

SEMÁFORO INTELIGENTE AUTÓNOMO  
TRABAJO DE FIN DE MÁSTER  
MASTER IN BIG DATA AND BUSINESS INTELLIGENCE

Esmeralda Arribas Calderón

Gisela María Gpe Bernal Avila

Judeh Samir Saba Ibarra

Director del Máster: Belarmino García

Tutor del TFM: Miguel Couto

Madrid, a 09 de Junio de 2025

Abstract

El crecimiento acelerado de la población urbana y el aumento de la congestión vehicular plantean importantes desafíos para la movilidad en las ciudades. Los sistemas de semáforos tradicionales, con tiempos fijos o controlados de forma reactiva, a menudo no logran optimizar el flujo del tráfico, lo que genera mayores tiempos de viaje, incremento en el consumo de combustible y un aumento en las emisiones de CO<sub>2</sub>. Adicionalmente, recientemente junto con la falta de electricidad en la totalidad de la península ibérica, en este contexto, los semáforos inteligentes autónomos, impulsados por Inteligencia Artificial (IA) y Big Data, surgen como una solución prometedora para mejorar la eficiencia del tráfico, la seguridad vial y la sostenibilidad en entornos urbanos.

Este estudio realiza una revisión bibliográfica y un análisis de factibilidad para el desarrollo de un sistema autónomo de semáforos basado en IA. La investigación identifica tecnologías clave (aprendizaje automático, visión por computadora, edge computing), bases de datos públicas de tráfico. Asimismo, examina los desafíos técnicos, legales y éticos, como la fiabilidad de los sensores, la privacidad de los datos y la equidad en la priorización del tráfico.

Se desarrolla un modelo de simulación en Python y PyGame para probar ajustes dinámicos en los semáforos según el flujo vehicular y peatonal en tiempo real. El sistema integra sensores basados en Raspberry Pi, detección de objetos con YOLO y datos abiertos de tráfico de Madrid para optimizar los tiempos de las señales.

El sistema propuesto en este trabajo, demuestra viabilidad técnica. El estudio concluye que los semáforos inteligentes pueden mejorar significativamente la movilidad urbana, sin embargo requieren mayor investigación.

INDICE

Introducción

En este capítulo introductorio se realiza un estudio de semáforos inteligentes autónomos, se analizan las oportunidades y desafíos para el proyecto, se plantean una serie de objetivos y una metodología matriz a seguir en su desarrollo.

Como se especifica en el Abstract, el problema de semáforos con tiempos fijos, provoca viajes con tiempos más largos, mayor consumo de combustible y un aumento de las emisiones de CO<sub>2</sub>. En este contexto, los semáforos inteligentes que utilizan inteligencia artificial (IA) y big data prometen la eficiencia del tráfico, la seguridad vial y la sostenibilidad en entornos urbanos. El mercado de la IA en el transporte se estima que alcanzará los 4,11 billones de dólares en 2024, lo que demuestra la creciente importancia económica de este campo.

Este trabajo presenta un estudio de la literatura y las bases de datos relevantes para el desarrollo de un modelo de semáforos inteligentes autónomos. El objetivo principal es identificar las tecnologías, los desafíos y las oportunidades de investigación en este campo, con el fin de crear un sistema que se gestione por sí mismo de manera eficiente y autónoma.

Antecedentes - Bases de Investigación

Semáforos Inteligentes e IA en la Gestión del Tráfico

La investigación en semáforos inteligentes se ha centrado en el desarrollo de sistemas que se adaptan dinámicamente a las condiciones del tráfico en tiempo real. Diversos enfoques, como la lógica difusa, el aprendizaje automático y las redes neuronales.

Estas soluciones garantizan el poder manejar grandes conjuntos de datos, que abarcan patrones de tráfico históricos, condiciones climáticas y eventos especiales como accidentes, para predecir escenarios de tráfico futuros. Esta capacidad predictiva permite a las autoridades de transporte implementar medidas proactivas, como ajustar los tiempos de las señales o desplegar recursos adicionales, para optimizar los tiempos de los semáforos y mejorar el flujo del tráfico.

Se están llevando a cabo varios proyectos de investigación de la aplicación de la IA en la gestión del tráfico en universidades y empresas, entre los cuales destacan los siguientes:

Surtrac: Proyecto de semáforos inteligentes de la Universidad Carnegie Mellon (Pittsburg,USA) [1]: Este proyecto piloto ha demostrado que los semáforos inteligentes pueden reducir significativamente las emisiones de los vehículos y los tiempos de viaje. El proyecto se centra en una tecnología de control de señales de tráfico adaptativo que permite que el tráfico se comuniquen entre sí y se adapten de forma colaborativa a las condiciones reales del tráfico en tiempo real.

City Brain: Sistema de gestión de tráfico de IA desarrollado por Alibaba (Hangzhou, China) [2]: El sistema utiliza el aprendizaje automático y la visión artificial para monitorear las condiciones del tráfico y gestionar el control de las señales. Con el sistema los vehículos pueden circular a velocidades algo más elevadas (+15%). Además, en Hangzhou se ha mejorado la precisión en la detección de incidentes en un 92%. Esto ha hecho que los tiempos de respuesta se reduzca notablemente en unos 3 minutos de media y un 50% de probabilidades de tras el accidente.

RHODESNG con IntellidriveSM: Proyecto basado en Modelos de IA [3]: Este proyecto de la Administración Federal de Carreteras de EE. UU. se centra en el desarrollo de una arquitectura y algoritmos de próxima generación para sistemas de control de tráfico que se adaptan continuamente a las condiciones cambiantes del tráfico a través de comunicaciones de alta velocidad con vehículos e infraestructura. El sistema utiliza algoritmos autoadaptativos para integrar los datos de posición, velocidad y cola recibidos de los vehículos y los sensores y transmisores de la infraestructura, realizar cálculos precisos de alta velocidad, hacer predicciones y ajustar continuamente sus parámetros críticos en función de los datos entrantes.

HumanLight [4]: Este proyecto de la Universidad de California en Berkeley utiliza la IA para priorizar los vehículos de alta ocupación en las intersecciones. El objetivo es optimizar el control de las señales de tráfico para maximizar el rendimiento de las personas, en lugar del rendimiento de los vehículos. HumanLight utiliza una técnica de IA, el aprendizaje por refuerzo, para gestionar el comportamiento dinámico de entornos de tráfico complejos.

Proyecto KI4LSA y KI4PED del Instituto IOSB [5]: Este proyecto utiliza IA para permitir el control inteligente de los semáforos con el fin de reducir los atascos de tráfico y mejorar la seguridad de los peatones en los cruces. El proyecto utiliza cámaras de alta resolución y sensores de radar para captar la situación real del tráfico con mayor precisión, lo que permite determinar con exactitud el número de vehículos que esperan en un cruce en tiempo real.

Una vez habiendo analizado diferentes tipologías de sistemas y la viabilidad de las mejoras implementadas, se realiza un estudio de fuentes de datos públicas con distintos formatos

### Bases de Datos Públicas

Las bases de datos encontradas han sido clasificadas en tres categorías: información de tráfico en tiempo real, patrones de movilidad, accidentes y condiciones climáticas.

#### Información de Tráfico en Tiempo Real

Existen varias bases de datos públicas que proporcionan información sobre el tráfico en tiempo real. Las más relevantes se muestran en la Tabla 1.

Tabla 1: Tipos de bases de datos de tráfico en tiempo real

#### Patrones de Movilidad

Las bases de datos, presentadas en la Tabla 2 ofrecen información valiosa sobre los patrones de movilidad.

Tabla 2: Tipos de bases de datos de patrones de movilidad

#### Accidentes y Condiciones Climáticas

Para obtener información sobre accidentes y condiciones climáticas, se pueden consultar las bases de datos recogidas en la Tabla 3.

Tabla 3: Tipos de bases de datos de accidentes y clima.

Teniendo en cuenta los casos de éxito previamente descritos, se plantea que el sistema para el control inteligente de semáforos debe combinar elementos de software y de hardware. Por un lado, la extracción de datos reales mediante elementos como sensores o cámaras. Por otro, el preprocesamiento de los datos, el procesamiento de estos y el algoritmo para predecir las condiciones del tráfico y los tiempos de los semáforos mediante algoritmos de aprendizaje automático. Por ejemplo mediante el empleo de random forest. Se propone que el programa siga una lógica de escalera donde clasifica el tráfico en volúmenes bajos, medios, altos y muy altos, y en base a esto, se modifica el flujo de tráfico.

#### Justificación

Los sistemas tradicionales, los cuales suelen operar con temporizadores fijos o sensores básicos sin capacidad de adaptación en tiempo real, generan congestión al no responder dinámicamente a las fluctuaciones del tráfico. Mientras que las nuevas tecnologías vienen para cambiar el paradigma actual. de un sistema de control inteligente de semáforos basado en tecnologías modernas de Inteligencia Artificial (IA) y Big Data permite una gestión dinámica capaz de responder a incidentes en tiempo real, patrones estacionales e incluso eventos imprevistos, maximizando la eficiencia de la red vial.

Además de la reducción de los atascos, también disminuyen el tiempo de viaje, el consumo de combustible y las emisiones de CO<sub>2</sub>, contribuyendo así a una ciudad más sostenible y respetuosa con el medio ambiente y reforzando el compromiso con los Objetivos de Desarrollo Sostenible (ODS) [6].

Actualmente, las ciudades líderes en movilidad inteligente, como Singapur, algunas metrópolis de EE.UU. (e.g., Los Ángeles, Pittsburgh) y China (e.g., Hangzhou, Shenzhen), ya han implementado sistemas de semáforos inteligentes con resultados demostrables: reducciones de hasta un 30% en los tiempos de desplazamiento y un 20% en emisiones contaminantes. Se selecciona Madrid, como capital innovadora, con potencial de posicionarse a la vanguardia de las smart cities europeas adoptando estas tecnologías, en línea con los objetivos de la Agenda 2030 y el Pacto Verde Europeo.

Este proyecto no solo mejoraría la calidad de vida de los ciudadanos, sino que también reforzaría el compromiso de Madrid con la digitalización y la sostenibilidad, transformando su infraestructura urbana en un referente global de eficiencia y reducción de la huella de carbono. La inversión en IA para la gestión del tráfico no es solo una apuesta por la innovación, sino una necesidad en un mundo donde la movilidad eficiente y ecológica es clave para el futuro de las ciudades.

