DOI: 10.15649/2346030X.743
Aibi revista de investigación, administración e ingeniería. Volumen 8, Número 2, Pág 133-141, Mayo - Agosto de 2020, ISSN 2346-030X

# Los accidentes de tránsito desde la perspectiva de la minería de datos. Una revisión de la literatura.
# Traffic accidents from the perspective of data mining A review of the literature.

Juan Pablo Henao-Pereira[1], Andrea Esperanza Tovar-León[2], Sandra Patricia Castillo-Landinez[3], Pablo Eduardo Caicedo-Rodriguez[4]
[1,2,3,4]Corporación Universitaria Autónoma del Cauca, Popayan - Colombia

Recibido: 30 de enero de 2020
Aprobado: 10 de abril de 2020

***Resumen*—** Se habla de un accidente de tránsito (AT) como un hecho inesperado que se presenta sobre las vías, condicionado por factores de naturaleza humana (imprudencia, descuido, problemas de salud) o también mecánica, involucrando por lo menos un vehículo en movimiento que puede ser automóvil, motocicleta o bicicleta; estos eventos provocan pérdidas de vidas o lesiones. Las cifras de los entes gubernamentales señalan que los accidentes de tránsito son la segunda causa de muerte violenta en Colombia, por lo que en este trabajo se indaga, cómo a través de técnicas de minería de datos es posible analizar los accidentes de tránsito, desde otra perspectiva, proponiendo un contexto inicial de investigación. Para tal fin, fueron recopilados trabajos de diferentes bases de datos como ScienceDirect, IEEE, ACM, Scielo, Redib y SpringerOpen, los cuales se clasificaron en tres ejes temáticos. Los resultados muestran que en una fase inicial de investigación de AT se deben desarrollar modelos de minería de datos de tipo descriptivo vinculando diferentes fuentes de datos.
**Palabras Claves**: accidente de tránsito, minería de datos, orígenes de datos, revisión bibliográfica**.**

***Abstract*—** We speak of a traffic accident (TA) as an unexpected event that occurs on the roads, conditioned by factors of human nature (recklessness, carelessness, health problems) or also mechanical, involving at least one vehicle in motion that can be car, motorcycle or bicycle; these events cause loss of life or injury. Figures from government agencies indicate that traffic accidents are the second cause of violent death in Colombia, so this paper explores how, through data mining techniques, it is possible to analyze traffic accidents from another perspective, proposing an initial research context. To this end, work was compiled from different databases such as ScienceDirect, IEEE, MCL, Scielo, Redib and SpringerOpen, which were classified into three thematic areas. The results show that in an initial phase of TA research, descriptive data mining models should be developed by linking different data sources.
**Keywords**: traffic accident, data mining, data sources, literature review.

.

*Autor de correspondencia
Correo electrónico: juan.henao.p@uniautoma.edu.co (Juan Pablo Henao Pereira)

Forma de citar: J. P. Henao-Pereira, A. E. Tovar-León, S. P. Castillo-Landinez y P. E. Caicedo-Rodriguez, “Los accidentes de tránsito desde la perspectiva de la minería de datos. Una revisión de la literatura”, Aibi revista de investigación, administración e ingeniería, vol. 8, no. 2, pp. 133-141, 2020 doi:10.15649/2346030X.743

# I. INTRODUCCIÓN

Se habla de un accidente de tránsito (AT) como un hecho inesperado que se presenta sobre las vías, condicionado por factores de naturaleza humana (imprudencia, descuido, problemas de salud) o también mecánica, involucrando por lo menos un vehículo en movimiento que puede ser automóvil, motocicleta o bicicleta; estos eventos provocan pérdidas de vidas o lesiones [1-2]. En la actualidad, los AT son una de las mayores causas de muerte a nivel mundial convirtiéndose en un problema de salud pública; de acuerdo con datos de la Organización Mundial de la Salud (OMS) publicados en 2018, en el mundo se presentan anualmente en promedio 1,35 millones de fallecidos por esta causa, y en el continente americano se registran cerca de 155000 víctimas, de ellos, el 34% son ocupantes de automóviles, 23% motociclistas y 22% peatones [3].

Los reportes del Observatorio Iberoamericano de Seguridad Vial (OISEVI) del año 2014 indican que el país con mayor número de fallecidos en AT por cada cien mil habitantes fue El Salvador y España reporta el valor más bajo para el mismo periodo; Colombia ocupa el octavo lugar. En la Figura 1 se observan las cifras para otros países de la región [4-5].

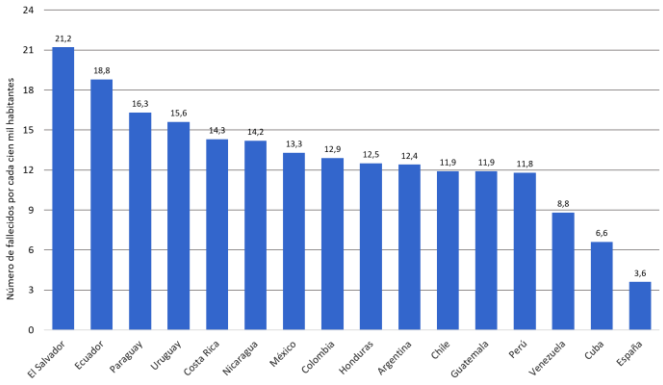


Figura 1: Tasa de fallecidos por cada cien mil habitantes durante 2014 en Iberoamérica.
Fuente: OISEVI.

En Colombia, durante 2019 según las cifras la Agencia Nacional de Seguridad Vial, los accidentes de tránsito ocasionaron 6826 víctimas fatales y 36812 lesionados, constituyendo la segunda causa de muerte violenta en el país después de los homicidios. En las Figuras 2 y 3 se comparan las cifras de actores viales involucrados en AT en 2018 y 2019.

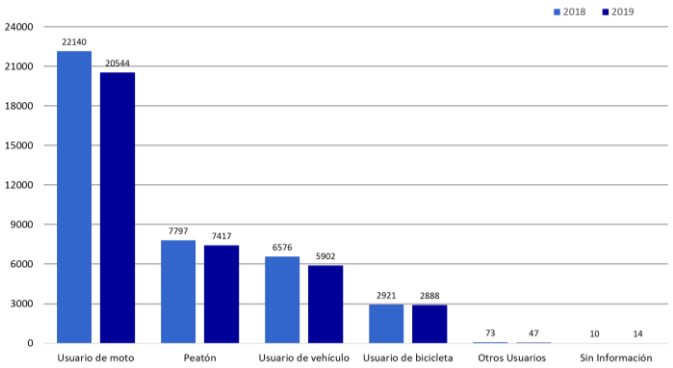


Figura 2: Lesionados en accidentes de tránsito en Colombia.
Fuente: ANS.

