Hallazgos**

La fase final de esta investigación contempló la difusión del conocimiento producido mediante la elaboración de un artículo científico de carácter publicable. El propósito de este es condensar y exponer de manera meticulosa los hallazgos encontrados. Con el fin de que puedan ser empleados como referencia por investigadores, profesionales y estudiantes interesados en la aplicación de técnicas de minería de datos en el ámbito de la seguridad vial y la gestión de accidentes de tránsito.

## **4. Revisión de Literatura**

Esta revisión de literatura tiene como propósito generar una caracterización de los accidentes de tránsito y los modelos de análisis de datos utilizados para la predicción de gravedad y frecuencia, para dar un marco de evidencia de los factores determinantes usados en los modelos expuestos por diversos autores generando un marco integral de análisis de dichos factores.

Dentro de este análisis se proporciona un análisis bibliométrico evaluando autores, fechas de publicación, palabras claves y países principales, este análisis fue desarrollado a partir de los resultados obtenidos luego de la aplicación de criterios de exclusión de los documentos obtenidos en Web of Science (WOS) usando el software de visualización VOSviewer.

### **4.1. Protocolo de búsqueda**

Para dar inicio al proceso de revisión de literatura se tomó como base una revisión de literatura gris, con el objetivo de analizar elementos y palabras claves para tener en cuenta en los artículos que serían pertinentes para el desarrollo de esta investigación dentro de las cuales se obtuvieron 10 palabras claves.

Teniendo en cuenta lo anterior se procedió a realizar una ecuación de búsqueda (Tabla 1) para aplicar en diferentes bases de datos, tales como *Web of Science* (WOS) se realizaron diferentes iteraciones teniendo en cuenta criterios de inclusión y exclusión con el propósito de limitar los artículos buscados ajustando elementos como el tipo de artículo, año y área de investigación.

**Tabla 1***Ecuación de búsqueda.*

Iteración	Ecuación	Resultados
1	TS = (("Forecasting" OR "Data Mining" OR "Predictive Modeling" OR "Machine Learning") AND ("Traffic Accidents" OR "Road Accidents" OR "Fatalities") AND ("Severity" OR "Risk Factors" OR "Frequency"))	415

Debido a la naturaleza de la investigación fue necesario realizar una segmentación efectiva excluyendo los tópicos relacionados con diferentes áreas que no están directamente relacionadas con el tema de estudio, tomando como criterios de inclusión y exclusión terminando con 162 artículos de los 415 obtenidos inicialmente (Ver Tabla 2).

**Tabla 2**

Criterios de Inclusión y Exclusión de Artículos.

Criterio	Descripción del Criterio	Justificación
<i>Año de Publicación</i>	2018 - 2024	Se realizó una segmentación de los últimos 5 años, con el propósito de obtener las investigaciones más recientes con respecto al tema.
<i>Tipo de Publicación</i>	Journal Article	Debido a la naturaleza del trabajo, se eliminaron fuentes como artículos con accesos anticipado, Conferencias, limitando la investigación solo a artículos publicados.
<i>Área de Investigación</i>	Relacionados con accidentes de tránsito, Ingeniería y análisis de datos	Con el propósito de evitar la inclusión de tópicos relacionados a medicina y ciencias anexas que no están directamente relacionadas a la investigación planteada.
<i>Accesibilidad</i>	Accesibilidad para la descarga de los artículos	Se descargaron los artículos excluyendo los artículos que no fueron posibles descargar en las bases de datos asociadas.

Con el fin de abordar más posibilidades, se realizó una búsqueda a través de *SCOPUS*, se realizó un ejercicio similar al realizado en *WOS*, se obtuvieron 155 resultados de los cuales no cumplían con los criterios planteados, Considerando la amplitud y complejidad del estudio, se optó por solo tomar como base de datos *WOS*.

Adicionalmente, para asegurarse que la segmentación realizada con los criterios de inclusión y exclusión se hizo una prelectura del resumen de cada uno de los 162 artículos para asegurarse que estuvieran alineados con el objetivo del trabajo, posterior a la descarga se cuenta con 105 artículos debido al último criterio de accesibilidad los cuales se usaron para el desarrollo de la revisión de literatura narrativa.

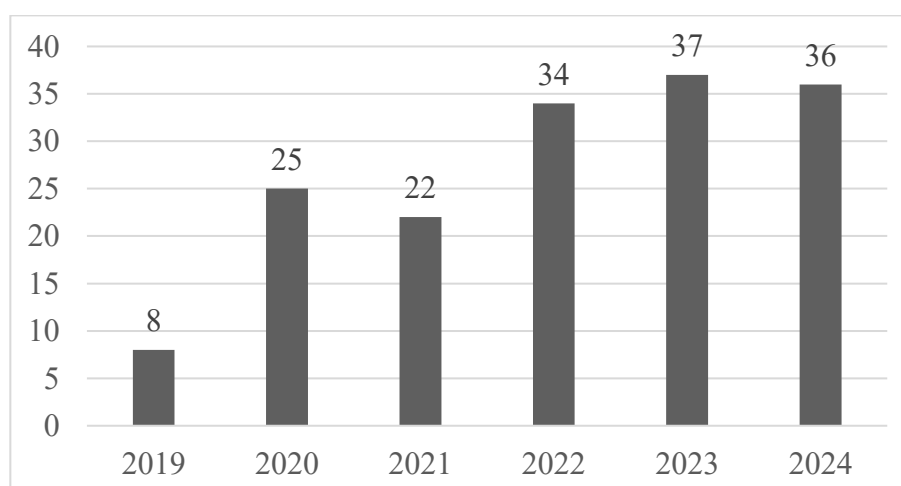
## 4.2. Análisis Bibliométrico

### 4.2.1. Publicaciones por Año

En la figura 2, se muestra un crecimiento en las investigaciones realizadas en los últimos 3 años sobre los modelos análisis para accidentes de tránsito, se evidencia una disminución entre los años 2021 y 2022, esto se puede deber a temas relacionados con el aislamiento por la pandemia del Covid-19.

**Figura 2**

*Publicaciones por año.*

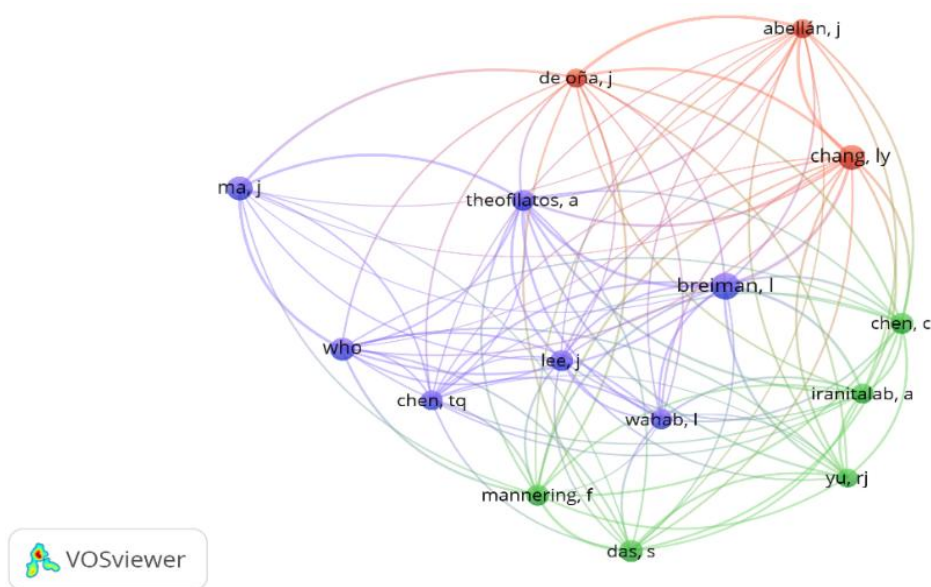


#### 4.2.2. Conexiones Entre Autores

En el análisis entre autores es posible evidenciar la existencia de tres clústeres o grupos principales, indicando grupos de investigación y áreas de trabajo similares, además de múltiples relaciones entre autores basadas en la cantidad de citas de los artículos revisados, siendo los más destacados Breiman y Chang teniendo la mayor cantidad de citas con 38 y 33 citas respectivamente, y una fuerte interconexión con autores como Abellán, Iranitalab y Chen. Por otro lado, Theofilatos y De Oña presentan una fuerte conexión con los autores mencionados anteriormente, a pesar de tener una menor cantidad de citas de 23 y 20 respectivamente, tal y como se observa en la red de la Figura 3.

#### Figura 3

*Relación entre autores.*



#### 4.2.3. Palabras Claves

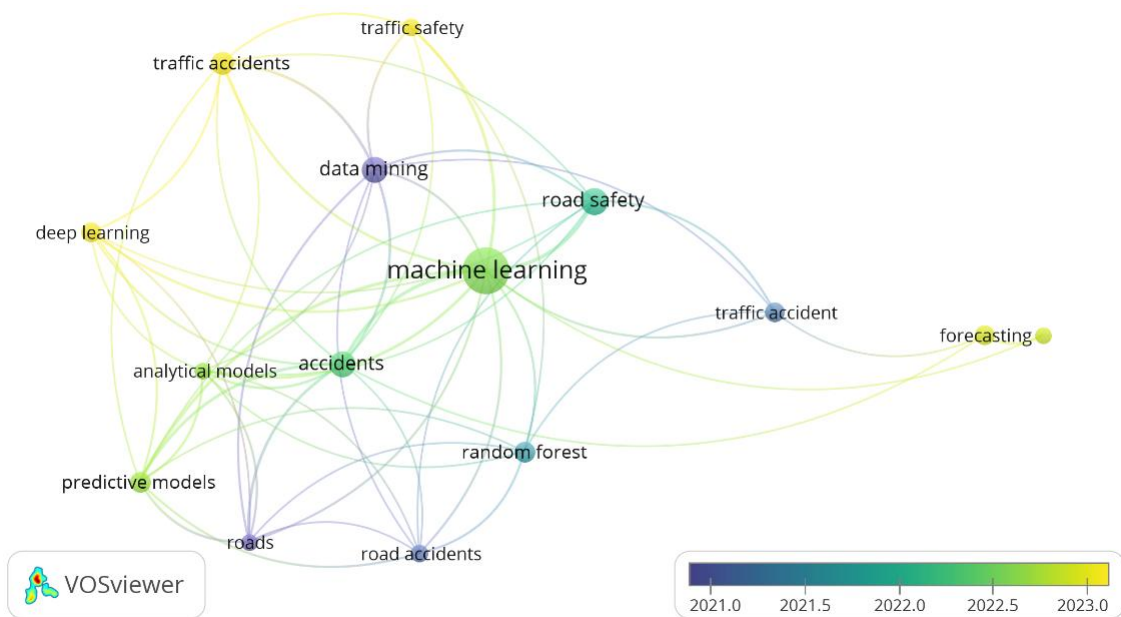
En la figura 4, se presenta la conexión y las relaciones entre las palabras clave descritas por los autores. Es posible evidenciar que destacan términos como “Machine Learning”, “Accidents”, “Predictive Models”, “Data Mining”, “Road Safety” y “Traffic Accidents” al presentarse con mayor frecuencia en los artículos analizados. Al ser una revisión exploratoria

de los factores, el relacionamiento entre palabras clave muestra coherencia con el propósito del estudio.

Es importante destacar la escala temporal analizada dentro de los resultados obtenidos en la figura 3, el cual abarca desde el año 2021 hasta el 2023. Dentro de esta escala se denota el aumento de interés investigativo asociado a las palabras “Traffic accidents”, “Traffic Safety” y “deeplearning”. Por otro lado, se puede evidenciar un crecimiento constante en el interés de palabras relacionadas a la minería de datos y modelos analíticos a lo largo de los años evaluados a la par de los relacionados con accidentes y siniestros viales.

