

Predicción del riesgo de lesiones mediante Big Data: el caso de Melbourne metropolitana

Ali Soltani^{1 2}, Betsabeh Tanoori³, Christopher J. Pettit⁴

¹ Estudios de lesiones, Universidad de Flinders, Bedford Park, SA, 5042, Australia

² UniSA Business, Universidad del Sur de Australia, SA, 5001, Australia

³ Departamento de Ingeniería Informática, Universidad de Shiraz, Shiraz, Irán

⁴ Centro de Investigación sobre el Futuro de las Ciudades, Universidad de Nueva Gales del Sur, Sídney, 2052, Australia

Correo electrónico para correspondencia (autor que presenta): ali.soltani@flinders.edu.au

Abstracto

Este trabajo de investigación investiga el uso de big data para analizar los accidentes de tráfico en el área metropolitana de Melbourne. Utilizando el conjunto de datos Crash y el modelo Random Forest, el estudio buscó determinar la relación entre diversos factores y la proporción de víctimas de accidentes que resultaron mutiladas o fallecidas. Los resultados demostraron que factores como los días de la semana, los meses y las condiciones meteorológicas pueden influir en la gravedad de los accidentes. El estudio proporciona a los responsables políticos y a las autoridades de transporte información valiosa para diseñar estrategias que mejoren la seguridad vial y reduzcan el riesgo de accidentes. Investigaciones adicionales pueden investigar otros posibles factores de gravedad de los accidentes.

Palabras clave: Big data; Accidentes de tránsito; Aprendizaje automático; Análisis de correlación; Victoria.

1. Introducción

El big data desempeña un papel fundamental en la identificación de tendencias de accidentes de tráfico, lo que fundamenta las políticas y medidas para mejorar la seguridad vial. Los investigadores pueden identificar patrones y tendencias en los accidentes de tráfico mediante el análisis de grandes volúmenes de datos procedentes de registros de accidentes, informes meteorológicos y datos sobre el volumen de tráfico (Pika et al., 2021). Esto puede implicar la identificación de cruces de alto riesgo, tramos de carretera y características que provocan accidentes, como el exceso de velocidad, la conducción distraída y el mal tiempo (Jiang et al., 2017). Los responsables políticos y las agencias de transporte pueden utilizar esta información para planificar medidas específicas que minimicen la frecuencia y la gravedad de los accidentes de tráfico, como la mejora de la infraestructura vial, una mayor aplicación de la ley o campañas de concienciación pública.

La tecnología y la disponibilidad de datos están ampliando el papel del big data en las colisiones y accidentes de tráfico. Los dispositivos conectados y el Internet de las Cosas (IdC) proporcionan diversos datos de transporte, como el volumen de tráfico, el clima y la velocidad de los vehículos (Kitchin, 2014). Las tecnologías y métodos de análisis de datos han agilizado y simplificado el procesamiento y análisis de grandes volúmenes de datos. Esto ayuda a académicos y responsables políticos a descubrir patrones y tendencias en los accidentes de tráfico. Las organizaciones de transporte y los responsables políticos pueden utilizar el big data para mejorar la seguridad vial y reducir la frecuencia y la gravedad de los accidentes. Por lo tanto, es probable que el papel del big data en la seguridad vial aumente en el futuro.

El uso de big data para la investigación en seguridad vial ofrece varias ventajas: al combinar y analizar grandes volúmenes de datos de diversas fuentes, el big data puede proporcionar una visión más completa y precisa de las tendencias en seguridad vial que los métodos tradicionales de recopilación y análisis de datos (Wang et al., 2018). El big data puede utilizarse para descubrir patrones y tendencias en los accidentes de tráfico, como las zonas de alto riesgo o los factores específicos que los causan. Al controlar el tráfico y las carreteras...

Actas de la ATRF 2023

Al analizar las condiciones en tiempo real, el big data puede mejorar los tiempos de respuesta ante emergencias (Papadimitriou et al., 2019). Los responsables políticos y las organizaciones de transporte pueden utilizar el big data para mejorar la seguridad vial identificando factores de riesgo específicos e implementando intervenciones específicas para abordar las causas principales de los accidentes. Mediante el análisis de big data, podemos determinar qué medidas de seguridad vial deben priorizarse para su financiación y dónde estos fondos tendrían el mayor impacto.

Los problemas de seguridad vial en Australia incluyen las largas distancias entre los principales centros de población, un porcentaje significativo de carreteras rurales y un elevado número de vehículos pesados (Naweed et al., 2014; Faulks, 2012). En comparación con países con regiones más densamente pobladas y distancias de viaje más cortas, estas características pueden provocar una mayor incidencia de accidentes y muertes.

A pesar de lo anterior, Australia ha logrado avances significativos en seguridad vial en las últimas décadas, pero se puede hacer más (Organización Mundial de la Salud, 2015). En 2020, en términos de tasa de mortalidad por cada 100.000 habitantes, Australia ocupó el puesto 20 entre 36 países, con una tasa de 4,26.

