

Digital Twin para la Gestión de Transporte Público: Análisis del Flujo de Articulados en TransMilenio por la Troncal Calle 26

Mónica-Juliana Pérez¹

¹ Pontificia Universidad Javeriana, Bogotá, Colombia
perezmonica@javeriana.edu.co

Resumen.

El presente documento detalla la construcción de un modelo de Digital Twin (DT) del sistema de transporte TransMilenio en Bogotá, enfocado específicamente en la Troncal de la Calle 26. Se aborda desde la adquisición y preparación de datos hasta la implementación y evaluación del modelo. Los datos se obtienen de fuentes GTFS y se manipulan para reflejar con precisión las estaciones y rutas de la troncal mencionada. Se realizan pruebas de estacionariedad y se emplea el análisis de series de tiempo para predecir patrones de tráfico, utilizando modelos ARIMA para estimar la congestión y flujo de pasajeros.

La implementación del DT se realiza en NetLogo, un entorno de modelado multiagente que facilita la simulación de sistemas complejos. En este entorno, se simulan las dinámicas de los autobuses y estaciones, utilizando un MAS (Sistema Multi-agente) que sigue un enfoque basado en la metodología GAIA. Este enfoque permite la coordinación de entidades decisionales locales (autobuses y estaciones) y globales (coordinadores del sistema), así como la integración de predicciones para adaptar la operación simulada a las condiciones cambiantes.

El modelo DT se evalúa demostrando su capacidad para replicar con precisión el comportamiento del sistema de transporte real, especialmente durante las horas pico. Los resultados muestran que el DT es capaz de reflejar de manera efectiva la dinámica operacional del TransMilenio, validando el enfoque de simulación y la integración de datos analíticos en la toma de decisiones y planificación estratégica

1 Introducción

TransMilenio, es el sistema de autobuses de tránsito rápido (BRT) de Bogotá, es un modelo pionero de movilidad urbana que ha transformado significativamente el panorama del transporte público no sólo en Colombia, sino que ha sentado un precedente para ciudades de todo el mundo. Su implantación marca un cambio fundamental a la hora de abordar los complejos retos del transporte urbano en zonas densamente pobladas. Inaugurado en el año 2000, TransMilenio se concibió como respuesta a la creciente congestión del tráfico y contaminación de Bogotá. Este sistema se diseñó para ofrecer una alternativa de transporte por la ciudad más rápida, eficiente

los servicios tradicionales de autobús. Su desarrollo se basó en la visión de mejorar la calidad de vida de los ciudadanos de Bogotá proporcionándoles un sistema de transporte público fiable y accesible.

TransMilenio tiene como metodología de funcionamiento el empleo de carriles exclusivos para autobuses y la adopción de un sistema de tarjetas inteligentes de prepago. Este sistema en comparación a lo que se tenía en la ciudad durante el siglo XX permite un sistema integrado de transporte. Además, Transmilenio ha implementado políticas como el uso de autobuses de gas natural comprimido (GNC) con el fin reducir la contaminación atmosférica y a mejorar la calidad del aire en la ciudad. Este sistema se concibió como una alternativa no sólo para agilizar los tiempos de desplazamiento, sino que también ha proporcionar una alternativa viable al uso del vehículo privado, contribuyendo así a reducir la congestión del tráfico.

A lo largo de los años, TransMilenio ha ampliado su red, intentando adaptarse a las crecientes necesidades de la población de Bogotá. Mediante esta política interna ha intentado llegar a todos los segmentos socioeconómicos de la población, incluidos los desfavorecidos, fomentando así una mayor inclusión social. TransMilenio ha desempeñado un papel crucial en el tejido social y económico de Bogotá. Según la estadísticas de Transmilenio para el mes de agosto del 2023, la distribución de usuarios refleja la extensa cobertura del sistema: 51.929.703 en el Componente Zonal, 46.589.710 en Troncal-Bogotá, 2.914.838 en Troncal-Soacha, 1.565.545 en Dual y 674.744 en el sistema de Cable.

Sin embargo, esta expansión no ha estado exenta de desafíos. Cuestiones como el hacinamiento, la seguridad y el mantenimiento han sido puntos conflictivos, lo que ha hecho necesarios esfuerzos continuos para mejorar y modernizar el sistema. Los esfuerzos para mejorar y modernizar el sistema TransMilenio han incluido la aplicación de medidas para aliviar el hacinamiento, como el aumento de la frecuencia de los autobuses y la adición de más rutas.

Para afrontar estos desafíos dentro sistema de transporte masivo durante los primero 8 meses del año 2023 ha tenido 103.674.540 validaciones en el sistema se deberá implementar soluciones desde las diferentes aristas de la sociedad. Uno de estos puntos de vista sería la inclusión de nuevas tecnologías. La incorporación de nuevas tecnologías en el sistema de transporte masivo puede mejorar enormemente su eficiencia y eficacia [1]. Por ejemplo, el uso de sistemas inteligentes de expedición de billetes y seguimiento en tiempo real puede agilizar el proceso de validación y proporcionar a los viajeros información precisa sobre rutas y horarios [2], [3]. Además, el aprovechamiento de la inteligencia artificial y el análisis de datos puede ayudar a identificar patrones y optimizar las operaciones, lo que mejora el rendimiento general del sistema. Al incorporar sistemas de transporte inteligentes, TransMilenio puede beneficiarse del análisis de datos en tiempo real para mejorar la planificación de rutas y reducir la congestión [4]. Además, las soluciones que implementan nuevas tecnologías de la Industria 4.0 pueden ayudar a minimizar la huella de carbono del sistema y contribuir a un medio ambiente más sustentable [5].

Alguna de las tecnologías que han sido utilizadas dentro de los contextos de la Industria 4.0 y nuevas tecnología dentro de sistemas logísticos y de transporte es el Digital Twin (DT). Los DTs son una representación de entidad física que describe el

estado y el comportamiento para simulación y análisis [6]. La NASA empleó una de las primeras aplicaciones DT para monitorear y simular el comportamiento de los satélites [7]. Hoy en día, el DT se define como una herramienta que soporta la integración de sistemas inteligentes [8]. Para este proyecto, el DT es una réplica virtual de un objeto o sistema físico, creada utilizando datos en tiempo real y tecnologías avanzadas como sensores, IoT e inteligencia artificial. El propósito de un gemelo digital es proporcionar información sobre el rendimiento, mantenimiento y optimización de la entidad física, permitiendo una mejor toma de decisiones y análisis predictivo.

2 Estado del arte

La logística urbana enfrenta numerosos problemas, como los altos tiempos dentro del tráfico, el transporte público eficaz y el crecimiento urbano sostenible. Como herramienta avanzada de modelado y predicción, los DT tiene la capacidad de proporcionar soluciones creativas a estos problemas [9]. Esta tecnología puede integrar datos urbanos a gran escala para producir modelos dinámicos en tiempo real que reproduzcan los sistemas existentes y pronostiquen resultados potenciales, lo que permitirá una planificación y gestión urbana más inteligente y eficaz. Para aprovechar plenamente su potencial, la tecnología de gemelos digitales en la logística urbana debe entenderse en términos de su alcance, impacto y técnicas de implementación. La tecnología de gemelos digitales en la logística urbana tiene el potencial de revolucionar la forma en que las ciudades planifican y gestionan sus sistemas de transporte. Al crear réplicas virtuales de activos y procesos físicos, como rutas de entrega y operaciones de almacén, los planificadores urbanos pueden simular diferentes escenarios y optimizar las operaciones logísticas para lograr la máxima eficiencia. Además, la tecnología de gemelos digitales puede permitir el monitoreo y control en tiempo real de las actividades logísticas, lo que permite una toma de decisiones proactiva y una respuesta rápida a eventos imprevistos. Sin embargo, la cooperación entre varias partes, incluidos el gobierno, los usuarios y las empresas de logística, es necesaria para el uso exitoso de la tecnología de gemelos digitales en la logística urbana. Para crear acuerdos para el intercambio de datos y garantizar que varios sistemas sean interoperables, estas partes interesadas deben cooperar. Además, la aplicación exitosa de la tecnología de gemelos digitales en la logística urbana depende de contar con infraestructura suficiente y conectividad confiable.

2.1 El concepto de gemelo digital (DT)

La idea del "gemelo digital" (DT, por sus siglas en inglés), no es concepto nuevo sino que data con una historia que comienza entre 1999 e inicios del siglo [10]. Como resultado del Internet de las cosas (IoT) y del rápido cambio hacia un ecosistema impulsado por la tecnología ha vuelto a cobrar protagonismo. La transición fluida entre los niveles físico y cibernético es esencial para el desarrollo de este ecosistema y el avance de la próxima ola de aplicaciones tecnológicas.

Un DT es un software digital equivalente a una entidad física de cualquier tipo que es responsable de reproducir recursos y características mientras mejora los comportamientos actuales con otros nuevos [11]. Las entidades físicas pueden ser tanto sistemas complejos, como sistemas de producción o cadenas logísticas, o entidades dentro del sistema, como humanos o máquinas.

Dado que cada DT es identificable individualmente y está estrechamente asociado con su contraparte física, se garantiza que representará a este último lo más cerca posible dentro de los parámetros de su contexto operativo. La representatividad de un dispositivo se puede medir de múltiples maneras: por sus atributos (datos de telemetría y configuraciones, por ejemplo), por sus comportamientos (como acciones que puede realizar el propio dispositivo físico) y por sus relaciones (conexiones entre dos activos), en el mismo espacio lógico, o entre diferentes subcomponentes de un mismo dispositivo, por ejemplo).

2.2 Uso de DT dentro de la Logística Urbana

El término DT se está volviendo cada vez más popular en la industria del transporte, aunque todavía existen dudas sobre su definición y uso, especialmente porque se superpone e interactúa con otras tecnologías relacionadas. Dentro de la logística urbana se pueden encontrar diferentes perspectivas que aportan a la movilidad de las ciudades desde varias perspectivas. El marco conceptual de Alam y Saddik [12] es un ejemplo digno de mención de un sistema ciber-físicos basado en la nube. Presenta una aplicación de asistencia a la conducción basada en telemática que integra computación, control y fusión de sensores y servicios. Con énfasis en la predicción de la intención del conductor y la prevención de la congestión del tráfico, Kumar et al. [13] desarrollaron esta estrategia con un modelo centrado en el gemelo digital que incorpora aprendizaje automático, computación de vanguardia, comunicación 5G y lagos de datos. En otra perspectiva, se observa al tomar DT como un “gemelo de comportamiento” donde se han desarrollado modelos que buscan replicar virtualmente, facilitando la predicción o simulación de los movimientos precisos de los individuos en una ciudad virtual casi en tiempo real [14]. Esta capacidad no sólo ayuda a maximizar la movilidad urbana, sino que también ofrece información invaluable para la creación de sistemas urbanos inteligentes basados en datos.

