

UNIVERSIDAD POLITÉCNICA DE MADRID

**ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN**



**GRADO EN INGENIERÍA DE
TECNOLOGÍAS Y SERVICIOS DE
TELECOMUNICACIÓN
TRABAJO FIN DE GRADO**

**DISEÑO E INTEGRACIÓN DE UN MODELO
DE PREDICCIÓN DE CONDICIONES
INICIALES DE TRÁFICO EN LA
HERRAMIENTA DE GESTIÓN DE TRÁFICO
PARA SMART CITIES SUMO**

ANDRÉS EMILIO FLORES REINA

2024

GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE INFORMACIÓN

TRABAJO FIN DE GRADO

Título: Diseño e Integración de un Modelo de Predicción de Condiciones Iniciales de Tráfico en la herramienta de Gestión de Tráfico para Smart Cities SUMO

Autor: D. Andrés Emilio Flores Reina

Tutor: D. Laura Fernández Galindo

Ponente: D. Javier Conde Díaz

Departamento: Departamento de Ingeniería de Sistemas Telemáticos (DIT)

MIEMBROS DEL TRIBUNAL

Presidente: D.

Vocal: D.

Secretario: D.

Suplente: D.

Los miembros del tribunal arriba nombrados acuerdan otorgar la calificación de:
.....

Madrid, a de de 20...

UNIVERSIDAD POLITÉCNICA DE MADRID

**ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN**



**GRADO EN INGENIERÍA DE TECNOLOGÍAS Y
SERVICIOS DE TELECOMUNICACIÓN
TRABAJO FIN DE GRADO**

**DISEÑO E INTEGRACIÓN DE UN MODELO
DE PREDICCIÓN DE CONDICIONES
INICIALES DE TRÁFICO EN LA
HERRAMIENTA DE GESTIÓN DE TRÁFICO
PARA SMART CITIES SUMO**

ANDRÉS EMILIO FLORES REINA

2024

RESUMEN

El constante crecimiento de la población mundial ha impuesto una presión significativa sobre las áreas urbanas. La población global ha alcanzado los 8.000 millones de personas en el año 2022, y las distintas proyecciones indican esta cifra siga aumentando exponencialmente, lo que plantea retos enormes para la planificación urbana y la gestión de recursos. La densificación de las áreas metropolitanas, haciendo énfasis en las grandes capitales del mundo, exige soluciones innovadoras que aseguren la sostenibilidad y la calidad de vida de los ciudadanos.

En este contexto, surge el concepto de ciudades inteligentes o *Smart Cities*, este hace referencia a un concepto en el que las ciudades utilizan tecnologías de información y comunicación con el objetivo de mejorar todos los aspectos que envuelven la gestión urbana. Este trabajo se centra en uno de los problemas más grandes que afectan a las grandes ciudades, el problema de la congestión de tráfico.

Debido a la importancia de esta problemática que supone una ciudad congestionada y atascada por el tráfico, los planificadores urbanos se han visto obligados a utilizar las nuevas tecnologías para poder tomar medidas en este asunto. La congestión vehicular es uno de los problemas más críticos que enfrentan las grandes ciudades, afectando a la economía, el medio ambiente y la calidad de vida de los ciudadanos. Con el objetivo de proponer una posible solución para mejorar un poco esta problemática, en este trabajo se utilizan técnicas de *Machine Learning* para analizar grandes volúmenes de datos e intentar predecir la intensidad de tráfico de una manera efectiva, anticipando posibles congestiones y facilitando la toma de decisiones en tiempo real.

La importancia de los datos en la gestión del tráfico es crucial. La recopilación y el análisis de datos de tráfico en tiempo real proporcionan una base sólida que ayudan al desarrollo de modelos predictivos precisos. La digitalización del transporte y el uso de datos abiertos fiables mejoran la eficiencia de las soluciones que se proponen hoy en día referentes a la movilidad, fomentando de esta manera la transparencia y la participación pública.

El objetivo final de este Trabajo Fin de Grado (TFG) es desarrollar una herramienta sencilla basada en *Machine Learning* que utilice datos de tráfico de la ciudad de Madrid para predecir la intensidad de tráfico en momentos específicos e integrarla en la herramienta de simulación de redes vehiculares SUMO (*Simulation of Urban MObility*) logrando la predicción de las condiciones iniciales de una simulación, es decir, se busca predecir las condiciones bajo las cuales la simulación se pueda iniciar, basándose en distintas variables que afectan directamente al tráfico. En resumen, se pretende desarrollar una herramienta que permita generar las condiciones iniciales de una simulación basándose en la predicción de un modelo de regresión lineal.

PALABRAS CLAVE

Tráfico; Aprendizaje Automático; SUMO; Simulación de tráfico; Intensidad de Tráfico.

SUMMARY

The constant growth of the world population has imposed significant pressure on urban areas. The global population reached 8 billion people in 2022, and this figure is expected to continue increasing exponentially, posing enormous challenges for urban planning and resource management. The densification of metropolitan areas, particularly in major capitals worldwide, demands innovative solutions to ensure sustainability and the quality of life for citizens.

In this context, the concept of smart cities or Smart Cities arises, using information and communication technologies to improve urban management. This work focuses on one of the biggest problems affecting large metropolises: traffic congestion.

Due to the importance of this problem, which results in congested and traffic-clogged cities, urban planners have been compelled to use new technologies to address this issue. Vehicular congestion is one of the most critical problems facing large cities, impacting the economy, the environment, and citizens' quality of life. To propose a potential solution to alleviate this problem, this work employs Machine Learning techniques to analyze large volumes of data and attempt to predict traffic intensity effectively, anticipating possible congestions and facilitating real-time decision-making.

The Importance of data in traffic management Is crucial. Collecting and analyzing real-time traffic data provides a solid foundation for developing accurate predictive models. The digitization of transport and the use of open and reliable data improves the efficiency of mobility solutions. Moreover, they promote transparency and public participation.

This Final Degree Project (TFG) aims to develop a tool based on Machine Learning that uses traffic data from the city of Madrid to predict traffic intensity at specific times and integrate it into the SUMO (*Simulation of Urban MObility*) vehicular network simulation tool. The aim is to develop a tool that generates the initial conditions of a simulation based on the prediction of a linear regression model.

KEYWORDS

Traffic; Machine Learning; SUMO; Traffic Simulation; Traffic Congestion

AGRADECIMIENTOS

Primero, doy gracias a Dios por estos años universitarios que tanto me han aportado.

A todas las personas que me ayudaron, de distintas maneras a lo largo de estos años, mi agradecimiento con todas ellas.

A mis padres, por dejar que un joven de 18 años cambiara de continente y se fuera de su casa a una temprana edad para emprender esta aventura. Gracias por estar siempre conmigo, aunque estemos lejos.

A mis abuelos, por la atención que siempre me prestan y lo mucho que me quieren. Gracias por siempre confiar en mí, estoy orgulloso de ustedes.

A mi hermano Santiago. Gracias por ser el mejor hermano que se pueda pedir. Pronto estaremos juntos.

A Carmen, que ha sido mi familia prácticamente desde que comencé mi etapa universitaria. Siempre ha estado ahí para apoyarme cuando más lo he necesitado, y ha creído en mí incluso cuando yo mismo dejé de hacerlo.

A mi tía, por ser el único familiar que he tenido en esta ciudad durante los años universitarios. Gracias por estar siempre ahí conmigo.

A mis amigos de Ruleco y Planchazo, que hacen que una tarde de estudio se vuelva mucho más que eso, el pasar tanto tiempo juntos me ha enseñado el significado de la amistad. Gracias por tantos momentos y porque cuando estaba a punto de caer, me recordaban que sí que se podía lograr.

Finalmente, gracias a mis amigos del colegio. Nos vinimos a Madrid con la edad de 18 años, y hemos estado juntos toda la vida. Vivir esta etapa estudiantil con ustedes me hizo en muchas ocasiones no sentirme solo, y recordar que siempre estamos el uno para el otro.

ÍNDICE DEL CONTENIDO

AGRADECIMIENTOS	6
1. INTRODUCCIÓN Y OBJETIVOS	1
1.1. Introducción.....	1
1.2. Objetivos.....	2
2. MARCO TEÓRICO	3
2.1. Gestión de tráfico en Smart Cities	3
2.1.1. Contextualización y relevancia en el entorno urbano moderno.....	3
2.1.2. Introducción a los simuladores de tráfico	4
2.1.3. Principales simuladores de tráfico	5
2.2. Machine Learning en la gestión de tráfico	7
2.2.1. Fundamentos del Machine Learning en Smart Cities.....	7
2.2.2. Aplicación a la predicción del tráfico	7
2.2.3. Fuentes de datos de las Smart Cities.....	8
2.2.4. Los datos en la gestión de tráfico	10
3. DISEÑO DE LA SOLUCIÓN.....	12
3.1. Diseño de la Arquitectura	12
3.2. Recolección y limpieza de datos.....	14
3.2.1. Fuente de los datos.....	14
3.2.2. Selección de los puntos de medida	17
3.2.3. Limpieza de datos, transformaciones realizadas y creación del dataset.....	19
3.3. Diseño del modelo de predicción	22
3.3.1. Selección del modelo	22
3.3.2. Análisis del Dataset	22
3.3.3. Entrenamiento del modelo	25
3.4. Integración con SUMO.....	27
3.4.1. Preparación de datos para SUMO.....	27
3.4.2. Diseño de la red de simulación.....	28
3.4.3. Integración de la simulación	31
4. RESULTADOS	34
4.1. Evaluación del módulo de predicción.....	34
4.2. Comportamiento del tráfico simulado	36
5. CONCLUSIONES Y LÍNEAS FUTURAS	39
5.1. Conclusiones.....	39
5.2. Líneas Futuras.....	40
6. BIBLIOGRAFÍA.....	41

ANEXO A: ASPECTOS ÉTICOS, ECONÓMICOS, SOCIALES Y AMBIENTALES	44
A.1 Introducción.....	44
A.2 Descripción de impactos relevantes relacionados con el proyecto.....	44
A.3 Análisis del Impacto Ambiental	44
A.4 Conclusiones.....	45
ANEXO B: PRESUPUESTO ECONÓMICO	46

ÍNDICE DE TABLAS

Tabla 1: Tabla comparativa de Simuladores	6
Tabla 2: Campos de los datos históricos de tráfico	15
Tabla 3: Ubicación de los puntos de medida del tráfico.....	16
Tabla 4: Parámetros del dataset diseñado	21
Tabla 5: Resultados de una simulación	36
Tabla 6: Presupuesto Económico.....	46

ÍNDICE DE FIGURAS

Figura 1: Diagrama de la Arquitectura del Sistema.....	13
Figura 2: Proceso de diseño del Dataset.....	16
Figura 3: Mapa de los puntos de medida de Madrid	17
Figura 4: Punto de medida con id 3840	18
Figura 5: Punto de medida con id 3841	18
Figura 6: Distribución cíclica de los meses	20
Figura 7: Comparación de la Intensidad media del tráfico en días festivos vs no festivos	23
Figura 8: Comparación de la intensidad media del tráfico con y sin partido	24
Figura 9: Intensidad media del tráfico por hora del día.....	24
Figura 10: Figura ilustrativa sobre el Rendimiento del Modelo.....	26
Figura 11: Exportación de datos OSM para la simulación	29
Figura 12: Red de simulación	30
Figura 13: Ejemplo de fichero de salida de una simulación	33
Figura 14: Comparación de la intensidad media de tráfico real y predicha a lo largo del día	35
Figura 15: Influencia de los eventos multitudinarios en el tráfico	35
Figura 16: Comportamiento de tráfico simulado.....	37

ACRÓNIMOS

SUMO *Simulation of Urban MObility*

ML *Machine Learning*

CCTV *Closed Circuit Television*

ITS *Intelligent Ttransport Systems*

MITMA *Ministerio de Transportes, Movilidad y Agentes Urbana*

QGIS *Quantum GIS*

KML *Keyhole Markup Language*

DGT *Dirección General de Tráfico*

OSM *Open Street Map*

GIS *Geographical Information System*

IA *Inteligencia Artificial*

XML *Extensible Markup Language*

EDIM *Espacio de Datos Integrados de Movilidad*

EMS/MSE *Error Cuadrático Medio*

OBD *On Board Devices/Dispositivos Abordo de los vehículos*

RL *Aprendizaje por refuerzo*

1. INTRODUCCIÓN Y OBJETIVOS

1.1. INTRODUCCIÓN

En la era de la urbanización acelerada y el crecimiento demográfico concentrado en las áreas metropolitanas, las ciudades enfrentan desafíos crecientes en términos de sostenibilidad, eficiencia y calidad de vida. Este contexto ha dado lugar al concepto de ciudades inteligentes o *Smart Cities*; este término [1] se refiere al empleo de tecnologías de información y comunicación para detectar, analizar e integrar información esencial de los distintos sistemas fundamentales que operan en las ciudades.

El aumento acelerado de la población global intensifica los retos urbanos a los que se enfrentan las ciudades, y por consecuencia, demandan un avance tecnológico en cuanto a la mejora de los entornos urbanos. Acorde con un informe de la ONU [2], la población mundial se ha triplicado desde mediados del siglo XX, llegando a alcanzar los 8.000 millones de personas en 2022. Las proyecciones indican un aumento de casi 2.000 millones de personas para los próximos 30 años, alcanzando los 9.700 millones para 2050. Este crecimiento demográfico impone presiones significativas en el diseño urbano y en la infraestructura de las ciudades. El aumento de la población intensifica la densidad en áreas metropolitanas y por ende exige un rediseño constante de las infraestructuras existentes, así como también de las políticas públicas para acomodarse al constante incremento del flujo de personas y de vehículos.

El rápido avance de la urbanización ha modernizado la vida de muchas personas, pero esto ha traído consigo desafíos notables, entre los que se destaca la gestión de tráfico, que lleva consigo un aumento en el consumo de combustible y en la emisión masiva de contaminantes [3]. Estos fenómenos impactan significativamente la salud y calidad de vida de los ciudadanos. En el artículo [4] se hace especial énfasis en la relación que existe entre la contaminación y la salud pública, esta es alarmante, pues según este estudio, se estima que aproximadamente 364.200 personas fallecen cada año prematuramente en Europa debido a la contaminación del aire, en la cual el transporte y el tráfico tienen gran parte de responsabilidad.

El aumento del número de vehículos impone una presión sin precedentes en el tema del diseño de la infraestructura de tráfico. A medida que pasa el tiempo, las infraestructuras viales demandan cada vez más expansiones para optimizar la circulación y poder responder eficientemente a las exigencias de un tráfico cada vez más denso. Esto implica la implementación de soluciones avanzadas que integren tecnologías inteligentes y se apliquen a los conceptos de las *Smart Cities*, con el objetivo de lograr una infraestructura vial que siempre esté en constante crecimiento. La modificación de la infraestructura vial debe basarse en utilizar las nuevas tecnologías para mejorar la conectividad, reducir los tiempos de viaje y minimizar el impacto ambiental, asegurando de esta manera una movilidad urbana más fluida y eficiente.

El *Machine Learning* se ha posicionado como una herramienta clave en la mejora de la infraestructura y la gestión de tráfico, ofreciendo soluciones innovadoras que abordan directamente los retos impuestos por la urbanización y el aumento de la densidad vehicular [5]. El estudio referenciado previamente habla sobre como los avances en el aprendizaje automático y en el surgimiento de nuevas fuentes de datos permiten examinar y predecir las condiciones del tráfico urbano con una precisión nunca antes vista.

Los sistemas de gestión de tráfico avanzados, alineados con tecnologías de *Machine Learning*, se traducen en beneficios tangibles para el diseño de las infraestructuras viales y urbanas. La optimización de carreteras y vías mediante el uso de estas tecnologías mejora la fluidez del tráfico y por consecuencia, disminuye la contaminación. Al disminuir los niveles de congestión vehicular, se reduce automáticamente la cantidad de emisiones contaminantes liberadas al ambiente, lo cual resulta en una mejora de la calidad del aire [5].

Con la integración de tecnologías inteligentes y el aprendizaje automático enfocados en la gestión de tráfico, se abren nuevas oportunidades para enfrentar los desafíos urbanos presentes en las ciudades de una manera más efectiva. La correcta implementación de estas innovaciones mejoran la eficiencia del tráfico, reducen la contaminación y contribuyen de manera significativa a la sostenibilidad urbana.

1.2.OBJETIVOS

El principal objetivo de este Trabajo Fin de Grado (TFG) es desarrollar un modelo de predicción que utilice datos de tráfico de la ciudad de Madrid para predecir la intensidad de tráfico en un momento específico y posteriormente integrar el modelo en la herramienta de gestión de tráfico SUMO.

Más concretamente, existen tres objetivos principales en este trabajo. El primero de ellos es el diseño y la creación de un *dataset* completo que recoja todo tipo de variables que puedan influir en el flujo del tráfico. Los datos se limpiarán y transformarán añadiendo nuevas variables y condiciones de entorno que nos ayuden a tener un conjunto de datos lo más completo posible. Este proceso se realiza con datos abiertos que están disponibles para todos los ciudadanos

El segundo objetivo es desarrollar un modelo de predicción utilizando algoritmos de *Machine Learning* para entrenarlo, utilizando la biblioteca *Scikit-Learn* de Python para entrenar el dataset diseñado en el primer objetivo.

El tercer objetivo es la integración del modelo de predicción entrenado con la herramienta de simulación de redes vehiculares SUMO (*Simulation of Urban Mobility*). Para cumplir este objetivo, se buscará que las condiciones iniciales de la simulación se basen en las predicciones del modelo.

2. MARCO TEÓRICO

2.1. GESTIÓN DE TRÁFICO EN SMART CITIES

Una ciudad inteligente o *Smart City* es un concepto de urbanización que utiliza tecnologías de información y comunicación (TIC) de forma amplia y estratégica con el objetivo de mejorar la calidad, el rendimiento y la interactividad de los servicios urbanos. En la era contemporánea, el concepto de *Smart Cities* o ciudades inteligentes se ha convertido en un foco central para abordar los crecientes desafíos urbanos. En este contexto, la gestión de tráfico es una de las áreas más críticas, donde la aplicación de soluciones tecnológicas avanzadas promete transformar la manera en que se mueven personas y vehículos en un entorno urbano. El uso del enfoque “Smart Cities” cuando se tratan problemas de gestión de tráfico tiene innumerables aplicaciones, como sistemas inteligentes de transporte, recopilación de datos en tiempo real, semáforos inteligentes, mejora en la fluidez del tráfico, entre otros [6]. Las *Smart Cities* utilizan sensores y datos en tiempo real para optimizar el flujo de tráfico, prevenir accidentes y reducir los niveles de emisiones contaminantes.

Desde el cambio de milenio, el desarrollo de las Ciudades Inteligentes ha cobrado gran importancia, la transición demográfica del campo a la ciudad durante el siglo XX y los primeros años del siglo XXI ha resultado en que más de la mitad de la población mundial reside en áreas urbanas, este aumento ha significado el auge de varios desafíos logísticos urbanos, como la gestión de desechos, la escasez de recursos esenciales, el deterioro de las infraestructuras esenciales y la congestión continua de vehículos.