**Figura 4**

*Relación entre Palabras Clave.*

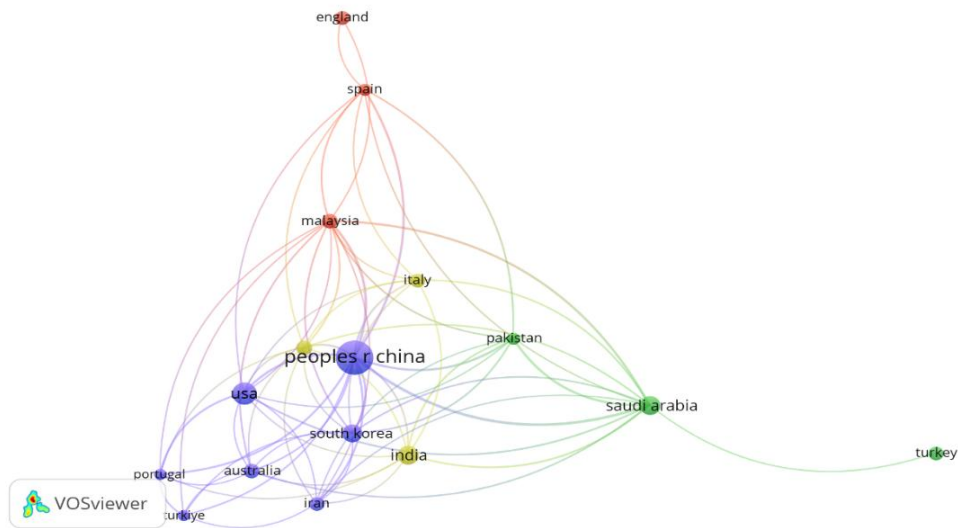


**4.2.4. Principales Países**

Basado en la Figura 5, se observa que países como China, destaca con 590 citaciones en un total de 40 de los artículos seleccionados, al igual que Estados Unidos, Portugal y Corea del Sur con 388, 352 y 137 citaciones correspondientemente.

**Figura 5**

*Países principales de los artículos arrojados por la ecuación de búsqueda.*



### 4.3. Análisis de Literatura

Dada la extensión de la información recopilada, en esta sección se presentan aspectos generales y elementos destacados de los estudios incluidos en la bitácora de investigación (ver Apéndice A), con el objetivo de ofrecer al lector una visión estructurada y útil.

La seguridad vial es una preocupación principal a nivel global, con un impacto significativo en la movilidad, la economía y la salud pública. Basado en la World Health Organization, (2023) los accidentes de tránsito son una de las principales causas de muerte y discapacidad en el mundo. Con más de 1.35 millones de muertes al año (Heydari et al., 2019).

A pesar de avances en tecnología, políticas de seguridad vial, diseño de infraestructura y tecnología vehicular enfocada en minimizar el impacto de accidentes, las tasas de accidentalidad se mantienen constantes, especialmente en países de bajos y medianos ingresos (Heydari et al., 2019) donde las deficiencias de los estándares viales y regulaciones agravan el problema.

Sin embargo, en los estudios realizados se evidencia variables de estudio similares a pesar de las diferencias contextuales y geográficas, pudiéndose generar una correlación de las variables determinantes en la frecuencia y severidad de los siniestros viales. Por ejemplo, el estudio realizado por Jeong et al., (2022) sobre accidentes de tráfico en Seúl, se propone una clasificación de los factores determinantes de los accidentes de tránsito los cuales son: factores humanos, factores ambientales y factores viales; como se evidenció en el estudio cada uno de estos factores influye en la posibilidad y severidad de los accidentes de tránsito, basado en los criterios de severidad establecidos por el estudio.

#### ***4.3.1. Factores Determinantes en la Ocurrencia y Severidad de Accidentes***

A partir de las revisiones, se puede establecer que los accidentes de tránsito son eventos influenciados por múltiples factores, dentro de los cuales se pueden destacar factores humanos, tecnológicos, ambientales y espaciales. Estos factores pueden ser agrupados de la siguiente manera:

- 1. Factores humanos:** Edad, género, experiencia del conductor, fatiga, uso del cinturón, consumo de alcohol y drogas. En particular, el manejo bajo efectos del alcohol y la fatiga tienen una relación con el aumento de la gravedad de los accidentes (Abuzinadah et al., 2024; Qi et al., 2024).
- 2. Factores vehiculares:** Dentro de este grupo se encuentra elementos como el tipo de vehículo, mantenimiento, año de fabricación y presencia de sistemas de seguridad activa y/o pasiva. (Farhangi et al., 2022; Hadjidimitriou et al., 2020).
- 3. Factores ambientales:** Condiciones meteorológicas, visibilidad, iluminación natural y artificial, hora del día (Jeong et al., 2022).
- 4. Factores viales y espaciales:** Tipo de vía, nivel de riesgo de la vía, presencia de intersecciones peligrosas, señalización y ubicación geográfica. (Ahmed et al., 2023; Farhangi et al., 2022)

A través del estudio realizado por Ahmed et al., (2023) se realizaron un análisis basado en modelos explicables de *machine learning*, a través del cual se pudieron relacionar variables como el tipo de vía, cantidad de vehículos involucrados y la categoría de las vías tienen una influencia directa sobre la severidad del accidente, dentro de su marco es posible tomar la caracterización establecida de variables, ya que comprende variables analizadas dentro de otros estudios revisados (Ver Figura 6).

**Figura 6**  
*Caracterización de Variables de Siniestros Viales*

Características Viales	Vehículo	Factores Humanos	Ambiente	Límites de Velocidad
Tipo de vía	Tipo de vehículo	Condición de licencia	Sol brillante	Límite de velocidad propuesto
Número de carriles	Área de daño	Tipo de licencia	Nublado	Límite de velocidad temporal
Carril de propósito especial	Uso del vehículo	Estado de la licencia	Crepúsculo	Límite de velocidad aconsejado
Características de la vía	Clasificación del vehículo (CC)	Licencia extranjera	Oscuridad	
Cruce de ferrocarril	Pasajeros delanteros totales	Fatiga	Lluvia	
Tipo de intersección	Pasajeros traseros totales	Consumo de drogas	Niebla o neblina	
Señalización de la vía	Velocidad inadecuada	Consumo de alcohol	Nieve	
Curvatura de la vía	Tipo de carga	Uso del cinturón de seguridad	Granizo o aguanieve	
Gradiente	Altura de la carga	Edad	Fino	
Tipo de superficie	Carga asegurada	Género	Viento	
Iluminación vial	Clasificación de seguridad	Etnicidad	Escarcha	
Tipo de barrera vial	Certificado de garantía	Estado mental		
Ubicación de la barrera	Tipo de permiso	Condición médica		
Categoría de la vía				

**Nota.** Adaptado de *Influencing Variables in Traffic Accidents*, de Ahmed et al., 2023, Transportation Research Interdisciplinary Perspectives.

#### 4.3.2. *Estrategias y Modelos Para la Gestión de la Seguridad Vial*

Diversos países han adoptado el enfoque de "sistema seguro", que reconoce que los errores humanos son inevitables, pero que el sistema vial debe estar diseñado para que estos errores no resulten en lesiones graves o muertes. Por ejemplo, Japón ha fijado una meta de reducir las muertes anuales por debajo de 2,300 para 2032; mientras que Dinamarca busca que no haya más de 90 muertes anuales (Ahmed et al., 2023).

En este contexto, se destacan estrategias como:

- **Fiscalización automatizada:** Uso de cámaras y sensores para control de velocidad y cruce de semáforos.
- **Educación y concienciación vial:** Campañas dirigidas a conductores y peatones.
- **Diseño vial seguro:** Incorporación de rotondas, reductores de velocidad, ciclovías, y zonas peatonales protegidas.
- **Tecnologías predictivas:** Herramientas de *big data* y aprendizaje automático para identificar puntos críticos y generar alertas.

En este contexto, se resalta la importancia de las tecnologías predictivas y sus metodologías se pueden destacar algunos modelos de manejo de data como los siguientes, siendo agrupados en modelos tradicionales, modelos de aprendizaje automático y modelos de inteligencia artificial.

Los modelos tradicionales se basan en la estadística convencional requiriendo una intervención humana considerable durante todo su proceso. En contraste, los modelos de aprendizaje automático automatizan parte de este proceso, reduciendo la necesidad de intervención humana durante el entrenamiento. Finalmente, los modelos de inteligencia artificial al ser una segmentación de los modelos de aprendizaje automático contienen la

capacidad de explicar modelos de alta complejidad con el propósito de facilitar la comprensión al usuario.

**Modelos Estadísticos Tradicionales.** Estos enfoques han sido útiles, pero su capacidad predictiva es limitada cuando se trabaja con grandes volúmenes de datos o relaciones no lineales entre variables, se destacan los siguientes con sus usos más comunes:

- **Regresión lineal:** Permite modelar la probabilidad de que ocurra un accidente grave en función de variables explicativas (Sahraei et al., 2022).
- **Modelos de Poisson y binomial negativo:** Apropriados para datos de conteo, como el número de accidentes por tramo o por día (Sangare et al., 2021).
- **Modelos multinivel:** Usados para capturar la variabilidad de los accidentes entre regiones o ciudades, como características viales o factores estructurales, permitiendo identificar factores anidados como el tipo de vía o políticas viales regionales (Congacha et al., 2020).

**Modelos de Aprendizaje Automático (Machine Learning).** A diferencia de los modelos tradicionales, presentan mejora respecto al manejo de grandes volúmenes de información, además de mejores resultados de precisión debido a la relación de variables no lineales entre sí, dentro de los cuales se destacan los siguientes:

- **Árboles de decisión y Random Forest:** Eficientes en clasificación de severidad y detección de variables críticas (Lee et al., 2020)
- **Support Vector Machines (SVM):** Utilizadas en escenarios donde se requiere alta precisión en clasificación binaria o multiclase, como nivel de lesión o severidad del accidente (Gatera et al., 2023).
- **Redes neuronales artificiales (ANN):** Capaces de procesar grandes volúmenes de datos con alta precisión, resultandos efectivos al momento de modelar la relación entre

los diferentes factores que toman lugar en los accidentes viales. Además de su capacidad para identificar patrones no lineales, permitiéndolos brillar en contextos de alta complejidad (Kushwaha & Abirami, 2023).

- **XGBoost y LightGBM:** Modelos de *boosting* ampliamente utilizados en predicción de accidentes de tránsito. Diversos estudios han demostrado su superioridad en tareas de clasificación y análisis de severidad (Dong et al., 2022a, 2022b; Li et al., 2022).

**Inteligencia Artificial Explicable (XAI)** La inteligencia artificial explicable refiere a un conjunto de métodos diseñados para hacer comprensibles los resultados de modelos complejos de inteligencia artificial, especialmente aquellos como las redes neurales profundas, ya que estos modelos pueden alcanzar niveles complicados para entender para los usuarios.

En el caso de herramientas como SHAP (*SHapley Additive exPlanations*) y LIME (*Local Interpretable Model-agnostic Explanations*) permiten interpretar el impacto de cada variable en la predicción final, esto es crucial para la toma de decisiones basada en evidencia facilitando la comprensión, por ejemplo, en el estudio de Ahmed et al. (2023) usaron SHAP para identificar qué variables tenían mayor peso en el modelo predictivo, mejorando la precisión y comprensibilidad del sistema.

En particular, el uso de SHAP permitió destacar que factores como tipo de vía y el número de vehículos involucrados eran determinantes clave en la predicción de gravedad, asimismo, permitió verificar interacciones específicas en variable tales como consumos en términos de drogas y la zona urbana/rural dando una perspectiva que enriqueció el análisis de los resultados

#### ***4.3.3. Análisis de los Modelos y sus Aplicaciones Dentro de la Literatura***

Dentro del desarrollo de esta revisión se lograron identificar algunos elementos de interés como la aplicación práctica de los modelos de predicción, el desempeño de algunos

modelos asociado al estudio de variables específicas, su predicción y precisión de los resultados, además de un análisis más detallado de factores influyentes. A través, de la revisión se destacan las siguientes aplicaciones, en las cuales se relacionan directamente las variables de estudio severidad y localización geográfica necesarias para el desarrollo de la investigación:

**Identificación de Zonas Críticas.** Técnicas de *clustering* como K-Means y DBSCAN han permitido localizar zonas urbanas con alta concentración de siniestros. Por ejemplo, Jeong et al. (2022) usaron estos modelos para analizar accidentes en Seúl, identificando intersecciones con mayor riesgo de lesiones graves. Esta información se ha usado para rediseñar semáforos, mejorar señalización y priorizar zonas para intervención policial.

