



UNIVERSIDAD TECNOLÓGICA ISRAEL
ESCUELA DE POSTGRADOS

TRABAJO DE TITULACIÓN EN OPCIÓN AL GRADO DE MAGISTER
PRODUCTO FINAL: ARTÍCULO CIENTÍCO

MAESTRÍA:
Telemática, mención Calidad en el Servicio.
LÍNEA DE INVESTIGACIÓN:
Telecomunicaciones y Sistemas Informáticos aplicados a la producción y la sociedad.
TÍTULO:
Minería de datos en el análisis de causas de accidentes de tránsito en el Ecuador.
AUTOR(A):
Ing. Alvaro Galilei Pumares Romero
TUTOR(A):
Dr. Edison Javier Guaña Moya

Quito, Ecuador

2019

INTRODUCCIÓN

En el Ecuador el campo automotor se ha incrementado sustancialmente, por los bajos costos e ingreso de fabricantes que están de la mano de la economía del país. La demanda de adquisición de un vehículo ha generado un tráfico en cada ciudad, por tal razón aumentado los accidentes de tránsito. Dentro de las estadísticas de siniestralidad de la Agencia Nacional de Tránsito registra diariamente siniestros de tránsito y sus causales siendo la principal causa de estos accidentes la distracción del conductor y la pérdida de control del vehículo por el uso del teléfono celular, mientras que la segunda más usual es no respetar los límites de velocidad en las carreteras y como tercera causa importante es el irrespeto a las señales de tránsito.

De acuerdo a la investigación realizada por 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,

Es importante describir 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.

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. Esto ha llevado a que el país reconozca el costo económico y social que representa este fenómeno.

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 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,

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 esto afirman que, Ecuador está ubicado como el séptimo país en el mundo con una mayor tasa de mortalidad por accidentes de tránsito (ANT). 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.

Dentro de los accidentes con mayor frecuencia, afirma 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.

Otro factor importante en los AT es el uso del celular, por ello, 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 ANT 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 uso del celular mientras se conduce, afirma 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, asevera 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 lo antes mencionado, 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 respetan las señales y sumado a un vehículo en mal estado, los conductores provocan lamentables accidentes de tránsito.

OBJETIVO GENERAL

1. Analizar las causas de tránsito en el Ecuador que determinen las mayores incidencias desde el año 2016-2018 para las posibles mejoras utilizando modelos predictivos de minería de datos.

OBJETIVOS ESPECÍFICOS

1. Identificar los siniestros, incidencias y causas de los accidentes de tránsito en las zonas del Ecuador
2. Obtener datos estadísticos reales de los accidentes de tránsito en el Ecuador de la Agencia Nacional de Tránsito para la utilización de minería de datos.
3. Estudiar las incidencias de tránsito en el Ecuador desde el 2016-2018 utilizando minería de datos
4. Determinar las posibles causas de mejoras de las incidencias que ocasionan los accidentes de tránsito en el Ecuador utilizando modelos predictivos en minería de datos

PERTIENENCIA DEL ARTÍCULO:

La importancia del presente tema de investigación radica en la necesidad de dar a conocer a la ciudadanía sobre el gran problema que se suscita a diario sobre las causas de accidentes de tránsito en el Ecuador.

En la investigación de tránsito uno de los aspectos que merece la especial atención de los técnicos es evitar los accidentes de tránsito en el sistema vial.

Si bien es cierto en el DMQ se crearon los Agentes Civiles de Tránsito quienes son encargados de controlar el tránsito y la circulación vehicular, pero que hacer para

disminuir los accidentes de tránsito es una de las preguntas que a diario se hace la Agencia Nacional de Tránsito.

En el Ecuador se ha emprendido en la investigación sobre los accidentes de tránsito registrados en los años 2016 - 2018.

La accidentalidad vial es un problema de salud pública, se ha tenido que recurrir a la información de la Agencia Nacional de Tránsito, analizando toda la información que en ellos se describen, con el objetivo de identificar el lugar, tipo y causa probable que ocasiona un evento de esta naturaleza.

En la investigación se llega a encontrar el número de accidentes del país, así como también se logró localizar las vías, intersecciones y tramos de vías en los que los accidentes son más recurrentes.

La información oportuna de los 14410 accidentes registrados en el período de análisis, ha permitido elaborar un análisis predictivo utilizando minería de datos y gráficos que facilitan el entendimiento de la problemática de la accidentalidad en el Ecuador.

Utilizando estos modelos predictivos no dan a conocer las posibles soluciones para mejorar la seguridad vial con el fin de reducir el índice de accidentalidad, mortalidad y morbilidad, que nos coloca como el séptimo país latinoamericano con más altos índices.

El presente Artículo de la Investigación radica en la necesidad de dar a conocer a la ciudadanía en general el gran problema que se suscita por el cometimiento de infracciones y contravenciones de tránsito, a su vez el artículo 55 de la Constitución de la República del Ecuador hace mención a que: Los gobiernos autónomos descentralizados municipales tendrán competencias exclusivas como planificar, construir y mantener la viabilidad humana y planificar, regular y controlar el tránsito y el transporte terrestre dentro de su circunscripción cantonal.

Por tal razón el presente Artículo de Investigación va encaminado a la sociedad ecuatoriana, para la reducción de los siniestros de tránsito en el Ecuador. Para ello se tomaron en cuenta varios conceptos relacionados a la materia de Tránsito con ciertos datos que servirán de ayuda para continuar con el presente tema de Investigación.

También el presente artículo va dirigido a las autoridades de control para la implementación de nuevas alternativas de modelamiento de indicadores de datos predictivos, utilizado minería de datos.

DESCRIPCIÓN DEL PROCESO INVESTIGATIVO

1. Contextualización del tema en el mundo profesional

El Plan Estratégico de Seguridad Vial 2015-2020, desarrollado por la Agencia Nacional de Tránsito en coordinación con actores gubernamentales, privados y de la sociedad civil, en consistencia con la proclamación del Decenio de Acción para la Seguridad Vial 2011-2020 realizada por Naciones Unidas, tiene como objetivo “Garantizar el derecho a las y los ciudadanos a la libre y segura movilidad terrestre, protegiendo la vida e integridad de los usuarios del transporte terrestre en el territorio ecuatoriano”, para lo cual se han establecido metas para cumplir con este objetivo que es disminuir en un 40% la siniestralidad nacional para el año 2020, de 16,4 vehículos accidentados en el año 2013 a 8,9 por cada 1.000 vehículos para el año 2020. Disminuir en el 40% las fatalidades en siniestros de tránsito hacia el año 2020; esto es, reducir la tasa de 19,47 fallecimientos en el 2013 a 11,7 fallecimientos por cada 100.000 habitantes en siniestros de tránsito; esto equivale a salvar más de 10.000 vidas en el periodo 2015-2020.

El accidente de tránsito es el daño o perjuicio que se le hace a una persona u objeto (bien material), durante el desplazamiento o conducción de un vehículo, provocando usualmente por la actitud negligente o irresponsable de los usuarios de la vía, sea este conductor o peatón, así como la acción adversa del medio ambiente o clima (neblina, lluvias torrenciales, inundaciones, deslaves, etc.), fallas mecánicas repentinas, cruce de animales en la vía pública, mala condición de la capa de rodadura o mala/nula señalización en las vías. Es decir, es importante considerar que su ocurrencia responde a una red de factores que aumentan la probabilidad de que sucedan, estando estos factores asociados a la persona, al vehículo y al ambiente.

Los países que menos índices de accidentes de tránsito presentan son aquellos que tienen ingresos altos en su economía: en ellos, dichas tasas oscilan entre 3,4 y 5,4 víctimas mortales por cada 100.000 habitantes y como ejemplo, tenemos a Suecia, Reino Unido y Países Bajos como líderes en esta área. En la ilustración 1, se puede observar la distribución de fatalidades en el mundo separada por regiones según la OMS.

Ecuador registra una tendencia creciente del número de muertes por accidentes de tránsito, siendo cerca de 18.600 el número de víctimas fatales acumulado en los últimos 12 años. Aunque se resalta varias acciones positivas tomadas por el gobierno, como son la legislación que restringe la velocidad a 50 km por hora en las zonas urbanas, los límites de alcoholemia que están entre 0.3 y 0.1 gramos por mililitro de sangre y el control al uso de celulares para los conductores.

El primer "INFORME SOBRE LA SITUACIÓN MUNDIAL DE LA SEGURIDAD

VIAL", publicado en el año 2009 por la Organización Mundial de la Salud (OMS), subraya dos hechos importantes:

- El 50% de las defunciones causadas por los siniestros de tránsito se producen entre los usuarios vulnerables de la vía pública, es decir, motociclistas, ciclistas y peatones.
- Solo el 15% de los países cuentan con una legislación integral sobre los principales factores de riesgo en materia de seguridad vial.

La Agencia Nacional de Tránsito es le encargada de controlar, coordinar y analizar los accidentes de tránsito en el Ecuador, los cuales se encargan de la reducción de siniestros de tránsito, por lo cual no llevan un análisis predictivo para bajar los índices de siniestralidad, la misma que no utiliza las herramientas necesarias para medir las causales de siniestros de tránsito.

En la actualidad nos encontramos inmersos en una época donde el volumen de datos generados está creciendo en volumen y a un ritmo relevante, de ahí la importancia que cobra la minería de datos, la evolución de la tecnología y la capacidad de computación nos permite adquirir conocimiento y ventajas competitivas mediante su explotación a través de la minería.

Los antiguos sistemas de almacenamiento no son capaces de proporcionar la capacidad suficiente para su gestión dando paso a sistemas capaces de atender las funciones de escritura y lectura con mayor capacidad y rapidez.

La evolución de la tecnología e infraestructuras permiten acceder y generar mayor cantidad de datos. La interpretación de los mismos mediante su manipulación, para encontrar asociaciones, relaciones, tendencias o patrones, permite convertirlos en información interpretable por los seres humanos.