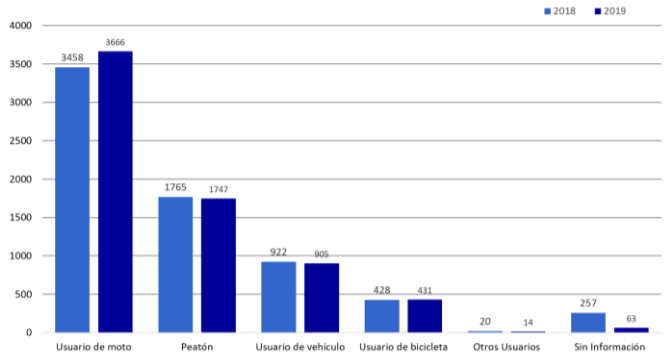


Figura 3: Fallecidos en accidentes de tránsito en Colombia.
Fuente: ANSV.

A nivel general se percibe un leve decremento en las cifras reportadas de fallecidos y lesionados comparado con el 2018 [6]. Un factor que influye en el número de accidentes es el aumento en el parque automotor en Colombia, que para el año 2019 creció un 2.7% en automóviles y un 10.6% en motocicletas con respecto a 2018, según la Asociación Colombiana de Vehículos Automotores [7].

El estudio de los AT implica indagar diferentes factores y procesar un número significativo de sucesos que se presentan diariamente, por lo que el análisis tradicional resulta poco eficiente y bastante tedioso en consideración a las complejas relaciones que pueden existir entre las diferentes variables estudiadas; el tratamiento de datos tiene sus raíces en la estadística pero esta técnica es limitada debido al aumento en el volumen de datos generados y su naturaleza diversa (imágenes, videos, coordenadas geográficas, etc.); buena parte de las variables a estudiar se pueden encontrar almacenadas en diferentes bases de datos digitales y la información más relevante no siempre se puede exponer mediante las herramientas matemáticas habituales. Gran parte de esta información es histórica, puede corresponder a millones de datos acumulados, por lo que encontrar manualmente patrones significativos que permitan la toma de decisiones de manera eficiente no es acertado [8-9]; por los aspectos mencionados anteriormente, la minería de datos se presenta como una solución viable al tratamiento de grandes cantidades de datos mediante el uso de técnicas y algoritmos robustos [10-11].

La minería de datos es el proceso de extracción de información a partir del procesamiento y análisis de grandes conjuntos de datos, para generar conocimiento que es usado en procesos que involucren toma de decisiones [10], éste puede ser representado mediante reglas, patrones o tendencias que no son evidentes a simple vista, y que pueden ser empleados en distintas áreas del conocimiento (medicina, biología, seguridad, genética, comercio, educación, etc.). Estos modelos son de tipo descriptivo o predictivo; los primeros buscan encontrar patrones interpretables para describir datos y engloba técnicas como: clustering (agrupación), descubrimiento de reglas de asociación, descubrimiento de patrones secuenciales, entre otras. Los modelos predictivos usan un conjunto de variables para estimar valores futuros o desconocidos y comprende técnicas como: clasificación, regresión y detección de la desviación [12-13]. El análisis de cluster permite construir grupos cuyos elementos tengan propiedades similares y obtener una caracterización de estos [14], el procedimiento resulta útil para registrar lugares donde se presenta mayor concentración de AT. Por otra parte, a partir de modelos basados en árboles de clasificación y regresión se puede establecer relaciones entre la gravedad de las lesiones y las características del conductor de un vehículo con otros factores externos. Cabe anotar que el principal inconveniente al realizar procesos de minería de datos es la calidad de los datos usados, ya que sus condiciones pueden generar un análisis poco profundo además de afectar notablemente en los resultados obtenidos [15-17].

De acuerdo con lo expuesto anteriormente, se busca determinar el potencial que ofrece el uso de técnicas de minería de datos en el

análisis de accidentes de tránsito para proyectar un estudio del caso colombiano. En tal sentido, se requiere establecer un contexto internacional en tres líneas temáticas: (i) el origen de datos utilizados en el desarrollo de este tipo de proyectos, (ii) el entorno geográfico que considera las características particulares de los países que han estudiado los AT (tipos de transporte o vehículos, condiciones de las vías, causas de los AT, etc.), (iii) indagar por los algoritmos utilizados a fin de identificar el tipo de estudio que se puede realizar (descriptivo, predictivo, prescriptivo o diagnóstico).

Este trabajo se encuentra organizado de la siguiente manera, inicialmente se presenta la metodología usada para la recolección y clasificación bibliográfica, detallando las bases de datos científicas, cadenas de búsqueda, criterios de selección de los artículos y categorías de clasificación. Posteriormente se estudian los documentos a la luz de tres ejes temáticos: países donde se han analizado AT usando técnicas de minería de datos, origen de los datos usados en las investigaciones y técnicas de minería de datos usadas para analizar AT. Finalmente, se presenta una discusión donde se confrontan otros artículos de revisión relacionados con el tema y esta investigación.

## II. METODOLOGÍA

La indagación literaria se realizó en diversas fuentes académicas con el propósito de identificar trabajos previos en los cuales se usaron algoritmos de minería de datos para la exploración y análisis de datasets relacionados con AT, siendo las más destacadas: IEEE, Scielo, ScienceDirect, SpringerOpen y ACM. En la Figura 4 se muestra el proceso realizado durante la revisión sistemática.

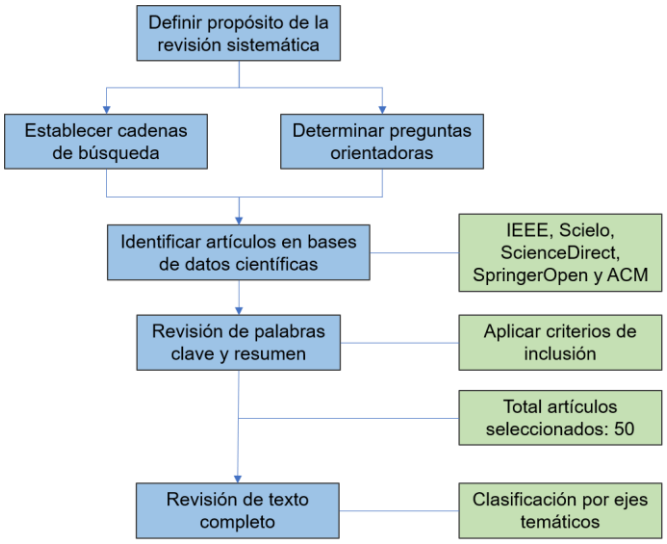


Figura.4: Proceso de búsqueda y selección de artículos.
Fuente: Elaboración propia.

Para guiar la búsqueda en las bases científicas se formularon las siguientes preguntas orientadoras:

- ¿En qué países o regiones se han realizado trabajos que utilicen técnicas de minería de datos para analizar AT?
- ¿De qué fuentes provienen los datos usados para desarrollar esos trabajos?
- ¿Cuáles son las principales técnicas o algoritmos de minería de datos empleadas para estudiar los AT?

La búsqueda sistemática de los documentos se realizó mediante cadenas de texto previamente definidas (Tabla.1).