**Predicción de Severidad de Accidentes.** Estudios como Hadjidimitriou et al. (2020) muestran que es posible predecir la severidad de un accidente con más del 90% de precisión utilizando solo 11 variables clave. En particular, estos modelos se han aplicado en el contexto de motocicletas, donde la separación del conductor y el vehículo representa un reto adicional para la clasificación del evento.

Otros estudios, como Malik et al. (2023) han explorado el uso de datos no estructurados (por ejemplo, tweets geolocalizados) para anticipar zonas de accidentes en tiempo real usando modelos de *deep learning* como Bi-LSTM con una precisión del 94.2%.

#### ***4.3.4. Aplicación de Diferentes Modelos y Estudio de Factores Relacionados a Siniestros Viales***

La precisión de los modelos de predicción ha superado el 95% en algunos casos gracias a modelos híbridos y arquitectura de redes profundas, se destaca el desempeño los modelos de *Machine Learning* frente a los estadísticos tradicionales en precisión predictiva. Por ejemplo, *Random Forest*, *LightGBM*, *XGBoost* y redes neuronales han mostrado mejoras del 10% al

25% en precisión respecto a modelos como regresión logística o Poisson (Jiang et al., 2020; Lee et al., 2020; Qi et al., 2024)

Por otro lado, modelos híbridos e inteligentes explicables (XAI) están ganando adopción. El uso de SHAP y LIME permite interpretar qué variables influyen más en la severidad, mejorando la confianza de los tomadores de decisiones (Ahmed et al., 2023; Cicek et al., 2023)

En el caso de modelos como *Random Forest Convolutional Neural Network* (RFCNN), *Bidirectional Long Short-Term Memory with Attention Mechanism* (BiLSTM-Attention), *Deep Learning SHapley Additive exPlanations* (DeepSHAP) y *Radial Basis Function Neural Network* (RBFNN) logran precisiones superiores al 95% al clasificar la severidad de un accidente en leve, grave o fatal (Malik et al., 2023; Pei et al., 2024; Sahraei et al., 2022). Mientras, algoritmos de *Deep Learning* como LSTM y RNN son útiles para predicción en tiempo real, especialmente en análisis de riesgo de accidentes con datos temporales o secuenciales (Gong et al., 2020; Pradhan & Ibrahim Sameen, 2020).

Otros estudios exploran el uso de análisis de texto y redes sociales permitiendo detectar zonas de riesgo en tiempo real. En el caso de Malik et al. (2023) lograron predecir ubicaciones de accidentes con 94.2% de precisión utilizando análisis de sentimientos en tweets geolocalizados.

Basado en esto, se puede establecer que los modelos más eficaces son aquellos que combinan precisión con interpretabilidad, además de facilidad para su adopción en diferentes contextos geográficos, permitiendo que los resultados proporcionen un entendimiento de la situación facilitando establecer medidas preventivas efectivas para la reducción de siniestros viales.

Como fue mencionado anteriormente, existen factores que coinciden dentro de las revisiones de los siniestros viales, dentro de estos factores se destacan los humanos debido a

que son las más influyentes, pero también los más difíciles de modificar, lo que justifica intervenciones mixtas (tecnológicas, educativas y regulatorias).

**Factores Humanos con Mayor Influencia.** Dentro de este factor, se tiene en cuenta la edad del conductor como un consistentemente predictor significativo. Conductores jóvenes (menores de 24 años) y mayores (mayores de 65 años) presentan tasas más altas de accidentes graves (Abuzinadah et al., 2024; Megnidio-Tchoukouegno & Adedeji, 2023).

Asimismo, el género masculino está asociado con mayor severidad, aunque las diferencias tienden a moderarse cuando se ajustan variables como tipo de vía y hora del accidente.

Por otro lado, factores como fatiga del conductor es uno de los factores humanos más críticos, sobre todo en zonas rurales o vías de baja iluminación, en el caso de la somnolencia predice con alta confiabilidad los accidentes nocturnos (Farhangi et al., 2022)

En el caso de uso del cinturón de seguridad y la presencia de bolsas de aire, disminuyen en más del 40% la probabilidad de lesiones graves o fatales, según análisis explicables (SHAP) según los estudios (Cicek et al., 2023; Pei et al., 2024), es importante aclarar que este factor tiene una característica híbrida, ya que depende de la existencia de estos elementos además de la disposición de usarla.

**Factores del Entorno y la Vía.** El tipo de vía, dividiéndolas entre rurales y urbanas, además del número de carriles impactan significativamente en la severidad. Zonas rurales concentran más muertes debido a velocidades más altas y menor tiempo de respuesta de emergencia (Qi et al., 2024; Sahraei et al., 2022) .

Asimismo, la presencia de intersecciones y señalización deficiente está altamente correlacionada con accidentes múltiples o en cadena, especialmente en contextos urbanos densos (Milenković & Glavić, 2018)

En el caso de condiciones meteorológicas, tales como lluvia, nieve, niebla o viento están asociadas a mayor riesgo de siniestros graves, pero su efecto es menor que el del factor humano (Jeong et al., 2022).

**Factores Vehiculares.** Dentro de este factor la principal segmentación establecida como tipo de vehículo es determinante, las motocicletas y vehículos de dos ruedas tienen la tasa más alta de letalidad por siniestro, debido a la exposición directa del conductor (Lee et al., 2020).

Asimismo, el número de ocupantes influye en el tipo y severidad del accidente. Vehículos con mayor cantidad de pasajeros presentan mayor riesgo de lesiones múltiples (Ahmed et al., 2023).

Relacionado con los factores de existencia y uso de elementos de seguridad preventivos como cinturones y bolsas de aire, como se mencionó previamente conllevan una disminución de más del 40% sobre la probabilidad de lesiones graves o fatales, basado en los análisis desarrollados con modelos SHAP (Cicek et al., 2023; Pei et al., 2024).

**Factores Contextuales y Espaciales.** Para la distribución geográfica de los siniestros muestra patrones repetitivos dentro de las intersecciones, entradas o salidas de autopistas y zonas escolares concentrando eventos con lesiones graves (Jeong et al., 2022).

En el caso de accidentes en clima adverso o baja visibilidad, se intensifican en temporadas específicas como el invierno. Sin embargo, pueden mitigarse con sistemas de alerta y reducción de velocidad dinámica (Jiang et al., 2020).

#### **4.4. Análisis de Literatura Complementario**

Con el propósito de dar un análisis más detallado de los elementos encontrados en la literatura se propone el desarrollo de esta sección de manera narrativa detallando en los tipos de predicción abordados por los estudios, los factores usados como elementos de predicción y

los modelos de minería de datos aplicados, adicionalmente de la relación de los tópicos anteriores entre sí.

#### 4.4.1. Tipo de predicción abordado

Dentro de la revisión de literatura, se clasificaron los estudios basados en la variable de predicción principal buscada teniendo en cuenta las siguientes características.

- **Severidad:** Aquellos modelos que buscan la predicción de nivel o tipo de lesión durante el accidente, usualmente clasificadas como leve, grave o fatal), sin embargo, también se englobaron estudios en los cuales se clasifican como fatal o no fatal
- **Frecuencia:** Aquellos modelos que se especializan en la determinar la cantidad de accidentes en un lugar o periodo de tiempo determinado.
- **Ambas:** Aquellos modelos que integran los dos componentes anteriores, teniendo como objetivo determinar la cantidad de accidentes en determinado lugar y la severidad de estos.

Basado en los criterios descritos anteriormente, se muestra la siguiente segmentación en los artículos revisados (Ver Tabla 3)

**Tabla 3**

*Segmentación de Artículos Basado en Tipo de Predicción*

<b>Criterio</b>	<b>Cantidad</b>	<b>Porcentaje</b>
<i>Severidad</i>	83	79.6%
<i>Frecuencia</i>	16	15.5%
<i>Ambos</i>	5	4.9%
Total	103	100%

*Nota.* Se excluyeron dos artículos debido a que eran revisiones bibliométricas y no un estudio aplicado que evaluara modelos en un ejercicio de campo

Se evidencia una tendencia hacia los modelos enfocados en la predicción de severidad, esto puede deberse a varios factores destacando la disponibilidad de datos etiquetados con niveles de lesión, facilitando el proceso de entrenamiento de modelos al tener datos previamente etiquetados. Por otro lado, modelos de frecuencia, requieren datos temporales y espaciales con diferentes niveles de agregación como semana, vía o zona.

Asimismo, modelos que combinan ambos factores usualmente se encuentran desarrollados en contextos con una mayor robustez de información disponiendo de múltiples fuentes de datos, en este contexto, es fundamental una revisión de los modelos aplicados y su construcción para la predicción de dichas variables.

#### ***4.4.2. Clasificación de Modelos Encontrados en la Literatura***

Dentro de la revisión de literatura y tomando como marco de referencia la clasificación propuesta por Faisal et al., (2025) donde se propone un marco de definición de los modelos de aprendizaje automático, fueron posteriormente clasificados los modelos propuestos.

A través de este marco se definieron modelos con necesidad de datos etiquetados para su entrenamiento como modelos de aprendizaje supervisado, modelos sin necesidad de datos etiquetados como no supervisados y modelos híbridos como aquellos que pueden ser clasificado como alguno de los dos anteriores basado en el entrenamiento y enfoque de la investigación. Adicionalmente, por su relevancia dentro del contexto de comprensión e interpretabilidad de los modelos, se mencionan aquellos que integran métodos de Inteligencia Artificial Explicable (XAI), obteniendo los siguientes resultados expuestos en la Tabla 4.

**Tabla 4***Segmentación de Artículos Basado en Tipo de Modelo*

<b>Clasificación</b>	<b>Cantidad</b>	<b>Porcentaje</b>
<i>Modelos supervisados</i>	126	72.4%
<i>Modelos No supervisados</i>	35	20.1%
<i>Modelos híbridos</i>	9	5.2%
<i>Modelos XAI</i>	3	1.7%
<b>Total</b>	<b>173</b>	<b>100%</b>

A partir de estos hallazgos se permite visualizar la predominancia de modelos supervisados para dentro de la muestra de literatura tomada, esto se puede deber a diferentes factores, sin embargo, no se encontró una razón concluyente sobre esta tendencia, estudios recientes como Behboudi et al., (2024) muestra un patrón, esto se puede deber a la existencia de bases de datos gubernamentales etiquetadas y a la facilidad para evaluar el desempeño del modelo.

Dentro de este análisis se destacan los siguientes modelos al ser los más recurrentes dentro de los artículos revisados (Ver Tabla 5), siendo los 10 modelos más empleados en los artículos analizados segregados basado su objetivo de predicción y clasificación. Se evidencia la dominancia de modelos supervisados siendo *Random Forest*, *Decision Tree* y *Logistic Regression* los tres más frecuentes, teniendo un enfoque en la predicción de gravedad, mientras modelos no supervisados, como el algoritmo *Apriori* o *LightGBM*, son menos frecuentes dentro de este enfoque.

**Tabla 5***Relación de los modelos más empleados y su objetivo de predicción*

<b>Modelo</b>	<b>Tipo</b>	<b>Gravedad</b>	<b>Frecuencia</b>	<b>Ambos</b>	<b>Total</b>
<i>Random Forest</i>	Supervisado	33	3	2	38
<i>Decision Tree</i>	Supervisado	23	2	2	27
<i>Logistic Regression</i>	Supervisado	20	0	2	22
<i>Support Vector Machine</i>	Supervisado	15	1	0	16
<i>XGBoost</i>	Supervisado	13	3	0	16
<i>Naive Bayes</i>	Supervisado	9	0	2	11
<i>Apriori Algorithm</i>	No Supervisado	9	0	1	10
<i>K-Nearest Neighbors</i>	Supervisado	10	0	0	10
<i>LightGBM</i>	No Supervisado	8	1	0	9
<i>Classification and Regression Trees</i>	Supervisado	8	0	0	8

Una vez clasificados los modelos de datos más usados en la literatura y los objetivos de predicción de estos, es pertinente el desarrollo de los factores usados como insumo en el proceso de entrenamiento de dichos modelos, por su pertinencia en el proceso de propuesta de un marco de implementación.