2. Campo teórico

La investigación se realizó mediante la metodología de extracción del conocimiento KDD, lo datos obtenidos se eliminan o corrigen; se consideran únicamente aquellos atributos que van a ser relevantes (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004). En la fase Modelado se aplica el modelo, la tarea, la técnica y el algoritmo C4.5 seleccionado para la obtención de reglas y patrones. Luego en la fase de Evaluación se comprueban los patrones y se analizan, se puede regresar a la fase Análisis del Problema en caso de querer perfeccionar los resultados. Finalmente, en la fase de Explotación se comparte el nuevo conocimiento con los interesados.

Para dichos análisis se toman los datos de la Agencia Nacional de Tránsito a nivel nacional. Se analizaron 14.410 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, además, se utilizó conceptos de varios autores como Arellano Aguilar, Thomson, I., & Bull, ORDOÑEZ ORDOÑEZ García, A. R. G., Bermúdez, P. R. S., Yáñez, I. D. R., & Claudio, O. M. T., Tomalá C., Ribadeneira E, para el análisis de puntos críticos de accidentabilidad, registros y causales de mortalidad.

ORIGEN Y MOTIVACIÓN

Fuera del ámbito informático, el término “minería” hace referencia al arte de extraer minerales de la corteza terrestre. Gracias a la minería se obtienen materiales preciados como, por ejemplo, el oro o la plata. Por su parte, el término “dato” se define formalmente como el valor que toma una variable, parámetro, atributo, característica, etc. De forma más general, un dato es la representación, mediante algún símbolo (número, letra, etc.), de un atributo de una determinada entidad. A modo de ejemplo, la altura (atributo) de una persona (entidad) puede ser, supóngase, de 165 (dato) centímetros. En este caso, el dato está representado por medio de un número. El concepto “dato” está estrechamente ligado a la informática, existiendo, dentro de esta, disciplinas que estudian su obtención, almacenamiento, tratamiento, etc. Sin embargo, en su origen, la “minería” es un arte que nada tiene que ver con la informática y que existe mucho antes que esta. La unión de ambos términos, sin embargo, da lugar a un nuevo término cada vez más conocido en el mundo de la informática. Ese término no es otro que “minería de datos”. La minería de datos es una disciplina de la informática que estudia el análisis de grandes cantidades de datos con el objetivo de obtener conocimiento a partir de ellos. De igual forma que, mediante la minería tradicional, se persigue la obtención de minerales a partir de la corteza terrestre o del mar, la minería de datos persigue la obtención de conocimiento a partir de los datos. Siguiendo con el símil, los datos analizados serían la corteza terrestre y el conocimiento obtenido sería el mineral valioso. La figura 1 ilustra de forma gráfica qué es la minería de datos y su comparación con la minería tradicional. Una de las cuestiones que se podría plantear alguien que desconoce la minería de datos, es ¿qué tipo de conocimiento se puede obtener mediante la minería de datos?

Para conocer el origen del término minería de datos hay que remontarse años atrás. En la década de 1960, los estadísticos utilizaron los términos data fishing (buscar o “pescar” en los datos) o data dredging (dragado de datos) para referirse al análisis de datos. Con el tiempo, esos términos desaparecieron, dejando lugar, en la década de 1990, al término data mining (minería de datos), con el que se conoce actualmente al campo de la informática que estudia la obtención de conocimiento a partir de los datos por medio de su análisis.

El nacimiento de la minería de datos en la década de 1990 tiene mucho que ver con el hecho de que, en ese momento, la concepción que se tenía de los datos comenzó a

cambiar sustancialmente. Hasta entonces, los datos, almacenados en las bases de datos de empresas e instituciones, eran el soporte para la operativa diaria. Por ejemplo, una empresa de venta de recambios de automóvil almacenaba en sus bases de datos información acerca de los pedidos, productos, trabajadores, clientes, proveedores, etc., lo que permitía gestionar su stock, sus ventas, las nóminas de sus empleados, etc. Sin embargo, esa concepción inicial de los datos como mero sustento para la operativa del día a día cambió, y los datos empezaron a considerarse como una importante fuente de conocimiento que podría ser útil para obtener beneficios. Por otra parte, en las últimas décadas la cantidad de datos almacenados ha crecido enormemente en diferentes vertientes:

- En el ámbito empresarial, las diferentes compañías (bancos, empresas de telecomunicaciones, etc.) han aumentado sensiblemente su volumen de negocio y, como consecuencia, el volumen de datos almacenados. •
- En el campo científico, se han producido importantes avances en disciplinas como la astronomía, la biología o la medicina. En esas y otras áreas, una gran cantidad de datos son registrados y almacenados a velocidades enormes por medio de sensores, satélites, telescopios, dispositivos de simulación, etc. El análisis de esos datos resulta fundamental para el avance de la ciencia.
- Finalmente, la irrupción de internet ha traído consigo un aumento de las transacciones a través de la web (comercio electrónico, e-mail, etc.) que también son almacenadas en bases de datos. En todos esos ámbitos, la cantidad de datos crece a un ritmo vertiginoso.

El cambio de concepción de los datos, unido a la gran cantidad de estos que se generan y almacenan continuamente, trajo consigo la aparición de la minería de datos como disciplina. La minería de datos resulta necesaria para extraer conocimiento cuando se dispone de gran cantidad de datos para analizar, ya que, es en esa circunstancia, cuando las técnicas manuales de análisis de datos resultan demasiado complejas de aplicar, ya que dicho análisis supera la capacidad humana. Para analizar grandes cantidades de datos se necesitan herramientas específicas de minería de datos, que son implementadas y distribuidas a los usuarios. Como se verá a lo largo de la asignatura, dichas herramientas ejecutan algoritmos muy complejos desde el punto de vista computacional. Afortunadamente, los últimos años también han sido testigos de un gran avance en las prestaciones y capacidad de cómputo de los computadores.

EL PROCESO DE DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS

Según la definición comúnmente aceptada, el proceso de KDD persigue la extracción automatizada de conocimiento no trivial, implícito, previamente desconocido y potencialmente útil a partir de grandes volúmenes de datos.

Tras la preparación, los datos pasan a la fase de minería de datos, en la que se aplican una serie de técnicas para obtener modelos.

TIPOS DE DATOS

En el campo de la informática, los datos se suelen almacenar en repositorios denominados bases de datos. Habitualmente, el tipo de bases de datos utilizadas para almacenar información son las bases de datos relacionales. Una base de datos relacional es un conjunto de relaciones, que se representan por medio de tablas. Cada una de esas tablas está compuesta por un conjunto de instancias (filas), de forma que cada instancia representa un objeto. De cada instancia se almacenan datos relativos a los diferentes atributos (columnas) de ella. Las instancias de una tabla están identificadas por una clave primaria, que puede ser un atributo o un conjunto de estos.

RELACIÓN CON OTRAS ÁREAS

A diferencia de otras disciplinas de la informática, la minería de datos no surge como un área completamente nueva, sino más bien como la mezcla de conceptos procedentes de otras muchas y diferentes disciplinas.

- Estadística. Muchas de las técnicas que se aplican en la minería de datos son o tienen su raíz en la estadística. De alguna manera, se podría decir que la estadística es la “madre” de la minería de datos, dado que muchos de los conceptos y técnicas de la estadística se aplican en minería de datos.
- Bases de datos. El proceso de KDD parte de datos que, habitualmente, se encuentran almacenados en bases de datos. Como se ha comentado anteriormente, dichos datos son preparados para su posterior análisis.
- Visualización. El objetivo final de la minería de datos es obtener conocimiento que sea útil. Para lograrlo, es un requisito fundamental que ese conocimiento pueda ser visualizado por los expertos de cada dominio. De ahí la importancia de las técnicas de visualización (diagramas, gráficos, resúmenes, etc.) en el campo de la minería de datos.
- Aprendizaje automático. El aprendizaje automático se encuentra profundamente ligado con la minería de datos, ya que ambos, de alguna manera, persiguen la obtención de modelos por medio de mecanismos automáticos.
- Otras. Además de las anteriores, la minería de datos también está relacionada con otras áreas como, por ejemplo, los sistemas de apoyo a la decisión, la recuperación de información, el tratamiento y procesamiento de señales, etc.

REPERCUSIÓN, DESAFÍOS Y TENDENCIAS EN MINERÍA DE DATOS

La minería de datos es una disciplina que, como su nombre indica, trata con datos. A lo largo de este epígrafe se ilustrará como, esa afirmación tan simple, en realidad, trae consigo importantes repercusiones y desafíos. En muchos casos, los datos que alimentan

al proceso de KDD son personales y privados. En realidad, se trata de información muy sensible de la que es obligación, tanto ética como legal, hacer un uso responsable.

Esa privacidad es especialmente crítica en casos en los que se trate con datos particulares como, por ejemplo, datos médicos de un conjunto de pacientes. Una forma típica de preservar la privacidad es eliminar aquellos atributos que identifican a cada individuo y que, en muchos casos, no son útiles para el análisis. Un ejemplo de este tipo de atributo es el DNI. Una vez garantizada la privacidad, condición fundamental para realizar minería de datos, se construyen modelos con esos datos. Otro aspecto importante que se plantea al respecto es la calidad de esos modelos y quién es responsable de ellos. Supóngase, por ejemplo, que se crea un sistema de minería de datos para predecir catástrofes naturales. ¿Qué ocurre si ese sistema falla al predecir un sismo y, como consecuencia de la falta de alerta, pierden su vida decenas de personas? ¿Quién es responsable de esa situación? Hay quien piensa que el responsable es quien usa el modelo, mientras que otros indican que el responsable es el creador del modelo.