Tabla 1: Cadenas de búsqueda usadas para recopilar los artículos.

| Base de Datos | Cadenas de búsqueda | Núm. Artículos |
|---|---|---|
| **Redib** | data mining and traffic accidents | 2 |
| **ACM** | data mining and traffic; road traffic and data mining; traffic accidents and data mining | 12 |
| **Scielo** | fatal accidents and data mining; homicides in traffic accidents and data mining; minería de datos and homicidios en accidentes de tránsito; procesamiento and metodología crisp; road traffic and data mining | 3 |
| **ScienceDirect** | traffic accidents and data mining; analysis of traffic accidents and data mining; data mining and accidents; data mining and traffic accidents; data mining techniques and accidents on road; road traffic and data mining | 5 |
| **SpringerOpen** | data mining and accident and techniques; data mining and traffic accidents | 3 |
| **IEEE** | accidents and data mining; analysis of traffic accidents and data mining; data mining and traffic accidents; data mining and accident and techniques; fatal accidents and data mining; homicides traffic and data mining; data mining and accident and analysis | 25 |

Fuente: Elaboración propia.

### Criterios de inclusión

En la revisión literaria inicial se reunieron 60 documentos; luego se seleccionaron 50 trabajos que cumplían los criterios de inclusión contenidos en la Tabla 2.

Tabla 2: Criterios de selección aplicados a los artículos revisados.

| Criterios de inclusión |
|---|
| Artículos publicados a partir de 2010. |
| En el artículo se detalla el uso de una o más técnicas de minería de datos. |
| El artículo abordar el análisis de accidentes de tránsito a partir del uso de técnicas de minería de datos. |
| El artículo ofrece respuesta por lo menos a una de las preguntas orientadoras. |

Fuente: Elaboración propia.

Finalmente, el material reunido fue clasificado de acuerdo con los tres ejes temáticos abordados en este trabajo y que son desarrollados a continuación.

## III. EJE TEMATICO 1: ORIGEN DE LOS DATOS USADOS EN LAS INVESTIGACIONES

Reportar la fuente de donde provienen los datos (Figura 5) resulta útil para que otros investigadores o personas interesadas en profundizar en la temática, puedan acceder a ellos y replicar o realizar nuevos trabajos en el mismo u otros tópicos.

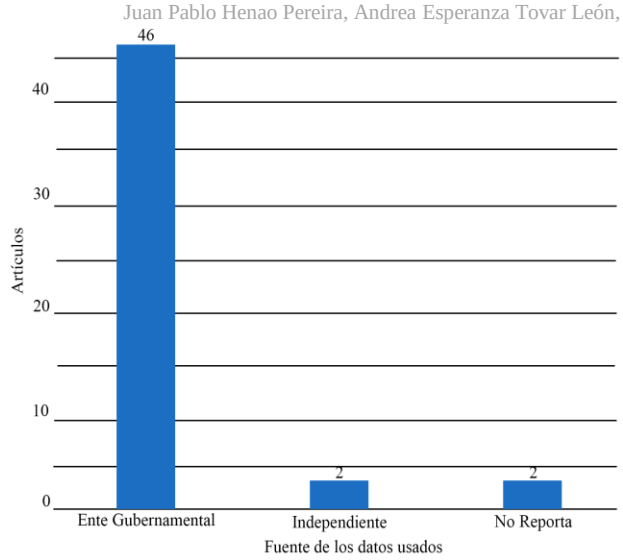


Figura.5: Origen de los datos usados en las publicaciones.
Fuente: Elaboración propia.

La calidad de los resultados de una investigación depende en gran medida del origen confiable, fidedigno y transparente de los datos, así como de los procedimientos aplicados durante su operacionalización. Buena parte de los trabajos inspeccionados en esta revisión bibliográfica hacen uso de datos que provienen principalmente de entidades gubernamentales, entre los que se destacan la policía, los departamentos de tránsito, observatorio de delitos, administración nacional de seguridad vial, ministerio de transporte, ministerio de salud, bases de datos hospitalarias y de cementerios, de donde se obtienen los reportes oficiales de los accidentes de tránsito. De igual manera, se encontraron estudios realizados a partir de cifras de personas involucradas en AT sin determinar la procedencia de los datos o la fuente que los reporta.

Los datos usados en los trabajos relacionados con AT normalmente se presentan en tablas estructuradas, que en su mayoría tienen 10 o más atributos, siendo los más comunes: fecha, día, hora, ciudad, municipio o localidad, lugar, género de la víctima, coordenadas del lugar del evento, estado del conductor, uso del cinturón de seguridad en automóviles, uso de casco en motociclistas, estado del clima, condiciones de la vía, condiciones de iluminación, inclinación de la vía, velocidad del vehículo, etc.

En India, se identificó un gran interés de las autoridades y los investigadores para analizar los AT a través de minería de datos, la mayor parte de los datos usados en estudios provienen de fuentes como la Policía, son recopilados en carreteras nacionales de alto tráfico como la National Highway (NH6), en donde ocurre un alto número de accidentes de tránsito, algunas variables consideradas fueron: hora del accidente, condiciones climáticas, lado de la vía donde se presentó el suceso (derecho, izquierdo, centro), condición de la vía, naturaleza del accidente (deslizamiento, parte trasera, colisión lateral), y tipo de accidente [33]; para el caso de la State Highway (SH) se examinaron datos que se consideran de influencia sobre accidentes como: tipo de intersección, carretera recta, si el accidente ocurrió cerca de una curva, tipo de lugar y tipo de accidente [32]. Otra fuente de información es el Instituto de Investigación y Gestión de Emergencias (Emergency Management and Research Institute) en donde se encuentran reportes de AT con atributos como la condición de iluminación de la vía, condición de la carretera (intersección, cuesta, curva), características en los alrededores del sitio del accidente, tipo de carretera, tipo de accidente [26] [37] [38].

En Estados Unidos, la Universidad de Alabama es la entidad que más información compila referente a AT, se reúnen atributos específicos del accidente, condiciones de la carretera, y datos relacionados con el medio ambiente [36]; en el trabajo de Shiau et al. [40] además de las características ya mencionadas, se consideran otras variables como la prueba de alcohol y de estupefacientes. En el caso de Ghomi et al. [31], estudiaron los AT donde estaban involucrados vehículos y ferrocarriles, usando datos de la administración federal de ferrocarriles, examinaron variables como la velocidad del tren, tipo de vehículo, tipo de accidente, condiciones del clima, iluminación en la vía, carretera pavimentada, y ángulo de colisión.

En España, la institución encargada de recopilar la información es la Dirección General de Accidentes de Tránsito del Ministerio de Transporte; el estudio de Abellán et al. [28] se ocupa de los accidentes registrados en áreas rurales de poblaciones españolas, los atributos elegidos fueron el tiempo (fecha y hora), tipo de vehículo, factores atmosféricos, condiciones de la vía, datos de la víctima (sexo, edad) y severidad del accidente, otros aspectos que resultan importantes para estudiar AT son la densidad del tráfico, superficie y la obstrucción de la visibilidad del conductor además de parámetros asociados a la geometría de la vía [29].

