Sistema inteligente Big Data para predecir siniestros viales

Intelligent Big Data System to Predict Road Accidents

Deysi Verónica Quintal Segura

Correo: deysegura18@gmail.com

ORCID:

Resumen

La atención de siniestros viales en carreteras federales representa un desafío crítico para las instituciones públicas encargadas de la seguridad y asistencia al viajero, debido a la naturaleza reactiva de los mecanismos tradicionales de reporte. En este contexto, el uso de datos generados en tiempo real a través de redes sociales se posiciona como una alternativa para fortalecer la gestión preventiva. El presente estudio analiza la viabilidad de un sistema inteligente basado en Big Data para la predicción temprana de siniestros viales, utilizando datos públicos provenientes de la red social X. La investigación adopta un enfoque cuantitativo, no experimental y de corte descriptivo-correlacional. Se analizaron más de 140,000 registros mediante técnicas de clasificación automática y análisis de patrones. Los resultados muestran niveles de precisión superiores al 85 % y una concordancia aceptable entre la clasificación automática y la validación experta, evidenciando la capacidad del modelo para identificar patrones asociados a la ocurrencia de siniestros y apoyar la toma de decisiones en la gestión pública vial.

Palabras clave: Big Data; siniestros viales; predicción; redes sociales; gestión pública.

Abstract

Road accident response on federal highways represents a critical challenge for public institutions responsible for traveler safety, mainly due to the reactive nature of traditional reporting mechanisms. In this context, the use of real-time data generated through social networks emerges as a valuable alternative to strengthen preventive management. This study analyzes the feasibility of an intelligent Big Data–based system for early road accident prediction using publicly available data from the social network X. The research follows a quantitative, non-experimental, descriptive-correlational approach. More than 140,000 records were analyzed using automatic classification and pattern analysis techniques. The results show accuracy levels above 85% and acceptable agreement between automated classification and expert validation, demonstrating the model's ability to identify patterns associated with road accidents and support decision-making processes in public road management.

Keywords: Big Data; road accidents; prediction; social networks; public management.

Introducción

Los siniestros viales constituyen una problemática estructural en la gestión del transporte y la seguridad pública, particularmente en contextos de alta movilidad como las carreteras federales de México. Estos eventos generan impactos significativos en términos humanos, económicos y operativos, lo que exige el desarrollo de estrategias orientadas no solo a la atención, sino también a la prevención (World Health Organization, 2023).

Tradicionalmente, la gestión de siniestros viales se ha basado en registros administrativos y reportes oficiales, los cuales presentan limitaciones en términos de inmediatez, cobertura y capacidad predictiva. En este sentido, diversos estudios han señalado que los enfoques reactivos resultan insuficientes para la gestión eficiente del riesgo, especialmente en entornos dinámicos (Janssen & Kuk, 2016).

En paralelo, el crecimiento del uso de redes sociales ha propiciado la generación de grandes volúmenes de datos en tiempo real, los cuales reflejan condiciones de tránsito, incidentes y alertas viales reportadas directamente por los usuarios. Estas plataformas se han consolidado como fuentes emergentes de información para la detección temprana de eventos, particularmente en situaciones de emergencia (Sakaki et al., 2010).

Desde la perspectiva del Big Data, el análisis de datos no estructurados provenientes de múltiples fuentes permite identificar patrones y tendencias que pueden ser utilizados para la toma de decisiones basada en evidencia (Kitchin, 2014). En el ámbito de la gestión pública, el uso de analítica avanzada ha demostrado su potencial para mejorar la eficiencia operativa y la asignación de recursos (Mergel et al., 2018).

No obstante, la integración de estas tecnologías en la gestión de siniestros viales aún presenta limitaciones, particularmente en lo relativo a la articulación entre fuentes de datos abiertas y sistemas institucionales. Esta brecha tecnológica plantea la necesidad de desarrollar modelos que permitan aprovechar el potencial del Big Data para la predicción temprana de riesgos.

En este contexto, el presente estudio analiza la viabilidad de un sistema inteligente basado en Big Data para la predicción de siniestros viales, utilizando datos públicos provenientes de la red social X, con el fin de fortalecer la gestión preventiva y la toma de decisiones en instituciones públicas de asistencia vial.

Objetivo del artículo

Analizar la viabilidad de un sistema inteligente basado en Big Data para la predicción temprana de siniestros viales en carreteras federales de México, utilizando datos públicos provenientes de la red social X.