La minería de datos es una disciplina relativamente reciente si la comparamos con otras áreas de la informática. Eso hace que haya muchos problemas abiertos y desafíos planteados para la comunidad. En este epígrafe se pretende dar una visión general de algunos de ellos, que son los que se detallan a continuación:

- El tratamiento de grandes volúmenes de datos. Cada vez más, comienzan a surgir dominios donde se generan grandes cantidades de instancias de datos. Cada una de esas instancias está compuesta por un número de atributos muy elevado, lo que dificulta su tratamiento.
- Logro de tiempos de respuesta bajos. Muy en relación con el desafío anterior, se encuentra el reto de conseguir tiempos de respuesta bajos en los sistemas de minería de datos. Estos sistemas, en ocasiones, se utilizan como sistemas de apoyo a la decisión en ámbitos donde el tiempo de respuesta es un factor esencial.
- Funcionamiento en línea y en tiempo real. En un entorno dominado por la web, resulta imprescindible contar con herramientas de minería de datos que funcionen en línea y en tiempo real. Es el caso, por ejemplo, de un supuesto sistema de detección de fraude en las operaciones realizadas con tarjetas de crédito.
- Tratamiento de datos con una estructura compleja. Cada vez son más los dominios que producen datos con una estructura no convencional. Por ejemplo, en medicina se generan datos relativos a pacientes que pueden ir desde un simple formulario de datos, hasta una imagen de radiodiagnóstico, pasando por un electroencefalograma. La obtención de conocimiento de este tipo de datos no convencionales es uno de los principales desafíos de la minería de datos.

- Automatización de tareas. En el proceso de KDD, tareas como el preprocesado siguen teniendo un importante componente artesanal. Esto hace que dicha fase consuma un porcentaje muy significativo del tiempo total de un proyecto de minería de datos. Si se consigue automatizar, aunque sea en parte, dichas tareas, se pueden obtener resultados en un tiempo mucho menor, reduciendo así el gasto y haciendo la minería de datos más accesible desde el punto de vista económico.

La tendencia en minería de datos es, precisamente, la propuesta de soluciones encaminadas a dar respuesta a los retos descritos anteriormente. Además, se empieza a intuir un giro importante en la concepción de la minería de datos. Normalmente, las empresas e instituciones suelen acudir a equipos de ingenieros expertos en minería de datos para que estos les ayuden a resolver algún problema mediante el análisis de datos. Las compañías entregan sus datos a los ingenieros y estos los analizan, generando resultados que se entregan a las empresas. El coste de este tipo de proyectos suele ser elevado.

Esta visión de la minería de datos parece estar evolucionando, en gran parte gracias a iniciativas como la propuesta por Kaggle, una plataforma on-line que propone competiciones de minería de datos.

CONCEPTOS BÁSICOS

- La minería de datos es una disciplina de la informática que estudia el análisis de grandes cantidades de datos con el objetivo de obtener conocimiento útil e interesante a partir de ellos.
- Descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases), es un término utilizado para referirse al proceso completo de descubrimiento de conocimiento a partir de un conjunto de datos. Desde una perspectiva purista, la minería de datos es una etapa de dicho proceso que, por ser la más importante, ha dado nombre a la disciplina entera.
- Un dato es, formalmente, el valor que toma una variable, y, más generalmente, una representación simbólica de una determinada característica de una entidad. En informática, los datos se suelen almacenar en repositorios denominados bases de datos. Los datos se pueden clasificar en: cuantitativos, que representan magnitudes; y cualitativos, que representan una categoría.
- Los atributos cuantitativos se dividen en: discretos, que son aquellos que pueden tomar un número limitado de valores; y continuos, aquellos para los que se cumple que, para cualquier par de valores, siempre se puede encontrar un valor intermedio.
- Los atributos cualitativos se clasifican en: ordinales, aquellos para los que existe una relación de orden entre las categorías; y nominales, para los que no existe

ese orden, sino que se asignan de forma arbitraria números o símbolos para cada categoría.

- Los atributos cuantitativos se dividen en: discretos, que son aquellos que pueden tomar un número limitado de valores; y continuos, aquellos para los que se cumple que, para cualquier par de valores, siempre se puede encontrar un valor intermedio.
- Los atributos cualitativos se clasifican en: ordinales, aquellos para los que existe una relación de orden entre las categorías; y nominales, para los que no existe ese orden, sino que se asignan de forma arbitraria números o símbolos para cada categoría.
- Las series temporales son sucesiones de valores que representan la evolución de una determinada característica a lo largo de un periodo de tiempo.

EL PROCESO DE KDD

proceso de KDD (Knowledge Discovery in Databases) se utiliza para referirse al proceso de extracción automatizada de conocimiento a partir de grandes volúmenes de datos. El conocimiento extraído por el proceso de KDD ha de poseer las cuatro características que se presentan a continuación:

El conocimiento extraído por el proceso de KDD ha de poseer las cuatro características que se presentan a continuación:

- No trivial. De nada sirve extraer conocimiento conocido por todos o que carezca de importancia.
- Implícito. Se encuentra oculto en los datos.
- Previamente desconocido. Nada nuevo se aporta si el conocimiento extraído ya había sido descubierto anteriormente.
- Útil. El conocimiento extraído debe servir para algo, de lo contrario no tiene ningún sentido invertir esfuerzos en extraerlo.
- Recopilación de datos. En esta fase, los datos, procedentes de diferentes fuentes, se integran en un mismo y único repositorio de datos, denominado almacén de datos, más conocido como data warehouse.
- Selección, limpieza y transformación de datos. Sobre los datos recopilados en el almacén de datos no es posible realizar aún data mining, debido a que dichos datos pueden no estar limpios, pueden contener atributos irrelevantes, etc. Precisamente, en la segunda fase del proceso de KDD, se realiza una selección de los datos integrados en el data warehouse. Dichos datos, además, se limpian y transforman de cara a fases posteriores.
- Data mining. Una vez se cuenta con una vista minable, el siguiente paso consiste en aplicar técnicas concretas de minería de datos para obtener modelos. El resultado de la fase de minería de datos es, precisamente, esos modelos.

- Interpretación y evaluación de modelos. Los modelos obtenidos en la fase de data mining han de ser evaluados. Una vez comprobada la calidad de los mismos, estos son interpretados y, a partir de ellos, se obtiene el conocimiento. El resultado de esta última fase es, precisamente, dicho conocimiento.

Una característica importante del proceso de KDD es su naturaleza iterativa. Esto significa que es posible tener que aplicar varias veces el proceso de KDD hasta obtener el conocimiento deseado. Por ejemplo, puede ocurrir que, tras obtener conocimiento en la primera iteración, este no sea de utilidad y sea necesario revisar alguna(s) de las fases anteriores. Una vez revisada(s) dicha(s) fase(s), se completaría de nuevo el proceso de KDD hasta obtener, de nuevo, conocimiento. Este proceso puede repetirse varias veces hasta que se extraiga conocimiento útil.

RECOPIACIÓN DE DATOS

El primer paso del proceso de KDD es tomar los datos que se desea analizar e integrarlos en un único repositorio del que partirán las siguientes fases.

Al realizar data mining, es posible que la empresa/ institución/ etc. propietaria de los datos disponga de múltiples fuentes de datos separadas. También es posible que, para el análisis, se vayan a utilizar otras fuentes de datos públicas como, por ejemplo, datos del censo de población de una determinada región. Por supuesto, antes siquiera de plantearse cualquier operación con los datos de las diferentes fuentes, es imprescindible aglutinarlos en un mismo almacén de datos o data warehouse. Para integrar las fuentes de datos en un mismo almacén, es necesario orquestar un proceso que lea los datos de las diferentes fuentes, los limpie y los adecúe a la estructura que tiene el data warehouse para su almacenamiento. Este tipo de proceso es llevado a cabo mediante un sistema conocido como sistema ETL (Extraction-Transformation-Load).

Una vez almacenados los datos, es posible tener que volver a repetir el proceso ETL para integrar nuevas fuentes de datos. Además, con cierta periodicidad, se deberán guardar copias de seguridad del contenido almacenado. Los almacenes de datos están diseñados de forma muy diferente a como lo están las bases de datos transaccionales que se pueden encontrar en cualquier empresa.

SELECCIÓN, LIMPIEZA Y TRANSFORMACIÓN DE DATOS

El objetivo de esta fase del proceso de KDD es tomar los datos almacenados en un data warehouse, sobre los cuales no es posible aplicar directamente técnicas de data mining, y transformarlos en un conjunto de datos sobre el que dichas técnicas sí se puedan aplicar. A ese conjunto de datos, resultado de esta fase, se le conoce con el nombre de vista minable.

SELECCIÓN DE DATOS

Las técnicas de selección tienen como objetivo filtrar aquellos datos que no son relevantes para el análisis posterior. El filtrado de esos datos se puede realizar a varios niveles:

- Filtrado de atributos. Es posible que algunos de los atributos de los datos a analizar no sean de interés. Por ejemplo, si se tiene información acerca de los clientes de una compañía y uno de los atributos es el DNI, es más que probable que dicho atributo deba ser filtrado ya que, de cara a un posible análisis de datos, ese atributo no aporta nada.
- Filtrado de registros. En ocasiones, el objetivo puede ser eliminar algunos de los registros almacenados. Por ejemplo, es posible que la empresa anterior solo esté interesada, en este momento, en los clientes cuya edad es mayor que un cierto umbral.