En un contexto global donde la competencia entre ciudades por atraer talento e inversión se centra en la digitalización y la sostenibilidad, Madrid no puede permitirse quedarse atrás. Implementar un sistema de semáforos inteligentes no solo mejoraría la movilidad, sino que posicionaría a la capital española como líder en la transformación hacia la smart city [7], siguiendo los ejemplos donde estas tecnologías ya han demostrado su valor económico y ecológico.

**Problema, Hipótesis y Objetivos**

El objetivo principal es contribuir al desarrollo de un sistema de semáforos inteligentes autónomo que sea seguro, eficiente, sostenible y ético, y que pueda mejorar la movilidad urbana en las ciudades del futuro. En relación a esto, se desarrollan objetivos específicos, los cuales se describen a continuación.

El desarrollo de semáforos inteligentes surge como respuesta a los graves problemas de congestión vehicular y peatonal en las ciudades modernas, donde los sistemas tradicionales de control de tráfico resultan inflexibles e ineficientes para manejar el flujo dinámico, lo que motiva la creación de un sistema adaptativo que optimice la movilidad urbana mediante IA y sensores en tiempo real.

La dependencia de los sistemas actuales de infraestructuras energéticas centralizadas, que los hace vulnerables ante fallos o desastres naturales, impulsa el objetivo de diseñar semáforos autosuficientes con paneles solares y baterías de respaldo, garantizando operatividad continua incluso en situaciones adversas.

Las altas emisiones de carbono generadas por el tráfico detenido y los motores en ralentí obligan a implementar una solución que reduzca activamente la contaminación mediante una gestión más fluida del tránsito, alineándose con los Objetivos de Desarrollo Sostenible (ODS) de ciudades sostenibles y energía limpia.

La limitación de los sistemas de tráfico convencionales, diseñados para ubicaciones específicas sin capacidad de adaptación a otras ciudades, plantea la necesidad de desarrollar un modelo escalable basado en parámetros ajustables y arquitecturas de IA transferibles, facilitando su replicación en distintas zonas geográficas.

La subutilización de los grandes volúmenes de datos recopilados por sensores urbanos, que no son aprovechados para una toma de decisiones inteligente, justifica la integración de técnicas de big data y análisis predictivo que permitan identificar patrones de congestión y optimizar rutas de forma proactiva.

Finalmente, la opacidad de muchos sistemas de IA, que operan como "cajas negras" sin posibilidad de auditar sus decisiones o detectar sesgos, exige que este proyecto priorice la transparencia mediante herramientas de explicabilidad y métricas de equidad, asegurando que los

beneficios del sistema sean distribuidos de forma justa entre todos los distritos.

#### Metodología

La metodología integra tres pilares fundamentales: Data Science para el modelado predictivo, Data Analytics para el procesamiento de información e Ingeniería de Datos para la arquitectura técnica, utilizando un conjunto robusto de herramientas que incluye Excel para manipulación básica, SQL para gestión de bases de datos, y Python con sus librerías especializadas (Pandas, Matplotlib, Seaborn) para análisis avanzado y visualización.

El proceso de análisis del proyecto sigue una metodología estructurada basada en el ciclo estándar de Data Analytics [8], diseñado para transformar datos en insights accionables. En la fase inicial ASK (Preguntar), se enfoca en definir con precisión los objetivos estratégicos y formular las preguntas clave que guiarán toda la investigación, tales como "¿Cómo optimizar los tiempos de los semáforos para reducir las congestiones vehiculares en horas pico?" o "¿De qué manera se puede minimizar las emisiones de CO<sub>2</sub> mediante una gestión inteligente del tráfico?". Este planteamiento fundamental establece el marco de referencia para todas las etapas posteriores.

La etapa PREPARE (Preparar) implica un trabajo meticuloso de recolección y evaluación de datos. El proyecto se basa en fuentes oficiales como el Ayuntamiento de Madrid para obtener datos de tráfico en tiempo real, y en la AEMET para información meteorológica detallada. Este proceso incluye una inspección minuciosa de la estructura de los datos, verificación de su integridad, y un filtrado riguroso para seleccionar únicamente la información relevante para los objetivos marcados anteriormente. La preparación de los datos contempla el diseño de una estructura de base de datos relacional en SQL Server que integra múltiples tablas interconectadas, permitiendo cruces analíticos sofisticados entre variables de tráfico, climatológicas y de accidentabilidad.

El core del proceso reside en la fase PROCESS (Procesar), donde se realiza una transformación profunda de los datos crudos. Esto incluye operaciones de limpieza para manejar valores faltantes o inconsistentes, normalización de formatos, y creación de variables derivadas mediante agregaciones estadísticas (cálculo de promedios móviles, identificación de valores máximos y mínimos por franja horaria, entre otros). Se implementan técnicas de enriquecimiento de datos para combinar múltiples fuentes y generar un conjunto unificado listo para el análisis.

En la fase ANALYZE (Analizar), se aplica un enfoque multidimensional que combina cuatro tipos de análisis complementarios. El análisis descriptivo proporciona una comprensión básica de las tendencias históricas, mientras que el diagnóstico profundiza en las relaciones causales subyacentes. La componente predictiva, apoyada en técnicas de inteligencia artificial, permite anticipar escenarios futuros, y el análisis prescriptivo deriva recomendaciones concretas para la optimización de los tiempos semaforicos. La visualización de resultados se realiza principalmente mediante librerías de Python, generando tanto gráficos estáticos como representaciones dinámicas que facilitan la comparación entre la Gran Vía y otras zonas de la ciudad, incluyendo mapas de calor para accidentes y series temporales para flujos vehiculares. Esta aproximación metodológica integral asegura que la solución desarrollada sea tanto técnicamente robusta como socialmente responsable.

El ciclo culmina con la etapa VISUALIZE/SHARE (Comunicar), donde se transforman los hallazgos complejos en representaciones accesibles. Se desarrollan visualizaciones estáticas creadas en Python (Matplotlib/Seaborn) para comunicar resultados específicos. Este componente es crucial para garantizar que los insights derivados del análisis sean comprendidos y aplicados efectivamente en la implementación del sistema de semáforos inteligentes.

#### Timeline (planificación)

A continuación se muestra la planificación desarrollada incluyendo su respectivo diagrama de Gantt.

Fase 1: Planificación y Revisión (Febrero - Marzo)

Febrero 26 - Marzo 7: Definir objetivos específicos y alcance del proyecto.

Delimitar el enfoque en alguna calle o av. como Gran Vía y la interacción con vehículos autónomos.

Identificar las variables a considerar (volumen de tráfico, tipos de vehículos, peatones, etc.).

Formular preguntas de investigación.

Marzo 8 - Marzo 21: Revisión exhaustiva de la literatura.

Buscar artículos, proyectos e informes relevantes sobre semáforos inteligentes, IA en la gestión del tráfico y big data en movilidad urbana.

Profundizar en los proyectos KI4LSA y HumanLight como referencia.

Contactar a instituciones como el Centro de Investigación del Transporte (TRANSyT) o el Instituto Universitario de Investigación del Automóvil (INSIA) para obtener información y asesoría.

Marzo 22 - Marzo 31: Selección de bases de datos y herramientas.

Identificar las bases de datos públicas con información relevante

Evaluar herramientas de simulación de tráfico como SUMO y plataformas de análisis de datos.

Evaluar la tecnología que vamos a usar FML.

Fase 2: Desarrollo y Simulación (Abril - Mayo)

Abril 1 - Abril 15: Diseño del modelo de semáforo inteligente.

Definir la arquitectura del sistema y los algoritmos de IA a utilizar.

Considerar la integración con vehículos autónomos y la optimización multimodal.

Abril 16 - Mayo 2: Implementación del modelo en la herramienta de simulación.

Configurar el entorno de simulación con las características de la Gran Vía.

Integrar los datos de tráfico y movilidad obtenidos de las bases de datos.

Mayo 3 - Mayo 16: Ejecutar simulaciones y analizar resultados.

Figura 1: Diagrama de Gantt

Marco teórico

En este capítulo, se hace una introducción a las tecnologías de gestión de bases de datos, al machine learning y a la visión artificial, conocimientos esenciales para la comprensión del desarrollo del proyecto.

Tecnologías de Gestión de Bases de Datos

Es importante mencionar la división de tipos de bases de datos que existen ya que de aquí parte la decisión de depositar la información dependiendo del origen de la información.

Las bases de datos no relacionales (NoSQL) representan un paradigma alternativo al modelo relacional tradicional, diseñado específicamente para manejar grandes volúmenes de datos no estructurados o semiestructurados con esquemas flexibles. A diferencia de las bases SQL que usan tablas rígidas, los sistemas NoSQL emplean diversos modelos de datos como documentos (JSON/BSON en MongoDB), pares clave-valor (Redis), columnas anchas (Cassandra) o grafos (Neo4j), permitiendo escalabilidad horizontal mediante arquitecturas distribuidas. Estas bases sacrifican algunas propiedades ACID (como consistencia fuerte) en favor del teorema CAP, donde priorizan disponibilidad y tolerancia a particiones, implementando eventual consistency mediante mecanismos como quórum de escritura o replicación multi-nodo. Su arquitectura sin esquema fijo (schema-on-read) las hace ideales para big data, IoT o aplicaciones web escalables, donde el volumen, velocidad y variedad de datos superan las capacidades de las bases relacionales tradicionales.

Por otro lado, las bases de datos relacionales como SQL Server son sistemas fundamentados en el modelo relacional que organizan la información en tablas interconectadas mediante relaciones lógicas. Estas tablas almacenan datos en estructuras bidimensionales compuestas por filas (registros) y columnas (campos), donde cada tabla representa una entidad específica (como clientes o productos) y las relaciones entre ellas se establecen mediante claves primarias y foráneas. SQL Server implementa este modelo con un estricto esquema predefinido (schema-on-write) que exige declarar la estructura de datos antes de su uso, garantizando consistencia e integridad.

El núcleo de SQL Server se basa en el motor de base de datos (SQLOS), que incluye componentes críticos como el Query Optimizer para generar planes de ejecución eficientes, el Storage Engine que gestiona el almacenamiento físico en páginas de 8KB (archivos .mdf para datos y .ldf para logs), y el Buffer Pool que cachea datos en memoria RAM para optimizar el rendimiento. La arquitectura sigue una jerarquía que va desde la instancia (que puede alojar múltiples bases de datos) hasta los filegroups y archivos físicos, permitiendo una administración granular de los recursos de almacenamiento.