De manera específica, el estudio busca:

- Identificar patrones temporales, temáticos y geográficos asociados a la ocurrencia de siniestros viales.
- Evaluar el desempeño de técnicas de clasificación automática aplicadas a datos no estructurados.
- Analizar el potencial de estos sistemas como herramientas de apoyo a la gestión preventiva y operativa de instituciones de asistencia vial.
- El artículo se concibe como un aporte empírico orientado a fortalecer la toma de decisiones basada en datos en el ámbito de la gestión pública del transporte y la seguridad vial.

Revisión de la literatura

Big Data y gestión pública

El Big Data se define como el conjunto de tecnologías y procesos orientados al análisis de grandes volúmenes de datos caracterizados por su variedad, velocidad y volumen (Kitchin, 2014). En el ámbito de la gestión pública, estas tecnologías han sido utilizadas para mejorar la toma de decisiones, optimizar procesos y fortalecer la prestación de servicios (Janssen & Kuk, 2016).

Diversos estudios han demostrado que el uso de analítica avanzada permite identificar patrones complejos en contextos dinámicos, lo que resulta particularmente relevante para la gestión del riesgo y la planificación operativa (Mergel et al., 2018).

Redes sociales como fuente de datos para la predicción de riesgos

Las redes sociales se han consolidado como una fuente relevante de datos en tiempo real, especialmente en el contexto de eventos críticos. Investigaciones previas han evidenciado que plataformas como X pueden ser utilizadas para la detección temprana de incidentes, mediante el análisis de contenido generado por los usuarios (Sakaki et al., 2010).

En el ámbito de la movilidad y el transporte, estos datos permiten complementar la información oficial, proporcionando una visión más amplia y actualizada de las condiciones en carretera (Zhang et al., 2018).

Analítica avanzada para predecir siniestros

La predicción de siniestros viales ha evolucionado desde modelos estadísticos tradicionales hacia enfoques basados en aprendizaje automático y análisis de grandes volúmenes de datos. Estos modelos permiten integrar múltiples variables y mejorar la capacidad predictiva de los sistemas (Chen et al., 2014).

En este sentido, el uso de Big Data en combinación con técnicas de clasificación y agrupamiento ha demostrado ser eficaz para la identificación de patrones asociados a la ocurrencia de eventos de riesgo (Kumar et al., 2017).

Metodología

La investigación se desarrolló bajo un enfoque cuantitativo, no experimental y de corte descriptivo-correlacional. La fuente de datos estuvo constituida por publicaciones públicas emitidas en la red social X, relacionadas con siniestros viales, accidentes, bloqueos y condiciones de tránsito en carreteras federales.

La muestra analizada incluyó más de 140 000 registros recolectados durante un periodo definido, seleccionados mediante criterios de relevancia temática, frecuencia y confiabilidad de la fuente. Las variables analizadas incluyeron tipo de incidente, temporalidad, localización geográfica y recurrencia temática.

El procesamiento de los datos se realizó mediante técnicas de clasificación automática y análisis de patrones, con el objetivo de identificar correlaciones significativas asociadas a la ocurrencia de

siniestros. Para la validación del sistema se compararon los resultados del modelo con una muestra evaluada por expertos del ámbito institucional, utilizando métricas de precisión y concordancia.

En términos éticos, se trabajó exclusivamente con datos públicos, garantizando la anonimización de la información y el uso académico de los resultados.

Resultados

Los resultados evidencian patrones claros en la ocurrencia de siniestros viales, particularmente en horarios y zonas específicas. El modelo alcanzó niveles de precisión superiores al 85 %, lo que demuestra su capacidad para identificar eventos relevantes.

Asimismo, se identificaron zonas de alta incidencia y horarios de mayor necesidad operativa, lo que permite apoyar la toma de decisiones operativas en la asignación de recursos.