Una de las razones de mayor peso para realizar filtrado de registros es que, en ocasiones, se dispone de un número demasiado elevado de estos. Con un subconjunto menor de registros, denominado muestra, se podría hacer un análisis igual de efectivo, pero mucho más eficiente desde el punto de vista computacional. En ese caso, se suelen aplicar técnicas de muestreo para seleccionar una muestra de los registros del conjunto completo de datos. Algunas de las técnicas de muestreo más importantes son las siguientes:

- Muestreo aleatorio simple. En este caso, todos los elementos del conjunto completo de datos tienen igual probabilidad de ser extraídos en la muestra.
- Muestreo aleatorio estratificado. El objetivo de este tipo de muestreo es que todos los grupos o estratos que conforman el conjunto de datos completo estén representados de forma equilibrada en la muestra.
- Muestreo de grupos. En este caso, solamente se seleccionan registros pertenecientes a un grupo determinado, por ejemplo, aquellos clientes de una compañía que consuman más de 200 euros mensuales en sus productos.

LIMPIEZA DE DATOS

La limpieza de datos es una tarea absolutamente necesaria en esta fase del proceso de KDD. Normalmente, los datos, que han sido almacenados en las bases de datos de una compañía durante la historia de esta, presentan diferentes problemas en cuanto a su calidad. Por ejemplo, es posible que, al dar de alta un cliente en la base de datos de una compañía telefónica, se le haya asignado como fecha de nacimiento el día 21 de junio de 1285. Es obvio que, en 1285, la telefonía ni siquiera existía.

Las tareas de limpieza de datos van, normalmente, encaminadas a resolver dos problemas bastante habituales:

- La ausencia de valores. Es muy habitual que, para muchos de los registros analizados, falte cierta información. A esos valores ausentes se les conoce con el nombre de valores faltantes o missing values. Por ejemplo, es posible que en los datos del data warehouse haya un atributo denominado Segundo_Apellido que, para un cliente anglosajón aparecerá vacío. Es importante que el ingeniero de minería de datos analice cautelosamente los valores faltantes ya que, en ocasiones, aportan información interesante. Por ejemplo, si una persona no tiene asociado un valor para el atributo Estado_Civil, puede deberse a que dicha persona no desea que se conozca dicha información.
En general, ante un valor faltante, se pueden tomar diferentes alternativas, siendo las siguientes algunas de las más habituales:
 - Pasar por alto el valor faltante y continuar con el análisis.
 - Filtrar toda la columna asociada a dicho atributo.
 - Filtrar el registro que contiene el valor faltante.
 - Asignar un valor al atributo en cuestión. En este caso, se puede intentar “adivinar” el valor que falta sustituyéndolo, por ejemplo, por el valor medio que para dicho atributo manifiestan el resto de registros.
- La existencia de valores erróneos. También suele ser común encontrar valores que, claramente, son erróneos. Por ejemplo, es posible encontrar un código postal inexistente en una dirección. Aunque existen diferentes técnicas para detectar este tipo de valores, la realidad es que, la mayoría de las veces, se realiza mediante procedimientos artesanales “ad-hoc”.

TRANSFORMACIÓN DE DATOS

En la fase posterior de data mining se ejecutarán una serie de técnicas o algoritmos sobre los datos. Esos algoritmos, normalmente, exigen que los datos se encuentren en un formato determinado. Por ejemplo, es posible que un algoritmo de data mining solo sea capaz de operar con datos cualitativos. Las técnicas de transformación de datos ofrecen soluciones a dicho problema, formateando los datos según se necesite. Existen múltiples técnicas de transformación de datos. Por su especial relevancia, se muestran a continuación algunas de las más aplicadas: • Numerización. Consiste en transformar un atributo de tipo cualitativo en un atributo equivalente de tipo cuantitativo. Por ejemplo, un atributo de tipo booleano se puede numerizar transformando cada aparición del valor Falso a 0, y del valor Cierto a 1.

Existen múltiples técnicas de transformación de datos. Por su especial relevancia, se muestran a continuación algunas de las más aplicadas:

- Numerización. Consiste en transformar un atributo de tipo cualitativo en un atributo equivalente de tipo cuantitativo. Por ejemplo, un atributo de tipo

booleano se puede numerizar transformando cada aparición del valor Falso a 0, y del valor Cierto a 1.

- Discretización. Consiste en transformar un atributo cuantitativo en un atributo cualitativo ordinal. Por ejemplo, la altura de una persona en centímetros se puede discretizar en los intervalos Alto (≥ 180 cm), Medio (entre 150 cm y 180 cm) y Bajo (≤ 150 cm).
- Creación de características. Consiste en la creación de un nuevo atributo en los datos, normalmente calculado como función de otros atributos ya existentes.
- Normalización. Consiste en la transformación del rango de valores que toma un determinado atributo.

DATA MINING

Una vez obtenida una vista minable de los datos, el siguiente paso del proceso de KDD consiste en aplicar técnicas de data mining para obtener modelos que representen a dichos datos.

- Tareas predictivas. Son aquellas tareas de data mining que se utilizan para predecir el valor desconocido de uno o varios atributos para uno o varios registros de la vista minable.
- Tareas descriptivas. Son aquellas tareas de data mining que generan modelos que, de alguna forma, describen los datos. El objetivo es, por tanto, no predecir ningún dato sino describir los existentes.

Algunas de las principales tareas predictivas de data mining son:

- Clasificación. La tarea de clasificación consiste en encontrar un modelo que, aplicado a un nuevo ejemplo sin clasificar, lo clasifique dentro de un conjunto predefinido de clases. Por ejemplo, los clientes de un banco se pueden clasificar en función de si devuelven sus préstamos o no. Para ello, basta con almacenar, para cada cliente, un atributo, por ejemplo, Devuelve_Préstamo, que puede tomar los valores {Sí, No}. Ante un nuevo cliente, es posible aplicar el modelo de clasificación generado para predecir si devolverá o no un préstamo en caso de serle concedido. Normalmente, el atributo a predecir, en este caso, Devuelve_Préstamo, es de tipo cualitativo. El atributo a predecir se denomina atributo de clase.
- Regresión. La tarea de regresión es similar a la de clasificación, con la diferencia de que, en este caso, el atributo a predecir no es un atributo cualitativo, sino que es un atributo cuantitativo.
- Clustering. La tarea de clustering pretende dividir una población heterogénea de objetos en grupos homogéneos, denominados clusters, de forma que los objetos de cada grupo sean muy similares entre sí. También se denomina segmentación o agrupamiento. Por ejemplo, es posible que una empresa desee conocer los tipos

(grupos) de clientes que posee, en función de su edad, sexo, número de hijos e ingresos de estos.

- Asociación. La tarea de asociación pretende encontrar reglas que muestran la relación que existe entre los distintos atributos de los datos analizados. Esas reglas se denominan reglas de asociación. Por ejemplo, es posible que los gerentes de un supermercado deseen obtener reglas que relacionen productos que los clientes suelen comprar juntos.
- Para abordar cada una de las tareas anteriores, existen numerosas técnicas o algoritmos. Por ejemplo, para realizar clasificación se puede emplear el algoritmo denominado C4.5; para realizar regresión, se pueden utilizar técnicas de regresión logística; para realizar clustering se puede utilizar el algoritmo K-medias; para detectar atípicos, uno de los enfoques típicos es el uso de distribuciones estadísticas.

INTERPRETACIÓN Y EVALUACIÓN

Una técnica algo más avanzada, a la vez que más utilizada, es la técnica de validación cruzada n-fold cross validation. En este caso, para validar un modelo, se elige aleatoriamente el n% de los datos como conjunto de prueba y con el (100-n)% restante, como conjunto de entrenamiento, se construye el modelo.

En general, para cada tarea de data mining se utilizan unas métricas específicas que miden la calidad de los modelos obtenidos con ellas:

- Clasificación. Para evaluar un modelo de clasificación se mide su precisión predictiva, es decir, el porcentaje de clasificaciones acertadas en el conjunto de prueba frente al total de las clasificaciones realizadas. Por ejemplo, supóngase que se tiene un conjunto de 500 datos con el que se desea construir un modelo de clasificación. Supóngase también que se utiliza la técnica 10-fold cross validation. En ese caso, se construye un modelo con los 450 datos del conjunto de entrenamiento (90 %), mientras que los 50 restantes conforman el conjunto de prueba (10 %). El objetivo de la evaluación es medir si esos 50 datos son correctamente clasificados. Este proceso se repite 10 veces. En total, por tanto, se realizan $10 \times 50 = 500$ clasificaciones.
- Regresión. En este caso, la medida típica que se suele emplear es el error cuadrático medio. Es decir, para cada dato del conjunto de prueba se mide la diferencia entre el valor real y el valor predicho por el modelo. Con esos valores de diferencia se calcula un valor medio que representa la calidad del modelo de regresión.
- Reglas de asociación. Para evaluar la calidad de una regla de asociación, se suele medir el porcentaje de las instancias que la regla predice correctamente.

- Clustering. La calidad de una segmentación de objetos en clusters se suele medir por medio de la cohesión de los clusters, es decir, mediante alguna métrica que, efectivamente, mida si los objetos de cada cluster son similares entre sí y diferentes a los objetos de otros clusters.

ÁRBOLES Y REGLAS DE DECISIÓN

En teoría de grafos, el término árbol se utiliza para referirse a un grafo para el que se cumple que, dos vértices cualesquiera, están conectados por exactamente un camino. Los árboles de decisión son estructuras en forma de árbol que se utilizan como modelos de predicción en muy diferentes áreas. Una de ellas es, sin duda, la minería de datos, donde los árboles de decisión se usan, principalmente, como herramientas de clasificación. Para ello, se utiliza el valor de los atributos conocidos del objeto para ir descendiendo por el árbol (cada nodo del árbol tiene una condición sobre dichos atributos conocidos, que determina la rama por la que descender) hasta llegar a un nodo hoja, que indica la clase dentro de la cual ha sido clasificado el objeto.

Los árboles de decisión son estructuras en forma de árbol que se utilizan como modelos de predicción en muy diferentes áreas. Una de ellas es, sin duda, la minería de datos, donde los árboles de decisión se usan, principalmente, como herramientas de clasificación. Para ello, se utiliza el valor de los atributos conocidos del objeto para ir descendiendo por el árbol (cada nodo del árbol tiene una condición sobre dichos atributos conocidos, que determina la rama por la que descender) hasta llegar a un nodo hoja, que indica la clase dentro de la cual ha sido clasificado el objeto.

