

## SEGURIDAD ESPACIAL DE LOS PEATONES EN LAS ZONAS ESCOLARES DE RIAD: UN ENFOQUE BASADO EN DATOS

**Evaluación de la seguridad espacial de los peatones en las zonas escolares de Riad mediante regresión lineal múltiple y aprendizaje automático: un enfoque basado en datos**

ALA HUSNI ALSOUD <sup>1</sup>, AHMAD H. ALOMARI <sup>2</sup> MOAMAR QRARAH <sup>3</sup>, ZAKI ABU AHMAD <sup>4</sup>

<sup>1</sup> Director ejecutivo, Departamento de Infraestructuras, Consolidated Consultants Group (CCG), Riad, Arabia Saudí. Correo electrónico: [a.alsoud@group-cc.com](mailto:a.alsoud@group-cc.com)

<sup>2</sup> Autor correspondiente, experto en tráfico y transporte, Consolidated Consultants Group (CCG). Correo electrónico: [a.alomari@group-cc.com](mailto:a.alomari@group-cc.com), y profesor de Ingeniería Civil, Universidad de Yarmouk, Irbid, Jordania. Correo electrónico: [alomarish@yu.edu.jo](mailto:alomarish@yu.edu.jo)

<sup>3</sup> Ingeniero sénior de carreteras y puentes, director de proyectos, Departamento de Infraestructuras, Consolidated Consultants Group (CCG), Riad, Arabia Saudí. Correo electrónico: [M.Qrarah@group-cc.com](mailto:M.Qrarah@group-cc.com)

<sup>4</sup> Ingeniero sénior de Tráfico y Transporte, director de Proyectos, Departamento de Infraestructuras, Consolidated Consultants Group (CCG), Riad, Arabia Saudí. Correo electrónico: [z.abuahmad@group-cc.com](mailto:z.abuahmad@group-cc.com)

### PALABRAS CLAVE

Zona escolar  
Seguridad vial  
Peatones  
Espacial  
Regresión  
Aprendizaje automático  
Inteligencia artificial

### RESUMEN

*Este estudio analiza de forma exhaustiva la densidad de atropellos de peatones (el número de atropellos por kilómetro cuadrado de superficie del distrito) en las zonas escolares de Riad, empleando técnicas avanzadas de inteligencia artificial (IA), como la regresión lineal múltiple (MLR) y el aprendizaje automático (ML), con el fin de mejorar la eficiencia urbana, la calidad de vida y la resiliencia. Se recopilaron datos de 884 zonas escolares distribuidas por Riad, que abarcaban diversos contextos infraestructurales, socioeconómicos y demográficos. El modelo MLR optimizado identificó predictores significativos, como la longitud de las carreteras del distrito, la gravedad media de los accidentes (EPDO), la renta media, la densidad de población y la disponibilidad de paradas de transporte público, que en conjunto explicaban aproximadamente el 65 % de la variabilidad de la densidad de accidentes. El modelo ML Bosque aleatorio mejoró aún más la precisión predictiva ( $R^2 \approx 0,88$ ), revelando interacciones complejas y no lineales entre variables clave, como el volumen de tráfico, los límites de velocidad, el número de carriles, la disponibilidad de pasos de peatones y la población estudiantil. Al integrar la regresión tradicional con metodologías ML de vanguardia, esta investigación proporciona información útil para los responsables políticos, los planificadores urbanos y los ingenieros, lo que permite intervenciones específicas y basadas en datos para mejorar la seguridad de los peatones y promover una movilidad urbana sostenible e inteligente en las zonas escolares de Riad.*

Recibido: 03 / 08 / 2025

Aceptado: 01 / 10 / 2025

## 1. Introducción

Las lesiones causadas por accidentes de tráfico se encuentran entre las principales causas de muerte y discapacidad en todo el mundo, y afectan a millones de personas cada año. Según la Organización Mundial de la Salud (OMS, 2023a), los accidentes de tráfico causan 1,19 millones de muertes y más de 50 millones de heridos cada año. El elevado número de muertes entre los usuarios vulnerables de la vía pública, como ciclistas, motociclistas y peatones, demuestra la necesidad de estrategias de intervención específicas (OMS, 2023a). Los niños de entre 5 y 19 años son los usuarios vulnerables de la vía pública que corren mayor riesgo. La principal causa de muerte de los niños de este grupo de edad son los accidentes, y las zonas escolares presentan peligros adicionales debido a la gran actividad peatonal y al limitado conocimiento del tráfico por parte de los jóvenes estudiantes (UNICEF, 2022).

Las investigaciones muestran que los niños de entre 5 y 19 años corren un alto riesgo de sufrir lesiones por accidentes de tráfico, especialmente cuando van y vuelven del colegio (DiMaggio y Li, 2013). Los programas Safe Routes to School (SRTS) en Estados Unidos, Canadá y Europa han demostrado que las medidas de pacificación del tráfico, los cruces peatonales designados y los límites de velocidad más bajos son eficaces para reducir estos riesgos (Lee et al., 2024). Sin embargo, muchos países en desarrollo siguen teniendo dificultades para aplicar medidas de seguridad en las zonas escolares, lo que se traduce en un mayor número de muertes y lesiones (Ehsani et al., 2023). Estos retos deben abordarse mediante modelos predictivos, la toma de decisiones basada en datos e intervenciones sistemáticas adaptadas al entorno urbano local.

En los últimos diez años se han logrado avances sustanciales en las medidas de seguridad vial de Arabia Saudita, ya que el país redujo las muertes por accidentes de tráfico de 28 a 18,5 por cada 100 000 habitantes entre 2016 y 2021, gracias a una mayor aplicación de la ley y a una mejor supervisión de la velocidad y la infraestructura vial (OMS, 2023b). La combinación de una alta dependencia del vehículo y una infraestructura peatonal inadecuada con la congestión del tráfico en las zonas escolares urbanas crea importantes riesgos para la seguridad de los peatones (Alharbi et al., 2024). La iniciativa nacional Visión 2030 trabaja para mejorar la seguridad vial mediante el desarrollo de tecnologías de transporte avanzadas y de infraestructuras, pero aún hay factores desconocidos sobre cómo las características particulares de las zonas escolares influyen en el riesgo de accidentes.

El peligro inmediato de accidentes y muertes por colisiones en zonas escolares crea retos adicionales para la salud pública y la movilidad urbana. Las investigaciones muestran que los entornos inseguros para desplazarse al colegio desincentivan que los niños vayan andando o en bicicleta al colegio, lo que conduce a un aumento de las tasas de obesidad y a una reducción de la actividad física (Universidad de Cambridge, 2019). La seguridad de los entornos escolares representa tanto una prioridad en materia de transporte como una necesidad de salud pública. Este estudio analiza los riesgos de accidentes en las zonas escolares de Riad, Arabia Saudita, con el fin de proporcionar conclusiones basadas en datos sobre los factores de los accidentes de tráfico en el entorno escolar, que respalden las decisiones de seguridad basadas en datos y las intervenciones específicas.

Esta investigación evalúa la densidad de atropellos de peatones (número de accidentes por kilómetro cuadrado de superficie del distrito) en las zonas escolares de Riad, Arabia Saudí, utilizando métodos de regresión lineal múltiple (MLR) y aprendizaje automático (ML). El análisis investiga un conjunto de variables, entre las que se incluyen la geometría de las carreteras, las características de las escuelas, las zonas de amortiguación circundantes y las características de todo el distrito, con el fin de determinar su efecto en la frecuencia de los atropellos de peatones. El estudio emplea MLR y ML como marcos analíticos para descubrir predictores estadísticamente significativos de la densidad de atropellos de peatones y establecer conocimientos basados en datos sobre los factores de riesgo para la seguridad en las zonas escolares.

Las siguientes secciones describen la estructura de este documento: *La sección 2* revisa las investigaciones existentes sobre la seguridad vial en las zonas escolares desde un punto de vista

mundial y regional, junto con los enfoques de modelización estadística para la evaluación del riesgo de atropello de peatones. *La sección 3* explica el área de estudio, las fuentes de datos y el marco analítico, que utiliza métodos de regresión MLR y ML para estudiar la densidad de atropellos de peatones y sus factores influyentes relacionados. *La sección 4* presenta los resultados del análisis exploratorio de datos, las evaluaciones de correlación y los modelos de regresión, que muestran la significación estadística de los diferentes predictores, el rendimiento del modelo MLR optimizado y los resultados del análisis de aprendizaje automático. La sección final de este documento (*sección 5*) resume las principales conclusiones y ofrece recomendaciones para mejorar la seguridad de los peatones en las zonas escolares de Riad.

## 2. Revisión bibliográfica

Garantizar la seguridad vial en las zonas escolares es una prioridad mundial debido a la mayor vulnerabilidad de los niños durante sus desplazamientos al colegio. Se han realizado numerosas investigaciones para explorar los factores que influyen en los accidentes en las zonas escolares, como el diseño de las carreteras, el comportamiento de los conductores, las políticas de aplicación de la ley y las disparidades socioeconómicas, con el fin de comprender su impacto en la seguridad de los peatones. Mientras que los estudios de los países desarrollados hacen hincapié en la eficacia de las intervenciones estructuradas, como los programas de rutas seguras al colegio (SRTS), las medidas de pacificación del tráfico y la aplicación automatizada de la ley, las investigaciones de los países en desarrollo ponen de relieve desafíos como la débil aplicación de la normativa, la alta dependencia de los vehículos y la infraestructura peatonal inadecuada, que contribuyen a aumentar desproporcionadamente los riesgos. Con la creciente aplicación de técnicas de modelización estadística, en particular el MLR, junto con el análisis espacial y la modelización predictiva, estudios recientes han tratado de cuantificar los riesgos de accidentes de peatones e identificar las zonas de alto riesgo. Esta revisión bibliográfica sintetiza las investigaciones existentes examinando los principales factores predictivos de la densidad de atropellos de peatones, evaluando las estrategias de intervención y abordando las lagunas de las metodologías actuales, en particular en el contexto de las zonas escolares en entornos urbanos en desarrollo como Riad.

### 2.1. Seguridad vial en las zonas escolares: panorama global

La seguridad de las zonas de tráfico alrededor de las escuelas sigue siendo una prioridad mundial, ya que los niños se enfrentan al mayor riesgo como usuarios de la vía pública cuando se desplazan hacia y desde la escuela. Kingham et al. (2011) observaron en Nueva Zelanda que la mejora de las medidas de seguridad condujo a una disminución de los accidentes de tráfico, pero los desplazamientos escolares seguían siendo peligrosos, lo que requería medidas de seguridad específicas. La combinación de una aplicación débil de la ley con una alta congestión del tráfico y una infraestructura peatonal insuficiente en las zonas escolares crea graves problemas de seguridad vial, según investigaciones realizadas en la India y Camerún (Lordswill et al., 2024; Tetali et al., 2016). Las investigaciones muestran que las medidas de seguridad para los peatones, como el control de la velocidad y las zonas de cruce designadas, reducen eficazmente los accidentes en los países ricos, pero las limitaciones en materia de aplicación de la ley e infraestructura obstaculizan el progreso en los países de ingresos bajos y medios (Bahrami et al., 2024; Rothman et al., 2017a). La diferencia significativa entre las regiones desarrolladas y en desarrollo demuestra la necesidad inmediata de soluciones de seguridad universales que unan el desarrollo de infraestructuras con la ejecución de políticas y la participación de la comunidad para proteger las zonas escolares.

### 2.2 Factores de riesgo clave que afectan a la seguridad de las zonas escolares

El diseño físico de las carreteras, junto con sus intersecciones, determina el grado de protección de los estudiantes frente a los riesgos. Según Zhao et al. (2015), las investigaciones han

demostrado que los dispositivos de control del tráfico deficientes, combinados con pasos de peatones insuficientes y carreteras de alta velocidad, crean condiciones peligrosas para los peatones cerca de las escuelas. Canadá y Corea del Sur llevaron a cabo investigaciones que demuestran que las carreteras anchas, combinadas con cruces sin control e intersecciones complicadas, provocan un mayor número de lesiones a los peatones, mientras que los pasos de peatones elevados y los controles de señales para peatones reducen el riesgo de lesiones (Lee et al., 2016; Rothman et al., 2015). El riesgo de accidentes de tráfico aumenta en las zonas comerciales con alta densidad de población que rodean las escuelas, mientras que los barrios de uso mixto con una sólida infraestructura peatonal dan lugar a una reducción de las tasas de lesiones (Yu y Zhu, 2016). Según Oh y Kim et al. (2025), las zonas escolares representan áreas de menor riesgo para los peatones que los destinos comerciales y de uso mixto que existen después del horario escolar.

Los accidentes en las zonas escolares se deben principalmente a las prácticas de conducción inseguras de los conductores que circulan a exceso de velocidad mientras su atención se divide entre la conducción y las infracciones de tráfico. La investigación de Flanagan y Morgan (2023) demuestra que las prácticas de conducción peligrosas, como el estacionamiento en doble fila, las paradas repentinas y las bajadas inseguras, provocan congestión del tráfico, lo que hace que caminar sea peligroso para los peatones. Hu et al. (2025) analizaron los accidentes en zonas escolares chinas durante las horas punta de la mañana y determinaron que las zonas de mayor riesgo se encontraban en los carriles superiores y exteriores cuando los estudiantes se reunían y durante las horas punta antes de desplazarse a los carriles centrales a medida que se disipaba el tráfico, siendo los cambios de carril la amenaza más peligrosa para la seguridad. La investigación observacional de Tetali et al. (2016) e Ivan et al. (2019) demostró que los niños de comunidades con ingresos bajos y medios muestran un comportamiento peatonal más arriesgado, ya que sus barrios carecen de medidas de seguridad adecuadas.