Un ejemplo de *Smart City* que podemos encontrar aplicado en España es el caso de la ciudad de Santander, que destaca como un modelo de ciudad inteligente en Europa, transformándose en un laboratorio urbano para la innovación tecnológica. Destaca por el uso de las nuevas tecnologías como Internet de las Cosas (IoT) y *Big Data*, ya que mediante su uso ha logrado mejorar la gestión de los distintos servicios públicos como el agua, la gestión de residuos, la movilidad, el alumbrado, entre otros, optimizando recursos y mejorando la calidad de vida de los ciudadanos. Este proceso se logró mediante el despliegue de miles de dispositivos que permiten conocer en tiempo real el estado de los servicios públicos mencionados anteriormente. Estos datos, son medidos en distintos puntos de control son puestos a disposición del ayuntamiento y de los ciudadanos, que participan activamente en la gestión urbana, reportando incidencias, así como también siguiendo su resolución en tiempo real, lo que fomenta una comunidad más conectada y proactiva [7].

2.1.1. CONTEXTUALIZACIÓN Y RELEVANCIA EN EL ENTORNO URBANO MODERNO

En la actualidad, el concepto de ciudades inteligentes ha evolucionado más allá de la simple implementación de la tecnología en las ciudades; se ha convertido en un modelo integral que recompensa la organización y la administración urbana. La gestión de tráfico, dentro de este marco, no es solo una cuestión de flujo vehicular, sino que, con el paso del tiempo y el aumento tanto de la población como de los vehículos en circulación, se ha convertido en una necesidad para fomentar entornos urbanos sostenibles y eficientes. Las ciudades inteligentes adoptan tecnologías avanzadas para transformar la infraestructura de los distintos elementos que conforman todo lo relacionado con el tráfico, integrando soluciones que van desde implementación de nuevos sensores hasta sistemas que trabajan conjuntamente para optimizar la movilidad y reducir las incidencias de congestión [8].

El marco fundamental de las ciudades inteligentes viene dado por las políticas aplicadas en cuanto a la implementación y operación de las tecnologías urbanas avanzadas. Estas políticas son diseñadas para asegurar que la adopción de soluciones innovadoras contribuya de manera efectiva a los objetivos de desarrollo sostenible, promoviendo una gestión eficiente de los recursos urbanos. A medida que las ciudades se transforman en ecosistemas cada vez más inteligentes, las políticas aplicadas deben abordar una variedad de desafíos y oportunidades, desde la integración tecnológica y la sostenibilidad hasta la inclusión social y la seguridad de los datos.

A continuación, exponemos las diferentes políticas que son aplicadas para un buen desarrollo de las *Smart Cities*:

- Las políticas de sostenibilidad en ciudades inteligentes incluyen la promoción del transporte sostenible y el manejo eficiente de recursos naturales, estas políticas tienen como objetivo minimizar el impacto ambiental de la urbanización y mejorar la calidad de vida urbana [8].
- Las políticas de movilidad y transporte son esenciales para crear sistemas y vías de transporte eficientes que logren en un futuro reducir la congestión y mejorar la calidad del aire [6].
- Las políticas de seguridad y privacidad de datos son necesarias ya que permiten establecer regulaciones estrictas para proteger la información personal de los ciudadanos contra usos indebidos de sus datos y ciberataques [1].

2.1.2. INTRODUCCIÓN A LOS SIMULADORES DE TRÁFICO

Los simuladores de tráfico son herramientas esenciales en lo que se refiere a la gestión urbana, estos cobran gran importancia tras el surgimiento del concepto de *Smart Cities*, donde actualmente las ciudades enfrentan retos significativos debido al aumento del comercio en línea, el incremento constante del parque vehicular y las restricciones de tráfico aplicadas en el centro de las ciudades. Estas herramientas permiten modelar y evaluar los impactos de diversas intervenciones de movilidad urbana, ofreciendo una base sólida para la toma de decisiones informadas y el desarrollo de soluciones innovadoras y sostenibles en temas relacionados con la logística urbana [9].

Un simulador de tráfico efectivo debe tener la capacidad de representar redes de transporte complejas, que incluyen la integración y comportamiento de las distintas entidades que conforman la red vehicular (vehículos, peatones, ciclistas y señales de tráfico). Una buena simulación de tráfico debe ser capaz de capturar la dinámica de los sistemas de transporte urbano y proporcionar resultados realistas en lo que se refiere al rendimiento del sistema, ajustándose a las condiciones reales del entorno.

Estos modelos son fundamentales para evaluar intervenciones de logística en las ciudades antes de llevarlas a cabo. Los simuladores permiten a los investigadores y planificadores urbanos realizar pruebas sobre distintos escenarios, proporcionando información valiosa sobre como las intervenciones propuestas podrían mejorar o empeorar la congestión y las emisiones, así como también la eficiencia del servicio de transporte en un contexto urbano [10].

2.1.3. PRINCIPALES SIMULADORES DE TRÁFICO

La elección de un simulador adecuado depende de las necesidades específicas de cada estudio. En la actualidad, existen muchas herramientas de código abierto a nuestra disposición que nos permiten realizar distintas simulaciones de tráfico.

Los simuladores de tráfico se clasifican en tres tipos principales:

- Simuladores de Movilidad: Estos simuladores modelan los flujos de vehículos de forma individual (microscópico) o en densidades (macroscópico), son ideales para estudiar el comportamiento del tráfico en diferentes escenarios urbanos o extraurbanos, proporcionando datos detallados sobre la interacción de vehículos individuales, así como de los factores externos que afectan su comportamiento [9].
- Simuladores de red: Evalúan el rendimiento de protocolos para la movilidad de nodos. Estos simuladores se centran en como los movimientos afectan a la topología de una red. Estos softwares permiten la simulación de sistemas de redes completos reproduciendo su comportamiento[11].
- Simuladores VANET: Combinan simuladores de movilidad y red para modelar tanto el comportamiento del tráfico como la comunicación entre los vehículos, este tipo de simuladores permiten estudiar tanto el flujo del tráfico como la transmisión de datos entre vehículos, así como también con la infraestructura vial [12].

En el desarrollo de este Trabajo de Fin de Grado (TFG), se ha optado por utilizar un simulador de movilidad debido a su capacidad de modelar de una forma detallada los flujos vehiculares individuales y las condiciones de tráfico en un entorno urbano. También se estudió la posibilidad de utilizar un simulador de tipo VANET, pero como buscamos la predicción de la intensidad de tráfico, no es necesario incluir un simulador de redes, con un simulador de movilidad es suficiente.

Como utilizaremos un simulador de movilidad, entraremos con un poco más de detalle en estos. A continuación, exponemos algunos de los principales simuladores de movilidad, así como su comparativa, esto de cara a elegir el simulador que más se adecue a nuestras necesidades [13]:

- AIMSUN (*Advanced Interactive Microscopic Simulator for Urban and Non-urban Networks*): Desarrollado por TSS (*Transport Simulation Systems*) en España, es un simulador de tráfico capaz de reproducir las condiciones reales de cualquier red de transporte, este software es utilizado para el desarrollo y prueba de sistemas de control de tráfico, reglas de gestión de tráfico, controles de acceso, peajes, redes de transporte público y carriles exclusivos. Además, puede integrarse con aplicaciones en tiempo real y con sistemas GPS. Una de sus principales ventajas es poder simular distintos tipos de redes en una sola simulación [14].
- ARCHISM: Desarrollado por el equipo de Modelado y Simulación del Instituto Nacional de Investigación en Transporte y Seguridad (INRETS), es un modelo de simulación de comportamiento desarrollado en Francia que implementa conceptos de multiagente, esto se refiere a que los conductores de los vehículos se simulan como agentes cuyas funciones se basan en procesos de percepción, decisión y acción, en este simulador hay dos tipos de componentes, agentes y objetos [15].

- **CORSIM:** Desarrollado por la Administración Federal de Carreteras (FHWA) en Estados Unidos, es un software de simulación de tráfico diseñado especialmente para los sistemas de señales, redes viales y autovías, consta de dos modelos que representan el entorno de la simulación completa, NETSIM, que se encarga de simular el tráfico en carreteras urbanas y FRESIM, que lo simula en carreteras fuera de poblado y autovías [16].
- **MatSim:** Desarrollado por el Instituto Politécnico de Zúrich, proporciona distintas herramientas para implementar una simulación basada en agentes a gran escala. Es un simulador capaz de simular el tráfico de regiones extensas a lo largo de un día completo, aunque no tiene enfoque en el comportamiento detallado de vehículos [17].
- **SUMO:** Es un simulador vehicular desarrollado por el Centro Espacial Alemán, es un simulador de código abierto que permite el modelado de sistemas de tráfico intermodales, incluyendo vehículos, transporte público y peatones[18].

Para la realización de este TFG, se seleccionó SUMO como simulador principal, esto se debe a diferentes razones, una de ellas es la capacidad de adaptación que posee, también consta de extensas opciones de personalización de simulaciones, lo que nos será muy útil en la etapa de desarrollo. Este simulador destaca por su arsenal de herramientas que facilitan la conversión de formatos de mapas y rutas, así como el análisis de los datos de salida de las simulaciones.

Una de las características principales de SUMO y una de las razones por las que se eligió utilizarlo es su capacidad de modelar una simulación microscópica de los flujos de tráfico, lo que es vital para predecir la intensidad de tráfico en entornos urbanos como en el que estamos trabajando. La naturaleza de código abierto de SUMO facilita la integración de herramientas de aprendizaje automático para el desarrollo de modelos de predicción de tráfico, más adelante detallaremos más propiedades de este simulador y su funcionamiento.

A continuación, se expone una tabla comparativa sobre los distintos simuladores anteriormente mencionados;

Simulador	Modelo	Categoría	Visualización	Infraestructura (Dificultad/Flexibilidad)	Vehículos y peatones	Sensores	GIS
AIMSUM	Micro/meso/macro	Comercial	2D, 3D	Difícil/Alta	Si	Cableados e inalámbricos	Si
ARCHISM	Microscópico	Comercial	2D	Difícil/Limitada	Si	Cableados	Parcial
CORSIM	Microscópico	Comercial	2D	Media/Limitada	Si	Cableados	Parcial
MATSim	Micro/Macro	Open Source	2D	Media/Limitada	Si	Cableados	Si
SUMO	Microscópico	Open Source	2D	Difícil/Limitada	Si	Cableados	Si

Tabla 1: Tabla comparativa de Simuladores

2.2.MACHINE LEARNING EN LA GESTIÓN DE TRÁFICO

2.2.1. FUNDAMENTOS DEL MACHINE LEARNING EN SMART CITIES

El *Machine Learning* (ML) se define como una rama de la inteligencia artificial, y se basa en la premisa de que los sistemas pueden aprender y mejorar a partir de la experiencia sin estar explícitamente programados para ello. Los sistemas de ML cuentan con la capacidad de aprender de grandes volúmenes de datos [21]. Antes se ha hablado de las *Smart Cities* y de la gran cantidad de datos que generan, lo que las hace perfectas para la aplicación del ML. Por ejemplo, en el artículo [19] se menciona como el *Machine Learning* rescataría el tráfico denso urbano en la ciudad de París, donde ya se están llevando a cabo proyectos de predicción de flujo de tráfico utilizando técnicas de ML.

En el contexto de *Smart Cities*, el ML tiene diversas aplicaciones, una de ellas es la aplicación en los sistemas de transporte inteligente (ITS), donde el uso de estas técnicas ayuda a optimizar las rutas, reducir congestiones y mejorar la seguridad vial. El ML también es fundamental en la gestión energética, donde se busca optimizar el consumo y la distribución en *Smart grids*, que se definen como los distintos sistemas de redes de distribución eléctrica que utilizan la tecnología digital para abastecer la demanda energética buscando la máxima eficiencia posible [20].

El *Machine Learning* se clasifica en tres categorías principales: aprendizaje supervisado, no supervisado y aprendizaje por refuerzo (*Reinforcement Learning*). Las distintas categorías de ML se utilizan en función de los datos con los que se entrena a un modelo, y utilizan distintos algoritmos con el objetivo de identificar patrones. A continuación, definimos las tres categorías del *Machine Learning* [20] [22]:

- **Aprendizaje Supervisado:** En esta modalidad del ML, los modelos se entrenan utilizando un conjunto de datos que incluye entradas y salidas deseadas. Los algoritmos aprenden una función que mapea entradas y salidas, lo que lo vuelve muy útil para tareas de predicción y clasificación.
- **Aprendizaje No Supervisado:** Esta modalidad del ML trabaja con datos que no están etiquetados, es decir, no tienen una salida conocida. Los algoritmos están diseñados para buscar patrones en los datos. Es fundamental para tareas como la segmentación de clientes en las empresas y para la detección de anomalías en el ámbito de la ciberseguridad.
- **Aprendizaje por Refuerzo:** RL es un tipo de ML donde un agente es entrenado para tomar decisiones secuenciales. Mediante prueba y error es capaz de aprender las consecuencias de sus acciones, y de esta manera es capaz de ajustarse, es útil en situaciones donde prevalecen las condiciones cambiantes de un entorno, y las decisiones deben adaptarse constantemente

2.2.2. APLICACIÓN A LA PREDICCIÓN DEL TRÁFICO

El *Machine Learning* se está convirtiendo con el paso del tiempo en una herramienta fundamental para la gestión moderna del tráfico en las ciudades, donde la congestión y la eficiencia del transporte se presentan como desafíos constantes a mejorar. Mediante la aplicación de algoritmos de ML como *Random Forest*, las ciudades pueden no sólo anticipar y mitigar problemas de congestión, sino que también pueden prepararse de cara a un posible corte de tráfico en algún punto analizado, este tipo de prácticas permiten a las autoridades y a los planificadores urbanos analizar grandes cantidades de datos de tráfico en tiempo real, desde condiciones meteorológicas hasta patrones de flujo vehicular según la hora del día y de esta manera poder predecir la congestión de tráfico [22].

Un ejemplo del uso de ML en temas referentes a la gestión de tráfico lo encontramos en la mega-ciudad de NEOM, un proyecto nacido de la visión de Arabia Saudita para el 2030. La investigación [6] propone un modelo robusto de predicción de tráfico que integra distintas técnicas de ML con la gestión de tráfico y propone soluciones para los problemas de congestión, los accidentes y los retrasos, los cuales causan incomodidades e incrementan la contaminación y el consumo de combustible.

El modelo propuesto en el estudio referenciado anteriormente aplica técnicas avanzadas de ML, incluyendo redes neuronales profundas y métodos de aprendizaje reforzado, para procesar y analizar datos de tráfico en tiempo real obtenidos a través de dispositivos abordo (OBD) en los vehículos. Estos dispositivos proporcionan información crucial como la ubicación GPS y la velocidad promedio de los vehículos, de esta manera, el modelo es capaz de realizar predicciones precisas sobre las condiciones del tráfico [6].

Además, el sistema utiliza una variedad de técnicas de aprendizaje automático para predecir diferentes aspectos del tráfico, por ejemplo, utiliza el algoritmo *Random Forest* para prever la congestión de tráfico en calles específicas. Estas técnicas permiten que el modelo ajuste sus predicciones basándose en una amplia gama de variables, desde parámetros climáticos hasta el estado de la carretera. Esta predicción avanzada de la congestión permite mejorar en muchos aspectos, siendo uno de ellos la toma de decisiones en lo que refiere a la gestión de tráfico, permitiendo a los vehículos de alta prioridad como ambulancias y otros servicios de emergencia, navegar de manera más eficiente por la ciudad. Esto se logra mediante la integración de un sistema de control de semáforos que ajusta los tiempos de luz roja y verde en tiempo real, basándose en los análisis del modelo de predicción.

Otro ejemplo donde se observa el uso de técnicas de ML para la predicción del tráfico es en el artículo [23], en concreto se explora la utilización del algoritmo *Random Forest* para predecir la congestión del tráfico en entornos urbanos, este es un algoritmo de clasificación que utiliza múltiples árboles de decisión para realizar predicciones buscando resultados de alta precisión, esta técnica se beneficia de características como la hora del día, condiciones meteorológicas, calidad de las carreteras, entre otros. Este documento nos enseña un enfoque que encuentra resultados efectivos en lo que se refiere a la predicción de tráfico.

En este trabajo se pretende desarrollar un modelo de predicción de tráfico utilizando datos históricos de una calle de Madrid para entrenar y validar el modelo, y una vez construido, permitirá prever la congestión en función de las condiciones actuales y pronosticadas. Se busca proporcionar una herramienta valiosa para la toma de decisiones proactivas en cuanto a la gestión de tráfico, más adelante entraremos con más profundidad en este tema.

La relevancia del ML en la gestión de tráfico no se limita solo a la prevención de la congestión; también se extiende a aspectos de seguridad, mantenimiento de infraestructuras y respuesta a emergencias [24], lo que demuestra su impacto integral en la creación de entornos urbanos más inteligentes y reactivos. Este enfoque tecnológico se está convirtiendo en un componente esencial de las *Smart Cities*, impulsando hacia un futuro donde la gestión del tráfico es dinámica, predicha y gestionada con una eficiencia sin precedentes.

2.2.3. FUENTES DE DATOS DE LAS SMART CITIES

En la actualidad, marcada por la globalización y el avance tecnológico, el acceso a la información se ha convertido en un derecho fundamental y en una necesidad práctica para todos los ciudadanos del mundo.

Por esta razón, los datos abiertos son de vital importancia en lo que se refiere a las *Smart Cities*, su disponibilidad fomenta la participación y la transparencia ciudadana, propulsa la innovación y el crecimiento económico. En este contexto, España ha tomado pasos significativos para garantizar a sus ciudadanos y a las entidades dentro de su territorio que puedan acceder y utilizar datos públicos de manera efectiva.

El portal datos.gob.es encarna esta visión al ofrecer una plataforma centralizada en la que se publican datos abiertos de múltiples administraciones públicas. Esta iniciativa, liderada por el Ministerio para la Transformación Digital y de la Función Pública junto con la Entidad Pública Empresarial Red.es, es un componente esencial de la estrategia nacional para promover la utilización de datos como recurso libre y abierto. El portal permite acceder a una amplia gama de datos de todos los sectores, lo que facilita el desarrollo de aplicaciones y servicios que pueden mejorar la calidad de vida de los ciudadanos y aumentar la eficiencia de las empresas, así como también de las administraciones públicas.

El proyecto Ley de Movilidad Sostenible, aprobado por el Consejo de Ministros en febrero de 2024, marca un paso significativo hacia un sistema de transporte más digital e innovador en España. Este proyecto de ley regula aspectos como el transporte según los niveles de demanda y el uso compartido de vehículos haciendo énfasis en la promoción y utilización de datos abiertos de movilidad proporcionados tanto por entidades públicas como por entidades privadas. Establece un marco para el desarrollo de un espacio común europeo de datos de movilidad, lo que posiciona a Europa y específicamente a España, en la vanguardia del desarrollo de sistemas de transporte inteligentes [25].