En China, una de las principales fuentes de datos lo constituyen los organismos de atención de AT que recopilan y posteriormente almacenan en bases de datos gubernamentales disponibles por entes públicos mediante el sistema de información de accidentes de tráfico administrado por el Ministerio de Seguridad Pública [20]; estos recursos sirven de base para el análisis por parte del Departamento de Policía de Tránsito, adicionalmente dichos reportes se encuentran disponibles en diferentes ministerios como por ejemplo, el Ministerio de Transportes y Comunicaciones [22], o la Agencia Nacional de Policía [40], en el caso de accidentes ferroviarios solo hasta el 2015 se hicieron públicos los datos de víctimas involucradas a través del Sistema Ferroviario de China [41].

En el caso de Colombia, la fuente que proporcionó los datos para realizar las investigaciones fue el Observatorio de Delitos del municipio de Pasto, los reportes muestran variables agrupadas por condiciones de tiempo, lugar e información de la persona involucrada en el accidente (edad, sexo, zona de residencia, ocupación, etc.) [16]; Timaran et al. [35] optaron por elegir otros atributos como el día, trimestre, barrio, comuna, y ubicación geográfica (latitud y longitud).

## IV.EJE TEMATICO 2: PAÍSES QUE HAN REALIZADO TRABAJOS DONSE SE ANALIZAN AT USANDO TÉCNICAS DE MINERÍA DE DATOS

Debido al impacto de los AT, el objetivo de diversas organizaciones tanto gubernamentales como independientes en todo el mundo, es establecer políticas preventivas para reducir significativamente pérdidas humanas, económicas y materiales. Usando minería de datos es posible analizar los datos recopilados de un evento, identificar e interpretar la relación entre ellos y generar un impacto en la sociedad, ya que los resultados permiten orientar campañas específicas y determinar situaciones comunes donde hay una mayor exposición a sufrir un accidente como: condiciones geográficas, ambientales, estado de la vía, características del conductor, etc. [18-19]. En el marco de esta revisión literaria se encontraron varios proyectos desarrollados en diferentes países, los cuales han empleado técnicas de minería de datos para estudiar las características asociadas a los AT, en la Figura 6 se detalla el número de estudios realizados de acuerdo con el país de origen.

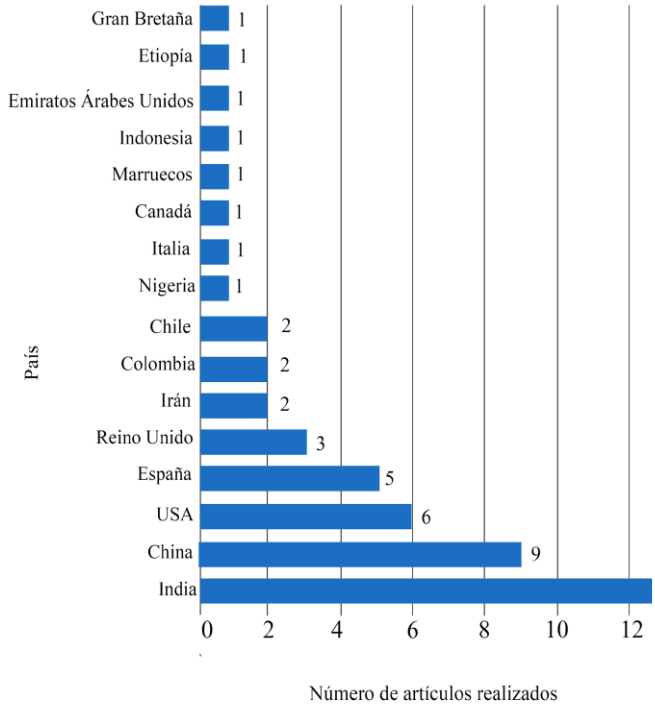


Figura.6: Países que han usado técnicas de minería de datos para analizar AT.
Fuente: Elaboración propia.

Unos de los temas más tratados por los investigadores son las causas externas [16] [20-24], estas son analizadas en países como India, China, Estados Unidos y otros. Se destacan dos artículos de la India que analizan accidentes donde estuvieron involucrados los vehículos más comunes en ese país conocidos como "auto-rickshaws" o "tuc-tuc" [25] y las motocicletas [26].

Otro foco de estudio es la gravedad de los accidentes [27-29], que resulta de interés en países como India, Estados Unidos y España. Aunque algunos trabajos exponen el mismo objetivo se pueden encontrar diferentes enfoques, por ejemplo, los autores Sakhare y Kasbe [30] se centran en encontrar las razones y riesgos que dan lugar a un accidente de tránsito considerando el nivel de la lesión, mientras que Gupta et al. [18] hacen una clasificación en tres categorías: fatal, lesión grave y lesión leve. Además, se destaca un artículo que estudia los factores de gravedad en las lesiones de conductores de vehículos accidentados en cruces de ferrocarril [31].

Algunas publicaciones de India, China y Colombia analizan y estiman los puntos o zonas propensas a accidentes, o condiciones en la malla vial que puedan aumentar la probabilidad de colisión [33-36]. De otra parte, el trabajo de Kumar y Toshniwal [37] se enfoca en predecir la forma de colisión de los vehículos.

## V. EJE TEMATICO 3: TÉCNICAS DE MINERÍA DE DATOS USADAS PARA ANALIZAR AT

En la literatura se encuentran reportados casos de estudio que usaron diferentes técnicas de minería de datos para analizar la accidentalidad en países con situaciones de tránsito complejas (aumento de parque automotor, infraestructura vial, entre otras). Esta revisión permitió identificar un amplio número de algoritmos para el estudio de AT, como se muestra en la Figura 7.

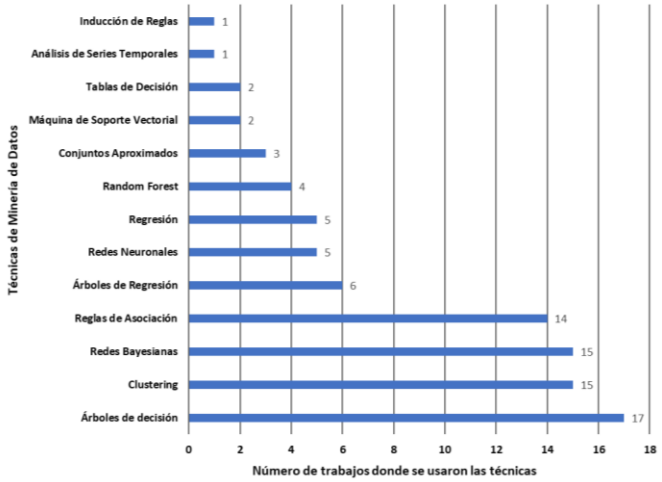


Figura 7: Técnicas de minería de datos usadas en el análisis de AT.
Fuente: Elaboración propia.

Buena parte de los resultados presentados en los artículos inspeccionados se basan en la identificación de patrones que permiten establecer factores que ocasionaron los accidentes en las vías [18] [40], aportan conocimiento valioso para establecer planes de prevención de accidentes y gestión de la seguridad del tránsito [21]. Cabe mencionar que un solo trabajo pudo haber hecho uso de una o varias técnicas de minería de datos para abordar un problema determinado, como se muestra en la Figura 8.