Los factores socioeconómicos tienen un impacto sustancial en las medidas de seguridad implementadas en las zonas escolares. Las investigaciones de Farid et al. (2024) y Rothman et al. (2017b) muestran que las escuelas estadounidenses y canadienses ubicadas en barrios de bajos ingresos enfrentan altas tasas de accidentes debido a infraestructuras deficientes, débil aplicación de la normativa y opciones limitadas para el cruce de peatones. La regulación de la aplicación de la normativa de tráfico difiere mucho entre los distintos países, ya que algunos implementan cámaras de velocidad automáticas, mientras que otros se enfrentan a mecanismos de aplicación débiles (Eun, 2023). Las investigaciones realizadas en Arabia Saudí y Filipinas demuestran que la seguridad en las zonas escolares mejora cuando las intervenciones basadas en políticas combinan sanciones más sustanciales con inversiones en infraestructura peatonal (Alharbi et al., 2024; Regidor et al., 2023).

### ***2.3 Intervenciones en materia de seguridad escolar y su eficacia***

La combinación de badenes con pasos de peatones elevados y balizas luminosas ha demostrado ser eficaz para reducir los accidentes en las zonas escolares. Las investigaciones indican que los programas Safe Routes to School (SRTS) en Estados Unidos reducen las tasas de lesiones de peatones gracias a la aplicación de medidas de pacificación del tráfico y la mejora de las infraestructuras de cruce (DiMaggio y Li, 2013; Lee et al., 2024). Actualmente se utilizan tecnologías de modelización predictiva y aprendizaje automático (ML) para evaluar los riesgos de accidente, mientras que las cámaras de velocidad automáticas reducen las infracciones por exceso de velocidad en países con sistemas de control estrictos (Eun, 2023; Zhang et al., 2024).

Los modelos de aprendizaje automático supervisados de las decisiones sobre el uso de los pasos de peatones en los densos corredores urbanos de Daca (Bangladesh) demostraron cómo el diseño de las infraestructuras y las operaciones de tráfico crean condiciones de cruce inseguras, lo que las auditorías de Star Rating for Schools (SR4S) pueden utilizar para determinar las medidas adecuadas cerca de las escuelas (Sakib et al., 2024). Otro estudio sobre el comportamiento de los escolares de Daca al cruzar la calle demostró que los programas educativos, combinados con

medidas de control y control de la velocidad del tráfico, pueden ayudar a minimizar los comportamientos peligrosos al cruzar la calle (Basunia et al., 2025).

Los protocolos del Programa Internacional de Evaluación de Carreteras (iRAP) utilizan el aprendizaje automático y la tecnología de visión artificial para llevar las capacidades de evaluación de riesgos a los países de ingresos bajos y medios (PIBM), lo que permite a las autoridades realizar evaluaciones de riesgos en las zonas escolares a gran escala y seleccionar los lugares más importantes para las mejoras de SR4S con el fin de aplicar medidas de seguridad de eficacia probada (iRAP, 2025).

El programa «AI&Me Safe School Zones» (Zonas escolares seguras con IA y yo) de Vietnam utiliza la inteligencia artificial (IA) y el aprendizaje automático para analizar miles de escuelas a través de imágenes satelitales y de Street View con el fin de identificar carreteras peligrosas, lo que luego dirige la financiación para implementar soluciones de ingeniería específicas para cada lugar con el fin de mejorar las zonas escolares en ciudades con recursos limitados (Fundación AIP, 2025).

El papel esencial de los programas de educación y sensibilización se pone de manifiesto en las investigaciones, que demuestran cómo la formación de los estudiantes en materia de seguridad y la participación de los padres crean condiciones más seguras para los desplazamientos escolares (Eun, 2023; Flanagan y Morgan, 2023).

## ***2.4 Lagunas en la investigación existente y orientaciones futuras***

La investigación sobre la seguridad en las zonas escolares ha dado lugar a numerosos hallazgos, pero los científicos aún necesitan comprender cómo la infraestructura urbana interactúa con los factores demográficos y las intervenciones políticas para afectar al riesgo de accidentes de peatones en ciudades en rápido desarrollo como Riad. La mayor parte de la investigación existente se basa en métodos estadísticos tradicionales con supuestos lineales, sin explorar enfoques avanzados de modelización predictiva que detecten relaciones no lineales complejas. Las investigaciones existentes carecen de marcos analíticos integrales que integren el flujo de tráfico con la geometría de las carreteras, el uso del suelo, las medidas de aplicación de la ley y las condiciones socioeconómicas.

La investigación aborda los vacíos de conocimiento existentes mediante la combinación de modelos de regresión lineal múltiple (MLR) con técnicas de aprendizaje automático (ML), utilizando modelos de Bosque aleatorio para incrementar la precisión de las predicciones y revelar relaciones no lineales entre los predictores. La capacidad predictiva de las técnicas de ML supera a la de MLR, mientras que MLR proporciona resultados estadísticos interpretables que ayudan a identificar los factores de riesgo. El estudio combina el análisis espacial con técnicas MLR y ML para crear un marco que produce recomendaciones basadas en pruebas y aplicables para intervenciones específicas de seguridad peatonal en las zonas escolares de Riad.

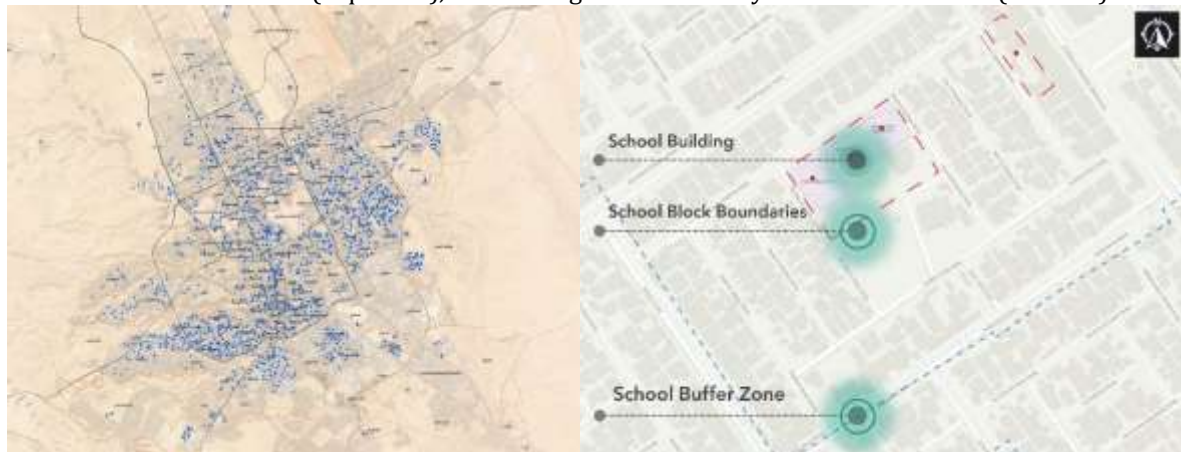
## **3. Metodología**

Riad, la capital de Arabia Saudita, es el centro político, económico y cultural del país. Con una población de 8 591 748 habitantes en 2022, es la ciudad más grande del país, caracterizada por una rápida urbanización, una extensa red de carreteras y una fuerte dependencia del transporte en vehículos privados (Datasaudi, 2025). En el marco de la Visión 2030, el Gobierno saudí ha trabajado activamente para mejorar la seguridad vial y la movilidad urbana sostenible, reconociendo la congestión del tráfico y la seguridad de los peatones como retos clave. La infraestructura educativa de Riad cuenta con 1 430 320 alumnos y 105 942 profesores, distribuidos entre sus escuelas (Datasaudi, 2025). La media de alumnos matriculados por escuela alcanza los 240, mientras que la media de profesores por escuela es de 17,8, lo que demuestra la alta densidad de escuelas y la necesidad de medidas específicas de seguridad vial en torno a estas instalaciones (Datasaudi, 2025). El aumento de la matriculación de alumnos y la expansión de las zonas urbanas hacen que sea esencial analizar los riesgos del tráfico en las zonas escolares de Riad para crear soluciones de seguridad basadas en datos.



Este estudio investiga 884 zonas escolares situadas en todos los distritos y municipios de la ciudad de Riad, como se muestra en la figura 1. Las escuelas elegidas representan diversos entornos socioeconómicos y de infraestructura, lo que proporciona una cobertura espacial completa de la ciudad. La extensa área geográfica permite a los investigadores estudiar la densidad de atropellos de peatones, al tiempo que examinan las características de las carreteras, los patrones de desarrollo urbano y las variaciones en las condiciones del tráfico. Esta investigación combina datos de múltiples entornos escolares para establecer conclusiones sobre la seguridad de las zonas escolares en toda la ciudad, lo que genera recomendaciones de políticas basadas en pruebas para abordar tanto los factores de riesgo específicos como las cuestiones generales de movilidad urbana.

**Figura 1.** Mapa de las zonas escolares seleccionadas y su distribución en todos los distritos y municipios de la ciudad de Riad (izquierda), zona de seguridad escolar y límites de manzana (derecha).



Fuente: Elaboración propia, 2025.

Las zonas escolares de Riad representan áreas de alto riesgo para la seguridad de los peatones, debido principalmente a la intensa interacción entre los vehículos y los escolares durante las horas punta. Para mitigar estos riesgos, se han establecido zonas de seguridad escolar que combinan áreas de alerta temprana con zonas de reducción de velocidad. El análisis tiene en cuenta los datos espaciales relevantes en un radio de 200 y 300 metros de los límites de las escuelas (Figura 1) para evaluar factores contribuyentes como la geometría de las carreteras, la infraestructura peatonal y las medidas de control del tráfico. La base de datos se desarrolló y organizó utilizando ArcGIS Pro (Esri, 2025). El conjunto de datos se caracterizó por atributos en cuatro niveles:

- **Nivel escolar:** número de clases, alumnos, profesores y administradores.
- **Nivel de distrito:** longitud de las carreteras, densidad de escuelas, uso del suelo (zona comercial/de uso mixto), densidad de atropellos de peatones, promedio de muertes/lesiones, ingresos medios, densidad de población y porcentaje de población de 6 a 18 años.
- **Nivel de la zona de seguridad escolar:** flujo de tráfico, velocidad de los vehículos, intersecciones y disponibilidad de paradas de transporte público.
- **Nivel de manzana adyacente a la escuela:** número de carriles y velocidad promedio.

La tabla 1 ofrece una lista detallada de las variables estudiadas, clasificándolas según su relevancia para las zonas escolares, las influencias a nivel de distrito y las características de las escuelas.

**Tabla 1.** Resumen de las variables estudiadas.

| Categoría                     | Nombre de la variable |                | Descripción                                            |
|-------------------------------|-----------------------|----------------|--------------------------------------------------------|
| Características de la escuela | 1                     | N.º de clases  | El número total de aulas de la escuela.                |
|                               | 2                     | N.º de alumnos | El número total de alumnos matriculados en la escuela. |

## Seguridad espacial de los peatones en las zonas escolares de Riad: un enfoque basado en datos

|                                                                |    |                                                                  |                                                                                                                                                                                                    |
|----------------------------------------------------------------|----|------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                | 3  | N.º de profesores                                                | El número total de profesores empleados en la escuela.                                                                                                                                             |
|                                                                | 4  | N.º de administradores                                           | El número total de personal administrativo de la escuela.                                                                                                                                          |
| <b>Características del distrito</b>                            | 5  | Longitud de las carreteras (km/superficie del distrito)          | Longitud total de las carreteras del distrito (km) por superficie del distrito (km <sup>2</sup> )                                                                                                  |
|                                                                | 6  | Densidad escolar (escuelas/área del distrito)                    | Número de escuelas por superficie del distrito (km <sup>2</sup> )                                                                                                                                  |
|                                                                | 7  | Área de uso comercial/mixto (%)                                  | Porcentaje del terreno del distrito destinado a usos comerciales o mixtos =<br>(Superficie total de uso comercial/mixto / Superficie total) x 100 %                                                |
|                                                                | 8  | Promedio de muertes/lesiones (EPDO/km <sup>2</sup> )             | Promedio de muertes y lesiones relacionadas con el tráfico en el distrito (años: 2018-2023) =<br>[12 x (número de muertes) + 5 x (número de heridos)] / Superficie del distrito                    |
|                                                                | 9  | Ingreso promedio (SAR/cápita al año)                             | Nivel de ingreso promedio de los residentes en el distrito.                                                                                                                                        |
|                                                                | 10 | Densidad de población (per cápita / km <sup>2</sup> )            | Número de residentes por unidad de superficie en el distrito =<br>Población (per cápita) / Superficie (km <sup>2</sup> )                                                                           |
|                                                                | 11 | Población de 6 a 18 años (%)                                     | Porcentaje de la población del distrito con edades comprendidas entre los 6 y los 18 años =<br>Población en el grupo de edad (6-18) años (per cápita) ÷ Población total del municipio (per cápita) |
|                                                                | 12 | Densidad de atropellos de peatones (Accidentes/km <sup>2</sup> ) | Número de atropellos de peatones por zona del distrito (años: 2018-2023) =<br>Accidentes con atropello de peatones / Área (km <sup>2</sup> )                                                       |
| <b>Características de la zona de seguridad escolar</b>         | 13 | Velocidad promedio en la zona de seguridad (km/hora)             | Velocidad promedio de los vehículos (km/hora) dentro de la zona de seguridad escolar.                                                                                                              |
|                                                                | 14 | Flujo promedio en la zona de seguridad (vehículos/hora)          | El volumen promedio de tráfico por hora dentro de la zona de seguridad escolar.                                                                                                                    |
|                                                                | 15 | Número de intersecciones                                         | Número de intersecciones dentro de la zona de seguridad escolar.                                                                                                                                   |
|                                                                | 16 | Número de paradas de transporte público                          | Número de paradas de transporte público dentro de la zona de seguridad escolar.                                                                                                                    |
| <b>Características de las calles de los bloques adyacentes</b> | 17 | Número promedio de carriles                                      | El número promedio de carriles en las calles adyacentes a la escuela.                                                                                                                              |
|                                                                | 18 | Velocidad media (km/hora)                                        | Velocidad media de los vehículos en las calles adyacentes a la escuela.                                                                                                                            |

Fuente: Elaboración propia, 2025.

