

Análisis de causas de accidentes de tránsito en el Ecuador utilizando Minería de Datos

Edwin León-Pluas¹, Alvaro Pumares-Romero², Javier Guaña-Moya³, Pablo Loayza-Valarezo⁴, Darwin Naranjo-Villota⁵, Nelson Salgado-Reyes⁶

eleonp@unemi.edu.ec, alvaro.pumares@ant.gob.ec, eguana953@puce.edu.ec, pablo.loayza@andinagestion.com, andres.darwin@educacion.gob.ec, nesalgado@puce.edu.ec

¹ Universidad Estatal de Milagro, 91050, Milagro, Ecuador.

² Agencia Nacional de Tránsito, 15101, Tena, Ecuador.

³ Pontificia Universidad Católica del Ecuador, 170525, Quito, Ecuador.

⁴ AndinaGestión S.A., 17053, Quito, Ecuador.

⁵ Ministerio de Educación Zona 1, 040302, Carchi, Ecuador.

⁶ Pontificia Universidad Católica del Ecuador, 170525, Quito, Ecuador.

Pages: 540–547

Resumen: El presente trabajo pretende realizar un análisis de las causas fundamentales de accidentes de tránsito en el Ecuador, justificado por las altas tasas de este tipo de eventos. Para ello, se emplea la metodología de extracción del conocimiento KDD, y el algoritmo de minería de datos C4.5 que nos permite realizar árboles de decisiones y extraer reglas relacionadas a las causas. Para dichos análisis se toman los datos de la Agencia Nacional de Tránsito a nivel nacional. Se analizaron 14410 accidentes ocurridos en el país entre los años 2016-2018 y se recogieron un total de 9 variables. El algoritmo ofrece una alternativa a modelos tradicionales, identificado los patrones más importantes en los datos y evaluando las posibles asociaciones entre las variables recogidas.

Palabras-clave: Accidentes de tránsito, incidentes, ANT, minería de datos.

Analysis of causes of traffic accidents in Ecuador using Data Mining

Abstract: The present work intends to carry out an analysis of the fundamental causes of traffic accidents in Ecuador, justified by the high rates of this type of events. To do this, we use the KDD knowledge extraction methodology and the C4.5 data mining algorithm that allows us to make decision trees and extract rules related to the causes. For these analyzes, the data of the National Traffic Agency at the national level are taken. 14410 accidents occurred in the country between the years 2016-2018 were analyzed and a total of 9 variables were collected. The algorithm offers an alternative to traditional models, identifying the most important patterns in the data and evaluating the possible associations between the variables collected.

Keywords: Traffic accidents, incidents, ANT, data mining.

1. Introducción

En los últimos años la generación de las grandes empresas y el desarrollo de producción en todo el mundo, ha permitido una nueva visión al progreso nacional, por lo cual, el crecimiento incesante de los países, estados, ciudades y comunidades han generado una saturación de la movilización (Arellano Aguilar, 2019), y es por esta razón que la necesidad de movilización constante ha obligado a que las personas adquieran vehículos de forma incontrolada, con el fin de moverse de manera rápida y cómoda, generando la adquisición excesiva de autos en el Ecuador, por lo que ha permitido un aumento en los accidentes de tránsito, mismos que son cada vez más frecuentes.

Por lo dicho, (García, 2009) asemeja que es necesario generar más campañas de concientización tanto para los conductores y peatones para el uso de seguridad al conducir ya que los decesos de las personas se deben por falta de conciencia en las dos partes y las causas principales podrían ser exceso de velocidad, conducción en estado etílico, no respetar las señales de tránsito por parte de los conductores y peatones.

Es importante analizar que los accidentes de tránsito son frecuentes tanto en países industrializados como en los que están en vías de desarrollo (Thomson, 2010), por lo cual, en Ecuador en los últimos años ha existido una alta demanda en la compra de vehículos, ya que es indispensable el transporte del usuario de forma rápida a sus actividades diarias, lo cual ha generado que el tránsito vial haya traído como consecuencia incrementos en la congestión vehicular, demoras en la movilización, accidentes de tránsito, problemas ambientales por contaminación excesiva, y este aumento de fenómenos se debe en parte al acceso indiscriminado de los automóviles en el país.

En el Ecuador, existen altas tasas de accidentes de tránsito, por lo que el país ha experimentado un incremento incontrolable en los últimos años, por ello (García A. R., 2018) especifica que se han realizado pocos estudios en seguridad y salud en el trabajo, por tal razón es importante cambiar el paradigma de información ya que es importante que cada uno este consiente de las leyes que regulan a los conductores y peatones; es por esto que, (Algora-Buenafé, 2017) especifica que hace dos años las incidencias de tránsito fueron mayores en las carreteras de Guayaquil con un 23.3% y en Pichincha hubo un 17.1%.