#### ***4.4.3. Clasificación de Factores Encontrados en la Literatura***

Una vez sistematizados los estudios, se inició el proceso de análisis de factores. Sin embargo, debido a la gran dispersión en la denominación de los factores y en los métodos utilizados para cuantificarlos, que varían según el modelo, se optó por realizar una clasificación basada en categorías temáticas, con el propósito de facilitar su análisis comparativo. Es importante resaltar que, para esta sección, las variables no son excluyentes y pueden presentarse simultáneamente varias en un mismo artículo o modelo.

La clasificación usada toma como referencia el estudio realizado por Ahmed et al., (2023) donde se agrupan los factores contribuyentes a siniestros viales dentro de 6 categorías principales:

- **Características viales:** Agrupando atributos físicos de las vías, como tipo y cantidad de carriles, intersecciones.
- **Características del Vehículo:** Incluye características técnicas y operativas del vehículo, como el tipo, clasificación de seguridad.
- **Factores Humanos:** Comprende atributos relacionados al comportamiento y caracterización de los ocupantes del vehículo como edad, género, estado mental, entre otros.
- **Factores Ambientales:** Relacionado a las condiciones ambientales como visibilidad, lluvia, viento, entre otros.
- **Límites de velocidad**
- **Otros:** Se incluyen relacionadas con la temporalidad y el uso del suelo, como el día de la semana, hora, si es feriado o día laboral,

A partir de los hallazgos dentro del proceso de revisión, se identificó la necesidad de desagregar la categoría "Otros" para obtener mayor precisión en el análisis posterior. Debido a esto, se evaluaron dos categorías adicionales a las iniciales:

- **Factores geográficos:** Incorporan elementos espaciales como el tipo de zona (urbana/rural), presencia de zonas escolares o vías de alta velocidad.
- **Variables temporales:** Incluyen información como el día de la semana, hora del día, mes o temporada, que pueden influir significativamente en la ocurrencia de accidentes.

La tabla 6 se muestra la frecuencia de las categorías anteriores dentro de los 103 artículos revisados, excluyendo las dos investigaciones no relacionadas directamente con un modelo experimental y los resultados asociados a su desempeño.

**Tabla 6***Frecuencia de categorías de factores en los artículos*

<b>Factores</b>	<b>Frecuencia (n)</b>	<b>Porcentaje</b>
<i>Características Viales</i>	84	81.6%
<i>Características del vehículo</i>	61	59.2%
<i>Factores Humanos</i>	63	61.2%
<i>Factores Ambientales</i>	79	76.7%
<i>Límite de Velocidad</i>	27	26.2%
<i>Factores Geográficos</i>	27	26.2%
<i>Variables Temporales</i>	73	70.9%
<i>Otros</i>	47	45.6%
<b>Total</b>	<b>103</b>	<b>100%</b>

Los resultados evidencian que los factores más utilizados dentro del contexto de predicción de gravedad y frecuencia de accidentes viales corresponden a las características viales (81.6%), factores ambientales (76.7%) y variables temporales (70.9%). Esta tendencia indica la orientación de los estudios hacia la comprensión del entorno y temporalidad en la que tienen ocurrencia los siniestros viales, siendo coherente con las disponibilidades de datos que suelen estar presentes en los registros gubernamentales y mediciones de tráfico.

De la misma manera, los factores humanos (61.2%) y las características de los vehículos (59.2%) también son elementos de interés para la comprensión de los siniestros, al integrar componentes conductuales y técnicos dentro de los modelos de análisis. Por otro lado, la presencia menos frecuente del límite de velocidad y factores geográficos (26.2%) puede deberse a la mutabilidad con la que estos factores son definidos dependiendo del contexto territorial y la dificultad para estructurar modelos complejos a niveles más específicos como zonas, carreteras o regiones cuya disponibilidad de datos es menor o no homogénea.

Dentro de la categoría otros (45.6%) se integraron variables adicionales que por su especificidad, no se ajustaban directamente a las clasificaciones revisadas, dentro de las cuales se detallan elementos como las condiciones del accidente como tipo de colisión, variables

legales, características del evento como la presencia de fuego, extracción de los sobrevivientes, variables no estructuradas y factores derivados o contruidos a partir de los sets de data usados como estadísticos o indicios compuestos.

#### 4.4.4. *Uso de Tipo de Categorías de Factores Según el Tipo de Predicción*

Los resultados mostrados en la Tabla 7 muestran las tendencias de uso de factores insumo para los modelos de predicción abordados. Como se evidencia en la Tabla 3, de los 103 artículos revisados, el 79.6% se enfocan en la predicción de los accidentes, el 15.5% en la predicción de frecuencia de los accidentes y el 4.9% abordan ambas predicciones de manera simultánea.

**Tabla 7**

*Tabla de comparación factores categorizados por tipo de predicción*

<b>Factores</b>	<b>Gravedad (79.6%)</b>	<b>Frecuencia (15.5%)</b>	<b>Ambos (4.9%)</b>	<b>Total</b>
<i>Características Viales</i>	71	8	5	84
<i>Características del vehículo</i>	53	3	5	61
<i>Factores Humanos</i>	57	2	4	63
<i>Factores Ambientales</i>	67	7	5	79
<i>Límite de Velocidad</i>	22	4	1	27
<i>Factores Geográficos</i>	19	6	2	27
<i>Variables Temporales</i>	60	8	5	73
<i>Otros</i>	38	8	1	47
<b>Total Artículos</b>	<b>82</b>	<b>16</b>	<b>5</b>	<b>103</b>

En términos de factores utilizados para los modelos de predicción de gravedad, se presenta una recurrencia en el uso de factores viales (71), ambientales (67), factores temporales (60) y temporales (57), esto sugiere un enfoque centrado en el contexto inmediato del siniestro y los factores comportamentales del conductor, estas categorías proporcionan información clave para comprender los móviles detrás de la severidad de los siniestros viales.

Por otro lado, la predicción de la frecuencia de accidentes presenta mayores limitaciones, ya que los factores de incidencia tienden a concentrarse en variables relacionadas con las en factores viales (8), temporales (8), ambientales (7) y otros factores (8), esta tendencia puede estar estrechamente relacionada a la necesidad de caracterizar patrones temporales generales, donde los detalles del conductor o el vehículo no tiene una relevancia completa. Sin embargo, la cantidad de factores bajo la segmentación de otros, presentan una gran variedad de datos que pueden ser complementarios para el desarrollo de estos modelos.

En el caso de los estudios que abordan ambas predicciones de manera simultánea, aunque en menor número presentan una combinación diversa de factores, lo que es proporcional a la complejidad metodológica y necesidad de modelos que integren una comprensión más detallada tanto de las tendencias temporales como de las variables que indiquen en la gravedad de los siniestros.

Adicionalmente se propone un análisis del uso de factores basado en el tipo de modelo previamente clasificado (Ver Tabla 8).

**Tabla 8**

*Tabla de comparación factores categorizados por tipo de modelo de predicción*

<b>Factores</b>	<b>Supervisado</b>	<b>No supervisado</b>	<b>Híbridos</b>	<b>XAI</b>
<i>Características Viales</i>	99	31	8	3
<i>Características del vehículo</i>	76	23	6	2
<i>Factores Humanos</i>	68	21	6	1
<i>Factores Ambientales</i>	93	26	4	2
<i>Límite de Velocidad</i>	27	8	5	2
<i>Factores Geográficos</i>	46	9	1	2
<i>Variables Temporales</i>	82	26	7	1
<i>Otros</i>	66	17	6	0
<b>Modelos Evaluados</b>	<b>126</b>	<b>35</b>	<b>9</b>	<b>3</b>

Los resultados evidencian una predominancia en los modelos supervisados hacia el uso de la mayoría de los tipos de factores, siendo coherente con la disponibilidad de datos

etiquetados para entrenamiento. En términos detallados, los factores viales (99), ambientales (93) y temporales (82) son los más empleados en los estudios donde se emplean modelos supervisados, lo que muestra una orientación hacia la comprensión del contexto físico y situacional del siniestro. Por otro lado, factores humanos (68) y de vehículo (76) también tienen una presencia significativa, demostrando una importancia en el proceso de predicción de gravedad.

Por su parte, los modelos no supervisados e híbridos tienen una cobertura equilibrada en el uso de los diferentes factores. Esto se debe a el propósito subyacente de estos modelos de generar relaciones latentes o patrones de datos no directos, por lo que su aplicación puede resultar en la necesidad de disponer mayor cantidad de datos para la comprensión de dichas relaciones en las diferentes situaciones viales.

Por su parte modelos de incorporación técnicas de Inteligencia Artificial (*XAI*), aunque escasos en números, muestran una tendencia parcial al uso de diversos factores con énfasis en aspectos viales, ambientales y geográficos, esto sugiere la necesidad de interpretabilidad dentro de variables estructurales más que en factores subjetivos como los de vehículos o humanos.

#### ***4.4.5. Análisis cruzado entre tipos de predicción, modelo y factores considerados***

El análisis cruzado entre los tipos de variable y los modelos según su objetivo de predicción muestra patrones claros en el caso de la predicción de gravedad, donde sobresale el uso de modelos como Random Forest (33), Decision Tree (23), Logistic Regression (20) y Support Vector Machine (17). Debido a la extensión de la matriz comparativa de variables elaborada para abordar esta sección, se optó por presentar una descripción narrativa de los hallazgos. Cabe señalar que dicha matriz mantiene la característica mencionada en la sección anterior de variables no excluyentes entre modelos o artículos (ver Apéndice B).

A nivel general, se observa una distribución homogénea en el uso de factores con predominios de los denominados en factores viales, ambientales y variables temporales, esto reflejando la importancia del contexto físico y temporal al determinar la severidad de los siniestros viales.

- *Random Forest* Y *Decision Tree* tienen un comportamiento similar en la evaluación de variables, donde los factores más considerados son factores viales (17–18%), ambientales (17–15%), lo cual es consistente con su capacidad para desarrollar múltiples variables categóricas y jerárquicas.
- *Support Vector Machine* tiene una mayor proporción hacia las características viales (19%) y ambientales (16%), lo cual coincide con su uso común en escenarios donde se busca maximizar la separación entre clases con variables estructurales.
- *Logistic Regression*, al ser un modelo más interpretativo, muestra una distribución balanceada, sin embargo, tiene una ligera reducción en el uso de factores humanos y geográficos, posiblemente por restricciones en simplicidad del modelo.

En el caso de los modelos enfocados a la predicción de frecuencia, no se hace evidente una preferencia clara en el uso de algoritmo. Sin embargo, se destaca el uso de *Random Forest*, *XGBoost* y *Decision Tree*, cada uno siendo utilizado en al menos dos o más estudios. Esta diversidad, puede deberse a la exploración de técnicas que puedan moldear series temporales, recurrencias espaciales o patrones estructurales de datos.

En términos de los factores observados, se presenta un comportamiento heterogéneo entre los modelos. Esto podría deberse al tamaño reducido de la muestra de estudios con este enfoque (16 estudios), lo que limita la posibilidad de establecer tendencias precisas. No obstante, en los modelos más empleados se identificó una tendencia hacia la incorporación de

variables geográficas y temporales en mayor proporción. En particular, el modelo *Random Forest*, a diferencia de su comportamiento en los modelos orientados a gravedad, muestra un uso más homogéneo de los factores observados, al igual que el modelo *Decision Tree*.

De manera general, en los modelos evaluados dentro de los estudios se observa un promedio de integración de 17% de los factores viales, 16% de factores ambientales, 15% de los factores geográficos y 16% de factores temporales. Adicionalmente, destaca la participación en la categoría de otros con un 20%, esto puede deberse a la necesidad de ahondar en variables no enmarcadas directamente dentro de las categorías propuestas para comprender de mejor manera la dinámica dentro de los modelos de frecuencia.