Esta investigación implementa el modelo MLR para realizar un análisis detallado de la densidad de atropellos de peatones (número de atropellos por km<sup>2</sup> de superficie del distrito) en las zonas escolares de Riad, Arabia Saudita. Este modelo de investigación utiliza métodos cuantitativos para examinar las relaciones entre las características de la carretera, los atributos

demográficos y los elementos de infraestructura, con el fin de determinar sus efectos en los incidentes de tráfico relacionados con los peatones.

La técnica de modelización estadística MLR permite a los investigadores establecer relaciones entre variables dependientes y múltiples variables independientes, según Kutner et al. (2005). La técnica demuestra una eficacia excepcional a la hora de determinar cómo múltiples factores afectan a los resultados, lo que la hace adecuada para la investigación sobre seguridad vial que implica diversos elementos ambientales, infraestructurales y demográficos que provocan accidentes (Washington et al., 2020). El MLR proporciona una gran interpretabilidad, junto con flexibilidad y capacidades de manejo de predictores continuos y categóricos, lo que permite a los investigadores evaluar los efectos variables sobre la densidad de accidentes con atropello de peatones a través de un marco estadístico claro (Washington et al., 2020). Las capacidades de inferencia estadística del MLR permiten a los investigadores determinar qué factores son estadísticamente significativos y cómo afectan a la frecuencia de los accidentes (Montgomery et al., 2021). Este estudio utiliza el MLR para analizar las relaciones entre diversas características y variables con la densidad de accidentes de peatones, ya que este enfoque ofrece un modelo explicable que orienta las recomendaciones políticas y las estrategias de planificación urbana para mejorar la seguridad en las zonas escolares.

Matemáticamente, un modelo MLR se expresa como (Montgomery, 2017; Alomari et al., 2016):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (1)$$

Donde:

- **Y** es la variable dependiente (densidad de atropellos de peatones en este estudio).
- **X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>k</sub>** son las variables independientes (predictoras), que representan diversos factores que influyen en los atropellos de peatones.
- **β** es la intersección, que representa la densidad de atropellos de peatones de referencia cuando todas las variables predictoras son cero.
- **β<sub>1</sub>, β<sub>2</sub>, ..., β<sub>k</sub>** son los coeficientes de regresión, que indican el cambio esperado en **Y** ante un cambio de una unidad en la variable correspondiente **X<sub>i</sub>**, manteniendo constantes todas las demás variables.
- **ε** es el término de error, que captura los factores no observados y la variabilidad aleatoria en los datos.

El primer paso del MLR requirió un análisis del factor de inflación de la varianza (VIF) para comprobar la multicolinealidad entre los predictores. El análisis eliminó todas las variables que mostraban valores de VIF superiores a 10 para reducir la multicolinealidad. La evaluación de las variables restantes se realizó mediante dos criterios, que incluían la significación estadística en  $p < 0,05$  y su capacidad para mejorar la interpretabilidad del modelo y su influencia en los valores del criterio de información de Akaike (AIC) y del criterio de información bayesiano (BIC). El proceso de selección del modelo consistió en eliminar los predictores no significativos o redundantes hasta que el ajuste del modelo ya no mejorara con eliminaciones adicionales, lo que dio como resultado la retención de solo predictores independientes y significativos. El proceso de optimización mejoró la estabilidad y la reproducibilidad y redujo la complejidad del modelo MLR.

Los resultados de la regresión lineal múltiple (MLR) recibieron un apoyo adicional mediante la implementación de modelos de regresión de aprendizaje automático (ML), que incluían los modelos K vecinos más cercanos (KNN), árbol de decisión y Bosque aleatorio. El método KNN utiliza valores observados de las k zonas escolares más similares en el espacio de características para predecir la densidad de accidentes (James et al., 2021). Los Árboles de Decisión utilizan umbrales de características para dividir los datos en grupos homogéneos más pequeños, lo que conduce a predicciones de resultados (Mienye y Jere, 2024).



Para los modelos de aprendizaje automático, se consideraron inicialmente todos los predictores disponibles. El proceso de desarrollo del modelo incluyó un procedimiento de validación cruzada de 10 pliegues, lo que ayudó a lograr un rendimiento sólido y a evitar el sobreajuste tanto durante el desarrollo del modelo como durante la optimización de los hiperparámetros. El conjunto de datos recibió una partición aleatoria que lo dividió en diez partes iguales para cada iteración, con nueve partes dedicadas al entrenamiento y una parte utilizada para la validación. El experimento se realizó diez veces y se combinaron los resultados de rendimiento de cada pliegue para obtener una medición imparcial de la precisión de la predicción. El marco de validación cruzada empleó la búsqueda por cuadrícula para determinar el conjunto óptimo de hiperparámetros (por ejemplo, el número  $e$  e de vecinos  $k$  en KNN, la profundidad máxima del árbol y el número de estimadores en Bosque aleatorio).

El modelo Bosque aleatorio sirvió como modelo principal de aprendizaje automático, ya que proporcionó los mejores resultados predictivos según Probst et al. (2020). El regresor Bosque aleatorio generó clasificaciones de importancia de las características, lo que ayudó a identificar las variables más críticas, al tiempo que se mantenía la interpretabilidad y se lograban buenos resultados predictivos y un cálculo eficiente.

Bosque aleatorio crea múltiples árboles de decisión que se entrenan a partir de muestras de datos obtenidas mediante *bootstrap* y subconjuntos aleatorios de variables predictoras. El método de promedio en conjunto produce una predicción estable al combinar las predicciones de todos los árboles, lo que conduce a una mayor precisión y a una reducción del sobreajuste. La predicción de Bosque aleatorio ( $\hat{Y}$ ) puede expresarse como (Pedregosa et al., 2011):

$$\hat{Y} = \frac{1}{B} \sum_{b=1}^B f_b(X) \quad (2)$$

El resultado de la predicción del árbol de decisión  $b^{\text{ésimo}}$  se denota por  $f_b(X)$ , mientras que  $B$  representa el número total de árboles en el conjunto. Bosque aleatorio destaca en la detección de patrones complejos y no lineales entre las variables, lo cual resulta fundamental para comprender los riesgos de atropello de peatones en diferentes entornos urbanos. El estudio seleccionó Bosque aleatorio como método principal de aprendizaje automático debido a su alta capacidad predictiva, su habilidad para proporcionar resultados interpretables a través de medidas de importancia de las variables y su resistencia al ruido.

#### 4. Análisis y resultados

El análisis de la densidad de atropellos de peatones en las zonas escolares de Riad se basa en el modelado MLR y ML en esta sección. El análisis comienza con el análisis exploratorio de datos (EDA) para presentar estadísticas esenciales, al tiempo que muestra patrones de distribución y detecta relaciones entre variables. Los resultados del análisis guían la selección de variables y métodos de transformación para el modelado predictivo. El estudio implementa el MLR para establecer relaciones entre las características de las carreteras, la demografía, las características de las zonas escolares y las condiciones del tráfico que afectan a la frecuencia de los atropellos de peatones.

En la siguiente sección se evalúa el rendimiento del modelo y la significación de las variables, utilizando el análisis del factor de inflación de la varianza (VIF) para abordar la multicolinealidad y garantizar la validez estadística. El proceso de optimización implica seleccionar de nuevo las variables y eliminar los elementos redundantes para mejorar tanto la precisión predictiva como la eficiencia. La comparación entre el modelo MLR original y su versión optimizada demuestra una mayor robustez del modelo, junto con una mejor interpretabilidad.

El rendimiento predictivo del modelo MLR se valida mediante técnicas de visualización y métricas estadísticas como el error absoluto medio (MAE), el error cuadrático medio (RMSE) y el R cuadrado ( $R^2$ ) para comparar las densidades de accidentes reales con las previstas. La información obtenida permite identificar las zonas escolares peligrosas, lo que respalda el

desarrollo de políticas basadas en datos empíricos para medidas específicas de seguridad peatonal.

#### 4.1 Análisis exploratorio de datos (EDA)

El análisis exploratorio de datos (EDA) incluyó resúmenes estadísticos completos de las características de los datos, análisis de distribución e identificación de correlaciones para variables clave. Los resultados del EDA condujeron a la selección de variables adecuadas para la ingeniería de características mediante la identificación de predictores relevantes para la predicción de la densidad de atropellos de peatones.

Este método permitió la selección de predictores importantes, lo que mejoró la precisión y la eficacia de los modelos resultantes para apoyar la toma de decisiones proactivas para intervenciones específicas de seguridad vial. La tabla 2 contiene estadísticas resumidas que describen las principales características del conjunto de datos (884 zonas escolares), lo que proporciona información crítica sobre las tendencias centrales, la variabilidad y la distribución de las variables utilizadas en el modelo predictivo.

**Tabla 2.** Estadísticas resumidas.

|    |                                                                                                         | Media    | Desviación | Mín.     | 0,25    | 0,5    | 0,75    | Máx.      |
|----|---------------------------------------------------------------------------------------------------------|----------|------------|----------|---------|--------|---------|-----------|
| 1  | Número de clases por escuela                                                                            | 25,10    | 16,0769    | 1,0      | 13      | 20     | 33      | 99,0      |
| 2  | Número de alumnos por escuela                                                                           | 610,0    | 519,5201   | 0        | 235     | 457,5  | 854,25  | 2960,0    |
| 3  | Número de profesores por escuela                                                                        | 43,97    | 29,9148    | 0        | 23      | 33,5   | 58      | 174,0     |
| 4  | Número de administradores por escuela                                                                   | 15,07    | 18,2166    | 0        | 2       | 6      | 20      | 80,0      |
| 5  | Longitud de las carreteras del distrito (km/superficie del distrito)                                    | 22,50    | 7,2965     | 2,36     | 19,9    | 22,92  | 25,32   | 56,68     |
| 6  | Densidad escolar del distrito (escuelas/área del distrito en km <sup>2</sup> )                          | 7,24     | 4,3320     | 0,0666   | 3,12    | 7,68   | 10,59   | 22,62     |
| 7  | Área de uso comercial mixto del distrito (%)                                                            | 19,21    | 11,8551    | 0,0005   | 8,80    | 18,56  | 28,1    | 84,31     |
| 8  | Promedio de muertes/lesiones por distrito (EPDO/superficie del distrito en km <sup>2</sup> )            | 21,89    | 12,5263    | 0        | 14,655  | 19,58  | 29,7    | 91,12     |
| 9  | Ingreso promedio del distrito (SAR/per cápita al año)                                                   | 91 461,9 | 24 337,81  | 24 000,0 | 81 600  | 93 600 | 104 400 | 248 400,0 |
| 10 | Densidad de población del distrito (Habitantes/Superficie del distrito en km <sup>2</sup> )             | 10719,7  | 7685,17    | 31,26    | 6232,74 | 9727,7 | 3025,54 | 49093,2   |
| 11 | Población del distrito de 6 a 18 años (%)                                                               | 17,79    | 8,0170     | 1,77     | 13,04   | 17,61  | 21,14   | 76,57     |
| 12 | Densidad de atropellos de peatones por distrito (accidentes/km <sup>2</sup> de superficie del distrito) | 1,81     | 1,5082     | 0        | 0,79    | 1,44   | 2,29    | 11,66     |
| 13 | Velocidad media del búfer (km/hora)                                                                     | 43,189   | 2,3989     | 40,0     | 41,50   | 42,63  | 44,38   | 55,17     |