Los accidentes de tránsito en el Ecuador representan una prioridad en el mejoramiento de la salud pública por que representa elevados costos económicos en educación vial y tienen un alto impacto social, por tal razón (Algora-Buenafé, 2017) y (Gómez García, 2016) atestiguan que, Ecuador está ubicado como el séptimo país en el mundo con una mayor tasa de mortalidad por accidentes de tránsito (AT). Esta afirmación está basada en estudios que afirman que, en el 2016, de 1976 muertes registradas, el 81,1% correspondió a hombres y el 18% a mujeres y los resultados dieron que la mayoría de las muertes corresponde a los conductores (52%), mientras que en 48% corresponden a decesos de mujeres.

Otro factor importante en los AT es el uso del celular, por ello (COMERCIO, 2018) describe que, según estadísticas de la Agencia Nacional de Tránsito (ANT),

a inicios del 2019 (enero - febrero del 2019), se han registrado alrededor de 1180 siniestros de tránsito en Quito, y asevera que la principal causa de estos accidentes son la distracción del conductor y la pérdida de control del vehículo por el uso del teléfono celular, mientras que la segunda más usual en AT es no respetar los límites de velocidad en las carreteras y como tercera causa importante es el irrespeto a las señales de tránsito.

En cuanto al usos del celular mientras se conduce, (telégrafo, 2018) certifica que, en los tres primeros meses del 2018, se registraron 6164 incidentes, de los cuales 1462 fueron causados por la distracción que genera el uso del teléfono celular mientras se conduce y producto de ese fenómeno 206 personas murieron. En cuanto al exceso de velocidad, el no respeto las señales de tránsito y conducir bajo la influencia de alcohol o drogas, (telégrafo, 2018) manifieste que en el año 2017 existieron 28967 accidentes de tránsito en el Ecuador; de los cuales, “el 17,66% fue por desatención del chofer, en su mayoría por usar el teléfono celular mientras conducían, por esa misma causa, la cifra ya sobrepasa el 23% en lo que fue el 2018”.

De acuerdo a los antes mencionado, (Mauricio., 2019) dice en el país no existe cultura a la hora de conducir, por lo cual se lo hace con efectos de alcohol, se utiliza el teléfono celular y no se respeta las señales y sumado a un vehículo en mal estado, los conductores provocan lamentables accidentes de tránsito.

2. Trabajos relacionados

De acuerdo a la investigación realizada por (Ordoñez Ordoñez, 2018), la Organización Panamericana de la Salud (OPS), asocia que las muertes por accidentes de tránsito son una de las causas primordiales en la mortalidad en la región de América Latina y que las lesiones causadas por estos accidentes representan la segunda causa de lesión externa en Latinoamérica.

Si se describe un accidente de tránsito, los conductores y peatones son parte de un conjunto de variables que podrían generar un accidente, es por esto que las personas que generan consciencia en estos accidentes tienden a modificar su comportamiento, y este fenómeno podría usarse para incitar a que las personas sean conscientes de los peligros que generan dichos accidentes y ellos creen un patrón para comportarse de manera más segura al conducir y al cruzar la calle (Wouters, 2000).

Es importante describir que, (Híjar-Medina, 2000) afirma que la tasa de motorización sigue patrones directamente proporcionales en cuanto a accidentes de tránsito, hecho que ha impactado al ámbito de la seguridad vial, así como también en la seguridad nacional de todas las personas, por ello, estos accidentes de tránsito se han convertido en un problema de Salud Pública reconocido a nivel mundial.

Dentro de los accidentes con mayor incidencia por, (Tomalá C., 2016) certifica que el 60% son por choques, el 17% por rozamientos, 10% a causa de atropellos y el 10% especifica que se dan por estrellamientos. Es importante aseverar que los servicios prestados por los vehículos son independientes a las víctimas que generan los AT, y el estado del vehículo es una parte importante en la generación de siniestros de tránsito.

Las consecuencias causadas por este tipo de accidentes, demanda una atención urgente del estado, ya que al ocurrir un incidente de tránsito el impacto que genera en el accidentado, la familia y la sociedad en general es devastador y lleva mucho tiempo en superar, afectando así fundamentalmente en la calidad de vida (Híjar-Medina, 2000). Esto ha llevado a que el país reconozca el costo económico y social que representa este fenómeno.