Al implementar los DT en el campo de la optimización del tráfico se ha visto como estos pueden ser entidades inteligentes para la restricción en tiempos de transporte [15]. Los gemelos digitales, mediante la generación de arquitecturas digitales complejas y el análisis de procesos del sistema, ofrecen una base integral para diseñar escenas de tráfico e identificar los patrones de tráfico más eficientes [16]. Para garantizar que las rutas elegidas sean ideales para los entornos urbanos modernos, este proceso está estrechamente ligado a factores del mundo real, como el estado de las carreteras y la incorporación de tecnologías IoT.

Algunas perspectivas más innovadoras en la aplicación de DT en la logística urbana incluyen la creación de redes aéreas urbanas, donde los modelos de gemelos digitales son esenciales. Facilitan el modelado y la simulación del sistema Urban Air Mobility (UAM), que es esencial para fusionar las redes aéreas con los servicios de transporte

terrestre convencionales [17]. Estos modelos también se pueden utilizar para caracterizar la infraestructura terrestre, que son necesarios para facilitar el acceso a los servicios de pasajeros y carga. La incorporación de gemelos digitales al Transporte Ferroviario Urbano pone de relieve su capacidad para transformar las redes de transporte público .

2.3 Simulación de sistemas multi-agentes

Los sistemas multi-agente (MAS) se basan en la interacción de entidades inteligentes llamados agentes. Estas entidades inteligentes se conceptualizan de manera algo ambigua porque se consideran cosas o entidades separadas que pueden usarse para diferentes propósitos[18]. Long [19], por ejemplo, implementó técnicas basadas en agentes para crear un modelo de toda la cadena de abastecimiento . Este modelo tenía como objetivo la réplica de las etapas de la logística como es el inventario, la producción y el transporte. Por el contrario, Mahunnah [20] realizó un modelo enfocado en el área de la salud que permitiera realizar el seguimiento de un embarazo. También se han utilizado técnicas basadas en agentes para el control de vehículos y la gestión del tráfico. En trabajos anteriores, Chen et al. [21] investigaron la detección y gestión del tráfico basada en agentes en tiempo real siguiendo las directrices de la Fundación para Agentes Físicos Inteligentes (FIPA). Al utilizar agentes compatibles con FIPA y el lenguaje de comunicación de agentes de FIPA (FIPA-ACL), que ofrece sintaxis y semántica exactas para las interacciones de los agentes, fue necesario implementar algoritmos de control para abordar circunstancias imprevistas. Estos ejemplos demuestran la amplitud y creatividad que aporta el uso de MAS dentro de una problemática en particular. El objetivo de la disciplina de la tecnología de agentes es proporcionar ideas y enfoques novedosos que ayuden en la creación y despliegue de sistemas complejos. Los agentes se basan en una estructura flexible pero con varios atributos clave que definen un uso óptimo de la tecnología de agentes en los sistemas. Estos atributos los define Parunak [22] así:

- **Modularidad:** Es el atributo en donde cada agente del sistema tiene su propia colección única de variables de estado, independientemente de su entorno. También debe definirse de forma clara y precisa el punto de interacción entre el agente y su entorno.
- **Descentralización:** Describe como la arquitectura del sistema permite su partición en entidades de software separadas. Sin supervisión u orientación continua de otra entidad de software, cada una de estas entidades puede llevar a cabo actividades importantes por su cuenta.
- **Dinamismo:** Esta característica de la arquitectura sistema que refleja la capacidad para ser resiliente y flexible a cambios rápidos y frecuentes.
- **Naturaleza no estructurada:** Propone que la información del sistema es fácilmente accesible y no está definida durante el diseño del sistema, lo que indica un grado de complejidad e imprevisibilidad en la arquitectura del sistema.

- **Complejidad:** El sistema exhibe una amplia gama de comportamientos variados que podrían interactuar de maneras complejas y sutiles, lo que apunta a una comunicación activa entre agentes.

Por lo anterior, este tipo de computación implica sistemas que comprenden el contexto y responden adecuadamente en función de circunstancias ambientales particulares.

2.4 Integración entre MAS y DT

Los DT proporcionan una contraparte virtual detallada que se asemeja mucho a las condiciones del mundo real y ofrecen una simulación dinámica y en tiempo real de sistemas físicos. La combinación de DT increíblemente precisos y con capacidad de respuesta con simulaciones de MAS abre un mundo de posibilidades. Las simulaciones de múltiples agentes, que son reconocidas por su capacidad para simular interacciones complejas y comportamientos de agentes individuales dentro de un sistema, pueden beneficiarse enormemente del rico contexto basado en datos de los DT.

Como resultado de esta integración, las simulaciones se vuelven más realistas y complejas, lo que mejora la precisión de las predicciones y la calidad de la toma de decisiones. Por ejemplo, Lv [23] incorpora DT para mejorar la seguridad de los sistemas ciberfísicos en el entorno de las ciudades inteligentes. En este modelo se utilizaron agentes y redes para datos dentro de los gemelos digitales, garantizando la seguridad de los datos críticos y manteniendo las ventajas de estos sistemas todo incluido.

Para Clemen [24] los planificadores urbanos y los responsables de la toma de decisiones pueden utilizar gemelos digitales conectados a redes de sensores y simulaciones de sistemas de tráfico en tiempo real. La capacidad de predecir y controlar el flujo de tráfico y los sistemas de tránsito de manera más efectiva es posible gracias a esta integración, que ofrece una imagen extraordinariamente realista y dinámica de los sistemas de transporte urbano.

Otra perspectiva es presentada por Nguyen [25], la idea de un "Swarm metaverse" presenta un uso creativo de esta integración. En este escenario, un mundo virtual construido por gemelos digitales se fusiona perfectamente con el mundo físico de enjambres de robots. Esta combinación reduce en gran medida la complejidad de gestionar enjambres masivos de robots. Al extender esta idea a la logística del transporte público urbano, se puede transformar completamente la forma en que se coordinan diversos agentes (como autobuses, trenes y taxis) dentro de una ciudad, proporcionando un nuevo grado de eficacia y eficiencia en la gestión de la movilidad urbana.

La planificación urbana, la seguridad y la gestión de sistemas complejos son sólo algunos de los ámbitos en los que la integración de la tecnología de DT con MAS puede tener efectos significativos.

3 Objetivos

En esta sección se presentan el objetivo general y específicos que se tuvieron en cuenta para desarrollar el proyecto planteado.

Desarrollar un modelo de Digital Twin de la troncal de la calle 26 en el sistema TransMilenio, utilizando la simulación basadas en agentes y análisis de datos.

- Realizar la extracción y análisis de los datos disponibles para la creación del Digital Twin de la troncal de la calle 26 en el sistema TransMilenio.
- Generar el modelo de predicción para flujo de persona dentro de troncal de la calle 26 en el sistema TransMilenio.
- Realizar el diseño conceptual del sistema de multi-agente en Netlogo para el análisis del flujo de articulados en TransMilenio por la troncal Calle 26.
- Implementar el Digital Twin basado en el sistema multi-agente propuesto.

4 Metodologías

Para realizar este proyecto, teniendo en cuenta los objetivos planteados en el numeral anterior se utilizaron dos enfoques complementarios para construir el DT para la línea troncal Calle 26 de TransMilenio, cada uno destinado a resolver componentes del proyecto diferentes pero igualmente importantes. Con la primera metodología, que se concentró en la recopilación y el análisis de datos, dado esto se escogió la metodología de conoció como CRISP-DM (Cross-Industry Standard Process for Data Mining) un enfoque bien establecido en el campo. de análisis de datos. CRISP-DM permitió manejar de manera efectiva y sistemática los datos del sistema TransMilenio, asegurando que la limpieza, preparación y modelado de los datos se realizaran de una manera que proporcionara información valiosa para la simulación del DT. Esta metodología fue fundamental para transformar los datos útiles y correctos para el DT, contribuyendo directamente al objetivo de desarrollar un modelo preciso y confiable. Mediante esta metodología se realizó la recopilación de datos extensos sobre la condición actual de la infraestructura, los patrones de flujo de pasajeros y el desempeño operativo. Este dio una base sólida sobre la cual construir una representación precisa del sistema.

Por otro lado, utilizamos la metodología GAIA [26], [27], diseñada específicamente para la construcción de MAS. GAIA permite la creación de un modelo de simulación basado en agentes que replica de forma dinámica y realista el comportamiento del sistema de transporte.

En la figura 1 se muestra el enfoque integrado que se plantea para el desarrollo de del DT mediante la combinación de elementos de las metodologías CRISP-DM y GAIA, mostrando un flujo de proceso armonizado. Las etapas CRISP-DM, comienzan con 'Comprensión del Negocio', donde se definen los objetivos del proyecto y los requisitos de datos. A esto le sigue 'Comprensión de los Datos', donde se recopilan datos y se comprenden sus propiedades. La tercera etapa, 'Preparación de los Datos', implica la limpieza y transformación de datos para modelar. A continuación, 'Modelado' se

refiere a la creación de modelos de datos que luego son evaluados en la etapa de 'Evaluación'. Finalmente, 'Despliegue' representa la implementación y la integración de este para el inicio de la siguiente metodología.

Para la metodología GAIA, el proceso comienza con los 'Requerimientos del sistema', que establecen las necesidades fundamentales para la funcionalidad del sistema. La segunda etapa es el 'Marco conceptual de la arquitectura', donde se define la estructura del sistema. Esto conduce a la 'Implementación del Digital Twin', donde el diseño conceptual se actualiza en forma de un Digital Twin funcional. La metodología GAIA se cierra con la 'Evaluación del modelo', donde se evalúa el desempeño del sistema y el cumplimiento de los requisitos.