La tasa de mortalidad de Australia disminuyó un 25,4% entre 2011 y 2020. Durante el mismo período, la tasa mediana de la OCDE disminuyó un 34,6% (BITRE, 2022).

Los accidentes de tránsito son una fuente importante de lesiones y muerte en Australia (Scott-Parker y Oviedo-Trespalacios, 2017), por lo que los estudios sobre accidentes de tránsito son cruciales. El Instituto Australiano de Salud y Bienestar informa que los accidentes de tránsito son la principal causa de muerte en jóvenes de 15 a 24 años y la segunda en jóvenes de 25 a 34 años. Además del costo humano, los accidentes de tránsito tienen un impacto económico significativo (Blincoe et al., 2002; Clements y Kockelman, 2017).

Estos datos demuestran la necesidad de Australia de mejorar la seguridad vial (Malik et al., 2020). Los estudios sobre accidentes de tráfico ayudan a los expertos a comprender las causas de los accidentes y a encontrar maneras de reducir su frecuencia y gravedad. Esta información puede ayudar a salvar vidas, reducir las lesiones y disminuir las consecuencias económicas de los accidentes de tráfico. Por lo tanto, los estudios sobre accidentes de tráfico mejoran la seguridad vial y protegen la salud y el bienestar de los australianos.

El propósito de este trabajo es investigar los elementos de seguridad vial en el área metropolitana de Melbourne, Victoria, mediante el análisis de 15 años (2006-2020) de datos sobre accidentes de tránsito para identificar patrones y tendencias. Por lo tanto, el tema de investigación puede formularse de la siguiente manera: ¿Cuáles son los patrones y tendencias más notables en los elementos de seguridad vial en el área metropolitana de Melbourne, Victoria, durante un período de 15 años, según los datos de accidentes de tránsito, y cómo se puede utilizar esta información para mejorar la seguridad vial?

1. Antecedentes

Debido a los numerosos factores que contribuyen a los accidentes de tráfico, el uso del análisis de accidentes mediante big data es cada vez más frecuente. Para predecir los factores que contribuyen a los accidentes de tráfico, se ha explorado la aplicación de técnicas de aprendizaje automático a gran escala con el framework Apache Spark y algoritmos de árboles de decisión. Según Ait-Mlouk et al. (2017), el modelo propuesto puede ayudar a los responsables de la toma de decisiones en el análisis y la mejora de la seguridad vial.

Además de la integración de datos, se han utilizado máquinas de vectores de soporte, máquinas de correlación y bondad multinomial para predecir datos de tráfico. Lokala et al. (2017) exploraron el uso de la metodología de SVM para clasificadores de texto basados en datos de tráfico. Mediante el preprocesamiento y la selección de datos, se ha utilizado el lenguaje de programación R para analizar datos y visualizaciones de tráfico, así como para seleccionar ubicaciones con alta frecuencia de accidentes. Chen (2017) también modeló datos mediante árboles de decisión, regresión lineal y algoritmos de bosque aleatorio, y validó la precisión del modelo.

Xie et al. (2017) destacaron la ventaja del análisis de big data al proporcionar estimaciones más precisas de los efectos de los factores de riesgo y facilitar la identificación de puntos críticos a gran escala. Park et al. (2016) recomiendan el marco Hadoop para el desarrollo de un modelo de resolución y predicción de problemas. Además, utilizaron una técnica de muestreo para equilibrar los datos, los corrigieron clasificándolos en grupos y aplicaron análisis de clasificación para mejorar la precisión de sus predicciones.

Shao et al. (2020) investigaron el efecto de la actualización de equipos ferroviarios mediante el coeficiente de información máxima (CIM) para examinar el análisis de big data de accidentes con equipos ferroviarios. La aplicación del análisis de big data ha producido resultados prometedores en la predicción y prevención de accidentes de tráfico. El uso de diversos algoritmos, marcos y técnicas en el análisis de datos ha permitido realizar predicciones más precisas, identificar puntos críticos y resolver desequilibrios en los datos. Estudios adicionales sobre la toma de decisiones mediante análisis de big data y su aplicación en la seguridad vial pueden proporcionar información valiosa a los responsables de la toma de decisiones en el sector del transporte.

Se han realizado numerosos estudios sobre el uso de big data en el análisis de accidentes, pero aún existen ciertas lagunas de investigación que deben abordarse (Marjani et al., 2017; fan et al., 2014; Wamba et al., 2015; Chen et al., 2015). La recopilación y el análisis de datos sobre accidentes no están estandarizados. Esto podría causar errores en los datos y dificultar la comparación de estudios. Esto dificulta estudios detallados y la precisión de los modelos de pronóstico. Si bien los datos sobre accidentes son esenciales para comprender los patrones de accidentes e identificar los factores de riesgo, la integración de otras fuentes de datos, como el clima, el tráfico y el comportamiento del conductor, puede proporcionar una visión más completa de las causas de los accidentes. La mayoría de los datos sobre accidentes se adquieren retrospectivamente, lo que limita los juicios e intervenciones en tiempo real. La recopilación y el análisis de datos en tiempo real pueden acelerar las intervenciones y minimizar los accidentes. Si bien la mayoría de los estudios sobre accidentes se centran en los accidentes automovilísticos, es necesario incluir a peatones y ciclistas en el análisis de accidentes. Esto puede detectar distintas variables de riesgo y mejorar las intervenciones de seguridad vial. Las ventajas del análisis de big data deben distribuirse equitativamente entre las poblaciones. Es importante incluir a diversos usuarios de la vía en la recopilación y el análisis de datos.

