

PREDICCIÓN DE ACCIDENTES VIALES EN CARTAGENA, COLOMBIA, CON ÁRBOLES DE DECISIÓN Y REGLAS DE ASOCIACIÓN

HOLMAN OSPINA-MATEUS Y LEONARDO
AUGUSTO QUINTANA JIMÉNEZ
Departamento de Ingeniería Industrial, Pontificia
Universidad Javeriana, Bogotá, Colombia

RESUMEN

El objetivo principal de esta investigación es predecir los factores asociados con la severidad en los accidentes viales de Cartagena (Colombia), la metodología está basada en técnicas de minería de datos como arboles de decisión (J48) y reglas de asociación (soporte, confianza, Lift). La investigación fue desarrollada con 10.053 registros de accidentes de tráfico entre 2016 y 2017, por medio del uso del Software WEKA (Waikato Environment for Knowledge Analysis). En el análisis, la severidad fue definida de bajo riesgo (daños materiales), y alto riesgo (victimas heridas y fatales), y su validación consideró la técnica transversal 10-fold. Entre los resultados más significativos, se evidenció que los motociclistas y ciclistas son los actores viales más vulnerables, además los moto-usuarios entre los 20-39 años son propensos a accidentes viales con alta severidad. Finalmente, los factores de accidentalidad vial identificados ayudan a promover contramedidas para mejorar la seguridad vial de la ciudad.

Palabras claves: Colombia, accidentalidad vial, severidad, minería de datos, predicción, reglas

Los autores agradecen al Departamento Administrativo de Tráfico y Transporte (DATT) por el acompañamiento y el apoyo con la información requerida para esta investigación. Los fondos para el primer autor fueron cubiertos por CEIBA-Gobernación de Bolívar (Colombia). Correo electrónico: Holman Ospina-Mateus, hospina@utb.edu.co. Recibido: 22 de noviembre de 2019; aceptado: 10 de diciembre de 2019.

Economía & Región, Vol. 13, No. 2, (Cartagena, diciembre 2019), pp. 83 - 115

Clasificaciones JEL: O18, R41

ABSTRACT

Prediction of road accidents in Cartagena, Colombia, using decision trees and association rules

The main objective of this research is to predict the factors associated with the severity of road accidents in Cartagena (Colombia); the methodology is based on data mining techniques such as decision trees (J48) and association rules (support, confidence, Lift). The research was developed with 10,053 traffic accident records between 2016 and 2017, using the WEKA (Waikato Environment for Knowledge Analysis) software. In the analysis, the severity was defined as low risk (material damage), and high risk (injured and fatal victims), and its validation considered the 10-fold cross-sectional technique. Among the most significant results, it was evidenced that motorcyclists, cyclists are the most vulnerable road users, and motorcyclists between the ages of 20-39 are prone to road accidents with high severity. Finally, the road accident factors identified help promote countermeasures to improve the city's road safety.

Key words: Road accidents, severity, data mining, prediction, rules.

JEL Classifications: O18, R41

I. INTRODUCCIÓN

De acuerdo con la Organización Mundial de la Salud (OMS), cada año 1,4 millones de personas pierden la vida y 50 millones son lesionados a causa de accidentes de tránsito (WHO 2018). Los accidentes de tránsito en el 2016 fueron la octava causa de muertes en el mundo y se estima que hubo 18,2 personas fallecidas por cada 100 mil habitantes (F/100M habitantes). Esto representa 3.700 muertes diarias en las vías. Los accidentes de tránsito son la causa número uno de muerte para personas entre los 15 a 29 años. Entre 2013 y 2016 las fatalidades en las vías aumentaron un 8% a nivel mundial. Las tasas de mortalidad en países de bajos ingresos (27,5 F/100M habitantes) son tres veces más altas que en los países de altos ingresos (8,3 F/100M habitantes). Aunque solo el 1% de los automotores se encuentran en países emergentes, el 13% de

las muertes ocurren en estas naciones. En el contexto regional para el 2016, África (26,6 F/100M habitantes) tuvo las cifras de accidentalidad más elevadas, seguido por el Sureste Asiático (20,7 F/100M habitantes). América y Europa tienen tasas de accidentalidad por debajo del promedio global. Europa es la región con mejor comportamiento en accidentalidad, comparando sus cifras con las de los países con altos ingresos.

A nivel mundial más del 50% de las muertes en accidentes de tránsito conciernen a usuarios vulnerables: peatones, ciclistas y motociclistas. Los peatones y ciclistas representan el 26%, mientras las motocicletas el 28%. Los ocupantes de automóviles registran el 29%. África tiene la mayor proporción (44%) de mortalidad de peatones y ciclistas. Las regiones del Sudeste Asiático y el Pacífico Occidental tienen una proporción del 43% y 36% de fatalidades en motociclistas, respectivamente. América evidencia una relación del 50% de actores vulnerables (motocicletas 23%, peatón 22% y ciclistas 3%).

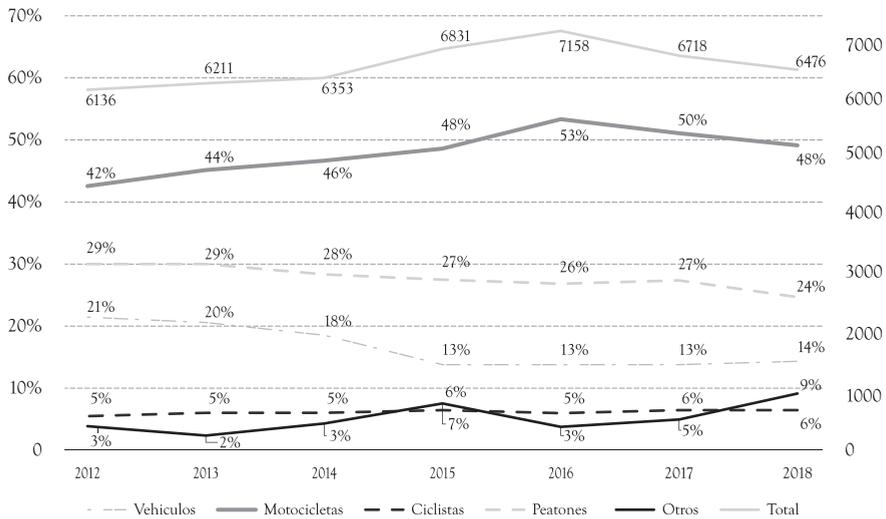
Según cifras de seguridad vial de Colombia informadas por la OMS en 2016, la tasa de fatalidades en el país es de 18,5 F/100M habitantes (WHO 2018). Esta cifra es muy cercana al promedio global (18,2 F/100M habitantes) y al promedio de naciones con ingresos medios (18,8 F/100M habitantes). Colombia se ubica en la décima posición en América, con un 21,7%, por encima de la media regional (15,2 F/100M habitantes). La proporción de accidentes en vehículos en Colombia en el 2016 corresponde a la mitad del promedio mundial. En accidentes de motociclistas, Colombia duplica aproximadamente el promedio global. Los ciclistas y peatones tienen cifras similares con la tendencia global.

Los accidentes de tránsito en los últimos siete años¹ en Colombia dejaron un promedio anual de 6.500 fatalidades y 42.000 víctimas heridas (ONSV 2019). El registro de fatalidades aumentó con una variación promedio del 4% entre 2012-2016. En el periodo 2017-2018 disminuyeron con una variación promedio del 5%. Las víctimas heridas en el periodo 2012-2015 aumentaron con variaciones promedio del 5%, mientras en el periodo 2016-2018 disminuyeron con una variación promedio del 7%.

En relación con los actores viales involucrados, los motociclistas corresponden en promedio al 47% de las víctimas fatales y al 53% de los lesionados. Los peatones corresponden al 27% y 21% de víctimas fatales y heridas respectivamente. Los usuarios de vehículos rondan el 16-19% de víctimas fatales y heridas. Las cifras más bajas las presentan los ciclistas con el 6% de accidentes de tránsito. Los Gráficos 1 y 2 muestran el número de víctimas discriminadas por actores viales para el periodo 2012-2018.

¹ Entre 2012 y 2018.