Por su parte los modelos relacionados con la predicción de frecuencia y gravedad de manera simultánea presentan un conjunto metodológico más completo, compuesto en este caso por una muestra de 5 artículos. Dicha complejidad se ve reflejada en la diversidad de algoritmos incluyendo enfoques tradicionales como *Random Forest*, *Decision Tree* y *Logistic Regression*, a la par de técnicas no supervisadas como *DBSCAN*, *A priori* y diferentes modelos de redes neuronales como *backpropagation*, *feedforward*, *recurrent*, entre otras.

Debido a las características multivariadas de este enfoque, se observa un uso proporcional y homogéneo de los factores, distribuidos entre factores viales (17 %), vehiculares (17 %), humanos (13 %), ambientales (17 %) y temporales (17 %), tomando en menor medida factores como límite de velocidad (4 %) y factores geográficos (8 %).

De manera general, se puede indicar que los modelos con predicción combinada no solo amplían el rango de algoritmos explorados, sino que también presentan una cobertura más amplia y equilibrada de factores explicativos. Esto podría deberse a la mayor necesidad de información para abordar dos objetivos predictivos de distinta naturaleza.

## **5. Recomendaciones para un Marco de Aplicación Efectivo**

A continuación, se contemplan recomendaciones para la implementación de modelos de minería de datos en la prevención de accidentes de tránsito con el propósito de proponer un marco de aplicación basado en los resultados observados dentro de la revisión de literatura.

### **5.1. Implementación Según Tipo de Predicción**

La primera consideración que se debe tener en el desarrollo de un marco de aplicación es definir de manera asertiva el objetivo de predicción, dentro de los estudios abordados se evidenciaron tres grandes enfoques el estudio de la frecuencia, gravedad de los siniestros y la predicción simultanea de ambas variables.

#### **5.1.1. Estudio de Frecuencia**

A pesar de que diferentes estudios desarrollen un mismo enfoque predictivo, la definición de este puede ser variable dentro del grupo. Por ejemplo, se puede estimar la cantidad de total de accidentes en un periodo de tiempo anual o mensual (Andrejiova, 2024; Meißner & Rieck, 2022), donde se evalúa la predicción de posibles accidentes futuros con base en datos históricos segmentados en periodos de tiempo, permitiendo desarrollar o reforzar recursos para la contingencia o prevención con anterioridad.

Asimismo, es posible desarrollar la frecuencia enfocada directamente en zonas específicas considerando las características propias de dicho territorio (Charm et al., n.d.; de Medrano & Aznarte, 2021; Farhangi et al., 2022; Wang et al., 2022), explorando segmentos de datos con alta siniestralidad y evaluación de modelos de regresión, obteniendo la posibilidad de intervenir factores de riesgos y elementos con alta incidencia dentro del modelo.

Adicionalmente, es posible desarrollar un enfoque basado en eventos o condiciones especiales como fines de semana, periodos vacacionales o condiciones meteorológicas, (Gorzelanczyk et al., 2023; Uğuz & Büyükgökoğlan, 2022) lo que muestra un enfoque en

variables contextuales y temporales, que habiliten la interpretación fluctuante de los siniestros, esto complementado con la posibilidad de ajustar los recursos y cuerpos de respuesta para contener estacionalidades.

### **5.1.2. Estudio de Gravedad**

Los hallazgos relacionados con la caracterización de gravedad presentan una mayor mutabilidad en comparación que la definición de la gravedad. Siendo abordada en diferentes niveles dependiendo de la disponibilidad de información para la estructuración de los modelos y la forma en la que estos presentan la información.

Principalmente clasificaciones de fatalidad siendo expresadas en clasificaciones binarias de fatalidad o no fatalidad. Se observó que particularmente los estudios desarrollados con datos históricos con mayor estructuración permitieron el desarrollo de modelos con mayor capacidad de caracterización de gravedad, siendo un ejemplo los que usaron el set de datos del Reino Unido (Esenturk et al., 2022; Ma et al., 2021), donde se encuentran registros multiclase de leve, grave y fatal, sin embargo, es fundamental disponer de información balanceada para la implementación de modelos multiclase. Adicionalmente, estos pueden representar un reto en el proceso de interpretación.

### **5.1.3. Estudio simultaneo de frecuencia y gravedad**

En el cuerpo de literatura evaluado, se encuentra un uso más limitado de este enfoque. Sin embargo, un desarrollo más sustancial y con una posible contribución más alta en la mitigación del desarrollo de accidentes viales, así como las consecuencias asociadas, este enfoque no solo permite evaluar la ocurrencia temporal de los posibles accidentes, sino que adicionalmente, caracteriza la gravedad asociada lo cual permite establecer un sistema de alertas tempranas.

Al igual que en los casos anteriores, se presenta una diversidad de aplicación metodológica, dentro de la cual se puede destacar estudios como el expuesto por Raja et al.

(2023) que se centra en la implementación de modelos de redes neuronales profundas para la predicción de la frecuencia, mientras que utilizan arquitecturas complejas de redes neuronales artificiales para clasificar la severidad, dentro de lo cual se obtiene una visibilidad no solo de los siniestros, sino de niveles de daño asociados a las víctimas y propiedades relacionadas, tomando como base datos temporales y del entorno.

Por su parte Kazmi et al., (2022) presentan un enfoque particular, evaluando y delimitando zonas clasificadas como críticas donde se emplean elementos clasificadores para la estimación de la gravedad del siniestro, con este enfoque se abarca la posibilidad de abordar el riesgo de accidentalidad, a través de variables espaciales y temporales.

Otros estudios adicionales mencionados previamente (Santos et al., 2021; Wang et al., 2022), demuestran la necesidad de integraciones de diferentes tipos de información, así como calidad asociada a dicha información, además de una selección de variables de predicción conjunta, cuya complejidad y valor agregado es resaltante.

## **5.2.Implementación Según Tipo Modelo**

Un factor fundamental para la implementación es determinar el modelo adecuado, dentro analizado es uno de los factores más variables y con mayor importancia para la obtención de resultados satisfactorios, ya que esta decisión representa comprender las capacidades técnicas y la información que se puede disponer, dependiendo del tipo de predicción que se desee explorar.

Dentro de lo encontrado en la literatura y la clasificación propuesta esta segmentación se puede evaluar agrupando los modelos supervisados, no supervisados, híbridos y aquellos que incorporen técnicas de Inteligencia Artificial Explicable (XAI).

### **5.2.1. Modelos Supervisados**

Están presentes en el 72.4% de los estudios abordados dentro de esta revisión, siendo la categoría predominante dentro de la literatura. Esta presencia puede estar directamente

relacionada a la disponibilidad de datos etiquetados recopilados por entidades o cuerpos gubernamentales, además de su aptitud metodológica para la evaluación del desempeño (Behboudi et al., 2024).

Este conjunto de técnicas engloba modelos y algoritmos como *Random Forest*, *Decision Tree* y *Logistic Regression*, con un amplio margen de aplicación en problemas de predicción de gravedad, donde resulta pertinente cuando se cuenta con estructuras claras en los sets de data y etiquetados de gravedad asociados o conteos históricos cuando se desea enfocar en la predicción de frecuencia.

La implementación de este enfoque de modelos parte de la disponibilidad de datos con alto nivel de integridad y una estructura adecuada, en orden de permitir una mejor capacidad predictiva, sino también una comprensión más profunda de variables relacionadas a los siniestros. En contextos donde no se cuente con bases de datos con un nivel adecuado de estructuración, puede requerir un esfuerzo adicional para corregir o recopilar esta información de manera correcta.

### **5.2.2. Modelos No Supervisados**

Por su parte, los modelos no supervisados tienen una presencia del 20.1% de los estudios abordados, siendo su característica principal el no requerimiento de datos etiquetado para su entrenamiento, esto les permite ser utilizados en contextos con una menor disponibilidad de datos.

Se destaca la aplicación de enfoques como el algoritmo *Apriori* y *LightGBM*, en busca de establecer patrones ocultos entre las variables analizadas, además de permitir la agrupación de datos con niveles de similitud, su aplicación dentro de la literatura suele estar relacionado con la segmentación de zonas de riesgo y comprensión de factores contextuales de los siniestros.

No obstante, es importante mencionar su limitación en la capacidad de predicción de variables complejas, como la frecuencia y la gravedad de los accidentes viales, siendo más fructíferos en la identificación de variables incidentes o en la caracterización de zonas clave con el propósito de determinar tendencias.

Asimismo, su efectividad y su eficacia es subjetiva a la calidad de la información usada para su proceso de entrenamiento, pueden ser una alternativa viable para el inicio de procesos de caracterización de siniestralidad, como base para la aplicación de modelos supervisados con mayor nivel de especificidad y precisión en relación con variables de estudio puntuales.

### ***5.2.3. Modelos Híbridos***

En el caso de los modelos que combinan modelos supervisados y no supervisados se encontró una representación menor en la literatura (5.2%), este enfoque presenta una capacidad prometedora, permitiendo adaptar el modelo según la naturaleza de los datos disponibles, siendo particularmente funcional en contextos con información parcial o con restricciones en la disponibilidad de datos etiquetados.

A pesar de su flexibilidad, también presentan una mayor necesidad técnica y una comprensión más detallada de las limitaciones de cada componente, Es fundamental considerar el flujo de integración de modelos, en el cual los modelos no supervisados pueden emplearse como herramientas exploratorias del contexto clasificando y generando relaciones preliminares que brinden una estructura más sólida al momento de aplicar modelos supervisados, que requieren una mayor comprensión de las variables utilizadas. Dentro de la literatura se observó un enfoque en escenarios metodológicos complejos, donde se buscaba extraer la mayor cantidad de conocimiento de bases de datos heterogéneas.

### ***5.2.4. Modelos de Inteligencia Artificial Explicable***

Su aplicación a lo largo de los diferentes estudios fue la más baja; sin embargo, presenta un componente significativo en el proceso de comprensión e interpretabilidad, siendo

particularmente útil en procesos que involucran modelos de aprendizaje profundo, generando una comprensión más profunda de las variables fundamentales en la predicción de severidad en los accidentes de tránsito. Particularmente, la capacidad de aceptación del modelo por parte de los usuarios técnicos y no técnicos, como entidades gubernamentales, responsables de políticas y organismos de seguridad vial.

Sin embargo, su adopción conlleva desafíos técnicos como redes de arquitectura de datos complejas, con una subyacente capacidad computacional dependiente no solo de los recursos disponibles, sino de mano de obra especializada en procesos de esta índole, para poder obtener el mayor retorno posible de la inversión que conlleva su implementación.

### **5.3. Implementación Según Tipo Categorías de Factores**

En relación con los factores estudiados, se componen de la selección y disponibilidad de la información explicativa, representando un elemento crítico para el marco de implementación de modelos de minería de datos como insumo de análisis y prevención de accidentes viales. Dentro del proceso realizado se optó por una clasificación en ocho categorías para englobar aquellos predominantes en la literatura, permitiendo una visión complementaria de las preferencias relacionadas con las diferentes metodologías.

Dentro de los hallazgos generales se puede resaltar factores relacionados con el contexto físico como las características viales, características ambientales y en relación con la temporalidad teniendo la más alta abundancia dentro de la literatura, siendo de alto valor en el momento de la implementación de técnicas con el objetivo de predecir la gravedad, dentro de su uso se observó una dominancia de registros estructurados provenientes de observatorios viales y entes gubernamentales asociados con la revisión de accidentalidades o la eficiencia de los sistemas viales, siendo una fuente confiable de información.

De la misma forma, los factores humanos y características del vehículo se presentaron como un elemento de estudio, permitiendo la exploración de información conductual y

caracterizaciones técnicas, siendo elementos de valor para modelos orientados a la predicción de gravedad con énfasis en determinar severidades de impacto, permitiendo captar diferentes riesgos asociados a atributos humanos o mecánicos. Sin embargo, es importante aclarar que la información utilizada dentro de estos marcos comprendía un alto nivel de granularidad y especificidad.