La imagen ilustra un proceso cíclico e iterativo donde las dos metodologías se cruzan, optimizando tanto los aspectos basados en datos como los basados en agentes del desarrollo del Digital Twin. Cada paso del proceso es esencial: CRISP-DM garantiza que los datos sean precisos y valiosos, mientras que GAIA se centra en el diseño estructural y operativo del sistema multi-agente, culminando en un robusto modelo de gemelo digital. La combinación de estas dos metodologías, CRISP-DM y GAIA, se alinea perfectamente con nuestro objetivo general: desarrollar un modelo de Gemelo Digital robusto y adaptable de la línea troncal Calle 26 de TransMilenio, integrando tanto la complejidad de los datos del sistema real como los basados en agentes.

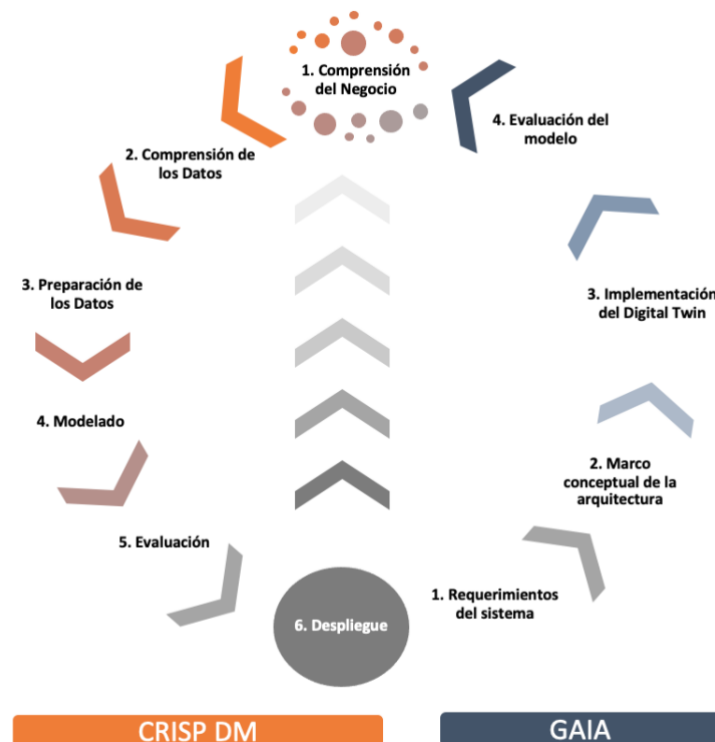


Fig. 1. Representación de los pasos que se llevan a cabo en el proyecto y la integración entre ambas metodologías.

4.1 Metodología CRISP-DM para Análisis de Datos

Para proyectos de análisis y minería de datos, el enfoque del Proceso Estándar Intersectorial para Minería de Datos, o CRISP-DM por sus siglas en inglés, es bien conocido. Ofrece una estrategia metódica para organizar y llevar a cabo un proyecto de minería de datos.

CRISP-DM es aplicable a proyectos de análisis de datos debido a su estructura y adaptabilidad. Se puede utilizar en una variedad de industrias, lo que permite a los equipos abordar metódicamente tareas desafiantes de minería de datos. La técnica fomenta la comprensión del problema empresarial y la profundización en los datos, que luego informan las etapas de preparación y modelado. Esta metodología garantiza que las ideas y resultados generados estén alineados con los objetivos del negocio y puedan integrarse efectivamente en los procesos organizacionales.

Debido a que CRISP-DM es iterativo, cualquier retroalimentación en cualquier momento podría provocar una revisión de etapas anteriores, lo que le permitirá adaptarse a nuevas circunstancias y demandas. Se ha convertido en el estándar de facto en el análisis de datos debido a su amplia aceptación y adaptabilidad, ayudando a los analistas y científicos de datos a proporcionar conocimientos prácticos y dirigir los procedimientos de toma de decisiones.

Etapas 1: Comprensión del negocio.

La fase de "Comprensión del Negocio" para TransMilenio se centra en entender cómo la demanda y la utilización de los servicios de transporte afectan la operación y planificación del Sistema.

TransMilenio S.A., es la entidad gestora del Sistema Integrado de Transporte Público de Bogotá, que fue designada con la tarea de controlar y mejorar la movilidad urbana poniendo énfasis en la calidad, el confort y la dignidad[28] El objetivo de la entidad es mejorar la calidad de vida de los usuarios, centrándose en la gestión eficiente del transporte, fomentando la intermodalidad y garantizando la sostenibilidad financiera y ambiental. Según los planes estratégicos para el año 2023 de la entidad uno de sus enfoques es lograr una cobertura total del sistema de transporte público de Bogotá manteniendo la sostenibilidad operativa, financiera y ambiental a través de políticas públicas y fuentes de financiamiento bien definidas[29].

Las tareas clave de TransMilenio S.A. incluyen la gestión, organización y planificación del servicio de transporte público masivo metropolitano, así como asegurar la inclusión de tecnología avanzada y combustibles de bajas emisiones en sus operaciones. La entidad también se centra en colaborar con otras entidades públicas o privadas para desarrollar actividades relacionadas con el transporte. Además, se compromete a adoptar Tecnologías de la Información y Sistemas Inteligentes de Transporte (ITS), con el fin de mejorar la supervisión y control operativo del Sistema Integrado de Transporte Público (SITP)[30].

En diciembre de 2022 [31], las estadísticas reveladas por TransMilenio muestran que el volumen total de entradas en los portales alcanzó las 493,927, con el Portal de las Américas registrando la mayor proporción de demanda. Tal concentración indica puntos de alta utilización que podrían beneficiarse de mejoras en la infraestructura o servicios.

Además, estaciones clave como San Mateo, Calle 100 y Banderas mostraron una demanda significativa, sumando 292,174 entradas diarias, lo que constituye el 68% de las validaciones totales del sistema[31]. Otros corredores importantes son la extensión con el municipio de Soacha y El Portal de El Dorado que conecta el Aeropuerto Internacional del Dorado.

La recuperación post-pandemia es otro factor crucial para el informe del 2022 de la entidad, especialmente en la extensión Soacha, que ha mostrado una notable resiliencia con un promedio de 120,000 entradas por día hábil y 27,500 durante la hora pico[31]. Estos números no solo reflejan una recuperación significativa en la demanda sino también la capacidad del sistema para adaptarse a cambios drásticos, un indicador positivo para la sostenibilidad a largo plazo del servicio. Esta tendencia se sigue viendo reflejada en las estadísticas del año 2023, donde para agosto se ve la mayor tasa de ingreso al sistema si se compara con los últimos 3 años.

Por último, los trasbordos representan el 18% de las tarifas pagadas, destacando la interconectividad y eficiencia del sistema al facilitar el acceso a múltiples modos de transporte a un costo reducido[31]. Esta característica es un punto clave en la evaluación de la usabilidad y accesibilidad del sistema TransMilenio, sugiriendo que la integración y la conveniencia son factores importantes para los usuarios.

Etapas 2: Compresión de Datos.

La etapa de "Comprensión de los Datos" dentro de la metodología CRISP-DM es un pilar crítico en la construcción del modelo de DT para la troncal de la calle 26 de TransMilenio. Entender los datos en este nivel permite sentar las bases para todas las fases analíticas subsecuentes, garantizando que cada paso, desde la preparación de los datos hasta el despliegue del modelo, esté construido sobre un entendimiento auténtico y detallado del ecosistema de datos de TransMilenio. Esta fase es esencial para asegurar que el modelo de DT sea una réplica virtual fiable y eficaz del sistema físico real, capaz de simular su comportamiento y anticipar su rendimiento bajo diversas condiciones.

Para el modelo que se plantea en el proyecto se tienen dos bases de datos. La primera son GTFS (General Transit Feed Specification) donde se encuentra los datos de posicionamiento y flujo de articulados por el sistema. Estos datos ofrecen una representación estructurada del servicio de transporte, incluyendo horarios, rutas, y paradas. La segunda base de datos utilizada son los datos de ingresos y salidas dentro del sistema. La obtención de esta data fue realizada la plataforma de datos abiertos de TransMilenio correspondiente al mes de agosto de 2023 [32].

La plataforma de datos abiertos, es una iniciativa regulada con la Ley 1712 de 2014 sobre Transparencia y Acceso a la Información Pública Nacional de Colombia. Esta plataforma proporciona un acceso a una amplia gama de datos públicos. Conforme a esta ley, los datos abiertos se definen como aquellos datos primarios, en formatos estándar e interoperables, que facilitan su acceso y reutilización. Estos datos están

disponibles sin restricciones, permitiendo su libre uso y promoviendo la creación de servicios derivados.

Exploración de los datos GTFS estáticos de Transmilenio

Este conjunto de datos está compuesto por varias subset, cada una contribuyendo con información clave para la comprensión y modelización del sistema de transporte urbano de Bogotá. Para este conjunto de datos se constituía de 7 subsets: agency, calendar_dates, calendar, fare_attributes, fare_rules, frequencies, routes, shapes, stop_times, stops, trips.

El subset de datos llamado agency del GTFS ofrece una visión detallada de las agencias de tránsito que operan dentro del sistema TransMilenio. Esta información incluye detalles como nombres, URLs, zonas horarias, idiomas y contactos. El conocimiento de la estructura organizativa y operativa de estas agencias es la que regiría a través de la estructura de datos de los siguientes subsets.

Los subsets calendar_dates y calendar tienen la información de la programación y variabilidad en la operación del sistema. calendar_dates especifica las excepciones a los horarios de servicio regulares, como los días festivos, mientras que calendar detalla la disponibilidad habitual de servicio.

Los subsets fare_attributes y fare_rules aportan una perspectiva económica al modelo, proporcionando información detallada sobre las estructuras tarifarias, incluyendo precios, tipos de moneda, métodos de pago y reglas aplicables.

La subset frequencies revela información sobre la disponibilidad de los servicios a lo largo del día.

Finalmente, las subsets routes, shapes, stop_times, stops y trips proveen una comprensión detallada de la operación física del sistema. Juntas, estos subsets describen el recorrido de las rutas, los horarios en cada parada, la información geográfica de las rutas y las paradas, así como los detalles específicos de cada viaje realizado. La integración de estos componentes es vital para cualquier simulación realista, permitiendo modelar con precisión cómo se mueven los usuarios y los vehículos a través del sistema TransMilenio.