2. Metodología

La metodología empleada en este artículo implica el uso de big data para analizar los accidentes de tráfico.

En concreto, utilizamos el conjunto de datos Victoria Crash, que contiene 65 535 muestras del área metropolitana de Melbourne. Nuestro objetivo fue evaluar la importancia de las características y desarrollar dos modelos. Se seleccionó Random Forest como el tipo adecuado para el modelado.

Random Forest es un enfoque común de aprendizaje conjunto para clasificación y regresión.

Se desarrollan múltiples árboles de decisión de forma independiente en un subconjunto aleatorio de datos de entrenamiento y características (Katuwal et al., 2018). Los árboles de decisión se entrenan mediante agregación

bootstrap, o bagging, en la que cada árbol de decisión se entrena en un subconjunto aleatorio independiente de los datos de entrenamiento.

La técnica Random Forest separa cada nodo del árbol de decisión en función de la característica con la mayor ganancia de información durante el entrenamiento (Breiman, 2001). La ganancia de información mide en qué medida una característica reduce la incertidumbre de la predicción. La técnica realiza esto recursivamente hasta que se satisface un requisito de parada, como una profundidad máxima del árbol o un número mínimo de muestras por nodo hoja. Una vez entrenados todos los árboles de decisión, el algoritmo predice la clase (clasificación) o el valor (regresión) agregando sus predicciones. La predicción más frecuente entre todos los árboles se utiliza en la clasificación. El promedio de las predicciones del árbol es la predicción final en la regresión. Random Forest supera a varios otros modelos de aprendizaje automático en precisión, escalabilidad y resistencia al sobreajuste. Puede gestionar conjuntos de datos de alta dimensión con muchas características (Hegelich, 2016).

3. Resultados

4.1 Estudios de correlación

En esta investigación, la característica objetivo fue PerNum1, que representa la proporción de personas lesionadas o fallecidas en un accidente. Para investigar la relación entre esta característica objetivo y

Se construyeron diversos factores, se construyeron diagramas de dispersión y se calcularon coeficientes de correlación de Pearson para las características Días de la semana, Meses y Aire acondicionado.

La Figura 1 muestra la correlación entre PerNum1 y los diversos valores de la característica de día laborable mediante un diagrama de dispersión. Los resultados muestran que los sábados (0,029) y los domingos (0,043) presentan los valores de correlación más altos con el número de personas lesionadas o fallecidas en accidentes, mientras que los lunes (0,002) y los jueves (0,017) presentan los valores de correlación más bajos con la gravedad de los accidentes, en igualdad de condiciones. Cabe destacar que la correlación negativa entre los viernes, jueves y martes y la gravedad de los accidentes sugiere que los accidentes en estos días son menos graves.