Este enfoque hacia los datos abiertos también se refleja en la creación del Espacio de Datos Integrado de Movilidad (EDIM), que se crea con el objetivo de brindar una gestión más eficaz y coordinada de la información de movilidad disponible. Esta plataforma integrará datos de diversas fuentes, mejorando la toma de decisiones y la planificación de nuevas infraestructuras y de nuevos servicios de transporte, con el objetivo de fomentar una movilidad más sostenible y eficiente. La Ley de movilidad sostenible busca transformar el panorama del transporte en España, mediante cuatro pilares fundamentales de los que hablaremos a continuación [26]:

- La Movilidad como un Derecho social: Este pilar reconoce la movilidad como un derecho fundamental, garantizando a todos los ciudadanos el acceso a medios de transporte seguros, asequibles y de alta calidad.
- Una movilidad limpia y saludable: Este pilar propone una transición hacia sistemas de transporte menos dependientes del vehículo privado y más orientados a alternativas sostenibles como el transporte público. Este enfoque busca reducir las emisiones de gases de efecto invernadero y con esto promover un estilo de vida más saludable dentro de la población.
- Un sistema de transporte digital e innovador: La digitalización y la innovación tecnológica son esenciales para este nuevo marco de movilidad. La ley fomenta el uso de datos abiertos y la implementación de tecnologías avanzadas para mejorar la gestión de transporte y la experiencia de usuario. Esto incluye la promoción de los vehículos autónomos, y la creación de un entorno experimental (*sandbox*) para probar nuevas soluciones a los problemas de movilidad.
- Invertir mejor al servicio de los ciudadanos: Este pilar se enfoca en asegurar que las inversiones en infraestructura y transporte sean efectivas y generen el máximo valor social posible. La ley propone una mejora en la planificación y financiación de la infraestructura de transporte, de esta forma se busca garantizar que los recursos se destinen a proyectos que realmente mejoren la movilidad y la calidad de vida de los ciudadanos.

Luego de haber visto un poco en que contexto nos encontramos actualmente en cuanto a las fuentes de datos en España, la forma en la que el público accede a los datos y el portal de datos públicos del gobierno, hablaremos sobre la forma en la que se recolectan los datos y su importancia, todo esto enfocado a las *Smart Cities*.

En lo que se refiere a las ciudades inteligentes, la adquisición, gestión y análisis de datos juegan un papel fundamental para la gestión efectiva y la toma de decisiones. Los datos son generados a partir de una gran variedad de fuentes, incluidos sensores IoT distribuidos por toda la ciudad, dispositivos de usuarios, sistemas de transporte, entre otros. Estos datos se recopilan de manera continua, proporcionando una visión en tiempo real del estado y el funcionamiento de la ciudad.

Los datos en las *Smart Cities* son gestionados y procesados a través de plataformas avanzadas que utilizan tecnologías de Big Data y aprendizaje automático para interpretar los datos y de esta manera poder utilizarlos de la manera más eficiente posible, permitiendo monitorear la infraestructura urbana anticipando distintos problemas y así optimizar los servicios de la ciudad [27].

Un aspecto muy importante a tener en cuenta con relación a los datos en las ciudades inteligentes es la seguridad, ya que es un componente crítico que enfrenta numerosos desafíos debido a la complejidad y diversidad de los datos manejados dentro de los ecosistemas urbanos. Con la integración de tecnologías avanzadas, las *Smart Cities* tienen el potencial de transformar la vida urbana mejorando y aumentando la sostenibilidad de los servicios urbanos, pero esto también eleva los riesgos relacionados con la privacidad y la seguridad de los datos.

El artículo [24] hace un enfoque de la tensión que hay entre la necesidad de compartir datos entre diferentes sectores de una *Smart City* y las estrictas leyes de protección de datos que buscan salvaguardar la privacidad de los individuos. Por ejemplo, las regulaciones como el Reglamento General de Protección de Datos (GDPR) imponen restricciones significativas al manejo de datos personales, lo que puede complicar los esfuerzos de integración y uso eficiente de los datos en las *Smart Cities*. Para abordar estos desafíos, el estudio mencionado anteriormente propone un marco sistemático para facilitar la distribución de datos entre los distintos sectores de manera que se cumplan en su totalidad las normativas de protección de datos. Este marco propuesto incluye el uso de tecnologías de cifrado para proteger datos sensibles, así como también la implementación de controles de acceso rigurosos.

2.2.4. LOS DATOS EN LA GESTIÓN DE TRÁFICO

Como se ha explicado previamente, la gestión eficaz del tráfico es fundamental en las ciudades modernas, en donde la congestión y los accidentes pueden causar enormes pérdidas económicas y humanas. Los sistemas de gestión de tráfico avanzados, impulsados por datos precisos y continuamente actualizados, son esenciales para abordar los desafíos que actualmente presenta el tráfico a nivel global.

Para hablar sobre el papel que juegan los datos en los temas referentes a la gestión de tráfico, tenemos que hablar sobre la recopilación de información detallada sobre el flujo vehicular, esto incluye datos sobre volúmenes de tráfico, velocidades, tipos de vehículos, entre otros. Estos datos son recolectados a través de una variedad de tecnologías, como cámaras CCTV, sensores de carretera, y datos proporcionados por vehículos conectados [28].

Cuando los datos son recogidos, son utilizados para alimentar modelos de simulación de tráfico, esenciales para mejorar aspectos de seguridad vial y de planificación urbana, las simulaciones permiten que los planificadores urbanos realicen la planificación óptima de rutas, logren la disminución de los tiempos de viaje y, sobre todo, estas simulaciones permiten evaluar los impactos que tendría la implementación de cambios físicos en la infraestructura vial antes de ser realizados [28].

La implementación de estrategias de gestión de tráfico basadas en datos incluye la operación en tiempo real de sistemas de tráfico inteligente (ITS), estos sistemas utilizan datos en tiempo real para ajustar semáforos y señales de tráfico, gestionar acceso a los carriles y controlar los distintos sistemas de transporte público para optimizar el flujo de tráfico. Por ejemplo, los sistemas adaptativos de control de tráfico pueden ajustar los tiempos de los semáforos basándose en los niveles de tráfico y congestión en tiempo real para minimizar las paradas y mejorar el flujo de tráfico [29].

La gestión de tráfico basada en datos beneficia la eficiencia y seguridad, así como también puede contribuir a la sostenibilidad. Al optimizar el flujo de tráfico, se reduce significativamente el consumo de combustible y las emisiones de gases. Los sistemas de tráfico inteligentes, al reducir las paradas y el tiempo de viaje, promueven lo que se llama conducción eficiente que es tanto económicamente como ambientalmente beneficiosa.

La digitalización en el transporte ha posicionado a los datos como un elemento central en diversas soluciones de movilidad y modelos de negocio de muchas empresas. Dentro del marco de la Estrategia de Movilidad Segura, Sostenible y Conectada 2030, se ha identificado que el MITMA (Ministerio de Transportes, Movilidad y Agenda Urbana) actúe como facilitador del uso de datos abiertos, fiables y de calidad en el sector de la movilidad [30].

En conclusión, los datos son el núcleo de la gestión moderna del tráfico, nos proporcionan las herramientas necesarias para entender, planificar y responder a las distintas situaciones que existen en lo que se refiere al tráfico urbano. A medida que las ciudades continúan creciendo y la tecnología avanza, la integración de los datos en la gestión de tráfico se volverá cada vez más crítica, ofreciendo oportunidades para innovaciones que transformen y mejoren la infraestructura de transporte urbana.

3. DISEÑO DE LA SOLUCIÓN

3.1. DISEÑO DE LA ARQUITECTURA

En el contexto actual, donde el volumen de vehículos en las ciudades sigue aumentando, la gestión eficiente del tráfico se ha convertido en una necesidad crítica. Este trabajo de Fin de Grado aborda directamente esta problemática mediante el desarrollo de un modelo predictivo de la intensidad de tráfico, integrado en un entorno de simulación utilizando SUMO (*Simulation of Urban MObility*). Se busca la realización de una herramienta que permita la integración de un modelo de predicción basado en datos actuales del tráfico de Madrid con un simulador de redes vehiculares. Este proyecto busca abordar las complejidades de la congestión de tráfico mediante la predicción de la intensidad de tráfico. Esta solución propuesta se articula en tres módulos que detallaremos individualmente más adelante en este trabajo:

- **Diseño del conjunto de datos:** El primer paso en la creación de una solución efectiva para la predicción de tráfico es el diseño y la configuración del conjunto de datos. Este paso comienza con la recopilación de datos históricos del tráfico de la ciudad de Madrid. Los datos proporcionados por el Ayuntamiento abarcan una amplia gama de parámetros, entre los que se encuentran los volúmenes de tráfico, los tiempos de ocupación de las calles y los niveles de carga de las vías. Cada variable es crucial para construir una representación precisa del comportamiento del tráfico. El desafío para esta fase es asegurarse de que los datos sean de alta calidad y estén bien estructurados para poder realizar distintos análisis sobre ellos, esta fase se dará por concluida con la obtención de un *dataset* preparado y listo para poder entrenar un modelo de predicción de tráfico utilizando herramientas de *Machine Learning*.
- **Realización del modelo de predicción:** Una vez obtenido el *dataset* final, luego de un proceso de recopilación y limpieza, el siguiente paso es desarrollar un modelo predictivo. El modelo que se busca desarrollar utilizará algoritmos de *Machine Learning* para identificar y analizar los distintos patrones de tráfico y predecir futuras condiciones del tráfico. Esta fase se dará por concluida cuando se obtenga un modelo que mediante distintas pruebas y evaluaciones de sus métricas, sea capaz de realizar predicciones coherentes y lo más precisas posibles. Las fases para la realización de este módulo son las de selección del modelo, entrenamiento del modelo y evaluación del modelo.
- **Integración con SUMO:** La fase final del diseño de la solución es la integración del modelo de predicción presentado anteriormente con una simulación de tráfico utilizando la herramienta SUMO, se busca que esta herramienta permita visualizar y simular el comportamiento del tráfico en diversos escenarios basados en las predicciones del modelo. Esta parte del trabajo es fundamental ya que nos permitiría ver como se comportaría el tráfico bajo condiciones específicas y modificaciones determinadas en la red de simulación. Esta fase consta de la creación de la red en donde se llevarán a cabo las simulaciones, la configuración de la simulación y su integración con el modelo de predicción.

Una vez diferenciados los tres módulos, procedemos a diseñar una arquitectura que cumpla con el objetivo de la integración de un modelo de predicción con una simulación de SUMO. El diseño de la arquitectura de este proyecto está contextualizado para permitir una integración fluida y funcional entre el modelo de predicción y la simulación en SUMO. A continuación, se detallan los componentes clave de la arquitectura propuesta y como estos interactúan entre sí para formar el sistema completo.

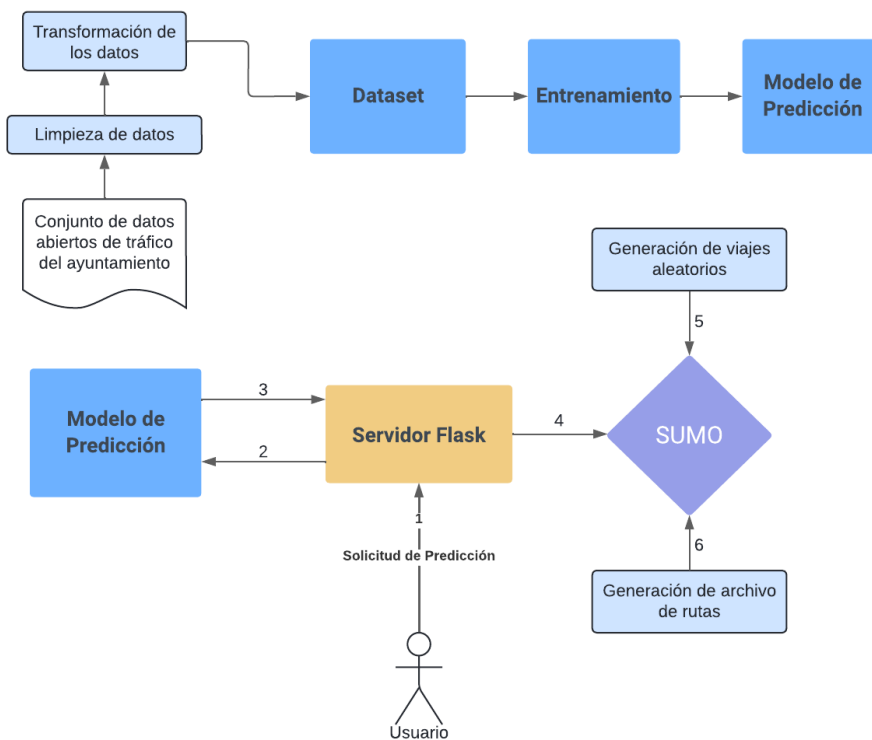


Figura 1: Diagrama de la Arquitectura del Sistema

El diagrama de la arquitectura (Figura 1) se divide en dos partes, la primera parte, la de arriba, hace referencia a la recolección de datos, su procesamiento y el entrenamiento del modelo, y la segunda parte hace referencia al funcionamiento del sistema una vez entrenado el modelo de predicción. A continuación, explicamos los principales componentes de este sistema

1. Dataset:

- Fuente de Datos: Se utilizan datos históricos de tráfico proporcionados por el ayuntamiento de Madrid. Estos datos incluyen intensidades de tráfico, tiempos de ocupación, niveles de carga, fecha en que se tomaron las medidas, entre otros.
- Limpieza y transformación de datos: Antes de su uso, los datos deben ser limpiados, y se realizan distintas transformaciones para que su procesamiento pueda derivar en resultados confiables. También se añaden distintas variables que se consideran importantes para complementar el diseño del conjunto completo de datos. El resultado obtenido luego de esta fase de limpieza y transformación es el *Dataset* final con el que se entrenará el modelo.

2. Modelo de predicción:

- Desarrollo y entrenamiento: Utilizando algoritmos de *Machine Learning* y la biblioteca *sklearn*, se desarrollará un modelo de regresión lineal para predecir la intensidad del tráfico en un momento determinado basado en el *dataset* que se desarrolló previamente.

3. Simulación con Sumo e integración en servidor:

- Se diseñará la red utilizando *OpenStreetMap*, una herramienta de código abierto que nos permite exportar mapas para poder diseñar la red en la que se predecirá y simulará la intensidad de tráfico.
- Mediante el uso de SUMO y las predicciones hechas por el modelo, se generan las rutas necesarias y los viajes aleatorios dentro de la red diseñada para la simulación.
- Un servidor *Flask* gestiona la interacción entre los módulos, automatizando el flujo de datos y asegurando que la información sea procesada y transmitida eficientemente entre ellos. El resultado es una aplicación que integra el modelo de predicción y la simulación realizada para que, mediante la elección de las condiciones iniciales, se inicialice una simulación de manera automática.

El flujo de los datos comienza en el módulo Dataset, ya una vez recopilados y preparados los datos. Con estos datos se entrena el modelo de predicción y se utiliza para generar las predicciones de tráfico. Estas predicciones se traducen en rutas y en parámetros para inicializar la simulación en SUMO, cuyos resultados pueden ser visualizados y analizados para ver el comportamiento del tráfico. Todo el proceso es gestionado mediante un servidor *Flask*, asegurando que cada paso se ejecute correctamente. El resultado es una aplicación sencilla en la que, modificando las condiciones iniciales a gusto del usuario, se generan distintos escenarios de simulación, sobre los que se pueden realizar pruebas y analizar los resultados en base a distintos datos que nos proporciona la herramienta SUMO.

La integración de los tres módulos mencionados será nuestro resultado final, teniendo una herramienta que nos permita predecir la intensidad de tráfico y generar las condiciones iniciales de una simulación, permitiendo el estudio del flujo del tráfico bajo distintas circunstancias. En las siguientes secciones se explicará el proceso de desarrollo de cada uno de los tres módulos que forman este sistema.

3.2. RECOLECCIÓN Y LIMPIEZA DE DATOS

3.2.1. FUENTE DE LOS DATOS

Para el desarrollo de este trabajo, se utilizaron datos proporcionados por el ayuntamiento de Madrid, concretamente de la Dirección General de Gestión y Vigilancia de la Circulación. Estos datos abiertos están formados por registros del tráfico desde el 2013, ofreciéndonos una perspectiva detallada y longitudinal del comportamiento del tráfico en diversos puntos de la ciudad. Los datos de tráfico utilizados en este trabajo son recolectados mediante sistemas de control de tráfico instalados en la ciudad de Madrid. Estos sistemas funcionan a través de una variedad de tecnologías de detección, principalmente sensores y cámaras, distribuidos en puntos estratégicos de la ciudad. Para la recolección de estos datos los distintos organismos de movilidad utilizan equipamientos que permiten el conteo de vehículos en una vía, así como también el grado de ocupación de las mismas. Estos sistemas están compuestos por lazos electromagnéticos que son colocados debajo del pavimento y son capaces de detectar la masa metálica de los vehículos que pasan sobre ellos, proporcionando los datos sobre el tráfico [31].

Utilizando los sensores distribuidos por la ciudad, se crea automáticamente un conjunto de tramas de carreteras que permite conocer con alta precisión la situación del tráfico en el momento de la realización de la medida. Con los datos recopilados por los sensores, se calculan las cargas de tráfico para cada vía y mediante algoritmos de control, se simplifica la información a un formato que los ciudadanos puedan entender fácilmente, visualizando el estado del tráfico en un momento dado.