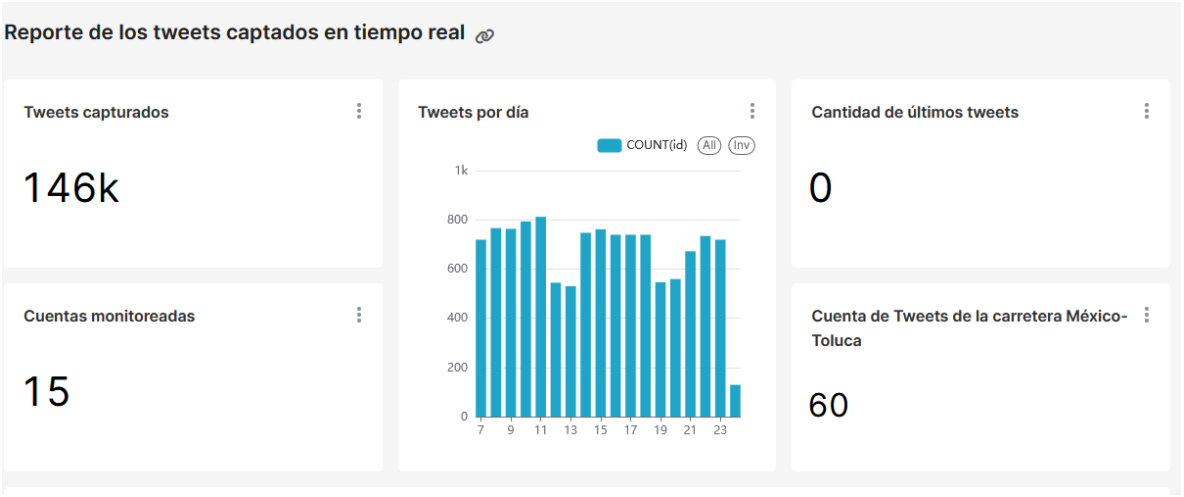


Ilustración 1. Reporte de tweets capturados en tiempo real.

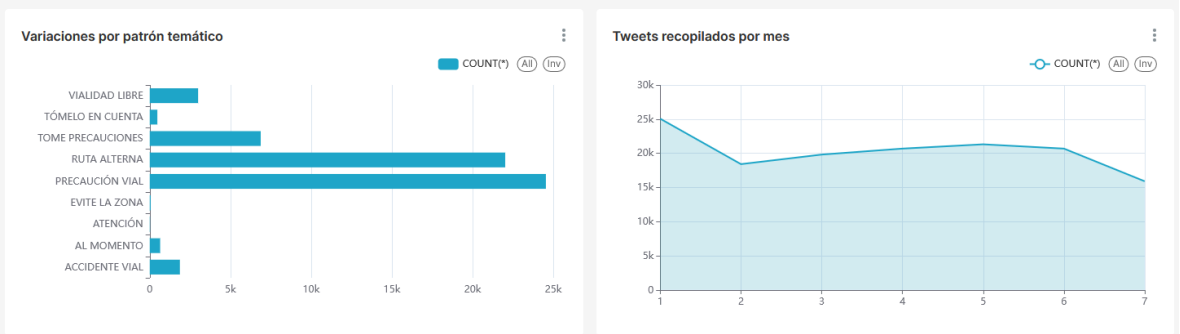


Ilustración 2. Patrón temático de tweets.

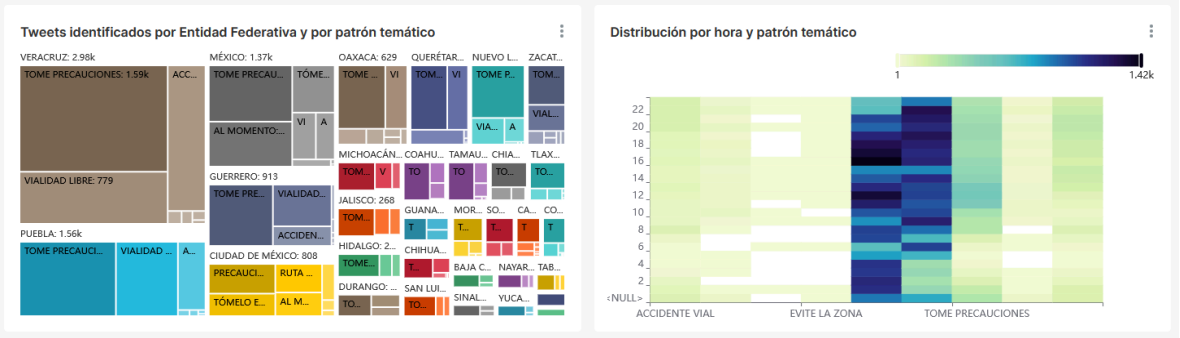


Ilustración 3. Zonas identificadas con alta incidencia de siniestros.