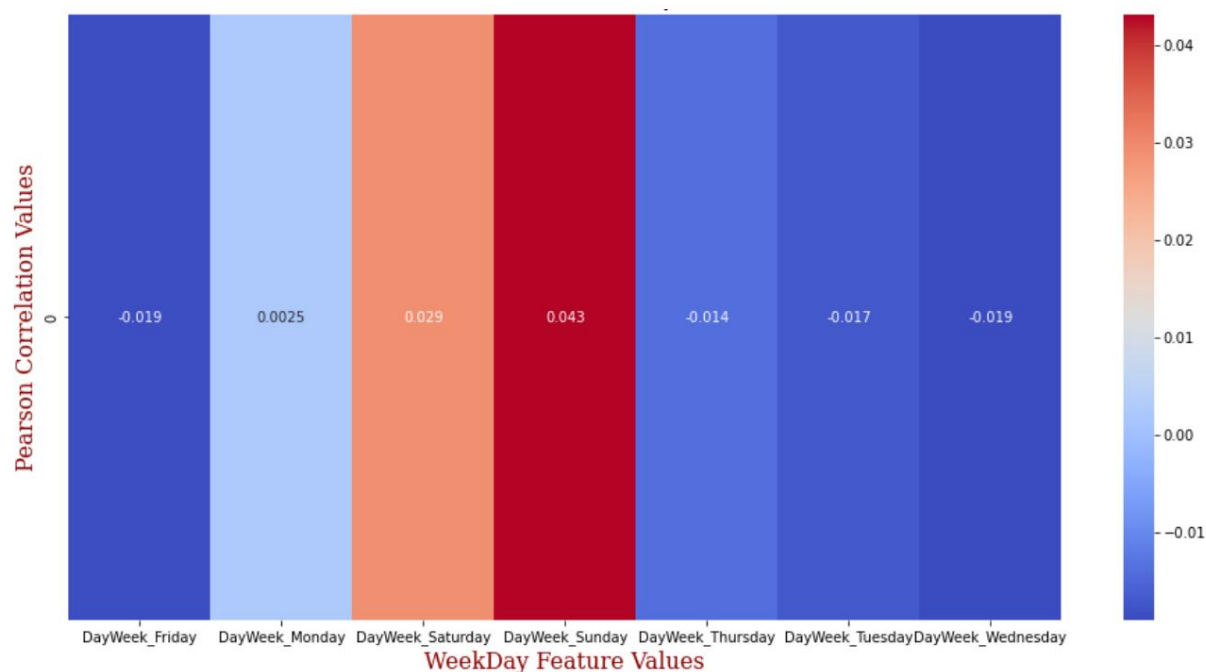


Figura 1: Correlación entre la gravedad del accidente y los días

La Figura 2 ilustra la correlación entre PerNum1 y los distintos meses. Los resultados indican que enero presenta la mayor correlación (0,018) con el número de personas lesionadas o fallecidas en accidentes, seguido de diciembre (0,014). Mayo (0,001), octubre (-0,003) y abril (0,001) son los meses con la menor correlación entre el mes y la gravedad del accidente.

Curiosamente, las correlaciones negativas entre los meses de agosto, febrero, julio, junio, octubre y septiembre implican que los accidentes en estas estaciones ocurren con menor gravedad, lo que se traduce en menores costos y daños.

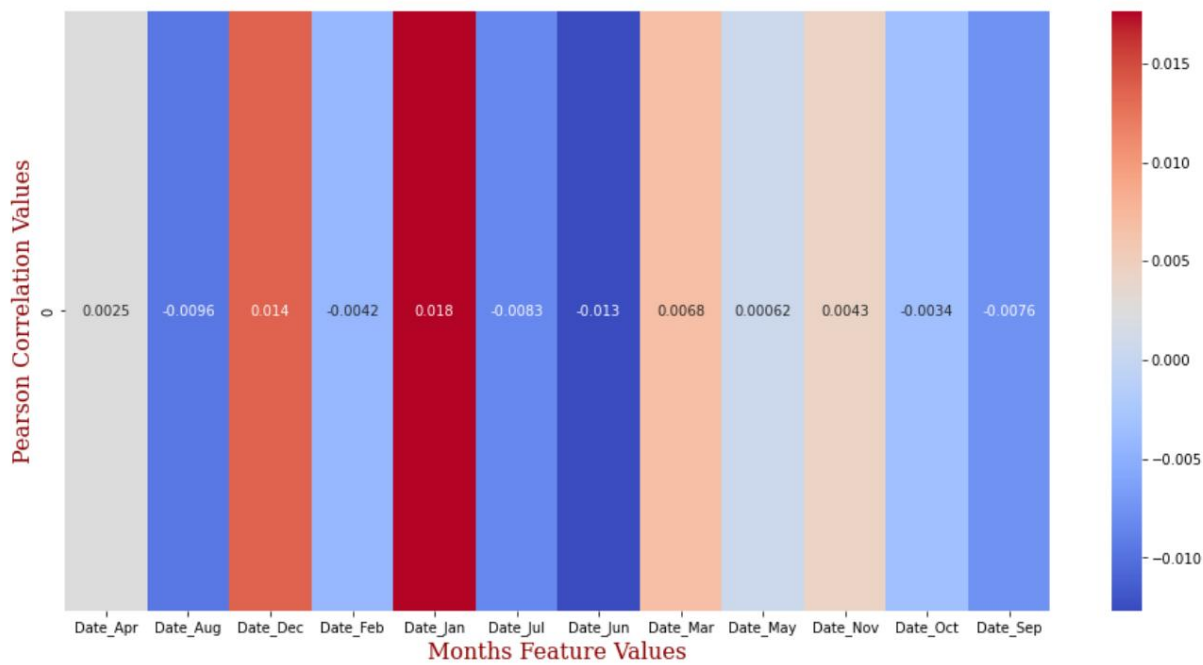


Figura 2: Correlación entre la gravedad del accidente y los meses

La Figura 3 muestra la correlación entre PerNum1 y los diversos valores de la característica Aire Acondicionado. Los resultados indican que las condiciones meteorológicas de Niebla (0,027), Lluvia (0,039) y Viento Fuerte (0,02) presentan la mayor correlación con el número de heridos o fallecidos en accidentes, lo que las convierte en las peores condiciones meteorológicas para la seguridad vial. Curiosamente, la correlación negativa entre el aire despejado y la gravedad de los accidentes sugiere que, con tiempo despejado, la gravedad de los incidentes es baja. Además, la nieve tiene una correlación débil con la gravedad de los accidentes, lo que podría deberse a la baja proporción de días nevados.

