

Revista Internacional de Ciencias de la Ingeniería e Investigación Tecnología

(Una revista en línea revisada por pares)

Factor de impacto: 5,164



Editor jefe Dr.
JB Helonde

Editor ejecutivo Sr.
Somil Mayur Shah

*1ZS Associates Pvt. Ltd., Evanston, Illinois, EE. UU.

²
Universidad de Carolina del Norte, Charlotte, Carolina del Norte, EE. UU.

DOI: 10.5281/zenodo.3609528

RESUMEN:

Este artículo analiza el impacto de las infracciones de tránsito en los accidentes de tráfico mediante técnicas de análisis de big data basadas en información sobre infracciones de tránsito actualizada diariamente en el condado de Montgomery, EE. UU. Al conocer el motivo de los accidentes de tránsito, se busca identificar las infracciones. Además, se propone utilizar los datos para calcular las primas de seguro para las aseguradoras. La hipótesis busca predecir futuras infracciones que provoquen accidentes y predecir la mortalidad, es decir, lesiones personales o daños materiales. Además, se utilizan los datos para calcular las primas de seguros de las compañías aseguradoras. En nuestro análisis se consideran diversas clasificaciones, modelos de agrupamiento y regresión, como árbol único, árboles aleatorios, agrupamiento de k-medias, regresión múltiple y Naive Bayes. El primer y el segundo modelo consideran los árboles aleatorios y los árboles únicos como los mejores algoritmos para nuestro caso de negocio debido a la importancia de una alta sensibilidad y... Una puntuación F1 alta. El Modelo III considera que la agrupación por K-medias y la clasificación de árbol único son los mejores algoritmos para generar agrupaciones con sus numerosos tipos de violación y número de lesiones.

1. INTRODUCCIÓN

Numerosos sectores de investigación e industriales han recurrido a diferentes técnicas de análisis y minería de datos durante la última década. [1], [2], [3], [4], [5]. Ampliar la cobertura de estas técnicas supone un reto para los científicos actuales. Actualmente, es común utilizar la minería de datos para la regulación y la seguridad del tráfico [1], [2], [3], [4].

Los accidentes de tránsito representan una amenaza mundial de gran magnitud que continúa causando pérdidas, lesiones y víctimas en las carreteras, con un enorme impacto socioeconómico. Anualmente, 1,27 millones de víctimas y hasta cincuenta millones de heridos en todo el mundo son consecuencia de accidentes de tránsito [5]. Por lo tanto, este problema global requiere mayor atención para reducir la frecuencia y la gravedad de los accidentes.

Los datos históricos sobre infracciones de tránsito previas representan una ardua oportunidad para que los investigadores reconozcan los factores esenciales de dichas infracciones. Las principales dificultades para extraer información de estos datos son su gran tamaño y dimensiones [1], [2], [3]. Últimamente, para la extracción de información útil de conjuntos masivos de datos sobre accidentes de tránsito, se utilizan eficientemente diversas técnicas de minería de datos [2], [4], [6]. Los investigadores de accidentes e infracciones de tránsito utilizan ampliamente la clasificación de datos para la minería. El objetivo principal de estas estrategias era construir clasificadores para la predicción de nuevos accidentes y su gravedad.

En este trabajo, tendemos a realizar nuestro análisis del conjunto de datos 'Infracciones de tránsito' siguiendo el proceso estándar público intersectorial para la minería de datos Figura 1, que permite colocar el problema de minería de datos en la estrategia general de resolución de problemas de una unidad de investigación [9], [8]. Las tareas de análisis del proyecto se desarrollaron dentro de la descripción de la asignación del proyecto [...]. Las tareas principales son elegir un conjunto de datos accesible en Internet, estudiarlo y extraer conclusiones preliminares de él, tras la formulación de las hipótesis iniciales basadas en el análisis de datos y el desarrollo del caso de negocio. Las otras tareas necesarias de la investigación son explorar el conjunto de datos y producir visualizaciones utilizando Tableau, revisar las hipótesis basadas en visualizaciones y producir conjunto(s) de datos que satisfagan las hipótesis creadas. La tarea final es idear tres técnicas de modelado para las hipótesis desarrolladas, implementar tres algoritmos para cada modelo y proporcionar recomendaciones estratégicas respaldadas por Con base en el marco CRISP-DM y las tareas de análisis del proyecto, la investigación y la comprensión de los datos, la visualización es

Posteriormente, se completan las fases de preparación y modelado de datos. De acuerdo con la tarea de análisis asignada, la parte de modelado se ha completado aplicando diversas técnicas conocidas de minería de datos.