En este trabajo podemos destacar el uso de dos grandes conjuntos de datos:

- Histórico de datos del tráfico desde 2013:** Este conjunto de datos comprende registros detallados de la intensidad de tráfico, tiempo de ocupación y niveles de carga en los diferentes puntos de medida distribuidos por la ciudad. Permite analizar la evolución del tráfico en cada punto específico. Las medidas están disponibles en archivos csv divididos por meses, y las muestras son tomadas sobre cada punto de medida específico cada 15 minutos. La estructura de los archivos csv es la siguiente:

Nombre	Tipo	Descripción
Id	Entero	Identificación del punto de medida, único e invariable
Fecha	Fecha	Fecha y hora oficiales de Madrid con formato dd/mm/yyyy hh:mi:ss
Tipo_elem	Texto	Nombre del punto de medida: Urbano o M30
Intensidad	Entero	Número de vehículos en el periodo de 15 minutos en veh/hora
Ocupacion	Entero	Porcentaje de tiempo que está un detector de tráfico ocupado por un vehículo
Carga	Entero	Parámetro de carga vial en el periodo de 15 minutos
vmed	Entero	Velocidad media de los vehículos en el periodo de 15 minutos
Error	Texto	Puede tomar 3 valores: N: no ha habido errores E: los parámetros de calidad de alguna de las muestras no son óptimos. S: Muestra totalmente errónea.
Periodo_integracion	entero	Número de muestras recibidas y consideradas para el período de integración

Tabla 2: Campos de los datos históricos de tráfico

- Ubicación de los puntos de medida del tráfico:** Contiene información precisa sobre la localización geográfica de cada sensor de tráfico, incluyendo datos específicos como el tipo de vía (Urbana o M30), y detalles adicionales que permiten una identificación exacta del punto en donde fueron realizadas las medidas, el proceso de identificación del punto de medida se explicará en la siguiente sección. La importancia de este conjunto de datos recae en poder relacionar la ubicación de los puntos de medida con los datos históricos del tráfico, para poder realizar un análisis espacial detallado garantizando la correcta interpretación de los datos. La estructura de los archivos csv de este conjunto de datos es la siguiente:

Nombre	Tipo	Descripción
Tipo_elem	Texto	Puede tomar dos valores: Urb: tráfico urbano M-30: tráfico interurbano
Distrito	Entero	Identifica el distrito al que pertenece el punto de medida.
Id	Entero	Identificación del punto de medida, único e invariable.
Cod_cent	Texto	Código centralizado del punto de medida.
Nombre	Texto	Denominación del punto de medida. Para los puntos de medida de tráfico urbano se identifica con la calle y orientación del sentido de la circulación. Para los puntos de vías rápidas y accesos a Madrid se identifica con el punto kilométrico, la calzada y si se trata de la vía central, vía de servicio o un enlace.
utm_x	Real	Coordenada X del punto que indica la representación del punto de medida en el sistema de referencia de coordenadas ETRS89 / UTM zone 30N (EPSG: 25830)
utm_y	Real	Coordenada Y del punto que indica la representación del punto de medida en el sistema de referencia de coordenadas ETRS89 / UTM zone 30N (EPSG: 25830)
longitud	Real	Longitud en el sistema de referencia de coordenadas WGS 84 (EPSG:4326)
latitud	real	Latitud en el sistema de referencia de coordenadas WGS 84 (EPSG:4326)

Tabla 3: Ubicación de los puntos de medida del tráfico

Una vez sabemos con qué datos contamos, procedemos a la elección del punto de medida en el que se va a basar el modelo, recordemos que el objetivo de este módulo es el diseño de un *Dataset* que tenga todos los datos del tráfico necesarios para entrenar nuestro modelo de predicción para unos puntos de medida determinados. En este trabajo, se decidió predecir el tráfico concretamente para la calle José Ortega y Gasset, utilizando el sentido del tráfico PºCastellana-Serrano y el sentido Serrano-PºCastellana, estos puntos son identificados con el id 3040 y 3041 respectivamente, campo que es mutuo en ambos conjuntos de datos mencionados anteriormente.

El siguiente diagrama ilustra el flujo de trabajo para la recolección y preparación de los datos:

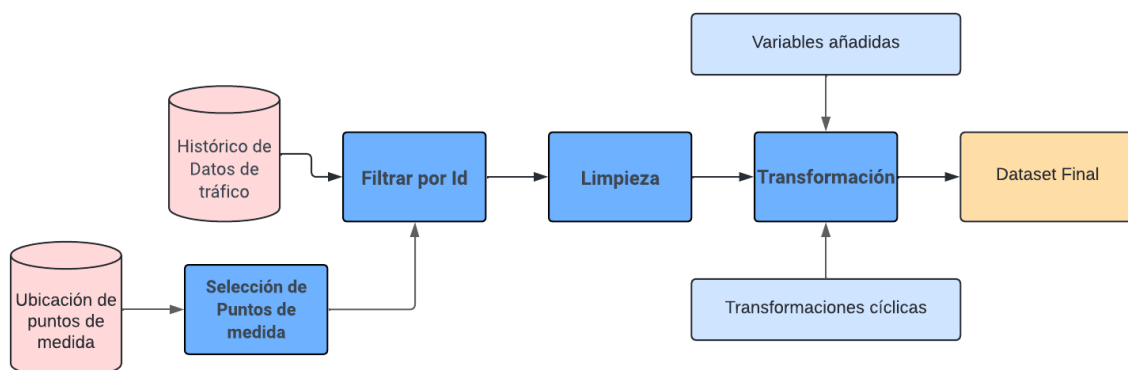


Figura 2: Proceso de diseño del Dataset

3.2.2. SELECCIÓN DE LOS PUNTOS DE MEDIDA

Para la correcta identificación de los puntos de medida de tráfico de Madrid, se empleó la herramienta QGIS, una herramienta avanzada de Sistemas de Información Geográfica (SIG). Es una herramienta de código abierto que permite la visualización, edición y análisis de mapas. Esta aplicación fue utilizada expresamente para ver la ubicación exacta de los puntos de medida procedentes del conjunto de datos. Este proceso se realiza trabajando con los archivos proporcionados por el Ayuntamiento de Madrid.

A través de QGIS, se generó un mapa inicial que, aunque integraba correctamente los puntos de medida, no incluía la visualización de las calles, lo cual era esencial para determinar la ubicación exacta de cada punto, e decir, podíamos acercarnos a un punto, pero no sabríamos su ubicación. A continuación, se enseña el mapa con la ubicación de los puntos de medición utilizando la herramienta QGIS:

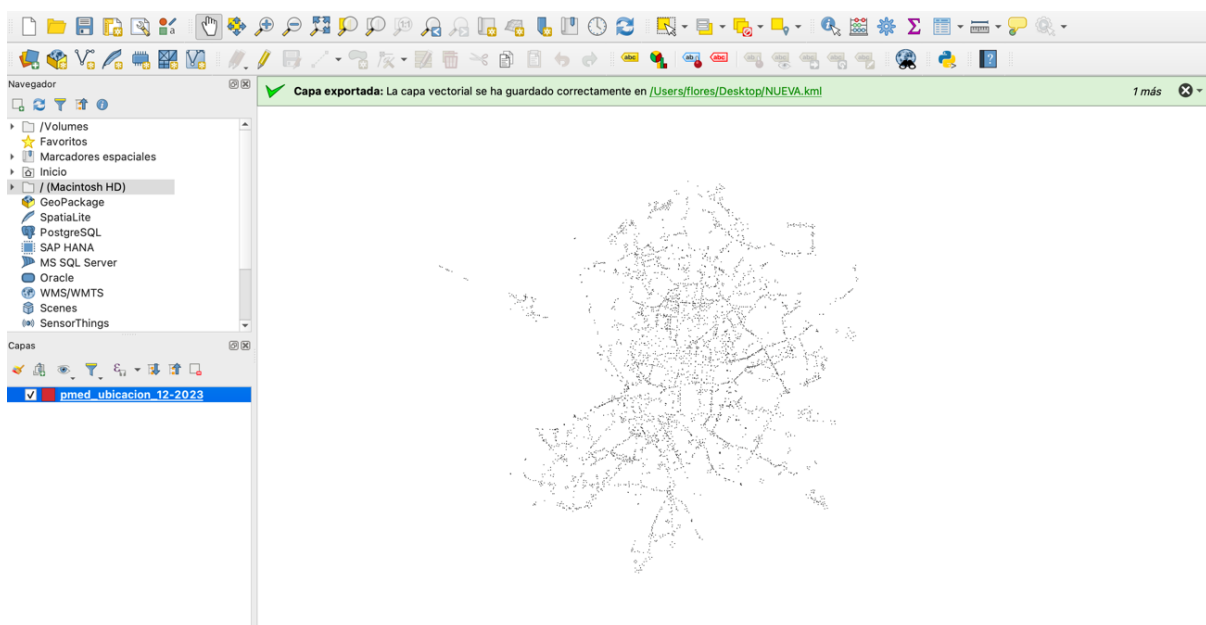


Figura 3: Mapa de los puntos de medida de Madrid

Para resolver esta problemática de la visualización de las calles y así completar el proceso de identificación de los puntos de medida deseados, se creó un archivo KML, compatible con Google Earth. Este formato permitió integrar el mapa de puntos de medida con el mapa detallado del entorno urbano de la ciudad de Madrid.

El procedimiento comienza con la exportación del mapa desde QGIS al formato KML, de esta manera conseguimos un formato que Google Earth puede integrar con sus mapas detallados. Esta transición a Google Earth facilitó una exploración visual de la ubicación de los puntos de medida, es decir, teníamos el mapa de los puntos de medida insertado directamente en Google Earth, lo que nos permitía observar el punto exacto de la ubicación de los puntos de medida. Introduciendo las coordenadas de los puntos deseados, se pudo verificar con exactitud la ubicación de los puntos de medida con id 3840 y 3841, situados en la calle José Ortega y Gasset, en los tramos que conectan con Paseo de la Castellana y Serrano.

Al visualizar las coordenadas de latitud y longitud, se confirmó la posición exacta de los puntos de medida, el conjuntos de datos proporcionado por el ayuntamiento marca la ubicación exacta de los puntos de medida con unas flechas rojas. Se muestran las fotos hechas desde Google Earth con los marcadores correspondientes, cada flecha roja hace referencia a la ubicación de un punto de medida.

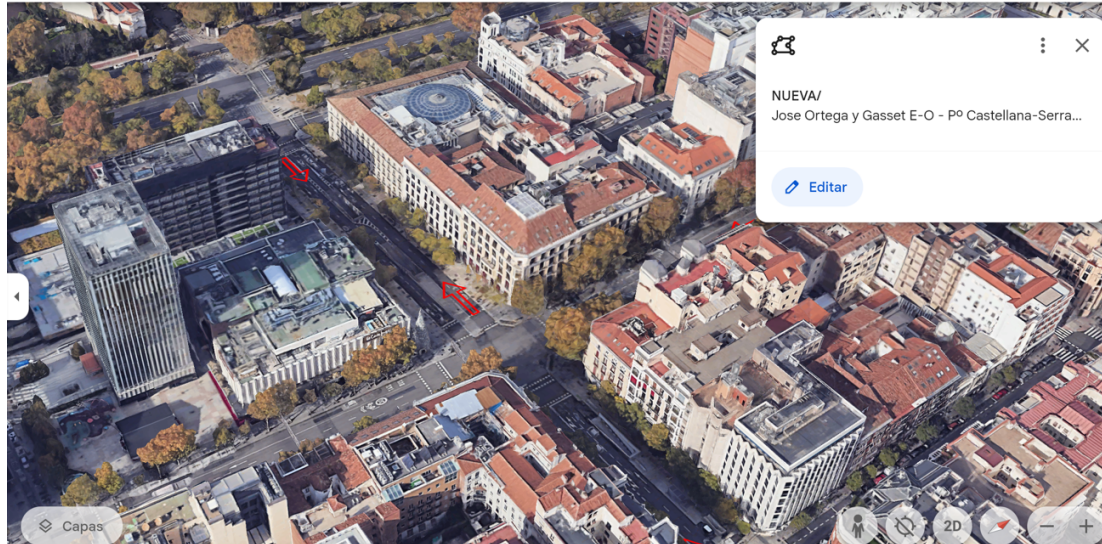


Figura 4: Punto de medida con id 3840

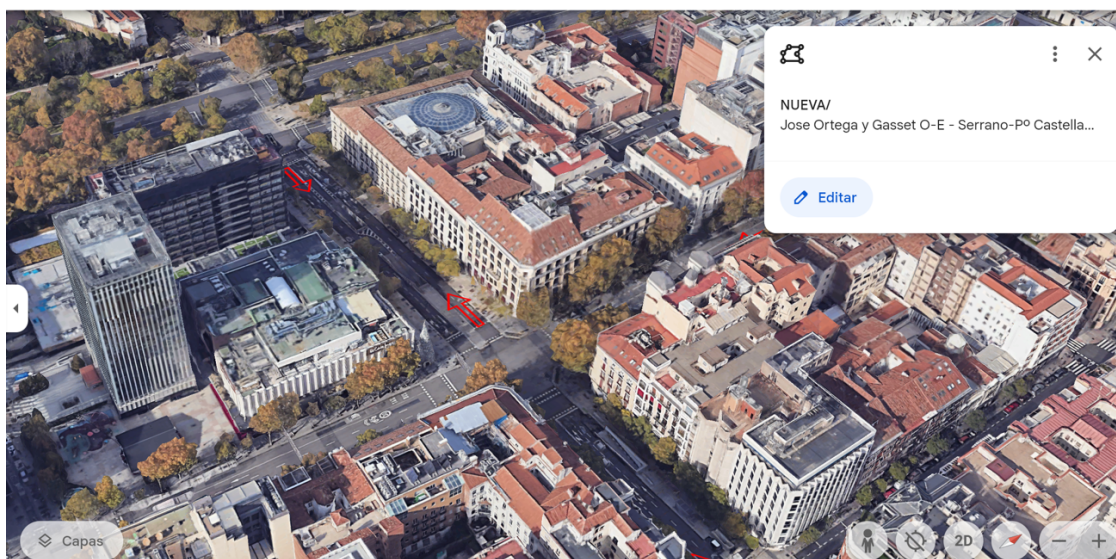


Figura 5: Punto de medida con id 3841

El análisis detallado de la ubicación de los puntos de medida es crucial, ya que nos permite definir la exactitud de los datos de tráfico que serán utilizados para diseñar nuestro *dataset*. Al confirmar estas ubicaciones, se garantiza que los datos tomados en estos puntos de medida reflejan las condiciones reales de estas áreas.

Utilizando Python como herramienta de procesamiento, se llevó a cabo un proceso de filtrado sencillo, los conjuntos de datos **históricos del tráfico** y **ubicación de los puntos de medida de tráfico** tienen

en común el campo “id”, por lo que procedió a unir dichos conjuntos de datos en un solo conjunto basándose en esta variable común. El resultado hasta este punto es un dataset con los datos de tráfico (intensidad, carga, etc...) de los puntos de medida con id 3840 y 3841.

3.2.3. LIMPIEZA DE DATOS, TRANSFORMACIONES REALIZADAS Y CREACIÓN DEL DATASET

Al tener los datos filtrados con los id de interés, se procede a la limpieza de estos, se eliminan las columnas que contienen errores de medición, la columna fecha se convierte a formato *datetime* para facilitar la extracción de los componentes específicos como el mes, el día de la semana y la hora, lo que es crucial para análisis temporales posteriores. Este proceso también incluye la extracción de la hora y los minutos específicos de cada registro, que se almacenan en nuevas columnas del *dataset*.

Dada la proximidad de la calle José Ortega y Gasset al estadio Santiago Bernabéu, se decidió considerar los eventos que se han llevado a cabo en ese lugar en los últimos dos años, ya que tienen un impacto significativo en los patrones de tráfico. Este estadio es sede de eventos de gran escala, incluyendo partidos de fútbol, conciertos y eventos corporativos. Estos acontecimientos atraen grandes multitudes y, en consecuencia, generan un aumento notable en el flujo vehicular en las zonas cercanas al estadio. Es por esto, que dada la proximidad de la calle José Ortega y Gasset con el estadio, se le añadirá al *dataset* una columna que refleje los días en los que hubo algún evento en el Bernabéu, de esta forma el modelo podrá ajustar sus predicciones para ese tipo de días y así mejorar considerablemente la precisión del mismo. Esta información permitirá al modelo predecir las fluctuaciones de tráfico significativas que se producen durante los días en los que dichos eventos tienen lugar.

Otra incorporación que se le realizó al *dataset* fue la de los días festivos, ya que los patrones de movilidad se ven afectados en estos días debido a la disminución de los vehículos en circulación. Los días festivos generalmente alteran los flujos de tráfico habituales debido a que muchas personas no trabajan y los niños no van al colegio, por lo que las rutinas diarias se modifican significativamente. Estos días presentan una reducción en el tráfico matutino, al ajustar el modelo para reconocer los festivos, se espera mejorar la capacidad de anticipar estos cambios, permitiendo una gestión más eficaz del tráfico.

A la hora de optimizar los datos de nuestro *dataset* con el objetivo de que el modelo entienda la forma en la que se le pasan los datos, se realizaron transformaciones cíclicas a las variables temporales. Los parámetros temporales que nos proporciona el ayuntamiento de Madrid son cíclicos en su totalidad. Estas transformaciones son cruciales para capturar la naturaleza cíclica del tiempo y de esta manera permitir que el modelo de predicción comprenda y utilice eficazmente la periodicidad de los datos.

A continuación, se exponen las transformaciones cíclicas realizadas para el diseño del *dataset*:

- **Meses**

El año tiene una estructura cíclica, luego de diciembre sigue enero, creando un ciclo continuo. La representación estándar de los meses (1 a 12) no refleja esta continuidad en los modelos de aprendizaje automático, donde diciembre (12) y enero (1) aparecerían distantes. La transformación a coordenadas cíclicas puede realizarse de una manera sencilla utilizando las funciones seno y coseno para mapear cada mes en un círculo unitario. Esto asegura que diciembre y enero, estén cercanos en la representación

transformada. A continuación, se muestran las ecuaciones utilizadas para modificar la variable de los meses.

$$Mes_sin = \sin \frac{2\pi \times Mes}{12}$$

$$Mes_cos = \cos \frac{2\pi \times Mes}{12}$$

La imagen que se muestra a continuación ilustra la distribución cíclica de los meses del año aplicada a través de las transformaciones mencionadas anteriormente. En esta imagen se visualiza una representación continua y cíclica para los meses:

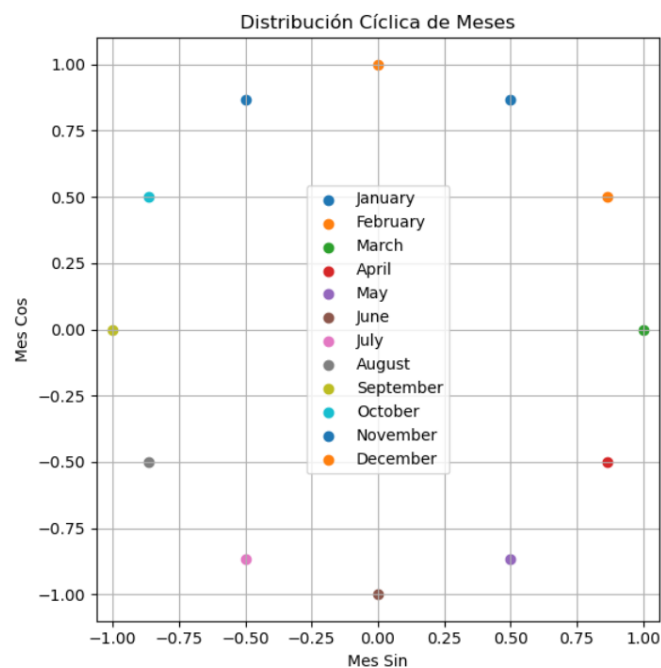


Figura 6: Distribución cíclica de los meses

- **Días de la semana**

De forma similar a los meses, los días de la semana forman un ciclo continuo, donde el lunes sigue al domingo. Representar estos números del 0 (lunes) al 6 (domingo) no refleja la proximidad del domingo al lunes, es por ello que, de manera similar a lo realizado con los meses, los días de la semana se mapean cíclicamente utilizando las funciones seno y coseno, utilizando las siguientes fórmulas:

$$Dia_sin = \sin \frac{2\pi \times Dia}{7}$$

$$Dia_cos = \cos \frac{2\pi \times Dia}{7}$$

- **Horas**

Las horas del día también presentan un comportamiento cíclico evidente, comenzando de nuevo luego de las 23 horas. Mapear las horas en un formato cíclico facilitará al modelo entender la transición entre la noche y la mañana. Este proceso se realiza de la misma forma que los dos anteriores, utilizando senos y cosenos, utilizando las siguientes fórmulas:

$$hora_sin = \sin \frac{2\pi \times hora}{24}$$

$$hora_cos = \cos \frac{2\pi \times hora}{24}$$

Con las transformaciones cíclicas realizadas y con las variables añadidas, se ha completado el proceso de preparación, diseño y creación del *dataset*. Este conjunto de datos ahora incorpora variables enriquecidas y transformadas cruciales para captar dinámicas temporales y estacionales del tráfico. Con este *dataset* terminado, se avanzará al siguiente módulo del diseño de la solución, la selección y el entrenamiento del modelo de predicción, nuestro *dataset* final tiene los siguientes campos:

Columna	Descripción
Intensidad	Intensidad del tráfico (vehículos/hora) escalada.
Carga	Indica el nivel de carga de la vía, valor de 0 a 10.
Fin_de_semana	Indica si el día es fin de semana (1) o no (0).
Hubo_evento	Indica si ese día hubo un evento en el Santiago Bernabéu (1) o no (0).
Festivo	Indica si el día es festivo (1) o no (0).
Periodo_dia_sin	Componente senoidal del período del día.
Periodo_dia_cos	Componente cosenoidal del período del día.
Mes_sin	Componente senoidal del mes.
Mes_cos	Componente cosenoidal del mes.
Dia_sin	Componente senoidal del día.
Dia_cos	Componente cosenoidal del día.
Hora_sin	Componente senoidal de la hora del día.
Hora_cos	Componente cosenoidal de la hora del día.
Minuto_sin	Componente senoidal del minuto.
Minuto_cos	Componente cosenoidal del minuto.