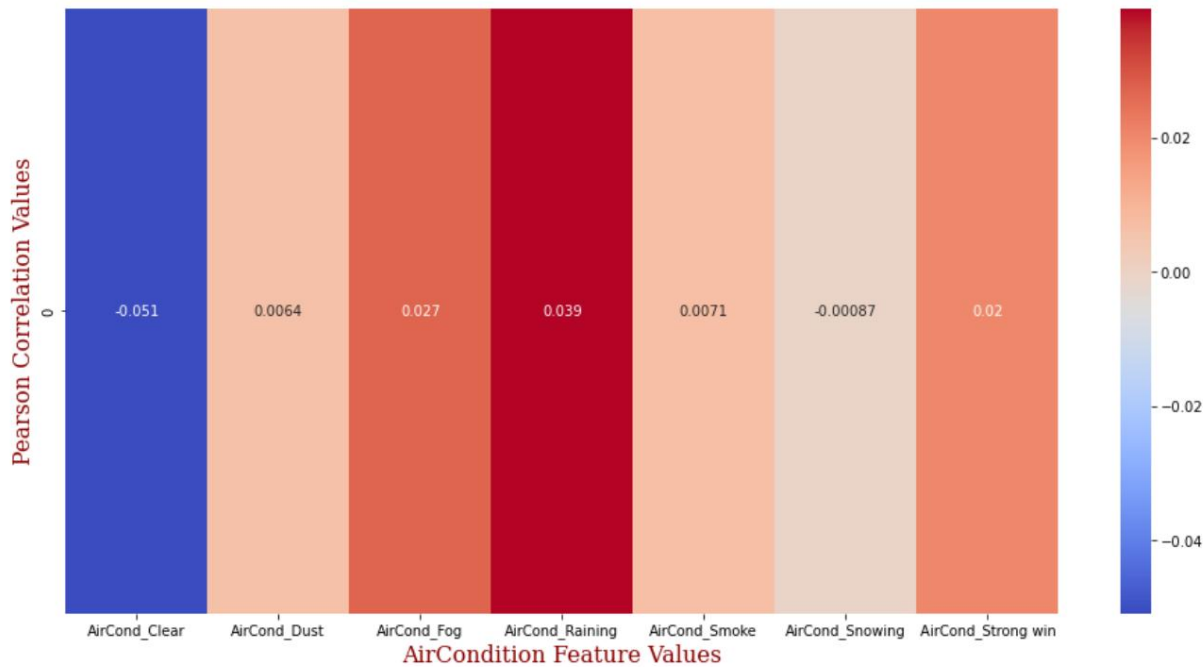


Figura 3: Correlación entre la gravedad del accidente y las condiciones climáticas

El modelo RF intenta predecir el porcentaje de personas heridas o muertas en el accidente (como índice de gravedad), que es una tarea de regresión que emplea la columna de destino PerNum1.

El rango de valores objetivo para esta tarea de regresión es [0,1]. Para completar estas tareas, seguimos un procedimiento de tres pasos que consiste en preprocesamiento, validación cruzada, y entrenamiento y evaluación. En las secciones siguientes del documento se describen los detalles de cada paso.

4.2 Preprocesamiento

Nuestro conjunto de datos, con características numéricas y categóricas mixtas, muestras duplicadas y valores faltantes, requirió gestión durante la fase de preprocesamiento. Primero, eliminamos las muestras duplicadas según el número de accidentes, lo que resultó en 58.982 muestras únicas. Posteriormente, abordamos los valores faltantes en características como LGA_NAME, Light, TypeInters, SpeedZone y AirCond, donde se consideraron faltantes valores indeterminados. Completar los valores faltantes es crucial para obtener un modelo preciso, por lo que utilizamos el enfoque más frecuente, reemplazando los valores faltantes con el valor más frecuente en toda la columna de características. Además, convertimos el atributo de tiempo en dos atributos, AM_Peak y PM_Peak, cada uno con un valor de uno entre las 6:30 y las 9:30 a. m. y las 15:30 y las 18:30 p. m. Tras completar la etapa de valores faltantes, las características categóricas se convirtieron a valores numéricos mediante codificación one-hot. Este método genera una nueva característica ficticia para cada valor único en la columna de características nominales, lo que resulta en un valor binario que especifica la ubicación de un ejemplo. Finalmente, se concatenaron características de codificación one-hot y características numéricas para generar un vector de características de longitud 67. La siguiente tabla enumera todas las características y sus respectivas categorías.

Utilizamos la codificación one-hot para convertir las características categóricas en valores numéricos y representarlas numéricamente. Este método requirió la creación de una nueva característica artificial para cada valor único en la columna de característica nominal. Convertimos la característica "Ubicación" en las siguientes cinco nuevas características: TOWNS, SMALL_TOWNS, MELB_URBAN, SMALL_CITIES y RURAL_VICTORIA. Posteriormente, se utilizaron valores binarios para especificar la ubicación precisa de una muestra. Por ejemplo, si el valor de la característica "Ubicación" de una muestra era RURAL_VICTORIA, se codificaría como RURAL_VICTORIA=1, SMALL_TOWNS=0, MELB_URBAN=0, SMALL_CITIES=0, TOWNS=0.