## Seguridad espacial de los peatones en las zonas escolares de Riad: un enfoque basado en datos

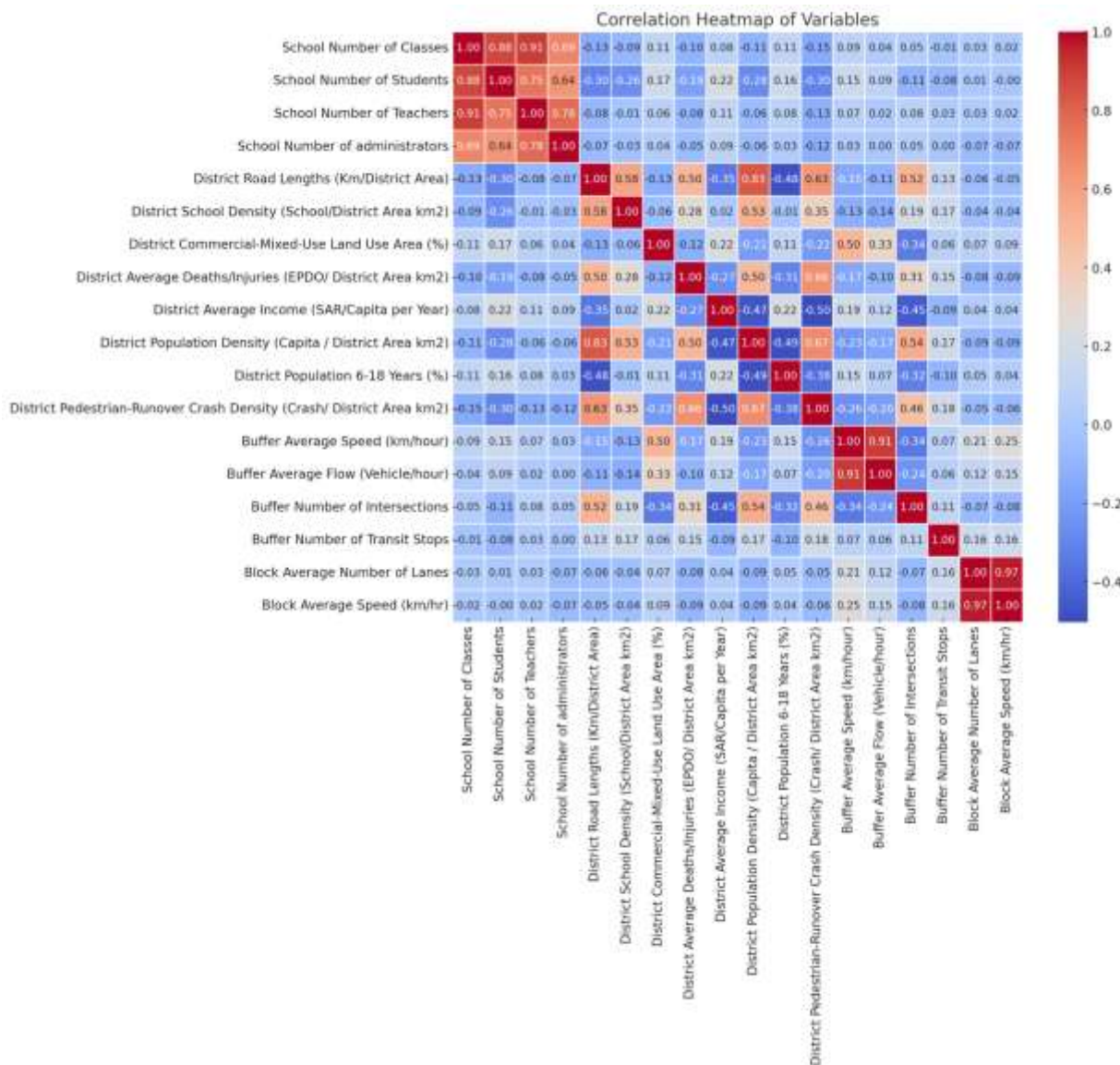
|    |                                                   |         |         |       |        |         |         |         |
|----|---------------------------------------------------|---------|---------|-------|--------|---------|---------|---------|
| 14 | Flujo promedio del búfer (vehículos/hora)         | 1619,09 | 1329,06 | 500,0 | 595,59 | 1278,79 | 1867,95 | 8730,41 |
| 15 | Número de intersecciones del búfer                | 41,344  | 15,9936 | 8,0   | 32     | 39      | 47      | 138,0   |
| 16 | Número de paradas de transporte público del búfer | 1,068   | 1,4324  | 0     | 0      | 0       | 2       | 7,0     |
| 17 | Número promedio de carriles por bloque            | 2,300   | 0,6053  | 2,0   | 2      | 2       | 2,5     | 8,0     |
| 18 | Velocidad media por bloque (km/h)                 | 41,711  | 3,6713  | 40,0  | 40     | 40      | 42,42   | 80      |

Fuente: Elaboración propia, 2025.

La tabla de estadísticas resumidas revela información fundamental sobre las características de las zonas escolares y sus entornos circundantes en Riad. En particular, la densidad de atropellos de peatones varía significativamente entre los distritos, oscilando entre 0 y aproximadamente 11,664 atropellos/km<sup>2</sup>, con una media de alrededor de 1,810 atropellos/km<sup>2</sup>, lo que indica una variabilidad sustancial en la frecuencia de los atropellos. Del mismo modo, el indicador de gravedad (EPDO) muestra una variación considerable, con una media de alrededor de 21,889 EPDO/km<sup>2</sup> y un máximo que alcanza los 91,120, lo que refleja una alta variabilidad en la gravedad de los atropellos de peatones en las zonas escolares. Además, los datos ponen de relieve diferencias sustanciales en los atributos relacionados con las escuelas (por ejemplo, número de alumnos, tamaño de las clases, número de profesores) y las condiciones del tráfico circundante (por ejemplo, flujo promedio, densidad de intersecciones). Estas variaciones observadas subrayan la importancia de intervenciones adaptadas y específicas al contexto para mejorar eficazmente la seguridad de los peatones en los alrededores de las escuelas en diversos entornos urbanos.

El análisis de correlación es un componente crítico del preprocesamiento de datos que mejora la modelización predictiva al identificar las relaciones entre las variables que influyen en la seguridad de los peatones y el tráfico en las zonas escolares. El mapa de calor de correlación que se muestra a continuación (Figura 2) ilustra las relaciones entre las variables numéricas clave del conjunto de datos, lo que proporciona información sobre los factores que influyen de manera más significativa en la densidad y la gravedad de los atropellos de peatones (EPDO) en las zonas escolares.

Figura 2. Mapa de calor de la matriz de correlación para todos los indicadores.



Fuente: Elaboración propia, 2025.

El mapa de calor de correlación proporciona información valiosa sobre las relaciones entre las diferentes variables que afectan a la densidad de atropellos de peatones en las zonas escolares de Riad. Una de las observaciones más significativas es la fuerte correlación positiva entre la densidad de atropellos de peatones en el distrito y la media de muertes/lesiones en el distrito (EPDO por km<sup>2</sup>) ( $r \approx 0,66$ ). Esto sugiere que las zonas con una alta densidad de atropellos de peatones también experimentan incidentes de tráfico más graves. Además, la densidad de atropellos de peatones muestra una notable correlación positiva con la densidad de población del distrito ( $r \approx 0,67$ ) y la densidad de escuelas del distrito ( $r \approx 0,35$ ), lo que indica que las zonas muy pobladas y aquellas con más escuelas por unidad de superficie tienden a tener una mayor frecuencia de atropellos de peatones. Esto es de esperar, ya que el aumento del tráfico peatonal, en particular de estudiantes y jóvenes peatones, aumenta naturalmente la probabilidad de accidentes relacionados con peatones. Por último, el número de intersecciones ( $r \approx 0,46$ ) y la longitud de las carreteras del distrito ( $r \approx 0,63$ ) se correlacionan positivamente con los atropellos de peatones, lo que significa que las zonas con más intersecciones y carreteras tienden a registrar más atropellos de peatones. Esto es lógico, debido a que las intersecciones y las carreteras

aumentan la exposición de los peatones a los vehículos, especialmente si los pasos de peatones y las medidas de seguridad son inadecuados.

Por otro lado, hay algunas correlaciones negativas dignas de mención. Por ejemplo, la renta media del distrito tiene una correlación negativa moderada con la densidad de atropellos de peatones ( $r \approx -0,50$ ), lo que implica que las zonas más ricas tienden a sufrir menos accidentes de peatones. Esto podría atribuirse a la mejora de las infraestructuras, las instalaciones para peatones o los comportamientos al volante. Otra correlación negativa interesante es la que existe entre la velocidad media por manzana y la densidad de escuelas del distrito ( $r \approx -0,26$ ), lo que sugiere que las zonas con mayor concentración de escuelas tienden a tener velocidades medias más bajas, probablemente debido a las medidas de pacificación del tráfico o al aumento de la congestión.

Además, el número promedio de carriles por manzana y la velocidad media ( $r \approx -0,05$  y  $-0,06$ , correlación débil) no parecen tener un impacto sustancial en los atropellos de peatones, lo que podría indicar que el número de carriles por sí solo no es un factor determinante para la seguridad de los peatones, sino más bien una combinación del ancho de los carriles, el volumen de tráfico y otras infraestructuras peatonales. Estos datos pueden informar a los urbanistas y responsables políticos sobre los principales factores de riesgo y orientar la aplicación de medidas específicas de seguridad peatonal.

#### 4.2 Modelo de regresión lineal múltiple (MLR)

Para predecir la densidad de atropellos de peatones en las zonas escolares de Riad, se desarrolló un modelo MLR. El modelo examina cómo diversos factores, como las características de la red viaria, la densidad de escuelas, los patrones de uso del suelo y las condiciones del tráfico, contribuyen a la frecuencia de los atropellos de peatones. La variable dependiente en el análisis es la densidad de atropellos de peatones (accidentes por km<sup>2</sup> de superficie del distrito), mientras que las variables independientes incluyen atributos relacionados con las escuelas, características a nivel de distrito, características de las zonas de seguridad escolares y factores a nivel de manzana adyacente a las escuelas. Este enfoque proporciona un modelo de referencia para comprender los riesgos de accidentes de peatones y sirve de base para modelos predictivos más avanzados.

El modelo MLR demuestra un gran poder explicativo, con un valor del coeficiente de determinación (R cuadrado) de 0,653, lo que indica que el 65,3 % de la varianza en la densidad de atropellos de peatones puede explicarse mediante las variables independientes seleccionadas. El valor R cuadrado ajustado de 0,646 confirma aún más que los predictores contribuyen de manera significativa a explicar la variable dependiente. Además, el modelo arroja una estadística F de 95,68 ( $p < 0,001$ ), lo que confirma que el modelo general es estadísticamente significativo. La tabla 3 resume las métricas de rendimiento del modelo.

**Tabla 3.** Rendimiento y ajuste del modelo.

| Métrica | R cuadrado | R cuadrado ajustado | Estadística F | Probabilidad (estadística F) | Log-verosimilitud | AIC    | BIC    |
|---------|------------|---------------------|---------------|------------------------------|-------------------|--------|--------|
| Valor   | 0,653      | 0,646               | 95,68         | <0,001                       | -1149,8           | 2336,0 | 2422,0 |

Fuente: Elaboración propia, 2025.

Las variables independientes del modelo tienen distintos grados de importancia a la hora de predecir la densidad de atropellos de peatones. La tabla 4 presenta los valores de los coeficientes, la importancia estadística (valores p) y los intervalos de confianza de todas las variables.