Por su parte factores relacionados a los límites de velocidad y factores geográficos, presentaron una menor frecuencia dentro de la literatura, esto puede estar relacionado a los diferentes contextos geográficos y la no homogeneidad de información, en particular de la segunda categoría, donde se engloban elementos geospaciales que debido a su especificidad pueden no tenerse disponibles en todos los lugares. Estos elementos fueron particularmente útiles en el proceso de desarrollar o reforzar modelos localizados, permitiendo el diseño de medidas e intervenciones particulares a nivel de estructura vial.

La categoría de otros excedió la expectativa debido a su alta aparición dentro de los estudios abordados, siendo usado numerosos datos no convencionalmente estructurados, abriendo la posibilidad de explorar con mayor profundidad estos elementos como insumo de los procesos de implementación. Dentro de los más llamativos se encontraron elementos como variables derivadas como índices compuestos, datos de reportes de características narrativas o registros no estructurados, cuya adopción requiere capacidades específicas dentro del diseño y pueden aportar al desarrollo de modelos exploratorios o enriquecer otras facultades del modelo.

Es importante resaltar que en función del tipo de predicción y de los modelos elegidos, se puede incorporar diferentes factores. En el caso de modelos de gravedad se observó un marco de factores asociados fuertemente a la situación y el desarrollo en un contexto específico, con el propósito de caracterizar las condiciones en las que se ocurrió el siniestro y las consecuencias asociados. En el caso de los modelos de frecuencia, su tendencia está regida

a variables temporales y geográficas facilitando reconocimiento de patrones agregados con una tasa mayor de presencia de accidentes viales.

Dentro de un proceso eficiente de implementación es fundamental contemplar la pertinencia de cada factor en relación con el modelo y los objetivos determinados, delimitando los datos con base en su calidad y disponibilidad. Asimismo, la integración de fuentes de información complementarias para ampliar la capacidad de predictiva del modelo. Para ello resulta indispensable un análisis de integridad, confiabilidad, estructura y capacidad de mantenimiento a lo largo del tiempo, buscando una integración de datos recientes para evitar una obsolescencia o disminución en la predicción a raíz de la falta de insumos para reflejar el contexto vial de estudio.

## **6. Marco de Gestión para la Implementación de Modelos Predictivos en Seguridad Vial Municipal**

Como parte del desarrollo de los objetivos de investigación, se procede a definir un marco de gestión eficaz para la seguridad vial municipal, integrando los modelos de minería de datos para la mitigación y reducción de los siniestros viales, tomando como referencia los elementos encontrados en la revisión de literatura.

El desarrollo de un proceso de integración eficaz de estas tecnologías en contextos municipales requiere un proceso metodológico que permita adaptar las capacidades técnicas y disponibilidad de información local a las necesidades específicas de la seguridad vial. A continuación, se propone una ruta estructurada, para facilitar el proceso de toma de decisiones e implementación.

### **6.1. Diagnóstico de Información Disponible.**

El primer paso para alcanzar una implementación exitosa de modelos de predictivos como recurso en la seguridad vial es la correcta identificación y evaluación de la información disponible, acotando las posibilidades metodológicas y técnicas que podrán adoptarse como

solución, el diagnóstico permite visualizar fortalezas y oportunidades de mejora, siendo una línea de partida para el posterior desarrollo de modelos acorde a la realidad de información.

Esta sección servirá de insumo para la elección de modelo y tipo de predicción, estableciendo una línea base de caracterización de variables y requerimientos de complementación o transformación.

### ***6.1.1. Identificación y Recopilación***

De manera general, proceso debe iniciar con el propósito de inventariar las fuentes de datos existentes y accesibles, en relación con accidentes de tránsito, información de infraestructura y datos meteorológicos. Dentro de las fuentes más usadas en la literatura, se incluyen:

- Registros oficiales de siniestros de tránsito, usualmente reportados por organismos de tránsito o cuerpos gubernamentales.
- Datos meteorológicos históricos de centros de observación.
- Información que incluya trazado de vías, señalización, límites de velocidad e infraestructura.

Estos datos pueden ser históricos usados como insumo para comprender los flujos históricos, sin embargo, la disponibilidad de datos recientes abre la posibilidad de tener mejores resultados de predicción.

### ***6.1.2. Clasificación Temática de Variables***

Para facilitar el análisis, puede resultar conveniente organizar las variables en categorías que reflejan su naturaleza y su posible utilidad para la predicción de gravedad o frecuencia de accidentes de tránsito, como fue realizado durante la revisión se proponen la siguiente clasificación (Ver Tabla 9).

**Tabla 9**

*Factores propuestos para durante el proceso de valoración.*

<b>Factores</b>	<b>Descripción</b>
<i>Características viales</i>	Agrupando atributos físicos de las vías, como tipo y cantidad de carriles, intersecciones.
<i>Características del Vehículo</i>	Incluye características técnicas y operativas del vehículo, como el tipo, clasificación de seguridad.
<i>Factores Humanos</i>	Comprende atributos relacionados al comportamiento y caracterización de los ocupantes del vehículo como edad, género, estado mental, entre otros.
<i>Factores Ambientales</i>	Relacionado a las condiciones ambientales como visibilidad, lluvia, viento, entre otros.
<i>Factores geográficos</i>	Incorporan elementos espaciales como el tipo de zona (urbana/rural), presencia de zonas escolares o vías de alta velocidad.
<i>Variables temporales</i>	Incluyen información como el día de la semana, hora del día, mes o temporada, que pueden influir significativamente en la ocurrencia de accidentes.

A través de caracterización se obtiene una mayor visibilidad de relaciones causales o correlacionales para la elección de variables objetivo, permitiendo comprender los elementos de interés para la integración de esta metodología.

### **6.1.3. Evaluación de la Calidad y Confiabilidad de los Datos**

Una vez que se tiene una base de clasificación de fuentes y el tipo de información contenida en ellas, es esencial limitar la cantidad con base en calidad y confiabilidad de dicha información, con esto en mente, la siguiente revisión puede agregar valor al proceso:

En el ámbito de la frecuencia y vigencia de la información, comprender la periodicidad de actualización de la información, evaluando si al integrar esta fuente de información como insumo del modelo, ¿Será posible obtener esta información actualizada con frecuencia?

Otro elemento que presenta un valor agregado es la granularidad de la información, en caso de contener información de eventos, ¿Qué nivel de granularidad resulta conveniente para el planteamiento deseado? Además de si los datos que se van a usar cuentan con el mismo nivel de detalle entre ellos.

Con respecto a la consistencia y confiabilidad, es fundamental comprender el nivel de estandarización de las variables en las fuentes de información, como punto de partida examinar incongruencias o ausencia de datos claves en los registros, mostrando falencias en el proceso de levantamiento y mantenimiento de dichas fuentes, habilitando el desarrollo de elementos de compensación o integración de fuentes complementaras para solventar estas.

Con lo anterior, se espera determinar qué tan viable es cada tipo de predicción y qué ajustes a nivel metodológico pueden resultar prácticos o necesarios.

#### ***6.1.4. Identificación de Vacíos y Estrategias de Mitigación***

Como parte de este proceso de diagnóstico, se proponen incluir los vacíos de información como elemento limitante de la capacidad y calidad de los modelos a implementar, planteándose estrategias para mitigar el impacto de variables no estandarizadas, ausentes o parcialmente ausentes que son de interés para el desarrollo del estudio.

Para esto, se sugieren algunos métodos de integración tales como la integración de fuentes externas como aplicaciones móviles de tránsito, datos de sensores o plataformas colaborativas de reporte ciudadano, siendo fuentes con un bajo costo de implementación.

Por otra parte, es posible a su vez evaluar la factibilidad de estudios de campo para la captación de variables de manera directa o indirectas, siendo una fuente de información primaria con un nivel de granularidad basada en los intereses de revisión.

Otro enfoque que puede resultar pragmático es la articulación interinstitucional con organismos en cuyas actividades generan o dispongan de información de valor, tales como, entes hospitalarios y aseguradoras vehiculares, que puede complementar o sustituir elementos faltantes en las fuentes gubernamentales principales.

## 6.2. Definición del Enfoque Predictivo

A partir del análisis anterior, teniendo una mayor comprensión de la información disponible, sus limitaciones y fortalezas, es necesario definir un enfoque de predicción pertinente.

### 6.2.1. Selección del Tipo de Predicción

La selección debe partir del tipo de fenómeno que se desea anticipar, en función de las necesidades institucionales, capacidades de respuesta y la proyección de intervención. Dentro de la literatura los tres enfoques con las necesidades de aplicación predominantes fueron los siguientes (Ver Tabla 10).

**Tabla 10**

*Consideraciones basadas en objetivo de predicción*

<b>Objetivo</b>	<b>Descripción</b>	<b>Requerimientos Típicos</b>
<i>Frecuencia</i>	Predicción del número de accidentes en un espacio/tiempo determinado	Datos agregados de las variables temporales y contextuales
<i>Gravedad</i>	Clasificación de la severidad asociada a los siniestros viales registrados	Datos etiquetados por nivel de gravedad, variables humanas, del entorno y del vehículo
<i>Predicción Simultánea</i>	Combinación de las dos anteriores, respondiendo a la cuantificación y la severidad de los siniestros.	Conjunto integral de los datos, modelos complejos y mayor nivel de procesamiento

### 6.2.2. Criterios para Elegir un Enfoque Adecuado

A manera de planteamiento, la selección puede tomar como referencia las siguientes preguntas como guía para el proceso de elección.

**Tabla 11***Preguntas para selección de enfoque de predicción*

<b>PREGUNTAS</b>
¿Cuál es la principal necesidad/preocupación del territorio: la frecuencia de los accidentes o su severidad?
¿Existen datos históricos con el suficiente nivel de detalle y etiquetado para modelar gravedad?
¿Se dispone de la información geográfica, contextual y temporal para el modelar frecuencia?
¿Cuál es la capacidad institucional para responder al escenario de riesgo identificado a partir del modelo?

### **6.2.3. Enfoque Escalonado y Adaptativos**

Es importante aclarar que la elección del enfoque no resulta excluyente, es posible explorar el proceso como un contexto de maduración donde se puede tomar una visión escalonada, comenzado con modelos de frecuencia identificando zonas críticas, integrando posteriormente modelos de gravedad que permitan priorizar los recursos según el nivel de riesgo de los resultados salida de los modelos.

Permitiendo una toma de decisiones a nivel táctico como el despliegue de elementos gubernamentales para dar cubrimiento a posibles siniestros, instalación de elementos de seguridad como medida preventiva en la gestión, así como una visión estratégica en la planificación urbana, el diseño de campañas preventivas para la reducción de factores con alta presencia en el desarrollo de accidentes o cambios normativos atacando elementos recurrentes.

### **6.3. Selección del Modelo de Minería de Datos**

Una vez claro el enfoque de predicción y los datos disponibles, lo siguiente es seleccionar el modelo más adecuado para alcanzar los objetivos definidos. Esta fase es

fundamental, la ejecución dependerá de la capacidad del sistema para entregar resultados confiables, adaptables y de utilidad para el contexto municipal.

### **6.3.1. Valoración de Condiciones Técnicas y Operativas**

De manera similar al proceso de las fuentes de información, resulta de utilidad comprender las capacidades operacionales existentes, así como sus oportunidades de mejora para la implementación deseada.

- **Infraestructura Computacional:** Capacidad de procesamiento, almacenamiento y disponibilidad de software
- **Capacidad Técnica:** Resulta indispensable realizar una autoevaluación del nivel de formación y experiencia del equipo de trabajo en áreas como ciencia de datos, gestión de bases de datos y estadística.
- **Flujo de información:** Frecuencia de actualización, accesibilidad e interoperabilidad entre las fuentes de datos seleccionadas.

La comprensión de estos factores permite trazar planes de suficiencia alineando las capacidades técnicas con la visión y expectativas de la implementación, destinando recursos específicos para la nivelación de infraestructura y capacidades del personal involucrado en el desarrollo.

### **6.3.2. Criterios de Selección del Modelo**

La selección del modelo es una tarea crucial en función del tipo de predicción, datos y capacidad disponibles para el desarrollo, algunas de las consideraciones generales obtenidas a partir de los elementos revisados en la literatura (Ver Tabla 12).