Tabla 1: Características utilizadas en los modelos de entrenamiento

características	Tipo	Tamaño de la característica numérica vector
NOMBRE LGA	categórico	19
Ubicación	categórico	5
Fecha	categórico	12
DíaSemana	categórico	7
Luz	categórico	6
TipoInters	categórico	8
Aire acondicionado	categórico	7
Zona de velocidad	numérico	1
AM_Peak	numérico	1
PM_Peak	numérico	1
Número total de funciones		67

4.3 Validación cruzada

Para ajustar y evaluar los modelos, se utilizó el método de validación cruzada de k-folds. El conjunto de datos se dividió aleatoriamente en k-folds sin reemplazo, con (k - 1) folds como conjunto de entrenamiento y un fold para evaluar el rendimiento del modelo. Este procedimiento se repitió k veces para obtener k modelos y estimaciones de rendimiento, siendo el rendimiento final del modelo la media de los k folds. Tanto los problemas de clasificación como los de regresión se resolvieron utilizando un valor k de 10.

4.4 Tarea de clasificación

El objetivo de la tarea de regresión fue predecir la proporción de personas que resultaron heridas o murieron en un accidente; para este propósito se utilizó el modelo Random Forest.

Se utilizó el criterio del error cuadrático para entrenar el modelo con una profundidad máxima de árbol de 10 y 100 árboles. Para evaluar el rendimiento del modelo, se aplicó una validación cruzada de 10 pasos, obteniendo 53 084 muestras de entrenamiento y 5898 muestras de prueba. La importancia promedio de las cuarenta características de la validación cruzada de 10 pasos se muestra en la Tabla 2, y la Figura 4 muestra las diez características más significativas. La validación cruzada de 10 pasos arrojó un error cuadrático medio (RMSE) promedio de 0,2021 y una desviación estándar (DE) de $\pm 0,01$, que se utilizó para evaluar el rendimiento del modelo.

Tabla 2: Importancia media de 40 características

NOMBRE DE LA FUNCIÓN	IMPORTANCIA
1) TIPO_COLISIÓN CON UN OBJETO FIJO	0.4940
2) TIPO_VEHÍCULO VOLCADO (SIN COLISIÓN)	0,1157
3) TIPO_SIN COLISIÓN Y SIN OBJETO GOLPEADO	0,1157
4) DCA_VEHÍCULO CHOCA CON VEHÍCULO ESTACIONADO EN A LA IZQUIERDA DE LA CARRETERA	0.0948
5) DCA_REAR END(VEHÍCULOS EN EL MISMO CARRIL)	0.0337
6) TIPO_COLISIÓN CON ALGÚN OTRO OBJETO	0.0191
7) DCA_OTHER EN LA RUTA	0.0138
8) ZONA DE VELOCIDAD	0.0099
9) TIPO_GOLPEADO ANIMAL	0.0097
10) DCA_DE CABEZA (NO ADELANTAMIENTO)	0.0090
11) DCA_FUERA DE CONTROL EN LA CALZADA (ENCENDIDO) DERECHO)	0,0035
12) DCA_ANIMAL GOLPEADO	0,0025
13) DCA_CARGA O VEHÍCULO IMPACTO CON MISIL	0,0025
14) TIPO_CAÍDA DESDE O DENTRO DE UN VEHÍCULO EN MOVIMIENTO	0,0024
15) TYPEINTERS_NOT EN LA INTERSECCIÓN	0.0019
16) DÍA SEMANA_MARTES	0.0019
17) DÍA SEMANA_SÁBADO	0.0019

Actas de la ATRF 2023

18) LUCES DE LA CALLE ENCENDIDAS	0,0018
19) LUZ_DÍA	0,0018
20) DÍASEMANA_DOMINGO	0,0017
21) INTERSECCIÓN TYPEINTERS_CROSS	0,0016
22) LGA_NAME_CASEY	0,0016
23) AIRE ACONDICIONADO_LLUVIAS	0,0015
24) DÍA SEMANA_VIERNES	0,0015
25) FECHA_JUL	0,0015
26) LGA_NOMBRE_MONASH	0,0015
27) DCA_SALIDA DE LA CARRETERA Y CHOQUE CONTRA UN OBJETO/ESTACIONADO VEHÍCULO	0,0015
28) DÍASEMANA_MIÉRCOLES	0,0014
29) FECHA_ENE	0,0014
30) LGA_NAME_PENÍNSULA DE MORNINGTON	0,0014
31) DÍASEMANA_JUEVES	0,0014
32) LGA_NOMBRE_FRANKSTON	0,0014
33) FECHA_JUN	0.0013
34) FECHA_OCT	0.0013
35) UBICACIÓN_RURAL_VICTORIA	0.0013
36) UBICACIÓN_MELB_URBAN	0.0013
37) AIRE ACONDICIONADO_DESPEJADO	0,0012
38) FECHA_SEP	0,0012
39) FECHA_AGO	0,0012
40) FECHA_ABR	0,0012