Un tema muy estudiado es el análisis de las causas externas que tienen influencia sobre los siniestros en la vía como se presenta en [14] [18] [23] [25] [30] [31] [35] [37] [38-46], los autores usaron técnicas como Clustering, Reglas de Asociación, Redes Bayesianas, Máquinas de Soporte Vectorial, Árboles de Decisión, Random Forest, entre otras. En la investigación de Kumar y Toshniwal [26] se compararon tres algoritmos de clasificación: Árboles de decisión, Naïve Bayes y Máquinas de Soporte Vectorial, se alcanzó una mayor precisión con el primero encontrando que las condiciones de iluminación en las noches son un factor relevante que puede causar los accidentes, teniendo un efecto similar las curvas y las pendientes.

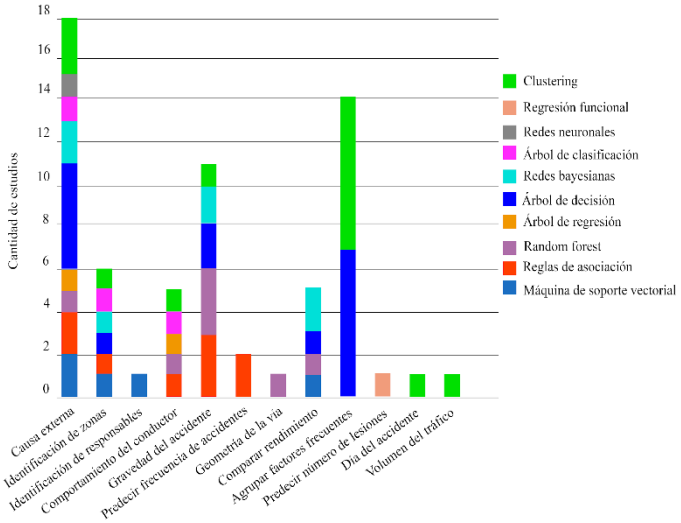


Figura 8: Técnicas de minería de datos usadas para el análisis de diferentes aspectos de los AT.
Fuente: Elaboración propia.

Otra temática muy discutida es la identificación de los factores generadores de accidentes de tránsito como las condiciones climáticas, características de la superficie, ubicación del siniestro vial, velocidad del automotor, estado del conductor y frecuencia de los AT, como se argumenta en [14] [18] [25] [30]; en los artículos [37] [38] [42] utilizaron conjuntamente algoritmos de Clustering y Reglas de

Asociación para analizar los datos, esta combinación es bastante usual, la primera técnica se usa preliminarmente para realizar una selección más homogénea de los datos y posteriormente mediante reglas de asociación se identifican patrones en cada agrupación, de igual forma, se emplearon reglas de asociación sobre el dataset completo a fin de comparar los resultados de ambos procedimientos.

La gravedad de los AT ha sido estudiada a través de técnicas como Reglas de Asociación, Random Forest, Árboles de Decisión, Redes Bayesianas, Random Forest y Clustering [9] [31] [45] [47-53]. Dadashova et al. [29] analizaron la relación existente entre la gravedad de los AT y la geometría de las vías, identificaron variables relevantes como el ancho de la vía, el peralte y la pendiente, también evidenció que la obstrucción de visibilidad del conductor aumenta la gravedad de los choques frontales, accidentes de ángulo, impactos laterales y accidentes de múltiples vehículos.

La identificación de zonas propensas a la ocurrencia de accidentes en la vía fue analizada en [34] [53-55], se emplearon algoritmos como Máquinas de Soporte Vectorial, Reglas de Asociación, Árboles de Decisión, Árboles de Clasificación y Clustering. En el trabajo de Muhammad et al. [43] desarrollaron un modelo predictivo que permitió ubicar los lugares críticos de los siniestros viales, las causas y los tipos de personas que pueden estar involucradas; por su parte, Manasa et al. [33] emplearon Árboles de Clasificación con datos espaciales, para identificar puntos críticos donde ocurren los accidentes y evidenciaron la importancia de este tipo de datos en el análisis de la gravedad de un accidente.

El comportamiento de los conductores, por ejemplo, manejo imprudente, desconocer las normas de tránsito, falta de experiencia al volante, o permanecer bajo los efectos del alcohol o sustancias sicotrópicas, fue objeto de estudio a través de algoritmos como como Reglas de Asociación, Árboles de regresión, Árboles de clasificación, Random Forest y Clustering [47-49] [56-57]. El estudio de Chen et al. [41] analizó los datos históricos ferroviarios y mediante Reglas de Asociación, identificó como motivo principal de los accidentes el error humano y causas externas como el clima; además se evidencio que los accidentes de colisión causan más víctimas que los accidentes de descarrilamiento. En la investigación de Pakgohar et al. [58] usaron la minería de datos para analizar la base de datos de la policía de tránsito a través de regresión logística multinomial y Árboles de Regresión y clasificación para predecir la gravedad de los accidentes de tráfico; los resultados mostraron además la incidencia de factores humanos como el permiso de conducir y el uso del cinturón en la gravedad de los accidentes viales.

La comparación de algoritmos de minería de datos busca identificar aquellos que generan resultados con mayor eficiencia, en [9] [26] [36] [49] [59] se equipararon técnicas como Árboles de Decisión, Naïve bayes y Reglas de Inducción. En el trabajo de Ramani y Shanthi [46] se confrontaron algoritmos de clasificación tales como Decision Stump (árboles de decisión de un nivel), Random Forest y Árboles de Decisión (generados con los algoritmos C4.5 y J48), con el propósito de construir modelos de predicción considerando aspectos como las características de los peatones, el comportamiento de los conductores, el estado de las carreteras y las condiciones climáticas relacionadas con la gravedad de los accidentes; los resultados mostraron que el algoritmo Random Forest alcanzó una mayor precisión.

En Colombia, la accidentalidad vial ha sido la causa de un significativo número de pérdida de vidas humanas, desde el sector público y privado se han realizado valiosos esfuerzos para sensibilizar a los ciudadanos sobre las causas más comunes de accidentalidad y sus consecuencias, a fin de reducir estas cifras [60-61]. En el país el uso de minería de datos aplicado al estudio de accidentalidad es poco común. En [17] se usó la herramienta Weka para establecer patrones a partir de registros de accidentalidad por causa externa con lesiones fatales y no fatales en la ciudad de Pasto, usando Árboles de decisión, los resultados mostraron que las personas mayores son la población más vulnerable en un accidente de tránsito.

Finalmente, otros temas estudiados fueron la identificación de responsables en los accidentes de tránsito mediante Máquinas de Soporte Vectorial multiclase [62], predicción del número de lesiones en accidentes de tránsito a través de un modelo de Regresión funcional [63], análisis del volumen vehicular y el día del siniestro vial usando técnicas de Clustering que permitieron identificar patrones de tráfico [54.

## VI. DISCUSION

Se identificaron siete revisiones literarias que abordan el uso de técnicas de minería de datos en el análisis de accidentes de tránsito [64-70], fueron publicados a partir del año 2015 y tienen como factor común la identificación de variables y patrones que permitan analizar y predecir la ocurrencia de AT, sin embargo, la estructura metodológica difiere en cada uno los artículos.