GRÁFICO 1
 Víctimas fatales en accidentes de tránsito en Colombia por tipo de actor vial, 2012-2018

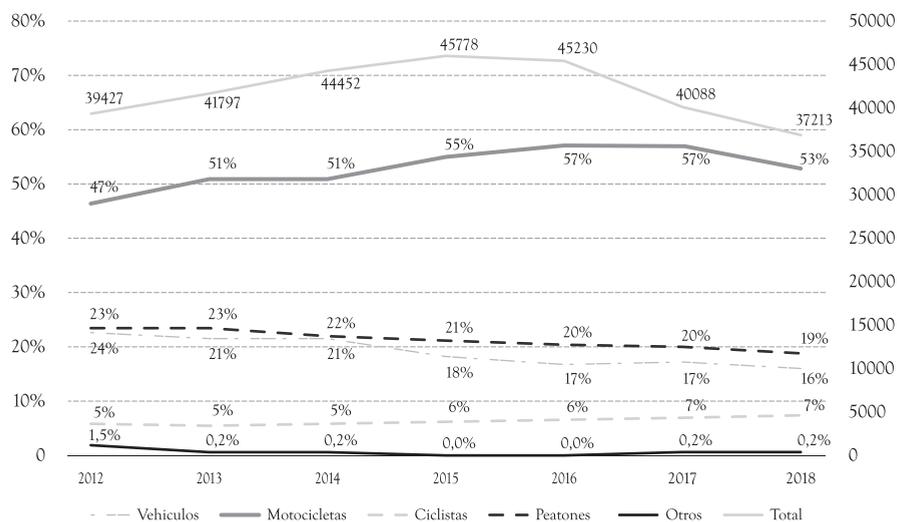


Fuente: ONSV (2019).

Aunque los accidentes de tránsito son de características multidimensionales y aleatorios, con uno o más usuarios fallando en un entorno vial común, el accidente, choque o colisión es el resultado de una cadena de eventos, que es única en sus condiciones y que vincula elementos comunes con las circunstancias del accidente (Montella *et al.* 2012). Conocer estas características y sus relaciones puede proporcionar información útil para el desarrollo de contramedidas efectivas (Ospina *et al.* 2019). En Colombia son reducidos los estudios relacionados con la causalidad y las condiciones de los eventos. Aunque en el país los entes de control emiten informes periódicos de accidentalidad vial, un análisis detallado de las víctimas mortales y lesiones en Colombia es una oportunidad para ampliar la gestión del conocimiento.

El objetivo de esta investigación es analizar los accidentes viales en Cartagena (Colombia), y los factores asociados con los choques y colisiones, así como su severidad. La finalidad es establecer reglas de decisión para una clara comprensión de la naturaleza del choque en el contexto local. La aplicación del método parte de información oficial por parte de los entes de control que

GRÁFICO 2
Heridos en accidentes de tránsito en Colombia por tipo de actor vial, 2012-2018



Fuente: ONSV (2019).

permitieron la aplicación de técnicas de minería de datos. Las reglas establecidas permitirán definir acciones sostenibles que puedan impactar en la seguridad vial.

Cartagena tiene aproximadamente 1,2 millones de habitantes, y en su malla vial circulan cerca de 130.000 vehículos. La ciudad tiene una extensión de 650 Km², y posee una red vial de más 730 Km. Sus principales actividades económicas giran en torno al turismo y las actividades del sector petroquímico y portuario. Cartagena entre el 2017 y 2018 se ha mantenido en los primeros lugares de los escalafones de ciudades capitales con mayor accidentalidad. En los últimos ocho años, Cartagena ha sido considerada ente las cinco ciudades más peligrosas en seguridad vial, después de Medellín, Cali, Bogotá y Barranquilla. En compañía de Barranquilla, Montería y Sincelejo, es una de las ciudades capitales de departamento en donde más se practica el mototaxismo (IEU 2019).

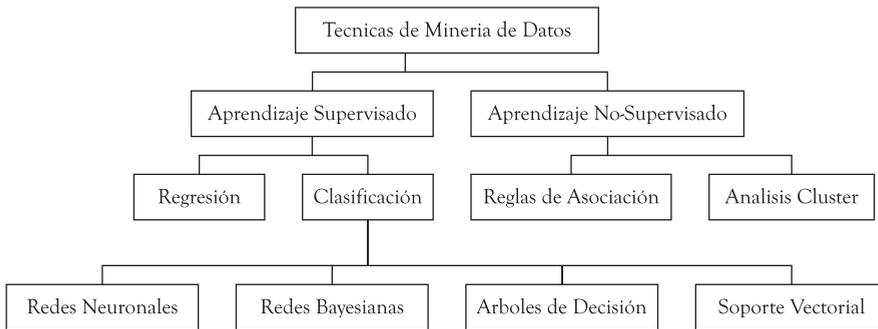
La investigación está organizada de la siguiente manera: la sección dos resume las técnicas de minería de datos aplicadas a los accidentes de tránsito y metodología. La sección tres brinda una introducción de los datos de

accidentalidad de la ciudad de Cartagena. La sección cuatro presenta el preprocesamiento y el análisis de los datos de accidentalidad. La sección cinco evidencia la parametrización de los métodos de Árbol de Decisión J48 y Reglas de Asociación. La sección seis presenta los resultados. Las secciones siete y ocho presentan la discusión y las conclusiones, respectivamente.

II. METODOLOGÍA

El método de investigación está fundamentado en un análisis inteligente de la información oficial registrada por los órganos de control en Cartagena con técnicas de minería de datos. La aplicación de estas técnicas está relacionada con técnicas de clasificación supervisadas como los árboles de decisión, así como la técnica no supervisada de reglas de asociación. En el Gráfico 3, se enseña la taxonomía de las técnicas de minería de datos.

GRÁFICO 3
Principales técnicas de minería de datos



Fuente: elaboración propia.

Las técnicas de minería de datos consisten en analizar y procesar grandes cantidades de información, para buscar comportamiento de los datos en un determinado contexto. La minería de datos en combinación con *machine learning* hacen parte de diversos campos, desde la estadística hasta las técnicas de inteligencia artificial y redes neuronales. Actualmente en la seguridad vial, se han implementado técnicas de minería de datos para el análisis general de

accidentalidad (Sohn y Shin, 2001; Martín *et al.* 2014; Ospina *et al.* 2019).

Las técnicas de procesamiento de datos en la predicción de eventos de tráfico han usado modelos regresivos, redes neuronales (Abdelwahab y Abdel-Aty, 2001; Delen *et al.*, 2006; Moghaddam *et al.*, 2011), inteligencia artificial (i.e. metaheurísticas) (Xi *et al.*, 2013; Jafari *et al.*, 2015; Hashmienejad y Hasheminejad, 2017), árboles de decisión (Sohn y Shin, 2001; Chang y Wang, 2006; Kashani y Mohaymany, 2011), redes de bayes (Huang *et al.*, 2008; Huang y Abdel-Aty, 2010; De Oña *et al.*, 2011; Cheng *et al.*, 2017), soporte vectorial (Li *et al.*, 2008) o alguna combinación de ellos (Taamneh *et al.*, 2017).

Los métodos orientados hacia el conocimiento como los árboles de decisión o las reglas de asociación permiten un mejor entendimiento de los resultados debido a la creación de reglas de decisión (sí-entonces) (Kashani *et al.*, 2014). El método propuesto pretende buscar los factores que interactúan en las condiciones que rodean el evento por medio de la aplicación de árboles de decisión con la técnica J48, así como con las reglas de asociación (soporte, confianza, y lift) de minería de datos. La aplicación de las técnicas fue desarrollada con el software WEKA.

Los árboles de decisión construyen modelos de clasificación de forma ramificada. Su implementación se basa en la inducción de reglas, y la iteración bajo el enfoque de división y segmentación. La aplicación de árboles de decisión con la técnica J48, o también conocida como C4.5, se presenta como un algoritmo que desarrolla estructura en forma de reglas basado en subconjuntos de datos extraídos del conglomerado total de datos de entrenamiento (Quinlan, 1993). De esta forma el subconjunto de reglas requiere ser evaluado en su bondad y precisión para garantizar la efectividad en sus predicciones. La aplicación del algoritmo requiere de monociclos que evalúan el beneficio de la forma cómo la información es clasificada de acuerdo con su afinidad. C4.5 o J48 es una mejora al algoritmo Id3 (Quinlan, 1986). El aprendizaje del árbol de decisión se usa ampliamente en la minería de datos. El resultado de este método es un modelo de clasificación que predice el valor de un atributo objetivo en función de los atributos de entrada. El árbol de decisión construye modelos de clasificación en forma de árboles. Cada nodo interior en estos árboles representa una de las variables de entrada, y tiene un número de ramas igual al número de valores posibles de esa variable de entrada. Cada nodo hoja tiene un valor del atributo de destino. El nodo hoja representa la decisión tomada en función de los valores de las variables de entrada desde la raíz hasta la hoja.

Para admitir la efectividad del algoritmo J48 se aplicó el uso de validación cruzada de 10-fold. Esta técnica permite evaluar que las predicciones son

independientes entre los subconjuntos de entrenamiento y de prueba. Esta técnica define de forma intrínseca la selección de grupos de datos para definir futuras predicciones, y así mismo validar su efectividad en grupos distintos y aleatorios.