SQL Server soporta el lenguaje SQL estándar con extensiones propietarias, dividiendo sus operaciones en tres categorías principales: DDL (para definir estructuras como tablas e índices), DML (para manipular datos con consultas SELECT, INSERT, etc.) y DCL (para gestionar permisos de acceso). Un aspecto clave es su cumplimiento de las propiedades ACID, donde destaca el uso de Write-Ahead Logging (WAL) para garantizar durabilidad, y mecanismos como los niveles de aislamiento (READ COMMITTED, SERIALIZABLE) para manejar transacciones concurrentes sin comprometer la consistencia.

Entre sus ventajas técnicas destacan su alto rendimiento para cargas OLTP, gracias a índices clustered y columnstore, y funciones avanzadas como AlwaysOn para alta disponibilidad. Sin embargo, presenta limitaciones en escalabilidad horizontal frente a alternativas NoSQL, y su modelo de licenciamiento por núcleo puede resultar costoso para implementaciones a gran escala. En proyectos como el sistema de semáforos inteligentes, SQL Server sería ideal para modelar datos estructurados de tráfico, ejecutar consultas complejas con joins entre tablas de sensores e intersecciones, e integrarse con herramientas de análisis como Power BI, aunque requeriría una planificación cuidadosa de recursos para manejar picos de carga eficientemente.

#### Machine Learning

En esta sección se va a hacer una breve introducción a los principales conceptos relacionados con la inteligencia artificial (IA) necesarios para comprender parte de la solución adquirida.

Dentro del campo de la inteligencia artificial se encuentra la rama denominada Machine Learning [9]. El aprendizaje automático (ML) consiste en la creación de modelos matemáticos que permiten a las máquinas aprender de los datos sin necesidad de ser explícitamente programadas. Por ejemplo, una red neuronal es uno de los algoritmos más conocidos y utilizados en el ámbito del machine learning. Al procesar grandes volúmenes de datos, estas máquinas pueden identificar patrones y hacer predicciones.

La estadística juega un papel fundamental en el ML, ya que es esencial para entender los datos empleados en el entrenamiento de los modelos y para interpretar los resultados obtenidos.

Figura 2: Esquema de lo que engloba Inteligencia Artificial

En este proyecto se ha empleado la técnica denominada Random Forest [10]. Un Random Forest en regresión es un modelo de aprendizaje automático que se utiliza para predecir valores numéricos continuos. Se basa en la construcción de múltiples árboles de decisión, cada uno entrenado con diferentes subconjuntos aleatorios de los datos y características. A diferencia de un solo árbol de decisión, que puede ser muy sensible a pequeñas variaciones en los datos, el Random Forest combina las predicciones de todos los árboles para obtener un resultado más robusto y preciso. En el caso de la regresión, la predicción final se obtiene promediando las salidas de todos los árboles del conjunto. Esta técnica ayuda a reducir el sobreajuste y mejora la capacidad de generalización del modelo, lo que la hace especialmente útil en problemas complejos con muchos datos y relaciones no lineales.

#### Visión artificial

En esta sección se hace una breve descripción de lo que es un sistema de visión artificial y sus etapas.

Un sistema de visión artificial es la unión de hardware y software que posibilita la captura y el procesamiento de datos de imágenes. Es decir, es el sistema que permite que las máquinas puedan emular la vista humana. El programa es capaz de procesar, analizar y medir información para controlar los procesos y tomar las decisiones correspondientes. En un sistema, se pueden diferenciar dos fases principales: obtención de imágenes y el análisis de estas. A continuación se van a explicar en detalle.

A pesar de contar con dos fases principales, la de análisis se puede dividir en otras cuatro categorías más. De esta forma, un ciclo completo de un sistema de visión consta de los siguientes pasos:

Adquisición de imágenes (etapa sensorial), consiste en conseguir una imagen. Consiste en emplear sensores y tener la capacidad para digitalizar la señal producida por el sensor (imagen). Transformaciones y filtrado de imágenes (etapa de preprocesado) para arreglar en la imagen las imperfecciones generadas o no suprimidas por el hardware con operaciones de mejora de la imagen. Otro de los objetivos es acondicionar la imagen en función del objetivo que haya. Las operaciones más importantes que se pueden aplicar a la imagen se puede dividir en tres apartados: operaciones básicas, transformaciones geométricas y algoritmos generadores de ruido y filtrado.

Aislamiento de los objetos de la imagen (etapa de segmentación) Segmentar la imagen es dividirla en zonas individualizadas y disjuntas, En otras palabras, consiste en separar las partes que la constituyen o los objetos que la forman y el fondo. Es importante conocer perfectamente los objetos puesto que la extracción de características dependerá en gran medida de la correcta detección de estos.

Cálculo de características (etapa de parametrización) se hace una selección de rasgos. Esto significa que en función de lo segmentado anteriormente se van a extraer unas características

concretas que definen el objeto: forma, textura, color, orientación, etc.

Reconocimiento e interpretación (etapa de clasificación) En base a los datos extraídos anteriormente se debe asignar una etiqueta y en función al resultado, interpretar y tomar una decisión o actuar en consecuencia.

Cuerpo central del Trabajo de campo/Desarrollo

En esta sección se describe el desarrollo principal del proyecto siendo este el nexo de unión entre los objetivos y el análisis de resultados del sistema.

En el contexto de la movilidad urbana inteligente, este trabajo se centra en optimizar los semáforos de la calle Gran Vía de Madrid - la zona más crítica de la ciudad por su alta densidad peatonal y vehicular - utilizando inteligencia artificial (IA) para mejorar tanto el flujo de peatones como la seguridad vial, priorizando a los transeúntes sin comprometer la eficiencia del tráfico. A través de un algoritmo basado en datos en tiempo real como densidad peatonal, volumen vehicular, eventos especiales y condiciones del entorno, se propone un sistema de semáforos autónomos que se adapte dinámicamente a las necesidades cambiantes de este emblemático corredor urbano.

La hipótesis plantea que, mediante un entrenamiento supervisado y la automatización integral comenzando por Gran Vía - para luego replicar el modelo en otras zonas - se puede reducir la congestión, minimizar tiempos de espera y crear un espacio más seguro e inclusivo. Bajo esta hipótesis, la imagen siguiente, muestra la distribución de ubicación de cámaras en todo Madrid (encontradas en la página del Ayuntamiento de Madrid), de las cuales, las que están marcadas con colores son las áreas de enfoque del desarrollo del trabajo.

Figura 3: Distribución de cámaras de la ciudad de Madrid

Factibilidad del Proyecto: Desafíos Técnicos, Legales y Éticos

La implementación de un modelo de semáforos inteligentes autónomos en un entorno real presenta varios desafíos:

Técnicos:

La integración de diferentes tecnologías de sensores y sistemas de IA.

La gestión de grandes volúmenes de datos en tiempo real.

La garantía de la fiabilidad y la seguridad del sistema.

Legales:

La responsabilidad en caso de accidentes.

La protección de la privacidad de los datos.

El cumplimiento de las normativas de tráfico.

Éticos:

La equidad en la asignación de tiempos de verde.

La discriminación potencial contra ciertos usuarios de la vía.

La transparencia en la toma de decisiones del sistema.

Los marcos éticos, como el utilitarismo y la deontología, pueden ayudar a guiar la toma de decisiones en los sistemas de semáforos autónomos. El utilitarismo se centra en maximizar el bien general, mientras que la deontología se basa en un conjunto de reglas predefinidas.

Se comenzó a trabajar en la detección de algunas variables que mientras está en desarrollo algunas irán cambiando, Hasta ahora, se detallan las siguientes:

Descripción del Flujo del Sistema de Monitoreo de Tráfico

A continuación se describe lo que sería el flujo del sistema inteligente de gestión de tráfico urbano que integra múltiples componentes tecnológicos para el monitoreo y control del flujo vehicular y peatonal.

El sistema comienza con la captura de datos a través de cámaras de tráfico que alimentan uno o varios modelos de procesamiento, los cuales analizan las condiciones del tráfico en tiempo real y generan información sobre diversos parámetros como el estado de los semáforos, la congestión vehicular y peatonal, además de detectar diferentes tipos de vehículos incluyendo buses, motocicletas, camiones, ambulancias y coches.

La información procesada se estructura en diferentes categorías de datos que incluyen identificadores de cámara con detalles sobre ubicación, congestión de tráfico y conteo de peatones y vehículos por tipo, así como datos del Ayuntamiento de Madrid relacionados con la congestión vehicular y peatonal. De esta manera, el sistema será capaz de determinar los tiempos de su estado actual (rojo, verde, amarillo), junto con datos de ubicación y referencia de cada semáforo en la red vial.

Adicionalmente, el sistema integra una red de sensores especializados que incluyen sensores de humedad, infrarrojos y pulsadores de emergencia, todos conectados a una CPU Raspberry Pi que procesa esta información adicional y genera estadísticas sobre el estado de la red de semáforos mediante análisis RGB. Todo este ecosistema tecnológico permite una gestión integral del tráfico urbano, combinando detección visual inteligente, monitoreo ambiental y control automatizado de la infraestructura vial para optimizar el flujo de tráfico y mejorar la seguridad ciudadana. En la figura de abajo se muestra el flujo.

Esquema 1: Flujo del sistema inteligente

Software basado en bases de datos del Ayuntamiento de Madrid

Fuente de Datos

Recurso disponible con Información Real: Portal de datos abiertos del Ayuntamiento de Madrid

Selección de Base de Datos Seleccionada y sus razones

Arquitectura de SQL Azure

Azure SQL Database [11] ofrece numerosas ventajas para proyectos de Big Data como el sistema de semáforos inteligentes de Madrid, destacando su perfecta integración con el ecosistema Microsoft. Al utilizar Power BI, permite conexiones nativas para visualizar datos de tráfico y métricas éticas en dashboards, mientras que Azure Machine Learning facilita el entrenamiento de modelos predictivos (como la previsión de congestión) directamente sobre los datos almacenados. Además, para una segunda fase del proyecto, Azure Data Factory simplifica la orquestación de procesos ETL para extraer información de sensores y cámaras. Su escalabilidad automatizada es ideal para ajustar recursos computacionales (DTUs) en horas pico, como durante eventos masivos en la ciudad.

