



Evaluación de la severidad de accidentes de tráfico de Autocares y Autobuses en Argentina y España mediante el estudio matemático con árboles de clasificación

Evaluation of the severity of traffic accidents of Coaches and Buses in Argentina and Spain by means of the mathematical study with classification trees

Avaliação da gravidade dos acidentes de trânsito de Autocarros e Autocarros na Argentina e Espanha por meio do estudo matemático com árvores de classificação

Víctor Miguel Toalombo-Vargas ^I
victormtv@gmail.com
<https://orcid.org/0000-0002-9479-6307>

Giovanny Pillajo-Quijia ^{II}
giovannypillajoq@gmail.com
<https://orcid.org/0000-0003-3569-288X>

Blanca del Valle Arenas-Ramírez ^{III}
blanca.arenas@upm.es
<https://orcid.org/0000-0003-0446-6417>

Correspondencia: victormtv@gmail.com

Ciencias Técnicas y Aplicadas
Artículos de investigación

***Recibido:** 16 de julio de 2021 ***Aceptado:** 30 de agosto de 2021 *** Publicado:** 30 de septiembre de 2021

- I. Escuela Superior Politécnica de Chimborazo, Sede Morona Santiago, Macas - Ecuador.
- II. Universidad Politécnica de Madrid INSIA- UPM, Madrid – España.
- III. Universidad Politécnica de Madrid-España.

Resumen

El transporte colectivo de personas en autobús y autocar es muy importante en la mayoría los países de Iberoamérica, y los accidentes de tráfico con implicación de estos vehículos involucran pérdidas elevadas si se consideran las pérdidas humanas, materiales y productividad de las empresas del transporte. En cualquier país del mundo en el que éstos eventos ocurren, tienen un alto impacto mediático y avivan la preocupación por la mejora de la seguridad y la protección de los ocupantes. En la región de América del Sur, la mejora de la seguridad vial necesita alcanzar un mayor nivel de desarrollo de los sistemas de información que permitan llevar a cabo una investigación detallada y en profundidad. Para suplir la escasez de información detallada de los accidentes con implicación de autobuses y autoca transporte colectivo de personas (AT-TCP), cinco equipos de investigadores de universidades: UPM de España, UPS de Ecuador y UTN, UNSa y UNTDF de Argentina, participaron en un proyecto cooperativo de investigación de accidentes en transporte co de pasajeros (COOPERAXVII-12) ocurridos en los tres países y desarrollaron metodología de análisis de datos tomados de publicaciones en medios de comunicación y datos oficiales disponibles, enfocado en el desarrollo de estudios de patrones y aplicación de modelos estadísticos.

En este trabajo se evaluó la severidad de una sub-muestra de 94 accidentes AT-TCP ocurridos en Argentina y España entre los años 2017 y 2018. Los datos fueron analizados con modelos de minería de datos: árboles de clasificación y bosques aleatorios. Los modelos permiten identificar los factores de influencia en la severidad del accidente y algunos patrones según el contexto y concurrencia de diferentes factores. De los resultados, se destaca que los accidentes de mayor severidad han ocurrido en vías interurbanas y son de tipo colisión entre vehículos, vuelco sobre la calzada y salidas de la vía con desbarranco.

Palabras claves: Severidad; accidentes; árboles de decisión.

Abstrac

The collective transport of people by bus and coach is very important in most Latin American countries, and traffic accidents involving these vehicles involve high losses if the human, material and productivity losses of transport companies are considered. In any country in the world in which these events occur, they have a high media impact and fuel concern for the improvement of safety and the protection of the occupants. In the South American region, the improvement of road safety needs to reach a higher level of development of information systems that allow a detailed and in-

depth investigation to be carried out. To make up for the scarcity of detailed information on accidents involving buses and buses, collective transport of people (AT-TCP), five teams of researchers from universities: UPM from Spain, UPS from Ecuador and UTN, UNSa and UNTDF from Argentina, participated in a cooperative project for the investigation of accidents in passenger transport (COOPERAXVII-12) that occurred in the three countries and developed a methodology for analyzing data taken from publications in the media and available official data, focused on the development of pattern studies. and application of statistical models.

In this work, the severity of a sub-sample of 94 AT-TCP accidents that occurred in Argentina and Spain between 2017 and 2018 was evaluated. The data were analyzed with data mining models: classification trees and random forests. The models make it possible to identify the factors influencing the severity of the accident and some patterns according to the context and concurrence of different factors. From the results, it stands out that the most severe accidents have occurred on interurban roads and are of the collision type between vehicles, overturning on the road and exits from the road with a ravine.