Además de la aplicación de técnica predictiva J48, se llevó a cabo un análisis con reglas de asociación. Las reglas de asociación, como lo propone Agrawal *et al.* (1993) permiten encontrar la conexión entre elementos de datos en una base de datos transaccional. La asociación rige los algoritmos de minería de datos con el fin de descubrir la asociación frecuente. Las reglas de asociación aplicadas fueron las de soporte, confianza y lift, para identificar elementos en común dentro del conjunto de datos. Para una correcta definición se tiene que:

- *I* Conjunto de items: $I = \{i_1, i_2, \dots, i_m\}$
- *Transacción*: subconjunto de items $t \subseteq I$
- *Base de datos de n transacciones*: $T = \{t_1, t_2, \dots, t_n\}$
- *Itemset es un subconjunto de items, X es un itemset, $X \subseteq I$*
- *Items sets X y Y así que $X \cap Y = \emptyset$*
- $X \rightarrow Y$, cuando X ocurra en una transacción entonces Y ocurre con una dada probabilidad
- X es el antecedente y Y la consecuencia de la regla
- $t(X)$ evento con todas las transacciones que contiene a X

El soporte de un conjunto de items X en una base de datos se define como la proporción de transacciones T que contiene el itemset X. El soporte es calculado con las ecuaciones 1 y 2. El valor de confianza de una regla, $X \rightarrow Y$, es la proporción de transacciones T, que contienen itemset X donde se incluye el itemset Y. El indicador de confianza mide qué tan frecuente es que la regla coincida con el conjunto general. Se calcula con la ecuación (3). El indicador Lift es la proporción del soporte observado con respecto al soporte teórico del conjunto de datos de forma independiente. El valor de la elevación es que considera tanto el soporte de la regla como el conjunto de datos general. El Lift se calcula con la ecuación (4).

$$\text{soporte}(X) = \#X/n = P(t(X)), \text{ entonces } X \rightarrow Y (X \cap Y = \emptyset): \quad (1)$$

$$\text{soporte}(X \rightarrow Y) = \#(X \cup Y)/n = \text{soporte}(X \cup Y) = P(t(X) \cap t(Y)). \quad (2)$$

$$\text{confianza}(X \rightarrow Y) = \#(X \cup Y)/\#X = \text{soporte}(X \cup Y) / \text{soporte}(X) = P(t(Y) | t(X)). \quad (3)$$

$$\text{lift}(X \rightarrow Y) = \text{soporte}(X \cup Y) / (\text{soporte}(X) * \text{soporte}(Y)) \quad (4)$$

Las reglas de asociación son llevadas dentro del Software WEKA, en la técnica Apriori. La aplicación de la técnica puede generar un número masivo de conjuntos candidatos, que representan la frecuencia y asociación de un conjunto de ítems. Apriori es el algoritmo más sencillo, que se emplea para extraer patrones repetitivos de la base de datos de transacciones, en bases de datos de conjuntos de elementos grandes.

El software WEKA (Waikato Environment for Knowledge Analysis) es una herramienta para el desarrollo de aprendizaje automático, como *Machine Learning* (Witten *et al.*, 2016). La herramienta permite el análisis y extracción de datos de manera rápida y flexible. El software está dotado de técnicas para: regresiones, clasificación, agrupamiento, reglas de asociación, concepción y selección de atributos. En este estudio, elegimos WEKA como herramienta dado que es intuitivo, robusto y tiene muchas características integradas e incorporadas que no requieren programación. El uso de WEKA es muy común en la comunidad académica para la investigación y la enseñanza. Es sencillo de usar en interfaz gráfica que permite una configuración y operación rápidas. WEKA permite trabajar con archivos planos o tabulares (xls o csv). El uso de la información dentro del software se puede administrar por variables numéricas (continua o discretas), nominales, y alfanuméricas o categóricas (Witten *et al.*, 2016).

Esta investigación se dirigió a buscar y predecir factores significativos (condiciones) en la severidad de los accidentes de tránsito en Cartagena. El proceso metodológico de la investigación se desarrolló en cinco etapas (Gráfico 4): a) preprocesamiento de la base de datos de accidentes viales de Cartagena; b) Aplicación de la técnica J48 con el software WEKA; c) aplicación de reglas de asociación con el software WEKA. d) análisis de resultados y métricas; e) definición de reglas de decisión y factores asociados en la accidentalidad.

GRÁFICO 4

Fases y etapas metodológicas seguidas para el análisis de accidentalidad vial



Fuente: elaboración propia.

La etapa de preprocesamiento comprendió la identificación de todas las instancias dentro del conjunto de datos y la determinación de las variables para su predicción. En el preprocesamiento se filtraron y depuraron las instancias con ruidos y con condición atípica. Luego, se realizó un análisis descriptivo e inferencial para analizar la posible incidencia de las variables declaradas sobre las futuras predicciones. En la etapa de aplicación de los árboles de decisión, se ingresó al software WEKA la información previamente preprocesada. En la etapa de aplicación de reglas de asociación se partió también de la información preprocesada. Esta información se ingresó dentro de la plataforma WEKA, con el uso de la técnica Apriori. En el análisis de los resultados se validó el algoritmo J48 en las métricas de precisión, exactitud y exhaustividad, así como ROC, con otros algoritmos en WEKA tales como: CART, *Naïve Bayes*, *Support Vector Machines*, y Redes Neuronales. Finalmente se combinaron los resultados obtenidos por medio del árbol de decisión y las reglas de asociación, para construir un conjunto de reglas de decisión alrededor de las condiciones de accidentalidad en las vías de Cartagena.

III. DATOS DE ACCIDENTALIDAD VIAL DE CARTAGENA

El registro de los accidentes corresponde a la base de datos de Cartagena (Colombia) entre enero 2016 y diciembre del 2017. El registro de la base de datos fue suministrado por el Departamento de Tránsito y Transporte de Cartagena (DATT). En total, 10.053 accidentes viales fueron registrados por los agentes de tránsito. Los datos recopilan información acerca de la temporalidad, tipo de accidente, localización, tipo de servicio, tipo de actor vial, género y edad de cada uno de los actores relacionados en el evento. En la fase de preprocesamiento de la información, se definieron 18 variables categóricas (Cuadro 1).

Las variables fueron clasificadas en cuatro categorías: 1) Actores viales involucrados en el accidente (carros, buses, motociclistas, ciclistas, peatones, tipo de servicio), 2) Datos de individuos involucrados (edad, género), 3) Condiciones climáticas (precipitaciones) y temporalidad (día, mes, tipo de día), y 4) Características de la colisión (clase y víctimas). La información climatológica fue establecida en relación con la fecha del accidente y los registros del clima por día en Cartagena (IDEAM, 2017). El nivel de predicción para la severidad del evento fue clasificado en: 1) Accidente de bajo riesgo, en el que solo hay daños materiales, menores, sin víctimas heridas o lesiones y; 2) Accidentes de alto riesgo, en los que se presentan víctimas heridas o fatales.

CUADRO 1
Descripción de las variables de accidentalidad vial y su codificación en categorías

Categorías	Descripción	Etiqueta	Definición [código] (valor)
Actores Viales	Tipo de Vehículo Involucrado 1	TV1	[1] Automóvil; [2] Bus, buseta o microbús; [3] Motocicleta; [4] Vehículos Pesados; [5] Bicicletas; [6] Animal-Carreta; [7] No registrados [8] No aplica
	Tipo de Vehículo Involucrado 2	TV2	
	Tipo de Servicio de Vehículo 1	TS1	[1] Particular; [2] Público; [3] Oficial, [4] Indefinido; [5] No aplica
	Tipo de Servicio de Vehículo 2	TS2	
Individual	Genero de Conductor vehículo 1	GCV1	[1] Hombre; [2] Mujer; [3] Hombres y mujeres; y [4] no definido
	Genero de Conductor vehículo 2	GCV2	
	Genero de Víctima	GV	
	Edad de Conductor vehículo 1	EC1	[1] (0-19); [2] (20-39); [3] (40-59); [4] (60-79); [5] (<79)
	Edad de Conductor vehículo 2	EC2	
Edad de Víctima en Accidente	EV		
Clima / Temporalidad	Trimestres	TM	[1] Enero-Marzo; [2] Abril-Junio; [3] Julio-Septiembre; y [4] Octubre-Diciembre
	Día de semana	DS	[1] Lunes; [2] Martes; [3] Miércoles; [4] Jueves; [5] Viernes; [6] Sábado; [7] Domingo
	Tipo de día	TD	[1] Normal; [2] Festivo
	Tipo de día comercial	TDC	[1] Normal; [2] Quincena
	Condiciones climáticas por precipitación (Precipitación, unidad: mm)	CCP	[1] (0-15); [2] (15-30); [3] (30-60); [4] (60-90); y [5] (<90).

Fuente: elaboración propia.

CUADRO 1 (Continuación)
*Descripción de las variables de accidentalidad vial y
 su codificación en categorías*

Categorías	Descripción	Etiqueta	Definición [código] (valor)
Característica del Accidente	Clase de accidente	CLA	[1] Choque; [2] Caída del vehículo; [3] atropello; [4] volcado; [5] incendiado y otros
	Víctima lesionado o Víctima fatal	VID	[0] Sin víctimas; [1] Una víctima; y [2] Dos víctimas; [3] Tres víctimas; [4] Cuatro víctimas; [5] Más de cinco víctimas
	Severidad	SEV	[1] Bajo-Riesgo (daños materiales); [2] Alto riesgo (víctimas heridas y fatales)

Fuente: elaboración propia.