En cuanto a seguridad, cumple con el GDPR mediante encriptación de datos en reposo y tránsito, crucial cuando se maneja información sensible de ciudadanos, e incluye auditorías integradas para garantizar transparencia en el acceso a los datos. También soporta formatos semiestructurados como JSON/XML, útil para datos de sensores o APIs, e incluso funciones de Graph DB para modelar relaciones complejas entre peatones, semáforos y calles.

Reduce la carga de mantenimiento gracias a parches automáticos gestionados por Microsoft y backups integrados que aseguran la recuperación ante fallos, como interrupciones en el sistema de semáforos.

En el esquema siguiente se muestra la base de datos diseñada para el cumplimiento de nuestros objetivos y procesamiento de predicciones detalladas en el análisis:

Esquema 2: Diagrama de la Base de Datos

Una vez definida, se comenzó a popular cada tabla con su correspondiente información proveniente de los CSV bajados del Web del Ayuntamiento. Debido a la seguridad de la base de Datos y a la longitud del CSV ya que contiene todos los dispositivos de Madrid, se hizo una limpieza y filtrado a éstos archivos(en las siguientes páginas se detallan). Al analizar profundamente el contenido Histórico de tráfico permanente se descubrió que la zona más concurrente y más transitada por el centro de la ciudad, es Gran Vía, por el cual a partir de aquí se decidió enfocar el filtrado a los Elementos Ids de la zona de Gran Vía (Anexo 8)

Análisis de Tráfico Vehicular

Debido a lo explicado anteriormente, filtrando la data para crear un semáforo inteligente en Gran Vía- Calle Clavel, se optó por usar los datos del tráfico de Gran Vía como macro entorno y Clavel como micro entorno. En ese sentido, esta tabla de "Autos\_traffic" contiene únicamente datos de Gran Vía y no de todo Madrid.

Figura 4. Evolución de la intensidad del Tráfico desde el 2019 en Gran Vía

Este gráfico muestra la evolución de la intensidad del tráfico a lo largo del tiempo desde 2019 hasta 2025, presentando tres series de datos: diaria (línea azul clara), semanal (línea naranja) y mensual (línea verde). Durante 2019, se observa una intensidad de tráfico relativamente alta y estable, con valores que oscilan entre 550 y 700 unidades, mostrando cierta variabilidad natural. El cambio más dramático ocurre a principios de 2020, cuando todas las series experimentan una caída abrupta y pronunciada, alcanzando mínimos históricos alrededor de 100-150 unidades de intensidad, lo que coincide temporalmente con el inicio de la pandemia de COVID-19 y las medidas de confinamiento implementadas globalmente.

Después de este mínimo histórico en 2020, el tráfico muestra una recuperación gradual pero estabilizándose en un nuevo nivel que se mantiene relativamente constante. Este nuevo nivel de equilibrio se sitúa aproximadamente entre 400 y 500 unidades de intensidad media,



permaneciendo por debajo de los niveles pre-pandémicos de 2019. Las tres series de datos muestran patrones similares, con las medias semanales y mensuales proporcionando una visión más suavizada de las tendencias que los datos diarios, que presentan mayor volatilidad y fluctuaciones naturales día a día.

Figura 5: Intensidad de Tráfico en los años 2023-2024

Este gráfico demuestra la regularidad diaria explicada en el gráfico anterior sólo habiendo una excepción en Agosto del 2023 que supone obras de mantenimiento o de peatonalización en esas fechas.

Mapa de calor en Calle Clavel

Figura 6. Mapa de calor en calle Clavel

Figura 7. Intensidad por año en calle Clavel

Figura 8. Intensidad Por hora, calle Clavel

Figura 9. Intensidad Por día de la semana, calle Clavel

Figura 10. Niveles de intensidad

Densidad de Tráfico en Tiempo Real en Madrid, Gran Vía y Clavel (en ambos sentidos de dirección)

Figura 11. Distribución de Intensidad del Tráfico

Intensidad de Tráfico en Tiempo Real en Madrid, Gran Vía y Clavel (en ambos sentidos de dirección)

Figura 12. Comparativo intensidad de Tráfico por Localización

BBDD tráfico peatonal

Figura 13. Flujo Peatonal en Gran Vía por dirección de movimiento

PROCESAMIENTO DE NA'S (Nulos)

Figura 14. Total de Peatones por mes y año

Figura 15. Total de Peatones por distrito

Figura 16. Heatmap de peatones: Día de la semana vs Hora del día

Este gráfico ilustra la densidad de tránsito peatonal a lo largo de la semana y las horas del día. La intensidad del color refleja la concentración de peatones, como las horas pico de días laborables (por ejemplo, entre las 8-9 a.m. y las 5-7 p.m.) son de mayor aglomeración, mientras que las madrugadas (0-5 a.m.) muestran una actividad mínima. Los fines de semana (Sábado y Domingo) presentan un patrón distinto, con mayor movimiento durante las tardes o noches debido a actividades recreativas. Esta visualización resulta valiosa para identificar tendencias de movilidad urbana, optimizar recursos en transporte o comercios, y planificar infraestructuras peatonales según la demanda horaria y semanal.

Análisis de mayor flujo peatonal

Figura 17. Mayor flujo peatonal en zona centro

BBDD bicicletas

Figura 18. Distribución de bicicletas por hora del día

La imagen representa la cantidad de bicicletas utilizadas o disponibles en diferentes horas del día. El gráfico probablemente ilustra patrones de uso, como picos durante las horas de mayor actividad (como las horas pico de la mañana o la tarde) y disminuciones en períodos de menor demanda (como la madrugada). Este tipo de visualización es útil para analizar tendencias y optimizar servicios relacionados con bicicletas, como sistemas de alquiler o infraestructura ciclista.

Figura 19. Bicicletas por Distrito

El distrito con mayor uso de Bicicletas es Latina, superando ampliamente las 950,000 bicicletas siguiéndole Chamberí por ser una de las zonas más céntricas.

BBDD accidentalidad

Figura 20. Mapa de calor de accidentes en Madrid

Como se ve en el mapa de calor ésta representa la distribución geográfica de los accidentes coincidiendo que, una de las áreas de mayor concentración de Accidentes (rojo intenso) es en la calle de Gran Vía.

Figura 21: Accidentes por hora en Madrid

Figura 22: Accidentes por Ubicación en Gran Vía

Figura 23: Accidentes por Año en Gran Vía

Figura 24. Accidentes por Hora Gran Vía

Figura 25. Densidad de accidentes por Hora en calle Clavel

Figura 26. Tipos de Accidentes en calle Clavel

Figura 27. Diferentes variables en Accidentes de calle Clavel

#### Modelo predictivo

Tras analizar los datos y obteniendo las conclusiones descritas anteriormente

El modelo predictivo de tráfico propuesto utiliza el algoritmo Random Forest para abordar dos objetivos principales: la clasificación del nivel de servicio del tráfico y la predicción de valores numéricos como la intensidad vehicular, la ocupación y la carga. Este enfoque dual permite no solo entender el estado general del tráfico, sino también obtener estimaciones precisas que son fundamentales para la optimización de semáforos. El Random Forest es especialmente adecuado para este problema debido a su capacidad para manejar múltiples variables, su resistencia al sobreajuste y su habilidad para trabajar con datos incompletos o con ruido. Además, proporciona información valiosa sobre la importancia de cada variable, lo que ayuda a identificar los factores que más influyen en el flujo del tráfico.

Las variables de entrada incluyen datos temporales como la hora del día, el día de la semana y el mes, así como información específica del tráfico como la intensidad, la ocupación, la carga y la velocidad media. También se consideran variables contextuales, como el tipo de vía y el distrito, que pueden ser determinantes en el comportamiento del tráfico. Antes de entrenar el modelo, es crucial realizar un preprocesamiento de los datos, que implica la conversión de columnas como intensidad y ocupación a un formato numérico, manejar valores faltantes o erróneos, y crear características derivadas como promedios móviles o valores de horas anteriores. Este paso asegura que los datos estén en un formato adecuado para el análisis.

Una vez preparados los datos, el modelo de clasificación predice el nivel de servicio del tráfico, categorizando en flujo, lento, con retenciones o congestionado. Paralelamente, el modelo de regresión estima valores numéricos exactos para la intensidad, la ocupación y la carga, lo que permite un ajuste más fino en la gestión de los semáforos. Estas predicciones se integran en un sistema de reglas que determina el tiempo óptimo de luz verde en los semáforos. Por ejemplo, en condiciones de tráfico fluido, se mantienen los tiempos estándar, mientras que en situaciones de congestión, se priorizan tiempos extendidos para aliviar la saturación.

La implementación del modelo en Python abarca desde la carga y limpieza de los datos hasta el entrenamiento y evaluación del modelo, utilizando herramientas como sklearn para el aprendizaje automático y matplotlib para la visualización de resultados. Para futuras mejoras, se puede incorporar datos externos como condiciones meteorológicas o eventos especiales, lo que podría aumentar la precisión del modelo. También se podría explorar el uso de modelos de series temporales avanzadas, como las LSTM, para capturar patrones temporales más complejos. Este enfoque integral no solo predice el estado del tráfico, sino que también sugiere acciones concretas para optimizar la movilidad urbana, adaptándose dinámicamente a las condiciones en tiempo real.

#### Software basado en visión computacional

En este apartado se describen los tres principales códigos empleados para el desarrollo del sistema: obtención de datos, análisis de datos y simulación de funcionamiento del sistema.

##### Obtención de imágenes reales (Web scraping)

En esta sección se describe el código que se ha empleado para la obtención de las imágenes del Ayuntamiento de Madrid. Se ha empleado una web del Ayuntamiento denominada Informo [18] y la técnica empleada es Web Scraping.

El código implementa tres funciones principales que trabajan de forma coordinada:

``download_image()`, `zip_images()` y `main()``, cada una con un rol específico en el proceso de recolección de imágenes.

La función ``download_image()`` es el núcleo del sistema. Recibe como parámetro el ID de una cámara (como "01315" siendo esta la de especial interés) y construye la URL correspondiente añadiendo un parámetro ``rand`` basado en el timestamp actual para evitar caché. Utiliza la librería ``requests`` para descargar la imagen y la guarda localmente con un nombre estructurado como ``camera_01315_20250507_1415.jpg``, donde se incluye el ID de cámara, fecha (año, mes, día) y hora (hora, minutos). La marca temporal se ajusta al uso horario de Madrid sumando una hora al UTC.