Los trabajos [64], [66] y [70] soportan su análisis en la elaboración de cuadros comparativos de las técnicas usadas, se asemejan a esta investigación que consideró una serie de factores externos como causantes de accidentes de tránsito, no obstante, los autores no realizaron un agrupamiento minucioso de los agentes externos estudiados respecto a cada técnica. Adicionalmente, en [66] y [70] realizaron comparativas entre las herramientas software más usadas, pero sin tener en cuenta su rendimiento respecto a las técnicas. Solo se encontraron tres casos [65], [67], [69], en los cuales se maneja un enfoque diferente, donde los investigadores analizan el rendimiento de las técnicas más usadas por otros autores para aplicarlas haciendo uso de datos propios.

En este trabajo se identificaron y analizaron diferentes técnicas de minería de datos asociando su uso con diferentes países que han trabajado esta temática; solamente en el trabajo de Gupta. et al. [68] se hace una revisión de técnicas y la metodología usada por los autores, pero se presentan de manera individual y aislada, lo que no brinda al lector profundidad en el análisis ni una correlación entre los casos para un estudio posterior; identifican la técnica con mejor rendimiento obtenido de acuerdo con las variables, sin embargo, omiten detalles como la relación entre las problemáticas abordadas en cada país y el algoritmo usado.

De manera general, se encontró que dentro de la literatura estudiada tienen referentes anteriores al año 2000, lo que no brinda información actualizada en el tema y tampoco realizaron un consolidado de países y el número de trabajos desarrollados como si se analizó en este trabajo.

Por último, se buscó identificar las fuentes de datos con las cuales los autores realizaron las investigaciones sobre accidentes de tránsito; los siete trabajos brindan una perspectiva de las técnicas usadas, pero hay poca claridad en el origen de los datos, sin embargo, se destacan [65] y [70] donde afirman que los datos usados tienen origen estatal.

Para el desarrollo de este trabajo los referentes bibliográficos fueron la base fundamental, debido a que los artículos de revisión requieren un alto nivel de investigación es notable el bajo número de fuentes de referencia usadas entre los artículos encontrados, en promedio estos siete artículos solo utilizan 16 trabajos y no presentan la metodología empleada para realizar la búsqueda de los artículos que sustentan su análisis, sin embargo, es de destacar que todos los artículos estudiados provienen la India, lo que demuestra la importancia de la investigación en ese país.

## VII. CONCLUSIONES

India es el país que origina la mayor cantidad de trabajos reportados seguido por China y Estados Unidos. Sin embargo, cada país tiene diferentes intereses investigativos, China e India se inclinan por identificar las principales causas de accidentes de tránsito mientras que en Estados Unidos y España la preocupación es determinar a priori la gravedad de los accidentes de tránsito. En Colombia el foco es la valoración de causas de accidentes de tránsito y su ubicación geográfica.

Este tema ha generado interés en las entidades gubernamentales vinculadas a la atención y gestión de AT, formulando políticas y planes de prevención; adicionalmente se tiene acceso a bases de datos y sitios web para que el público en general tome conciencia de la gravedad de los eventos y además pueda consultar información de estos incidentes, en portales como el observatorio del delito de la Policía Nacional o Datos Abiertos Colombia.

Las principales técnicas que los autores utilizaron fueron los Árboles de Decisión, Clustering, Reglas de Asociación, y algoritmos para predicción de evento futuros como Redes Bayesianas, Maquinas de Soporte Vectorial, Árboles de Clasificación, Random Forest y Redes Neuronales. Sin embargo, existen pocos reportes de Análisis de Series de Tiempo usando ARIMA o sus variantes, y técnicas de Regresión, entre otras. Una posible causa de esta carencia es el tipo de variables utilizadas en los proyectos; entre ellas: la hora, fecha, ubicación, las condiciones climáticas y de iluminación, el estado de la malla vial, características del conductor, visibilidad, velocidad del vehículo, la geometría y el gradiente de las vías.

El informe global sobre seguridad vial de la OMS [3] revela que India tiene un rango de ingresos similar a Colombia (US$1.006 - US$12.235), un producto interno bruto per cápita mucho menor (Colombia: US$6.320 – India: US$1.680) y un número de muertes por accidentes de tránsito 21 veces mayor (Colombia: 7158 – India: 150785), sin embargo, desde 2010 se encontraron 13 investigaciones que analizan los AT a partir de técnicas de minería de datos. En razón a lo anterior, India constituye un referente para examinar sus experiencias de gestión de tráfico para adaptarlas a las realidades del contexto colombiano.

Para iniciar un estudio de AT en Colombia, un punto de partida son los datasets disponibles en portales gubernamentales y el uso de algoritmos descriptivos (reglas de asociación, cluster). No obstante, para análisis de tipo predictivo, muchos de los conjuntos de datos abiertos carecen de variedad y completitud: no es posible establecer el lugar exacto donde ocurrió un accidente debido a la ausencia de coordenadas, o información del estado de la vía, o características de las personas involucradas en el hecho. En el mejor de los casos, estos análisis se pueden realizar para ciudades específicas, pero no para todo el territorio nacional.

Para trabajos futuros se propone que el contexto de la investigación este soportado en técnicas descriptivas de minería de datos, utilizando datos abiertos para ciudades específicas de Colombia, como el caso de Neiva (33 variables disponibles), adaptando propuestas de investigaciones realizadas en India. Es importante vincular diferentes fuentes de datos que resulten complementarias, permitiendo de esta forma subsanar los problemas de completitud y variedad.

## VIII. REFERENCIAS


[1] Ministerio de Transporte, “Manual para el diligenciamiento del formato del informe policial de accidentes de tránsito adoptado según resolución 004040 del 28 de diciembre de 2004 modificada por la resolución 1814 del 13 de julio de 2005.” 2006, p. 4. [Online] Disponible en: http://web.mintransporte.gov.co/rnat/app/ayudas/Resolucion_0011268_2012.pdf

[2] S. Kumar y D. Toshniwal, “Analysing road accident data using association rule mining,” in 2015 International Conference on Computing, Communication and Security (ICCCS), pp. 1–6, 2015.

[3] World Health Organization, “Global status report on road safety 2018: summary” 2018. [Online] Disponible en: https://www.who.int/violence_injury_prevention/road_safety_status/2018/English-Summary-GSRRS2018.pdf?ua=1

[4] OISEVI, “VII Informe Iberoamericano de Seguridad Vial” 2016, p. 22. [Online] Disponible en: https://www.oisevi.org/a/images/files/informes/info-7.pdf

[5] División de Población de las Naciones Unidas, “Population Total” The World Bank Group, 2017. [Online]. Disponible en: https://data.worldbank.org/indicator/SP.POP.TOTL?name_desc=false.