De acuerdo con el anterior enfoque general, las variables fueron convertidas a categóricas, como lo evidencia su codificación en el Cuadro 1. Se destacan aquellos que representan vehículos o individuos, o, por ejemplo, edad, tiempo y clima. La codificación categórica permite una clasificación y agrupación adecuadas de las variables condicionales para la definición de reglas de decisión en la metodología propuesta.

IV. FASE DE PREPROCESAMIENTO

En la primera etapa relacionada con el preprocesamiento y ajuste del conjunto de datos, así como las definiciones de las variables relacionadas con la condición del accidente, se limpió la información dentro de la minería de datos. El conjunto de datos o instancias con información atípica, defectuosa, incompleta, ruidosa y duplicada fue discriminado con el método de eliminación mal clasificado (*Remove Misclassified*) del software WEKA. El conjunto de datos fue reducido en 9.276 instancias. En el Cuadro 2 se resumen las variables usadas y su relación con la severidad.

CUADRO 2
Estadísticas de accidentalidad vial en Cartagena del Data-set, 2016–2017

Etiqueta	Código	%Total	Severidad [1]	Severidad[2]
TV1	[1]	65,0%	75,2%	24,8%
	[2]	12,4%	69,4%	30,6%
	[3]	14,0%	11,5%	88,5%
	[4]	7,7%	81,6%	18,4%
	[5]	0,2%	22,2%	77,8%
	[6]	0,0%	60,0%	40,0%
	[7]	0,6%	51,6%	48,4%
TV2	[1]	48,6%	84,3%	15,7%
	[2]	9,9%	88,6%	11,4%
	[3]	18,3%	24,4%	75,6%
	[4]	6,8%	88,7%	11,3%
	[5]	1,0%	10,3%	89,7%
	[6]	0,2%	68,2%	31,8%
	[7]	10,6%	44,9%	55,1%
	[8]	4,6%	11,9%	88,1%
TS1	[1]	59,9%	61,7%	38,3%
	[2]	37,6%	72,3%	27,7%
	[3]	0,8%	63,9%	36,1%
	[4]	1,7%	68,9%	31,1%
TS2	[1]	54,2%	63,5%	36,5%
	[2]	30,4%	85,9%	14,1%
	[3]	0,8%	65,1%	34,9%
	[4]	10,0%	41,5%	58,5%
	[5]	4,6%	12,0%	88,0%

Fuente: elaboración propia.

CUADRO 2 (Continuación)
Estadísticas de accidentalidad vial en Cartagena del Data-set, 2016–2017

Etiqueta	Código	%Total	Severidad [1]	Severidad[2]
GCV1	[1]	91,0%	65,4%	34,6%
	[2]	7,0%	79,6%	20,4%
	[3]	0,0%	0,0%	0,0%
	[4]	2,0%	35,5%	64,5%
GCV2	[1]	81,4%	70,8%	29,2%
	[2]	5,2%	77,4%	22,6%
	[3]	8,8%	40,8%	59,2%
	[4]	4,6%	11,9%	88,1%
GV	[1]	21,7%	0,2%	99,8%
	[2]	12,4%	0,0%	100,0%
	[3]	0,2%	0,0%	100,0%
	[4]	65,8%	100,0%	0,0%
EC1	[1]	0,8%	31,6%	68,4%
	[2]	43,4%	61,1%	38,9%
	[3]	26,8%	66,2%	33,8%
	[4]	16,9%	73,5%	26,5%
	[5]	10,7%	73,2%	26,8%
	[6]	1,4%	72,9%	27,1%
EC2	[1]	1,4%	33,6%	66,4%
	[2]	41,8%	63,0%	37,0%
	[3]	27,1%	70,9%	29,1%
	[4]	15,9%	71,3%	28,7%
	[5]	12,6%	59,4%	40,6%
	[6]	1,2%	79,2%	20,8%

Fuente: elaboración propia.

CUADRO 2 (Continuación)
Estadísticas de accidentalidad vial en Cartagena del Data-set, 2016–2017

Etiqueta	Código	%Total	Severidad [1]	Severidad[2]
EV	[0]	65,8%	100,0%	0,0%
	[1]	1,0%	0,0%	100,0%
	[2]	13,5%	0,1%	99,9%
	[3]	9,6%	0,2%	99,8%
	[4]	6,7%	0,1%	99,9%
	[5]	2,5%	0,0%	100,0%
	[6]	0,9%	0,0%	100,0%
TM	[1]	23,9%	65,0%	35,0%
	[2]	28,1%	64,5%	35,5%
	[3]	29,1%	65,9%	34,1%
	[4]	18,9%	68,5%	31,5%
DS	[1]	13,8%	63,9%	36,1%
	[2]	14,9%	66,5%	33,5%
	[3]	30,1%	66,8%	33,2%
	[4]	0,0%	0,0%	0,0%
	[5]	15,7%	74,3%	25,7%
	[6]	14,8%	64,6%	35,4%
	[7]	10,8%	53,8%	46,2%
TD	[1]	95,9%	66,0%	34,0%
	[2]	4,1%	61,3%	38,7%
TDC	[1]	51,7%	66,1%	33,9%
	[2]	48,3%	65,4%	34,6%

Fuente: elaboración propia.

CUADRO 2 (Continuación)
 Estadísticas de accidentalidad vial en Cartagena del Data-set, 2016–2017

Etiqueta	Código	%Total	Severidad [1]	Severidad[2]
CCP	[1]	38,8%	65,0%	35,0%
	[2]	9,3%	65,0%	35,0%
	[3]	11,6%	67,2%	32,8%
	[4]	8,4%	63,8%	36,2%
	[5]	31,9%	67,0%	33,0%
CLA	[1]	96,0%	68,4%	31,6%
	[2]	0,4%	2,2%	97,8%
	[3]	3,3%	0,0%	100,0%
	[4]	0,2%	44,4%	55,6%
	[5]	0,1%	75,0%	25,0%
VID	[0]	65,8%	100,0%	0,0%
	[1]	23,6%	0,2%	99,8%
	[2]	9,5%	0,0%	100,0%
	[3]	0,9%	0,0%	100,0%
	[4]	0,2%	0,0%	100,0%
	[5]	0,1%	0,0%	100,0%

Fuente: elaboración propia.

El análisis del conjunto de datos (*data-set*) situó el 66% de las instancias en bajo riesgo y el 34% en alto riesgo. En el análisis descriptivo de la información, la mayoría de los accidentes involucran al menos un automóvil (57% en promedio). En promedio, el 11% de las instancias conciernen al menos un bus, buseta o microbús. En promedio, el 16% involucran al menos a una motocicleta. En promedio, el 8% corresponde a casos con un vehículo de carga pesada, como camiones, tractores o furgones. Cuando se analiza la relación de colisión entre los actores viales, se encuentra que el 33% de los accidentes ocurren entre carros, y el 23% suceden entre un carro y un vehículo grande (pesado o buses). El 17% entre motocicletas y carros. El 3% entre motocicletas, y finalmente el

3% entre motocicletas y vehículos grandes (pesados o buses).

Al analizar el tipo de servicio del vehículo, el 40% de accidentes relacionan un vehículo particular y uno público. El 32% ocurren entre vehículos particulares. Finalmente, el 11% relacionan a accidentes entre vehículos de transporte público. En el género de los conductores se encontró que los hombres son predominantes. Los accidentes en los que se involucra al menos un motociclista tienen severidad con más alto riesgo (entre 75,6% y 88,5%), es decir, que se presentan al menos heridos o víctimas fatales.

La edad de los conductores que predomina se encuentra entre los 20 a 39 años (en promedio 42%). La mayoría de los accidentes se presentan entre días de semana (75%) y en días no festivos (96%), sin embargo, los fines de semana la severidad es más alta (en promedio 41%). El 53% de los accidentes se registraron en condiciones de precipitaciones mayores de 30mm. El tipo de accidente más frecuente es la colisión (96%). El más severo de este tipo de accidentes son los atropellos (100%), en los que al menos se presenta una víctima herida o fatal.

Después del análisis descriptivo, se realizó un análisis inferencial de las variables propuestas para predecir la severidad de los accidentes viales. En el Anexo A se muestran los resultados por medio del análisis Friedman ANOVA. El análisis evidenció que todas las variables tienen significancia estadística en las predicciones de la accidentalidad vial de Cartagena (p -valor $< 0,05$). Se esperan encontrar afinidad en las variables para la definición de predicciones y la construcción de reglas de decisión.

V. PARAMETRIZACIÓN DE ALGORITMO DE ÁRBOL DE DECISIÓN J48 Y REGLAS DE ASOCIACIÓN

Después del preprocesamiento de la información se parametrizó la plataforma WEKA, con el fin de ajustar y configurar las condiciones de predicción del algoritmo. Las condiciones de parametrización se muestran en el Cuadro 3.