La función ``zip_images()`` se encarga de consolidar todas las imágenes descargadas. Recorre el directorio ``camera_images`` y comprime su contenido en un archivo ZIP llamado ``camera_images.zip`` usando la librería ``zipfile``. Este paso es crucial para manejar eficientemente múltiples imágenes antes de su descarga masiva.

La función `main()` orquesta todo el proceso. Contiene la lista de cámaras (como "01315") y controla las iteraciones del sistema. En cada ciclo (5 por defecto), itera sobre todas las cámaras llamando a `download_image()` con una pausa de 2 segundos entre cada descarga para evitar saturación. Entre iteraciones completa espera exactamente 294 segundos (4 minutos y 54 segundos) para compensar el tiempo de ejecución y mantener intervalos precisos de 5 minutos. La sección de configuración inicial define variables clave como `base_url_template` con el formato de URL base y `output_dir` para el almacenamiento local. El bloque final del código verifica si es ejecutado como script principal para iniciar el proceso llamando a `main()`, que a su vez invoca `zip_images()` al terminar todas las iteraciones, generando el archivo comprimido listo para descargar mediante la integración con Google Colab.

#### Detección de vehículos y coches en imágenes

En este apartado se describe el código que se emplea para la detección de elementos en las imágenes extraídas del Ayuntamiento de Madrid.

El código implementa un sistema de detección y conteo de vehículos y peatones utilizando el modelo YOLO, estructurado en una función principal `detect_and_count()` que realiza todo el procesamiento de imágenes.

La función `detect_and_count()` recibe como parámetro la ruta de una imagen y sigue un flujo bien definido. Primero verifica la existencia del archivo usando `os.path.exists()` y carga la imagen con `cv2.imread()`. Si la carga es exitosa, redimensiona la imagen a 1280x720 píxeles usando `cv2.resize()` para optimizar el procesamiento. Luego realiza la detección de objetos llamando al modelo YOLO con umbrales de confianza (0.4) e IoU (0.45) equilibrados.

El código inicializa contadores para vehículos (`vehicle_count`) y peatones (`pedestrian_count`), luego procesa los resultados de la detección. Para cada objeto detectado, obtiene su clase mediante `box.cls` y la traduce a texto usando `model.names`. Los vehículos se identifican cuando la etiqueta coincide con 'car', 'truck', 'bus' o 'motorcycle', mientras los peatones se detectan con la etiqueta 'person'. Cada detección válida incrementa los contadores correspondientes y se marca en la imagen con un rectángulo delimitador (bounding box) usando `cv2.rectangle()`, usando color verde para peatones y rojo para vehículos.

Finalmente, la función guarda la imagen anotada añadiendo '\_detected' al nombre original mediante `cv2.imwrite()` y retorna los conteos junto con la ruta de la imagen procesada. La sección de ejemplo al final del código demuestra su uso, iterando sobre una lista de rutas de imágenes (`image_paths`) y mostrando los resultados por consola, incluyendo los conteos y ubicación de las imágenes procesadas.

El código utiliza principalmente las bibliotecas `ultralytics` para el modelo YOLO y `OpenCV (cv2)` para el procesamiento de imágenes, mostrando un enfoque modular donde la función principal encapsula toda la lógica de detección y conteo, mientras la parte de demostración maneja la ejecución concreta y presentación de resultados.

#### Código de simulación

En esta sección se describe el código que se ha empleado adaptándolo al sistema de estudio para hacer una simulación del funcionamiento del sistema [19].

Debido a la imposibilidad de acceso a la información en tiempo real (tanto los datos de intensidad de tráfico/peatonal como de cámaras de vigilancia de tráfico), se genera un código para realizar una simulación de cómo podría funcionar el sistema de forma aproximada. El escenario que se simula es similar al de la realidad en una vía donde existen dos direcciones vehiculares (izquierda y derecha) y donde existen otras dos direcciones peatonales (arriba y abajo). El lenguaje de programación empleado es Python y las herramientas principales para el desarrollo del sistema son la biblioteca Pygame para la visualización del sistema y el módulo threading para gestionar procesos concurrentes.

El objetivo principal es replicar de la manera más realista posible, el comportamiento de vehículos y peatones para ver cómo el sistema actúa mediante el semáforo ante diferentes escenarios. Para ello, se definen diferentes vehículos y personas (variando su velocidad), se incluyen distancias de seguridad entre vehículos y se monitoriza el comportamiento de los semáforos para su posterior análisis. Es decir, la simulación no solo genera un entorno dinámico para observar el flujo de elementos del sistema, sino que también recopila datos cuantitativos que permiten evaluar la eficiencia del sistema de control de tráfico implementado.

En cuanto a la estructura general, el programa se organiza en cuatro bloques principales interrelacionados. Para comprender mejor la interrelación, se puede observar el Esquema 2.

### Esquema 3: Desglose de los diferentes componentes del programa

En primer lugar, se define una configuración inicial para establecer los parámetros fundamentales que gobiernan la simulación, como por ejemplo las velocidades asignadas a diferentes tipos de vehículos y peatones y las coordenadas espaciales que definen las rutas de circulación. A continuación, dos clases centrales definen la lógica fundamental del sistema. Por una parte, la clase TrafficSignal modela el comportamiento de los semáforos, mientras que la clase Elements, que hereda de pygame.sprite.Sprite, no solo gestiona la representación visual de los vehículos/peatones sino que define también su comportamiento.

Finalmente, se definen un conjunto de tres funciones necesarias para el funcionamiento del sistema (initialize(), generateElements() y simTime()). Finalmente, la clase que orquesta la ejecución del sistema (Main) donde se ejecutan los procesos de manera concurrente, incluyendo la generación continua de vehículos y peatones (de forma aleatoria), la alternancia sincronizada de los semáforos y la visualización interactiva de toda la simulación. A continuación en la Figura se puede observar en el código de forma esquemática los tres últimos bloques mencionados con anterioridad (siguiendo el código de colores del Esquema 2 para facilitar su identificación).

Figura 28. Muestra de Código Python

### Hardware

En esta sección, inicialmente se describen brevemente los elementos que pueden englobar un sistema como el descrito anteriormente, los elementos seleccionados para la creación de un prototipo físico de sistema y, para concluir, con el estudio de la autonomía del sistema.

Se debe tener en cuenta la infraestructura que cuenta la localización donde se vaya a situar el sistema. En el caso de Madrid, está altamente capacitada para implementar este sistema sin tener que realizar modificaciones significativas al contar con sensores y cámaras ya instalados a lo largo de toda la ciudad. No obstante, es importante tener en cuenta que es posible que sean necesarias modificaciones en la infraestructura donde se quiera instalar dicho sistema.

En la Figura XXX, se puede observar un resumen de los elementos que pueden añadirse al sistema, tal y como se ha comentado en el Apartado 3.1. .

Figura 29: Elementos finales del sistema

### Elementos para el prototipo

En base al esquema anterior, se seleccionan elementos en función de su rendimiento, calidad y precio para generar el mejor sistema posible. En la Tabla 1 se pueden encontrar un resumen de los elementos seleccionados incluyendo las fuentes de obtención de estos.

Tabla XX: Descripción de elementos del sistema

### Análisis de autonomía operacional

Uno de los objetivos principales del proyecto es que el sistema sea lo más autónomo y sostenible posible para evitar el colapso del tráfico en casos de situaciones de emergencia como problemas en la red eléctrica.

Teniendo en cuenta esa premisa se analizan diferentes tipos de soluciones [22]: batería grande externa, generador de diésel o gasolina, energía eólica y panel solar con batería.

Primeramente, la batería grande es descartada debido al coste elevado y a la necesidad de recarga externa. En cuanto al generador con diésel o gasolina es excluido debido a que requiere un alto coste de combustible y de mantenimiento además es ruidoso y contaminante. En lo referente a la energía eólica desechada debido a que es inconsistente, requiere una turbina y espacio y su mantenimiento es complejo y costoso.

Finalmente, el panel solar proporciona una energía gratuita e ilimitada que es favorecida por climas como el de Madrid, necesita un mantenimiento bajo, es silencioso y limpio. Además, es escalable en función de las necesidades energéticas (se pueden instalar cuantos paneles sean necesarios en función de la demanda de energía). Como desventaja es que es dependiente del clima y durante la noche depende de batería o de la red eléctrica.

### Consumo estimado del sistema del semáforo

Raspberry Pi (~3W (0.6A, 5V)) + cámara (~1.5W) + sensores (~1W): ~5.5W (1.1A, 5V)

Luces LED del semáforo (estimado): ~9.5W (dependiendo del modelo)

Total de consumo de elementos: ~15W continuos

Consumo diario de elementos:

### Generación solar necesaria

Se busca un panel que se ajuste a la demanda necesaria. Se selecciona un panel de 80W que según las especificaciones que se muestran en la Figura XX, genera hasta 360 Wh.

Figura XX: Especificaciones técnicas del panel 80W

Esa generación de energía depende de factores externos como las condiciones de uso, corrientes de carga y descarga, ubicación, orientación, inclinación e instalación de paneles solares y de la época del año. Suponiendo unas condiciones óptimas de generación de energía (considerando 5h de sol equivalente) se generarían 400Wh ().

Balance energético

En un día soleado de condiciones óptimas:

=

Autonomía con batería

Tras una búsqueda se determina que dependiendo de las condiciones climáticas de la ciudad se recomienda usar dos tipos de baterías: de 560 Wh o de 1280 Wh. Se realizan cálculos con ambas para determinar cuál es más viable para el sistema descrito.

Autonomía (consumiendo únicamente de batería)

En el caso de la ciudad de Madrid se opta por la batería de 560 Wh debido a que es una ciudad con mucho sol anualmente y para situaciones en las que no se pueda obtener energía solar, la autonomía seguiría siendo de día y medio. Para prevenir posibles inconvenientes, se hará una conexión al sistema eléctrico por si el sistema autónomo falla o si el periodo en el que no hay sol se alarga.

Teniendo en cuenta que se elige un panel de 80W y una batería de 560 Wh, se opta por seleccionar un kit () cuyo datasheet se incluye en el Anexo 17. En la Figura 30, se muestra un resumen de la energía que se genera en función de la climatología, en un día nublado, durante el día el sistema podría estar funcionando prácticamente sin necesidad de la batería.