Además, se ha comparado la eficiencia de diversos algoritmos de minería de datos para las hipótesis inicialmente desarrolladas. Finalmente, las fases de validación e implementación se llevan a cabo según el marco CRISP-DM.

Paralelamente se han elaborado las recomendaciones estratégicas.

2. COMPRENSIÓN DE LA INVESTIGACIÓN

El objetivo del proyecto es investigar los datos de infracciones de tránsito del estado de Maryland para mejorar el conocimiento de la Policía del Capitolio de Maryland (MCP) y ayudarlos a disminuir la cantidad de infracciones respaldadas por su ubicación, lesiones, tasas de víctimas y variedad de daños a la propiedad.

Por otro lado, MCP puede compartir la información sobre infracciones de conductores de otros estados con su respectivo condado de regulación de tráfico. MCP aloja los datos en un dominio público en data.montgomerycountymd.gov [10] para crear conciencia y proporcionar a los entusiastas de los datos datos primarios para aplicar la ciencia y lograr un transporte más seguro en las carreteras. Creemos que es necesario informar a la Policía del Capitolio de Maryland sobre el número de infracciones que se producen en diversos lugares, respaldadas por la información, y la frecuencia de infracciones específicas a lo largo del tiempo.

Además, podemos predecir la probabilidad de infracciones en un accidente que haya resultado en lesiones o daños materiales. Podemos predecir las infracciones susceptibles a la temporada específica, al descubrir datos como en qué temporada se produjo la mayor cantidad de infracciones y si se produjeron lesiones personales o daños materiales. Finalmente, podemos analizar una parte de las infracciones relacionadas con vehículos comerciales. Esta información puede proporcionar sugerencias para la modificación de las normas y regulaciones de tránsito exclusivamente para vehículos comerciales. MCP puede utilizar el análisis para identificar áreas relevantes en la gestión del tráfico y ejercer control sobre ellas para... reducir el número de violaciones.

3. COMPRENSIÓN DE DATOS

El conjunto de datos seleccionado, "Infracciones de tránsito", se obtuvo de <https://data.montgomerycountymd.gov> [10] y se actualiza diariamente. Cada infracción electrónica ocurrida entre 2012 y 2018 se registra en el conjunto de datos del condado de Montgomery, Maryland. Estos datos proporcionan información sobre todas las infracciones de tránsito en el condado, su impacto y el lugar donde se cometieron. Además, proporciona datos asociados con el vehículo y su conductor. Para identificar las variables con mayor impacto en el accidente, creamos tres

Se formularon preguntas de investigación mediante el análisis de los atributos del conjunto de datos mostrados en la búsqueda de datos y se creó un modelo para verificar la pregunta. Antes del modelado, la búsqueda de datos evalúa el impacto en los accidentes de parámetros como el uso del cinturón de seguridad, la influencia del alcohol, los dispositivos portátiles y el tipo de vehículo. Se diseñaron modelos para verificar las preguntas de investigación utilizando algoritmos que utilizan técnicas de big data.

Derivamos las preguntas/hipótesis del análisis inicial subsiguiente a través del conjunto de datos.

- Se pueden predecir futuras infracciones utilizando atributos como la probabilidad de accidentes y las coordenadas geográficas de la ubicación, utilizando los datos de infracciones de tránsito recopilados durante seis años. Esta predicción sería útil para las autoridades de transporte, quienes pueden revisar y reformar las leyes actuales de control de tránsito para mejorar la seguridad vial.
- El conjunto de datos seleccionado se puede analizar para obtener información sobre el número estimado de futuras infracciones que pueden resultar en lesiones personales o fatales y daños materiales. Esta información es útil para las compañías de seguros, quienes pueden proporcionar sus cotizaciones con base en datos estadísticos históricos.

Las agencias de tránsito y de seguridad optimizan sus operaciones considerando las infracciones, los cargos y su frecuencia.

Las preguntas de investigación primarias someten a actualización los siguientes pasos de preprocesamiento y visualización de datos. En base a estas hipótesis se implementaría un modelado utilizando diferentes algoritmos.

Las preguntas de investigación primarias someten a actualización los siguientes pasos de preprocesamiento y visualización de datos. En base a estas hipótesis se implementaría un modelado utilizando diferentes algoritmos.