3. Metodología

La investigación se realizó a través de un estudio del arte referente a los accidentes de tránsito más comunes en el Ecuador, basado en información de artículos de revistas indexadas, así como también de fuentes de los diarios impresos como el comercio, el telégrafo, entre otros (Gómez García, 2016) (Tomalá C., 2016) (COMERCIO, 2018) (telégrafo, 2018) (Mauricio., 2019).

La recogida, análisis y procesamiento de datos se realizó por etapas, usando la metodología para la extracción del conocimiento “Knowledge Discovery in Databases” (KDD) mostrado en la Figura 1 y en particular la etapa de minería de datos presente en el proceso de KDD.

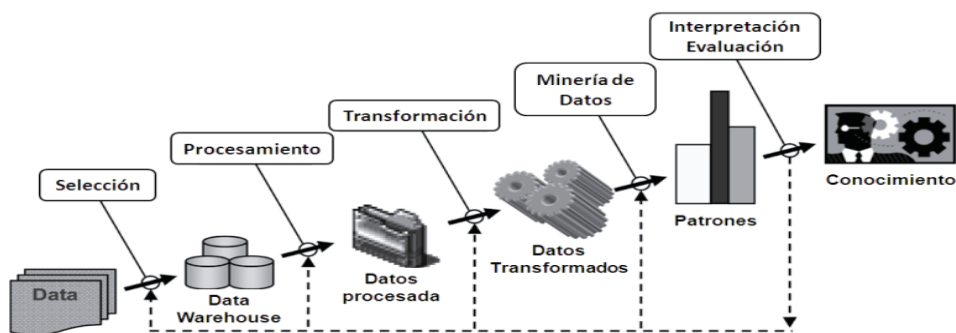


Figura 3.2 Proceso de Descubrimiento del Conocimiento (KDD)

Figura 1 – Etapa del proceso KDD

Fuente: (Larose, 2006)

Para el análisis de datos con el algoritmo C4.5 se generarán los árboles de decisión a partir de los datos mediante particiones realizadas recursivamente. El árbol se construye mediante la estrategia de profundidad-primero (depth-first).

El algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información. Para cada atributo discreto, se considera una prueba con n resultados, siendo n el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo en los datos. En cada nodo, el sistema debe decidir cuál prueba escoge para dividir los datos, estas pruebas propuestas están divididas en las siguientes:

- a. La prueba “estándar” para las variables discretas, con un resultado y una rama para cada valor posible de la variable.

- b. Una prueba más compleja, basada en una variable discreta, en donde los valores posibles son asignados a un número variable de grupos con un resultado posible para cada grupo, en lugar de para cada valor.
- c. Si una variable A tiene valores numéricos continuos, se realiza una prueba binaria con resultados $A \leq Z$ y $A > Z$, para lo cual debe determinarse el valor límite Z

Se desarrollaron las 5 fases básicas en la resolución de problemas de la Minería de datos, que son los siguientes: filtrado de datos, selección de variables, extracción del conocimiento, interpretación y evaluación.

3.1. Filtrado de datos

Para los análisis de datos se tomó la información de la Agencia Nacional de Tránsito a nivel nacional. Se analizaron 14410 accidentes ocurridos en el país entre los años 2016-2018 y se recogieron un total de 9 variables: Año en el que ocurre el accidente, día de la semana, hora en la que se registra el accidente, zonas de accidente, se clasifica según la causa, según la tipología, el tipo de carretera, y el factor clima.

3.2. Selección de variables

Una vez filtrada la base de datos y los accidentes objetos de estudio se colocaron en cada una de las variables descritas en el párrafo anterior, para determinar según su categoría definida por la ANT, el mayor índice que representa en el estudio a realizar, esto permite una caracterización de los mismos. Las variables recogidas, se presentan a continuación:

Nombre	Descripción	Categorías
<i>Año</i>	Año en el que ocurre el accidente	2016, 2017, 2018
<i>Día</i>	Día de la semana	Lunes, martes, miércoles, jueves, sábado y domingo
<i>Horarios</i>	Agrupado según hora en la que se registra el accidente	Día, tarde y noche
<i>Zona</i>	Basada en las 9 zonas del Ecuador	Zona1, Zona2, Zona3, Zona4, Zona5, Zona6, Zona7, Zona8, Zona 9 (rural, urbana)
<i>Sub-Zonas</i>	Subzonas establecidas en el país, estipuladas por la ATN	Azuay, Bolívar, Cañar, Carchi, Chimborazo, Cotopaxi, DMQ, E Oro, Esmeraldas, Galápagos, Imbabura, Loja, Manabí, Morona Santiago, Napo, Orellana, Pastaza, Pichincha, Santo Domingo, Sucumbíos, Tungurahua, Zamora.
<i>Causa</i>	Según la clasificación de la ATN	Adelantamiento inadecuado/invadir carril, bajarse o subirse de vehículos en movimiento sin precaución, calzada resbaladiza o granillo, casos fortuitos, conduce con falta de atención a las condiciones del tránsito, daños mecánicos, dormirse manejando (Impericia), encandilamiento, estado de embriaguez, exceso de velocidad, factor climático, fallas de iluminación en la vía, falta de atención en la conducción, falta de señalización, imprudencia peatón, mal estado de la vía, no ceder el derecho de vía o preferencia de paso a vehículos, no ceder el derecho de vía o preferencia de paso al peatón, no respetar las señales reglamentarias de tránsito, obstáculos en la vía, presencia de animales en la vía, transita bajo influencia de alcohol-Peatón)

Nombre	Descripción	Categorías
<i>Tipologia</i>	Según clasificación de la ATN	Arrollamiento, atropello con animal o con personas, caída de pasajero, choque frontal excéntrico, choque frontal longitudinal, choque lateral angular, choque lateral perpendicular, choque por alcance, colisión, encunetamiento, estrellamiento, pérdida de carril, pérdida de pista, roce negativo, roce positivo, rozamiento, volcamiento lateral y volcamiento longitudinal
<i>Tipo de carretera</i>		Adoquinado, asfalto, empedrada, hormigón, lastrado y tierra
<i>Clima</i>	Condición climática en la que sucede el accidente	Despejado, lluvia, neblina, nublado

Tabla 1 – Variables relacionadas a los accidentes

3.3. Extracción de conocimiento

En esta fase, se utiliza la técnica de la minería de datos basada en árboles de decisión C4.5 es un modelo de clasificación consistente en un diagrama que representa, en forma secuencial, las condiciones que se consideran en primer lugar, en segundo lugar y así sucesivamente hasta clasificar un patrón. Los árboles de decisión son normalmente contruidos a partir de la descripción de la narrativa de un problema. Así proveen una visión gráfica de la toma de decisión necesaria, especifican las variables que son evaluadas, las condiciones a considerar y su orden. Cada vez que se evalúa un dato con un árbol de decisión, sólo un camino será seguido hasta su clasificación dependiendo del valor actual de la variable evaluada. Para los datos de la La Tabla 1 de variables relacionadas a los accidentes, está fue dividida de forma aleatoria en dos conjuntos: conjunto de entrenamiento conformado por el 70% (10087 accidentes) de la muestra y el conjunto de verificación del modelo por el 30% (4323) restante.

Construcción del modelo. Utilizando el conjunto de entrenamiento se construye el modelo, utilizando para ello el software libre, “R”. El árbol de decisión se obtiene mediante el método de clasificación C4.5. La precisión del modelo construido fue de un 95%.

Una vez que se ha construido el árbol, se procede a la extracción de las reglas de decisión y, por último, se procede a la verificación de estas reglas utilizando para ello el conjunto de verificación.

Verificación del modelo. Se eliminan los patrones aleatorios de los datos, dejando aquellas reglas de entrenamiento que cumplen los parámetros establecidos.

La tabla 2 muestra un resumen de las reglas según nivel de soporte y confianza.

Regla Nº	Soporte	Confianza
1	64,8%	86,2%
2	50,3%	77,2%
3	51,8%	67,1%

Tabla 2 – Resultados del modelo de clasificación

3.4. Interpretación de Resultados

A partir de los resultados del modelo se pueden establecer 3 reglas importantes.

La regla con mayor soporte y confianza esta asociada a la causa relacionada con la conducción con falta de atención a las condiciones de tránsito. En este sentido, los accidentes se dan en zonas urbanas del Chimborazo, los domingos, en horas de la noche, tipo de carretera de asfalto y clima despejado; por lo que este último no constituye un agravante para esta causa. Además, en esta misma regla, se obtiene que el choque lateral es la tipología asociada a esta causa.

Por otro lado, la segunda regla obtenida indica que los accidentes ocurridos por estado embriaguez, cuya tipología son los choques laterales y estrellamientos; se presentan en zonas urbanas, carretera de asfalto y fundamentalmente en la noche. En este caso igualmente se presentan los fines de semana (sábados y domingos) con un clima despejado.