Tabla 4. Variables y su influencia.

|    | Variable                                                                     | Coeficiente | Valor p          | IC inferior del 95 % | IC superior del 95 % | Interpretación                                                                                                             |
|----|------------------------------------------------------------------------------|-------------|------------------|----------------------|----------------------|----------------------------------------------------------------------------------------------------------------------------|
| 1  | Constante                                                                    | 1,7867      | 0,291            | -1,533               | 5,107                | La densidad de colisiones de referencia cuando todos los predictores son cero.                                             |
| 2  | Número de clases por escuela                                                 | 0,0115      | 0,088            | -0,002               | 0,025                | Un mayor número de clases puede aumentar ligeramente los atropellos de peatones, aunque no de forma muy significativa.     |
| 3  | Número de alumnos por escuela                                                | -0,0003     | <b>0,041</b>     | -0,001               | -1,29e-05            | Un mayor número de alumnos se asocia con una densidad de accidentes ligeramente inferior.                                  |
| 4  | Número de profesores por escuela                                             | -0,0022     | 0,463            | -0,008               | 0,004                | Sin impacto significativo en los atropellos de peatones.                                                                   |
| 5  | Número de administradores escolares                                          | -0,0037     | 0,184            | -0,009               | 0,002                | Sin impacto significativo en los atropellos de peatones.                                                                   |
| 6  | Longitud de las carreteras del distrito (km/superficie del distrito)         | 0,0264      | <b>0,003</b>     | 0,009                | 0,043                | Una mayor longitud de carretera se asocia con una mayor densidad de atropellos de peatones.                                |
| 7  | Densidad escolar del distrito (Escuela/Área del distrito km <sup>2</sup> )   | 0,0170      | 0                | -0,004               | 0,038                | Un mayor número de escuelas por área podría aumentar la densidad de accidentes, pero no es estadísticamente significativo. |
| 8  | Área de uso comercial mixto del distrito (%)                                 | -0,0056     | 0,088            | -0,012               | 0,001                | Un mayor uso comercial del suelo puede reducir ligeramente la densidad de accidentes.                                      |
| 9  | Promedio de muertes/lesiones por distrito (EPDO por km <sup>2</sup> )        | 0,0467      | <b>&lt;0,001</b> | 0,041                | 0,052                | Una mayor gravedad de los accidentes se correlaciona con un mayor número de atropellos de peatones.                        |
| 10 | Ingreso promedio del distrito (SAR/cápita al año)                            | -1,359e-05  | <b>&lt;0,001</b> | -1,69e-05            | -1,03e-05            | Las zonas con ingresos más altos tienden a tener menos accidentes con peatones.                                            |
| 11 | Densidad de población del distrito (habitantes/km <sup>2</sup> del distrito) | 3,287e-05   | <b>&lt;0,001</b> | 1,65e-05             | 4,92e-05             | Una mayor densidad de población aumenta los atropellos de peatones.                                                        |
| 12 | Población del distrito de 6 a 18 años (%)                                    | -0,0091     | 0,066            | -0,019               | 0,001                | Sin impacto significativo en los atropellos de peatones.                                                                   |
| 13 | Velocidad promedio del búfer(km/hora)                                        | 0,0158      | 0,678            | -0,059               | 0,091                | La velocidad no influye significativamente en la densidad de atropellos de peatones.                                       |

|    |                                                     |            |       |        |          |                                                                                                           |
|----|-----------------------------------------------------|------------|-------|--------|----------|-----------------------------------------------------------------------------------------------------------|
| 14 | Flujo promedio del búfer (vehículos/hora)           | -8,636e-05 | 0     | -0,000 | 3,25e-05 | El flujo de tráfico no tiene un impacto significativo en la densidad de accidentes de peatones.           |
| 15 | Número de intersecciones en el búfer                | 0,0014     | 0,576 | -0,004 | 0,007    | Un mayor número de intersecciones no influye significativamente en la densidad de atropellos de peatones. |
| 16 | Número de paradas de transporte público en el búfer | 0,0438     | 0     | -0     | 0,088    | Un mayor número de paradas de transporte público aumenta ligeramente los atropellos de peatones.          |
| 17 | Número promedio de carriles por bloque              | 0,2758     | 0,168 | -0,117 | 0,669    | Sin impacto significativo en los atropellos de peatones.                                                  |
| 18 | Velocidad promedio por bloque (km/h)                | -0,0429    | 0,199 | -0,108 | 0,023    | Sin impacto significativo en los atropellos de peatones.                                                  |

Fuente: Elaboración propia, 2025.

Los resultados indican que varios factores influyen significativamente en la densidad de atropellos de peatones en las zonas escolares. Los predictores más impactantes ( $p < 0,05$ ) incluyen el promedio de muertes/lesiones del distrito, la longitud de las carreteras del distrito, la densidad de población del distrito y el ingreso promedio del distrito, donde una mayor gravedad de los accidentes, redes de carreteras más largas y poblaciones más densas contribuyen a una mayor densidad de accidentes, mientras que las áreas de mayores ingresos se asocian con menores riesgos de accidentes. Por el contrario, otras variables, como la velocidad de amortiguación, el número de carriles y el número de intersecciones, no muestran efectos estadísticamente significativos ( $p > 0,05$ ), lo que sugiere que estas características de la infraestructura pueden no determinar directamente los riesgos de accidentes con peatones. Las variables relacionadas con las escuelas muestran resultados dispares: el número de alumnos por escuela presenta una correlación ligeramente negativa con la densidad de accidentes, posiblemente debido a las medidas de seguridad reforzadas en las zonas con alta densidad de alumnos, mientras que el número de clases por escuela muestra una débil asociación positiva, lo que indica que las aulas más dispersas pueden aumentar la exposición de los peatones. Estos resultados proporcionan información valiosa sobre los factores clave que contribuyen al riesgo de accidentes de peatones y ponen de relieve la necesidad de intervenciones específicas para mejorar la seguridad en las zonas escolares.

Se evaluó la multicolinealidad utilizando el factor de inflación de la varianza (VIF) para determinar si las variables independientes estaban altamente correlacionadas, lo que podría afectar a la fiabilidad del modelo. Los resultados del VIF ponen de relieve que algunos predictores presentan problemas de multicolinealidad, en particular las variables relacionadas con las escuelas y los atributos de la red de carreteras. Los valores altos del VIF ( $>10$ ) sugieren una fuerte correlación entre estas variables, como se muestra en la tabla 5.

**Tabla 5.** Consideraciones y comprobación de la multicolinealidad.

| Variable                               | VIF          |
|----------------------------------------|--------------|
| Número de clases de la escuela         | <b>12,95</b> |
| Número de profesores de la escuela     | 8,62         |
| Número de alumnos por escuela          | 6,70         |
| Velocidad promedio en el búfer         | 9,12         |
| Flujo promedio en el búfer             | 7,10         |
| Número promedio de carriles por bloque | <b>16,07</b> |
| Velocidad media del bloque             | <b>16,43</b> |
| Todas las demás variables              | < 5,0        |

Fuente: Elaboración propia, 2025.

Dado que varios valores del factor de inflación de la varianza (VIF) superaban 5, lo que indicaba una posible multicolinealidad entre las variables independientes, se eliminaron sistemáticamente todas las variables con un VIF superior a 10 para mejorar la estabilidad y la fiabilidad del modelo. Este refinamiento garantiza que cada predictor contribuya de forma independiente al modelo de regresión, lo que reduce la redundancia y mejora la interpretabilidad. La tabla 6 presenta el resumen del modelo MLR optimizado, destacando los coeficientes refinados y la significación estadística de las variables restantes. Además, la tabla 7 ofrece un análisis comparativo del modelo MLR original frente al optimizado, lo que demuestra las mejoras en la eficiencia del modelo, la reducción de la multicolinealidad y la solidez predictiva general.

**Tabla 6.** Resumen del modelo MLR optimizado

|    | Variable                                                                       | Coeficiente | Error estándar | t      | P> t             | 0,025     | 0,975     |
|----|--------------------------------------------------------------------------------|-------------|----------------|--------|------------------|-----------|-----------|
| 1  | Constante                                                                      | 0,751       | 1,476          | 0,509  | 0,611            | -2,146    | 3,648     |
| 2  | Número de alumnos por escuela                                                  | -0,0001     | 0              | -1,143 | 0,253            | 0         | 8,35E-05  |
| 3  | Número de profesores por escuela                                               | 0,0017      | 0,002          | 0,824  | 0,41             | -0,002    | 0,006     |
| 4  | Número de administradores escolares                                            | -0,005      | 0,003          | -1,85  | 0,065            | -0,01     | 0         |
| 5  | Longitud de las carreteras del distrito (km/superficie del distrito)           | 0,0259      | 0,009          | 2,965  | <b>0,003</b>     | 0,009     | 0,043     |
| 6  | Densidad escolar del distrito (escuelas/área del distrito en km <sup>2</sup> ) | 0,0194      | 0,011          | 1,791  | 0,074            | -0,002    | 0,041     |
| 7  | Área de uso comercial mixto del distrito (%)                                   | -0,0054     | 0,003          | -1,67  | 0,095            | -0,012    | 0,001     |
| 8  | Promedio de muertes/lesiones por distrito (EPDO por km <sup>2</sup> )          | 0,0467      | 0,003          | 16,276 | <b>&lt;0,001</b> | 0,041     | 0,052     |
| 9  | Ingreso promedio del distrito (SAR/cápita al año)                              | -1,43E-05   | 1,62E-06       | -8,843 | <b>&lt;0,001</b> | -1,75E-05 | -1,11E-05 |
| 10 | Densidad de población del distrito (habitantes/km <sup>2</sup> del distrito)   | 3,30E-05    | 8,31E-06       | 3,97   | <b>&lt;0,001</b> | 1,67E-05  | 4,93E-05  |
| 11 | Población del distrito de 6 a 18 años (%)                                      | -0,009      | 0,005          | -1,806 | 0,071            | -0,019    | 0,001     |
| 12 | Velocidad promedio del búfer (km/hora)                                         | 0,0147      | 0,037          | 0,4    | 0,689            | -0,057    | 0,087     |
| 13 | Flujo promedio en el búfer (vehículos/hora)                                    | -8,90E-05   | 5,93E-05       | -1,501 | 0,134            | 0         | 2,74E-05  |
| 14 | Número de intersecciones en el búfer                                           | 0,0016      | 0,003          | 0,629  | 0,53             | -0,003    | 0,007     |
| 15 | Número de paradas de transporte público en el búfer                            | 0,0442      | 0,022          | 1,996  | <b>0,046</b>     | 0,001     | 0,088     |

Fuente: Elaboración propia, 2025.

**Tabla 7.** Comparación entre el modelo MLR original y el optimizado.

| Métrica                             | Modelo original | Modelo optimizado |
|-------------------------------------|-----------------|-------------------|
| <b>R cuadrado</b>                   | 0,65255507      | 0,65061962        |
| <b>R cuadrado ajustado</b>          | 0,64573456      | 0,64499093        |
| <b>Estadística F</b>                | 95,6753733      | 115,590029        |
| <b>Probabilidad (estadística F)</b> | 2,99E-185       | 2,74E-187         |
| <b>AIC</b>                          | 2335,65255      | 2334,56324        |
| <b>BIC</b>                          | 2421,77277      | 2406,3301         |

Fuente: Elaboración propia, 2025.

El modelo MLR optimizado mantiene una gran capacidad predictiva al tiempo que aborda los problemas de multicolinealidad. El valor R cuadrado permanece prácticamente sin cambios ( $0,653 \rightarrow 0,651$ ), lo que indica que el modelo sigue explicando aproximadamente el 65 % de la varianza en la densidad de atropellos de peatones. Del mismo modo, el valor R cuadrado ajustado se mantiene estable ( $0,646 \rightarrow 0,645$ ), lo que confirma que la eliminación de las variables con un VIF elevado no ha afectado significativamente a la capacidad explicativa del modelo. Esta estabilidad sugiere que el modelo perfeccionado sigue captando las relaciones clave entre las variables independientes y los atropellos de peatones, al tiempo que elimina la redundancia.

En términos de eficiencia del modelo, la estadística F aumentó de 95,68 a 115,59, lo que indica una significación estadística más sustancial y una mayor eficiencia en la explicación de la variable dependiente. Además, el criterio de información de Akaike (AIC) disminuyó ligeramente ( $2335,65 \rightarrow 2334,56$ ), lo que sugiere que el modelo optimizado logra un mejor rendimiento con menos predictores, lo que lo hace más parsimonioso. Cabe destacar que el valor p de la estadística F sigue siendo muy significativo ( $p < 0,001$ ), lo que refuerza la fiabilidad y solidez generales del modelo optimizado.

La mejora más significativa se produce gracias a la reducción de la multicolinealidad. Los valores del factor de inflación de la varianza (VIF) son ahora inferiores a 5, lo que confirma que se han mitigado con éxito los problemas de multicolinealidad. La mejora refuerza tanto la estabilidad como la fiabilidad de los coeficientes estimados, ya que cada predictor contribuye ahora de forma independiente al modelo sin una correlación excesiva. El modelo se vuelve más interpretable y estadísticamente válido tras eliminar el número de clases de la escuela, el número promedio de carriles por bloque y la velocidad promedio por bloque, lo que da como resultado una herramienta más fiable para identificar las zonas de alto riesgo de atropellos de peatones.

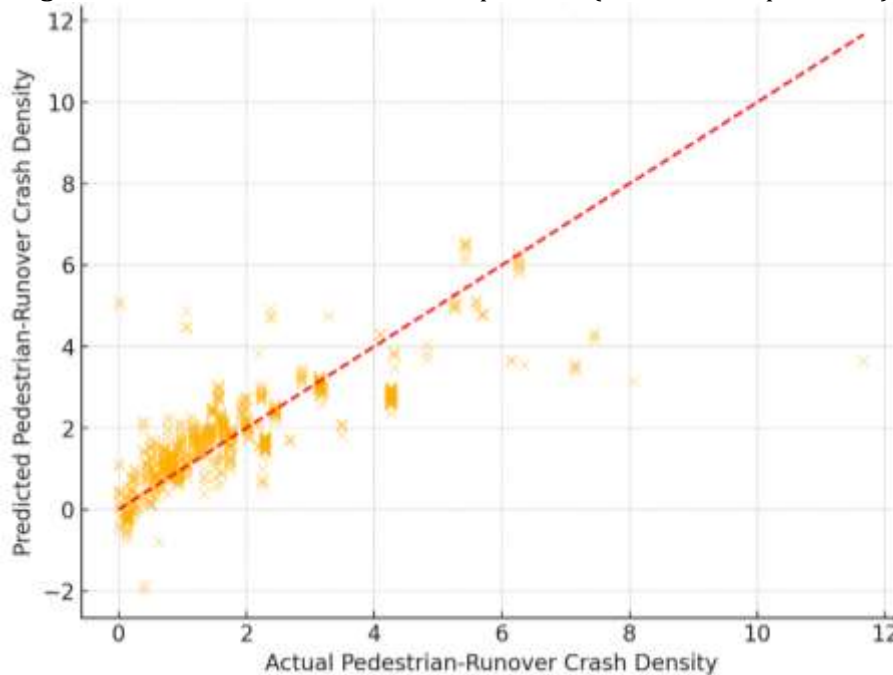
#### **4.2.1. Evaluación y validación del modelo MLR optimizado**

La evaluación del rendimiento del modelo de regresión requiere una comparación entre los valores reales y los valores previstos. La comparación entre los valores reales y los valores previstos ayuda a determinar la capacidad del modelo MLR optimizado para generalizar las predicciones de la densidad de atropellos de peatones en el mundo real. El modelo demuestra una comprensión eficaz de las variables independientes y los atropellos de peatones cuando los valores reales coinciden con los valores previstos. El modelo demuestra un sobreajuste, un subajuste o la falta de variables predictivas que afectan a la densidad de atropellos de peatones cuando hay desviaciones significativas entre los valores reales y los previstos. La visualización de los valores reales frente a los previstos permite una evaluación cuantitativa e intuitiva de la fiabilidad y la precisión del modelo, lo que orienta las futuras mejoras del modelo y su aplicación en el mundo real.