CUADRO 3
Parametrización del algoritmo J48

Parametrización	Descripción	Valor
Particiones Binarias	Si se deben usar divisiones binarias en atributos nominales al construir los árboles.	FALSO
Número Mínimo de Objetos	Número mínimo de instancias por hoja.	2
Pliegues numéricos	Determina la cantidad de datos utilizados para la reducción de errores reducidos.	3
Factor de Confianza	El factor de confianza utilizado para la poda.	0,25
Sin podar	Si se realiza la poda.	FALSO

Fuente: elaboración propia.

La estimación de la precisión de un algoritmo supervisado es muy importante para asegurar la futuras predicciones. Este aspecto influye de manera determinante en la confiabilidad de los resultados. El método tradicional de validación consiste en separar los datos para el entrenamiento y los conjuntos de pruebas. Dentro del esquema de validación del algoritmo se usó la validación transversal o cruzada 10 fold (*10-fold cross-validation*). Esta técnica se fundamenta en aleatorizar todas las instancias del conjunto de datos y dividirlo en k subconjuntos distintos de instancias.

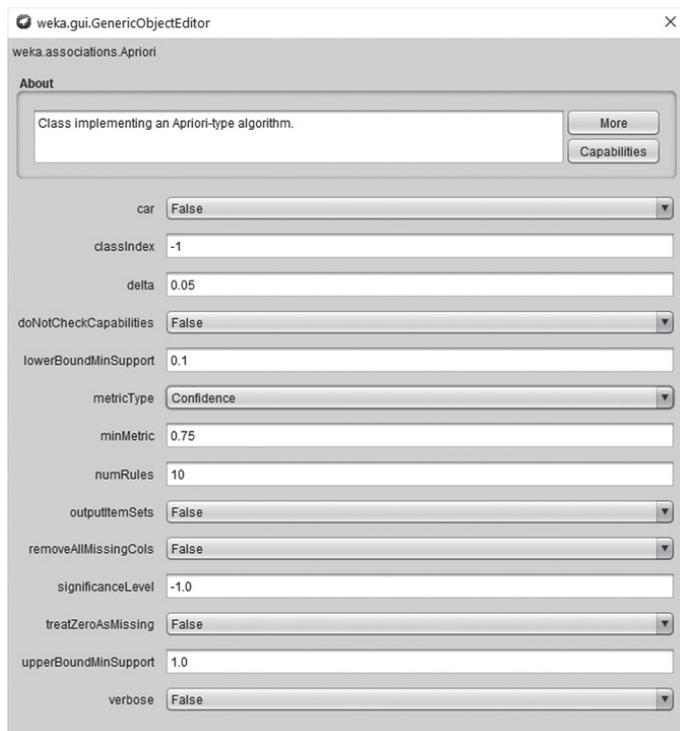
La operación se establece entrenando al clasificador en el número $k-1$ de registros y registrando en un subconjunto restante. El entrenamiento seguido por las pruebas se repite k veces, en las que un segmento es utilizado para validación y el resto de las muestras se usan en el conjunto de datos de entrenamiento. La precisión de la clasificación general del clasificador se calcula usando simplemente un promedio de las k pruebas individuales realizadas. En este estudio se usó esta técnica asignando a k un valor de 10 (es decir, validación cruzada 10 veces). Esta técnica ha demostrado que 10 parece ser un valor apropiado.

Con el propósito de contrastar los resultados del algoritmo J48 y otras técnicas de minería de datos, se hace uso de la matriz de confusión para calcular las métricas de precisión, exactitud y exhaustividad (también conocidas en inglés como “accuracy”, “precision”, y “recall” respectivamente). Finalmente, la configuración y ajuste para determinar las reglas de asociación para la extracción

de reglas alrededor de los datos de accidentalidad se visualizan en el Gráfico 2. Los valores de corte para la confianza se ajustaron, el superior a un valor de 0,75, y el valor de soporte inferior a 0,1.

CUADRO 5

Configuración de reglas de asociación dentro del Software WEKA



Fuente: elaboración propia.

VI. RESULTADOS

Luego de la fase de preprocesamiento, cuando se establecieron 18 variables que componen el conjunto de datos de accidentalidad de Cartagena, las instancias depuradas de la base de datos, y las configuraciones y ajustes, se ejecutó el algoritmo de árbol de decisión J48 y la regla de asociación con la

técnica Apriori. En el Cuadro 4 se muestran los resultados desde las métricas establecidas para el algoritmo de árbol de decisión J48. La técnica fue aplicada con técnica de validación cruzada 10-fold.

CUADRO 4

Medidas de rendimiento de algoritmo de árbol de decisión J48 en la técnica de validación cruzada 10-fold

Severidad	Instancias Clasificadas		Accuracy	Precision	Recall	Área ROC	Tiempo (seg)
	Correctas	Incorrectas					
Baja	5.282	667	0,888	0,862	0,888	0,856	0,040
Alta	2.484	834	0,747	0,788	0,747	0,856	
Total	7.766	1.501	0,837	0,836	0,837	0,856	

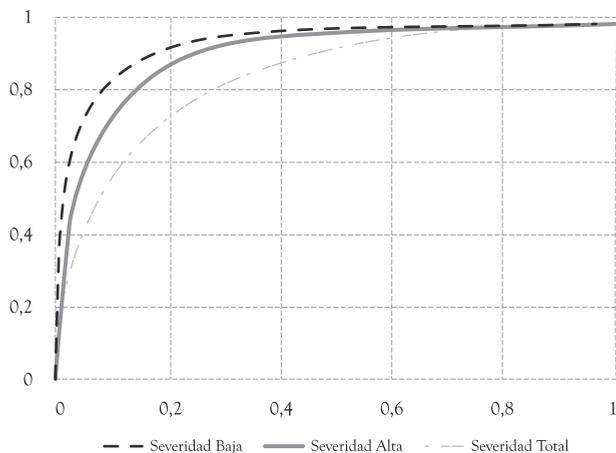
Fuente: elaboración propia.

En el algoritmo J48 la exactitud y exhaustividad de la predicción, basada en el conjunto de datos para severidad baja y alta de accidentes, tuvieron valores de 88,8% y 74,7% respectivamente. En el global, la exactitud y exhaustividad se enmarcaron en 83,7%. La precisión en severidad baja fue de 86,2% y en alta fue de 78,8%. En el global, la precisión fue 83,6%. El área bajo la curva en las predicciones fue de 85,6%. En el Gráfico 6 se establece la curva ROC de la técnica aplicada.

Con la finalidad de ratificar la efectividad del Método DT-J48, en el Anexo B se contrastaron los resultados con las diferentes técnicas de predicción PART, *Naïve Bayes* (NB), *Support Vector Machines* (SVM), y Redes Neuronales (MLP, sigla de *multilayer perception*), por medio del software WEKA.

La importancia de este estudio radica en su desarrollo de nuevas ideas relacionadas con los accidentes de tráfico en Cartagena. Estas ideas proporcionan una valiosa ayuda en el desarrollo de métodos para mejorar la seguridad vial, particularmente en la fase de elegir los medios apropiados y las asignaciones presupuestarias de recursos. El Cuadro 5 presenta una salida parcial de las reglas DT J48. El número de reglas generadas para severidad baja y alta de los accidentes fue de 486, con 613 atributos dentro del árbol de decisión.

GRÁFICO 6
Curvas ROC del árbol de decisión J48



Fuente: elaboración propia.

Debido a limitaciones de espacio, el Cuadro 6 presenta solo los resultados depurados para las reglas más significativas halladas en esta investigación.

CUADRO 5
Lista de reglas identificadas con el algoritmo de árbol de decisión J48

Número	Regla identificada (código)
1	Sí (TV1 o TV2 = 3) entonces Severidad =Alta
2	Sí (TV1 o TV2 ≠ 3 o 5) entonces Severidad =Baja
3	Sí (TV1 =3, TV2 =1, GCV1 =1, TM=4) entonces Severidad =Alta
4	Sí (TV1=3, CCP>3) entonces Severidad =Alta
5	Sí (TV1 o TV2=5) entonces Severidad =Alta
6	Sí (TV1 o TV2=3, GCV1 o GCV2=1, EC1 o EC2=2, DS>5) entonces Severidad =Alta

Fuente: elaboración propia.

CUADRO 5 (Continuación)
Lista de reglas identificadas con el algoritmo de árbol de decisión J48

Número	Regla identificada (código)
7	Si (TV1=1 , TV2=3, GCV2=2, EC2=3) entonces Severidad= Baja
8	Si (CLA=1 , TV1 o TV2 ≠ 3 o 5) entonces Severidad= Baja
9	Si (CLA= 2 o 3) entonces Severidad= Alta
10	Si (TV1=3 , TV2=2 o 4, CCP>4) entonces Severidad= Alta
11	Si (CLA=1 , TV1=1, TV2 =4, TM=4, CCP=4) entonces Severidad= Alta
12	Si (TV1=1, TV2 =2, GCV1=2, TDC=2) entonces Severidad= Alta
13	Si (CLA=1,TV1=1, TV2 =1) entonces Severidad= Baja
14	Si (CLA=1,TV1=3, TV2 =3, VID>2) entonces Severidad= Alta
15	Si (TV1 o TV2 = 1 o 2 o 4 , TS1 o TS2 =2) entonces Severidad =Baja

Fuente: elaboración propia.