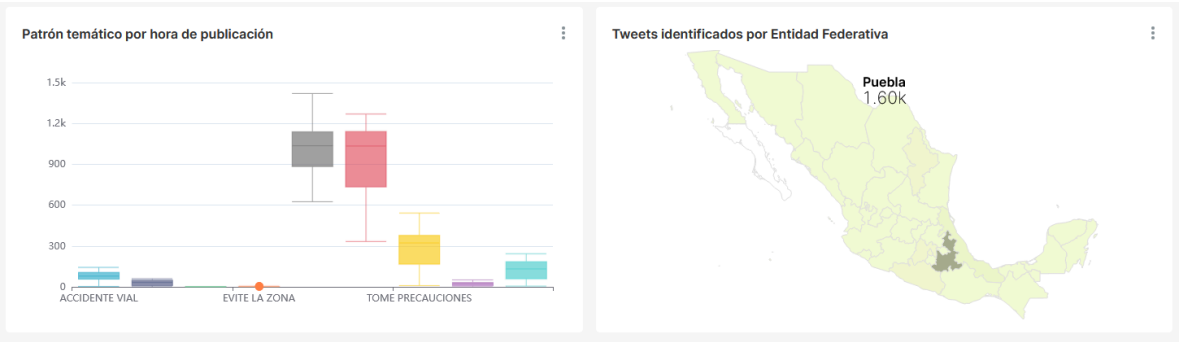


Ilustración 4. Horarios de mayor necesidad operativa.

Discusión

Los resultados obtenidos confirman el potencial del Big Data y de las redes sociales como herramientas complementarias para la gestión de siniestros viales. En concordancia con estudios previos, el análisis de datos en tiempo real permite mejorar la capacidad de anticipación frente a eventos de riesgo (Sakaki et al., 2010).

Asimismo, los niveles de precisión alcanzados en el modelo coinciden con investigaciones que destacan la efectividad de técnicas de clasificación automática en el análisis de datos no estructurados (Chen et al., 2014). Esto refuerza la validez del enfoque metodológico adoptado en el presente estudio.

En comparación con los enfoques tradicionales basados exclusivamente en registros oficiales, el modelo propuesto amplía la capacidad de detección al incorporar información generada por los usuarios, lo que coincide con lo señalado por Zhang et al. (2018) respecto al valor de las redes sociales como fuente complementaria de datos.

No obstante, se identifican limitaciones asociadas a la dependencia de una sola plataforma y a posibles sesgos en la información, lo cual ha sido documentado en estudios sobre calidad de datos en entornos digitales (Kitchin, 2014).

En este sentido, los resultados sugieren que la integración de múltiples fuentes de datos podría fortalecer la robustez del modelo y mejorar su aplicabilidad en contextos reales de gestión pública.

Conclusiones

El artículo demuestra la viabilidad de un sistema inteligente basado en Big Data para la predicción temprana de siniestros viales a partir de datos públicos de redes sociales. Los resultados evidencian que estos sistemas pueden fortalecer la gestión preventiva y la toma de decisiones operativas en instituciones de asistencia vial.

Desde una perspectiva práctica, el modelo propuesto ofrece una herramienta adaptable a distintos contextos carreteros y organismos públicos. Desde el ámbito académico, el estudio aporta evidencia empírica relevante para el análisis de la aplicación del Big Data en la gestión pública del transporte.

Finalmente, se concluye que la incorporación progresiva de analítica predictiva en la gestión vial representa una oportunidad estratégica para mejorar la seguridad en las carreteras y optimizar el uso de los recursos públicos.

Referencias

Ahuja, R., & Shakeel, H. (2017). *Mining social media: A brief introduction to the tools and techniques of analyzing social media data*. Springer.

Bertalanffy, L. von. (1968). *General system theory: Foundations, development, applications*. George Braziller.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*.

García, J., & Martínez, P. (2020). Innovaciones tecnológicas para la gestión de incidentes viales: Desafíos y oportunidades. *Revista de Ingeniería del Transporte*.

Hernández-Sampieri, R. (2022). *Metodología de la investigación* (7.ª ed.). McGraw-Hill.

Janssen, M., & Kuk, G. (2016). Big data in the public sector: A systematic review. *Government Information Quarterly*.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. SAGE Publications.

Kumar, S., et al. (2017). Predictive analytics in transportation systems.

López, A., & Sánchez, M. (2021). Sistemas de información centralizados y su impacto en la predicción de eventos críticos. *Journal of Data Science and Technology*.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Mergel, I., Rethemeyer, R. K., & Isett, K. (2018). Big data in public affairs. *Public Administration Review*.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors.

World Health Organization. (2023). *Global status report on road safety*.

Zhang, Y., et al. (2018). Social media data for traffic analysis.