Figura 30: Gráfica de generación de energía en función de la climatología

Con el fin de ser energéticamente más eficientes, se decide optimizar el consumo energético. Por ejemplo, durante la noche, se implementa en modo de bajo consumo (reduciendo ligeramente la iluminación del semáforo). Adicionalmente, en el caso de una hora valle donde el tráfico sea bajo, dejar que el sistema funcione únicamente con los sensores, desactivando la cámara.

Problemas Técnicos y Tecnológicos

En esta sección se exponen las diferentes situaciones a las que el equipo se ha enfrentado y cómo ante dichas dificultades técnicas, se han resuelto de manera puntual y eficiente. A continuación se enlistan algunas de ellas.

Dificultades encontradas durante el ETL

Para el ETL de "Autos\_Traffic", la primera dificultad presentada fue el tamaño de cada archivo. Cada CSV mide entre 700 y 800 MB, cerca de 1'150,000,000 registros por archivo. Esto impactó en el tiempo de carga a la Base de Datos y en el tamaño de almacenamiento ya que en un principio se había propuesto un RDS en AWS. Los costos se elevaron desde un inicio por lo que, se analizó más detalladamente el conjunto de datos descubriendo que al final, para efectos del del proyecto, el enfoque estaba sólo en los IDS de la calle Gran Vía.

Así que, mediante código de python pandas, se filtraron los archivos y se generaron archivos CSV mucho más pequeños.

PROCESO ETL:

Python Pandas Colab

Para el ETL de la tabla de "PM\_Trafico\_Tpp\_Real" Proyecto en Visual Studio para la carga de los archivos XML que actualiza la base de datos cada 5 mins. Link GitHub: Anexo 9 (pendiente subirlo)

Limpieza de Datos: La mayoría de los datos del archivo de BICICLETAS y ACCIDENTES, contenían caracteres especiales que dificultan la subida de los archivos CSV a la base de Datos.

Anexo 11: Descripción de fallas en los datos.

Bloqueo de la Suscripción de Azure por mecanismos de seguridad:

Microsoft bloquea la suscripción debido a mecanismos de seguridad automáticos que detectaron patrones inusuales en el acceso a tu base de datos SQL. Al configurar el acceso restringido únicamente a direcciones IP públicas específicas, cualquier intento de conexión desde una IP no autorizada - especialmente al correr los scripts Python que se ejecutan desde una conexión con IP dinámica puede activar alertas de seguridad.

El problema se agrava por el patrón de uso de los procesos ETL diseñados. Los scripts Python suben datos de forma masiva en intervalos irregulares (creando picos abruptos de carga), los algoritmos de detección de anomalías de Azure lo marcan como comportamiento sospechoso, similar a intentos de exfiltración de datos. Para evitar futuros bloqueos, en la segunda fase del

proyecto, se planea considerar el uso de Azure Private Link para crear conexiones seguras mediante puntos de acceso privados, desde una VM con IP estática dedicada. Mejorar la autenticación implementando "Managed Identity" o "Service Principals" con certificados.

Ejecución de queries bastante lentos:

Al requerir crecimiento la base de datos y requería mejor performance en la ejecución de queries para traer los datos. Esto impactó en el desempeño de ejecución por lo cual, se tuvo que mejorar el Pricing Tier de General Purpose - Serverless: Gen5, 4 vCores a General Purpose - Serverless: Gen5, 8 vCores.

Modificación de código Python:

Se reemplazaron los queries para llenar los Data Frames por ejecución de Store Procedures (En Anexo 14 se muestra el código de su creación de los SPs).

Multicolinealidad

Al llenarse los dataframes y comenzar los análisis de predicción, la matriz de correlación en el caso de "Autos\_Traffic" (abajo representada) arrojan resultados importantes que se explican después de la figura 31.

Figura 31: Comparativo de Matriz de Correlación

El resultado son correlaciones fuertes ( $>0.7$ ) en ocupación y nivel\_servicio: 0.78, indicando que a mayor ocupación, mayor es el nivel de servicio, nivel\_servicio e intensidad (0.75) - Correlación fuerte positiva Correlaciones Moderadas (0.5-0.7): ocupación e intensidad: (0.52) Correlación moderada positiva entre ocupación e intensidad

"Timestamp" muestra correlaciones muy débiles con todas las demás variables lo que sugiere que el tiempo (timestamp) no tiene una relación lineal significativa con otras métricas del sistema por lo que se concluye que no hay patrones temporales claros en los datos, lo que podría indicar que el sistema no sigue ciclos horarios o estacionales marcados.

Lo que llama la atención del resultado de la regresión es que el  $R^2$  es igual a uno.

Figura 32. Resultado de Modelo de Regresión

Las variables operacionales (ocupación, nivel de servicio, intensidad) están interrelacionadas de tal manera que pudiese ser estadísticamente correcto?

La correlaciones con el timestamp, se observan relaciones muy fuertes entre las variables principales de tráfico: ocupación, carga e intensidad presentan correlaciones altas entre sí (0.85-0.86), lo cual es lógico ya que estas métricas están intrínsecamente relacionadas y tienden a aumentar o disminuir conjuntamente cuando hay más o menos tráfico, pero en el siguiente punto se detalla algo a simple vista no se había considerado al principio.

Falta de acotación del sistema:

Inicialmente, se trata de generar un sistema con todos los elementos de medición para que pueda ser un sistema general. Tras hacer varias pruebas de matriz de correlación, se observa una correlación muy baja entre variables. Mientras se busca la causa de esa falta de relación, se verifica la matriz de correlación empleando un único id de dispositivo de medición.

Tras comprobar que la relación entre las variables de entrada y la variable objetivo mejora, se determina que lo más conveniente es realizar la regresión específicamente por dispositivo. De esta manera, la naturaleza vial de la carretera no pierde su esencia y en consecuencia, no se vuelve impredecible. Es decir, el considerar varios dispositivos que engloban patrones de funcionamiento distintos, afecta negativamente a la regresión.

Factores que afectan a la detección

Tras realizar varias pruebas, se observa que existen varios factores, mayormente climatológicos (luz solar y la lluvia), que afectan a la detección de imágenes. No obstante, también se pueden dar casos de elementos de la naturaleza como ramas de árboles que pueden dificultar el desarrollo del sistema.

Por una parte, cuando los rayos del sol inciden directamente en la lente de la cámara, junto con la suciedad acumulada, genera unos reflejos y destellos que afectan a la visión computacional. A continuación, en las siguientes figuras, se muestran unos ejemplos de esa casuística.

Figura 33: Cámara de Neptuno ID 01-303 Figura 34: Cámara Gta. Ruiz Jimenez ID 01-333

Figura 35: Cámara de A.Aguilera-B.Garay Figura 36: Cámara de Callao- Gran Vía

ID 07-311 ID 01-314

Por otra parte, cuando hay demasiada lluvia, el suelo se queda mojado y genera reflejos como si fuese un espejo. Los destellos de los elementos como farolas o vehículos deslumbra, desdibujando las siluetas y dificultando la detección de vehículos y peatones. La falta de precisión

impacta de forma directa en el rendimiento del sistema. Seguidamente, en las siguientes figuras, se muestran unos ejemplos.

Figura 37: Cámara de Arturo Soria ID15-318 Figura 38: Cámara Marqués de Salamanca ID 04-313

Figura 39: Cámara de Gran Vía-San Bernardo Figura 40: Cámara de Goya- Velázquez ID 01-335 ID 04-314

Adicionalmente, también se ha observado que mientras llueve, las gotas pueden tapar el objetivo de la cámara, distorsionando completamente las imágenes como se muestra en las Figuras 39 y 40. Finalmente, en la Figura 41 o 42, se puede observar un ejemplo de cómo la naturaleza estática puede imposibilitar la detección de elementos en el sistema.

Figura 41: Cámara de Marq. de Salamanca Figura 42: Cámara Castellana- Marq. Riscal

Figura 43: Cámara General Perón- Orense Figura 44: Cámara Bravo Murillo - Ávila

#### Análisis y resultados

En este capítulo, se analizan los resultados obtenidos de ambos sistemas para determinar la viabilidad de las soluciones, las limitaciones de estos y las posibles líneas futuras de investigación para la mejora de estos. Adicionalmente, se hace una estimación de costes monetarios y el impacto de huella de carbono que implicaría ahorrar 5 minutos de espera en semáforos al día en una ciudad como Madrid.

#### Sistema basado en modelo predictivo

En esta sección, se analizan los resultados obtenidos del algoritmo desarrollado con Random Forest del device de Gran Vía '4308'. En la Figura XX, se expone la matriz de correlación, considerando "timestamp" vs "Componentes Separados"(día, año, hora). En dichas matrices se puede observar el fenómeno descrito como multicolinealidad descrito anteriormente.

Figura 45: Comparativo Matriz de Correlaciones.

Los modelos mostraron excelente rendimiento:

Modelo con timestamp completo:  $R^2=0.988$ , MAE=18.881

Modelo con componentes separados:  $R^2=0.989$ , MAE=18.558

Ambos modelos explican aproximadamente el 99% de la varianza en los datos ( $R^2$  cercano a 1), con un error absoluto alrededor de 18-19 vehículos. En el siguiente gráfico se muestra este excelente rendimiento que, aplicando hiper parámetros puede mejorar.

Figura 46: Rendimiento del Modelo Random Forest

El análisis revela un desempeño excepcional del modelo con características muy positivas en términos de precisión y confiabilidad. La línea diagonal punteada representa la predicción perfecta, donde los valores reales y predichos serían idénticos. La distribución de los puntos azules muestra una alineación muy estrecha con esta línea de referencia, lo que indica que el modelo logra predicciones muy precisas a lo largo de todo el rango de valores, desde aproximadamente 0 hasta 1600 unidades.

El modelo demuestra una capacidad predictiva consistente en diferentes niveles de intensidad de tráfico. En los valores bajos (0-400), medios (400-1000) y altos (1000-1600), la precisión se mantiene elevada sin mostrar sesgos sistemáticos hacia la sobre-estimación o sub-estimación.

La dispersión de los puntos es mínima, con muy pocos valores atípicos que se alejen significativamente de la línea diagonal. Los casos donde existe mayor variabilidad aparecen principalmente en los extremos superiores del rango (valores cercanos a 1500-1600), lo cual es común en modelos predictivos debido a la menor frecuencia de estos eventos extremos en los datos de entrenamiento.