[6] Observatorio nacional de seguridad vial, “Cifras para Colombia Fallecidos y Lesionados en hechos de tránsito” 2018, pp. 2,6. [Online] Disponible en: http://ansv.gov.co/observatorio/public/documentos/boletin.pdf

[7] Asociación Colombiana de vehículos automotores, “Andemos: Informe Sector Automotor Colombia febrero 2018, ¿recuperación a la vista?” 2018. [Online] Disponible en: http://www.andemos.org/index.php/2018/03/02/andemos-informe-sector-automotor-colombia-febrero-2018-recuperacion-a-la-vista/.

[8] W. Hasperué, “Extracción de conocimiento en grandes bases de datos utilizando estrategias adaptativas”, Universidad Nacional de La Plata, 2012.

[9] S. I. Kabeer, “Analysis of Road accident in Leeds”, in Leeds MSc Research Project Data Analytics, 2016.

[10] E.A. Oviedo-Carrascal, A.I. Oviedo-Carrascal y G.L. Velez-Saldarriaga, “Minería multimedia: hacia la construcción de una metodología y una herramienta de analítica de datos no estructurados,” Rev. Ing. Univ. Medellín, vol. 16, no. 31, pp. 125–142, Dec. 2017.

[11] E.J. Hernández-Leal, N.D. Duque-Méndezy J. Moreno-Cadavid, “Big Data: una exploración de investigaciones, tecnologías y casos de aplicación”, TecnoLógicas, vol. 20, no. 39, mayo -agosto, 2017.

[12] J. Hernández-Orallo, M.J. Ramírez-Quintana y C. Ferri-Ramírez, Introducción a la minería de datos, 1st ed. Pearson Education, 2004.

[13] J.E. Rodríguez-Rodríguez, "Fundamentos de mineria de datos. Universidad Distrital Francisco José de Caldas", 2010.

[14] J. Hernández-Cáceres, “Clustering basado en el algoritmo K-means para la identificación de grupos de pacientes quirúrgicos,” in Congreso Académico UDI 2016, 2016.

[15] S. Kumar y D. Toshniwal, “A data mining approach to characterize road accident locations” J. Mod. Transp., vol. 24, pp. 62–72, 2016.

[16] G.R. Macías, N. Almeida-Filho y M. Alazraqui, “Análisis de las muertes por accidentes de tránsito en el municipio de Lanús, Argentina, 1998-2004,” Salud Colect., vol. 6, no. 3, p. 313, Dec. 2010.

[17] R. Timaran-Pereira, A. Calderón-Romero y A. Hidalgo-Troya, “Aplicación de los árboles de decisión en la identificación de patrones de lesiones fatales por causa externa en el municipio de Pasto, Colombia,” Univ. y Salud, vol. 19, no. 3, p. 388, Dec. 2017.

[18] M. Gupta, V. Kumar-Solanki y V. Kumar-Singh, “A Novel Framework to Use Association Rule Mining for classification of traffic accident severity,” Ing. Solidar., vol. 13, no. 21, pp. 37–44, Jan. 2017.

[19] M. Hassinger-Rodríguez, M. Ramírez-Quintana y C. Ferri-Ramírez, “Aplicación de técnicas de minería de datos en

accidentes de tráfico" p. 3, 2014.
[20] T. Chen, C. Zhang y L. Xu, "Factor analysis of fatal road traffic crashes with massive casualties in China" Adv. Mech. Eng., vol. 8, p. 1, 6, 2016.
[21] Y. Yanbin, Z. Lijuan, L. Mengjun y S. Ling, "Early Warning of Traffic Accident in Shanghai Based on Large Data Set Mining" 2016 Int. Conf. Intell. Transp. Big Data Smart City, p. 19, 2016.
[22] L.Y. Chang, H.C. Chu, D.J. Lin y P. Lui, "Analysis of freeway accident frequency using multivariate adaptive regression splines" Procedia Eng, vol. 45, pp. 824–829, 2012.
[23] R. Tian, Z. Yang y M. Zhang, "Method of Road Traffic Accidents Causes Analysis Based on Data Mining" in 2010 International Conference on Computational Intelligence and Software Engineering, pp. 1,2. 2010.
[24] J. Wang y Y. Ohsawa, "Evaluating model of traffic accident rate on urban data" in 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), p. 185, 2016.
[25] S. Vasavi, "Extracting Hidden Patterns Within Road Accident Data Using Machine Learning Techniques" p. 14, 19. 2018.
[26] S. Kumar y D. Toshniwal, "Severity analysis of powered two wheeler traffic accidents in Uttarakhand, India," Eur. Transp. Res. Rev., vol. 9, no. 2, p. 24, jun. 2017.
[27] H. Al-Najada y I. Mahgoub, "Big vehicular traffic Data mining: Towards accident and congestion prevention" in 2016 International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 256,257,260. 2016.
[28] J. Abellán, G. López y J. De Oña, "Analysis of traffic accident severity using Decision Rules via Decision Trees" Expert Syst. Appl., p. 6047,6053,6054. 2013.
[29] B. Dadashova, B.A. Ramírez, J.M. McWilliams y F.A. Izquierdo, "The Identification of Patterns of Interurban Road Accident Frequency and Severity Using Road Geometry and Traffic Indicators" Transp. Res. Procedia, vol. 14, pp. 4122,4123,4128. 2016.
[30] A.V. Sakhare y P.S. Kasbe, "A review on road accident data analysis using data mining techniques," in 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017, pp. 1–5.
[31] H. Ghomi, L. Fu, M. Bagheri y L.F. Miranda-Moreno, "Identifying vehicle driver injury severity factors at highway-railway grade crossings using data mining algorithms" in 2017 4th International Conference on Transportation Information and Safety (ICTIS), p. 1054. 2017.
[32] G. Kaur y E.H. Kaur, "Prediction of the cause of accident and accident prone location on roads using data mining techniques" 8th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT pp. 1, 4, 2017.
[33] J.M. Manasa, S. Bhattacharjee, S.K. Ghosh y S. Mitra, "Spatial Decision Tree for Accident Data Analysis" 9th International Conference on Industrial and Information Systems (ICIIS). pp. 1,2. 2014.
[34] L. Martín, L. Baena, L. Garach, G. López y J. de-Oña, "Using Data Mining Techniques to Road Safety Improvement in Spanish Roads" Procedia - Social and Behavioral Sciences, vol. 160. p. 607. 2014.
[35] R. Timaran-Pereira, G. Hernandez y N. Quemá-Taimbud, "Identificación Georreferenciada de Patrones de Lesiones no Fatales con Técnicas de Aprendizaje no Supervisado," in Proceedings of the 15th LACCEI International Multi-Conference for Engineering, Education, and Technology: "Global Partnership for Development and Engineering Education," 2017.
[36] S.Shanthi, D. Ramani, "Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms" Int. J. Comput. Appl., vol. 35, p 30. 2011.
[37] S. Kumar y D. Toshniwal, "A data mining framework to analyze road accident data" J. Big Data, vol. 2, pp. 6,7. 2015.
[38] S. Kumar, D. Toshniwal y M. Parida, "A comparative analysis of heterogeneity in road accident data using data mining techniques" Evol. Syst., vol. 8, pp. 150,151. 2017.
[39] S. Shanthi y R.G. Ramani, "Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques" Proc. World Congr. Eng. Comput. Sci, vol. 1, pp. 1,2. 2012.
[40] Y.R. Shiau, C.H. Tsai, Y.H. Hung y Y.T. Kuo, "The Application of Data Mining Technology to Build a Forecasting Model for Classification of Road Traffic Accidents" Math. Probl. Eng., vol. 2015, pp. 1. 2015.
[41] D. Chen, C. Xu y S. Ni, "Data mining on Chinese train accidents to derive associated rules" Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit, vol. 231, p. 5, 2015.
[42] L. Li, S. Shrestha y G. Hu, "Analysis of road traffic fatal accidents using data mining techniques" IEEE 15th Int. Conf. Softw. Eng. Res. Manag. Appl., pp. 363,365. 2017.
[43] L.J. Muhammad et al., "Using Decision Tree Data Mining Algorithm to Predict Causes of Road Traffic Accidents, its Prone Locations and Time along Kano – Wudil Highway" Int. J. Database Theory Appl., vol. 10, pp. 197,202. 2017.
[44] Y. Castro y Y.J. Kim, "Data mining on road safety: Factor assessment on vehicle accidents using classification models" Int. J. Crashworthiness, vol. 21, pp. 1,2,3,7. 2016.
[45] J. Xi, Z. Gao, S. Niu, T. Ding y G. Ning, "A hybrid algorithm of traffic accident data mining on cause analysis" Math. Probl. Eng., vol. 2013, p. 4. 2013.
[46] R.G. Ramani y S. Shanthi, "Classifier prediction evaluation in modeling road traffic accident data" 2012 IEEE Int. Conf. Comput. Intell. Comput. Res., pp. 1–4, 2012.
[47] A.T. Kashani, A. Shariat-Mohaymany y A. Ranjbari, "A Data Mining Approach To Identify Key Factors of Traffic Injury Severity" Prelim. Commun. Saf. Secur. Traffic, vol. 23, pp. 11,13,16. 2011.
[48] A. Jain, G. Ahuja, Anuranjana y D. Mehrotra, "Data mining approach to analyse the road accidents in India" 5th International Conference on Reliability, Infocom Technologies and Optimization, ICRITO 2016: Trends and Future Directions, pp. 175,176. 2016.
[49] B. Atnafu y G. Kaur, "Survey on analysis and prediction of road traffic accident severity levels using data mining techniques in Maharashtra, India" Int. J. Curr. Eng. Technol., vol. 7, pp. 1973–1978, 2017.
[50] A. Castro, "Análisis de accidentes de tránsito en zonas urbanas y rurales usando minería de datos difusa" Pontificia Universidad Católica de Valparaíso, pp. 42,86. 2012.
[51] A.M. Addi, A. Tarik y G. Fatima, "An approach based on association rules mining to improve road safety in Morocco" International Conference on Information. pp 1-6. 2016.
[52] F. Babic y K. Zuskacova, "Descriptive and predictive mining on road accidents data" in IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 91. 2016.
[53] A. Irfan, R. Al-Rasyid y S. Handayani, "Data mining applied for accident prediction model in Indonesia toll road," in 4th International Conference on Engineering, Technology, and Industrial Application (ICETIA) 2017, 2018.
[54] G. Gecchele, R. Rossi, M. Gastaldi y A. Caprini, "Data Mining methods for Traffic monitoring data analysis: A case study" Procedia - Soc. Behav. Sci., vol. 20, pp. 455,463. 2011.
[55] T. Beshah y S. Hill, "Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in Ethiopia" AAAI Spring Symp. - Tech. Rep., vol. SS-10-01, no. 1997, p. 14. 2010.
[56] X.F. Zhang y L. Fan, "A Decision Tree Approach for Traffic Accident Analysis of Saskatchewan Highways" 26th IEEE Can. Conf. Electr. Comput. Eng., p. 2. 2013.
[57] S. An, T. Zhang, X. Zhang y J. Wang, "Evolution of Traffic Flow Analysis under Accidents on Highways Using