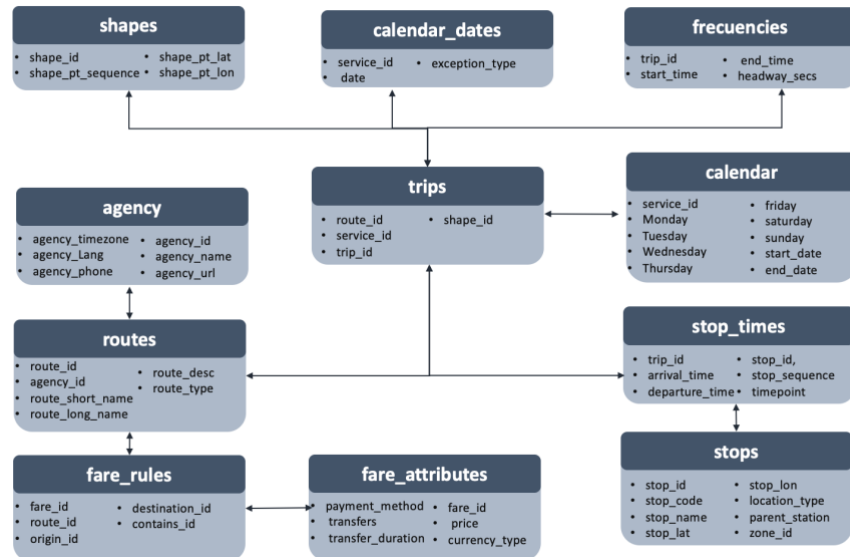


Fig. 2. Modelo relacional de atributos de la base de datos GTFS

En la tabla 1 se encuentra un resumen de la descripción de cada una de esas bases de datos y el análisis del contenido de estas. Por otra parte, en la figura 2 se encuentra el modelo relacional de la base de datos donde se encuentran los atributos de esta.

Componente GTFS	Descripción	Análisis Exploratorio
agency	Contiene información sobre las agencias de tránsito.	5 agencias con nombres, URLs, zonas horarias, idiomas y números de teléfono.
calendar_dates	Define excepciones para los IDs de servicio definidos en el calendario.	140 entradas definiendo excepciones de servicio en una fecha específica.
calendar	Define las fechas en las que el servicio está disponible para una o más rutas.	7 entradas definiendo la disponibilidad de servicio en diferentes días de la semana con fechas de inicio y fin.
fare_attributes	Define información de tarifas para varios aspectos de la agencia de tránsito.	7 atributos de tarifas definiendo precios, tipos de moneda, métodos de pago y transferencias.
fare_rules	Define reglas para aplicar la información de tarifas.	1224 reglas definiendo cómo se aplican las tarifas en rutas, orígenes y destinos.
frecuencias	Define el intervalo de tiempo (frecuencia) para	2 frecuencias definiendo los segundos de intervalo para los viajes.

	rutas con frecuencia variable.	
routes	Define información sobre rutas dentro del sistema de tránsito.	1224 rutas con nombres cortos y largos, descripciones, colores y tipos.
shapes	Define la forma física (trayecto) de una ruta.	Más de 1.4 millones de puntos de forma definiendo la latitud y longitud de los trayectos de las rutas.
stop_times	Define los tiempos en los que un vehículo llega y sale de paradas para cada viaje.	Más de 9.5 millones de tiempos de parada definiendo los tiempos de llegada y salida en las paradas para los viajes.
stops	Contiene información sobre ubicaciones individuales donde los vehículos recogen o dejan pasajeros.	8362 paradas con códigos, nombres, latitudes, longitudes y IDs de zona.
trips	Define viajes para cada ruta, incluyendo carteles de viaje y la secuencia de rutas recorridas.	230805 viajes definiendo rutas, IDs de servicio, IDs de viaje y IDs de forma.

Table 1. Descripción general de los documentos que componen el elemento de GFTS estático que se extrajo de TransMilenio.

Exploración de los datos de entradas y salidas del sistema

Las bases de datos en estudio representa un compendio exhaustivo de la actividad de usuarios dentro del sistema TransMilenio, particularmente enfocado en las entradas y salidas a lo largo de la fase constructiva y operativa de la red de transporte. Es importante resaltar que son dos bases de datos diferenciadas, una para entradas y otra para salidas. A pesar de esto, la estructura de entrega de dicha información el mismo formato y dado esto se van a explicar las dimensiones que tienen ambas y después se diferenciara la composición de estas para evaluar la interacción diaria de los usuarios con el sistema de tránsito. Las bases de datos se componen de archivos de Excel, se compone de 36 atributos. En la tabla 2 se muestra la descripción de estos atributos.

Index	Atributo	Descripción	Tipo de variable
0	Fase de Construcción	Etapas de desarrollo en la que se encuentra la línea específica de TransMilenio.	Categórica
1	Línea de TransMilenio	Identificación de la línea de TransMilenio a la que pertenece la estación.	Categórica
2	Estación	Nombre o código identificativo de la estación en la línea de TransMilenio.	Categórica

3	Acceso de la Estación	Especificación del acceso utilizado dentro de la estación para el registro de entradas.	Catórica
4	Intervalo	Hora de inicio del intervalo en que se registraron las entradas / salidas.	Tiempo
5 a 37	Fechas del Mes (01/08/23 - 31/08/23)	Datos diarios de la cantidad de personas que validaron su tarjeta e ingresaron al sistema para cada día del mes.	Númerico
6	Total General	Sumatoria total de las entradas registradas en el intervalo a lo largo del mes.	Númerico

Table 2. Descripción general de los atributos de las bases de ingresos y salidas del sistema de Transmilenio

La primera dimensión, la 'Fase', categoriza cada entrada o salidas del sistema según la etapa de desarrollo de la línea de TransMilenio correspondiente. Este atributo es de naturaleza catórica y permite segmentar el análisis según las distintas fases evolutivas del sistema, desde su concepción hasta su operación plena.

La siguiente dimensión, 'Línea de TransMilenio', clasifica las entradas o salidas según la línea específica a la que pertenecen las estaciones. La 'Estación' es otro atributo catórico que refiere a la ubicación específica donde los usuarios acceden o salen del sistema, proporcionando un nivel granular de detalle geográfico.

'Acceso de la Estación' se refiere al punto específico de la estación donde se registra la entrada, lo cual permite una evaluación detallada de la distribución de la demanda dentro de una única infraestructura. El 'Intervalo' representa el marco temporal específico en el que se contabilizan las entradas, siendo un atributo temporal que delimita el flujo de usuarios en segmentos manejables para análisis y modelización. Este atributo tiene una duración de 15 minutos durante las 24 horas del día.

Los atributos que comprenden las fechas del mes, del 1 al 31 de agosto para la muestra tomada, son de tipo númerico y se reflejan la cantidad de usuarios que validan su ingreso al sistema o su salida en cada intervalo específico. Estos datos son el núcleo del análisis diario de flujo de pasajeros y son fundamentales para la modelación predictiva y la toma de decisiones basada en patrones de uso.

Por último, el 'Total General' es un atributo númerico que resume el volumen total de entradas durante el mes en estudio para cada intervalo, ofreciendo una visión sintética del tráfico en cada acceso.

Con el fin de realizar una revisión de la data adquirida se realizó una evolución de la data, donde no se encontraron datos nulo o incongruentes. Posterior a ellos se realizó una visualización de esta que para evaluar las decencias según la hora, los días y las estaciones. La figura 3 ilustra la fluctuación diaria en las entradas y salidas, permitiendo identificar patrones de tráfico y anomalías. La comparación directa entre las líneas de 'Entradas' y 'Salidas' sugiere una correlación, aunque con variaciones que podrían indicar diferencias en los patrones de viaje, como la acumulación de usuarios en ciertas estaciones o tiempos de espera prolongados.

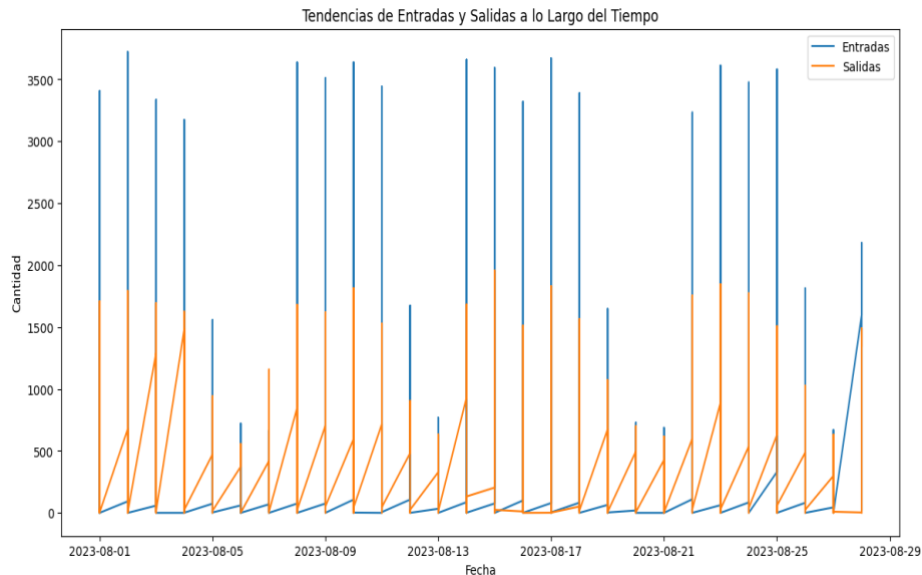


Fig. 3. Gráfico de barras que visualiza las tendencias de entradas y salidas a lo largo del mes de agosto del 2023

En la figura 4 se muestra el análisis de entradas y salidas por día de la semana revela diferencias significativas en el uso del sistema entre días laborables y fines de semana. La distribución de las barras en el gráfico refleja un patrón cíclico semanal y es indicativa de la rutina habitual de los pasajeros. El día menos transitado es el día domingo, mientras que se muestra que hay un número que se mantiene constante de pasajeros a lo largo de la semana desde el martes hasta el día viernes.

