

Análisis Descriptivo y Predictivo Basado en Datos del Tráfico Vehicular en Antofagasta: Un Enfoque a partir de Reportes de Conductores

Richard Peña Bonifaz

8 de enero de 2025

Dedicatoria

A nadie.

Reconocimientos

A mí.

Aquí está la versión corregida, usando el tiempo pasado, acorde a un trabajo ya realizado:

Resumen

El presente proyecto tuvo como propósito desarrollar una herramienta efectiva para analizar y predecir, con alta confiabilidad, el comportamiento del tráfico vehicular en la ciudad de Antofagasta mediante el uso de datos provenientes de la

plataforma Waze Cities. Waze, a través de su comunidad de usuarios, ofrece información en tiempo real que permitió obtener una visión detallada de los eventos de tráfico que ocurren en la ciudad. Esta investigación generó información relevante para la gestión del tráfico, facilitando la toma de decisiones por parte de las autoridades locales con el fin de mejorar la seguridad vial y optimizar la eficiencia del flujo vehicular. A través del análisis y explotación de estos datos, se identificaron patrones y tendencias que, integrados en la planificación urbana, contribuyeron a optimizar rutas críticas, reducir la congestión y minimizar la probabilidad de accidentes.

En este estudio se emplearon técnicas de análisis de datos y métodos geoespaciales mediante el uso de GeoPandas, además de la aplicación de series temporales para proporcionar visualizaciones claras y comprensibles dirigidas al usuario final. Adicionalmente, se implementaron técnicas de aprendizaje automático para realizar predicciones de tráfico, anticipándose a eventos y contribuyendo al desarrollo urbano eficiente y seguro (Barceló & Casas, 2005).

1. Introducción

1.1. Descripción del problema

Antofagasta, una ciudad con más de 106,000 vehículos en circulación (de Seguridad de Tránsito, 2023), enfrenta desafíos significativos en la gestión de su tráfico vehicular. Durante el año 2023, se registraron 1,715 accidentes, los cuales resultaron en 31 fallecidos y 102 heridos graves (de Seguridad de Tránsito, 2023). La infraestructura vial limitada, sumada a la alta concentración de vehículos en un número reducido de arterias principales, agravó la congestión y elevó el riesgo de accidentes. A pesar de la existencia de estos problemas, no se contaba con sistemas de monitoreo en tiempo real que permitieran gestionar el tráfico de manera

proactiva. Por ello, se aprovecharon fuentes de datos alternativas, como Waze, para recolectar información valiosa que facilitara la toma de decisiones en materia de tráfico (Chen et al., 2015).

1.2. Objetivo general

Realizar un análisis exhaustivo del comportamiento del tráfico en la ciudad de Antofagasta basado en los eventos reportados por los conductores en la plataforma Waze. El objetivo final fue generar información relevante que contribuyera a la gestión eficiente del tráfico, mejorando la seguridad vial y optimizando el flujo vehicular.

1.3. Objetivos específicos

- Obtener datos suficientes y representativos sobre los eventos de tráfico en Antofagasta mediante la API de Waze Cities.
- Realizar un análisis descriptivo de los datos recolectados para identificar patrones y tendencias relevantes en el comportamiento del tráfico.
- Identificar los factores clave que influyen en la seguridad vial y la eficiencia del tráfico en la ciudad.
- Presentar información visualmente comprensible y útil para las autoridades de gestión vial, facilitando la implementación de políticas y acciones basadas en datos (Auld & Mohammadian, 2009).
- Utilizar los datos disponibles para desarrollar modelos predictivos de tráfico que permitieran anticipar eventos y tomar decisiones proactivas en la gestión vial.

2. Marco teórico

El tráfico vehicular en entornos urbanos presenta un comportamiento complejo e impredecible, lo que dificulta su gestión eficiente. No obstante, el avance de las tecnologías móviles y la popularidad de aplicaciones como Waze permiten disponer de datos en tiempo real generados por los propios usuarios. Este proyecto se apoyó en técnicas de análisis de datos y aprendizaje de máquinas (Machine Learning) para convertir esta información en herramientas útiles para la gestión vial. La utilización de datos geoespaciales, junto con la automatización de los procesos de recolección, análisis y visualización, constituyó una solución costo-efectiva para mejorar la planificación del tráfico (Barceló & Casas, 2005).

2.1. Metodología

Una de las principales limitaciones en el análisis de fenómenos es la ausencia de datos suficientes y debidamente estructurados. Por ello, se diseñó una estrategia de recolección de datos que garantizara conclusiones con un nivel adecuado de certeza. La API de Waze Cities, que proporciona información en tiempo real sobre eventos activos, fue la fuente principal de datos. Para lograr un volumen de datos representativo, se implementó un servidor encargado de recopilar y almacenar esta información de manera continua.