Tabla 4: Parámetros del dataset diseñado

3.3.DISEÑO DEL MODELO DE PREDICCIÓN

3.3.1. SELECCIÓN DEL MODELO

En este trabajo se ha utilizado la biblioteca *Scikit-learn* de Python para la implementación de métodos de aprendizaje automático, debido a su extensa funcionalidad y la amplia adopción que ha tenido en la comunidad científica y tecnológica. *Scikit-learn* proporciona una interfaz consistente para aplicar una variedad de técnicas de modelado de datos y aprendizaje automático, lo que la convierte en una herramienta esencial para la investigación y el desarrollo en el campo del aprendizaje automático.

Scikit-learn es una biblioteca de aprendizaje automático para el lenguaje de programación Python que se ha convertido en una herramienta estándar para muchos que trabajan en el campo de la ciencia de datos. Es una herramienta muy eficiente en lo que se refiere a la implementación de una gama amplia de algoritmos de aprendizaje automático. Esta herramienta facilita la realización de experimentos con diferentes modelos, proporcionando a los usuarios distintas herramientas para realizar comparaciones entre distintos modelos. Como nos exponen en [32], paquete *Scikit-learn* abarca cuatro puntos principales relacionados al *Machine Learning*, la transformación de los datos, el aprendizaje supervisado, el aprendizaje no supervisado y la evaluación y selección de modelos.

Dentro del mundo del aprendizaje automático, se pueden destacar dos grandes ramas, el aprendizaje supervisado y el aprendizaje no supervisado, en este trabajo de fin de grado se utilizó únicamente el aprendizaje supervisado, este se refiere a un subconjunto de algoritmos de aprendizaje automático que tienen como objetivo realizar un mapeo entre las variables de características y sus variables objetivo correspondientes [32].

Este módulo se realizó utilizando técnicas de aprendizaje supervisado para modelar y predecir la intensidad de tráfico a partir de las otras variables que componen nuestro *dataset*, es decir, a partir de las características de los datos históricos del tráfico. En nuestro análisis, tenemos unas variables características, que serían la carga de la vía, la fecha, y todas las variables del dataset, y tenemos por otro lado la variable objetivo, que sería la intensidad. Para nuestro caso, el aprendizaje supervisado es ideal, ya que se requiere que las variables independientes (características) y la variable dependiente (objetivo) estén claramente definidas. Las variables características se mapean contra la variable objetivo, es decir contra la intensidad. Este enfoque permite que este problema sea tratado como un problema de regresión, dado que la variable objetivo es una cantidad continua, se adecua perfectamente a los métodos de regresión dentro del aprendizaje supervisado. Se predicen valores numéricos basados en patrones aprendidos durante el entrenamiento.

3.3.2. ANÁLISIS DEL DATASET

El tráfico en Madrid ha experimentado una evolución significativa desde principios del siglo XX, marcando una era creciente de congestión vehicular y generando nuevos desafíos en la gestión urbana.

Acorde con el análisis histórico [33] el aumento de la población y la expansión del transporte urbano han provocado que los desplazamientos de personas y vehículos se vuelvan cada vez difíciles.

En esta sección, se realiza una comparación entre las variables de entrada con las variables de salida, analizando y representando las características importantes del *dataset* y su comportamiento con relación a la intensidad de tráfico (variable objetivo). Se examinarán aspectos específicos como la influencia de los partidos, los fines de semana, los días festivos y otros factores que se consideran importantes para el análisis de tráfico.

Según la DGT [34], la ciudad de Madrid sufre cortes temporales de tráfico los fines de semana y los días festivos para ampliar el espacio de los peatones. Estas medidas incluyen dispositivos especiales de la Policía Municipal y agentes de Movilidad y hacen que la intensidad de tráfico disminuya durante estos cortes.

El siguiente gráfico presenta una comparación de la intensidad media de tráfico en días festivos versus días no festivos. Podemos observar que la intensidad de tráfico es menor en días festivos respecto a los no festivos, y esto es consistente con la disminución de actividades comerciales, laborales y escolares en dichos días. Esta observación indica que el modelo será sensible a las variaciones estacionales y será capaz de ajustar sus predicciones en consecuencia:

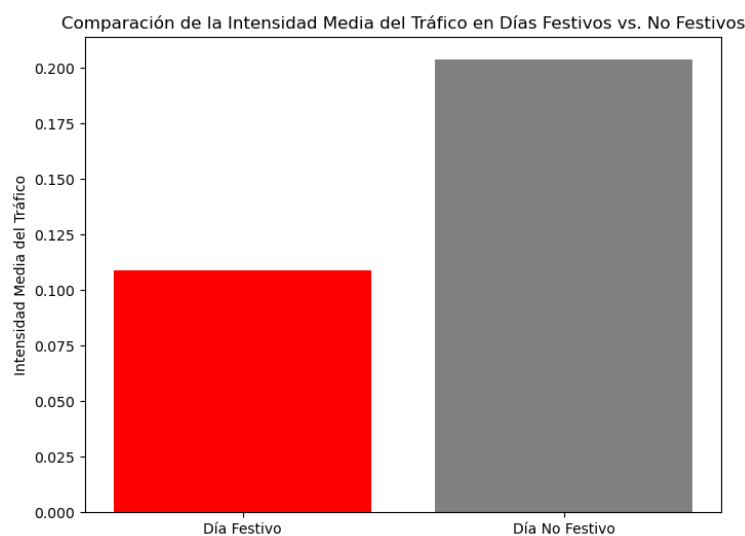


Figura 7: Comparación de la Intensidad media del tráfico en días festivos vs no festivos

La gestión de tráfico en Madrid ha tenido que considerar el impacto de eventos multitudinarios, como partidos de fútbol y manifestaciones, que alteran significativamente los patrones de tráfico. La capacidad de las autoridades para manejar este flujo de personas y vehículos es crucial para mantener el orden y la seguridad en la ciudad [35].

El siguiente gráfico presenta una comparación de la intensidad media del tráfico en los días en los que hay partido en el estadio Santiago Bernabéu y los días en los que no lo hay. La mayor intensidad de tráfico se obtiene en los días que hay partido, como se era de esperar. Esto confirma que un evento deportivo de este tipo afecta considerablemente a la intensidad de tráfico. En la gráfica se puede observar que en un día de partido la intensidad media de tráfico es un 25% mayor que un día sin partido, lo que permitirá al modelo captar esta variación.

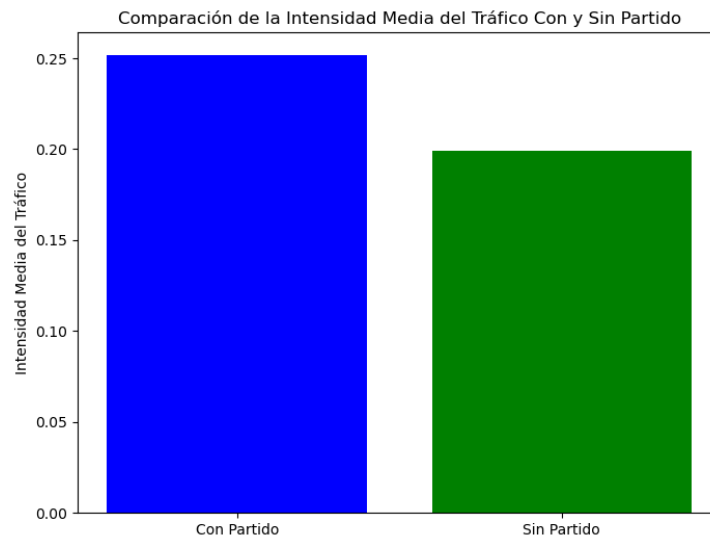


Figura 8: Comparación de la intensidad media del tráfico con y sin partido

El análisis de las variables temporales del *dataset* revela que estas se comportan de manera coherente con las expectativas basadas en el comportamiento típico del tráfico urbano. Podemos comprobar que el comportamiento de las horas pico se reflejan correctamente en la intensidad del tráfico, según un estudio realizado por la comunidad de Madrid [36], se registran tres intensidades punta que coinciden aproximadamente con los horarios de entrada y salida laboral.

Según se ve en el siguiente gráfico, las horas de mayor intensidad de tráfico concuerdan con lo expuesto en el estudio mencionado anteriormente, donde las horas de mayor congestión coinciden con los periodos esperados de mayor actividad vehicular, como lo son alrededor de las 14:00 horas y las horas de la mañana y tarde que corresponden a los horarios de entrada y salida laboral.

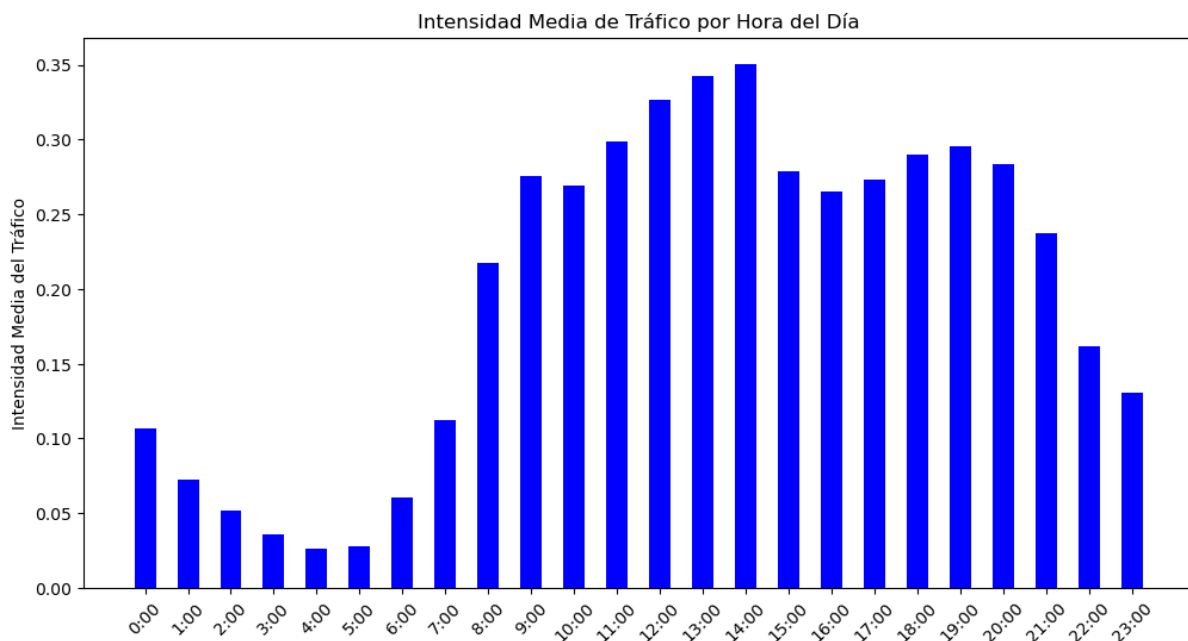


Figura 9: Intensidad media del tráfico por hora del día

Este mismo proceso se realizó con los días de la semana y con los meses, observando que los patrones de la intensidad de tráfico son consistentes con la disminución esperada durante los fines de semana y el aumento durante los días laborales.

De esta manera, se validan las características añadidas al *dataset*, se demuestra que efectivamente las variables temporales capturan adecuadamente las tendencias del tráfico luego de la realización de las transformaciones, es decir, estas transformaciones se realizaron de la manera correcta, y los patrones temporales son captados de manera cíclica, cosa fundamental para el entrenamiento de nuestro modelo.

3.3.3. ENTRENAMIENTO DEL MODELO

En el desarrollo de este proyecto, se optó por abordar el problema de la predicción de la intensidad de tráfico como un problema de regresión lineal. Esta decisión se basó en su simplicidad y efectividad para establecer una relación lineal entre las variables independientes y la variable dependiente. También se realizaron pruebas utilizando modelos más como *Random Forest*, pero se seleccionó la regresión lineal debido a su eficiencia computacional y su capacidad para proporcionar buenos resultados.

El proceso de la implementación del modelo comenzó con la carga y preparación de los datos, todo el proceso se realiza y se valida utilizando *Jupyter Notebook*. El proceso de entrenamiento comienza con el uso de la librería *Pandas* para importar el conjunto de datos (*dataset*) desde un archivo CSV. Este paso es fundamental para que los datos se carguen correctamente y utilizando la función *head()* de *Pandas*, comprobamos que la estructura del *DataFrame* es la adecuada y que los datos están listos para su futuro procesamiento.

Posteriormente, procedemos a la selección de características y la definición de la variable objetivo. En este caso, la variable objetivo es la intensidad de tráfico, y las características incluidas en el análisis son todas las que se encuentran en el conjunto de datos menos la intensidad y algunas características que fueron descartadas ya que no contribuían al modelo o eran redundantes. Este paso determina los datos que influyen directamente en la predicción.

En lo que se refiere a la eficacia de los modelos predictivos en el ámbito del aprendizaje automático, es crucial la división del conjunto de datos completo en subconjuntos de entrenamiento y prueba, esto se realiza utilizando la función *train_test_split*, proporcionada también por *Scikit-learn*. Este proceso se realiza para evaluar la capacidad del modelo para predecir datos no vistos durante el entrenamiento. Los datos se dividieron en una proporción 70/30, es decir, el 70% de los datos de uso para entrenamiento y el otro 30% se utilizó para realizar pruebas, con esto aseguramos que el modelo no solo se ajuste a los datos con los que se entrena, sino que también mantenga su precisión en datos externos, minimizando de esta manera el riesgo de sobreajuste.

En la fase de entrenamiento del modelo, se utilizó el subconjunto de entrenamiento para ajustar los parámetros del modelo. La regresión lineal busca encontrar una línea que mejor se ajuste a los datos, minimizando la suma de las diferencias cuadráticas entre los valores observados reales y los valores predichos por el modelo [37]: Este proceso se conoce como minimización del Error Cuadrático Medio (ECM), que se formula como:

$$ECM = \frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2$$

Donde Y_i son los valores reales y Y'_i son los valores resultado de una predicción, siendo n el número de observaciones realizadas.

Además, se calcula también el coeficiente de determinación (R^2), que mide la cantidad de variación en la variable dependiente (predecible) a partir de las variables independientes. El R^2 es un indicador del ajuste del modelo, donde un valor cercano a 1 sugiera que el modelo es capaz de explicar en gran proporción la variabilidad de los datos observados, concretamente, un valor de (R^2) igual a 1 significa un ajuste lineal perfecto, significa que la totalidad de la variación de la variable Y es explicada por el modelo [38]. El coeficiente de determinación (R^2) se formula como:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - Y'_i)^2}{\sum_{i=1}^n (Y_i - Y''_i)^2}$$

Donde Y'' es el promedio de los valores observados.

Para visualizar la eficacia del modelo, se generó un gráfico que compara los valores reales con los predichos en función de los días de la semana, de esta forma podemos evaluar visualmente cuales son los resultados del modelo. Más adelante, en la sección de resultados, profundizaremos en el análisis del rendimiento del modelo, discutiendo las métricas específicas del Error Cuadrático Medio (ECM) obtenido y el coeficiente de determinación (R^2).

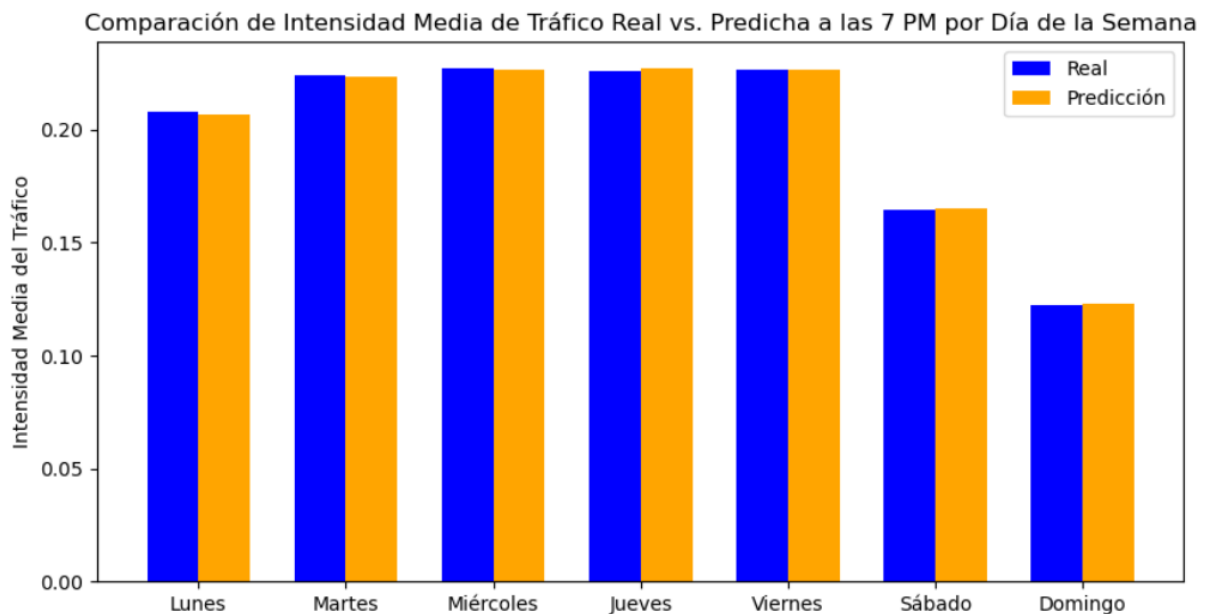


Figura 10: Figura ilustrativa sobre el Rendimiento del Modelo

3.4. INTEGRACIÓN CON SUMO

3.4.1. PREPARACIÓN DE DATOS PARA SUMO

Para la realización de este trabajo, se optó por usar SUMO (*Simulation of Urban MObility*), un software de simulación de tráfico de código abierto ampliamente reconocido por su capacidad de modelar el tráfico de manera microscópica, es decir, simula cada vehículo y las interacciones que tienen con la red de forma individual. La elección de SUMO se debe principalmente a su accesibilidad como herramienta de código abierto, su comunidad activa que garantiza un soporte continuo y una actualización constante de sus funcionalidades. A continuación, se habla un poco de lo que es SUMO y sus principales características, así como también de los aspectos necesarios para realizar una simulación de manera adecuada.