El diagrama de dispersión de la densidad real frente a la prevista de atropellos de peatones (Figura 3) representa visualmente la relación entre las densidades de atropellos observadas y las predicciones generadas por el modelo. Idealmente, si el modelo fuera perfecto, todos los puntos de datos se situarían exactamente en la línea de referencia roja discontinua 1:1, lo que significaría

que los valores previstos coincidirían perfectamente con los valores reales. En este gráfico, la mayoría de los puntos de datos se agrupan muy cerca de esta línea ideal, lo que demuestra que el modelo realiza predicciones razonables y precisas. Sin embargo, se observan algunas desviaciones, especialmente en las densidades de accidentes más altas, lo que indica que, aunque el modelo capta eficazmente las tendencias generales, puede tener algunas limitaciones en casos extremos. Esta información ayuda a evaluar la coherencia de las predicciones del modelo en diferentes distritos, lo que refuerza su utilidad a la hora de priorizar las zonas de alto riesgo de atropello de peatones.

**Figura 3.** Valores reales frente a valores previstos (modelo MLR optimizado).



Fuente: Elaboración propia, 2025.

Las métricas de evaluación del modelo optimizado cuantifican aún más el rendimiento predictivo del modelo, como se muestra en la tabla 8. El error absoluto medio (MAE) de 0,59 indica que, en promedio, la densidad de accidentes con atropello de peatones prevista se desvía de los valores reales en aproximadamente 0,59 accidentes por km<sup>2</sup>, lo que proporciona una medida del error fácil de interpretar. El error cuadrático medio (MSE) de 0,79 enfatiza el impacto de los errores más grandes al elevarlos al cuadrado, mientras que el error cuadrático medio raíz (RMSE) de 0,89 convierte este error al cuadrado de nuevo a la escala original de la variable dependiente, lo que lo convierte en un punto de referencia útil para comprender la magnitud del error esperado. El valor R cuadrado ( $R^2$ ) de 0,651 sugiere que el modelo explica aproximadamente el 65 % de la variabilidad en la densidad de atropellos de peatones, lo que demuestra una gran capacidad predictiva, aunque deja margen para futuras mejoras.

**Tabla 8.** Métricas de evaluación del modelo optimizado.

| Métrica                       | Valor  |
|-------------------------------|--------|
| Error absoluto medio (MAE)    | 0,5945 |
| Error cuadrático medio (MSE)  | 0,7938 |
| Error cuadrático medio (RMSE) | 0,8909 |
| R cuadrado ( $R^2$ )          | 0,6506 |

Fuente: Elaboración propia, 2025.

El modelo MLR optimizado funciona correctamente como sistema implementado, lo que ha demostrado su precisión predictiva mediante análisis visuales y evaluaciones de rendimiento. El



diagrama de dispersión «Real frente a previsto» muestra que el modelo se ajusta a las densidades reales de accidentes de peatones, ya que se agrupa alrededor de la línea de referencia 1:1, lo que indica una gran precisión predictiva. El modelo presenta desviaciones menores en niveles de densidad de accidentes más altos, pero mantiene su capacidad para detectar zonas escolares peligrosas. Las métricas de rendimiento demuestran la eficacia del modelo a través de su MAE (0,59), que indica un error medio bajo, y su RMSE (0,89), que proporciona una medida de error comprensible. El valor  $R^2$  de 0,651 muestra que el modelo explica una cantidad significativa de varianza en la densidad de accidentes de peatones, lo que lo hace útil para los planificadores urbanos y los analistas de seguridad vial.

#### **4.2.2 Discusión**

Los resultados de este estudio coinciden con las investigaciones internacionales sobre lo que hace que las zonas peatonales sean más seguras. Este análisis respalda a Rothman et al. (2015) en Toronto e Ivan et al. (2019) en Bucarest, ya que muestra que la densidad de atropellos de peatones depende de la densidad de población y de la infraestructura viaria, incluida la longitud de las carreteras y la velocidad media. La investigación confirma que el aumento del riesgo de atropellos de peatones se da en zonas urbanas con alta densidad poblacional y extensas redes de carreteras. El estudio confirma los efectos socioeconómicos sobre la seguridad de los peatones, lo que coincide con las conclusiones de Rothman et al. (2017b) sobre cómo las desigualdades socioeconómicas determinan las tasas de lesiones de peatones en edad escolar en Toronto. La relación constante entre los niveles de ingresos y los resultados en materia de seguridad de los peatones demuestra por qué es necesario incorporar los factores socioeconómicos en los planes de mejora de la seguridad en las zonas escolares.

El presente estudio difiere de investigaciones anteriores en lo que respecta a los efectos específicos de las características de las carreteras. El estudio no encontró significación estadística en Riad en cuanto a los efectos de la velocidad del tráfico y la geometría de las carreteras en los atropellos de peatones, aunque Zhao et al. (2015) y Forward et al. (2025) demostraron su importancia en sus investigaciones. Las diferencias entre esta investigación y los estudios previos probablemente se deben a las normas de planificación urbana, los estándares de infraestructura, los métodos de control del tráfico y las políticas locales de planificación urbanas específicas de Riad. Los estudios realizados por Yu y Zhu (2016) y Bahrami et al. (2024) hacen hincapié en la congestión del tráfico y la complejidad de las intersecciones como predictores esenciales, pero este estudio demostró que estos factores no afectaban a la densidad de atropellos de peatones en Riad. Los resultados demuestran la necesidad de desarrollar intervenciones específicas para cada lugar, en lugar de utilizar soluciones genéricas procedentes de otras zonas.

El modelo MLR optimizado establecido en esta investigación produce resultados sólidos e interpretables que coinciden con estudios anteriores. El modelo respalda estrategias políticas para mejorar las instalaciones peatonales, la aplicación de los límites de velocidad y los programas de mejora de los cruces que reducen eficazmente los riesgos de accidentes en las zonas escolares.

#### **4.3 Análisis de aprendizaje automático**

En esta sección, proponemos un modelo de aprendizaje automático (ML) basado en la regresión para predecir la densidad continua de accidentes de peatones en cada zona escolar, en lugar de agruparlos en niveles de riesgo generales. De esta manera, las autoridades pueden asignar prioridades a las zonas de riesgo según la magnitud del mismo y, a continuación, asignar sus recursos de seguridad (Shuai y Kwon, 2025). El uso de este tipo de modelos predictivos ha sido respaldado por las últimas investigaciones. Por ejemplo, un estudio a nivel nacional en los Estados Unidos empleó un modelo Bosque aleatorio para predecir el número de accidentes en determinados lugares con gran precisión y obtener información útil para intervenciones preventivas de seguridad vial (Yamarthi et al., 2025). Nuestro análisis sigue este enfoque

examinando varios modelos de regresión de aprendizaje automático para seleccionar el mejor para estimar la densidad de accidentes en zonas escolares.

Algunos de los algoritmos de regresión que se han utilizado ampliamente en la investigación sobre seguridad en el transporte incluyen el modelo K vecinos más cercanos (KNN), el árbol de decisión y Bosque aleatorio. Estos modelos van desde la simple predicción basada en instancias (KNN) hasta las reglas de decisión de un solo árbol y los métodos basados en conjuntos de árboles. Se eligió el método de conjunto Bosque aleatorio porque ha demostrado ser eficaz en el manejo de relaciones e interacciones no lineales en los datos y se ha comprobado que produce mejores predicciones que los modelos estadísticos tradicionales (Donnell et al., 2020). El conjunto de datos disponible de las zonas escolares se utilizó para el entrenamiento y la validación de los modelos, y cada zona se caracterizó por sus características (por ejemplo, volumen de tráfico, atributos de las carreteras y factores ambientales) y la densidad de accidentes de peatones observada. Se realizó un ajuste de hiperparámetros para cada modelo (por ejemplo, el número de vecinos  $k$  en KNN, la profundidad máxima del árbol, etc.) utilizando la validación cruzada para garantizar que los modelos se compararan de forma justa. El rendimiento de los modelos se evaluó utilizando el coeficiente de determinación ( $R^2$ ) y métricas de error como el error cuadrático medio en un conjunto de pruebas retenido. El modelo que obtuvo mejores resultados en el conjunto de pruebas —en este caso, el modelo Bosque aleatorio— fue elegido para un análisis más detallado de sus predicciones y los factores subyacentes.

#### 4.3.1 Métodos

Para la tarea de regresión, el conjunto de datos se estructuró de manera que cada entrada correspondiera a una zona escolar específica, descrita por diversas características predictivas y la variable objetivo de densidad de accidentes de peatones (por ejemplo, accidentes por unidad de superficie o alguna exposición estandarizada). Las variables predictivas clave incluían métricas de exposición al tráfico (como el volumen medio diario de tráfico), características geométricas de la carretera (número de carriles, presencia de pasos de peatones, límite de velocidad, etc.) y características del entorno circundante (uso del suelo, población estudiantil, etc.). Todas las características se normalizaron o codificaron adecuadamente para su uso en los modelos de aprendizaje automático. Dividimos los datos en conjuntos de entrenamiento y prueba para permitir una evaluación imparcial del rendimiento predictivo de cada modelo.

Entrenamos tres regresores de ML en los datos de entrenamiento utilizando (1) K vecinos más cercanos (KNN) para predecir la densidad de accidentes de zonas escolares similares en el espacio de características, (2) árbol de decisión para aprender una jerarquía de reglas «si-entonces» para la partición de datos, y (3) Bosque aleatorio como un conjunto de muchos árboles de decisión que votan sobre la predicción. Los hiperparámetros del modelo se optimizaron utilizando la búsqueda por cuadrícula y la validación cruzada en el conjunto de entrenamiento. Se ajustó el número de vecinos  $k$  en KNN, así como la profundidad y el número de árboles en Bosque aleatorio para equilibrar el sesgo y la varianza. También empleamos técnicas como la detención temprana o la poda (para los árboles) y reservamos una parte de los datos de entrenamiento para la validación, con el fin de evitar el sobreajuste. A continuación, se evaluó el rendimiento de cada modelo final en el conjunto de prueba utilizando  $R^2$  (que indica la proporción de varianza explicada en la densidad de accidentes) y métricas de error como el error absoluto medio (MAE) o el error cuadrático medio (RMSE).

La tabla 9 resume el rendimiento de los tres modelos de regresión en el conjunto de datos de prueba. El modelo Bosque aleatorio logró la mayor precisión con un  $R^2$  de aproximadamente 0,88 y el menor error de predicción ( $RMSE \approx 0,56$ ). El modelo árbol de decisión mostró un rendimiento moderado ( $R^2 \sim 0,68$ ), mientras que el modelo KNN quedó rezagado con un  $R^2$  de alrededor de 0,56. Esta clasificación es coherente con las expectativas y la bibliografía, ya que los métodos de árboles ensamblados tienden a superar a los modelos más simples en las tareas de predicción de accidentes (Donnell et al., 2020). Dado su rendimiento superior, centramos nuestros resultados posteriores en el modelo Bosque aleatorio.

**Tabla 9.** Rendimiento del conjunto de pruebas de diferentes modelos de regresión para predecir la densidad de accidentes de peatones en zonas escolares.

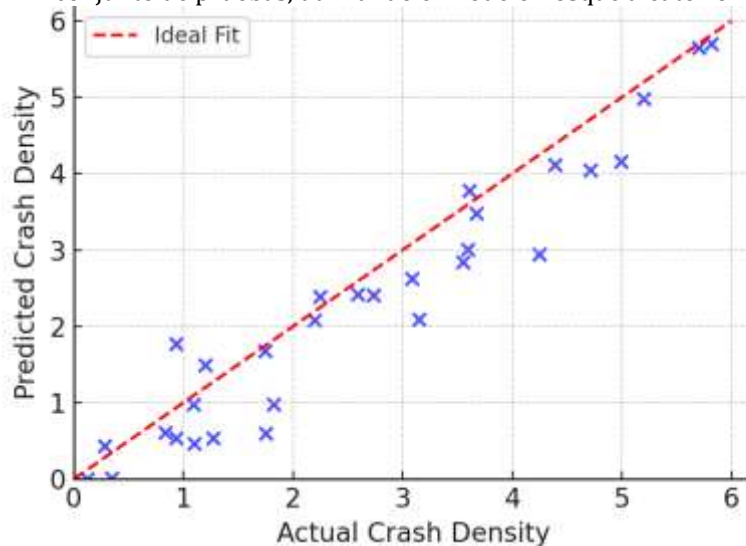
| Modelo de regresión    | R <sup>2</sup> (prueba) | RMSE (prueba) |
|------------------------|-------------------------|---------------|
| K vecinos más cercanos | 0,56                    | 1,01          |
| Árbol de decisión      | 0,68                    | 0,85          |
| Bosque aleatorio       | <b>0,88</b>             | <b>0,56</b>   |

Fuente: Elaboración propia, 2025.