En total se hallaron 10 reglas para la predicción de accidentes con severidad alta y 5 reglas relacionadas con accidentes con severidad baja. La regla 1, señala que si al menos hay una motocicleta involucrada, la severidad es alta. La regla 2, indica que cuando no hay motocicletas y bicicletas involucradas, la severidad es baja. La regla 3, expresa que cuando hay un accidente entre motocicletas y automóviles, el motociclista es hombre, y el evento se presenta en el último trimestre del año, la severidad es alta. La regla 4, estipula que si hay una motocicleta involucrada en condiciones de precipitación superior a 30mm, la severidad es alta. La regla 5, indica que si al menos una bicicleta involucrada en el accidente, la severidad es alta. La regla 6, estipula que si hay un motociclista hombre con edad entre los 20 a 39 años, en un fin de semana, la severidad es alta. La regla 7, señala que si hay un accidente entre un vehículo y una motocicleta, y la motociclista es mujer entre 30-59 años, la severidad es baja. La regla 8, indica que cuando el accidente es un choque sin motocicletas o bicicletas, la severidad es baja. La regla 9, expresa que si el accidente es caída o atropello, la severidad es alta. La regla 10, señala que si el accidente es entre una motocicleta y un vehículo grande (Bus o pesado), en condiciones de lluvias intensas, la severidad es alta. La regla 11, expresa que si es un choque entre un automóvil y un vehículo pesado, en el trimestre final del año y en condiciones de lluvias

intensas, la severidad es alta. La regla 12, representa que si es accidente entre un automóvil, con conductora mujer y un bus, en un día de quincena, la severidad es alta. La regla 13, indica que si es un choque entre dos vehículos, la severidad es baja. La regla 14, expresa que si el choque se presenta entre dos motocicletas, se presentarán más de dos víctimas heridas, siendo este evento de severidad alta. La regla 15, señala que si el accidente se presenta entre automóviles, buses, o vehículos pesados, todos de servicio público, la severidad es baja.

Luego de conocer las reglas de decisión en la accidentalidad vial de Cartagena, a continuación se muestran las reglas de asociación entre las variables definidas. En el Cuadro 6 se evidencian las 10 reglas más recurrentes para el conjunto de datos.

CUADRO 6
Lista de asociaciones identificadas

Número	Asociaciones identificadas (código)
1	Si (GCV1 o GCV2 =1, TD=1) entonces CLA =1
2	Si (GCV1 =1) entonces GCV2=1
3	Si (TV1=1, TV2=1) entonces VID =0
4	Si (TV1=1 o 2 o 4, TV2= 3) entonces VID>1
5	Si (TV1 o TV2=1 o 2) entonces GCV1 o GCV2 =1
6	Si (TS1 o TS2 = 2) entonces GCV1 o GCV2 =1
7	Si (EC1 o EC2 =2) entonces GCV1 o GCV2 =1
8	Si (TV1 o TV2 =3) entonces DS> 5
9	Si (TV1 = 2 o 4, TV2 = 3) entonces VID > 2
10	Si (TV1 o TV2 =3) entonces EV>3

Fuente: elaboración propia.

Los resultados de las reglas de asociación permiten reforzar los vínculos evidenciados previamente con el algoritmo J48. En los hallazgos se tiene: Asociación 1, indica que los hombres en días normales (no festivos) presentan más choques. Asociación 2, señala que los hombres tienden a accidentarse con otros hombres. Asociación 3, expresa que los accidentes entre automóviles no dejan víctimas. Asociación 4, representa que los choques de automóviles, buses

o vehículos pesados, con motocicletas dejan más de una víctima. Asociación 5, representa que los accidentes en automóviles y buses se producen por hombres. Asociación 6, indica que en los accidentes los vehículos de servicio público son manejados por hombres. Asociación 7, señala que la edad de los 20 a 39 años en hombres es la más propensa para accidentes. Asociación 8, expresa que las motocicletas se accidentan con mayor probabilidad los fines de semana. Asociación 9, indica que los accidentes entre motocicletas y vehículos grandes (buses o pesados) dejan más de dos víctimas. Asociación 10, señala que la mayoría de los motociclistas relacionados en accidentes son menores de 39 años.

VII. ANÁLISIS

Esta investigación permitió analizar los registros de accidentes de Cartagena (Colombia) con técnicas de minería de datos. Las reglas de decisión identificadas son esenciales para la caracterización de los accidentes de tráfico. El método propuesto con las variables establecidas relaciona factores asociados con las condiciones del evento. Las reglas contribuyen a la definición de políticas y estrategias para la reducción de la gravedad de los accidentes de acuerdo con el contexto y la información disponible para ser reproducida. El método propuesto relaciona dos tipos de accidentes (1: daños materiales, 2: víctimas lesionadas y muertes). La aplicación del método propuesto depende del conjunto de datos disponible y se registran las variables que caracterizan cada evento.

Los resultados muestran a los ciclistas y motociclistas como los usuarios más vulnerables de la carretera. Los motociclistas hombres y mujeres de entre 20 y 39 años predicen accidentes de alta gravedad. Cuando no hay motocicletas o ciclistas involucrados en el accidente, la gravedad probable es baja. Además, la colisión entre dos motocicletas se considera de alta gravedad. Si el accidente es un atropello, la gravedad es alta y se infiere que la víctima es un peatón. Finalmente, si el choque se produce entre vehículos con sistemas de protección rígidos, como automóviles, autobuses o vagones, el accidente disminuye su gravedad.

Las reglas de decisión (algoritmo J48) evidencia que más del 50% de las reglas definidas están relacionados con usuarios de motocicletas. La presencia de usuarios vulnerables de la carretera como los motociclistas fueron aspectos predictivos en la gravedad de los accidentes, así como lo evidencian resultados de Delen *et al.* (2017), y Hashmienejad y Hashmienejad (2017). Los motociclistas

son una población con un crecimiento significativo debido a algunas condiciones de movilidad, transporte, deportes y otras actividades económicas mencionadas en Ospina-Mateus y Quintana (2019).

Una contramedida basadas en el presente estudio es definir planes de control de seguridad vial según la temporada. Las reglas 3 y 11 contrastan que hay meses más exigentes en el control de carreteras. Otra medida puede ser definir límites de velocidad condicionados por la intensidad de la lluvia. Las reglas 10 y 11 muestran que las lluvias moderadas e intensas aumentan la posibilidad de accidentes. También, los accidentes de motociclistas se pueden evitar si se reduce la interacción con otros usuarios. Esto se logra con carriles de circulación única. Las reglas 3, 6, 7, 10, y 14 relacionan los automóviles cuando la motocicleta interactúa con otro vehículo (por ejemplo, camiones y automóviles). Finalmente, las reglas 6 y 7 confirman que la edad del motociclista influye en la gravedad del accidente. Estas reglas sugieren recomendaciones en el sentido de definir planes de licencia según la edad. Por ejemplo, en Europa existen restricciones de circulación, velocidad, desplazamiento y violaciones según la edad. La regla 9 está estrechamente relacionada con la alta gravedad en el tipo de accidente, que se puede generar por los abusos, violaciones o maniobras en las vías. Para definir contramedidas efectivas a partir de estas reglas, sería bueno incluir nuevas variables. Estos deben centrarse en aspectos adicionales como la ubicación del accidente, la intersección, la ubicación de la carretera, el tipo de carretera, la señalización, la iluminación, el tiempo, entre otros.