Keywords: Severity; accidents; decision trees.

Resumo

O transporte coletivo de pessoas por ônibus é muito importante na maioria dos países latino-americanos, e os acidentes de trânsito envolvendo esses veículos envolvem grandes perdas se consideradas as perdas humanas, materiais e de produtividade das transportadoras. Em qualquer país do mundo onde esses eventos ocorram, eles têm grande repercussão midiática e alimentam a preocupação com a melhoria da segurança e proteção dos ocupantes. Na região da América do Sul, a melhoria da segurança viária precisa atingir um nível mais alto de desenvolvimento de sistemas de informação que permitam uma investigação detalhada e aprofundada. Para suprir a escassez de informações detalhadas sobre acidentes envolvendo ônibus e ônibus, transporte coletivo de pessoas (AT-TCP), participaram cinco equipes de pesquisadores de universidades: UPM da Espanha, UPS do Equador e UTN, UNSa e UNTDF da Argentina. em projeto cooperativo de investigação de acidentes no transporte de passageiros (COOPERAXVII-12) ocorrido nos três países e desenvolveu metodologia de análise de dados retirados de publicações na mídia e de dados oficiais

disponíveis, com foco no desenvolvimento de estudos de padrões. e aplicação de modelos estatísticos.

Neste trabalho, avaliou-se a gravidade de uma subamostra de 94 acidentes AT-TCP ocorridos na Argentina e na Espanha entre 2017 e 2018. Os dados foram analisados com modelos de mineração de dados: árvores de classificação e florestas aleatórias. Os modelos permitem identificar os fatores que influenciam a gravidade do acidente e alguns padrões de acordo com o contexto e a simultaneidade dos diferentes fatores. Dos resultados, destaca-se que os acidentes mais graves ocorreram em vias interurbanas e são do tipo colisão entre veículos, capotamento na via e saídas da via com ravina.

Palavras-chave: Severidade; acidentes; Árvores de decisão.

Introducción

Los accidentes de tráfico que implica autobuses y autocares están involucrados en todo lugar que exista medios de transporte terrestre y presente en toda clase social, provincia o país. Dichos sucesos superan los millones en todo el mundo y cada vez va en aumento, lo cual tiene consecuencias graves en el aspecto humano y económico dentro de la colectividad.

En referencia a pérdidas humanas en España, durante el año 2017, los diferentes miembros policiales notificaron 102.233 accidentes con víctimas. Estos accidentes ocasionaron 1.830 fallecidos (en el momento del accidente o hasta 30 días después del mismo); además, 9.546 personas fueron ingresadas en un centro hospitalario y 129.616 resultaron heridas no hospitalizadas. (DGT, 2018)

Sobre la actividad asistencial de las unidades de soporte vital avanzado a accidentes de tráfico en Asturias en el año 2010, indica que en España los accidentes de tráfico son la segunda causa de muerte, y la primera dentro de la población juvenil comprendida entre 15 a 29 años, siendo uno de los principales problemas de Salud Pública. (Villar Fernández, 2012)

Según la Asociación Civil Luchemos por la vida de Argentina en el año 2017, reporta accidentes de tráfico con 5.420 personas fallecidas (en el lapso de los 30 días posteriores al accidente). Con 6.626 heridos graves hospitalizados; 51.945 con heridas leves.

Argentina reporta:5.472 víctimas fatales; 113.805 heridos, de los cuales 8.174 son graves que requerían hospitalización, en un total de 81.592 siniestros ocurridos. Las principales causas de muerte por accidentes de tránsito fueron: la velocidad y el consumo de bebidas alcohólicas. Para

los peatones y los conductores, el problema principal es la distracción. (Anuario estadístico de siniestralidad vial)

Fundamentación

Dentro del ámbito de la investigación de accidentes de tráfico de TCP este trabajo tiene como principal utilidad identificar las variables que influyen en la severidad de los accidentes en la cual están implicados autobuses y autocares en Argentina y España.

El desarrollo de este estudio, que permitirán analizar y extraer de una manera más profunda los fenómenos causales, y así adoptar las mejores decisiones sobre el origen que incurre y reducir el nivel de accidentabilidad o reducir sus consecuencias.