Por otra parte, la figura 5 y la figura 6 resume el número total de entradas y salidas por estación de la troncal de la 26 que es el foco de este trabajo. Mediante este gráfico se puede observar una perspectiva de la carga de tráfico que cada una soporta. La variación en la altura de las barras ilustra la distribución desigual de la demanda a través de la red. La estación que más flujo de personas tiene es la de Estación del ElDorado. Si se mezcla esta perspectiva con que esta Estación recibe el flujo tanto del Aeropuerto como de los alimentadores que provienen de la localidad de Fontibón se puede observar que este es un punto de intereses bastante significativo para la ciudad.

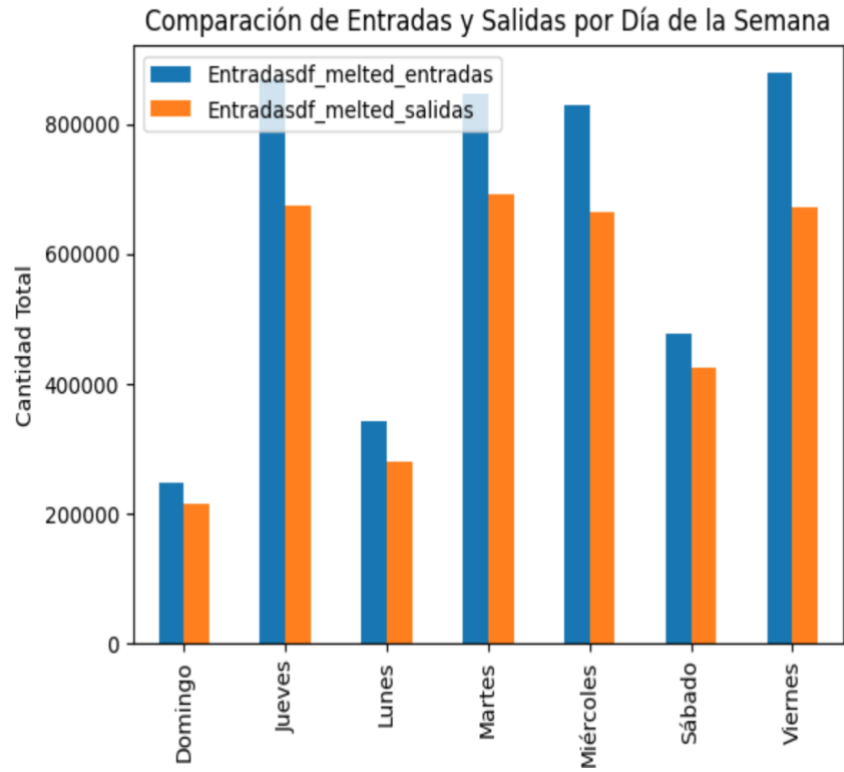


Fig. 4. Gráfico de barras de la comparación entre los días de la semana para entradas y salidas.

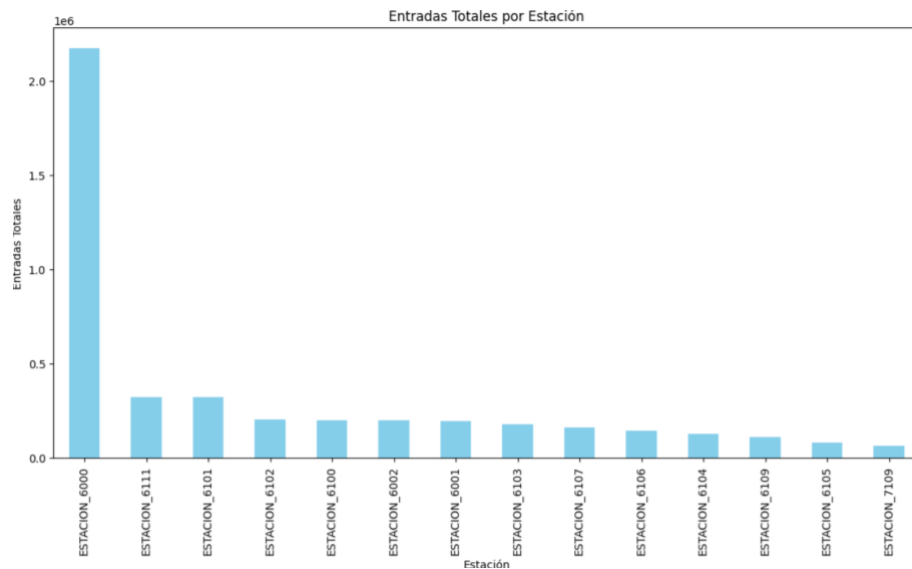


Fig. 5. Validaciones de entradas efectivas al Sistema de TransMilenio de la Troncal de la Calle 26

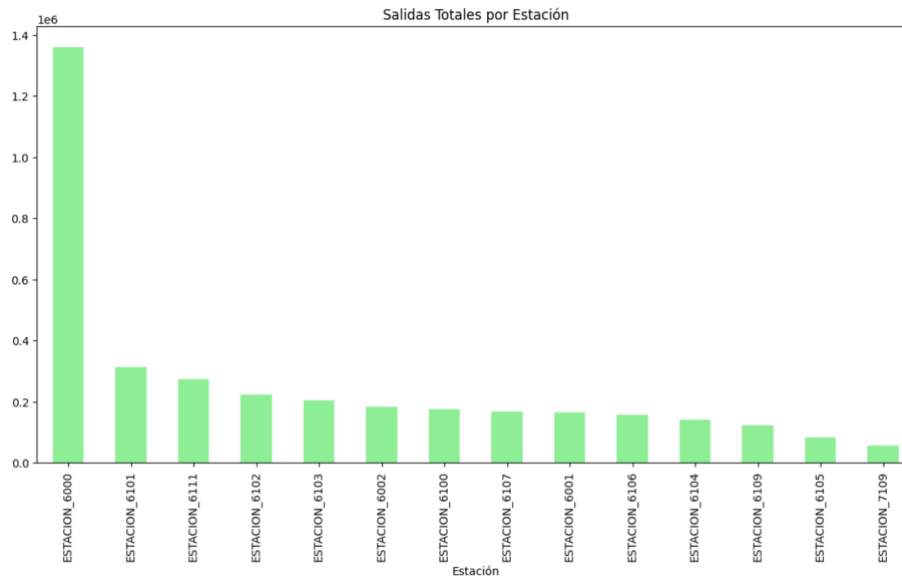


Fig. 6. Salidas al Sistema de TransMilenio de la Troncal de la Calle 26

En concordancia con lo anterior, se observa en la figura 7 una curva prominente, que representa la 'ESTACION_6000', que es la Estación ElDorado. Esta curva sugiere un notable volumen de entradas, destacando significativamente sobre las demás estaciones. Podemos observar que esta estación en particular se ve altamente afectada cuando en días de fines de semana donde se nota un decremento en el flujo de personas bastante significativo si se compara con otros días.

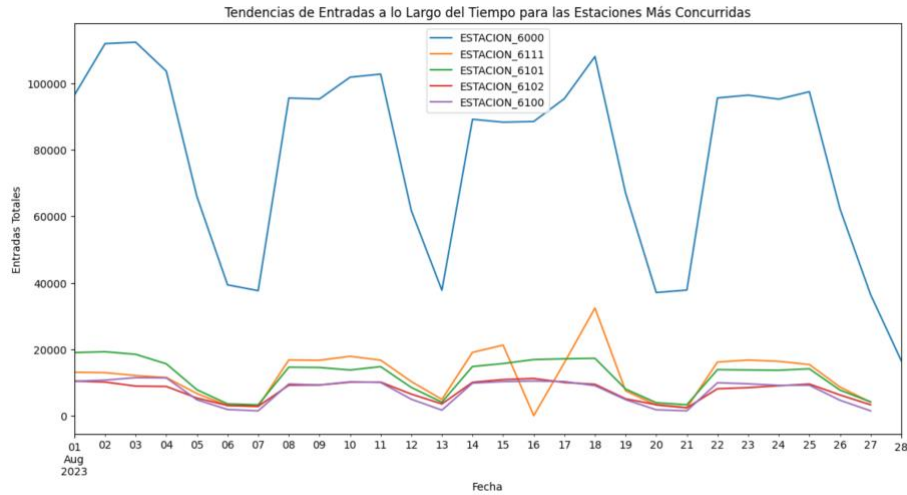


Fig. 7. Tendencias de Entradas a lo Largo del Tiempo para las Estaciones Más Concurridas del Sistema de TransMilenio de la Troncal de la Calle 26

Las demás estaciones, aunque presentan un volumen menor de tráfico en comparación, muestran una consistencia que puede indicar un uso regular y predecible, esencial para el mantenimiento de un flujo constante y la optimización de los recursos del sistema.

Etapa 3: Preparación de los datos

Durante esta etapa, se llevan a cabo procedimientos tales como la limpieza de datos, que incluye la corrección de errores, el manejo de valores faltantes y la eliminación de duplicados. Además, se realiza la estandarización de formatos y la normalización de rangos para homogeneizar los conjuntos de datos que provienen de diferentes fuentes o formatos.

La preparación de datos para la construcción del modelo de DT de la troncal de la 26 se inició con la selección meticulosa de los nombres de las paradas pertinentes. Se compiló una lista detallada de las paradas de interés, que para nuestro caso son las estaciones de Transmilenio de la Troncal calle 26. Dado lo anterior se excluyeron las paradas auxiliares de SITP y de las rutas alimentadoras. Utilizando esta lista, se procedió a filtrar los DataFrames de GTFS para aislar solo aquellos registros relacionados con la troncal de la 26. Se empleó el método `isin` para filtrar `stops`, `stop_times`, `trips`, y `routes` que coinciden con las paradas seleccionadas, creando subconjuntos de datos focalizados y relevantes para el DT. Este enfoque asegura que el análisis posterior se centre en datos pertinentes a la troncal específica, evitando así el ruido y la distorsión que podrían introducir datos irrelevantes.

Además, se realizó un proceso de combinación (`merge`) para enriquecer los `trips` con nombres de rutas, vinculándolos con los códigos de parada correspondientes. Esto

no solo mejoró la legibilidad de los datos sino que también permitió una identificación más intuitiva de las rutas y su secuencia de paradas a lo largo de la troncal.

Por otra parte, para adaptar los datos al contexto geográfico del modelo DT, se incorporó la librería `geopandas` junto con `pyproj` para transformar las coordenadas geográficas a un sistema de coordenadas proyectadas. Esta transformación es crucial para la representación precisa de las paradas en un plano XY, permitiendo su integración en un modelo espacial que refleja con exactitud la geografía de la troncal de la 26.