**Tabla 12***Consideraciones para la selección de un modelo*

<b>Tipo de modelo</b>	<b>Modelos frecuentes</b>	<b>Características principales</b>	<b>Aplicabilidad sugerida</b>
<b><i>Supervisados</i></b>	Random Forest, Decision Tree, Logistic Regression, XGBoost, SVM, Naive Bayes, KNN	Requieren datos etiquetados para el proceso de entrenamiento	Predicción de gravedad, estimación de frecuencia con datos históricos.
<b><i>No Supervisados</i></b>	Apriori, K-Means, DBSCAN, LightGBM	Caracterización de patrones latentes. Útiles para agrupación, segmentación y análisis exploratorio.	Identificación de zonas críticas, agrupación de patrones de siniestralidad.
<b><i>Híbridos</i></b>	Deep Learning con componentes supervisados/no supervisados	Combinan características de ambas categorías	Modelos conjuntos de gravedad y frecuencia, o detección y clasificación con un mayor índice de complejidad.

Es importante resalta que el uso de modelos de Inteligencia Artificial Explicable (XAI), no fue incluida, debido a que su funcionalidad no se relaciona directamente con la predicción de una de las variables. Sin embargo, resulta de gran valor al establecer modelos complejos cuya comprensión de procesamiento puede resultar en un reto, su implementación amplia la interpretabilidad de los resultados, permitiendo una transparencia del desempeño del modelo.

Debido a la naturaleza de esta sección y a la variabilidad presente entre los modelos, no se proporcionará una descripción detallada de los procesos de entrenamiento; no obstante, se recomienda mantener un proceso metodológico estructurado que permita validar los resultados y verificar su capacidad de interpretación dentro del contexto. Asimismo, es fundamental contar con documentación relacionada con configuraciones, métricas y observaciones, que garantice la trazabilidad y facilite la implementación de futuras actualizaciones del modelo.

#### **6.4. Aplicaciones Estratégicas de los Modelos Predictivos**

A través de la implementación de los modelos predictivos explorados los entes gubernamentales y organismos de control de la seguridad vial obtienen más insumos para la toma de decisiones basado en datos, resultando en medidas preventivas y correctivas de los sucesos viales.

Dentro de esta sección se busca abordar algunos de los posibles beneficios resultantes de la integración de modelos predictivos en la seguridad vial y ejes de gestión que pueden ser abordados.

### ***6.4.1. Priorización de Zonas y Momentos Críticos***

Dentro de las ventajas derivadas de esta integración se encuentra la capacidad de identificar y priorizar de manera efectiva zonas de alto interés por alta presencia de accidentes viales, además de periodos temporales con un mayor índice de ocurrencia de accidentes viales, estos análisis pueden resultar en procesos más efectivos de administración enfocados en optimización de recursos y elementos de respuesta con base en datos resultantes de las predicciones.

En este sentido, la aplicación de modelos de frecuencia puede tener una tasa de utilidad mayor, definiendo zonas de alta siniestralidad y siendo reforzado con detección de patrones en factores espaciotemporales, siendo un insumo de valor para la administración eficaz de recursos. Asimismo, análisis de temporalidad asociada con condiciones climáticas y festividades, franjas horarias o días de la semana, resulta en la posibilidad de ajustar variables de respuesta de acuerdo a una estructura dinámica como asignación de personal de tránsito en función de horarios y zonas, desarrollo de campañas de control localizadas, intervenciones físicas como refuerzos de la señalización, reductores de velocidad e incluso despliegues operativos viales, mejorando la eficiencia institucional de los organismos asociados.

#### ***6.4.2. Intervención Focalizada de Factores de Riesgo***

A través de identificación de patrones que presentan un alto índice en la ocurrencia y severidad, es posible abordar estratégicamente el diseño de medidas de intervención enfocadas y focalizadas a las condiciones de un entorno específico.

Con base en los hallazgos de patrones encontrados se dispone de información contextualizada y concreta, permitiendo la priorización eficaz de recursos en el proceso de modificación e intervención de infraestructura, como intersecciones mal diseñadas, zonas con ausencia de señalización, iluminación o curvas cerradas. Estos patrones pueden ser de insumo para elaborar medidas o ajustes normativos detallados para grupos específicos, por ejemplo, conductores en un rango de edad que se desplazan en motocicletas, cuando los resultados del modelo asocien riesgo en estos perfiles.

Adicionalmente, la estructuración de un sistema que permita comprender estos patrones permite el desarrollo de valoraciones de la efectividad de las políticas de manera objetiva, estimando si la intervención específica presenta un índice de mejoría sobre el escenario base de estudio, permitiendo una visibilidad detallada que no recae completamente en indicadores agregados de carácter administrativo, complementado con la capacidad de retroalimentar los modelos con nuevos datos en tiempo real o en ciclos periódicos.

## 7. Conclusiones

La revisión de literatura exploratoria de tipo narrativo sobre técnicas de minería de datos en el contexto de la seguridad vial permitió evidenciar que esta es un insumo potencial de alto valor para el análisis y la predicción de accidentes de tránsito, particularmente en relación con su severidad y frecuencia. Se identificaron patrones subyacentes, factores incidentes en el desarrollo de accidentes viales y metodologías recurrentes aplicadas en diversos contextos, junto con sus respectivas variaciones.

A partir de la consolidación de la información recopilada, fue posible establecer tendencias en la literatura respecto al uso de modelos supervisados, como *Random Forest*, *Decision Tree* y *Logistic Regression*, asociados a la disponibilidad de datos etiquetados. Estos modelos demostraron un alto rendimiento en la predicción de la gravedad de los siniestros, siendo dicho rendimiento proporcional a la calidad de la información utilizada.

Aunque con menor frecuencia, la predicción simultánea se presenta como un enfoque con un potencial significativo para el desarrollo de sistemas de alerta e intervención focalizada preventiva, permitiendo estimar tanto un rango de tiempo para la ocurrencia del suceso como la comprensión de sus consecuencias y ajustando recursos para respuestas más eficaces.

Por otra parte, fue posible determinar los factores predominantes dentro de la predicción de la frecuencia y severidad asociada de los accidentes como características viales, vehiculares, temporales, geográficos y contextuales son clave para el desarrollo de estas herramientas predictivas, elementos adicionales como la disponibilidad de información son dependientes del objetivo de estudio, sin embargo, resulta clave comprender las fortalezas y carencias de las fuentes previo al proceso de implementación.

En el marco propuesto se propone la implementación de los modelos predictivos en contextos municipales a través de una estrategia progresiva, partiendo del diagnóstico de la

información disponible, definición de los objetivos de predicción de manera clara y la selección de modelos en concordancia con la disponibilidad de información.

La integración de los modelos de minería de datos como facilitadores de la toma de decisiones a nivel de administración vial conlleva oportunidades en muchos niveles, como la identificación de zonas y aspectos críticos en el desarrollo de los siniestros, permitiendo la toma de decisiones basadas en la comprensión profunda de estos fenómenos, esto se traduce en una gestión más eficiente de los recursos, la planificación de las políticas preventivas, además de la evaluación de los impactos de las medidas en relación de la seguridad vía, contribuyendo a la reducción de los siniestros y reducción del impacto de los mismos.

Es importante establecer que, a pesar de su potencial también conlleva desafíos y limitaciones dentro de las cuales la más grande recae sobre la accesibilidad a información de calidad, confiable y estandarizada que presente información pertinente para desarrollo de este tipo de herramientas, dentro de los desafíos también se encuentra generar interconexión entre los resultados y su aplicabilidad en la seguridad vial, así como la interoperabilidad entre instituciones para habilitar marcos de gestión más robustos apoyados en la minería de datos.

## **8. Recomendaciones**

Con base en los hallazgos obtenidos a lo largo del presente trabajo, se proponen una serie de recomendaciones orientadas en fortalecer y facilitar la implementación de modelos de minería de datos en el ámbito de la seguridad vial municipal.

En primer lugar, la puesta en marcha de este tipo de iniciativas conlleva un esfuerzo organizacional, donde acciones como consolidar procesos de estructuración y estandarización de las fuentes de datos gubernamentales, así como su mantenimiento periódico representa una gran parte del éxito en la adopción de estas tecnologías, con énfasis en elementos como integridad, cobertura y precisión de los registros donde se cuente con información actualizada

de manera sistemática, facilita que la implementación se centren establecer un modelo predictivo duradero y que genere un valor agregado tanto para las instituciones que lo adopten como para los habitantes de los municipios donde se apliquen.

De la misma manera, se recomienda iniciar con modelos de menor complejidad como aquellos con un enfoque de frecuencia, los cuales pueden establecer lineamientos base para el avance progresivo hacia enfoques multivariados que integren predicción de gravedad o simultáneos. Esta puesta en marcha a través de manera progresiva permite fortalecer y solventar falencias en las capacidades institucionales, validar hipótesis sobre los siniestros locales y optimizar la formación de estos nuevos procesos.

También se considera relevante promover la adopción de modelos interpretables, a través de herramientas de inteligencia artificial explicable y visualización de resultados, con el propósito de facilitar la interpretación para los equipos técnicos como para los encargados de la gestión, traduciendo los resultados y encuentros complejos a nivel técnico a un lenguaje que reduzca la complejidad para comprender los hallazgos y justificando las intervenciones futuras.

Finalmente, se recomienda invertir en formación técnica y en el fortalecimiento en la infraestructura de datos, ya que la sostenibilidad de los sistemas predictivos dependerá en gran medida de la existencia de recursos humanos que puedan darle mantenimiento y uso, así como la disponibilidad de tecnologías adecuadas para el almacenamiento, procesamiento y visualización de los datos.

### Referencias bibliográficas

- Abuzinadah, N., Aljrees, T., Chen, X., Umer, M., Ibrahim Aboulola, O., Tahir, S., Abdulmajid Eshmawi, A., Alnowaiser, K., & Ashraf, I. (2024). Improving Traffic Accident Severity Prediction Using Convolved Features and Decision-Level Fusion of Models. *Transportation Research Record*, 2678(8), 731–744. <https://doi.org/10.1177/03611981231220656>
- Ahmed, S., Hossain, M. A., Ray, S. K., Bhuiyan, M. M. I., & Sabuj, S. R. (2023). A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance. *Transportation Research Interdisciplinary Perspectives*, 19. <https://doi.org/10.1016/j.trip.2023.100814>
- Andrejiova, M. (2024). Application of a time series to analyse the evaluation of road traffic accidents in Slovakia. *Acta Logistica*, 11(1), 119–129. <https://doi.org/10.22306/al.v11i1.467>
- Angarita-Zapata, J. S., Maestre-Gongora, G., & Calderín, J. F. (2021). A bibliometric analysis and benchmark of machine learning and automl in crash severity prediction: The case study of three colombian cities. *Sensors*, 21(24). <https://doi.org/10.3390/s21248401>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2019). *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*. <http://arxiv.org/abs/1910.10045>
- Behboudi, N., Moosavi, S., & Ramnath, R. (2024). *Recent Advances in Traffic Accident Analysis and Prediction: A Comprehensive Review of Machine Learning Techniques*. <http://arxiv.org/abs/2406.13968>