SUMO (*Simulation of Urban MObility*) es un simulador de tráfico microscópico de código abierto que permite modelar de forma detallada el comportamiento de cada vehículo dentro de grandes redes vehiculares. Es una herramienta diseñada para la investigación del tráfico a través de simulaciones microscópicas, donde cada vehículo y sus dinámicas se manejan de manera individual.

Este simulador se distingue por ser un software de código abierto y está disponible bajo licencia pública de Eclipse. Ha sido adoptado globalmente y es descargado más de 35.000 veces al año, lo que refleja su relevancia y utilidad en la comunidad de investigación del tráfico. SUMO destaca por su gran facilidad de uso y su capacidad para manejar distintos escenarios de tráfico que pueden ser descargados, compartidos y utilizados por investigadores, para facilitar de esta manera la implementación de estudios viales sin necesidad de configurar un escenario completo desde cero[39].

Esta herramienta permite la creación de redes de tráfico detalladas que incluyen datos de infraestructura como semáforos y peatones, así como el modelado del transporte público, lo que permite simular escenarios de tráfico realistas y complejos, todos los aspectos de SUMO son detallados en su manual de usuario [18].

Estructura de una simulación

La simulación en SUMO requiere de varios ficheros de configuración en formato XML que describen la red de tráfico, las rutas de los vehículos y las configuraciones personalizadas de la simulación, los ficheros necesarios son los siguientes:

- **Fichero de red (.net.xml):** Define la estructura de la red vial, incluyendo intersecciones (*nodes*), carreteras (*edges*), y la topología del entorno. Este archivo especifica detalles como la velocidad permitida, el número de carriles, las señales de tráfico y las rutas que siguen los vehículos dentro de la red.
- **Fichero de rutas (.rou.xml):** En este fichero se definen las rutas y los viajes que sigue cada vehículo, se define en que instante de la simulación aparecerán los vehículos dentro de la red. En este fichero se definen también los tipos de vehículos y sus parámetros específicos.

- **Fichero de configuración (.sumocfg):** Es el fichero básico donde se lanza la simulación, en este archivo se hace el llamado al archivo de red que utilizará la simulación, el archivo de rutas correspondiente y los archivos adicionales que se quieran añadir para personalizar la simulación. En este fichero se establece la duración de la simulación.
- **Ficheros adicionales:** Los ficheros adicionales pueden incluir configuraciones de elementos externos como detectores de tráfico, definiciones de infraestructura adicional como semáforos, entre otros. Son ficheros que se añaden para poder realizar una simulación más completa y son llamados en el archivo de configuración.

Para la creación de estos ficheros, SUMO viene acompañado de distintas herramientas que se discutirán más adelante, siendo las siguientes las principales:

- **netedit:** Una herramienta básica que nos permite desde la creación hasta la modificación de redes de tráfico, aquí se modifican los cruces, la infraestructura, los caminos, entre otras muchas opciones que nos da la herramienta.
- **netconvert:** Permite la conversión de mapas externos en un formato de red que SUMO pueda utilizar para realizar una simulación.

Para el desarrollo de este trabajo, se utilizan todos estos archivos partiendo de una intensidad de tráfico predicha en un momento concreto seleccionado por el usuario. En las siguientes secciones, se hablará del proceso detallado de creación de la simulación y de la integración de la misma con el modelo de predicción.

3.4.2. DISEÑO DE LA RED DE SIMULACIÓN

Uno de los procesos fundamentales para el desarrollo de este trabajo, es el diseño y la creación de una red de simulación precisa que refleje las condiciones reales de la calle José Ortega y Gasset en la ciudad de Madrid. En esta sección se explica el proceso detallado de la creación de la red de simulación, desde que los datos de la zona son exportados utilizando una herramienta externa (*OpenStreetMap*) hasta que son transformados y validados a un formato en el que SUMO pueda entenderlo. Los pasos para la correcta creación de la simulación son los siguientes:

- **Exportación del mapa de la zona utilizando OpenStreetMap (OSM):**

OpenStreetMap es un proyecto colaborativo que permite la creación de mapas editables libres de ciudades y regiones de todo el mundo. Los datos geográficos de OSM son gratuitos y accesibles, y poseen información detallada sobre las características urbanas como calles, caminos, semáforos, y más. Para este paso, se selecciona la zona en la que se va a realizar la simulación, que es la zona en la que se recopilaban los datos y se entrenó el modelo de predicción previamente, y se descargan los datos en un formato OSM XML, que contiene todos los detalles de la zona.

Nuestra zona de simulación es la calle José Ortega y Gasset, se muestra una imagen de la herramienta *OpenStreetMap* en la zona seleccionada:

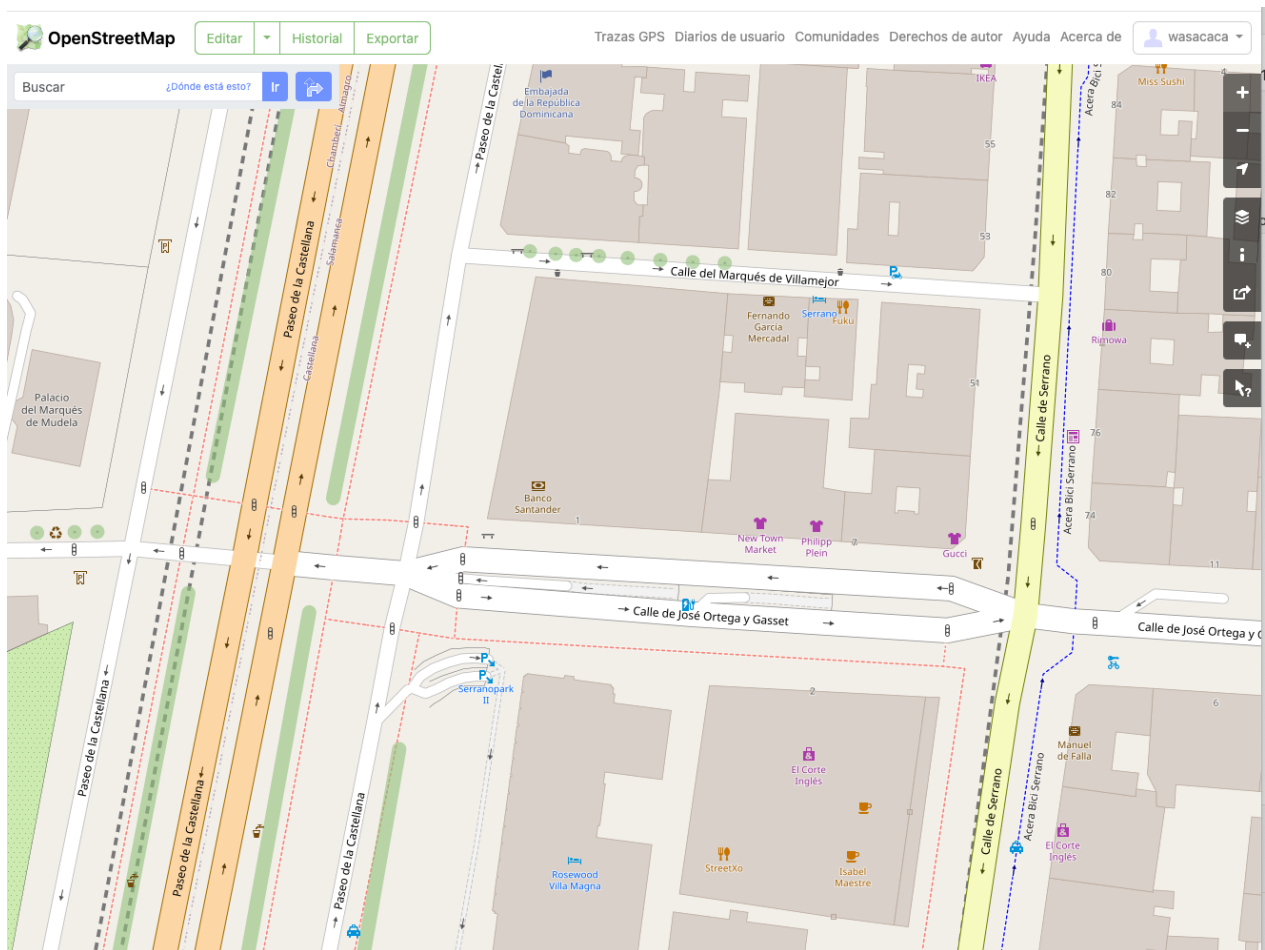


Figura 11: Exportación de datos OSM para la simulación

- **Transformación de un archivo OSM a un archivo .net.xml:**

Una vez obtenido el archivo OSM, el siguiente paso es convertir este archivo a un formato que sea compatible con SUMO, específicamente un archivo .net.xml que describa la red vehicular. Para esta fase, se utilizó la herramienta *netconvert*, proporcionada por el simulador, que es capaz de interpretar los datos que están en un formato OSM y traducirlos a un archivo .net.xml, archivo que SUMO es capaz de procesar para realizar una simulación.

Con el archivo .net.xml correctamente definido, se utiliza la herramienta *netedit*, también proporcionada por el simulador, para revisar las conexiones de la red. Esta herramienta permite visualizar y modificar interactivamente la red de tráfico, facilitando la identificación y corrección de errores de carriles, o problemas en la definición de rutas. Este paso es muy importante para garantizar el correcto funcionamiento de todos los elementos de la red.

Para el diseño de esta red, se escogió la calle para la que el modelo de predicción es capaz de predecir el tráfico, y se añadieron las calles cercanas para poder obtener información de los cruces con las calles próximas, que son Paseo de la Castellana y la calle de Serrano, de esta manera con la intensidad predicha, se podrá observar y estudiar el comportamiento del tráfico en las proximidades de la zona.

A continuación, se muestra una imagen de la red completa con los nodos correctamente editados para el correcto flujo del tráfico dentro de la misma:

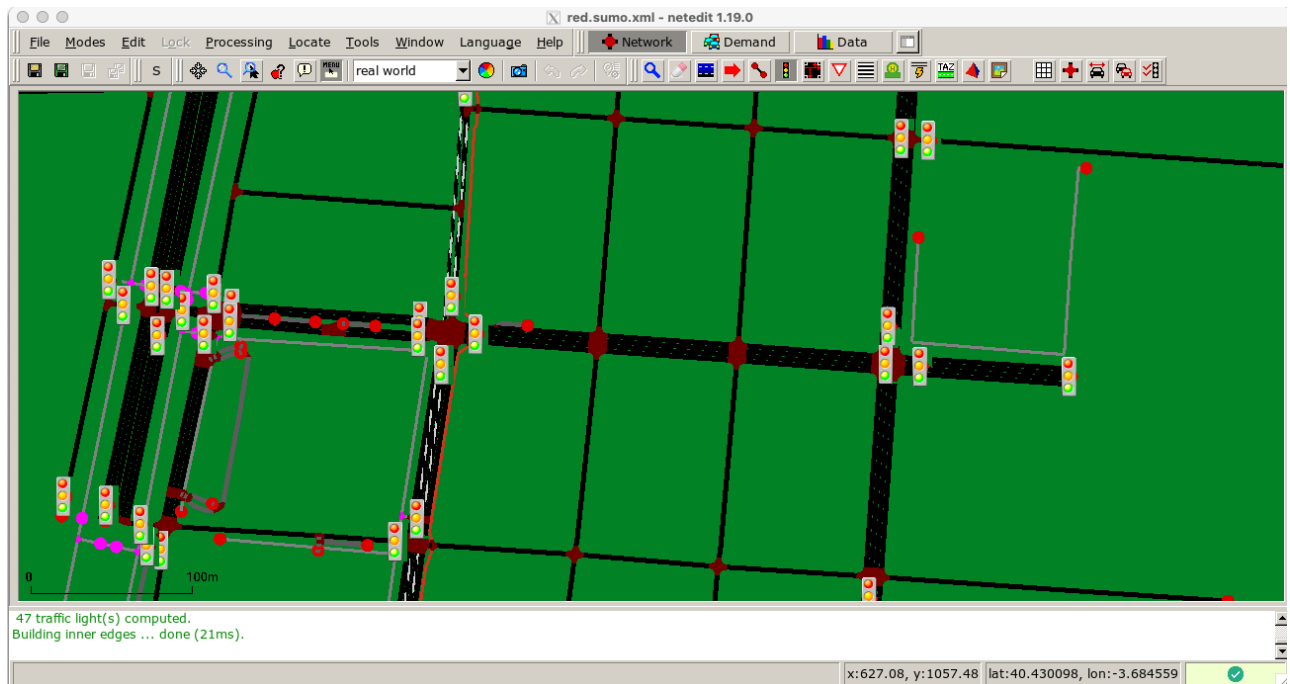


Figura 12: Red de simulación

Una vez diseñada la red de la simulación, es decir, el archivo *.red.xml*, el siguiente paso es la creación de la simulación, para llevar a cabo esta simulación se sigue un proceso detallado que integra la generación de viajes aleatorios, la creación de rutas y la ejecución de la simulación. Este proceso se realiza mediante la utilización de distintas herramientas en forma de comando proporcionadas por el simulador, como se detalla a continuación:

Generación de demanda de tráfico

El primer paso para la creación de la simulación es generar la demanda de tráfico, esto hace referencia a los viajes individuales que los vehículos realizarán en la red. Para esto, se utiliza el script *randomtrips.py* que viene incluido en las herramientas de SUMO. Este script utiliza el archivo *.rou.xml* de la red definida para generar rutas y viajes aleatorios. El comando utilizado es:

```
python $SUMO_HOME/tools/randomTrips.py -n red.sumo.xml -e 3600 -p {interval:.2f} -o demanda.trips.xml
```

- **-n red.sumo.xml:** Especifica el archivo de red de la simulación a utilizar.
- **-e 3600:** Define el tiempo total de la simulación en segundos. En esta simulación la duración es de 1 hora.
- **-p {Interval:.2f}:** Este es el periodo entre la salida de los vehículos, se obtiene proveniente de la intensidad de tráfico que el modelo predice para unas condiciones iniciales determinadas. Este intervalo es el que asegura que la generación de los vehículos sea acorde con la intensidad de tráfico predicha. Es un número entero y se mide en segundos.

- **-o demanda.trips.xml:** Indica el archivo de salida donde se guardarán los viajes generados

Creación de rutas

Una vez generado el archivo de la demanda de viajes, el siguiente paso es convertir estos viajes en rutas específicas que los vehículos seguirán dentro de la red una vez iniciada la simulación. Esto se realiza mediante el uso de *duarouter*, también proporcionado por SUMO. El comando para esta conversión es el siguiente:

```
$SUMO_HOME/bin/duarouter --route-files demanda.trips.xml --net-file red.sumo.xml --output-file rutas.rou.xml
```

- **--route-files demanda.trips.xml:** Especifica el archivo de viajes generados anteriormente.
- **--net-file red.sumo.xml:** Hace referencia al archivo de red utilizado para la simulación.
- **--output-file rutas.rou.xml:** Es el archivo resultante que contiene las rutas calculadas para cada viaje, es llamado desde el archivo de configuración de SUMO para iniciar la simulación.

Ejecución de la simulación

Como paso final, la simulación se ejecuta utilizando el simulador principal de SUMO. El comando para iniciar la simulación es:

```
$SUMO_HOME/bin/sumo -c archivo.sumocfg
```

- **-c archivo.sumocfg:** Este archivo de configuración SUMO contiene referencias a todos los archivos que se generaron anteriormente y es el archivo que ejecuta la simulación, incluye la red, las rutas, la duración de la simulación y cualquier ajuste adicional realizado. Este archivo asegura que todos los componentes de la simulación estén correctamente enlazados y configurados. Al ejecutar este comando se lanza la simulación directamente, en el caso de que quisiéramos ver la simulación de una forma visual, se sustituiría *sumo -c* por *sumo-gui* para abrir la interfaz gráfica de SUMO y poder observar el recorrido de los vehículos de forma interactiva.

3.4.3. INTEGRACIÓN DE LA SIMULACIÓN

Para la integración entre el módulo de predicción y la simulación de tráfico, se utilizó un servidor *Flask*. Esta integración permite a los usuarios mediante un formulario seleccionar distintos parámetros que influyen en las condiciones del tráfico de la zona, y mediante una predicción del modelo el usuario obtiene una predicción de la intensidad de tráfico de la zona y le permite lanzar una simulación basada en la predicción realizada. En esta sección, se detalla el proceso paso a paso, incluyendo las herramientas utilizadas y el flujo de interacción entre el usuario y la aplicación.

Herramientas utilizadas

- **Flask:** Es un *microframework* de Python utilizado para mejorar solicitudes HTTP, renderizar plantillas HTML y gestionar sesiones de usuario. Facilita la creación de aplicaciones web ligeras y escalables.
- **Joblib:** Es una biblioteca de Python utilizada para serializar objetos, en este caso, se utiliza para cargar el modelo de predicción de tráfico entrenado.
- **numpy:** Es una biblioteca de Python, que se emplea para realizar cálculos matemáticos, en este trabajo se utiliza para la transformación de los datos introducidos por el usuario a coordenadas cíclicas para que el modelo pueda realizar la predicción.
- **subprocess:** Es un módulo de Python que permite ejecutar comandos en la terminal desde el código de Python. En este trabajo se utiliza para ejecutar los comandos de SUMO automáticamente desde la aplicación.
- **xml.etree.ElementTree:** Es una biblioteca de Python que se utiliza para manipular y procesar archivos csv, se utiliza para leer y mostrar los resultados de la simulación.

Flujo del proceso

El proceso empieza con el usuario accediendo a la aplicación web y se encuentra con un formulario en el que puede ingresar datos relevantes sobre las condiciones de tráfico que quiere simular, este formulario a rellenar contiene indicadores binarios para identificar si es fin de semana o no, si hubo algún evento en el estadio Santiago Bernabéu, si es festivo o no, además se solicita la fecha específica, concretamente el día de la semana y el mes, así como también la hora del día a la que se quiere predecir la intensidad.