Con el fin de ajustar la geolocalización de las paradas a un sistema de referencia que se alinea con el marco de trabajo del DT, se escaló la representación espacial a los límites de un plano XY específico. Este ajuste meticuloso asegura que la representación espacial en el modelo sea una traducción fiel de la realidad física del sistema de transporte.

En cuanto a las bases de entradas y salidas del sistema se llevó a cabo inicialmente la estandarización de los nombres de las columnas en la base de datos de salidas, esto se realizó incluyendo una columna llamada 'stop_id' que permite integrarse fácilmente con las convenciones preexistentes y facilitar la manipulación y el análisis. Posteriormente, se procedió a eliminar la columna 'Total general' de ambos conjuntos de datos, entradas y salidas, para centrar la atención en los datos individuales de entradas por intervalo, eliminando así sumatorias que podrían obstruir el análisis granular.

La agrupación de los datos se realizó según 'stop_id' e 'Intervalo', sumando las entradas y las salidas para cada combinación única de identificador de parada e intervalo de tiempo, lo cual fue realizado con la función `groupby`. Este paso fue crucial para condensar los datos y prepararlos para la transformación subsiguiente, donde las columnas de fechas se convierten en una sola columna.

Para corregir el formato de tiempo y normalizar la representación de los intervalos, se convirtió la columna 'Intervalo' a un objeto `datetime.time` utilizando la librería `pandas`. Esta conversión es esencial para permitir comparaciones temporales y la manipulación de series de tiempo dentro del análisis.

Finalmente, se enriquecieron los datos con información temporal adicional. Se extrajo la hora del intervalo y se calculó el día de la semana correspondiente a cada fecha, lo cual se tradujo a un nombre de día legible en el idioma local. Esta adición de datos de tiempo contextualiza aún más el conjunto de datos y abre la puerta a análisis específicos por hora del día y por día de la semana, lo cual es vital para entender los patrones de tránsito y la dinámica de uso del sistema TransMilenio.

Etapas 4: Modelado.

En esta etapa se muestra el proceso del modelado para estimar la congestión dentro de las estaciones de TransMilenio de la Troncal Calle 26. Dado la naturaleza de los datos se plantea el uso del análisis de series de tiempo. El análisis de las series de tiempo representa una herramienta muy útil, especialmente en la predicción de patrones temporales en contextos como el sistema de transporte público. Este enfoque analítico se sustenta en la premisa de que los patrones históricos de datos pueden ser extrapolados para prever tendencias futuras.

Para utilizar cualquier herramienta del análisis y predicción de series de tiempo primero se debe validar la aplicabilidad de esta sobre los datos que se tienen son estacionales. Para revisar la estacionariedad de la serie temporal se implemento la prueba de Dickey-Fuller aumentada (ADF). Esta prueba se fundamenta en un modelo autoregresivo y evalúa la presencia de una raíz unitaria, la cual es indicativa de una tendencia estocástica o un paseo aleatorio en los datos, contrapuesta a la estacionariedad deseada.

$$H_0: y=0 \quad (1)$$

$$H_1: y<0 \quad (2)$$

La hipótesis nula (1) de la prueba ADF postula que la serie de tiempo posee una raíz unitaria y, por lo tanto, no es estacionaria. Por otro lado, la hipótesis alternativa (2) sostiene la ausencia de dicha raíz unitaria, implicando que la serie es estacionaria.

Los resultados de la prueba de Dickey-Fuller aumentada (ADF) aplicada a la serie de tiempo de entradas del sistema TransMilenio demuestran de manera concluyente la estacionariedad de los datos. El estadístico de prueba obtenido, se pueden ver en la tabla 3, está significativamente por debajo de los valores críticos para el 1%, 5% y 10%. Además, el valor p extremadamente pequeño, prácticamente nulo, refuerza el rechazo de la hipótesis nula de una raíz unitaria presente en la serie de tiempo.

Métrica	Valor
Estadístico de Prueba	-17,89
Valor p	2,98E-15
Valor Crítico (1%)	-3,43
Valor Crítico (5%)	-2,86
Valor Crítico (10%)	-2,57

Table 3. Resultados de las prueba Dickey-Fuller aumentada (ADF)

Estos hallazgos sugieren que la serie de tiempo es estacionaria, lo que implica que las propiedades estadísticas de la serie no dependen del tiempo y, por lo tanto, se presta a un modelado predictivo efectivo utilizando el modelo de series de tiempo. ARIMA, que significa Autoregressive Integrated Moving Average, es un modelo estadístico prominente utilizado para analizar y predecir series de tiempo. Este modelo es particularmente útil para datos estacionarios y se ha aplicado en el análisis de flujo de pasajeros de transporte público [33], [34].

Un modelo ARIMA se caracteriza por tres parámetros fundamentales, denotados como ARIMA(p, d, q):

1. p - Autoregresión (AR): Este parámetro representa el componente autoregresivo del modelo. Se refiere al número de retrasos, o lags, de la serie de tiempo que se

utilizarán como predictores. El componente AR modela la correlación entre un punto de observación y un número específico de puntos anteriores. Intuitivamente, puede pensarse como la influencia que los valores pasados tienen sobre el valor actual.

2. d - Integración (I): El parámetro de integración indica el número de diferencias necesarias para convertir la serie de tiempo en una serie estacionaria. Diferenciar es el proceso de restar el valor actual de la serie con su valor anterior. Este proceso se repite 'd' veces hasta que la serie alcanza la estacionariedad. La integración es fundamental para remover la tendencia o la estacionalidad en los datos, permitiendo que el modelo se ajuste a una estructura que refleje las fluctuaciones puras en los datos sin tendencias de largo plazo o patrones cíclicos.
3. q - Media Móvil (MA): El último parámetro, q, está asociado con el componente de media móvil del modelo ARIMA. Este componente utiliza el error de la predicción de un modelo de media móvil aplicado a los retrasos de la serie de tiempo. Esencialmente, se centra en el ruido blanco de los errores de predicciones pasadas y los utiliza para mejorar la precisión de las predicciones futuras.

Dado que ya se comprobó que los datos pasan el supuesto de estacionariedad, se puede usar como parámetro d el número 0. Para el resto de parámetros se van a ser gráficos que ayuden a su ajuste. Los gráficos de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) son herramientas analíticas esenciales en el diagnóstico de la naturaleza y estructura de las series de tiempo, fundamentales para la selección de modelos ARIMA adecuados. La interpretación conjunta de ambos gráficos facilita la identificación precisa de los parámetros p y q del modelo ARIMA.

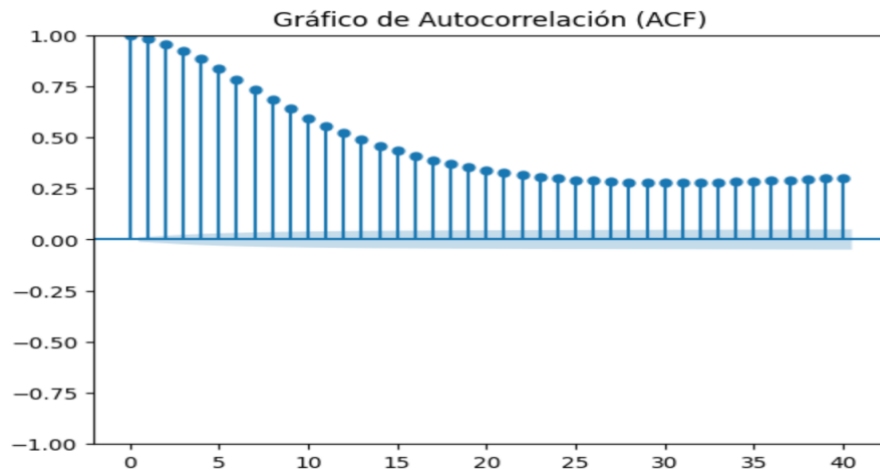


Fig. 8. Gráfico de Autocorrelación del modelo

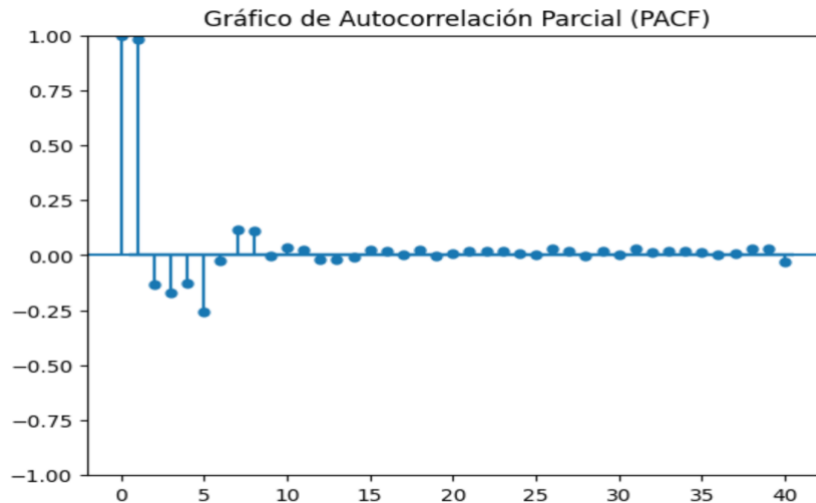


Fig. 9. Grafico de Autocorrelación parcial

El gráfico ACF , que se encuentra en la figura 8, proporciona una medida de la correlación entre puntos de una serie de tiempo y sus lags. En el gráfico presentado, observamos un decaimiento gradual de los picos de autocorrelación a medida que aumentan los lags, lo que podría sugerir un componente autoregresivo significativo en los datos.

Por otro lado, el gráfico PACF en la figura 9 muestra la correlación parcial entre la serie y sus lags, controlando los valores de los demás lags intermedios. La presencia de picos significativos en los primeros lags, seguidos de correlaciones que se acercan a cero, puede ser indicativa de la orden autoregresiva ‘p’ en un modelo AR(p). En el gráfico proporcionado, el marcado descenso tras el primer lag sugiere que un modelo AR(1) podría ser apropiado para los datos en cuestión.