VISUALIZACIÓN DE LA INFORMACIÓN. Se desarrollan

numerosas visualizaciones para ampliar la comprensión del conjunto de datos inicial, comunicar perspectivas y hallazgos fascinantes. Tableau, una herramienta de análisis de datos muy conocida, realiza la visualización necesaria.

Infracciones vs. Horas del día: El número de infracciones observadas muestra que una parte considerable de ellas se produjeron durante la medianoche, en comparación con cualquier otro momento del día. Además, durante la madrugada, supera el promedio de infracciones de tránsito. La visualización indica que se deben implementar medidas para extremar la precaución durante estas horas específicas (Gráfico 1).

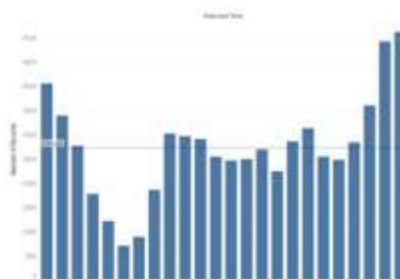
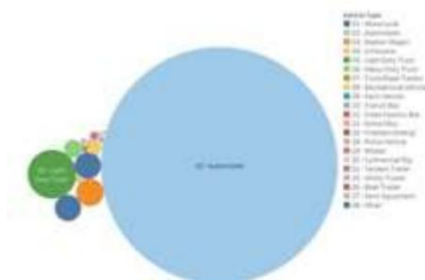


Gráfico 1. Infracciones de tránsito en diferentes horas del día entre 2012 y 2017[11]

Infracciones por tipo de vehículo: El gráfico de burbujas compactas muestra que cuanto más grande es el círculo, mayor es el número de infracciones causadas por ese tipo de vehículo. En este caso, el automóvil causa la mayor cantidad de infracciones, seguido por la camioneta ligera en segundo lugar. Gráfico 2



Gráfica 2. Infracciones de tránsito por tipo de vehículo entre 2012-2017 [11]

Infracciones de tránsito por temporada: El gráfico de árbol representa la tasa de infracciones por temporada en EE. UU., donde cuanto más oscuro es el color, mayor es la tasa. El promedio de infracciones no muestra cambios significativos entre las diferentes temporadas. El mayor número de infracciones se registró en otoño, seguido del invierno. Gráfico 3

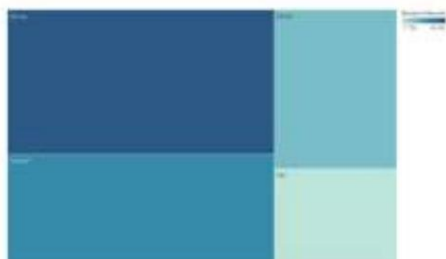


Gráfico 3. Infracciones de tránsito resumidas a nivel de temporada entre 2012 y 2017[11]

Infracciones debidas a la presencia de un niño: Este gráfico lineal muestra el número de infracciones que incluyen la presencia de un niño. <http://www.ijesrt.com> Revista internacional de ciencias de la ingeniería y tecnología de investigación

Según la descripción, estas infracciones se deben al manejo inadecuado de un menor al volante. Tras una ligera caída en 2014 y 2015, la cifra se recuperó y alcanzó un máximo histórico en 2017. Gráfico 4



Gráfico 4. Infracciones de tránsito con participación de un menor entre 2012 y 2017 [11]

Infracciones por uso del teléfono: El gráfico indica un aumento constante en el número de infracciones causadas por el uso de teléfonos móviles. La línea de tendencia muestra que, tras estar por debajo del promedio hasta 2013, el porcentaje total de infracciones mantuvo una tendencia ascendente, superior al promedio general. El uso de teléfonos móviles contribuyó al 29% de las infracciones, alcanzando su máximo en 2017. Gráfico 5

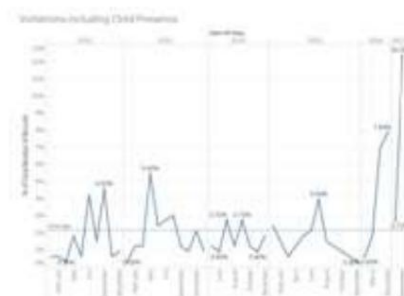


Gráfico 5. Infracciones de tránsito que implicaron el uso de un teléfono móvil entre 2012 y 2017 [11]