Una vez que el usuario rellena el formulario y pulsa el botón predecir, se envían mediante una solicitud POST y *Flask* procesa los datos ingresados. Debido a que el modelo debe recibir los datos cíclicos en forma de seno y coseno, estos datos son transformados en coordenadas cíclicas mediante funciones trigonométricas. Estos datos transformados, junto con los otros datos de entrada, se utilizan como entrada para llamar al modelo de predicción de tráfico que se desarrolló anteriormente. El modelo previamente entrenado es cargado utilizando *joblib*. El modelo predice la intensidad de tráfico en vehículos por hora, y esta predicción se almacena en la sesión del usuario para su uso posterior.

Una vez obtenida y almacenada la predicción de tráfico, el usuario puede ejecutar una simulación de tráfico mediante otra solicitud POST. *Flask* utiliza la predicción de la intensidad de tráfico para calcular el intervalo de tiempo entre la generación de vehículos de la simulación. Se ejecutan los comandos necesarios automáticamente utilizando *subprocess* para interactuar con las herramientas de SUMO necesarias para lanzar la simulación. En este punto, la simulación ya ha sido correctamente ejecutada y los resultados se guardan en un archivo XML (*summary.xml*).

Finalmente, *Flask* lee el archivo de resultados *summary.xml* para extraer los datos relevantes de la simulación. Estos datos incluyen parámetros como el tiempo, la cantidad de vehículos insertados, las colisiones, la velocidad media de los vehículos, el tiempo medio de espera, entre otros parámetros. La aplicación presenta al usuario una tabla donde se puede observar de una forma detallada el estado final de tráfico simulado, es decir, la última fila del archivo *summary.xml* que hace referencia al último instante de la simulación.

```

<> summary.xml
27 <summary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://sumo.dlr.de/xsd/summary_file.xsd">
235 <step time="207.00" loaded="9" inserted="7" running="2" waiting="0" ended="5" arrived="5" collisions="0" teleports="0" halting="1" stopped="0" meanWaitingTime="0.86" mear
236 <step time="208.00" loaded="9" inserted="7" running="2" waiting="0" ended="5" arrived="5" collisions="0" teleports="0" halting="0" stopped="0" meanWaitingTime="0.86" mear
237 <step time="209.00" loaded="9" inserted="7" running="2" waiting="0" ended="5" arrived="5" collisions="0" teleports="0" halting="0" stopped="0" meanWaitingTime="0.86" mear
238 <step time="210.00" loaded="9" inserted="7" running="2" waiting="0" ended="5" arrived="5" collisions="0" teleports="0" halting="0" stopped="0" meanWaitingTime="0.86" mear
239 <step time="211.00" loaded="9" inserted="7" running="2" waiting="0" ended="5" arrived="5" collisions="0" teleports="0" halting="0" stopped="0" meanWaitingTime="0.86" mear
240 <step time="212.00" loaded="9" inserted="7" running="2" waiting="0" ended="5" arrived="5" collisions="0" teleports="0" halting="0" stopped="0" meanWaitingTime="0.86" mear
241 <step time="213.00" loaded="9" inserted="7" running="2" waiting="0" ended="5" arrived="5" collisions="0" teleports="0" halting="0" stopped="0" meanWaitingTime="0.86" mear
242 <step time="214.00" loaded="9" inserted="7" running="2" waiting="0" ended="5" arrived="5" collisions="0" teleports="0" halting="0" stopped="0" meanWaitingTime="0.86" mear
  
```

Figura 13: Ejemplo de fichero de salida de una simulación

4. RESULTADOS

4.1. EVALUACIÓN DEL MÓDULO DE PREDICCIÓN

Para evaluar el módulo de predicción, se utilizó un conjunto de datos de prueba que contiene características relacionadas con el tráfico, como se menciona en la sección 3.3 del trabajo, el modelo fue evaluado mediante métricas de regresión como lo son el error cuadrático medio (MSE) y el coeficiente de determinación (R^2).

El error cuadrático medio (MSE O EMS) es una métrica fundamental utilizada para evaluar la precisión de los modelos de regresión. Se calcula como la media de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo. En los análisis de regresión, el error cuadrático medio se refiere al error medio de predicción. Un MSE cercano a cero indica que el modelo tiene una buena capacidad de predecir con precisión, ya que las diferencias entre los valores reales y predichos son menores, por el contrario, un MSE alto sugiere que hay una gran diferencia entre los valores reales y los valores predichos [40].

El modelo entrenado logró un MSE de 0.007521, lo que indica que el promedio de los errores al cuadrado entre las predicciones y los valores reales es muy bajo. Mediante esta métrica podemos afirmar que el modelo predice con alta precisión y que los errores cometidos son pequeños [37].

Otra métrica utilizada para evaluar el rendimiento del modelo fue el coeficiente de determinación (R^2), el modelo obtuvo un R^2 de 0.753, lo que implica que alrededor del 76% de la variabilidad de la intensidad de tráfico es explicada por el modelo. Este alto valor del coeficiente de determinación indica que el modelo captura bien las distintas relaciones entre las variables de entrada y la intensidad del tráfico [38]. El R^2 es una medida sin unidades que hace referencia a la correlación de las variables características con la variable objetivo (en este modelo, a la intensidad de tráfico), esta medida, cuanto más se acerque a 1, las variables de entrada pueden explicar en mayor proporción la varianza de la variable objetivo [41].

Uno de los análisis realizados para medir la eficacia del modelo fue el análisis de la intensidad de tráfico predicha vs real a lo largo de las 24 horas del día. En el gráfico que se muestra a continuación, las barras azules representan la intensidad media del tráfico real, mientras que las barras naranjas representan la media de las predicciones del modelo a cada hora del día. Este gráfico revela los patrones típicos de tráfico a lo largo del día. Se observa que la intensidad de tráfico es baja a primeras horas de la madrugada, y luego empieza a aumentar cuando empieza el horario laboral y escolar, vemos también que el pico del tráfico diario se alcanza a las dos de la tarde. Se siguen los parámetros esperados

La cercanía entre las barras de predicción y las barras de los valores reales a lo largo de todas las horas del día demuestra la capacidad del modelo para predecir con precisión la intensidad de tráfico. Se observa que durante las horas pico las predicciones del modelo son altamente precisas, lo que es crucial para aplicaciones prácticas, ya que estas horas son las más críticas en lo que se refiere a la gestión de tráfico.

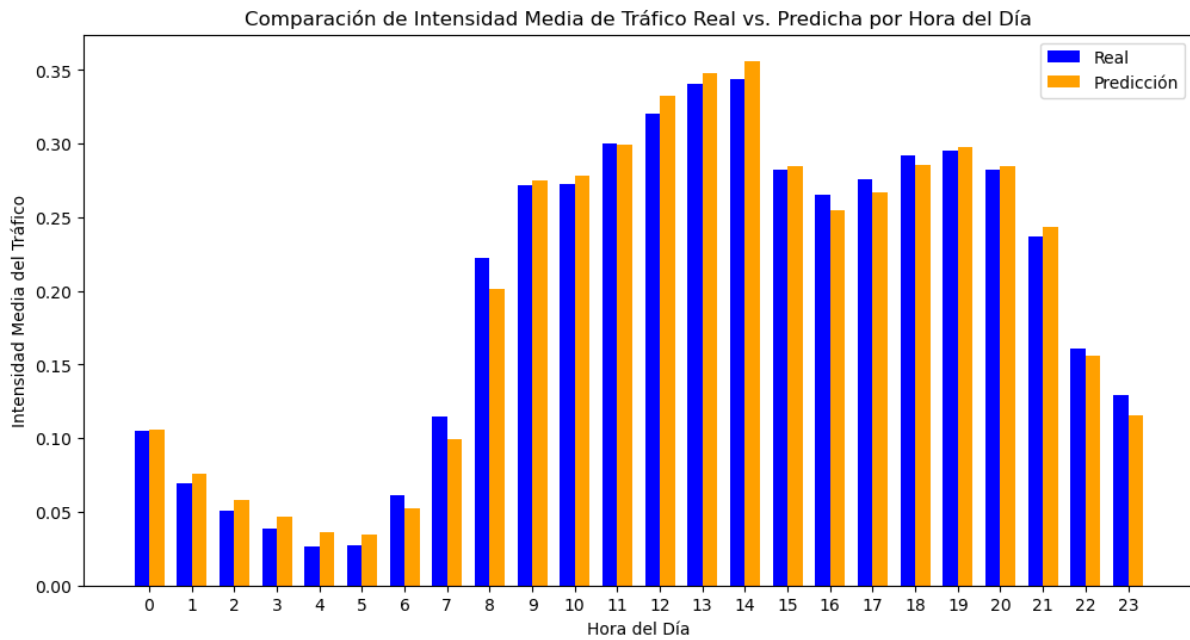


Figura 14: Comparación de la intensidad media de tráfico real y predicha a lo largo del día

Adicionalmente, se examinó la influencia que los eventos especiales tienen en la intensidad de tráfico. Para este análisis, se creó un *dataframe* que contenía las intensidades reales y predichas, junto con una columna indicando la presencia de un partido en el Santiago Bernabéu. Luego, se calcularon las medias de estas intensidades para ambos escenarios y se generó el gráfico de barras que se enseña a continuación:

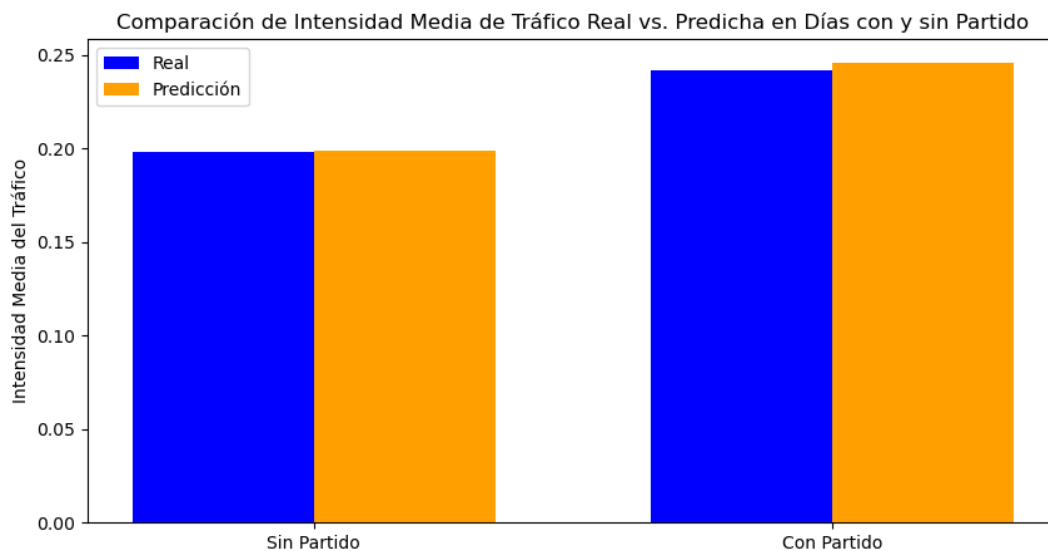


Figura 15: Influencia de los eventos multitudinarios en el tráfico

El gráfico muestra claramente como la intensidad de tráfico varía en los días con y sin partido, demostrando la capacidad del modelo para capturar estos cambios y reflejando su efectividad en la predicción de la intensidad de tráfico bajo diferentes condiciones.

4.2.COMPORTAMIENTO DEL TRÁFICO SIMULADO

En esta sección se analiza el comportamiento del tráfico simulado obtenido a partir de una predicción realizada por el modelo de predicción entrenado. Una vez el usuario rellena los campos de entrada para establecer las condiciones bajo las cuales quiere que se realice la simulación, el modelo se encarga de predecir la cantidad de tráfico en vehículos por hora que habrá para el instante seleccionado. Esta predicción hace referencia a la cantidad de vehículos que se espera que circulen por la calle José Ortega y Gasset en ese momento.

La simulación tiene una duración de 3600 segundos (una hora) y se recolectan distintas métricas en intervalos de un segundo. El simulador SUMO proporciona un archivo de resultados llamado *summary.xml*, el cual contiene información detallada sobre diversos parámetros de tráfico en cada instante de tiempo, a continuación, se muestra una tabla que describe los valores resultado de una simulación[42]:

Nombre	Tipo	Descripción
tiempo	Segundos	Segundo donde son tomados los resultados
Cargados	Entero	Número de vehículos cargados desde archivos de entrada hasta este instante de tiempo
Insertados	Entero	Número de vehículos insertados hasta este instante de tiempo
En movimiento	Entero	Número de vehículos que estaban en movimiento en ese instante de la simulación
Esperando	Entero	Número de vehículos que esperaban ser insertados (no pudieron ser insertados) en este instante de tiempo
Finalizados	Entero	Número de vehículos que han llegado a su destino o que fueron eliminados de otra manera de la simulación
llegado	Entero	Número de vehículos que han llegado a su destino hasta este instante de tiempo
detenidos	Entero	Número de vehículos en la red con velocidad inferior a 0.1m/s (que no están esperando en una parada) en este instante de tiempo
Colisiones	Entero	Número de vehículos que estuvieron involucrados en una colisión hasta este instante de tiempo
Teletransportados	Entero	Número de vehículos en la red que fueron teletransportados hasta este instante de tiempo
Tiempo medio de espera	Segundos	Tiempo medio que todos los vehiculos han tenido que esperar para ser insertados hasta este instante de tiempo
Tiempo medio de viaje	Segundos	Tiempo medio de viaje de todos los vehículos que han salido de la simulación en este instante de tiempo
Velocidad media	m/s	Velocidad media de todos los vehículos en la red en este instante de tiempo
Velocidad media relativa		Velocidad media de todos los vehiculos en la red en relación con el límite de velocidad en este instante de tiempo
duración	Milisegundos	Tiempo de cálculo para estos resultados durante este paso de simulación en milisegundos

Tabla 5: Resultados de una simulación

Todos los parámetros descritos en la tabla anterior se calculan y registran para cada instante de tiempo en la simulación. Esto permite un análisis detallado y en tiempo real del comportamiento del tráfico simulado, proporcionando una visión precisa de como varían estas métricas a lo largo del tiempo. Cada valor representa el estado de un aspecto específico del tráfico en un momento particular, facilitando así

la identificación de patrones y tendencias en el flujo vehicular durante la duración completa de la simulación.

SUMO ofrece una amplia gama de salidas que pueden resultar difíciles de interpretar y visualizar. Para facilitar este proceso, el simulador proporciona varias herramientas que permiten visualizar los resultados de una simulación de manera adecuada, estas herramientas generalmente se basan en la utilización de la biblioteca *matplotlib*, para más información sobre todas las opciones de visualización existentes se puede consultar el manual de visualización de SUMO [43].

A continuación, se presenta un ejemplo de un gráfico que muestra la cantidad de vehículos que se encuentran en movimiento a lo largo de una simulación aleatoria realizada, mediante la observación de este gráfico, se puede entender como varía el flujo de tráfico durante la simulación, permitiendo identificar patrones y posibles cuellos de botella en la red de tráfico:

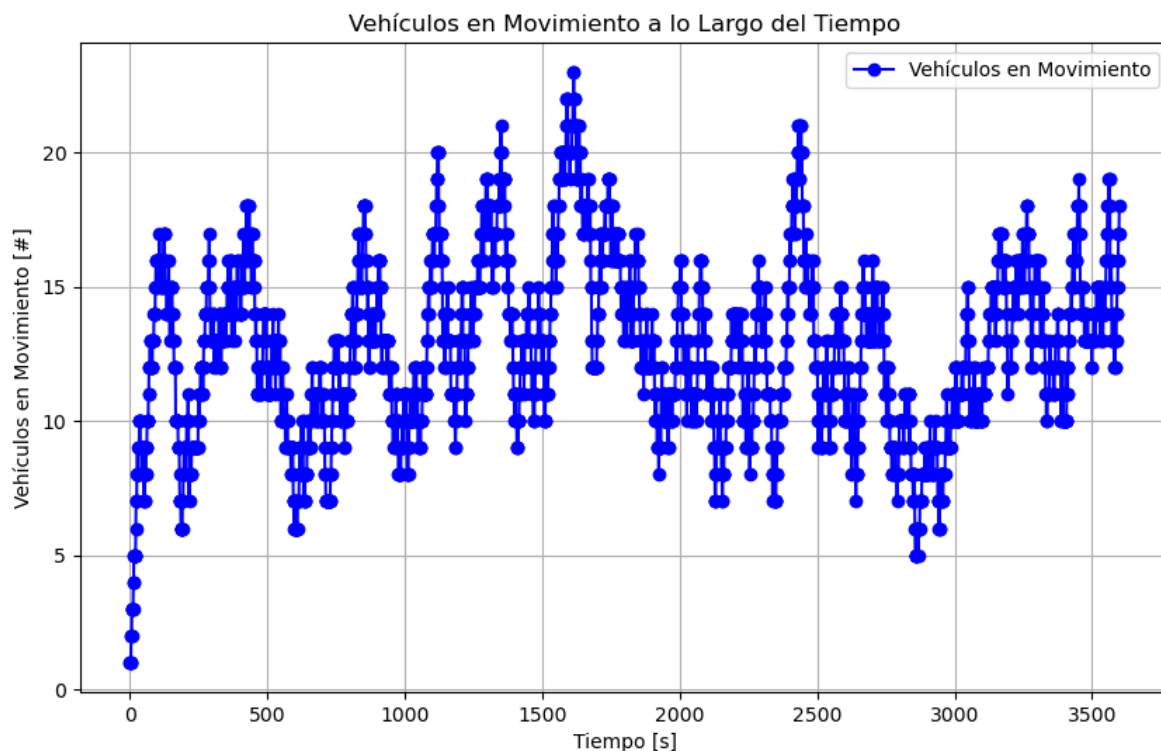


Figura 16: Comportamiento de tráfico simulado

En conclusión, estas visualizaciones permiten analizar de manera detallada el comportamiento del tráfico simulado bajo diferentes condiciones, proporcionando una visión clara y comprensible de los datos generados. Estas técnicas facilitan la evaluación del desempeño de diferentes estrategias de gestión de tráfico.

Las aplicaciones en lo que se refiere a la visualización de los datos generados por la simulación son inmensas, mediante el uso de herramientas de visualización es posible mostrar los tiempos de espera a lo largo de la simulación, pudiendo identificar bajo que circunstancias de tráfico este se maximiza. Por ejemplo, se pudiera observar que en ciertas intersecciones experimentan tiempos de espera más altos durante las horas pico, lo que indicaría la necesidad de ajustar los tiempos de los semáforos o rediseñar la infraestructura de las intersecciones.

En general, la capacidad de generar diversos tipos de gráficos y visualizaciones a partir de los datos de la simulación de tráfico es una herramienta poderosa para analizar y comprender el comportamiento del tráfico. Los resultados de las simulaciones proporcionan una base sólida para la toma de decisiones informadas por parte de los planificadores urbanos y la implementación de medidas efectivas para mejorar la eficiencia y la seguridad del sistema de transporte.