- Cai, Q. (2020). Cause Analysis of Traffic Accidents on Urban Roads Based on an Improved Association Rule Mining Algorithm. *IEEE Access*, 8, 75607–75615. <https://doi.org/10.1109/ACCESS.2020.2988288>
- Castillo, D., Coral, C., & Salazar Méndez, Y. (2020). Econometric Modeling of Traffic Accidents in Ecuador. *Revista Politecnica*, 46(2), 21–28. <https://doi.org/10.33333/rp.vol46n2.02>
- Charm, T., Assistant, G. R., Wang, H., Zuniga-Garcia, N., Ahmed, M., Kockelman, K. M., & Greer, D. (n.d.). *PREDICTING CRASH OCCURRENCE AT INTERSECTIONS IN TEXAS: AN OPPORTUNITY FOR MACHINE LEARNING*.
- Cicek, E., Akin, M., Uysal, F., & Topcu Aytas, R. M. (2023). Comparison of traffic accident injury severity prediction models with explainable machine learning. *Transportation Letters*, 15(9), 1043–1054. <https://doi.org/10.1080/19427867.2023.2214758>
- Congacha, A. E., Espinoza, Lady, Barba, L., & Morales, C. H. (2020). Pronóstico de siniestros viales en el Ecuador basado en descomposición de valores singulares multinivel - MSVD. *3C Tecnología\_Glosas de Innovación Aplicadas a La Pyme*, 45–63. <https://doi.org/10.17993/3ctecno/2020.v9n4e36.45-63>
- de Medrano, R., & Aznarte, J. L. (2021). A New Spatio-Temporal Neural Network Approach for Traffic Accident Forecasting. *Applied Artificial Intelligence*, 35(10), 782–801. <https://doi.org/10.1080/08839514.2021.1935588>
- Dong, S., Khattak, A., Ullah, I., Zhou, J., & Hussain, A. (2022a). Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. *International Journal of Environmental Research and Public Health*, 19(5). <https://doi.org/10.3390/ijerph19052925>

- Dong, S., Khattak, A., Ullah, I., Zhou, J., & Hussain, A. (2022b). Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. *International Journal of Environmental Research and Public Health*, 19(5). <https://doi.org/10.3390/ijerph19052925>
- Esenturk, E., Wallace, A. G., Khastgir, S., & Jennings, P. (2022). Identification of Traffic Accident Patterns via Cluster Analysis and Test Scenario Development for Autonomous Vehicles. *IEEE Access*, 10, 6660–6675. <https://doi.org/10.1109/ACCESS.2021.3140052>
- Faisal, A., Jhanjhi, N., Ashraf, H., Ray, S. K., & Ashfaq, F. (2025). *A Comprehensive Review of Machine Learning Models: Principles, Applications, and Optimal Model Selection*. <https://doi.org/10.36227/techrxiv.174285687.71966152/v1>
- Farhangi, F., Sadeghi-Niaraki, A., Nahvi, A., & Razavi-Termeh, S. V. (2022). Spatial modelling of accidents risk caused by driver drowsiness with data mining algorithms. *Geocarto International*, 37(9), 2698–2716. <https://doi.org/10.1080/10106049.2020.1831626>
- Gatera, A., Kuradusenge, M., Bajpai, G., Mikeka, C., & Shrivastava, S. (2023). Comparison of random forest and support vector machine regression models for forecasting road accidents. *Scientific African*, 21. <https://doi.org/10.1016/j.sciaf.2023.e01739>
- Ghandour, A. J., Hammoud, H., & Al-Hajj, S. (2020). Analyzing factors associated with fatal road crashes: A machine learning approach. *International Journal of Environmental Research and Public Health*, 17(11). <https://doi.org/10.3390/ijerph17114111>

- Gong, Y., Abdel-Aty, M., Yuan, J., & Cai, Q. (2020). Multi-Objective reinforcement learning approach for improving safety at intersections with adaptive traffic signal control. *Accident Analysis and Prevention*, 144. <https://doi.org/10.1016/j.aap.2020.105655>
- Google Inc. (2024). *Clasificación: Exactitud, recuperación, precisión y métricas relacionadas*. [https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?utm\\_source=chatgpt.com&hl=es-419](https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?utm_source=chatgpt.com&hl=es-419)
- Gorzelańczyk, P., Kalina, T., Jurkovič, M., & Mohanty, M. (2023). Forecasting the number of road accidents in Poland by day of the week and the impact of pandemics and pandemic-induced changes. *Cogent Engineering*, 10(1). <https://doi.org/10.1080/23311916.2023.2185956>
- Hadjidimitriou, N. S., Lippi, M., Dell'amico, M., & Skiera, A. (2020). Machine Learning for Severity Classification of Accidents Involving Powered Two Wheelers. *IEEE Transactions on Intelligent Transportation Systems*, 21(10), 4308–4317. <https://doi.org/10.1109/TITS.2019.2939624>
- Heydari, S., Hickford, A., McIlroy, R., Turner, J., & Bachani, A. M. (2019). Road safety in low-income countries: State of knowledge and future directions. In *Sustainability (Switzerland)* (Vol. 11, Issue 22). MDPI. <https://doi.org/10.3390/su11226249>
- Jeong, H., Kim, I., Han, K., & Kim, J. (2022). Comprehensive Analysis of Traffic Accidents in Seoul: Major Factors and Types Affecting Injury Severity. *Applied Sciences (Switzerland)*, 12(4). <https://doi.org/10.3390/app12041790>
- Jiang, F., Yuen, K. K. R., Ming Lee, E. W., & Ma, J. (2020). Analysis of run-off-road accidents by association rule mining and geographic information system techniques

- on imbalanced datasets. *Sustainability (Switzerland)*, 12(12).  
<https://doi.org/10.3390/SU12124882>
- Kazmi, S. S. A., Ahmed, M., Mumtaz, R., & Anwar, Z. (2022). Spatiotemporal Clustering and Analysis of Road Accident Hotspots by Exploiting GIS Technology and Kernel Density Estimation. *Computer Journal*, 65(2), 155–176.  
<https://doi.org/10.1093/comjnl/bxz158>
- Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. In *Journal of the Royal Society of Medicine* (Vol. 96, Issue 3, pp. 118–121). Royal Society of Medicine Press Ltd.  
<https://doi.org/10.1258/jrsm.96.3.118>
- Kushwaha, M., & Abirami, M. S. (2023). Intelligent Model for Avoiding Road Accidents Using Artificial Neural Network. *International Journal of Computers, Communications and Control*, 18(5). <https://doi.org/10.15837/ijccc.2023.5.5317>
- Kuyumcu, Z. C., Aslan, H., & Yurtay, N. (2023). Identifying Interrelated Factors of Fatal and Injury Traffic Accidents Using Association Rules. *Turkish Journal of Civil Engineering*, 34(5), 55–79. <https://doi.org/10.18400/tjce.1322965>
- Lee, J., Yoon, T., Kwon, S., & Lee, J. (2020). Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms: Seoul city study. *Applied Sciences (Switzerland)*, 10(1). <https://doi.org/10.3390/app10010129>
- Li, K., Xu, H., & Liu, X. (2022). Analysis and visualization of accidents severity based on LightGBM-TPE. *Chaos, Solitons and Fractals*, 157.  
<https://doi.org/10.1016/j.chaos.2022.111987>
- Ma, Z., Mei, G., & Cuomo, S. (2021). An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. *Accident Analysis and Prevention*, 160. <https://doi.org/10.1016/j.aap.2021.106322>

- Malik, N., Altaf, S., Tariq, M. U., Ahmed, A., & Babar, M. (2023). A Deep Learning Based Sentiment Analytic Model for the Prediction of Traffic Accidents. *Computers, Materials and Continua*, 77, 1659–1615. <https://doi.org/10.32604/cmc.2023.040455>
- Maza, E. (2023). *Aplicación de técnicas de agrupamiento para caracterizar patrones de siniestros viales en Ecuador en el año 2021*.
- Megnidio-Tchoukouegno, M., & Adedeji, J. A. (2023). Machine Learning for Road Traffic Accident Improvement and Environmental Resource Management in the Transportation Sector. *Sustainability (Switzerland)*, 15(3). <https://doi.org/10.3390/su15032014>
- Meißner, K., & Rieck, J. (2022). Strategic planning support for road safety measures based on accident data mining. *IATSS Research*, 46(3), 427–440. <https://doi.org/10.1016/j.iatssr.2022.06.001>
- Milenković, M., & Glavić, D. (2018). Analysis of Relations Between Freeway Geometry and Traffic Characteristics on Traffic Accidents. In *Lecture Notes in Networks and Systems* (Vol. 28, pp. 539–548). Springer. [https://doi.org/10.1007/978-3-319-71321-2\\_48](https://doi.org/10.1007/978-3-319-71321-2_48)
- Monroy, S., & Díaz, H. (2018). *Modelo de predicción de gravedad de accidentes de tránsito: un análisis de los siniestros en Bogotá, Colombia*.
- Observatorio Nacional de Seguridad Vial. (2023). *Reporte de Accidentes de Tránsito*.
- Organización Mundial de la Salud. (2023). *Traumatismos causados por el tránsito*. <https://www.who.int/es/news-room/fact-sheets/detail/road-traffic-injuries>
- Pei, Y., Wen, Y., & Pan, S. (2024). Road Traffic Accident Risk Prediction and Key Factor Identification Framework Based on Explainable Deep Learning. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3451522>

- Pérez-Sala, L., Curado, M., Tortosa, L., & Vicent, J. F. (2023). Deep learning model of convolutional neural networks powered by a genetic algorithm for prevention of traffic accidents severity. *Chaos, Solitons and Fractals*, 169. <https://doi.org/10.1016/j.chaos.2023.113245>
- Pradhan, B., & Ibrahim Sameen, M. (2020). Modeling Traffic Accident Severity Using Neural Networks and Support Vector Machines. In *Advances in Science, Technology and Innovation* (pp. 111–117). Springer Nature. [https://doi.org/10.1007/978-3-030-10374-3\\_9](https://doi.org/10.1007/978-3-030-10374-3_9)
- Qi, Z., Yao, J., Zou, X., Pu, K., Qin, W., & Li, W. (2024). Investigating Factors Influencing Crash Severity on Mountainous Two-Lane Roads: Machine Learning Versus Statistical Models. *Sustainability*, 16(18), 7903. <https://doi.org/10.3390/su16187903>
- Raja, K., Kaliyaperumal, K., Velmurugan, L., & Thanappan, S. (2023). Forecasting road traffic accident using deep artificial neural network approach in case of Oromia Special Zone. *Soft Computing*, 27(21), 16179–16199. <https://doi.org/10.1007/s00500-023-08001-6>
- Rezashoar, S., Kashi, E., & Saeidi, S. (2024). A hybrid algorithm based on machine learning (LightGBM-Optuna) for road accident severity classification (case study: United States from 2016 to 2020). *Innovative Infrastructure Solutions*, 9(8). <https://doi.org/10.1007/s41062-024-01626-y>
- Rodríguez, D. Y., Fernández, F. J., & Acero Velásquez, H. (2003). Road traffic injuries in Colombia. *Injury Control and Safety Promotion*, 10(1–2), 29–35. <https://doi.org/10.1076/icsp.10.1.29.14119>

- Sahraei, M. A., Çodur, M. K., Çodur, M. Y., & Tortum, A. (2022). Forecasting the Accident Frequency and Risk Factors: A Case Study of Erzurum, Turkey. *Tehnicki Vjesnik*, 29(1), 190–199. <https://doi.org/10.17559/TV-20200620164552>
- Sangare, M., Gupta, S., Bouzefrane, S., Banerjee, S., & Muhlethaler, P. (2021). Exploring the forecasting approach for road accidents: Analytical measures with hybrid machine learning. *Expert Systems with Applications*, 167. <https://doi.org/10.1016/j.eswa.2020.113855>
- Santos, D., Saias, J., Quaresma, P., & Nogueira, V. B. (2021). Machine learning approaches to traffic accident analysis and hotspot prediction. *Computers*, 10(12). <https://doi.org/10.3390/computers10120157>
- Singh, G., Pal, M., Yadav, Y., & Singla, T. (2020). Deep neural network-based predictive modeling of road accidents. *Neural Computing and Applications*, 32(16), 12417–12426. <https://doi.org/10.1007/s00521-019-04695-8>
- Uğuz, S., & Büyükgökoğlan, E. (2022). A Hybrid CNN-LSTM Model for Traffic Accident Frequency Forecasting During the Tourist Season. *Tehnicki Vjesnik*, 29(6), 2083–2089. <https://doi.org/10.17559/TV-20220225141756>
- Wang, F., Wang, J., Zhang, X., Gu, D., Yang, Y., & Zhu, H. (2022). Analysis of the Causes of Traffic Accidents and Identification of Accident-Prone Points in Long Downhill Tunnel of Mountain Expressways Based on Data Mining. *Sustainability (Switzerland)*, 14(14). <https://doi.org/10.3390/su14148460>
- World Health Organization. (2023). *Global status report on road safety 2023*.