Lesiones personales versus cinturón de seguridad: La ayuda visual muestra que el cinturón de seguridad juega un papel importante en determinar si una infracción resultará en lesiones personales. Al observar el gráfico, de todas las infracciones que resultaron en lesiones, el 45 % de los registros no incluyen el uso del cinturón de seguridad. Esto enfatiza la importancia del cinturón de seguridad al conducir y permite a la MCP considerar su uso. Gráfico 6

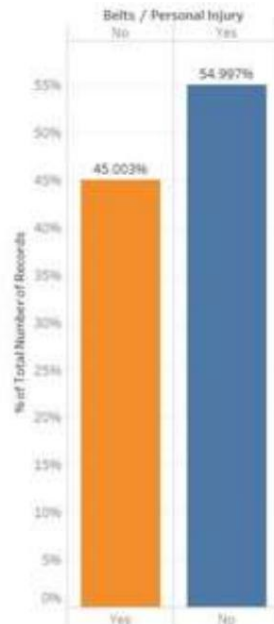


Gráfico 6. Relación entre el uso del cinturón de seguridad y las lesiones personales entre 2012 y 2017 [11]

Una vez analizadas las ayudas visuales se pueden plantear las siguientes hipótesis:

- El número de infracciones relacionadas con el uso del teléfono ha alcanzado un nivel extremo, y la tendencia indica que ha aumentado continuamente a lo largo de los años. ¿Qué medidas puede tomar MCP para reducir aún más esta tendencia? Entre las 22:00 y la 01:00 y las 08:00 y las 09:00, el número de infracciones es muy superior al promedio.
- ¿Cuáles son las principales razones de ello y qué medidas puede adoptar el MCP para reducir esta cifra?

Un análisis minucioso de la información y de los gráficos anteriores en conjunto con las preguntas de investigación primaria produce las tres hipótesis siguientes:

- 1) Predecir la probabilidad de que una infracción resulte en un accidente, dado un conjunto de variables predictoras.
- 2) Para valores dados de variables predictoras, predecir si la infracción que provocó un accidente es fatal.
- 3) Para una combinación dada de factores, predecir la geolocalización donde probablemente ocurrirá una violación particular, identificando además las ubicaciones con el mayor número de violaciones específicas.

Se debe completar la preparación de datos para modelar estas hipótesis. Esto permite introducir las variables objetivo, utilizar variables predictoras y tratar el sesgo dentro del conjunto de datos.

5. PREPROCESAMIENTO DE DATOS

A. En datos consistentes

Como parte del preprocesamiento de datos, observamos que pocas variables tienen el mismo valor en todos los registros, lo que las hace inutilizables desde el punto de vista del análisis de datos. Estas variables son "Agencia", que representa la agencia que emitió la citación, advertencia o aviso, y "Accidente", que indica la ocurrencia del accidente en el momento de la parada. En el caso de "Agencia", dado que MCP recopila todos los datos, solo tiene un valor, mientras que "Accidente", que debería tener valores booleanos "sí/no", solo tiene un valor para todos los registros: "no". Al considerar estas variables como univalentes, se eliminan del conjunto de datos.

B. Datos faltantes

Desde la perspectiva de la manipulación de datos, identificar y gestionar valores nulos es un paso vital. Esto se puede solucionar ignorando la tupla (eliminando la tupla) o añadiendo el valor de forma manual o automática. Estos valores pueden ser la media de un atributo, cualquier constante o valor más probable calculado mediante fórmulas bayesianas.

XLMiner identifica registros con valores nulos y los procesa. Los valores nulos representaban menos del 1% de los datos primarios, por lo que XLMiner los descartó.

C. Normalización/Estandarización Como no

había necesidad de una convergencia estable del peso de los atributos para apoyar nuestro análisis y La hipótesis y la mayoría de las variables en el conjunto de datos son categóricas, el proceso consume el conjunto de datos tal como está.

D. Nuevos atributos. Con

base en la comprensión del conjunto de datos, surge la necesidad de inducir nuevas variables. Hemos introducido tres variables binarias: "Uso del teléfono" como variable predictora, mientras que "Contribución al accidente" y "Fatal" son variables objetivo. Estas variables tienen un valor de "1" para todos los casos verdaderos y "0" para falsos, según el valor del atributo "Descripción". Por ejemplo, "Uso del teléfono" se incluye en el conjunto de datos y es "1" cuando el atributo "Descripción" de un registro indica el uso de un dispositivo portátil; de lo contrario, es "0".