#### 4.3.2 Resultados de la regresión

El modelo de regresión Bosque aleatorio, que presentó el mejor desempeño, se utilizó para predecir las densidades de accidentes de peatones en todas las zonas escolares del conjunto de pruebas (Figura 4). En general, las predicciones del modelo muestran una elevada concordancia con los valores observados. El coeficiente de determinación  $R^2 \approx 0,88$  sugiere que aproximadamente el 88 % de la variabilidad en la densidad de accidentes en las zonas escolares se explica por el modelo, una proporción notablemente alta para los datos de accidentes, que a menudo son ruidosos. En términos prácticos, este nivel de precisión significa que el modelo puede distinguir de forma fiable las zonas de mayor riesgo (con densidades de accidentes elevadas) de las de menor riesgo en un espectro continuo, lo que proporciona una estimación matizada del riesgo para cada ubicación.

Cada «x» azul representa una zona escolar; los puntos situados en la línea discontinua roja indican una predicción perfecta (predicho = real). La agrupación de puntos alrededor de la línea diagonal refleja el buen rendimiento del modelo: la densidad de accidentes prevista para la mayoría de las zonas se aproxima al valor observado. Se pueden observar algunos valores atípicos en los que el error de predicción es mayor (puntos más alejados de la línea). Sin embargo, no se aprecia ningún sesgo sistemático (los errores se distribuyen de forma relativamente simétrica alrededor de la línea). Esto sugiere que el modelo Bosque aleatorio no sobreestimó ni subestimó de manera sistemática las densidades de accidentes, sino que capturó adecuadamente la tendencia general. Las desviaciones ocasionales probablemente se deben a zonas con condiciones particulares y difíciles de predecir.

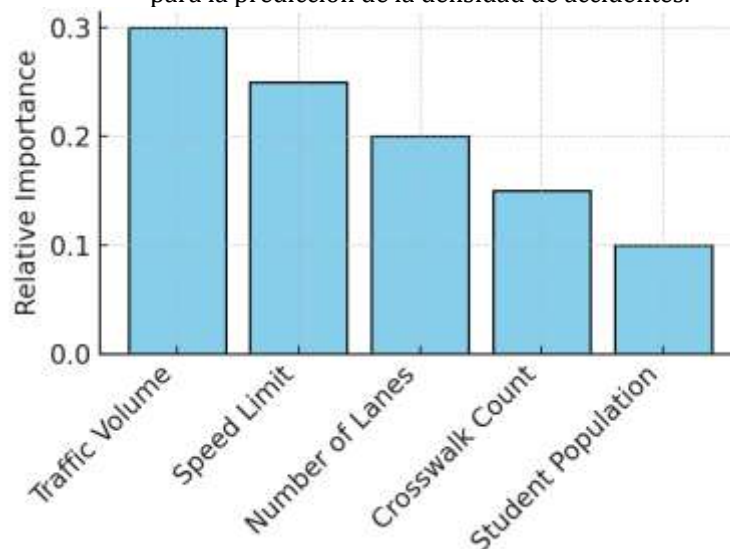
**Figura 4.** Densidad de accidentes de peatones prevista frente a la real para cada zona escolar del conjunto de pruebas, utilizando el modelo Bosque aleatorio.

Fuente: Elaboración propia, 2025.

Más allá de la precisión general, examinamos el modelo Bosque aleatorio para identificar qué características eran más influyentes en la predicción de la densidad de accidentes. El conjunto de árboles del modelo permite calcular los valores de importancia de las características, lo que indica

en qué medida cada predictor contribuye a reducir el error de predicción. La figura 5 presenta las características mejor clasificadas según su importancia relativa. El eje horizontal muestra la importancia relativa (contribución a la reducción del error de predicción) de las cinco características principales. El volumen de tráfico (flujo de vehículos a través de la zona escolar) es el predictor más fuerte, lo que refleja el aumento de la probabilidad de accidentes con una mayor exposición de los vehículos (Turner et al., 2017). El límite de velocidad es el segundo factor más importante: los límites de velocidad más altos suelen correlacionarse con un mayor riesgo y gravedad de los accidentes, ya que los conductores tienen menos tiempo de reacción y mayores distancias de frenado (Shuai y Kwon, 2025). El número de carriles también tiene una influencia considerable; las carreteras más anchas (con más carriles) fomentan velocidades más altas y crean distancias de cruce más largas para los peatones, lo que aumenta el riesgo. El número de pasos de peatones en una zona tiene una importancia moderada, lo que sugiere que, si bien los pasos de peatones son elementos de seguridad, su presencia se asocia con zonas de concentración de movimiento peatonal (y, por lo tanto, de accidentes, si no se protegen adecuadamente). La población estudiantil de la escuela también contribuye, en consonancia con la idea de que las zonas con más estudiantes (y, presumiblemente, más peatones) experimentan una mayor exposición de los peatones y, por lo tanto, más posibles accidentes.

**Figura 5.** Clasificación de importancia de las características del modelo de regresión Bosque aleatorio para la predicción de la densidad de accidentes.



Fuente: Elaboración propia, 2025.

Como era de esperar, predominan las medidas de exposición al tráfico y las características de la carretera. El volumen de tráfico (flujo de vehículos a través de la zona) es el factor más importante, seguido del límite de velocidad y del número de carriles. Estos factores se alinean intuitivamente con los mecanismos de seguridad: mayores volúmenes de tráfico y carreteras de varios carriles con alta velocidad tienden a aumentar los puntos de conflicto potenciales y la gravedad de los accidentes para los peatones, lo que conduce a tasas más elevadas de ocurrencia de accidentes (Turner et al., 2017). El modelo también identificó contribuciones notables de características como el número de pasos de peatones en la zona y la población estudiantil (o el número de alumnos matriculados) de la escuela. Un mayor número de pasos de peatones señalizados podría indicar una respuesta a la demanda de los peatones (y, por lo tanto, una exposición inherentemente mayor) o mitigar potencialmente el riesgo: la inclusión de esta característica en el modelo sugiere que las zonas con más cruces tienen patrones de accidentes distintos. Por otra parte, una mayor población estudiantil puede aumentar la actividad peatonal en los alrededores de las escuelas, lo que eleva el riesgo de exposición, lo que concuerda con los

hallazgos de que las zonas con más estudiantes o escuelas tienden a tener una mayor frecuencia de accidentes de peatones (Moradi et. al, 2016).

#### 4.3.3 Discusión

Nuestro análisis de regresión demuestra que los modelos de aprendizaje automático pueden predecir eficazmente el riesgo continuo de accidentes en las zonas escolares, proporcionando una métrica de riesgo granular para cada ubicación. El modelo Bosque aleatorio superó significativamente a modelos más simples, como KNN y árbol de decisión único, lo que refuerza la ventaja de las técnicas de conjunto para capturar las relaciones complejas y no lineales inherentes a los datos de accidentes (Donnell et al., 2020). Este hallazgo concuerda con otras investigaciones sobre predicción de seguridad vial, en las que modelos avanzados, como Bosque aleatorio y árboles potenciados por gradientes, han logrado alta precisión en la predicción de la frecuencia de accidentes (Shuai y Kwon, 2025). Cabe destacar que la precisión de Bosque aleatorio en nuestro estudio ( $R^2 \approx 0,88$ ) se encuentra a la par con el rendimiento reportado en esfuerzos similares de predicción de accidentes (normalmente  $R^2$  entre 0,8 y 0,9) (Shuai y Kwon, 2025), lo que indica que nuestro enfoque es competitivo con la tecnología más avanzada, a pesar de centrarse en un contexto específico: las zonas escolares.

Desde un punto de vista práctico, la predicción de la densidad de accidentes proporciona información más detallada que la simple clasificación de riesgos. La cuantificación del riesgo de accidentes permite a los responsables de la toma de decisiones priorizar las zonas de forma más eficaz, por ejemplo, asignando más recursos a las zonas en las que se prevé una densidad de accidentes significativamente mayor. El análisis de la importancia de las características proporciona más información para las intervenciones específicas; nuestros hallazgos destacan el volumen de tráfico y los límites de velocidad, lo que sugiere estrategias como la pacificación del tráfico, el control de la velocidad o el desvío del tráfico (Turner et al., 2017). Los factores de diseño de las carreteras, como el número de carriles y la presencia de pasos de peatones, también subrayan la eficacia potencial de las medidas de ingeniería (por ejemplo, la reducción del número de carriles y los refugios para peatones). Además, el tamaño de la población estudiantil indica la necesidad de mejorar la supervisión y las medidas educativas en las escuelas más grandes.

En cuanto a las limitaciones y el trabajo futuro, el modelo Bosque aleatorio, aunque muy preciso, sigue siendo una «caja negra», lo que limita su interpretabilidad directa. Hemos abordado esta cuestión parcialmente mediante la importancia de las características, pero recomendamos aplicar métodos avanzados de interpretabilidad (por ejemplo, valores SHAP, gráficos de dependencia parcial) para aclarar aún más los efectos no lineales. La incorporación de variables adicionales, como características detalladas del entorno construido, comportamientos de los conductores o datos de tráfico en tiempo real, puede mejorar las predicciones. Las investigaciones futuras podrían explorar modelos como XGBoost o redes neuronales, aunque es poco probable que se produzcan mejoras sustanciales más allá de Bosque aleatorio. Por último, a pesar del buen rendimiento estadístico, la validación en terreno mediante evaluaciones in situ es fundamental para confirmar las mejoras prácticas en materia de seguridad.

## 5. Conclusiones

Los resultados de este estudio demuestran que la densidad de atropellos de peatones en las zonas escolares de Riad está significativamente influenciada por las características demográficas y de infraestructura a nivel de distrito, incluyendo la densidad de población, la longitud de la red de carreteras y las condiciones socioeconómicas. El análisis MLR reveló que el aumento de la densidad de la red de carreteras y la mayor densidad de población se correlacionan significativamente con el aumento de la frecuencia de los accidentes. Por el contrario, una mayor renta media del distrito se asocia con una reducción del riesgo de atropellos de peatones. El análisis ML utilizando el modelo Bosque aleatorio mejoró aún más la precisión predictiva, identificando eficazmente factores críticos como el volumen de tráfico, los límites de velocidad, el



número de carriles, la disponibilidad de pasos de peatones y la población estudiantil. Al integrar métodos estadísticos tradicionales y técnicas de IA de vanguardia, este estudio proporciona información sólida y basada en datos que respalda intervenciones urbanas más inteligentes, en consonancia con los objetivos de mejorar la eficiencia urbana, la calidad de vida y la resiliencia.

A pesar de los sólidos resultados analíticos, este estudio tiene limitaciones. La dependencia de datos históricos sobre accidentes, que pueden estar sujetos a infravaloración e inexactitudes, es una restricción notable. Además, dado el contexto urbano y socioeconómico único de Riad, es necesario considerar cuidadosamente la generalización de estos hallazgos a otros entornos urbanos.

Las investigaciones futuras deberían aprovechar las herramientas avanzadas de IA, los simuladores de gemelos digitales y los conjuntos de datos detallados sobre el comportamiento, incluidos los comportamientos en tiempo real de los peatones y los conductores, la eficacia de la aplicación de la ley y la calidad de la infraestructura peatonal, para refinar aún más la precisión predictiva. La incorporación de la dinámica espacio-temporal mediante modelos geoespaciales avanzados puede optimizar aún más el rendimiento predictivo y la eficacia de las intervenciones.

Los responsables políticos, los urbanistas y los ingenieros pueden utilizar estos conocimientos mejorados por la IA para aplicar intervenciones de seguridad específicas y equitativas. Los gestores municipales deben centrarse en las zonas escolares de alto riesgo que el modelo Bosque aleatorio ha identificado para realizar mejoras de ingeniería inmediatas que minimicen los accidentes a un coste razonable. Los paneles de control GIS municipales reciben datos de mapas de riesgo generados por el modelo, lo que permite a los funcionarios asignar presupuestos y programar el mantenimiento basándose en el análisis de los datos. La implementación de sistemas de control automatizados, como cámaras de velocidad o señales de radar, debe establecerse en zonas escolares con límites de velocidad bajos que se encuentren en áreas con alta densidad de población y registros de accidentes graves. El despliegue de campañas de rutas seguras al colegio y programas de guardias de cruce en áreas con alta población estudiantil requiere la colaboración activa de las autoridades educativas locales. El sistema requiere actualizaciones periódicas del modelo con nuevos datos sobre accidentes y tráfico para seguir siendo eficaz en entornos urbanos en evolución, como parte del desarrollo de la ciudad inteligente de la Visión 2030 de Riad. En última instancia, el enfoque de este estudio respalda el objetivo de la ciudad inteligente de crear entornos escolares más seguros, resilientes y sensibles a los datos en Riad y en contextos urbanos comparables.