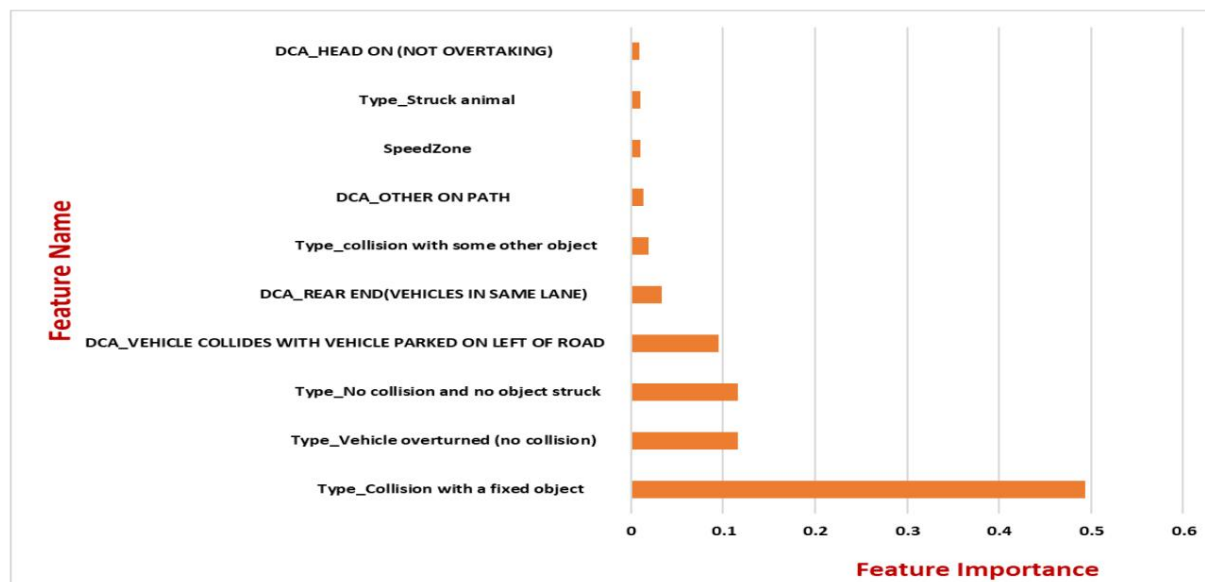


Figura 4: Las 10 características más importantes

5. Conclusión

La contribución de esta investigación radica en la aplicación del análisis de big data al análisis de colisiones viales y a la evaluación de la importancia de diversos factores para predecir la gravedad de los accidentes. En concreto, la investigación demuestra el uso del conjunto de datos Victorian Crash para examinar la relación entre el porcentaje de personas lesionadas o fallecidas en un accidente y factores como los días laborables, los meses y las condiciones meteorológicas. Además, el estudio construye y evalúa un modelo de Bosque Aleatorio para predecir el porcentaje de víctimas de accidentes que resultaron heridas o fallecidas.

Los hallazgos del estudio pueden ayudar a los responsables políticos y a las autoridades de transporte a diseñar estrategias para mejorar la seguridad vial y reducir la probabilidad de accidentes. Además, el estudio destaca el potencial del análisis de big data para analizar datos sobre accidentes de tráfico e identificar factores significativos que contribuyen a su gravedad. El estudio también identifica varias líneas de investigación futuras, como la investigación de otros factores que pueden contribuir a la gravedad de los accidentes, como las condiciones de la carretera, el comportamiento del conductor y el tipo de vehículo.

Referencias

- Ait-Mlouk, A., Gharnati, F. y Agouti, T. (2017). Aplicación del análisis de big data con árboles de decisión para accidentes de tráfico. *Revista India de Ciencia y Tecnología*, 10(29), 1-10.
- Instituto Australiano de Salud (2012). Una imagen de los niños australianos en 2012. AIHW.
- Oficina de Investigación Económica de Infraestructura y Transporte (BITRE) (2022). Comparaciones internacionales de seguridad vial 2020 BITRE, Canberra, ACT.
- Blincoe, LJ, Seay, AG, Zaloshnja, E., Miller, TR, Romano, EO, Luchter, S. y Spicer, RS (2002). El impacto económico de los accidentes automovilísticos, 2000 (No. DOT-HS-809-446).
- Estados Unidos. Administración Nacional de Seguridad del Tráfico en las Carreteras.
- Breiman, L. (2001). Bosques aleatorios. *Aprendizaje automático*, 45, 5-32.
- Chen, C. (2017). Análisis y pronóstico de macrodatos de accidentes de tráfico. En *ITM Web of Conferences* (Vol. 12, p. 04029). EDP Sciences.

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, AV y Rong, X. (2015). Minería de datos para el internet de las cosas: revisión bibliográfica y desafíos. *Revista Internacional de Redes de Sensores Distribuidos*, 11(8), 431047.