E. Valores atípicos

En el conjunto de datos inicial, se detectaron valores atípicos y se reemplazaron imputando o eliminando el registro. El atributo de año de fabricación del vehículo, es decir, "Año", tiene un valor inferior a 1953 y superior a 2018 en algunos registros. En estos casos, todos los registros con un valor de "Año" inferior a 1953 se reemplazaron por el valor constante de 2000, mientras que los registros superiores a 2018 se eliminaron, ya que representan datos de un año de investigación.

F. Filtrado de datos. Los

datos imparciales son un requisito previo para una predicción precisa. Se puede aplicar el filtrado para lograr este objetivo. Utilizamos R. script que filtra y transforma datos en un conjunto de datos imparcial.

El conjunto de datos principal se filtra para predecir la variable objetivo 'Contribución_al_Accidente' de la hipótesis, donde se seleccionan todos los registros con valor verdadero para ella junto con una cantidad igual de registros con valor falso. La hipótesis tres utilizó nuevamente el mismo subconjunto. Para la hipótesis dos, se repite el mismo proceso de selección de subconjuntos, con la variable objetivo «Fatal» .

G. Bins y variables ficticias Para cualquier

método de análisis de datos que incorpora clasificación, la agrupación de los datos en intervalos y la creación de variables ficticias son tareas esenciales [9]. Muchos algoritmos eligen variables categóricas agrupadas en intervalos en lugar de las numéricas continuas. [9]. Aquí, la agrupación remodela la variable "Año" en una variable categórica con diez intervalos iguales.

Para los atributos 'Cinturones', 'Materiales peligrosos', 'Vehículo comercial', 'Alcohol', 'Uso de teléfono' y el agrupado 'Año', se crea una variable ficticia para lograr variables categóricas con valores '1' y '0' según la descripción del Apéndice A.

6. MODELADO

La sección de modelado se completó utilizando diferentes técnicas de modelado e investigando el desempeño de muchos algoritmos de clasificación para predecir la contribución a un accidente, la mortalidad por accidente y la geolocalización, donde sea que probablemente ocurra una infracción en función de predictores específicos elegidos de los registros de infracciones de tránsito recopilados por la Policía Estatal de Maryland durante el período de seis años de 2012 a 2018.

A. Modelo I

Para predecir la "Contribución al Accidente" en una situación particular, el modelo I utiliza tres algoritmos: árbol único, árboles aleatorios y bayesiano ingenuo. El modelo I utiliza la variable objetivo "Contribución al Accidente", generada durante el preprocesamiento de datos. Las variables categorizadas "Cinturones", "Alcohol" y "Uso del Teléfono", junto con las variables agrupadas "Tipo de Vehículo" y "Año", se utilizan como predictores. El árbol completo muestra que los mejores predictores para este modelo son las variables "Cinturones" y "Uso del Teléfono" . El rendimiento del modelo desarrollado se puede evaluar comparando los parámetros que se muestran en la Tabla I.

Tabla I Lecturas de los parámetros de rendimiento del modelo i obtenidas utilizando xl miner (porcentaje) [11]

| | Precision | F1-score | Specificity | Sensitivity |
|-------------|-----------|----------|-------------|-------------|
| SingleTree | 62.6 | 38.3 | 83.6 | 27.5 |
| RandomTrees | 50.3 | 66.9 | 2.3 | 99.6 |
| NaïveBayes | 62.1 | 37.3 | 83.8 | 26.6 |

La tabla muestra que el algoritmo de árbol único ofrece la mayor precisión, mientras que los algoritmos de árboles aleatorios presentan la mayor sensibilidad y puntuación F1.

Es esencial recordar el significado de los parámetros anteriores (precisión, sensibilidad, especificidad y F1-puntuación) siguiendo métricas de evaluación estándar [4], [6], [7], [5], [12], para evaluar el rendimiento. Pensamos en diferentes evaluadores de rendimiento para evaluar el modelo y su resultado en función del requisito de predicción. Según la Tabla I, el algoritmo de árboles aleatorios tiene la mayor sensibilidad, lo que muestra la proporción de contribución a los casos de accidentes que se reconocen adecuadamente. Cuando se modela para la predicción de 'Contribución_al_Accidente', la sensibilidad se convierte en un parámetro vital. Informar sobre la proporción de casos de 'Contribución_al_Accidente' correctamente identificados, el algoritmo de árbol aleatorio con una puntuación F1 máxima, proporciona la conjunción más efectiva de precisión y sensibilidad. Al mismo tiempo, el algoritmo de árbol único tiene la mayor precisión, lo que se vuelve más útil para el rendimiento de la predicción de casos generales.