## **6. Agradecimientos**

El presente estudio se inscribe en el marco del proyecto del Ayuntamiento de Riad «Mejora de la seguridad vial en torno a las escuelas y mezquitas de Riad (Fase 1)».

## **7. Declaración de conflicto de intereses**

Los autores confirman que no existe ningún conflicto de intereses en relación con la publicación de este artículo.

## **8. Declaración de ética de la investigación**

Esta investigación cumple con las directrices de buenas prácticas del Comité de Ética Editorial (COPE) para la investigación y la publicación responsables.

## Referencias

- AIP Foundation. (2025). Smart Solutions for Safer Schools: Vietnam Leads with AI-Driven Road Safety Big Data Screening across the country. <https://www.aip-foundation.org/smart-solutions-for-safer-schools-vietnam-leads-with-ai-driven-road-safety-big-data-screening-across-the-country/>
- Alharbi, R., Alghamdi, A., Al-Jafar, R., Almuwallad, A., & Chowdhury, S. (2024). *Identifying the key characteristics, trends, and seasonality of pedestrian traffic injury at a major trauma center in Saudi Arabia: a registry-based retrospective cohort study, 2017–2022*. BMC emergency medicine, 24(1), 135. <https://doi.org/10.1186/s12873-024-01051-5>
- Alomari, A. H., Al-Deek, H., Sandt, A., Rogers Jr, J. H., & Hussain, O. (2016). *Regional evaluation of bus rapid transit with and without transit signal priority*. Transportation Research Record, 2554(1), 46-59. <https://doi.org/10.3141/2554-06>
- Bahrami, V., Lavrenz, S., & Ahmed, M. M. (2024). *Severity Analysis of Pedestrian and Bike Crashes in School Buffer Zones*. Transportation Research Record, 03611981241297682. <https://doi.org/10.1177/03611981241297682>
- Basunia, A., Anchal, T. H., Tasnim, J., Ahmed, N., Mahzabeen, T. Z., & Rifaat, S. M. (2025). Exploring factors influencing jaywalking to promote safe and active travel to school among Dhaka city adolescent students. Journal of Road Safety, 36(2), 47-62. <https://doi.org/10.33492/JRS-D-25-1-2470690>
- Datasaudi. (2025). *A unified platform to present and analyze the latest economic and social data for the Kingdom*. Ministry of Economy & Planning. Kingdom of Saudi Arabia. <https://datasaudi.sa/en>
- DiMaggio, C., & Li, G. (2013). *Effectiveness of a safe routes to school program in preventing school-aged pedestrian injury*. Pediatrics, 131(2), 290-296. <https://doi.org/10.1542/peds.2012-2182>
- Donnell, E. T., Hanks, E., Porter, R. J., Cook, L., Srinivasan, R., Li, F., ... & Eccles, K. A. (2020). *The Development of Crash Modification Factors: Highway Safety Statistical Paper Synthesis*. No. FHWA-HRT-20-069. United States. FHWA, Federal Highway Administration.. <https://www.fhwa.dot.gov/publications/research/safety/20069/20069.pdf>
- Ehsani, J. P., Michael, J. P., & MacKenzie, E. J. (2023). *The future of road safety: challenges and opportunities*. The Milbank Quarterly, 101(Suppl 1), 613. <https://doi.org/10.1111/1468-0009.12644>
- Esri. (2025). *ArcGIS Pro, The world's leading desktop GIS software*. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>
- Eun, S. J. (2023). *Effects of tougher school zone laws on road traffic safety in school zones for children in South Korea*. Journal of Transport & Health, 32, 101687. <https://doi.org/10.1016/j.jth.2023.101687>
- Farid, A., Lin, E., & Pande, A. (2024). *Analysis of School Zone Crash Severities with an Equity Lens: A Random Parameters Modeling Approach*. Transportation Research Record, 03611981241295716. <https://doi.org/10.1177/03611981241295716>
- Flanagan, R. and Morgan, R. (2023). *Improving traffic safety during arrival and dismissal for students at the Quinsigamond School*. Project Report. Worcester Polytechnic Institute, MA, United States. <https://digital.wpi.edu/downloads/rr172143p>
- Forward, S., Henriksson, P., Silvano, A. P., Miyoba, T., Sinkala, S., Mawele, S., & Mwamba, D. (2025). *Increasing traffic safety at schools in Zambia: a before and after study*. Reg. No., VTI: 2022/0296-8.3 <https://urn.kb.se/resolve?urn=urn:nbn:se:vti:diva-21493>
- Hu, X., Deng, H., Liu, H., Zhou, J., Liang, H., Chen, L., & Zhang, L. (2025). *Assessment of the collision risk on the road around schools during morning peak period*. Accident Analysis & Prevention, 210, 107854. <https://doi.org/10.1016/j.aap.2024.107854>
- iRAP. (2025). What is AiRAP. <https://irap.org/project/ai-rap/>

- Ivan, K., Benedek, J., & Ciobanu, S. M. (2019). *School-aged pedestrian-vehicle crash vulnerability*. Sustainability, 11(4), 1214. <https://doi.org/10.3390/su11041214>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R*. (2nd ed.). Springer.
- Kingham, S., Sabel, C. E., & Bartie, P. (2011). *The impact of the 'school run' on road traffic accidents: A spatio-temporal analysis*. Journal of transport geography, 19(4), 705-711. <https://doi.org/10.1016/j.jtrangeo.2010.08.011>
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2005). *Applied Linear Statistical Models*. McGraw-Hill.
- Lee, G., Park, Y., Kim, J., & Cho, G. H. (2016). *Association between intersection characteristics and perceived crash risk among school-aged children*. Accident Analysis & Prevention, 97, 111-121. <https://doi.org/10.1016/j.aap.2016.09.001>
- Lee, I. J., Sagar, S., Agarwal, N., Srinivasan, S., & Steiner, R. (2024). *Data-Driven Approach to Develop a Master Plan to Prioritize Schools for the Safe Routes to School Program*. Transportation Research Record, 03611981241250019. <https://doi.org/10.1177/03611981241250019>
- Lordswill, N. T., Jean-Francois, W., Fondzenyuy, S. K., Feudjio Tezong, S. L., Ndonue, A. R., Usami, D. S., & Persia, L. (2024). *Assessment and countermeasures selection for safer roads to schools in the city of Yaoundé: progressive evaluation using surveys and iRAP methodology*. Transportation Research Procedia, 1-8. AIIT 4th International Conference on Transport Infrastructure and Systems (TIS ROMA 2024), 19th - 20th September 2024, Rome Italy. [https://iris.uniroma1.it/bitstream/11573/1722401/1/Ndingwan\\_Assessment-and-countermeasures-selection\\_2024.pdf](https://iris.uniroma1.it/bitstream/11573/1722401/1/Ndingwan_Assessment-and-countermeasures-selection_2024.pdf)
- Mienye, I. D., & Jere, N. (2024). *A survey of árbols: Concepts, algorithms, and applications*. IEEE access. <https://doi.org/10.1109/ACCESS.2024.3416838>
- Montgomery, D. C. (2017). *Design and analysis of experiments*. John Wiley & sons.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Moradi, A., Soori, H., Kavousi, A., Eshghabadi, F., & Jamshidi, E. (2016). *Spatial factors affecting the frequency of pedestrian traffic crashes: A systematic review*. Archives of trauma research, 5(4), e30796. <https://doi.org/10.5812/at.30796>
- Oh, J., & Kim, J. (2025). *Potential risk factors of child pedestrian crashes after-school hours in Seoul, Korea*. Journal of Transport Geography, 123, 104084. <https://doi.org/10.1016/j.jtrangeo.2024.104084>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). *Scikit-learn: Machine learning in Python*. the Journal of machine Learning research, 12, 2825-2830.
- Probst, P., Wright, M. N., & Boulesteix, A.-L. (2020). *Hyperparameters and tuning strategies for Bosque aleatorio*. WIREs Data Mining and Knowledge Discovery, 10(3), e1371. <https://doi.org/10.1002/widm.1371>
- Regidor, J. R. F., Kamid, S. A., Latonero, G. S. D., Abao, N. S. A., & Sigua, R. D. (2023). *Evaluation and Improvement of Road Safety In the Vicinity of Schools*. Philippine Institute of Civil Engineers 2023 National Convention. [https://www.researchgate.net/publication/378077600\\_Evaluation\\_and\\_Improvement\\_of\\_Road\\_Safety\\_in\\_the\\_Vicinity\\_of\\_Schools](https://www.researchgate.net/publication/378077600_Evaluation_and_Improvement_of_Road_Safety_in_the_Vicinity_of_Schools)
- Rothman, L., Buliung, R., Howard, A., Macarthur, C., & Macpherson, A. (2017a). *The school environment and student car drop-off at elementary schools*. Travel Behaviour and Society, 9, 50-57. <https://doi.org/10.1016/j.tbs.2017.03.001>
- Rothman, L., Howard, A., Buliung, R., Macarthur, C., Richmond, S. A., & Macpherson, A. (2017b). *School environments and social risk factors for child pedestrian-motor vehicle collisions: A case-control study*. Accident Analysis & Prevention, 98, 252-258. <https://doi.org/10.1016/j.aap.2016.10.017>

- Rothman, L., Macarthur, C., To, T., Buliung, R., & Howard, A. (2015). *Motor vehicle-pedestrian collisions and walking to school: the role of the built environment*. *Pediatrics*, 133(5), 776-784. <https://doi.org/10.1542/peds.2013-2317>
- Sakib, N., Paul, T., Ahmed, M. T., Al Momin, K., & Barua, S. (2024). Investigating factors influencing pedestrian crosswalk usage behavior in Dhaka city using supervised machine learning techniques. *Multimodal Transportation*, 3(1), 100108. <https://doi.org/10.1016/j.multra.2023.100108>
- Shuai, Z., & Kwon, T. J. (2025). *Analyzing Winter Crash Dynamics Using Spatial Analysis and Crash Frequency Prediction Models with SHAP Interpretability*. *Future Transportation*, 5(1), 17. <https://doi.org/10.3390/futuretransp5010017>
- Tetali, S., Edwards, P., Murthy, G. V. S., & Roberts, I. (2016). *Road traffic injuries to children during the school commute in Hyderabad, India: cross-sectional survey*. *Injury prevention*, 22(3), 171-175. <https://doi.org/10.1136/injuryprev-2015-041854>
- Turner, S., Sener, I. N., Martin, M. E., Das, S., Hampshire, R. C., Fitzpatrick, K., ... & Wijesundera, R. K. (2017). *Synthesis of methods for estimating pedestrian and bicyclist exposure to risk at areawide levels and on specific transportation facilities*. No. FHWA-SA-17-041. United States. Department of Transportation. Federal Highway Administration. Office of Safety. <https://highways.dot.gov/sites/fhwa.dot.gov/files/2022-06/fhwasa17014.pdf>
- UNICEF. (2022). *Technical guidance for child and adolescent road safety*. New York: United Nations Children's Fund.
- University of Cambridge. (2019). *Children who walk to school less likely to be overweight or obese*. ScienceDaily. ScienceDaily, 21 May 2019. [www.sciencedaily.com/releases/2019/05/190521101344.htm](http://www.sciencedaily.com/releases/2019/05/190521101344.htm).
- Washington, S., Karlaftis, M. G., Mannering, F., & Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC.
- WHO, World Health Organization. (2023a). *Global status report on road safety 2023*. Geneva: WHO Press. <https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023>
- WHO, World Health Organization. (2023b). *Reducing Road Crash Deaths in the Kingdom of Saudi Arabia*. <https://www.who.int/news/item/20-06-2023-reducing-road-crash-deaths-in-the-kingdom-of-saudi-arabia>
- Yamarthi, D., Raman, H., Parvin, S. (2025). *United States Road Accident Prediction using Machine Learning Algorithms*. arXiv:2505.06246v1 28 Apr 2025. <https://arxiv.org/pdf/2505.06246v1>
- Yu, C. Y., & Zhu, X. (2016). *Planning for safe schools: Impacts of school siting and surrounding environments on traffic safety*. *Journal of Planning Education and Research*, 36(4), 476-486. <https://doi.org/10.1177/0739456X15616460>
- Zhang, K., Tamakloe, R., Cao, M., & Kim, I. (2024). *Exploring fatal/severe pedestrian injury crash frequency at school zone crash hotspots: using interpretable machine learning to assess the micro-level street environment*. *Journal of Transport Geography*, 121, 104034. <https://doi.org/10.1016/j.jtrangeo.2024.104034>
- Zhao, X., Li, J., Ding, H., Zhang, G., & Rong, J. (2015). *A generic approach for examining the effectiveness of traffic control devices in school zones*. *Accident Analysis & Prevention*, 82, 134-142. <https://doi.org/10.1016/j.aap.2015.05.021>