Existen múltiples algoritmos de construcción de árboles de decisión. Algunos de los más representativos son Id3, CART, C4.5, C5.0 y CHAID. Cada uno de ellos tiene ciertas particularidades, si bien es cierto que todos siguen un procedimiento similar de carácter iterativo:

ALGORITMO C4.5

El algoritmo C4.5 genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente. El árbol se construye mediante la estrategia de profundidad-primerero (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.

Los tres tipos de pruebas posibles propuestas por el C4.5 son:

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

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.

CARACTERÍSTICAS DEL ALGORITMO C4.5

- Permite trabajar con valores continuos para los atributos, separando los posibles resultados en 2 ramas $A_i \leq N$ y $A_i > N$.
- Los árboles son menos frondosos, ya que cada hoja cubre una distribución de clases no una clase en particular.
- Utiliza el método "divide y vencerás" para generar el árbol de decisión inicial a partir de un conjunto de datos de entrenamiento.

MEJORAS DEL ALGORITMO C4.5

- Evitar Sobreajuste de los datos.
- Determinar qué tan profundo debe crecer el árbol de decisión.
- Reducir errores en la poda.
- Condicionar la Post-Poda.
- Manejar atributos continuos.
- Escoger un rango de medida apropiado.
- Manejo de datos de entrenamiento con valores faltantes.
- Manejar atributos con diferentes valores.
- Mejorar la eficiencia computacional.

FUNCIONAMIENTO

El algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que le haya generado 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 (1,0) sobre cada uno de los valores que toma el atributo en los datos. En cada nodo, el sistema debe decidir que prueba escoge para dividir los datos. Según Espino (2005) los tres tipos de pruebas posibles propuestas para el C4.5 son:

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

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 determinar el valor límite Z.

SINIESTROS DE TRÁNSITO

Un accidente vial o siniestro automovilístico es un suceso imprevisto y ajeno al factor humano que altera la marcha normal o prevista del desplazamiento en las vialidades. Especialmente es aquel suceso en el que se causan daños a una persona o cosa, de manera repentina ocasionada por un agente externo involuntario. El perjuicio ocasionado a una persona o bien material, en un determinado trayecto de movilización o transporte, debido (mayoritaria o generalmente) a factores externos e imprevistos que contribuyen la acción riesgosa, negligente o irresponsable de un conductor, de un como pueden ser fallos mecánicos repentinos, condiciones ambientales desfavorables (sismos o cambios climáticos bruscos y repentinos) y cruce de animales durante el tráfico o incluso la caída de un árbol por fuertes vientos en la calle o carretera.

Un hecho, siniestro o incidente vial es aquella colisión entre uno o más sectores de la vialidad (peatones, ciclistas, automóviles, autobuses, camiones, tractores) en el cual si hay víctimas (tanto con lesiones leves o graves) se redefine como agresión vial, si se da con daños materiales se le dice “daños de tráfico”. Estos no son aleatorios ni imprevisibles, y usualmente están acompañados por corresponsabilidades, como puede ser falta de señalización adecuada, carencia de iluminación en las calles o la mala construcción de una avenida, falta de planeación o la ejecución de proyectos mal planeados, carencia de responsabilidad al manejar (conducir en estado de ebriedad, utilizar el celular mientras se maneja o conducir a exceso de velocidad).

Si bien, en la mayoría de los siniestros no se generaliza la culpabilidad, aunque no hay intención de lastimar hay culpa. Por ejemplo, un conductor en estado de ebriedad atropella peatones, sin embargo, sabe es ilegal manejar en ese estado, así como el hecho de encontrarse en estado etílico reduce sus capacidades de maniobra, por lo que el hecho vial deja de ser impredecible o inevitable.

LEY ORGÁNICA DE TRANSPORTE TERRESTRE TRÁNSITO Y SEGURIDAD VIAL determina en sus artículos 196, 88 y 20 mencionan los siguientes: “Art. 196.- El Director Ejecutivo de la Comisión Nacional y los Directores de las Comisiones Provinciales, serán los encargados de elaborar y supervisar los planes, programas, proyectos y campañas de prevención, educación y seguridad vial, la realización de estudios, formulación de soluciones y ejecución de acciones para la reducción de la accidentabilidad, con base en los factores y causas de incidencia.”

Art. 88 .- En materia de tránsito y seguridad vial, la presente Ley tiene por objetivo, entre otros, los siguientes: a) La organización, planificación y regulación de la movilidad peatonal, circulación, seguridad vial, uso de vehículos a motor, de tracción humana, mecánica o animal, y la conducción de semovientes: b) La prevención, reducción

sistemática y sostenida de los accidentes de tránsito y sus consecuencias, mortalidad y morbilidad; así como aumentar los niveles de percepción del riesgo en los conductores y usuarios viales; c) El establecimiento de programas de capacitación y difusión para conductores, peatones, pasajeros y autoridades, en materia de seguridad vial, para la creación de una cultura y conciencia vial responsable y solidaria;

Art. Art. 20.- Serán competencias de los responsables de unidad, además de las que determine el Director Ejecutivo de la ANT, las siguientes:

1. Elaborar los estudios regionales y provinciales, bajo los parámetros técnicos emitidos por la Agencia Nacional de Tránsito, que sirvan de insumo para la expedición del Plan Nacional de Rutas y Frecuencias;
2. Mantener los registros respectivos con la información que se genere en sus jurisdicciones; para estos efectos, los GADs que hayan asumido las competencias señaladas en la Ley, ¿deberán proporcionar a las unidades la información correspondiente a sus jurisdicciones;
4. Supervisar, en coordinación con los GADs, el cumplimiento del plan o planes de transporte terrestre, ¿tránsito y seguridad vial elaborados y autorizados por el organismo rector;
5. Emitir licencias de conducir para conductores profesionales y no profesionales, maquinaria agrícola y equipo caminero; y, 6. Coordinar operativos de control con los agentes de tránsito que correspondan.

Art. Art. 5.- El Estado, a través de la Agencia Nacional de Regulación y Control del Transporte Terrestre, Tránsito y Seguridad Vial, controlará y exigirá la capacitación integral, permanente, la formación y tecnificación a conductoras y conductores profesionales y no profesionales y el estricto cumplimiento del aseguramiento social”.

3. Investigaciones previas realizadas

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 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, generado 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 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 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 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, especifica que hace dos años las incidencias de transito fueron mayores en las carreteras de Guayaquil con un 23.3% y en Pichincha hubo un 17.1%.

La Agencia Nacional de Tránsito a través de la Dirección de Estudios y Proyectos (DEP) mensualmente recopila, analiza y presenta productos de siniestralidad vial, convirtiéndose en una herramienta que contribuye a todas las Direcciones Provinciales de la institución y a los entes de control operativo a nivel nacional a identificar factores o datos específicos, influyentes y reincidentes que se presentan en los siniestros de tránsito de cada provincia.

Para el análisis de la investigación se tomaron los datos de la Agencia Nacional de Tránsito a nivel nacional ocurridos en el país entre los años 2016-2018, así como otras investigaciones relacionas con las variables de la presente investigación, en ese sentido se tiene el trabajo de Hernández J Ramírez (2016) titulado “Introducción a la Minería de Datos”, en el mismo que (Esta Unidad didáctica se realizará una descripción exhaustiva del proceso de descubrimiento de conocimiento en base de datos, cuyo objetivo es conocer el proceso de KDD y las diferentes etapas que lo componen).

El Trabajo de Nicholas J. Horton (2016) titulado “Using R y RStudio”, en el mismo que (Nos indica las herramientas del software R con la utilización de vectores, matrices, estadísticas predictivas, dando como objetivo aprender su utilización)

El trabajo Gómez García, A. R., Russo Puga, M., Bermúdez, S., Roberto, P., Ortega, C., Alexander, F., ... & González Jijón, L. A. (2016), titulado “Caracterización de la mortalidad por accidentes de tránsito en Ecuador”, en el mismo que (En su resumen nos menciona las estadísticas realizadas por 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.)

García, A. R. G., Bermúdez, P. R. S., Yáñez, I. D. R., & Claudio, O. M. T. (2018). titulado” Accidentes de tránsito relacionados con el trabajo: una prioridad en salud laboral y pública para Ecuador”, en el mismo que nos indica en resumen que el Ecuador registra una tendencia creciente den el número de muertes por accidentes de tránsito, siendo cerca de 18.600 el número de víctimas fatales acumulado en los últimos 12 años. Aunque se resalta varias acciones positivas tomadas por el gobierno, como son la legislación que restringe la velocidad a 50 km por hora en las zonas urbanas, los límites de alcoholemia que están entre 0.3 y 0.1 gramos por mililitro de sangre y el control al uso de celulares para los conductores.



UNIVERSIDAD TECNOLÓGICA ISRAEL

ESCUELA DE POSGRADOS

MAESTRIA EN TELEMÁTICA,

MENCIÓN: CALIDAD EN EL SERVICIO

(Aprobado por: RPC-SO-19-No.300-2016-CES)

TRABAJO DE TITULACIÓN EN OPCIÓN AL GRADO DE MAGISTER

Título:
MINERÍA DE DATOS EN EL ANÁLISIS DE CAUSAS DE ACCIDENTES DE TRÁNSITO EN EL ECUADOR.
Autor/a:
ALVARO GALILEI PUMARES ROMERO
Tutor/a:
EDISON JAVIER GUAÑA MOYA

Quito - Ecuador

2019

Minería de datos en el análisis de causas de accidentes de tránsito en el Ecuador.

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.

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 [1], 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, generado 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 [2], 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 [3], 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.