Temporal Data Mining" Intell. Syst. Des. Eng. Appl. (ISDEA), 2014 Fifth Int. Conf., pp. 454–457. 2014.

[58] A. Pakgohar, R.S. Tabrizi, M. Khalili y A. Esmaeili, "The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach," Procedia Comput. Sci., vol. 3, pp. 764–769, 2011.

[59] S. Krishnaveni y M. Hemalatha, "A Perspective Analysis of Traffic Accident using Data Mining Techniques" Int. J. Comput. Appl., vol. 23, p. 40. 2011.

[60] El tiempo, "Muertes en Colombia por accidentes de tránsito" 2018. [Online]. Disponible en: https://www.eltiempo.com/colombia/otras-ciudades/muertes-en-colombia-por-accidentes-de-transito-en-lo-que-va-del-2018-264096.

[61] MinTransporte, "MinTransporte y ANSV anuncian estrategia para reducir siniestros viales en el país" 2018. [Online]. Disponible en: https://mintransporte.gov.co/Publicaciones/mintransporte_y_ansv_anuncian_estrategia_para_reducir_siniestros_viales_en_el_pais

[62] E.A. Mohamed, "Predicting Causes of Traffic Road Accidents Using Multi-class Support Vector Machines" Journal of Communication and Computer 11(2014) 441-447 vol. 11, p. 441. 2014.

[63] C. Montt, N. Rodríguez, A. Valencia, L. Barba y J. Rubio, "Regresión funcional para predecir lesionados en accidentes de tránsito de la región de Valparaíso" no. June 2014, p. 2. 2015.

[64] A. Prasath y M. Punithavalli, "A Review on Road Accident Detection Using Data Mining Techniques," Int. J. Adv. Res. Comput. Sci., vol. 9, no. 2, pp. 881–885, 2018.

[65] P. Kasbe y A.V Sakhare, "A Review On Road Accident Data Analysis Using Data Mining Techniques," Int. Conf. Innov. Inf. Embed. Commun. Syst. A, pp. 2–6, 2017.

[66] M. Singh y A. Kaur, "A Review on Road Accident in Traffic System using Data Mining Techniques," Int. J. Sci. Res., vol. 5, pp. 2319–7064, 2016.

[67] R. Saravanya y M. Mangayarkarasi, "A Study and Analysis of Road Accident in Tamilnadu using Data mining Technique," Int. J. Recent Eng. Sci., vol. 2, no. 2349–7157, 2015.

[68] M. Gupta, V.K. Solanki y V.K. Singh, "Analysis of Datamining Technique for Traffic Accident Severity Problem: A Review," Proc. Second Int. Conf. Res. Intell. Comput. Eng., vol. 10, no. June, pp. 197–199, 2017.

[69] N. Divya, R. Preetam, A.M. Deepthishree y V.B. Lingamaiah, Analysis of Road Accidents Through Data Mining, vol. 500. Springer Singapore, 2019.

[70] B. Atnafu y G. Kaur, "Survey on analysis and prediction of road traffic accident severity levels using data mining techniques in Maharashtra, India," Int. J. Curr. Eng. Technol., vol. 7, no. 6, pp. 2277–4106, 2017.