Etapas 5: Evaluación del modelo

En la presente investigación, se ha procedido a la evaluación comparativa de dos modelos ARIMA, concretamente ARIMA(1,0,1) y ARIMA(1,0,0), para pronosticar la afluencia de pasajeros en el sistema TransMilenio, enfocándonos en la troncal de la calle 26. La evaluación se ha basado en criterios estadísticos, incluyendo la verosimilitud logarítmica, el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC). Los resultados son expuestos en la tabla 4. El modelo ARIMA(1,0,1) exhibió un desempeño superior. Los valores expuestos en la tabla 4 indican un ajuste más favorable en comparación con el modelo ARIMA(1,0,0). Además, la significancia de los residuos en ambos modelos fue corroborada por la prueba Ljung-Box, reflejando la independencia temporal de los mismos y apoyando así la idoneidad de los modelos en el análisis predictivo.

Criterio / Modelo	ARIMA(1,0,1)	ARIMA(1,0,0)
<i>Log Likelihood</i>	-184 664.43	-184 891.47
<i>AIC</i>	369 336.85	369 788.95
<i>BIC</i>	369 370.502	369 814.19

Table 4. Resultados de los modelos de ARIMA.

Teniendo en cuenta lo anterior, se elige el modelo de ARIMA (1,0,1) para el diseño del DT del proyecto.

Etapla 6: Implementación del modelo

En el marco de esta investigación, se ejecutaron el pronóstico de 96 tipos a lo largo del día de simulación para pronosticar el flujo de entradas y salidas del sistema TransMilenio, específicamente para la troncal de la calle 26. Estas simulaciones se efectuaron tanto para días entre semana como para fines de semana, con el propósito de capturar la dinámica diferenciada que caracteriza los patrones de movilidad en estos períodos. Cada corrida de pronóstico representa una ventana temporal de 15 minutos, abarcando así las 24 horas del día y permitiendo la construcción de un modelo integral que refleje la actividad completa de un día típico. Este enfoque meticuloso, que distingue entre los días laborables y no laborables, es crucial para la precisión del modelo Digital Twin (DT) desarrollado, ya que reconoce y adapta el análisis a las variaciones en la demanda de transporte asociadas a la rutina diaria de los usuarios del sistema que se vio que sucedía. Basado en el pronóstico y en la data ya procesada se puede implementar dentro del DT.

4.2 Metodología GAIA para MAS

La metodología GAIA, aplicada en el desarrollo del Digital DT del sistema TransMilenio a través de un MAS en NetLogo, aborda la construcción de un modelo complejo que emula la dinámica de transporte público de la troncal de la calle 26. En este modelo, la composición individual de los agentes y su interacción colectiva se delinean meticulosamente para reflejar las operaciones del mundo real.

El uso de MAS para la construcción de un DT es beneficioso dada la interacción que puede generarse entre las entidades inteligentes que lo componen las cuales pueden simular de manera asertiva la complejidad de las interacciones del mundo físico. En este capítulo se montaran las etapas que se realizaron para la construcción del modelo de DT teniendo en cuenta los datos procesados y el modelo para la predicción de ingresos y salidas.

Etapla 1: Requerimientos del Sistema de agentes

En esta etapa se definirán los requerimientos según los cuales se basa el modelo de DT que se presenta en este proyecto. Los requerimientos son extraídos tanto de las etapas anteriores del modelo analítico y la evaluación de la data que se realizó, como

de un entendimiento de los objetivos del proyecto y del negocio. Para esto se plantearon cuatro requerimientos:

- *Fidelidad Operativa*: Cada agente en el sistema debe estar modelado con atributos detallados que reflejan sus contrapartes físicas, asegurando una representación precisa del sistema de transporte.
- *Sincronización Temporal*: El DT se mantiene actualizado con datos reales y predicciones generadas por modelos estadístico, lo que permite simular escenarios futuros y responder dinámicamente a cambios en la demanda de transporte.
- *Interactividad y Autonomía de Agentes*: Los agentes poseen la capacidad de tomar decisiones y realizar acciones basadas en su estado y en la información recibida del entorno, lo que incluye la adaptación a condiciones variables y la interacción con otros agentes.
- *Integración y Escalabilidad*: El sistema es diseñado para integrarse fácilmente con nuevas fuentes de datos y puede escalar para incorporar más rutas, vehículos y patrones de tráfico, lo que lo hace adaptable a la evolución del sistema TransMilenio.

Etapas 2: Marco conceptual de Sistema de agentes

Esta sección explica los elementos conceptuales del modelo de DT. El modelo está compuesto por entidades de decisión que interactúan con tres niveles de software para el comportamiento conjunto del sistema. Esta configuración permite que el sistema funcione. El modelo sugerido está compuesto por un conjunto de entidades de toma de decisiones autónomas y reactivas.

Agentes como entidades decisionales

Para efectuar el marco conceptual del MAS, se va a delimitar a utilizar el término de entidad decisional para delimitar los agentes del sistema. Las entidades decisionales tienen un comportamiento predeterminado que contribuyó a las capacidades reactivas y cooperativas del sistema [35]. Cada entidad decisional contiene diferentes propiedades que le permiten actuar y ser identificada. Estas propiedades incluyen algoritmos de toma de decisiones, bases de conocimiento y protocolos de comunicación. Al combinar estas propiedades, las entidades de decisión pueden analizar y procesar información, tomar decisiones basadas en reglas predefinidas o patrones aprendidos e interactuar con otras entidades de manera coordinada. Esto permite que el sistema responda eficazmente a entornos dinámicos y alcance sus objetivos de manera eficiente. Cada entidad decisional tiene cuatro características principales:

- La meta u objetivo de la entidad decisional es un indicador medible que está asignada a cada entidad por individual o puede ser compartida entre varias entidades. Esta meta puede ser una ubicación, estado, destino e incluso un indicador incluye su forma de actuar y tomar decisiones, así mismo.
- La forma de toma de decisiones es la forma de alcanzar una meta, para ello se evalúan las opciones entre sí, y se elige la mejor opción de acuerdo a los parámetros

y restricciones previamente dadas. Una técnica de toma de decisiones debe estar basada en la información que recopilan del ambiente o de los parámetros asignados.

- Los parámetros de las entidades decisionales son información específica de la entidad y pueden ser compartidos entre un grupo de entidades llamadas razas. Estos parámetros que ya tengan establecidos desde el principio de la interacción y pueden ser variables tanto estáticas como modificables.
- La comunicación se refiere a que las entidades interactuaran con otras entidades que están dentro del mismo entorno. Esta interacción incluye, pero no se limita a, compartir el valor del parámetro o de las metas, si así lo requieren. Por medio de esta interacción se puede tener una efectiva forma de toma de decisiones.

Entidades decisionales del modelo

Con el fin de generar un DT que permita reflejar las operaciones del mundo real de TransMilenio. En este modelo, la composición individual de las entidades decisionales y su interacción colectiva se delinean por medio de dos tipos: globales y locales.

- Entidad decisional global: Este tipo de entidad, que se utiliza para coordinar la interacción entre las entidades locales y el modelo predictivo. Por tanto, el objetivo de esta entidad de decisión es tratar la información y el estado del sistema para que este pueda llevar a cabo las actividades designadas
- Entidades decisionales locales: Se consideran las entidades que interactúan para llegar a un objetivo dentro del sistema. Para este modelo estas entidades tiene asignadas razas que comparten entre si la caracterización del sistema. Los agentes en este modelo se clasifican en dos categorías esenciales: estaciones y autobuses. El objetivo principal de las estaciones es de servir a los autobuses de la información de flujo de personas. Por otro lado, el objetivo de los autobuses es completar la ruta designada.

Arquitectura de las entidades decisionales

La arquitectura del modelo multi-agente se compone de dos entornos diferenciados que coexisten. El primero, es el entorno físico donde se encuentran todos los elementos de entorno, las personas, estaciones, los puntos de geolocalización, etc. Por otra parte, está el entrono virtual donde se distinguen tres capas principales que trabajan en concierto: la capa de DT, la capa decisional global y el modelo predictivo.

La capa de DT alberga las entidades decisionales locales, que son los autobuses y las estaciones, actuando como réplicas virtuales de sus contrapartes físicas. Estas entidades ejecutan la simulación basándose en datos integrados del entorno físico, que incluye los vehículos reales, pasajeros y la infraestructura geográfica del sistema de transporte.

Por encima de la capa de DT, se sitúa la capa decisional global, que tiene la función de coordinar las actividades de las entidades locales. Esta capa es crucial para asegurar que el sistema funcione como un todo cohesivo, dirigiendo y coordinando las operaciones.

Finalmente, el modelo predictivo, situado en el tercer nivel, se encarga de analizar tendencias y datos, suministrando proyecciones y conocimientos que influyen en la

toma de decisiones. Utiliza métodos estadísticos para ver patrones de demanda, flujo de tráfico y otras variables críticas, proveyendo así una retroalimentación esencial a la capa decisional global y permitiendo adaptaciones proactivas en la operación simulada.

En conjunto, estas tres capas forman un sistema integrado que permite al DT del TransMilenio funcionar como una herramienta de análisis.

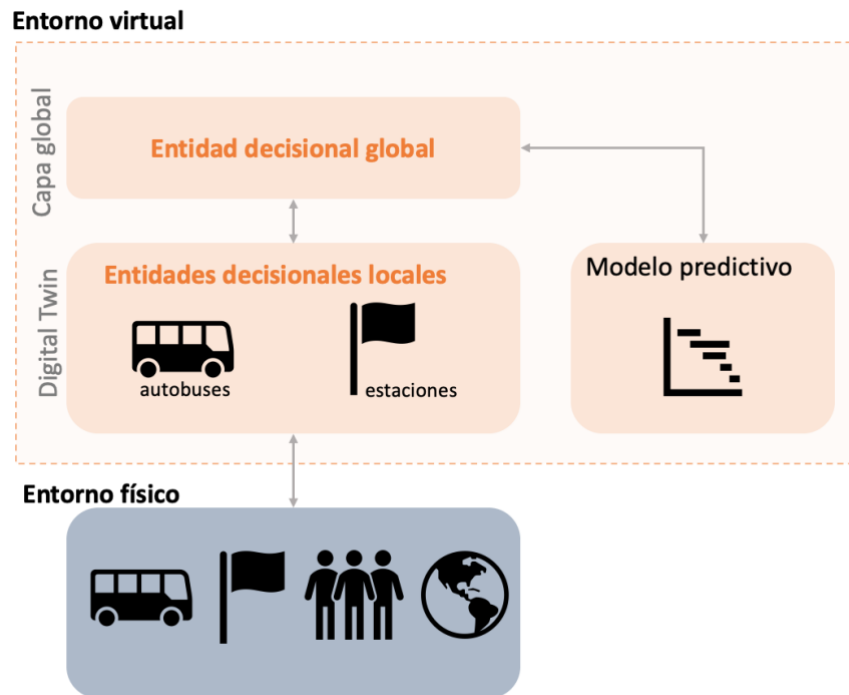


Fig. 10. Arquitectura de las entidades decisionales para la construcción del DT

Etapas 3: Implementación de Digital Twin

En esta sección se va a mostrar el detalle de la implementación del modelo del DT. En esta sección la metodología GAIA, aplicada en el desarrollo del Digital Twin (DT) del sistema TransMilenio a través de un MAS muestra el detalle de la implementación realizada.