De acuerdo a la investigación realizada por [4], 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, [5].

Es importante describir que, [6] 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.

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, [6]. Esto ha llevado a que el país reconozca el costo económico y social que representa este fenómeno.

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 [7] 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 [8],

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 esto [8] [9] afirman 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.

Dentro de los accidentes con mayor frecuentes, [10] afirma 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.

Otro factor importante en los AT es el uso del celular, por ello [11], 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, [12], afirma 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, [12], asevere 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, [13], 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. 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 y el telégrafo.

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.

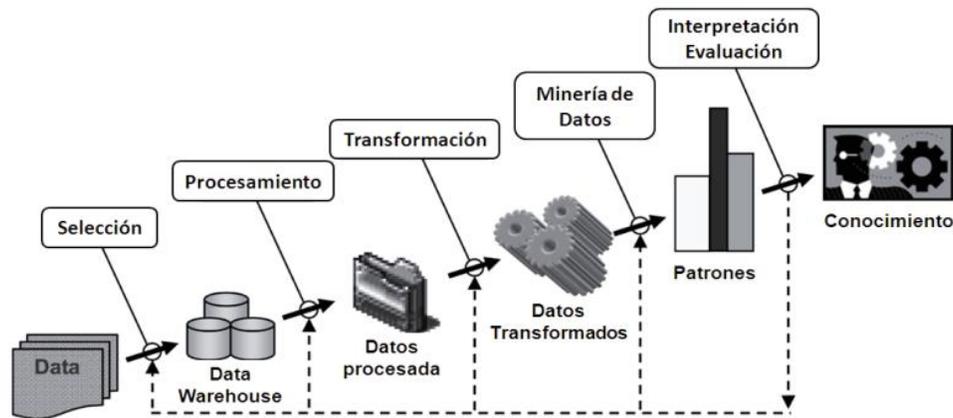


Figura 1 – Etapa del proceso KDD

Fuente: [14]

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

2.1. 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, dando un 100% de la recolección de datos y se recogieron un total de 9 variables.

2.2 Selección de variables

Una vez filtrada la base de datos y recogido los accidentes objetos de estudio se recogieron cada una de las variables que permiten realizar una caracterización de los mismos. Las variables recogidas, se presentan a continuación:

Tabla 1 – Variables relacionadas a los accidentes

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	Zona 1, 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/invasión 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)
<i>Tipología</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

2.3 Extracción de conocimiento

En esta fase, se utiliza la técnica de minería de datos basada en árboles de decisión. La base obtenida en la fase anterior 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 58%.

Con el árbol construido, procedemos a la toma de decisión en base a las reglas ya mencionadas, finalmente verificamos estas reglas para la verificación en conjunto.

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.

Tabla 2 –Resultados del modelo de clasificación

Regla N°	Soporte	Confianza
1	64,8%	86,2%
2	50,3%	77,2%
3	51,8%	67,1%

2.4 Interpretación

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

La regla con mayor soporte y confianza está 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.

3. Conclusiones

- Se identificaron las causas de los accidentes de tránsito utilizando los algoritmos de predicción.
- El algoritmo de minería de datos C 4.5, nos ayuda a organizar y clasificar la información sobre los accidentes de tránsito, en base a las causas de la misma.
- Esta metodología ofrece una técnica precisa haciendo referencia a los tradicionales modelos paramétricos, la misma que nos permite obtener patrones en la información.
- Por otro lado, estas técnicas 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.

4. Referencias bibliográficas.

- [1]. 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.).
- [2]. 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.
- [3]. Thomson, I., & Bull, A. (2001). La congestión del tránsito urbano: causas y consecuencias económicas y sociales. CEPAL.
- [4]. ORDOÑEZ ORDOÑEZ, F. A. 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 (p. 132).
- [5]. Wouters, P. I., & Bos, J. M. (2000). Traffic accident reduction by monitoring driver behaviour with in-car data recorders. *Accident Analysis & Prevention*, 32(5), 643-650.

- [6]. Híjar-Medina, M. C., Carrillo-Ordaz, C. E., Flores-Aldana, M. E., Anaya, R., & López-López, M. (1999). Factores de riesgo de lesión por accidentes de tráfico y el impacto de una intervención sobre la carretera. *Revista de saúde Pública*, 33, 505-512.
- [7]. García, A. R. G., Bermúdez, P. R. S., Yáñez, I. D. R., & Claudio, O. M. T. (2018). Accidentes de tránsito relacionados con el trabajo: una prioridad en salud laboral y pública para Ecuador. *NOTAS Y REFLEXIONES*, 112.
- [8]. Algora-Buenafé, A. F., Suasnavas-Bermúdez, P. R., Merino-Salazar, P., & Gómez-García, A. R. (2017). Epidemiological study of fatal road traffic accidents in Ecuador. *Australasian Medical Journal (Online)*, 10(3), 238.
- [9]. Gómez García, A. R., Russo Puga, M., Bermúdez, S., Roberto, P., Ortega, C., Alexander, F., ... & González Jijón, L. A. (2016). Caracterización de la mortalidad por accidentes de tránsito en Ecuador, 2015.
- [10]. Tomalá C., Ribadeneira E. (2016). Análisis Estadístico de las Causas de Accidentes de Tránsito en la Ciudad de Guayaquil.
- [11]. Diario EL COMERCIO. (2018). El celular es la principal causa de accidentes de tránsito en Quito, tomado de: <https://www.elcomercio.com/actualidad/celular-causa-accidentes-transito-quito.html>.
- [12]. El telégrafo. (2018). El 23% de accidentes se debe al uso de celular, tomado de: <https://www.eltelegrafo.com.ec/noticias/sociedad/6/accidentes-transito-uso-celular-Ecuador>
- [13]. Maya Mauricio. (2019). Siat alerta causas de accidentes de tránsito en invierno. Tomado de: <https://www.ministeriointerior.gob.ec/siat-alerta-causas-de-accidentes-de-transito-en-invierno/>
- [14]. Larose, Daniel. (2006). *Data Mining: Methods and Models*. New Jersey: John Wiley & Sons, Inc

CONCLUSIONES

1. Se logró identificar las causales de accidentes de tránsito con mayor incidencia en la provincia de Napo.
2. Se obtuvieron 14410 datos del año 2016 al 2018, de la Agencia Nacional de Tránsito para generar patrones de árboles de decisiones y como resultado reglas predictivas utilizando minería de datos.
3. El algoritmo de minería de datos C 4.5, nos ayudó al estudio y clasificación de la información sobre los accidentes de tránsito ocurridos en 2016 al 2018, con técnicas precisas, haciendo referencia a los tradicionales modelos paramétricos, la misma que nos permite obtener patrones en la información.
4. Se generó resultados y se determinó las posibles causas de incidentes de tránsito, para minimizar los siniestros de tránsito en el Ecuador.



Yo, Alvaro Galilei Pumares Romero portador(a) de la C.I. 1500616527 autor del trabajo de graduación titulado:

Minería de datos en el análisis de causas de accidentes de tránsito en el Ecuador, previo a la obtención del título de Magíster en: Telemática Mención Calidad en el Servicio.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el artículo 144 de la Ley Orgánica de Educación Superior, de difundir el respectivo trabajo para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de graduación con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Quito, 09 de septiembre de 2019

Firma

C.I. 1500616527

ANEXOS

Tabla 1. Cuadro comparativo de provincias con mayores siniestros de tránsito, años 2016, 2017

SUBZONA	2016	2017	DIFERENCIA
AZUAY	49	38	-11
BOLIVAR	189	167	-22
CAÑAR	219	174	-45
CARCHI	131	55	-76
CHIMBORAZO	742	788	46
COTOPAXI	431	503	72
DMQ	321	377	56
EL ORO	381	269	-112
ESMERALDAS	296	205	-91
GALAPAGOS	4	4	0
IMBABURA	196	103	-93
LOJA	82	77	-5
LOS RIOS	464	253	-211
MANABI	549	296	-253
MORONA SANTIAGO	156	176	20
NAPO	114	93	-21
ORELLANA	44	102	58
PASTAZA	68	54	-14
PICHINCHA	513	521	8
STO DOMINGO	373	272	-101
SUCUMBOS	73	76	3
TUNGURAHUA	314	307	-7
ZAMORA	101	84	-17
Total	5810	4994	-816

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito

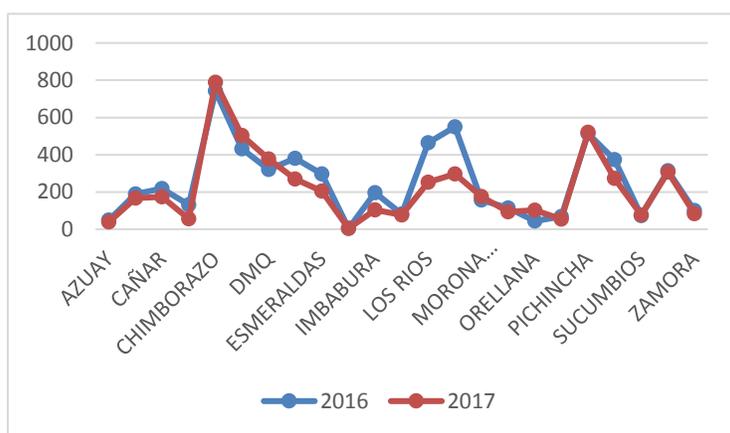


Gráfico 1

Tabla 2. Cuadro comparativo de provincias con mayores siniestros de tránsito, años 2017, 2018