B. Modelo II

El modelo II utiliza el mismo conjunto de algoritmos que el modelo I para predecir si el accidente es fatal para un escenario proporcionado. Aquí, la variable objetivo es "Fatal".

Inicialmente, junto con las variables predictoras seleccionadas en el modelo I, se selecciona el "Tipo de vehículo". El modelo II utiliza evaluadores de rendimiento idénticos a los del modelo I (Tabla II), tras un análisis similar. El algoritmo de árbol único detecta los mejores parámetros de sensibilidad y puntuación F1 para el modelo II, mientras que el algoritmo de árboles aleatorios presenta la mejor precisión para el modelo II.

Tabla II Lecturas de los parámetros de rendimiento del modelo II obtenidas utilizando xl miner (porcentaje) [11]

| | Puntuación F1 | Precisión | Sensibilidad | Especificidad |
|-------------------------|---------------|-----------|--------------|---------------|
| Árbol único 57.9 | 66.8 | 79.0 | 40.4 | — |
| Árboles aleatorios 63.9 | 47. | 38.3 | 77.6 | — |
| Bayes ingenuo 57.9 | 8 56.1 | 54.3 | 59.0 | — |

El árbol completo muestra que los principales predictores de este modelo son las variables "Año" y "Alcohol" agrupadas en agrupaciones. Estos resultados se pueden obtener debido a la gran cantidad de registros del conjunto de datos, pertenecientes a los intervalos de años noveno y décimo agrupados. Según nuestras tareas de predicción iniciales, la clasificación obtenida no es muy útil, ya que no proporciona una visión clara ni un conocimiento preciso de las situaciones específicas que pueden provocar accidentes mortales. Asimismo, el MCP no puede utilizar los años de los vehículos para tomar medidas específicas que reduzcan el número de accidentes.

Se omite la variable "Año" para aumentar la capacidad de predicción del modelo II. Debido a datos sesgados, "Vehículo Comercial" parece ser el principal predictor de "Fatal" en el árbol mejor podado. Dado que "Fatal" no depende del "Tipo de Vehículo" según los objetivos de predicción, se elimina. Tras repetir el proceso de modelado, las variables "Alcohol", "Uso de Teléfono" y "Materiales Peligrosos" parecen ser los principales predictores en el árbol completo. El MCP puede aumentar su vigilancia y controlar infracciones específicas utilizando los resultados del modelo II. El modelo II puede evaluarse utilizando predictores de rendimiento idénticos a los del modelo I (Tabla III).

Tabla III Lecturas de los parámetros de rendimiento (actualización) para el modelo II obtenidas utilizando xl miner (porcentaje) [11]

| | Puntuación F1 | Precisión | Sensibilidad | Especificidad |
|-------------------------|------------------|-----------|--------------|---------------|
| Árbol único 50.9 | 67,5 100,0 y 0,0 | 5,7 92,0 | — | — |
| Árboles aleatorios 51,9 | 67.0 | 98.0 | — | — |
| Bayes ingenuo 50.0 | 12.0 | 7.0 | — | — |

C. Modelo III

En el modelo III, el conjunto de datos se agrupa mediante la agrupación K-Medias y se elige "cluster_ID" como variable objetivo. Posteriormente, se asignan variables de predicción para determinar el "Fatal" en un escenario. El modelado se realiza mediante los mismos algoritmos que en el modelo I, junto con la regresión lineal múltiple. El modelo III utiliza la variable objetivo "cluster_ID" para identificar el clúster más buscado para un conjunto de predictores.

El modelo de predicción generado utilizando el algoritmo de regresión lineal múltiple tenía un valor R-cuadrado ligeramente ajustado. Los árboles aleatorios y el método bayesiano ingenuo ya no son imparciales en cuanto al resultado. Por otro lado, un algoritmo de árbol único, capaz de clasificar los clústeres formados, muestra resultados más determinados. Para el modelo III, se elige el algoritmo de árbol único para las predicciones de calidad, ya que presenta como resultado un árbol de un solo nodo con la mejor poda.