Figura 47: Importancia de las Características

La variable "carga" domina completamente el modelo con una importancia cercana a 0.95, lo que indica que es el factor más determinante para explicar el comportamiento del tráfico. Esto es lógico desde el punto de vista del análisis de tráfico, ya que la carga vehicular (porcentaje de ocupación de la vía) es una medida directa de la congestión y representa fielmente la intensidad del flujo vehicular en tiempo real. Le sigue "ocupación" con una importancia mucho menor (aproximadamente 0.05), lo cual es interesante porque aunque estas dos variables están relacionadas, "carga" tiene un poder predictivo significativamente superior.

Figura 48: Análisis de Residuos

Este análisis revela características muy favorables que confirman la robustez del modelo desarrollado. La distribución de los residuos muestra una concentración muy densa alrededor de la línea roja horizontal (residuo = 0), lo que indica que la mayoría de las predicciones tienen errores

muy pequeños. Esta concentración es particularmente notable en el rango de predicciones entre 400 y 1200 unidades.

La homocedasticidad del modelo es excelente, ya que la varianza de los residuos permanece relativamente constante a lo largo de todo el rango de predicciones. No se observan patrones de "embudo" o cambios sistemáticos en la dispersión de los errores conforme aumentan los valores predichos, lo cual es una señal muy positiva de que el modelo no sufre de heterocedasticidad. Se identifican algunos residuos extremos, con el más notable siendo un valor cercano a +600 unidades alrededor de las 800 predicciones, pero estos casos excepcionales representan una fracción muy pequeña del total de observaciones.

Resultado del Grid Search para optimización:

Figura 49: Resultado Grid Search

Sistema basado en visión computacional

En esta sección, se analizan los resultados obtenidos del análisis de imágenes de la cámara situada en la intersección entre Gran Vía y Clavel (01-315 ID de cámara). Adicionalmente, se analizan dos ejemplos de otras cámaras, donde se pretende argumentar la escalabilidad del sistema.

Con los datos obtenidos de las imágenes del Ayuntamiento, se realiza un análisis exhaustivo de las imágenes adquiridas para poder analizar las fortalezas y las debilidades del sistema, para conocer las limitaciones del sistema y fortalecer estas últimas en las líneas futuras de investigación. Se debe tener en cuenta que el sistema no siempre va a funcionar bajo las mismas casuísticas. En el Anexo 15 se pueden observar algunos ejemplos de situaciones anómalas que pueden suceder en la vida útil del sistema como cortes de tráfico por una carrera, manifestación o evento así como obras. Adicionalmente, se pueden encontrar diferentes ejemplos abordando distintas situaciones en el Anexo 16.

Como se puede observar en las Figuras 50 y 51 de Gran Vía, el sistema cumple con su cometido de cara a la detección de elementos en la imagen obtenida desde la cámara del Ayuntamiento.

Figura 50: Cámara Gran Vía Figura 51: Cámara Gran Vía

Sin embargo, en situaciones adversas como grandes volúmenes de personas, el algoritmo no es capaz de detectar dicha magnitud como se muestra en el apartado de líneas futuras de investigación. Otro detalle que se debe corregir es que en ocasiones, los motoristas/ciclistas se consideran tanto vehículo como persona.

Figura 52: Grandes volúmenes Figura 53: Motorista detectado como peatón

Para comprobar la escalabilidad del sistema se hacen pruebas en otras zonas de Madrid cuya casuística es diferente. En dichas pruebas, se detecta que en ocasiones, los vehículos que se encuentran a cierta distancia resultan complicados de detectar para el algoritmo. Esto se puede deber a la resolución de la imagen. Teniendo en cuenta que la imagen se trata de una matriz de píxeles, cuanto peor resolución, más complicado es para el algoritmo la detección de rasgos.

Figura 54: Detección distancia larga Figura 55: No detección distancia larga

Finalmente, como se muestra en las Figuras 54 y 55, la cuestión que más preocupa de cara a la escalabilidad del sistema se trata de cómo hacer que el algoritmo pueda discernir entre coches que afectan al flujo de movimiento y aquellos que están aparcados.

Figura 56: Cámara Goya-Velázquez Figura 57: Cámara General Perón - Orense

Análisis de costes

En esta sección se realiza una primera estimación de los costes que pueden suponer el desarrollar un proyecto de semáforos inteligentes, teniendo en cuenta diferentes posibilidades de tecnologías desde algunas más asequibles hasta la tecnología puntera en el sector.

El coste de implementar un sistema de semáforos inteligentes basado en IA y big data puede variar significativamente según la escala del proyecto y la tecnología utilizada.

Los costes iniciales incluyen:

Costes de instalación: Los costes de instalación de sensores y dispositivos inteligentes pueden oscilar entre 100.000 y 500.000€.

Costes de software: El desarrollo o la licencia de software de IA puede requerir una inversión de entre 20.000 y 150.000€.

Costes de infraestructura: La configuración de la infraestructura necesaria para el análisis de datos puede costar entre 15.000 y 100.000€.

Costes de hardware: Teniendo en cuenta los dispositivos de comunicación inalámbrica, los sistemas de control, los sistemas de alimentación de reserva puede costar entre 35.000 y



170.000€ (en función del tipo de elementos adaptados a las necesidades de cada ciudad).

Los costes de mantenimiento incluyen:

Tarifas de licencia de software: Pueden oscilar entre 10.000 y 100.000 € anuales, dependiendo de los servicios que se utilicen.

Costes de análisis de datos: La utilización de análisis avanzados puede suponer un coste de entre 20.000 y 150.000 € al año.

Salarios del personal: Los salarios anuales de los especialistas en gestión del tráfico pueden oscilar entre 60.000 y 120.000 €.

Gastos de servicios públicos: Las facturas mensuales de servicios públicos y la infraestructura de comunicaciones pueden añadir entre 10.000 y 50.000€ al presupuesto.

Pese a que la inversión inicial sea elevada debido a la necesidad de monitorización completa del sistema, la retribución se ve compensada posteriormente y no tan solo monetariamente, sino con la salud y satisfacción de los ciudadanos.

Análisis de reducción de huella de carbono

En esta sección, se analiza brevemente el impacto que puede generar al cabo de un año la reducción de espera en los semáforos de tan sólo 5 minutos.

El consumo de CO<sub>2</sub> de un carro frenado en un semáforo con el motor encendido depende de varios factores, como el tipo de motor, la cilindrada y el tipo de combustible. Sin embargo, podemos hacer una estimación general:

Cálculo aproximado del CO<sub>2</sub> emitido en 1 minuto de ralentí (marcha mínima):

Un automóvil promedio con motor a gasolina consume entre 0.6 y 1 litro por hora en ralentí, se escoge la media de 0.8L .

Un litro de gasolina genera aproximadamente 2.31 kg de CO<sub>2</sub> al quemarse.

Cálculo para 1 minuto:

Si el coche consume 0.8 litros por hora, en 1 minuto consumirá:

$0.8 \text{ L/h} / 60\text{min} = 0.0133\text{L/min}$

Emisión de CO<sub>2</sub> en 1 minuto:

$0.0133\text{L/min} \times 1\text{min} \times 2.31 \text{ kg CO}_2/\text{L} = 0.03 \text{ kg CO}_2$

Impacto en la huella de carbono:

Si un solo coche emite 0.03 kg de CO<sub>2</sub> en 1 minuto, teniendo en cuenta una ciudad con 1 millón de coches, si cada uno pasa 5 minutos diarios en semáforos, se podría reducir la huella de carbono en:

$1,000,000 \times 0.03 \times 5 = 150,000 \text{ kg CO}_2/\text{día} = 150 \text{ toneladas diarias de CO}_2$

Por tanto, teniendo en cuenta una posible reducción de 150T al día, al cabo de un año, la reducción de huella de carbono alcanzaría casi las 55.000T.

Propuesta de líneas futuras de investigación

En este capítulo, se pretende determinar cuáles han sido las limitaciones del sistema o ideas que se consideran interesantes de desarrollar de cara a generar un sistema más eficiente y robusto.

La principal línea de investigación consiste en un análisis de más posibilidades de algoritmos de machine learning para determinar el que mejor se ajusta al sistema descrito anteriormente. Cuanto más robusto sea el sistema, más factores se van a poder incluir como variables de medición de contaminación del aire para hacer un sistema más eficiente.

En cuanto a la detección de elementos en la imagen, se debe estudiar los diferentes filtros, resoluciones y/o algoritmos que permitan eliminar algunas de las erratas previamente descritas.

Por otra parte, se considera interesante, generar un sistema de semáforos inteligentes interrelacionados para hacer un flujo vehicular coherente a lo largo de la ciudad y probarlo en una ciudad piloto como Madrid para poder detectar posibles dificultades antes de escalarlo a otras ciudades.

Uno de los factores de mayor interés se trata de la seguridad del sistema frente a malware o ciberataques. Se cree conveniente realizar una investigación de cómo proteger a un sistema tan crucial como el de la gestión del tráfico de personas que quieran emplearlo para sus propios intereses y beneficios.

Finalmente, se cree fundamental, desarrollar una investigación enfocada puramente en la ética del algoritmo y de cómo convertir este en lo más transparente y accesible para los usuarios, así como generar una ley donde se regule las responsabilidades ante diferentes sucesos.

Conclusiones

Los semáforos inteligentes autónomos tienen el potencial de revolucionar la gestión del tráfico urbano y mejorar la movilidad, la seguridad y la sostenibilidad. Sin embargo, la implementación de estos sistemas presenta desafíos técnicos, legales y éticos que deben abordarse cuidadosamente, tal y como se han descrito en el Análisis de resultados y en las Líneas futuras de investigación. La investigación futura debe centrarse en la optimización mediante distintos modelos y mejores elementos (como cámaras de mayor resolución), la integración de nuevas variables que otorguen de mayor eficiencia al sistema, la gestión de los incidentes que presenta actualmente el sistema, la seguridad y la ética del mismo.

A pesar de los desafíos, las ventajas de los semáforos inteligentes son significativas y justifican la inversión en esta tecnología. Al abordar los desafíos y aprovechar las oportunidades de investigación, podemos crear sistemas de tráfico más inteligentes, eficientes y sostenibles para las ciudades del futuro.