Se realizó un análisis geoespacial con el objetivo de identificar puntos críticos, como vías principales, calles secundarias y zonas de alto tráfico. Este análisis se llevó a cabo utilizando GeoPandas, una herramienta de Python especializada en operaciones geoespaciales. Además, se analizaron series temporales para detectar patrones y tendencias del tráfico, identificando estacionalidades en el comportamiento. Los resultados se presentaron mediante técnicas de visualización que permitieron interpretar las tendencias y puntos de interés de manera efectiva.

El pipeline de datos incluyó una base de datos relacional para almacenar los datos, un ETL (Extract, Transform, Load) para procesarlos y un servidor web para visualizarlos. Se utilizó PostgreSQL como base de datos, Apache Airflow para la orquestación de tareas y Dash para la creación de visualizaciones interactivas. Este enfoque permitió automatizar el flujo de datos y garantizar la actualización constante de la información.

Adicionalmente, se entrenó un modelo de clasificación para determinar la probabilidad de ocurrencia de accidentes en diferentes puntos de la ciudad. El modelo seleccionado fue XGBoost, el cual fue seleccionado utilizando técnicas de validación cruzada y optimización de hiperparámetros. Los resultados obtenidos permitieron identificar las variables más influyentes en la ocurrencia de accidentes, así como la probabilidad de ocurrencia en diferentes condiciones de tráfico. Para la selección de variables se utilizó GridSearchCV, una técnica de búsqueda de hiperparámetros que permite encontrar la mejor combinación de variables para el modelo, se compararon modelos de regresión logística, árboles de decisión y XGBoost, siendo este último el que presentó el mejor desempeño en términos de precisión y sensibilidad.

Para la gestión del modelo, en cuanto a su implementación, mantenimiento y actualización, se utilizó la herramienta MLflow, que permite gestionar el ciclo de vida de los modelos de aprendizaje automático, desde su entrenamiento hasta su despliegue en producción y versionado.

En la parte de backend, para las diferentes solicitudes y eventos que se generan en la aplicación, se utilizó Flask, un framework de Python que permite crear APIs sencillas, con esto se generó la capacidad de poder responder a eventos generados desde Airflow para la actualización de los datos, renderizado, y actualización de los modelos.

La perspectiva general del flujo de datos se muestra en la Figura 1.

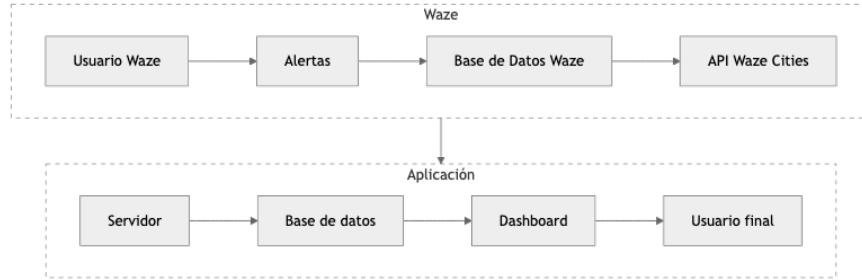


Figura 1: Pipeline general de datos

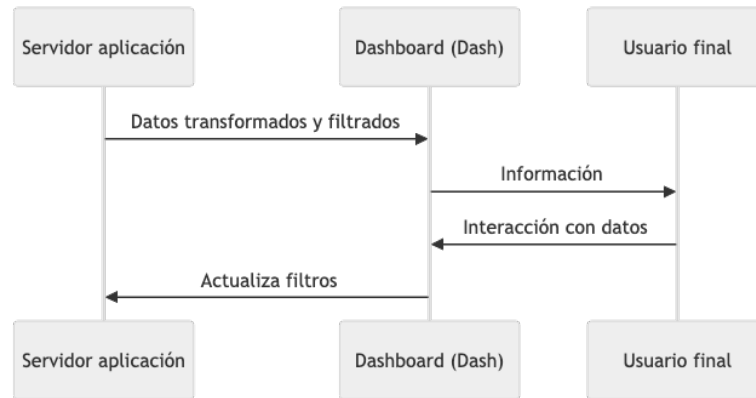


Figura 2: Flujo de trabajo de dashboard

Referencias

- Auld, J., & Mohammadian, A. (2009). Framework for modeling activity planning as a sequence of planned activities. *Transportation Research Part B: Methodological*, 43(6), 924-937.
- Barceló, J., & Casas, J. (2005). Dynamic network simulation with AIMSUN. *Simulation approaches in transportation analysis*, 57-98.
- Chen, C., Zhang, J., & Antoniou, C. (2015). Data-driven approaches to traffic prediction and management. *Transportation Research Procedia*, 10, 12-24.

de Seguridad de Tránsito, C. N. (2023). Informe de accidentes de tránsito [Antofagasta, Chile.]. <https://www.conaset.cl>