Entre otras políticas, estrategias y contramedidas para mejorar la seguridad vial del motociclista y las recomendaciones de la WHO (2017), se deben considerar: promover la cultura y la educación en seguridad vial para motociclistas (Segui-Gomez y Lopez-Valdes, 2007); analizar los informes de accidentes en motocicletas para conocer los factores de riesgo asociados (Schneider IV *et al.* 2012); campañas de seguridad vial en los usuarios más vulnerables en las edades de 15 a 30 años (Zambon y Hasselberg, 2006; Ivers *et al.*, 2016); promover el uso de elementos protectores (casco, guantes, espinilleras, coderas, *airbags*, protectores lumbares) (Li *et al.*, 2008; Donate-López *et al.*, 2010; Bourdet *et al.*, 2012); restringir y sancionar la conducción bajo la influencia de drogas y alcohol (Peek-Asa y Kraus, 1996; Albalate y Fernández-Villadangos, 2010); Controlar la velocidad según el tipo de carretera (Clabaux *et al.*, 2012); mejorar la calidad de las carreteras o diseñar carriles exclusivos (Daniello y Gabler, 2011; Sager *et al.*, 2014); mejorar las condiciones mecánicas y el mantenimiento (e.g. frenos ABS, neumáticos de calidad, sistema antichoque y estabilización) (Langley *et al.*, 2000; Fowler *et al.*, 2016; Rizzi *et al.*, 2016); mejorar la visibilidad del motociclista (por ejemplo, con

el usos de chalecos, reflectantes, o sistemas de iluminación) (Cercarelli *et al.*, 1992., Clarke *et al.*, 2007); mejorar las condiciones de seguridad vial, como la señalización, la iluminación y la infraestructura (Quddus *et al.*, 2001, Haque *et al.* 2009; Schneider IV *et al.*, 2010; Pai, 2011); prohibir el transporte y la exposición de niños en motocicletas (Brown *et al.*, 2018); penalizar las infracciones y las conductas de riesgo (i.e. exceso de capacidad, maniobras) (Elliott *et al.*, 2007); y restringir la manipulación de dispositivos electrónicos mientras se conduce (Truong *et al.*, 2019).

Entre las limitaciones que presentan los métodos predictivos se encuentran la estabilidad de las condiciones temporales, que se asumen como supuestos en el enfoque de estos modelos (Mannering, 2018). Por ejemplo, con el tiempo, las políticas públicas de seguridad vial podrían afectar las tasas de accidentalidad. Otras limitaciones del estudio están relacionadas con la disponibilidad de información sobre accidentes, que depende del contexto. También puede existir subregistro de datos por parte de las entidades de control de tráfico, así como la posibilidad de obtener información causal del evento, que no está registrada por las entidades de control. Si se incluyen más variables causales, se puede lograr la creación de reglas significativamente más estratégicas para la seguridad vial.

VIII. CONCLUSIONES

Esta investigación permitió analizar los registros de accidentes de Cartagena (Colombia) con técnicas de minería de datos. Las reglas contribuyen a la definición de contramedidas, enfocadas en usuarios vulnerables para la reducción de la gravedad de los accidentes. La definición de reglas a partir de la minería de datos es más efectiva que analizar información con un simple análisis estadístico descriptivo. Debido a que el análisis de la información se realiza de forma correlacional, esto contribuye a obtener resultados que son más fáciles de entender y aplicar. Técnicas como el análisis multivariado o las técnicas de Caja negra requieren pasos adicionales para el análisis de la información.

En este estudio, los motociclistas en edades de adultos jóvenes están relacionados con factores predictivos de gravedad a nivel de lesión o fatalidad. Los registros mundiales en 2016 colocaron a Colombia en la décima posición mundial, la tercera en la región y la segunda en América del Sur en accidentes de motociclistas. En 2018, Colombia tenía 8,3 millones de motocicletas registradas. En los últimos siete años (2012-2018), la proporción de motociclistas muertos

y heridos es cercana al 50% de los usuarios de la carretera. Analizar la tasa de accidentalidad de motociclistas en Colombia y su causalidad son investigaciones futuras esenciales para mejorar la seguridad vial. Las limitaciones del estudio son el subregistro de datos por parte de las entidades de control de tráfico. Si se incluyen más variables causales, se puede lograr la creación de reglas significativamente más estratégicas para la seguridad vial (por ejemplo, estado y condiciones de la carretera).

REFERENCIAS

- Abdelwahab, Hassan., y Abdel-Aty, Mohamed (2001), "Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections", *Transportation Research Record: Journal of the Transportation Research Board* Vol. 1746, No. 1.
- Agrawal, Rakesh., Imieliński, Tomasz., y Swami, Arun (1993), "Mining association rules between sets of items in large databases", *Acm sigmod record*, ACM Vol. 22, No. 2.
- Albalate, Daniel y Fernández-Villadangos, Laura (2010), "Motorcycle injury severity in Barcelona: The role of vehicle type and congestion", *Traffic Injury Prevention* Vol. 11, No. 6.
- Bourdet, Nicolas., Deck, Caroline., Tinard, Violaine., y Rémy. Willinger (2012), "Behaviour of helmets during head impact in real accident cases of motorcyclists", *International Journal of Crashworthiness* Vol. 17, No. 1.
- Brown, Julie., Schonstein, Lisa., Ivers, Rebecca., y Keay, Lisa (2018), "Children and motorcycles: a systematic review of risk factors and interventions", *Injury Prevention* Vol. 24, No. 2.
- Cercarelli, L. Rina., Arnold, P.K., Rosman, D.L., Sleet, David., y Thornett, M.L. (1992), "Travel exposure and choice of comparison crashes for examining motorcycle conspicuity by analysis of crash data", *Accident Analysis and Prevention* Vol. 24, No. 4.
- Chang, Li., y Wang, Hsiu (2006), "Analysis of traffic injury severity: An application of non-parametric classification tree techniques", *Accident Analysis & Prevention* Vol. 38, No. 5.
- Cheng, Wen., Gill, Gurdiljot., Sakrani, Taha., Dasu, Mohan., y Zhou, Jiao (2017), "Predicting motorcycle crash injury severity using weather data and alternative Bayesian multivariate crash frequency models", *Accident Analysis and Prevention* 108.

- Clabaux, Nicolas., Brenac, Thierry., Perrin, Christophe., Magnin, Joël., Canu, Bastien., y Van Elslande, Pierre (2012), "Motorcyclists' speed and 'looked-but-failed-to-see' accidents", *Accident Analysis and Prevention* 49.
- Clarke, David., Ward, Patrick., Bartle, Craig., y Truman, Wendy (2007), "The role of motorcyclist and other driver behaviour in two types of serious accident in the UK", *Accident Analysis and Prevention* Vol. 39, No. 5.
- Daniello, Allison y Gabler, Hampton (2011), "Fatality risk in motorcycle collisions with roadside objects in the United States", *Accident Analysis and Prevention* Vol. 43, No. 3.
- De Oña, Juan., Mujalli, Randa., y Calvo, Francisco (2011), "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks", *Accident Analysis & Prevention* Vol. 43, No. 1.
- Delen, Dursun., Sharda, Ramesh., y Bessonov, Max (2006), "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks", *Accident Analysis & Prevention* Vol. 38, No. 3.
- Delen, Dursun., Tomak, Leman., Topuz, Kazim., y Eryarsoy, Enes (2017), "Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods", *Journal of Transport & Health* 4.
- Donate-López, Carolina., Espigares-Rodríguez, Elena., Jiménez-Moleón, Juan., Luna-del-Castillo, Juan., Bueno-Cavanillas, Aurora., y Lardelli-Claret, Pablo (2010), "The association of age, sex and helmet use with the risk of death for occupants of two-wheeled motor vehicles involved in traffic crashes in Spain", *Accident Analysis & Prevention* Vol. 42, No. 1.
- Elliott, Mark., Baughan, Christopher., y Sexton, Barry (2007), "Errors and violations in relation to motorcyclists' crash risk", *Accident Analysis and Prevention* Vol. 39, No. 3.
- Fowler, Graeme., Ray, Rose., Huang, Su-Wei., Zhao, Ke., y Frank, Todd (2016), "An examination of motorcycle antilock brake systems in reducing crash risk", *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* Vol. 2, No. 2.
- Haque, Md., Chin, Hoong., y Huang, Helai (2009), "Modeling fault among motorcyclists involved in crashes", *Accident Analysis and Prevention* Vol. 41, No. 2.
- Hashmienejad, Seyed., y Hasheminejad, Seyed (2017), "Traffic accident severity prediction using a novel multi-objective genetic algorithm", *International journal of crashworthiness* Vol. 22, No. 4.
- Huang, Helai y Abdel-Aty, Mohamed (2010), "Multilevel data and Bayesian analysis in traffic safety", *Accident Analysis & Prevention* Vol. 42, No. 6.