Clements, LM, y Kockelman, KM (2017). Efectos económicos de los vehículos automatizados.

Registro de investigación en transporte, 2606(1), 106-114.

Fan, J., Han, F. y Liu, H. (2014). Desafíos del análisis de big data. *National Science Review*, 1(2), 293-314.

Faulks, IJ, Lane, M., e Irwin, JD (agosto de 2012). Vigilancia del tránsito y seguridad vial para individuos y poblaciones. En la Conferencia Nacional de la ACRS: «Un sistema seguro: ampliando el alcance».

Hegelich, S. (2016). Árboles de decisión y bosques aleatorios: técnicas de aprendizaje automático para clasificar eventos raros. *European Policy Analysis*, 2(1), 98-120.

Jiang, Z., Shekhar, S., Jiang, Z. y Shekhar, S. (2017). Big Data espacial. *Ciencia de Big Data espacial: Técnicas de clasificación para imágenes de observación de la Tierra*, 3-13.

Katuwal, R., Suganthan, PN y Zhang, L. (2018). Un conjunto de árboles de decisión con redes de enlaces funcionales vectoriales aleatorios para la clasificación multiclase. *Applied Soft Computing*, 70, 1146-1153.

Kitchin, R. (2014). La revolución de los datos: Big data, datos abiertos, infraestructuras de datos y sus consecuencias. Sage.

Lokala, U., Nowduri, S. y Sharma, PK (2017). Minería y visualización de big data de accidentes de tráfico mediante máquinas de vectores de soporte. *Academia Mundial de Ciencias, Ingeniería y Tecnología - Revista Internacional de Ingeniería de Sistemas y Computación*, 10(8).

Malik, S., Swapan, MSH y Khan, S. (2020). Movilidad sostenible mediante carreteras más seguras: Adaptación de la estrategia de seguridad vial al contexto local en Australia Occidental. *Sustainability*, 12(21), 8929.

Marjani, M., Nasaruddin, F., Ghani, A., Karim, A., Hashem, IAT, Siddiqi, A. y Yaqoob, I.

(2017). Análisis de datos de Big IoT: arquitectura, oportunidades y desafíos de investigación abierta. *IEEE acceso*, 5, 5247-5261.

Naweed, A., Balakrishnan, G. y Thomas, MJ (2014). Desafíos del uso de simuladores en la investigación del transporte: lecciones de un estudio de caso sobre seguridad vial. *Road & Transport Research: Revista de Investigación y Práctica de Australia y Nueva Zelanda*, 23(2), 60-70.

Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C. y Yannis, G.

(2019). Revisión y clasificación de los factores de riesgo de accidentes relacionados con la infraestructura vial. *Análisis y Prevención de Accidentes*, 125, 85-97.

Park, SH, Kim, SM y Ha, YG (2016). Predicción de accidentes de tráfico mediante análisis de big data VDS. *The Journal of Supercomputing*, 72, 2815-2831.

Pika, A., ter Hofstede, AH, Perrons, RK, Grossmann, G., Stumptner, M., y Cooley, J.

(2021). Uso de big data para mejorar el rendimiento de la seguridad: una aplicación de la minería de procesos para optimizar la visualización de datos. *Big Data Research*, 25, 100210.

Scott-Parker, B. y Oviedo-Trespalacios, O. (2017). Conductas de riesgo de conductores jóvenes y predictores del riesgo de accidentes en Australia, Nueva Zelanda y Colombia: ¿Igualdad, pero diferencia?. *Accident Analysis & Prevention*, 99, 30-38.

Shao, F., Yang, S., Sun, B., Jia, L., Dong, Y. y Wang, D. (2020). Análisis de big data de accidentes de equipo ferroviario basado en el coeficiente de información máxima. *Journal of Transportation Safety & Security*, 12(7), 959-976.

Soltani, A. y Askari, S. (2017). Exploración de la autocorrelación espacial de los accidentes de tráfico según su gravedad. *Injury*, 48(3), 637-647.

Soltani, A., Azmoodeh, M. y Qadikolaie, MR (2023). Accidentes de tráfico en la región metropolitana de Adelaida, las consecuencias del COVID-19. *Revista de Transporte y Salud*, 30, 101581.

Actas de la ATRF 2023

Wamba, SF, Akter, S., Edwards, A., Chopin, G. y Gnanzou, D. (2015). Cómo el «big data» puede generar un gran impacto: Hallazgos de una revisión sistemática y un estudio de caso longitudinal. *Revista internacional de economía de la producción*, 165, 234-246.

Wang, Y., Kung, L., Wang, WYC y Cegielski, CG (2018). Un análisis integrado de big data habilitado

Organización Mundial de la Salud. (2015). Informe sobre la situación mundial de la seguridad vial 2015. Organización Mundial de la Salud.

Xie, K., Ozbay, K., Kurkcu, A. y Yang, H. (2017). Análisis de accidentes de tráfico con peatones involucrados mediante big data: Investigación de factores contribuyentes e identificación de puntos críticos.

Análisis de riesgos, 37(8), 1459-1476.