La otra regla de interés resulta la causa asociada a la falta de atención en la conducción con un 67,1% de confianza de suceder. La misma establece que los siniestros ocurridos son fundamentalmente por choques laterales, sobre carreteras de asfalto en horarios de la tarde y noche; fundamentalmente los fines de semana (sábado y domingo). De igual forma que en las reglas anteriores el clima no influye en el accidente pues ocurren en climas despejados.

4. Conclusiones

Los resultados obtenidos con el modelo de clasificación por árboles de decisión, indican que este es capaz de generar modelos consistentes con la realidad observada y el respaldo teórico, basándose únicamente en los datos que se encuentran almacenados en los datos de la ANT (Agencia Nacional de Tránsito).

Los porcentajes de instancias correctamente clasificadas presentados en el árbol como en la matriz de confusión muestran que el modelo tiene una precisión buena y por consiguiente es confiable para clasificar nuevos casos, especialmente los problemas que se presentan en la tipología según la clasificación de la ANT. El proceso de pospoda del árbol y el factor de confianza y soporte establecidos no permitió generar reglas para las muertes no intencionales.

El algoritmo de minería de datos C 4.5, nos ha permitido clasificar los datos de accidentes de tráfico en base a las causas de la misma. Por ello esta técnica ofrece una alternativa a los tradicionales modelos paramétricos, dada su capacidad de identificar patrones en los datos, sin la necesidad de establecer de antemano una relación entre las distintas variables que conforman el estudio. Por otro lado, estos modelos permiten obtener las posibles interacciones entre diferentes variables que no serían posibles por métodos tradicionales. El algoritmo nos ha identificado aquellas reglas de mayor utilidad y que pueden ser utilizadas por el personal encargado de analizar los datos de seguridad vial. El modelo construido presenta un buen nivel de precisión, aunque como trabajos futuros se pueden utilizar otras técnicas más complejas como redes neuronales o incorporar elementos de lógica difusa.

Referencias

- Algora-Buenafé, A. F.-B.-S.-G. (2017). Epidemiological study of fatal road traffic accidents in Ecuador. *Australasian Medical Journal (Online)*, 238.
- Arellano Aguilar, L. E. (2019). Modelo de predicción de puntos de exceso de velocidad en la autopista General Rumiñahui . *Bachelor's thesis, Universidad de las Fuerzas Armadas ESPE. Carrera de Ingeniería en Sistemas e Informática*.
- COMERCIO, D. E. (2018). El celular es la principal causa de accidentes de tránsito en Quito. *Diario EL COMERCIO*.
- García, A. R. (2018). Accidentes de tránsito relacionados con el trabajo: una prioridad en salud laboral y pública para Ecuador. *Notas y reflexiones*, 112.
- García, P. (2009). Registro de datos en temas de seguridad ciudadana, el caso del OMSC. Un lenguaje colectivo en construcción: el diagnóstico de la violencia. 49-70.
- Gómez García, A. R. (2016). Caracterización de la mortalidad por accidentes de tránsito en Ecuador, 2015.
- Híjar-Medina, M. C.-O.-A.-L. (2000). Factores de riesgo de lesión por accidentes de tráfico y el impacto de una intervención sobre la carretera. *Revista de saúde Pública*, 505-512.
- Larose, D. (2006). *Data Mining: Methods and Models*. New Jersey: John Wiley & Sons, Inc.
- Mauricio., M. (2019). Siat alerta causas de accidentes de tránsito en invierno.
- Ordoñez Ordoñez, F. A. (2018). Determinación de los Factores Influyentes en Accidentes de Transporte de Carga Un análisis exploratorio para la identificación de patrones I. In *XXI Simposio Internacional de Métodos Matemáticos Aplicados a las Ciencias*, 132.
- telégrafo, E. (2018). El 23% de accidentes se debe al uso de celular. *El telégrafo*.
- Thomson, I. &. (2010). La congestión del tránsito urbano: causas y consecuencias económicas y sociales. *CEPAL*.
- Tomalá C., R. E. (2016). Análisis Estadístico de las Causas de Accidentes de Tránsito en la Ciudad de Guayaquil.
- Wouters, P. I. (2000). Traffic accident reduction by monitoring driver behaviour with in-car data recorders. . *Accident Analysis & Prevention*, 643-650.

© 2020. This work is published under
<https://creativecommons.org/licenses/by-nc-nd/4.0/>(the
“License”). Notwithstanding the ProQuest Terms and
Conditions, you may use this content in accordance with the
terms of the License.