El árbol completo indica el desempeño de la puntuación de clasificación para el modelo III. En el grupo cinco, el desempeño indica que se produjeron algunas infracciones bajo la influencia del alcohol, junto con lesiones personales en el otro grupo. Nueve recibieron múltiples violaciones del cinturón de seguridad además de lesiones personales.

7. CONCLUSIÓN

De la investigación realizada anteriormente se concluye que 'Belts' y 'Phone_usage' son dos parámetros críticos. La negligencia hacia ellos resultó en la mayoría de las infracciones y, posteriormente, en accidentes. Además, el uso del teléfono aumentó la probabilidad de que un accidente fuera mortal y causara víctimas. Además, se debe notificar al MCP sobre las regiones significativas donde es probable que se produzcan infracciones. Los grupos cinco y nueve son áreas. donde las violaciones de alcohol condujeron a lesiones personales y las violaciones de múltiples cinturones con lesiones respectivamente.

El MCP, junto con el control de tráfico, puede trabajar para concienciar sobre estas variables. Además, para impulsar la seguridad vial, puede optar por reformar las leyes vigentes que las regulan. Se pueden dar recomendaciones similares a las compañías de seguros para que reflexionen sobre sus políticas para los conductores en las zonas analizadas.

REFERENCIAS [1]

- J. Abellan, G. Lopez, D. O ~na, y J., "Análisis de la gravedad de accidentes de tránsito utilizando reglas de decisión Árbol de decisión. Sistemas expertos con", Aplicaciones, vol. 40, págs. 6047–6054, 2013.
- [2] LY Chang y JT Chien, "Análisis de la gravedad de las lesiones de los conductores en accidentes con camiones involucrados utilizando un modelo de árbol de clasificación no paramétrico", Safety Science, vol. 51, n.º 1, págs. 17-22, 2015.
- [3] WH Chen y P. Jovanis, "Método para identificar los factores que contribuyen a la gravedad de las lesiones del conductor en accidentes de tráfico", Transportation Research Record, págs. 1–9, 2012.
- [4] AT Kashani, R. Rabieyan y MM Besharati, "Un enfoque de minería de datos para investigar los factores que influyen en la gravedad de los accidentes de los pasajeros de motocicletas", Journal of Safety Research, vol. 51, págs. 93–98, 2014.
- [5] D. O ~na, LJ, G. y J. Abellan, "Extracción de reglas de decisión de informes de accidentes policiales a través de árboles de decisión", Análisis y prevención de accidentes, vol. 50, págs. 1151–1160, 2013.
- [6] OH Kwon, W. Rhee y Y. Yoon, "Aplicación de algoritmos de clasificación para el análisis de dependencias de factores de riesgo de seguridad vial", Accident Analysis and Prevention, vol. 75, págs. 1–15, 2015.
- [7] Y. Xie, Y. Zhang y F. Liang, "Análisis de la gravedad de las lesiones por accidentes utilizando modelos probit ordenados bayesianos", Revista de Ingeniería de Transporte ASCE, vol. 135, no. 1, págs. 18–25, 2009.
- [8] SM S y RS T, Sistema de predicción de credibilidad crediticia basado en un algoritmo de árbol de decisión, 9 de 2015. [En línea]. Disponible en: <http://dx.doi.org/10.17577/IJERTV4IS090708> TC D, Ed. Larose:
- [9] "Descubriendo conocimiento en datos: una introducción a la minería de datos", D. Wiley, 2da edición: Wiley.
- [10] "Datos de Montgomery". [En línea]. <https://data.montgomerycountymd.gov/Public-Safety/Traffic-Violations/4mse-ku6q>
- [11] A. Raghuvanshi, D. Mehta, R. Bikmetov, J. Park, S. Pothina y S. Narsaraj, "Proyecto a plazo para Big data Análisis para la ventaja competitiva. UNCC", 2018.



[Raghuvanshi, et al., 9(1): enero de 2020]
Valor IC™: 3,00

ISSN: 2277-9655
Factor de Impacto: 5.164
CODEN: IJESS7

[12]MO Mujalli, O ~na y J., "Un método para simplificar el análisis de la gravedad de las lesiones por accidentes de tránsito en carreteras de dos carriles utilizando redes bayesianas", Journal of Safety Research, vol. 42, págs. 317–326, 2011.