- Huang, Helai., Chin, Hoong., y Haque, Md (2008), "Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis", *Accident Analysis & Prevention* Vol. 40, No. 1.
- IDEAM (2017), "Precipitación mensual por año para Cartagena". Recuperado de www.ideam.gov.co
- IEU (2019), El mototaxismo sigue aumentando en Colombia. U. N. d. Colombia. Recuperado de <http://ieu.unal.edu.co/medios/noticias-del-ieu/item/el-mototaxismo-sigue-aumentando-en-colombia>
- Ivers, Rebecca., Sakashita, Chika., Senserrick, Teresa., Elkington, Jane., Lo, Serigne., Boufous, Soufiane y De Rome, Liz (2016), "Does an on-road motorcycle coaching program reduce crashes in novice riders? A randomised control trial", *Accident Analysis and Prevention* 86.
- Jafari, Seyed., Jahandideh, Sepideh., Jahandideh, Mina., y Asadabadi, Ebrahim (2015), "Prediction of road traffic death rate using neural networks optimised by genetic algorithm", *International journal of injury control and safety promotion* Vol. 22, No. 2.
- Kashani, Ali y Mohaymany, Afshin (2011), "Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models", *Safety Science* Vol. 49, No. 10.
- Kashani, Ali., Rabieyan, Rahim., y Besharati, Mohammad (2014), "A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers", *Journal of Safety Research* 51.
- Langley, John., Mullin, Bernadette., Jackson, Rodney., y Norton, Robyn (2000), "Motorcycle engine size and risk of moderate to fatal injury from a motorcycle crash", *Accident Analysis and Prevention* Vol. 32, No. 5.
- Li, Xiugang., Lord, Dominique., Zhang, Yunlong y Xie, Yuanchang (2008), "Predicting motor vehicle crashes using support vector machine models", *Accident Analysis & Prevention* Vol. 40, No. 4.
- Mannering, Fred (2018), "Temporal instability and the analysis of highway accident data", *Analytic methods in accident research* 17.
- Martín, Luis., Baena, Leticia., Garach, Laura., López, Griselda., y De Oña, Juan (2014), "Using data mining techniques to road safety improvement in Spanish roads", *Procedia-Social and Behavioral Sciences* 160.
- Moghaddam., F. Rezaie., Afandzadeh, Shariar., y Ziyadi, Max (2011), "Prediction of accident severity using artificial neural networks", *International Journal of Civil Engineering* Vol. 9, No. 1.
- Montella, Alfonso., Aria, Massimo., D'Ambrosio, Antonio., y Mauriello, Filomena (2012), "Analysis of powered two-wheeler crashes in Italy by

- classification trees and rules discovery”, *Accident Analysis and Prevention* 49.
- ONSV (2019), “Boletín 2018 (Parcial), Fallecidos, lesionados y hechos de tránsito en Colombia por INMLCF.”
- ONSV (2019), “Boletines Estadísticos, Fallecidos, Lesionados y hechos de tránsito en Colombia- Histórico Años 2012-2017.”, Recuperado de <https://ansv.gov.co/observatorio/?op=Contenidos&sec=63>.
- Ospina-Mateus, Holman y Quintana, Leonardo (2019), “Understanding the impact of physical fatigue and postural comfort experienced during motorcycling: A systematic review”, *Journal of Transport & Health* 12.
- Ospina-Mateus, Holman., Quintana, Leonardo., Lopez-Valdes, Francisco., y Salas-Navarro, Katherine (2019), “Bibliometric analysis in motorcycle accident research: a global overview”, *Scientometrics* Vol. 121, No. 2.
- Ospina-Mateus, Holman., Jiménez, Leonardo., López-Valdés, Francisco., Morales-Londoño, Natalie., y Salas-Navarro, Katherine (2019), Using data-mining techniques for the prediction of the severity of road crashes in Cartagena, Colombia. Workshop on Engineering Applications, Springer.
- Pai, Chih (2011), “Motorcycle right-of-way accidents - A literature review”, *Accident Analysis and Prevention* Vol. 43, No. 3.
- Peek-Asa, Corinne y Kraus, Jess (1996), “Alcohol use, driver, and crash characteristics among injured motorcycle drivers”, *Journal of Trauma - Injury, Infection and Critical Care* Vol. 41, No. 6.
- Quddus, Mohammed., Chin, Hoong., y Wang, J. (2001), “Motorcycle crash prediction model for signalised intersections”, *WIT Transactions on the Built Environment* 52.
- Quinlan, J. Ross (1986), “Induction of decision trees”, *Machine learning* Vol. 1, No. 1.
- Quinlan, J. Ross (1993), *C4. 5: Programming for machine learning*, San Francisco, CA.: Morgan Kauffmann.
- Rizzi, Matteo., Strandroth, Johan., Holst, Jan., y Tingvall, Claes (2016), “Does the improved stability offered by motorcycle antilock brakes (ABS) make sliding crashes less common? In-depth analysis of fatal crashes involving motorcycles fitted with ABS”, *Traffic Injury Prevention* Vol. 17, No. 6.
- Sager, Bertrand., Yanko, Matthew., Spalek, ., Thomas., Froc, David., Bernstein, Daniel., y Dastur, Farhad (2014), “Motorcyclist’s lane position as a factor in right-of-way violation collisions: A driving simulator study”, *Accident Analysis and Prevention* 72.
- Schneider IV, William., Savolainen, Peter., Van Boxel, Dan., y Beverley, Rick (2012), “Examination of factors determining fault in two-vehicle motorcycle

- crashes”, *Accident Analysis and Prevention* 45.
- Schneider IV, William., Savolainen, Peter., y Moore, Darren (2010), “Effects of horizontal curvature on single-vehicle motorcycle crashes along rural two-lane highways”, *Transportation Research Record* Vol. 2194, No. 1.
- Segui-Gomez, María y Lopez-Valdes, Francisco (2007), “Recognizing the importance of injury in other policy forums: The case of motorcycle licensing policy in Spain”, *Injury Prevention* Vol. 13, No. 6.
- Sohn, So y Shin, Hyungwon (2001), “Pattern recognition for road traffic accident severity in Korea”, *Ergonomics* Vol. 44, No. 1.
- Taamneh, Madhar., Alkheder, Sharaf., y Taamneh, Safah (2017), “Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates”, *Journal of Transportation Safety & Security* Vol. 9, No. 2.
- Truong, Long., Nguyen, Hang., y De Gruyter, Chris (2019), “Mobile phone use while riding a motorcycle and crashes among university students”, *Traffic injury prevention* Vol. 20, No. 2.
- World Health Organization (WHO) (2017), *Seguridad de los vehículos de motor de dos y tres ruedas: manual de seguridad vial para decisores y profesionales*. Génova: World Health Organization,
- WHO (2018), *Global status report on road safety 2018*. Génova: World Health Organization.
- Witten, Ian., Frank, Eibe., Hall, Mark., y Pal, Christopher (2016), *Data Mining: Practical machine learning tools and techniques*, San Francisco, CA.: Morgan Kaufmann.
- Xi, Jianfeng., Gao, Zhenhai., S. Niu, T. Ding y G. Ning (2013), “A hybrid algorithm of traffic accident data mining on cause analysis”, *Mathematical Problems in Engineering* 2013.
- Zambon, Francesco y Hasselberg, Marie (2006), “Socioeconomic differences and motorcycle injuries: age at risk and injury severity among young drivers. A Swedish nationwide cohort study”, *Accident Analysis and Prevention* Vol. 38, No. 6.

ANEXO A

Análisis Inferencial para la predicción de severidad en accidentes viales de Cartagena

CUADRO A1
Friedman ANOVA

Etiqueta Variables	Suma de cuadrados	Df	Cuadrado medio	F-Ratio	P-Valor
TV1	1209,34	1	1209,34	82483,31	0,00
TV2	381,385	1	381,385	26012,42	0,00
TS1	12,177	1	12,177	830,55	0,00
TS2	23,158	1	23,158	1579,51	0,00
GCV1	8,336	1	8,336	568,57	0,00
GCV2	24,266	1	24,266	1655,06	0,00
GV	888,388	1	888,388	60592,65	0,00
EC1	151,052	1	151,052	10302,51	0,00
EC2	95,815	1	95,815	6535,08	0,00
EV	323,59	1	323,59	22070,48	0,00
TM	20,346	1	20,346	1387,71	0,00
DS	13,375	1	13,375	912,27	0,00
TD	19,956	1	19,956	1361,08	0,00
TDC	52,070	1	52,070	3551,42	0,00
CCP	0,696	1	0,696	47,46	0,00
CLA	5,444	1	5,444	371,29	0,00
VID	61,475	1	61,475	4192,88	0,00
Modelo	3290,87	17			

Fuente: elaboración propia.

ANEXO B

CUADRO B1

Comparación de otras técnicas de predicción dentro del Software WEKA basado en los indicadores de rendimiento y con la técnica de validación cruzada 10-fold

Método	Instancias Clasificadas		Accuracy	Precision	Recall	Área ROC	Tiempo (seg)
	Correctas	Incorrectas					
DT-J48	7.766	1.510	83,7%	83,6%	83,7%	85,6%	0,04
PART	7.722	1.554	83,2%	83,1%	83,2%	86,4%	0,55
NB	7.573	1.703	81,6%	81,6%	81,6%	86,4%	0,02
SVM	7.814	1.462	84,2%	84,1%	84,2%	82,1%	30,15
MLP	7.433	1.843	80,1%	80,0%	80,1%	85,1%	142

Fuente: elaboración propia.

CUADRO B2

Parámetros considerados en las técnicas aplicadas en el software WEKA

Método	Parametrización
DT-J48	weka.classifiers.trees.J48 -C 0.25 -M 2
PART	weka.classifiers.rules.PART -M 2 -C 0.25 -Q 1
NB	weka.classifiers.bayes.NaiveBayes;Use kernel estimator F, Use supervised; discretization
SVM	weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"
MLP	weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Fuente: elaboración propia.