Materiales y métodos

Metodología de recolección de la información

El estudio toma como base el proyecto COOPERA-TCP que estandarizó una herramienta de recolección de datos de accidentes de tráfico de autocares y autobuses (Arenas et al., 2018), y ahora se toma en cuenta vehículos adicionales como microbús, combi, autocar de un piso, microbús y autocar de dos pisos; La recolección de datos se realizó entre 2017 – 2018, relacionados con víctimas (heridos graves y fallecidos) en los países de Argentina y España; para la cual se utilizó la metodología de búsqueda, localización y análisis de medios de comunicación masivos. Esta información se contrastó con las bases de datos disponibles en el país y la información se recogió en el instrumento destinado para la recopilación de datos en formato digital. En la base disponible para este trabajo se han recogido datos de mas de 200 accidentes ocurridos en España y Argentina, lo cual da a notar la potencialidad de las herramientas propuestas. Siendo uno de los objetivos de este trabajo la evaluación de la metodología planteada, con los resultados obtenidos se pretende demostrar que puede ser aplicada a otros colectivos de interés (motos, ciclistas, peatones) en contextos en los que la información de accidentes es limitada y poco accesible para estudio de investigación científica. En la figura 1 se presentan los factores que se tienen en cuenta en el formulario del proyecto COOPERA-TCP.

Número de factor	Factores de influencia y otros	Tipo de factor de acuerdo a MICAA
1	Datos generales	Datos del momento y espacio. Geo-referenciación.
2	Tipo de vía	Infraestructura
3	Tipo de servicio del autobús o autocar / TCP	Vehículos
4	Lugar	Dato de tiempo y espacio
5	Tipo de accidente	
6	Condiciones en el momento del accidente	Condiciones ambientales y del tráfico
7	Vehículos	Vehículos
8	Datos de los ocupantes (conductor, pasajeros, otros) y víctimas	Actores del tráfico
9	Datos del conductor del autobús o autocar / TCP	Actores del tráfico
10	Evacuación de las personas lesionadas	Otros datos - aportes de datos para la investigación del siniestro
11	Presuntas infracciones del conductor	Actores del tráfico
12	Posible responsabilidad del accidente	Actores del tráfico
13	Factores que pudieron afectar la atención y presuntos errores del conductor	Actores del tráfico
14	Descripción del accidente	Otros datos - aportes de datos para la investigación del siniestro
15	Daño estructural en el autobús o autocar / TCP	Otros datos - aportes de datos para la investigación del siniestro
16	Esquema de daños en los vehículos	Otros datos - aportes de datos para la investigación del siniestro

Fig. 1 _ Resumen Factores de recolección de datos en un siniestro en formularios de accidentes para Transporte colectivo de personas _ Proyecto COOPERAXVII -12 (TCP). (Arenas et al., 2018)

El formulario utilizado para la ejecución del proyecto reúne los datos de accidentes, el cual contiene campos con las variables primordiales para abordar estudios científicos .

El análisis de modelos estadísticos permite profundizar en el conocimiento de la situación accidentalológica de los países participantes y realizar a partir de ello la identificación de factores de influencia para la obtención de un buen diagnóstico, riguroso y completo dentro de los límites que proporcionan los datos disponibles.

La muestra de análisis consta de 94 accidentes de tráfico con implicación de autobuses y autocares y constituye la base de datos que contiene más de 80 variables. Se ha realizado una selección de variables con varios criterios (como evitar repetición de la información, valores nulos o vacíos,

etc.) resultando del proceso un total de 13 variables explicativas que se han agrupado en 4 grupos: factor infraestructura, vehículo, entorno y factor humano. Con las variables seleccionadas, se desarrollará un modelo de análisis de la severidad del accidente.

Del conjunto de 4 factores se han seleccionado las variables de mayor importancia mediante Random Forest (RF) considerando su error de clasificación. Se seleccionan las 13 variables superiores en la clasificación con las que se elabora un modelo final.

Metodología estadística para el análisis

Modelos de bosques aleatorios (Random Forest (RF))

Es una versión sofisticada del procedimiento bagging, señalada por (Breiman, 2001), que replica subconjuntos de registros y elige al azar un subconjunto de las variables de entrada (Hastie, Tibshirani, & Friedman, 2008). Esta metodología también se usa para análisis de sensibilidad (Grömping, 2009). Además, se indica que es la técnica de conjuntos de árboles tienen enfoque clásico o frecuentista, con el algoritmo RF (Random Forest de R Core, 2016). Este algoritmo optimiza la precisión en la clasificación mediante la incorporación de aleatoriedad en la construcción de cada clasificador individual. La herramienta indicada se enfoca en la técnica de aprendizaje automático estándar denominada “árbol de decisiones. Para la selección y optimización se utilizan dos índices: criterio de Gini y de clasificación.