SUBZONA	2017	2018	DIFERENCIA
AZUAY	38	21	-17
BOLIVAR	167	129	-38
CAÑAR	174	101	-73
CARCHI	55	37	-18
CHIMBORAZO	788	705	-83
COTOPAXI	503	113	-390
DMQ	377	320	-57
EL ORO	269	74	-195
ESMERALDAS	205	121	-84
GALAPAGOS	4	6	2
IMBABURA	103	91	-12
LOJA	77	65	-12
LOS RIOS	253	223	-30
MANABI	296	132	-164
MORONA SANTIAGO	176	158	-18
NAPO	93	83	-10
ORELLANA	102	119	17
PASTAZA	54	55	1
PICHINCHA	521	373	-148
STO DOMINGO	272	167	-105
SUCUMBIOS	76	72	-4
TUNGURAHUA	307	353	46
ZAMORA	84	88	4
Total	4994	3606	-1388

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito

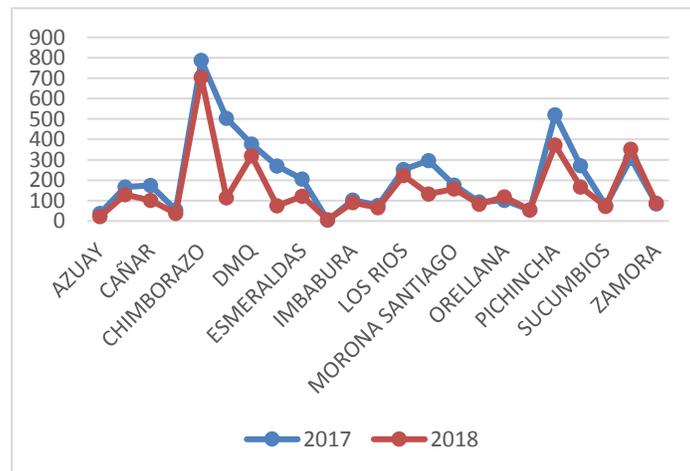


Gráfico 2

Tabla 3. Cuadro comparativo de provincias con mayores fallecidos de tránsito, años 2016, 2017

SUBZONA	2016	2017	DIFERENCIA
AZUAY	10	10	0
BOLIVAR	33	24	-9
CAÑAR	47	38	-9
CARCHI	36	14	-22
CHIMBORAZO	77	88	11
COTOPAXI	61	131	70
DMQ	60	61	1
EL ORO	40	34	-6
ESMERALDAS	49	67	18
GALAPAGOS	0	2	2
IMBABURA	58	48	-10
LOJA	24	14	-10
LOS RIOS	59	82	23
MANABI	104	77	-27
MORONA SANTIAGO	36	31	-5
NAPO	13	24	11
ORELLANA	19	24	5
PASTAZA	11	10	-1
PICHINCHA	94	98	4
STO DOMINGO	32	19	-13
SUCUMBIOS	41	35	-6
TUNGURAHUA	50	47	-3
ZAMORA	11	17	6
Total	965	995	30

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito

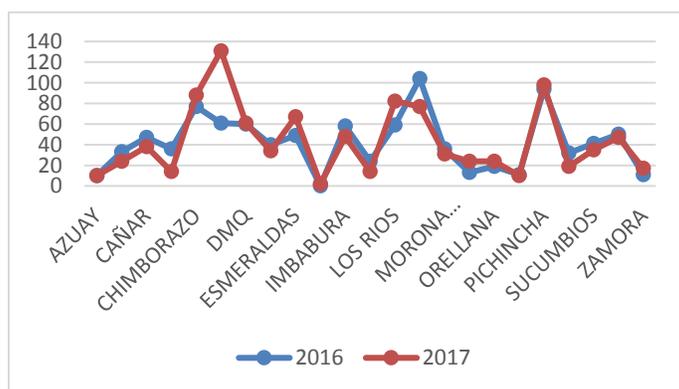


Gráfico 3

Tabla 4. Cuadro comparativo de provincias con mayores fallecidos de tránsito, años 2017, 2018

SUBZONA	2017	2018	DIFERENCIA
AZUAY	10	8	-2
BOLIVAR	24	36	12
CAÑAR	38	53	15
CARCHI	14	16	2
CHIMBORAZO	88	72	-16
COTOPAXI	131	102	-29
DMQ	61	75	14
EL ORO	34	22	-12
ESMERALDAS	67	64	-3
GALAPAGOS	2	3	1
IMBABURA	48	24	-24
LOJA	14	17	3
LOS RIOS	82	73	-9
MANABI	77	52	-25
MORONA SANTIAGO	31	27	-4
NAPO	24	19	-5
ORELLANA	24	35	11
PASTAZA	10	18	8
PICHINCHA	98	105	7
STO DOMINGO	19	29	10
SUCUMBIOS	35	27	-8
TUNGURAHUA	47	49	2
ZAMORA	17	11	-6
Total	995	937	-58

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito

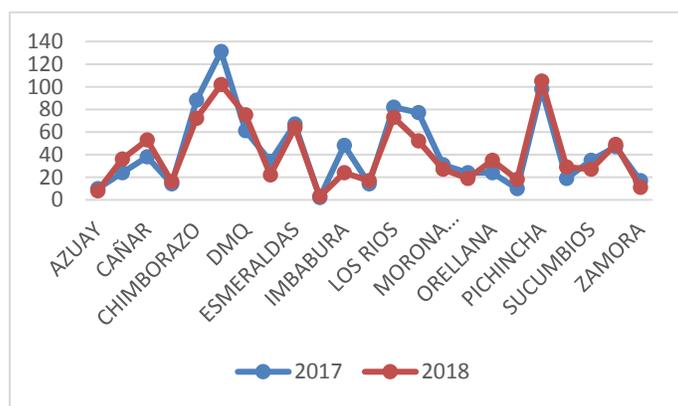


Gráfico 4

Tabla 5. Cuadro comparativo de provincias con mayores heridos de tránsito, años 2016, 2017

SUBZONA	2016	2017	DIFERENCIA
AZUAY	59	34	-25
BOLIVAR	218	243	25
CAÑAR	183	229	46
CARCHI	140	65	-75
CHIMBORAZO	290	293	3
COTOPAXI	258	278	20
DMQ	283	337	54
EL ORO	316	229	-87
ESMERALDAS	313	212	-101
GALAPAGOS	2	6	4
IMBABURA	161	102	-59
LOJA	116	92	-24
LOS RIOS	420	287	-133
MANABI	578	348	-230
MORONA SANTIAGO	213	189	-24
NAPO	102	119	17
ORELLANA	39	118	79
PASTAZA	91	81	-10
PICHINCHA	318	353	35
STO DOMINGO	335	280	-55
SUCUMBIOS	83	66	-17
TUNGURAHUA	244	220	-24
ZAMORA	84	85	1
Total	4846	4266	-580

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito

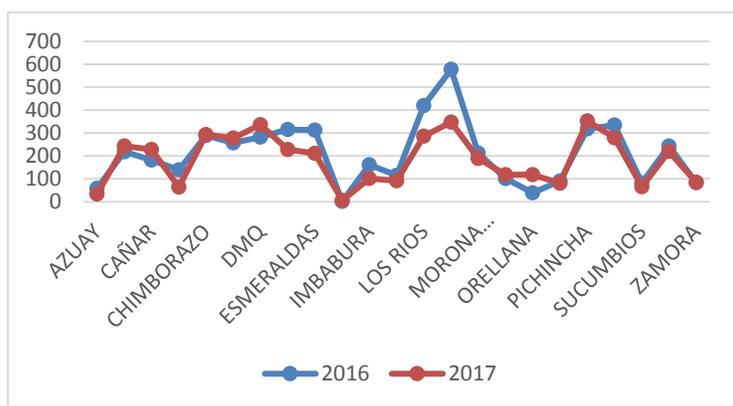


Gráfico 5

Tabla 6. Cuadro comparativo de provincias con mayores heridos de tránsito, años 2017, 2018

SUBZONA	2017	2018	DIFERENCIA
AZUAY	34	9	-25
BOLIVAR	243	176	-67
CAÑAR	229	119	-110
CARCHI	65	27	-38
CHIMBORAZO	293	333	40
COTOPAXI	278	44	-234
DMQ	337	199	-138
EL ORO	229	49	-180
ESMERALDAS	212	102	-110
GALAPAGOS	6	7	1
IMBABURA	102	95	-7
LOJA	92	72	-20
LOS RIOS	287	256	-31
MANABI	348	130	-218
MORONA SANTIAGO	189	181	-8
NAPO	119	75	-44
ORELLANA	118	126	8
PASTAZA	81	62	-19
PICHINCHA	353	233	-120
STO DOMINGO	280	142	-138
SUCUMBIOS	66	53	-13
TUNGURAHUA	220	177	-43
ZAMORA	85	99	14
Total	4266	2766	-1500

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito

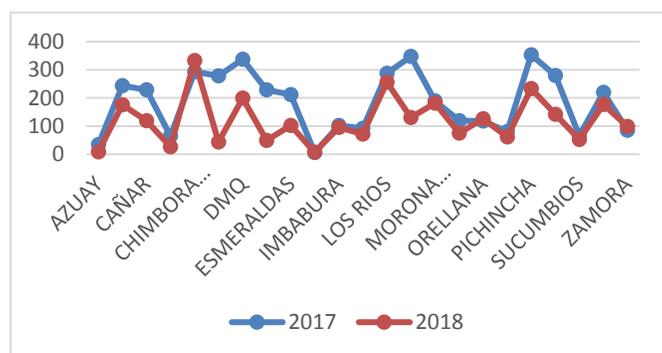


Gráfico 6

Tabla 7. Cuadro Causalidad 2016

ZONA	ACCIDENTES	FALLECIDOS	HERIDOS	SUBZONA	ACCIDENTES	FALLECIDOS	HERIDOS	DISTRITO	ACCIDENTES	FALLECIDOS
Total	2528	391	2230	Total	2528	391	2230	Total	2528	391
ZONA 1	349	90	355	AZUAY	29	5	46	ALAUSI	12	5
ZONA 2	314	59	222	BOLIVAR	79	9	78	AMBATO NORTE	18	0
ZONA 3	633	66	429	CAÑAR	115	22	94	AMBATO SUR	22	4
ZONA 4	334	57	372	CARCHI	60	21	83	ARAJUNO	3	0
ZONA 5	331	32	298	CHIMBORAZO	300	22	94	ATACAMES	18	1
ZONA 6	199	42	225	COTOPAXI	170	29	184	AZOGUES	53	6
ZONA 7	240	28	224	DMQ	128	17	105	BABAHYO	79	5
ZONA 9	128	17	105	EL ORO	163	18	146	BAÑOS	2	0

