



Fusión Semántico-Espacio-Temporal de Datos de Redes Sociales Y Gubernamentales para el Análisis de Crimen y Seguridad Pública

Leon Felipe Davis Coropuna

Orientador: Prof Ana Maria Cuadros Valdivia

Plan de Tesis presentado la Escuela Profesional Ciencia de la Computación como paso previo a la elaboración de la Tesis Profesional.

**UNSA - Universidad Nacional de San Agustín de Arequipa
Junio de 2025**

1. Motivación y Contexto

En la actualidad, el análisis de datos provenientes de redes sociales ha cobrado gran relevancia como herramienta para comprender fenómenos sociales complejos. Este interés ha sido impulsado por la posibilidad de combinar información espacial, temporal y textual, lo que permite una visión más integral de los eventos sociales [Kibanov, 2019], al considerar no solo dónde y cuándo ocurren, sino también qué expresan los usuarios en sus publicaciones. En este sentido, el análisis de contenidos y la minería de opiniones aplicados a redes sociales han emergido como enfoques prometedores para monitorear amenazas a la seguridad pública, detectar eventos críticos en tiempo real y comprender las reacciones sociales frente a fenómenos disruptivos [Md Suhaimin et al., 2023].

A pesar de los avances recientes, persisten desafíos significativos en la integración efectiva de las dimensiones semántica, espacial y temporal en el análisis de redes sociales. La complejidad, heterogeneidad y volumen de los datos [Almaslukh et al., 2024], junto con la escasez de herramientas y modelos capaces de abordar estas dimensiones de manera conjunta, limitan la capacidad de extraer conocimiento útil para la toma de decisiones en ámbitos como la seguridad y el análisis delictivo. Estudios recientes destacan la necesidad de enfoques más integrales: por un lado, la mayoría de las propuestas tienden a procesar estas dimensiones de forma secuencial [Hanny and Resch, 2024, G. Almatar et al., 2020]; por otro, persiste una marcada fragmentación de los datos, especialmente en contextos como el crimen y la violencia urbana, debido a su procedencia de múltiples fuentes [Vivek and Prathap, 2023, Ahmed et al., 2022].

Esta situación pone en evidencia un vacío metodológico y tecnológico que justifica el desarrollo de nuevas propuestas orientadas a la integración semántico-espacio-temporal en el análisis de redes sociales y datos gubernamentales, especialmente en contextos relacionados con el crimen y la seguridad pública. En este trabajo, se propone una metodología que combina técnicas de integración de datos heterogéneos [Ahmed et al., 2022], con el objetivo de analizar información proveniente tanto de redes sociales como de fuentes oficiales en relación con eventos delictivos urbanos, integrando sus dimensiones semántica, espacial y temporal.

2. Definición del Problema

Actualmente, los enfoques existentes para el análisis de datos criminales enfrentan limitaciones debido a la fragmentación y heterogeneidad de la información proveniente de diversas fuentes. Esta dispersión genera registros superpuestos, incompletos o inconsistentes, lo que dificulta consolidar una visión unificada de fenómenos delictivos.

3. Justificación

El análisis de datos de redes sociales se enmarca en el campo de la inteligencia urbana y la seguridad pública, área que permite estudiar dinámicas sociales mediante la integración de información geográfica, temporal y semántica [Martí et al., 2019]. Esta disciplina ayuda a determinar patrones de comportamiento colectivo e identificar focos de riesgo en ciudades. En este sentido, se requieren enfoques capaces de procesar grandes volúmenes de datos heterogéneos [Ahmed et al., 2022], correlacionar variables espacio-temporales con contenido semántico y ofrecer visualizaciones interpretables [Pongpaichet et al., 2024]. Frente a esto, el presente trabajo propone una metodología que combina técnicas de fusión de datos heterogéneos y modelos de procesamiento de lenguaje natural especializados en el análisis de eventos delictivos. Esta propuesta pretende superar la fragmentación actual de datos criminales integrando las tres dimensiones semántico-espacio-temporal, con el fin de establecer un marco metodológico replicable para futuros estudios sobre análisis social contextualizado en seguridad urbana.

4. Objetivos

4.1. Objetivo General

Desarrollar una metodología para el análisis de datos criminales urbanos que integre de manera conjunta las dimensiones semántica, espacial y temporal, a partir de la fusión de información heterogénea proveniente de redes sociales y fuentes oficiales, con el fin de generar una visión unificada y coherente que contribuya a la toma de decisiones en seguridad pública.

4.2. Objetivos Específicos

- Recolectar datos heterogéneos de redes sociales y fuentes oficiales sobre eventos delictivos.
- Preprocesar y limpiar los datos para reducir fragmentación e inconsistencias.
- Aplicar procesamiento de lenguaje natural para extraer información semántica relevante.
- Implementar modelos para integrar dimensiones espacial y temporal.
- Desarrollar visualizaciones que faciliten la interpretación y toma de decisiones.

Referencias

- [Ahmed et al., 2022] Ahmed, S., Gentili, M., Sierra-Sosa, D., and Elmaghraby, A. S. (2022). Multi-layer data integration technique for combining heterogeneous crime data. *Information Processing Management*, 59(3):102879.
- [Almaslukh et al., 2024] Almaslukh, A., Almaalwy, A., Allheeib, N., Alajaji, A., Almukaynizi, M., and Alabdulkarim, Y. (2024). Top-k sentiment analysis over spatio-temporal data. *PeerJ Computer Science*, 10:e2297.
- [G. Almatar et al., 2020] G. Almatar, M., Alazmi, H. S., Li, L., and Fox, E. A. (2020). Applying gis and text mining methods to twitter data to explore the spatiotemporal patterns of topics of interest in kuwait. *ISPRS International Journal of Geo-Information*, 9(12).
- [Hanny and Resch, 2024] Hanny, D. and Resch, B. (2024). Multimodal geo-information extraction from social media for supporting decision-making in disaster management. *AGILE: GIScience Series*, 5:28.
- [Kibanov, 2019] Kibanov, M. (2019). Social network mining for analysis of social phenomena.
- [Martí et al., 2019] Martí, P., Serrano-Estrada, L., and Nolasco-Cirugeda, A. (2019). Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74:161–174.
- [Md Suhaimin et al., 2023] Md Suhaimin, M. S., Ahmad Hijazi, M. H., Moun, E. G., Nohuddin, P. N. E., Chua, S., and Coenen, F. (2023). Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions. *Journal of King Saud University - Computer and Information Sciences*, 35(9):101776.
- [Pongpaichet et al., 2024] Pongpaichet, S., Sukosit, B., Duangtanawat, C., Jamjongdamrongkit, J., Mahacharoensuk, C., Matangkarat, K., Singhajan, P., Noraset, T., and Tuarob, S. (2024). Camelon: A system for crime metadata extraction and spatiotemporal visualization from online news articles. *IEEE Access*, 12:22778–22802.
- [Vivek and Prathap, 2023] Vivek, M. and Prathap, B. R. (2023). Spatio-temporal crime analysis and forecasting on twitter data using machine learning algorithms. *SN Computer Science*, 4(4):383. Accessed: 2023/05/06.