El mecanismo de construcción de RF establece un baremo que prioriza cada variable en la predicción final. Para ello se calcula el error de la muestra. Para cada variable de la muestra, se intercambia un par de elementos y se calcula el error de la muestra permutada. El resultado debería ser peor que para la muestra original. Este procedimiento se realiza para todos los valores de cada variable y se calcula el promedio. El proceso es realizado para todas las variables. Las variables de menor importancia deberían alterar menos la diferencia entre el error y el error permutado, que las variables importantes.

Mediante la función Rpart se analiza el modelo de clasificación con las 13 variables seleccionadas previamente en RF y se ejecuta la poda del árbol en función del % de error relativo y del índice de complejidad de división CP óptimo para generar el árbol de clasificación final.

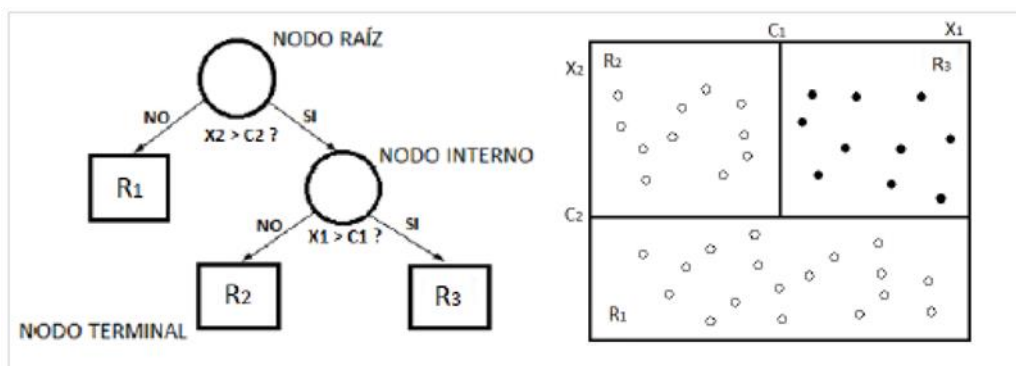
Los modelos de Árboles de clasificación es una técnica de aprendizaje controlado, no paramétrica, de segmentación binaria, es decir, las particiones se hacen de forma permanente hasta alcanzar un

criterio de parada, es decir que el árbol se construye dividiendo los datos repetidamente, en cada nueva división los datos son divididos otra vez en dos grupos.

Mutualmente excluyentes. Por lo que se aplican para modelar relaciones entrada – salida. El enfoque algorítmico más habitual para el CART creado por (Breiman, Friedman, Stone, & Olshen, 1984) produce de manera inicial un árbol muy grande, para consecutivamente podarlo; es decir, recorta ramas que no incrementen su capacidad predictiva. El grado de poda se define en función de un compromiso entre bondad de ajuste y confusión del árbol, o incluyendo también criterios de precisión de una validación cruzada.

La figura 2 muestra un nodo raíz o madre, que se divide después en nodos hijos, y finalmente el procedimiento de partición se aplica a cada nodo hijo por separado, estas divisiones se seleccionan, de modo que la impureza de los grupos hijos sea menor que la del grupo madre, este procedimiento tiene por objetivo, discriminar la respuesta en grupos homogéneos, de tal manera que se pueda mantener el árbol relativamente pequeño. (Pillajo et al., 2018).

Fig. 2 – Descripción de objetos de un árbol de clasificación (Zhang / Singer, 2010)



Resultados y discusión

Se analizan los datos concernientes a una muestra de 94 accidentes de tráfico de autocares y autobuses ocurridos en las carreteras de Argentina y España implicados entre los años 2013 – 2018.

Tabla 1 – Descripción accidentes de TCP en Argentina y España

Argentina				Total	España				Total
Autocar/Micro		Colectivo/Autobús			Autocar/Micro		Colectivo/Autobús		
N.º	%	N.º	%		N.º	%	N.º	%	
30(0.6)	31.91	20(0.4)	21.28	50(1)	37(0.84)	39.36	7(0.16)	7.45	44(1)

Se ha encontrado dentro del análisis en referencia a accidentes de tráfico que están implicados Autocar/Micro que es mayor en España, con un total de 39.36% en comparación con Argentina que posee el 31.91%. Por otra parte, en lo que concierne a accidentes de tráfico que involucra Colectivo/Autobús se ha hallado mayor incidencia en Argentina con el 21.28 % en relación de España con el 7.45%. Los porcentajes se han tomado en base a los 94 accidentes en los dos países. Con respecto a Argentina aparece más accidentes de tráfico en los que están involucrados Autocar/Micro con el 60% en comparación con Colectivo/Autobús que posee el 40%. Lo mismo sucede en España que el mayor número de accidentes están implicados Autocar/Micro con el 84%. Se puede identificar que en cualquier relación ya sea en función de los 94 accidentes o por país, en España existe mayor número de accidentes en referencia a Autocar/Micro.