CAUSALIDAD	ACCIDENTES	FALLECIDOS	HERIDOS	TIPOLOGIA	ACCIDENTES	FALLECIDOS	HERIDOS	ZONA ACC	ACCIDENTES	FALLECIDOS
Total	2528	391	2230	Total	2528	391	2230	Total	2528	391
Adelantamiento Inadecuado / ...	31	6	38	Arrollamiento	32	27	15	RURAL	984	255
Calzada Resbaladiza o Granillo	4	3	15	Atropello (con ...	2	0	1	URBANA	1544	136
Casos Fortuitos	4	1	2	Atropello (con ...	328	82	281	HORARIO	ACCIDENTES	FALLECIDOS
Conduce con falta de atenci...	1344	181	1155	Caida de Pasajero	26	7	24	Total	2528	391
Danos Mecanicos	35	10	65	Choque Frontal ...	283	63	317	DIA	605	101
Dormirse Manejando (Impericia)	22	3	47	Choque Lateral ...	754	53	614	NOCHE	1075	179
Encandilamiento	3	0	2	Choque por Alc...	208	14	136	TARDE	848	111
Estado de Embriaguez	239	15	189	Colision	57	2	22	DIASE...	ACCIDENTES	FALLECIDOS
Exceso de Peso y Volumen	2	0	1	Encunetamiento	4	0	0	Total	2528	391
Exceso de Velocidad	211	51	206	Estrellamiento	304	28	176	lunes	279	41
Factor Climatico	81	5	59	TIPO CARRE...	ACCIDENTES	FALLECIDOS	HERIDOS	martes	282	60
Fallas de Iluminacion en la Via	3	0	1	Total	2528	391	2230	miercoles	268	37
Falta de Atencion en la Con...	250	24	214	Adoquinado	126	14	88	jueves	261	35
Falta de Senalizacion	4	0	2	Asfalto	2182	312	1887	viernes	386	57
Imprudencia Peaton	106	35	78	Empedrada	48	14	72	sábado	515	72
Mal Estacionamiento	4	0	3	Hormigon	112	21	108	domingo	537	89
Mal Estado de la Via	19	1	25	Tierra	60	30	75			
No ceder el derecho de via ...	129	52	111							
Obstaculos en la Via	11	1	5							
Presencia de Animales en la ...	23	1	11							
Transita Bajo Influencia de A...	3	2	1							

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito

Tabla 8. Cuadro Causalidad 2017

ZONA	ACCIDENTES	FALLECIDOS	HERIDOS	SUBZONA	ACCIDENTES	FALLECIDOS	HERIDOS	DISTRITO	ACCIDENTES	FALLECIDOS	HERIDOS
Total	2092	368	1776	Total	2092	368	1776	Total	2092	368	1776
ZONA 1	221	69	204	AZUAY	22	6	17	24 DE MAYO	2	0	0
ZONA 2	253	45	194	BOLIVAR	74	8	114	ALAUSI	4	2	2
ZONA 3	672	90	383	CAÑAR	81	14	85	AMBATO NORTE	17	2	2
ZONA 4	265	39	258	CARCHI	29	4	29	AMBATO SUR	21	3	3
ZONA 5	158	36	226	CHIMBORAZO	286	22	96	ATACAMES	8	5	5
ZONA 6	162	32	175	COTOPAXI	240	46	126	AZOGUES	45	5	5
ZONA 7	225	27	200	DMQ	136	30	136	BABAHYOY	33	11	11
ZONA 9	136	30	136	EL DRO	147	13	114	BAÑOS	13	2	2

CAUSALIDAD	ACCIDENTES	FALLECIDOS	HERIDOS	TIPOLOGIA	ACCIDENTES	FALLECIDOS	HERIDOS	ZONA ACC	ACCIDENTES	FALLECIDOS
Total	2092	368	1776	Total	2091	368	1774	Total	2092	368
Adelantamiento Inadecuado / ...	24	3	31	Arrollamiento	28	19	11	RURAL	846	214
Calzada Resbaladiza o Granillo	8	0	15	Atropello (con ...	2	0	1	URBANA	1246	154
Casos Fortuitos	20	8	11	Atropello (con ...	282	95	215	HORARIO	ACCIDENTES	FALLECIDOS
Conduce con falta de atenci...	1141	159	905	Caida de Pasajero	25	2	23	Total	2092	368
Conducir en sentido contrari...	1	1	0	Choque Frontal ...	236	64	260	DIA	452	66
Danos Mecanicos	23	12	53	Choque Frontal ...	6	7	5	NOCHE	953	187
Dormirse Manejando (Impericia)	14	1	10	Choque Lateral ...	586	32	460	TARDE	687	115
Encandilamiento	1	0	1	Choque Lateral ...	13	1	9	DIASE...	ACCIDENTES	FALLECIDOS
Estado de Embriaguez	201	19	167	Choque por Alc...	188	20	134	Total	2092	368
Exceso de Peso y Volumen	1	0	0	Colision	49	1	23	lunes	242	44
Exceso de Velocidad	122	27	141	TIPO CARRE...	ACCIDENTES	FALLECIDOS	HERIDOS	martes	254	52
Factor Climatico	74	10	78	Total	2092	368	1776	miercoles	241	39
Falta de Atencion en la Con...	198	47	185	Adoquinado	63	7	39	jueves	257	36
Imprudencia Peaton	88	27	60	Asfalto	1842	302	1536	viernes	277	46
Mal Estacionamiento	1	0	0	Empedrada	38	16	40	sábado	372	63
Mal Estado de la Via	18	3	21	Hormigon	90	25	78	domingo	449	88
No ceder el derecho de via ...	5	1	5	Tierra	59	18	83			
No ceder el derecho de via ...	128	43	77							
No guardar la distancia lateral	1	0	0							
No respetar las senales ma...	1	0	0							
No respetar las senales regl...	6	3	6							

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito

Tabla 9. Cuadro Causalidad 2018

ZONA	ACCIDENTES	FALLECIDOS	HERIDOS	SUBZONA	ACCIDENTES	FALLECIDOS	HERIDOS	DISTRITO	ACCIDENTES	FALLECIDOS	HE
Total	1620	386	1211	Total	1620	386	1211	Total	1620	386	1
ZONA 1	152	54	122	AZUAY	3	1	2	24 DE MAYO	3	5	
ZONA 2	284	81	162	BOLIVAR	75	13	100	ALAUSSI	10	2	
ZONA 3	508	94	271	CAÑAR	57	22	67	AMBATO NORTE	24	6	
ZONA 4	148	43	139	CARCHI	19	6	11	AMBATO SUR	23	2	
ZONA 5	157	39	182	CHIMBORAZO	306	21	158	ATACAMES	7	2	
ZONA 6	121	34	160	COTOPAXI	47	41	26	AZOGUES	21	7	
ZONA 7	112	18	92	DMQ	138	23	83	BABAHOYO	24	3	
ZONA 9	138	23	83	EL ORO	34	9	18	BAÑOS	7	2	

CAUSALIDAD	ACCIDENTES	FALLECIDOS	HERIDOS	TIPOLOGIA	ACCIDENTES	FALLECIDOS	HERIDOS	ZONA ACC	ACCIDENTES	FALLECIDOS
Total	1620	386	1211	Total	1620	386	1211	Total	1620	386
Adelantamiento Inadecuado / ...	15	8	17	Arrollamiento	36	28	13	RURAL	710	233
Calzada Resbaladiza o Granillo	6	1	3	Atropello (con ...	3	0	3	URBANA	910	153
Casos Fortuitos	12	8	10	Atropello (con ...	234	96	160	HORARIO	1620	386
Conduce con falta de atenci...	1063	213	805	Caida de Pasajero	14	2	10	DIA	359	63
Danos Mecanicos	11	0	5	Choque Frontal ...	148	51	167	NOCHE	718	196
Dormirse Manejando (Impericia)	10	1	5	Choque Frontal ...	36	14	45	TARDE	543	127
Encandilamiento	3	2	2	Choque Lateral ...	357	46	239	DIASE...	1620	386
Estado de Embriaguez	125	22	87	Choque Lateral ...	80	8	61	Total	1620	386
Exceso de Velocidad	50	12	46	Choque por Alc...	144	17	73	lunes	214	52
Factor Climatico	39	9	34	Colision	38	0	18	martes	219	51
Fallas de Iluminacion en la Via	2	2	0	TIPO CARRE...	1620	386	1211	miercoles	168	37
Falta de Atencion en la Con...	92	30	63	Total	1620	386	1211	jueves	167	31
Falta de Senalizacion	4	0	2	Adoquinado	54	1	44	viernes	244	53
Imprudencia Peaton	78	30	51	Asfalto	1460	352	1057	sábado	315	88
Mal Estado de la Via	7	5	7	Empedrada	30	12	25	domingo	293	74
No ceder el derecho de via ...	6	2	4	Hormigon	46	12	54			
No ceder el derecho de via ...	93	39	63	Lastrado	2	0	2			
No respetar las senales regl...	1	1	0	Tierra	28	9	29			
Presencia de Animales en la ...	2	0	7							
Transita Bajo Influencia de A...	1	1	0							

FUENTE: Aplicación SGP_Tránsito de la Agencia Nacional de Tránsito