#### Ventajas

**Reducción de la congestión:** Los semáforos inteligentes pueden reducir significativamente la congestión del tráfico, lo que se traduce en tiempos de viaje más cortos y una mayor eficiencia del combustible.

**Mejora de la seguridad:** Los semáforos inteligentes pueden mejorar la seguridad vial al detectar y responder a posibles peligros, como peatones que cruzan la calle o vehículos que se aproximan a gran velocidad. Además, siendo un sistema autónomo, puede velar por la seguridad de los ciudadanos ante algunas adversidades como apagones generalizados.

**Cumplimiento de los ODS con la reducción de las emisiones:** Al reducir la congestión y los tiempos de inactividad de los vehículos, los semáforos inteligentes pueden contribuir a la reducción de las emisiones de CO<sub>2</sub> y la mejora de la calidad del aire.

**Mayor eficiencia del transporte público:** Los semáforos inteligentes pueden dar prioridad al transporte público, lo que mejora la velocidad y la fiabilidad de los autobuses y trenes.

**Mejora de la planificación urbana:** Los datos recopilados por los semáforos inteligentes pueden utilizarse para mejorar la planificación urbana y el desarrollo de infraestructuras.

**Reducción de las infracciones de tráfico:** Al optimizar el flujo de tráfico y reducir la frustración de los conductores, los semáforos inteligentes pueden ayudar a disminuir la tentación de infringir las normas de tráfico, lo que contribuye a una mayor seguridad vial. Aproximadamente el 90% de los accidentes de tráfico se producen principalmente debido a infracciones de tráfico.

#### Desventajas

**Costes elevados:** La implementación y el mantenimiento de los semáforos inteligentes pueden ser costosos.

**Ciberseguridad:** Los sistemas de semáforos inteligentes son vulnerables a los ciberataques.

**Complejidad ética:** La programación de los semáforos inteligentes para que tomen decisiones éticas en situaciones complejas es un desafío.

En definitiva, los semáforos inteligentes autónomos representan un avance transformador para las ciudades, combinando innovación tecnológica con beneficios tangibles en movilidad, seguridad y sostenibilidad. Su implementación responsable y continuo desarrollo, acercará a la humanidad a un futuro urbano más eficiente, seguro y alineado con las necesidades de la sociedad y el planeta.

#### Bibliografía

##### Antecedentes

Ambiente Soluciones. (s.f.). Desarrollo: Sistema autónomo off / grid.

CAR Group. (2017). MDOT international practices [Documento PDF].

Carnegie Mellon University. (2013). Surtrac [Documento PDF].

eHangzhou Government. (2020, 23 de octubre). City Brain.

InfoBierzo. (s.f.). Así funcionan los nuevos semáforos con inteligencia artificial: Menos tiempo de espera y más seguros.

Pacto Mundial Red España. (s.f.). ¿Qué puedes hacer tú? ODS.

PMC. (2025). Sensors on Internet of Things Systems for the Sustainable Development of Smart Cities: A Systematic Literature Review. PubMed Central.

Wu, J., Zhang, Y., & Chen, L. (2023). HumanLight: Incentivizing ridesharing via human-centric deep reinforcement learning in traffic signal control. arXiv preprint arXiv:2304.03697.

##### Metodología

CareerFoundry. (s.f.). The data analysis process step by step.

Marco Teórico

Centro México Digital. (s.f.). Diferencia machine learning inteligencia artificial.

Desarrollo

Fuentes de Bases de Datos Tabuladas

Agencia Estatal de Meteorología (AEMET). (s.f.). Acceso general AEMET.

Ayuntamiento de Madrid. (s.f.). Accidentes de tráfico de la ciudad de Madrid.

Ayuntamiento de Madrid. (s.f.). Aforo de tráfico en la ciudad de Madrid permanente del mes actual vigente que se actualiza cada 5 mins.

Ayuntamiento de Madrid. (s.f.). Aforos de peatones y bicicletas.

Ayuntamiento de Madrid. (s.f.). Aforos de tráfico en la ciudad de Madrid permanentes.

Ayuntamiento de Madrid. (s.f.). Pestañas accidentes de tráfico.

Microsoft Azure. (s.f.). SQL Azure Architecture.

Fuentes de Base de Datos de Imágenes

Ayuntamiento de Madrid. (s.f.). Informo Madrid - Tiempo real.

Desarrollo de Código

Towards Data Science. (s.f.). Traffic intersection simulation using pygame.

Hardware

Ambiente Soluciones. (s.f.). Desarrollo: Sistema autónomo off/grid.

IoT Sensor & Applications Online. (2025). Low Code IoT Platform.

TEDx Talks. (Año). Título del video [Video]. YouTube.

ANEXOS

ANEXO 1

ANEXO 2

ANEXO 3

ANEXO 4

Este conjunto de datos esta relacionado con el de , indicando este último para cada punto de medida, los vehículos que pasan.

La infraestructura de puntos de medida, disponible en la ciudad de Madrid se corresponde con:

- 7.360 detectores de vehículos con las siguientes características:
  - 71 incluyen dispositivos de lectura de matrículas
  - 158 disponen de sistemas ópticos de visión artificial con control desde el Centro de Gestión de Movilidad
  - 1.245 son específicos de vías rápidas y acceso a la ciudad
  - y el resto de los 5.886, con sistemas básicos de control de semáforos.
  - Más de 4.000 puntos de medida:
  - 253 con sistemas para el control de velocidad, caracterización de los vehículos y doble lazo de lectura
  - 70 de ellos conforman las estaciones de toma de aforos específicas de la ciudad.
  - Sistemas de control automático de toda la información obtenida de los detectores con el contraste continuo con patrones de comportamiento esperado, así como el seguimiento de las instrucciones marcadas por el Comité Técnico de Normalización AEN/CTN 199 "Equipamiento para la Gestión del Tráfico" y en particular las aplicaciones específicas del SC3 relativas a "Detectores y estaciones de toma de datos" y SC15 relativas a "Calidad de datos".
- En este mismo portal puede encontrar otros conjuntos de datos relacionados como son:
- Con la información en tiempo real (actualizada cada 5 minutos)
  - , con la misma información en formato KML, y con la posibilidad de verla en Google Maps o Google Earth.
  - Y otros conjuntos de datos relacionados con el tráfico. Puede buscarlos, poniendo en el buscador (arriba a la derecha), la palabra "Tráfico".

ANEXO 5

ANEXO 6

ANEXO 6.1

Manual básico para la visualización de la ubicación de dispositivos (PDF)

ANEXO 6.2

Visualización interactiva de puntos de medida desde Google Earth:

ANEXO 7

ANEXO 8

import numpy as np

```

import pandas as pd
import os
import sys
from google.colab import drive
drive.mount('/content/drive', force_remount=True)
path1="/content/drive/MyDrive/TFM_DetecciondeTrafico_SemInt/Datos/Trafico_en_tpo_real/05-20
25.csv" #cambia el mes según el que se suba
filename="05_2025" #cambia el mes según el que se suba
df=pd.read_csv(path1, sep=";")
Gran_Via_IdLocators = [10844, 10843, 10847, 10845, 11207, 10388, 4264, 3475, 10860, 10859,
4301, 10387, 4294, 4289, 4297, 4308,4307, 3730, 10386, 3672, 4292]
df_filtered = df[df['id'].isin(Gran_Via_IdLocators)]
from datetime import datetime
from datetime import time
#it was showing a warning, so this line is to improve the copy on writing mode.
pd.options.mode.copy_on_write = True
df_filtered['fecha']= pd.to_datetime(df_filtered['fecha'])
#separate the fecha field for date and time columns
df_filtered['date'] = df_filtered['fecha'].dt.date
df_filtered['time'] = df_filtered['fecha'].dt.time
#saving dataframe to CSV file
rutacsv="Gran_Via_" + filename+".csv"
df_filtered.to_csv(rutacsv, index=False)
try:
df_filtered.to_csv(rutacsv, index=False) # Set index=False to avoid saving the DataFrame index
print("Successfully saved " + rutacsv + ".csv")
except Exception as e:
print("An error occurred while saving the file: {e}")
from google.colab import files
files.download(rutacsv)

```

ANEXO 9

Links de Código GitHub

ANEXO 10

ANEXO 11: Procesamiento de variables

Procesamiento - Limpieza de Datos

Accidentes

Procesamiento:

Un null es "null" y otro es el null del SQL

AUTOS\_TRAFFIC

BICICLETAS

DATOS CLIMA

Se consume una API que existe en el web de mediante una aplicación en Python para subir la data por trimestre, ya que sólo así lo permite la API. (Anexo 9)

PM\_TRAFFIC\_TPO\_REAL

PEATONES

ANEXO 13: Ejemplo de análisis de nulos en peatones

Análisis de nulos por tiempo y dispositivo de medición. Se observa una gran cantidad de nulos inicialmente en 2019, especialmente en 5 dispositivos.

ANEXO 14: Código SQL para la creación de

Store Procedures

ANEXO 15: Casuísticas especiales

Obras en la calle/calle cortada por carrera manifestación u otro evento

ANEXO 16: Ejemplos de detección exitosa y fallidas

Mala calidad de las imágenes por la baja resolución de la cámara

Caso fallido con una resolución baja Caso de éxito con resolución baja

Caso fallido detección de riders Éxito pero detecta como persona y la bicicleta

Caso de éxito de detección de peatones desde ángulo diferente de cámara

Caso de éxito de detección de elementos

Buena detección de vehículos especiales como el de la policía, buses turísticos o plantas del ayuntamiento

Caso de éxito de detección con elementos externos (gota de agua o planta de árbol)

Casos fallidos en la detección nocturna

Casos de éxito en la detección nocturna

Caso de elementos aparcados y en movimiento

Caso de éxito de detección de paraguas pero detecta a persona como boca de incendio

Caso fallido de detección de moto como coche o detección de señalización de velocidad pintada en el suelo detectada como coche

Caso de detección de pantalla pero detecta edificio como pantalla también

Caso especial de persona en publicidad de autobús detectado como peatón

Caso especial de carretilla detectada como camión

Detección de vaca de Alehop como peatón

Caso de detección de grandes volúmenes de personas mal detectados

Casos fallidos de detección de vehículos a distancia

ANEXO 17: Datasheet paneles solares

ANEXO 18: Otros gráficos importantes

Accidentes por Mes