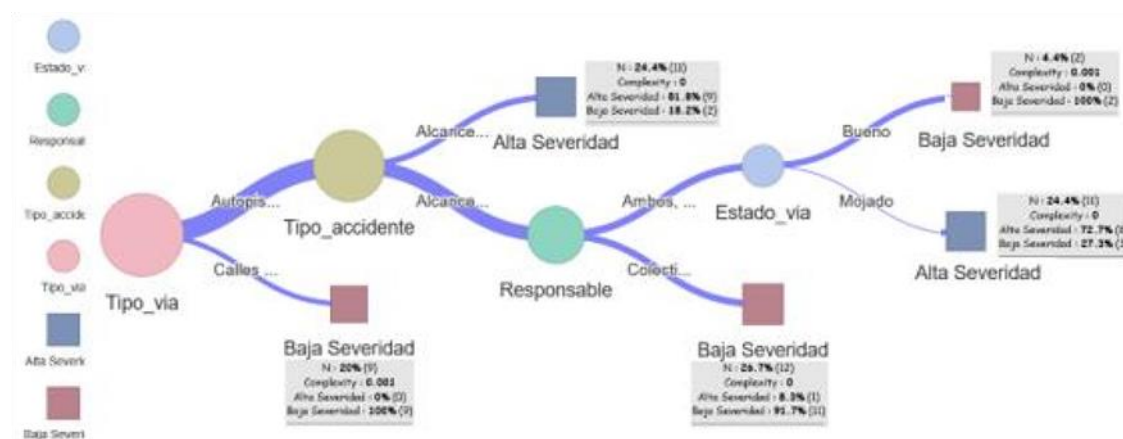
Se genera dos graficas de árbol de clasificación (Fig. 3 y 4) para Argentina y España respectivamente, Con respecto a Argentina indica que la variable más importante es el Tipo de arcén, en segundo lugar el tipo de accidente, tercer lugar el tipo de vía y posterior las restantes, mientras que para España la variable más importante es el Tipo de vía, seguido de Tipo de accidente y en tercer lugar Responsable; con lo cual se aprecia que dentro de los tres primeros lugares están presentes en los dos países las variables Tipo de accidente y Tipo de vía, lo cual indica que estos dos son la factores que más inciden dentro de la severidad de accidentes de tráfico en autocares y autobuses.

Figura 3. Árbol de severidad de Argentina



En síntesis, en Argentina cuando el tipo de arcén es natural, o a su vez, cuando está en mal estado, pavimentado u otros, según el tipo de accidente es por alcance, e implicado en una intersección resultaría una alta lesividad.

Figura 4. Árbol de severidad de España



Para finalizar, en España cuando el tipo de vía es autopista o carretera de doble sentido y tipo de accidente es por alcance al auto bus, salidad de la vía con desbarranco, salida de la vía u vuelco a severidad es alta, similar a que si el tipo de accidentes, es por alcance del autobús, choque o vuelco sobre la calzada, ambos vehículos implicados son responsables y el estado de la vía es mojado.

Referencias

Libro

1. HOYLE, B.S. y KNOWLES, R.D. (1992). Modern Geography. Belhaven Press, London.

Artículo de revista

2. SOLOMON, K.T. (1978). Sand in concrete mix. Australian Road Research 7(3), pp. 2730.

Un capítulo de un autor recogido en un libro editado por otros

3. EULER, G.W. (1992). Intelligent vehicle-highway systems. En: J.L. PLINE (ed.) Traffic Engineering Handbook, 4th edition. Prentice Hall, Englewood Cliffs, N.J. pp. 448-463.

Ponencia en Congreso

4. O'BRIEN, A. (1992). Safety audit for rural intersections. Workshop on Road Safety Audit, 14-16 April 1993, ponencia 6. Monash Transport Group, Clayton, Victoria.

Series

5. DEMPSEY, B. Climatic Effects of Airport Pavement Systems: State of the Art. Report DOT-RD-75-196. FHWA. U.S. Department of Transportation, 1976.

© 2021 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons

Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

(<https://creativecommons.org/licenses/by-nc-sa/4.0/>)