5. CONCLUSIONES Y LÍNEAS FUTURAS

5.1. CONCLUSIONES

Este trabajo se centró en el desarrollo e integración de un modelo de predicción de tráfico en el simulador SUMO. En el proyecto se abordó la problemática que la congestión vehicular tiene en entornos urbanos, y se utilizaron técnicas de *Machine Learning* para ello. La implementación del módulo de predicción ha demostrado ser efectiva para predecir la intensidad de tráfico en la calle José Ortega Y Gasset en la ciudad de Madrid, proporcionando resultados precisos y coherentes con la situación de tráfico actual.

El uso de datos históricos de tráfico junto con la incorporación de variables significativas, han sido cruciales para lograr el diseño de un dataset completo que pudiera representar de una manera precisa las condiciones de tráfico de esta calle. El modelo entrenado con dicho dataset es capaz de captar las variaciones temporales y espaciales de tráfico, validando así sus características. Las distintas transformaciones a las variables temporales realizadas fueron fundamentales para el desarrollo del modelo. La correcta predicción de las horas pico y patrones del tráfico bajo distintas circunstancias confirman la capacidad del modelo de obtener resultados coherentes en lo que se refiere al manejo de escenarios complejos y dinámicos.

La integración con el simulador SUMO ha permitido simular distintos escenarios de tráfico, todos ellos basados en una predicción realizada por el modelo entrenado, integrándose en un servidor de fácil uso para los usuarios. La intención de este trabajo siempre fue ofrecer una herramienta, o por lo menos la base de una herramienta útil para la planificación y toma de decisiones en temas de gestión urbana, lo único que se necesita para utilizar la aplicación es tener instalado el simulador localmente en el ordenador y cada usuario podrá realizar las predicciones que quiera.

Si bien este trabajo se enfocó solamente en la predicción del tráfico en una calle de la ciudad de Madrid, la metodología desarrollada es completamente extrapolable y escalable para crear conjuntos de datos que abarquen las características de calles adicionales. Cuando hablamos de que es extrapolable, quiere decir que la metodología aplicada para predecir el tráfico de la calle José Ortega y Gasset puede ser utilizada para otras calles de la ciudad. Al aplicar este enfoque a diversas vías, se podrían realizar simulaciones paralelas, proporcionando una herramienta más amplia para predecir el tráfico de toda la ciudad de Madrid. Un enfoque de este tipo propondría una posible solución a la reducción de la congestión vehicular, así como de la contaminación ambiental.

El proyecto destaca la importancia de combinar técnicas de *Machine Learning* con herramientas de simulación, la ventaja de la propuesta realizada en este trabajo es la inicialización de las variables de entorno del simulador, es decir, las condiciones bajo las cuales se inicia una simulación.

En definitiva, el sistema desarrollado busca mejorar las situaciones de congestión de tráfico, por un lado, se enfoca en la detección y análisis de patrones de tráfico a partir de datos públicos. Por otro lado, una vez identificados estos patrones, el modelo desarrollado es capaz de integrar múltiples variables del entorno para ofrecer predicciones lo más precisas posibles, todo ello integrado en una interfaz de fácil manejo para los planificadores e investigadores urbanos. De esta manera, se logra predecir el tráfico en una calle de Madrid.

5.2.LÍNEAS FUTURAS

Debido a la importancia de la meteorología en el comportamiento del tráfico, se espera incluir en el *dataset* nuevas variables que puedan de alguna manera contribuir a una predicción mucho más precisa, dándole de esta manera al modelo otra variable de la que pudiera depender la intensidad de tráfico, así como la inclusión de nuevas variables como podrían ser las obras de construcción. El enriquecimiento del conjunto de datos con estas nuevas variables permitiría aumentar la precisión de las predicciones dándole al modelo nuevas variables para identificar nuevos patrones de tráfico, mejorando así la capacidad predictiva del modelo.

Por otro lado, se espera poder crear nuevos conjuntos de datos más completos que incluyan varias calles y distritos, extendiendo el alcance del sistema a toda la ciudad de Madrid. La implementación de distintas simulaciones de manera conjunta permitiría la realización de una simulación más grande, con esto logrado se podrían controlar aspectos como la coordinación de semáforos, la gestión de eventos, el control de emergencias, los cortes por obras, entre otras muchas aplicaciones referentes a la gestión urbana.

6. BIBLIOGRAFÍA

- [1] F. P. Appio, M. Lima, and S. Paroutis, “Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges,” *Technol Forecast Soc Change*, vol. 142, pp. 1–14, May 2019, doi: 10.1016/j.techfore.2018.12.018.
- [2] “Una Población en Crecimiento.” Accessed: Jun. 01, 2024. [Online]. Available: <https://www.un.org/es/global-issues/population>
- [3] D. Mage *et al.*, “Urban Air pollution in Megacities of the World,” 1996.
- [4] N. Brusselaers, C. Macharis, and K. Mommens, “The health impact of freight transport-related air pollution on vulnerable population groups,” *Environmental Pollution*, vol. 329, Jul. 2023, doi: 10.1016/j.envpol.2023.121555.
- [5] A. M. Nagy and V. Simon, “Survey on traffic prediction in smart cities,” *Pervasive and Mobile Computing*, vol. 50. Elsevier B.V., pp. 148–163, Oct. 01, 2018. doi: 10.1016/j.pmcj.2018.07.004.
- [6] F. H. Alshamrani, H. F. Syed, and M. A. Elhusein, “Machine Learning Based Model for Traffic Prediction in Smart Cities,” 2019.
- [7] E. Córtes Luque, “Santander: Una ciudad con cerebro,” <https://blogthinkbig.com/peoplefirst/santander-smart-city>.
- [8] A. Caragliu and C. F. Del Bo, “Smart innovative cities: The impact of Smart City policies on urban innovation,” *Technol Forecast Soc Change*, vol. 142, pp. 373–383, May 2019, doi: 10.1016/j.techfore.2018.07.022.
- [9] O. Altintas and Institute of Electrical and Electronics Engineers, *Luxembourg SUMO Traffic (LuST) Scenario: 24 Hours of Mobility for Vehicular Networking Research*.
- [10] Y. Su, H. Ghaderi, and H. Dia, “The role of traffic simulation in shaping effective and sustainable innovative urban delivery interventions,” *EURO Journal on Transportation and Logistics*, vol. 13. Elsevier B.V., Jan. 01, 2024. doi: 10.1016/j.ejtl.2024.100130.
- [11] F. Trinidad and C. Talay, “Consideraciones metodológicas para la investigación con simuladores de red,” *Informes Científicos Técnicos - UNPA*, vol. 11, no. 3, pp. 211–235, Nov. 2019, doi: 10.22305/ict-unpa.v11.n3.803.
- [12] Rocío Murcia Hernández, Joan García Haro, and Esteban Egea López, “Evaluación de herramientas de simulación de Redes Vehiculares.”
- [13] M. Saidallah, A. El Fergougui, and A. E. Elalaoui, “A Comparative Study of Urban Road Traffic Simulators”, doi: 10.1051/0.
- [14] “AIMSUN.” [Online]. Available: <https://www.aimsun.com/es/>
- [15] J.-M. Auberlet, “Behavioral Road Traffic Simulation with ARCHISM,” 2001. [Online]. Available: <https://www.researchgate.net/publication/245583342>
- [16] “TSIS-CORSIM 2024.” [Online]. Available: <https://mctrans.ce.ufl.edu/tsis-corsim/>
- [17] “MatSim.” [Online]. Available: <https://www.matsim.org/>
- [18] “SUMO”, [Online]. Available: <https://sumo.dlr.de/docs/index.html>
- [19] “El machine learning al rescate del tráfico denso urbano”, [Online]. Available: <https://www.theagilityeffect.com/es/article/como-el-machine-learning-puede-aumentar-la-fluidez-en-una-zona-urbana-de-trafico-denso/>
- [20] Z. Ullah, F. Al-Turjman, L. Mostarda, and R. Gagliardi, “Applications of Artificial Intelligence and Machine learning in smart cities,” *Computer Communications*, vol. 154. Elsevier B.V., pp. 313–323, Mar. 15, 2020. doi: 10.1016/j.comcom.2020.02.069.

- [21] J. M. Cárdenas, “EL MACHINE LEARNING A TRAVÉS DE LOS TIEMPOS, Y LOS APORTES A LA HUMANIDAD DENNIYE HINESTROZA RAMÍREZ.”
- [22] N. Alqudah and Q. Yaseen, “Machine Learning for Traffic Analysis: A Review,” in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 911–916. doi: 10.1016/j.procs.2020.03.111.
- [23] Y. Liu and H. Wu, “Prediction of road traffic congestion based on random forest,” in *Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017*, Institute of Electrical and Electronics Engineers Inc., Feb. 2018, pp. 361–364. doi: 10.1109/ISCID.2017.216.
- [24] A. Joyce and V. Javidroozi, “Smart city development: Data sharing vs. data protection legislations,” *Cities*, vol. 148, May 2024, doi: 10.1016/j.cities.2024.104859.
- [25] “Impulso a la promoción de los datos abiertos a través del Proyecto de Ley de Movilidad Sostenible.” Accessed: May 21, 2024. [Online]. Available: <https://datos.gob.es/es/noticia/impulso-la-promocion-de-los-datos-abiertos-traves-del-proyecto-de-ley-de-movilidad>
- [26] “Ley de Movilidad Sostenible,” 2024, Accessed: May 21, 2024. [Online]. Available: <https://www.transportes.gob.es/el-ministerio/campanas-de-publicidad/ley-de-movilidad-sostenible-y-financiacion-del-transporte>
- [27] S. E. Bibri, J. Huang, and J. Krogstie, “Artificial Intelligence of Things for Synergizing Smarter Eco-City Brain, Metabolism, and Platform: Pioneering Data-Driven Environmental Governance,” *Sustain Cities Soc*, p. 105516, May 2024, doi: 10.1016/j.scs.2024.105516.
- [28] D. Gao, J. Wu, and L. Niu, “A Method for Comprehensive Ability Assessment of Smart City Construction from the Perspective of Big Data,” in *Proceedings - 2021 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 51–54. doi: 10.1109/ICITBS53129.2021.00021.
- [29] “Sistemas Inteligentes de Transportes (ITS).” [Online]. Available: <https://www.dgt.es/muevetecon-seguridad/tecnologia-e-innovacion-en-carretera/sistemas-inteligentes-de-transporte-its/>
- [30] “La importancia del uso de datos en la movilidad.” [Online]. Available: <https://esmovilidad.transportes.gob.es/noticias/la-importancia-del-uso-de-datos-en-la-movilidad>
- [31] “DESCRIPCIÓN DEL CONJUNTO DE DATOS DE TRÁFICO: UBICACIÓN DE LOS PUNTOS DE MEDIDA DEL TRÁFICO.” Accessed: May 23, 2024. [Online]. Available: <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=ee941ce6ba6d3410VgnVCM1000000b205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>
- [32] J. Hao and T. K. Ho, “Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language,” *Journal of Educational and Behavioral Statistics*, vol. 44, no. 3. SAGE Publications Inc., pp. 348–361, Jun. 01, 2019. doi: 10.3102/1076998619832248.
- [33] N. Rodríguez-Martín, “La aparición del «problema de la circulación» y los orígenes de la gestión del tráfico urbano en España, 1900-1936,” *Historia Contemporanea*, no. 55. Universidad del País Vasco, pp. 483–516, 2017. doi: 10.1387/hc.17824.
- [34] “Cortes de tráfico los fines de semana y festivos.” [Online]. Available: <https://www.madrid.es/portales/munimadrid/es/Inicio/Movilidad-y-transportes/Incidencias-de-Trafico/Principales-incidencias-en-la-ciudad/Cortes-de-trafico-los-fines-de-semana-y-festivos/>
- [35] S. Florencia Torrego, “Tráfico Urbano en Madrid,” *Geographica; Madrid*, 1974.
- [36] Comunidad de Madrid, “Estudio del tráfico,” Madrid. [Online]. Available: <https://www.comunidad.madrid/sites/default/files/doc/medio-ambiente/rd6352-anexovi.pdf>
- [37] E. L. Lehmann, *Theory of Point Estimation (2nd edición)*. 1998.

- [38] E. M. Rodríguez, “Errores frecuentes en la interpretación del coeficiente de determinación lineal.”
- [39] IEEE ITSS, Institute of Transportation Engineers, and Institute of Electrical and Electronics Engineers, *2018 IEEE Intelligent Transportation Systems Conference : November 4-7, Maui, Hawaii*.
- [40] A. Botchkarev, “Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology.”
- [41] “Puntuación de modelos de Regresión.” Accessed: Jun. 01, 2024. [Online]. Available: https://help.qlik.com/es-ES/cloud-services/Subsystems/Hub/Content/Sense_Hub/AutoML/scoring-regression.htm
- [42] “SUMO Summary.” [Online]. Available: <https://sumo.dlr.de/docs/Simulation/Output/Summary.html>
- [43] “SUMO visualization.” [Online]. Available: <https://sumo.dlr.de/docs/Tools/Visualization.html>
- [44] “Salario medio para Ingeniero Telecomunicaciones en España, 2024.” [Online]. Available: <https://es.talent.com/salary?job=ingeniero+telecomunicaciones>

ANEXO A: ASPECTOS ÉTICOS, ECONÓMICOS, SOCIALES Y AMBIENTALES

A.1 INTRODUCCIÓN

El objetivo principal de este proyecto es desarrollar un modelo de predicción de tráfico utilizando técnicas de *Machine Learning* y datos de tráfico de la ciudad de Madrid, integrándolo posteriormente en la herramienta de simulación de tráfico SUMO. Este trabajo busca abordar la creciente problemática de la congestión vehicular en áreas urbanas, mejorando así la planificación y gestión del tráfico.

A.2 DESCRIPCIÓN DE IMPACTOS RELEVANTES RELACIONADOS CON EL PROYECTO

Durante la fase de selección y descripción de impactos, se han identificado varios aspectos clave relacionados con la sostenibilidad social, económica y ambiental. Los impactos sociales incluyen la mejora de la calidad de vida de los ciudadanos al reducir los tiempos de desplazamiento y la contaminación acústica. Económicamente, la optimización del tráfico puede resultar en ahorros significativos en combustible y en una disminución del tiempo perdido en atascos, lo cual tiene un impacto positivo en la productividad económica. Ambientalmente, la reducción de la congestión vehicular disminuye las emisiones de gases de efecto invernadero y otros contaminantes, contribuyendo así a una mejor calidad del aire. Los principales grupos de interés identificados incluyen a los ciudadanos, autoridades de transporte, urbanistas, y empresas del sector de movilidad.

A.3 ANÁLISIS DEL IMPACTO AMBIENTAL

La congestión vehicular es una fuente significativa de contaminación del aire en las ciudades. Al implementar un modelo de predicción de tráfico eficiente, es posible anticipar y mitigar los puntos críticos de congestión, reduciendo así las emisiones de dióxido de carbono y otros contaminantes. Este proyecto, que pretende aportar a la mejorara del problema de la gestión del tráfico, puede contribuir a una disminución sustancial en los niveles de contaminación, lo cual tiene efectos directos en la salud pública.

Facilitar una movilidad más eficiente mediante la reducción de los tiempos de viaje, hace que la gente tenga cada vez una mayor tendencia a utilizar el transporte público, las bicicletas compartidas, y cualquier tipo de medio de transporte que ayude a tener una movilidad más sostenible en la sociedad. Observamos que hay una tendencia en el mundo actual a reducir los vehículos privados para aumentar los vehículos públicos, esto lo vemos ya en las grandes ciudades, donde el acceso al centro de estas, es decir, la zona más densa en cuanto a densidad de población se refiere, está total o parcialmente regulado.

Al reducir la congestión de tráfico evitando cuellos de botella, modificando los tiempos de los semáforos y promoviendo una movilidad pública, se está aportando directamente al consumo de menos combustible, lo que deriva en menos gases liberados a la atmósfera y en una mejora importante en lo que se refiere al impacto ambiental.

A.4 CONCLUSIONES

Las nuevas tecnologías son un gran paso en lo que se refiere a la implementación de soluciones innovadoras para afrontar los problemas de la sociedad actual, especialmente con los crecientes desafíos que presentan las ciudades. Los datos abiertos de acceso público permiten que se realicen avances para tratar problemas como la congestión vehicular con una velocidad nunca antes vista, ya que mientras más personas puedan acceder a estos datos, más soluciones podrán ser propuestas. Estudiar y analizar los patrones de tráfico mediante el uso de estas nuevas tecnologías genera patrones de pensamiento en la sociedad científica que permiten el diseño de soluciones avanzadas con el objetivo de aportar a una movilidad más sostenible.

En este trabajo, se logró predecir de una manera precisa la intensidad de tráfico en una calle de Madrid, una calle que se ve directamente afectada por la cantidad de eventos multitudinales que se desarrollan en el estadio Santiago Bernabéu, y con el creciente flujo de eventos que tendremos en esta zona para los próximos años, el tema de la congestión vehicular en los alrededores del estadio cobrará mayor importancia, habrá que modificar las políticas de movilidad de la zona y habrá que rediseñar parte de la infraestructura. En este trabajo se muestra como las herramientas de ML y los datos abiertos pueden aportar mucho a estas controversias que se presentan. La zona del Bernabéu será como otro centro de la ciudad, y habrá que regular la circulación por esa zona, así como también habrá que aplicar políticas estrictas de movilidad, todo esto para disminuir en la medida de lo posible la congestión de tráfico que hay en esa zona.

ANEXO B: PRESUPUESTO ECONÓMICO

En este anexo, se presenta la estimación del presupuesto económico necesario para la elaboración del TFG. Esto se realiza teniendo en cuenta la equivalencia entre créditos ECTS y horas de actividad (1 ECTS = 25 horas) y sabiendo que el TFG tiene un valor de 12 créditos, el número total de horas requerido para realizar el desarrollo de este trabajo ha sido de 300 horas. El salario de un ingeniero de teleco ronda los 15 euros la hora [44].

El ordenador utilizado para realizar este trabajo ha sido un Macbook Air de apple con chip M1, con CPU de 8 núcleos, GPU de 7 núcleos y Neutral Engine de 16 núcleos

COSTE DE MANO DE OBRA (coste directo)

Horas	Precio/hora	Total
300	15 €	4.500 €

COSTE DE RECURSOS MATERIALES (coste directo)

	Precio de compra	Uso en meses	Amortización (en años)	Total
Ordenador personal (Software incluido)	1.300,00 €	6	4	162,50 €

COSTE TOTAL DE RECURSOS MATERIALES				162,50 €
---	--	--	--	-----------------

GASTOS GENERALES (costes indirectos)	15%	sobre CD	699,38 €
BENEFICIO INDUSTRIAL	6%	sobre CD+CI	321,721 €

SUBTOTAL PRESUPUESTO			5.683,59 €
IVA APLICABLE	21%		1.193,55 €
TOTAL PRESUPUESTO			6.877,14 €

Tabla 6: Presupuesto Económico