Atributos de los agentes y variables globales

En este modelo, la composición individual de los agentes y su interacción colectiva se delinean meticulosamente para reflejar las operaciones del mundo real. Los agentes en este modelo se clasifican en dos razas: estaciones y autobuses. Cada estación, representada por la familia `stations`, posee atributos tales como `stop_id`, `stop_code`, `stop_name`, `location_type` y `parent_station`, que corresponden a la información extraída de los archivos GTFS. Estos atributos permiten modelar con fidelidad las propiedades físicas y operativas de cada punto de parada en la red de transporte.

Parámetro	Descripción
stop_id	Identificador único de la estación, según la especificación GTFS.
stop_code	Identificador corto y textual para la estación, usualmente un número o código simple.
stop_name	Nombre legible de la estación.
location_type	Indica el tipo de estación; por ejemplo, si es una parada o una estación.
parent_station	Identificador de la estación principal si la estación es parte de un complejo mayor.

Table 5. Parámetros de las Estaciones

Los autobuses, pertenecientes a la familia `autobuss`, se caracterizan por atributos como `ruta_id`, que indica la ruta específica del autobús según la definición del archivo `trips.txt`, y `posicionEnRuta`, que denota la posición actual del autobús en su ruta programada. Estos agentes también poseen un registro `ParadaArray`, que es una lista de las próximas paradas, y un estado `enParada`, que indica si el autobús está actualmente detenido en una estación. Adicionalmente, cada autobús tiene un `tiempoDeEspera`, que es un contador que gestiona el tiempo de estancia en cada parada.

Parámetro	Descripción
ruta_id	Identificador de la ruta que sigue el autobús, obtenido de los datos GTFS de viajes.
posicionEnRuta	Posición actual del autobús en su ruta, que puede ser un índice en la lista de paradas.
ParadaArray	Array que representa la secuencia de paradas para el autobús.
enParada	Bandera booleana que indica si el autobús está actualmente en una parada.
start_time	Hora programada para el inicio de la ruta del autobús.
status	Estado actual del autobús, como 'en movimiento', 'esperando' o 'no iniciado'.
tiempoDeEspera	Contador que indica el tiempo de espera en una parada.
siguienteParada	Referencia a la próxima parada en la ruta del autobús.

Table 6. Parámetros de los Autobuses

Las variables globales del entorno de simulación donde los autobuses se mueven de acuerdo a un reloj global `clock`. En la tabla 7 se pueden observar las características y la descripción de estas variables dentro del sistema.

Variable	Descripción
clock	El reloj del sistema, que avanza en unidades de tiempo simulado.
pas_sim	El paso de tiempo de la simulación, donde cada tick corresponde a una fracción de segundo.
step	La distancia avanzada por las entidades por tick en la simulación, afectando tanto las coordenadas x como y.

Table 7. Variables Globales

Sistema multi-agente

La figura 11 proporciona una visión integral del diagrama UML para la implementación de un Digital Twin (DT) del sistema TransMilenio. En este esquema, las entidades decisionales locales, autobuses y estaciones, son coordinadas por una entidad decisional global que establece los parámetros iniciales y asegura que la réplica virtual refleje con precisión el layout físico. Esto se logra mediante la geolocalización escalada de las estaciones y la integración de datos proyectados por el modelo predictivo para influir en el comportamiento autónomo de los agentes. Este modelo multiagente integra los datos analíticos obtenidos del modelo ARIMA, alinear los tiempos de llegada y partida de los autobuses con las predicciones de demanda de pasajeros. Las predicciones de entradas y salidas generadas por ARIMA informan la frecuencia de los autobuses y los patrones de flujo de pasajeros, optimizando así la eficiencia del sistema y la experiencia del usuario.

Una vez que el DT está completamente configurado, se inicia la simulación siguiendo un horario preestablecido. Los autobuses parten del punto inicial de sus rutas y transitan sucesivamente por las estaciones designadas. Al arribar a cada estación, los autobuses ajustan su volumen de pasajeros, descargando y recogiendo personas según la capacidad del vehículo y el flujo previsto en el modelo.

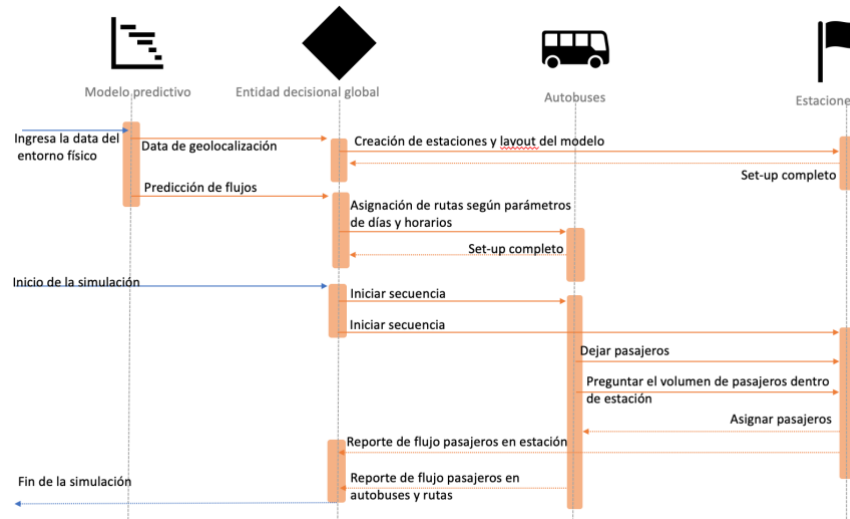


Fig. 11. Diagrama UML del comportamiento del modelo

Tanto los autobuses como las estaciones, en su rol de entidades locales, generan reportes continuos hacia la entidad global. Estos reportes incluyen información sobre los flujos de pasajeros y la ocupación de los autobuses, permitiendo un monitoreo detallado y en tiempo real de la dinámica del sistema. Esta retroalimentación es vital para mantener la precisión del DT y para proporcionar insights que podrían ser utilizados para futuras mejoras operativas y estratégicas en el sistema de transporte TransMilenio.

Etapas 4: Evaluación del modelo.

La etapa de evaluación del Digital Twin para el sistema TransMilenio revela una operatividad eficaz y una representación fidedigna del sistema de transporte en cuestión. En la figura 12, se presenta el layout final donde se destacan las estaciones organizadas conforme a su distribución geoespacial, lo que permite una visualización clara de la infraestructura de la Troncal de la calle 26. La figura 13, por su parte, ofrece una visión dinámica del Digital Twin durante un periodo de alta demanda, evidenciando la capacidad del modelo para simular la realidad operacional en horas pico con un alto grado de detalle y precisión.



Fig. 12. Layout y organización de las estaciones en Netlogo

Este desarrollo se ha llevado a cabo en NetLogo, un entorno de modelado multiagente robusto y accesible que permite la simulación de sistemas complejos y dinámicos. NetLogo se caracteriza por su interfaz intuitiva y su lenguaje de programación de alto nivel, facilitando la implementación y experimentación con modelos computacionales que representan fenómenos naturales y sociales. esta herramienta se integró con Python para la creación del DT, permitiendo la integración de datos, la creación de entidades autónomas y la visualización de estas entidades.



Fig. 13. Perspectiva de la simulación de día entre semana

En la figura 14 se puede observar un heatmap que muestra la distribución del tráfico de pasajeros por hora y estación durante los días entre semana. Las áreas con colores más claros indican una mayor cantidad de tráfico en ciertas horas, lo que sugiere picos de actividad, probablemente durante las horas de inicio y fin de la jornada laboral. Este patrón se repite con consistencia, lo que podría indicar un flujo de pasajeros predecible y regular en estos horarios.

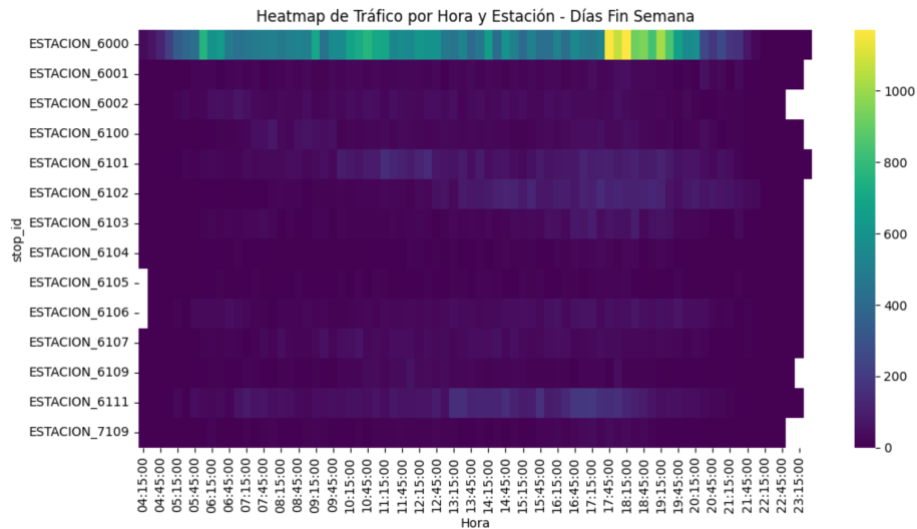


Fig. 14. Mapa de calor de tráfico por hora por estación en fines de semana

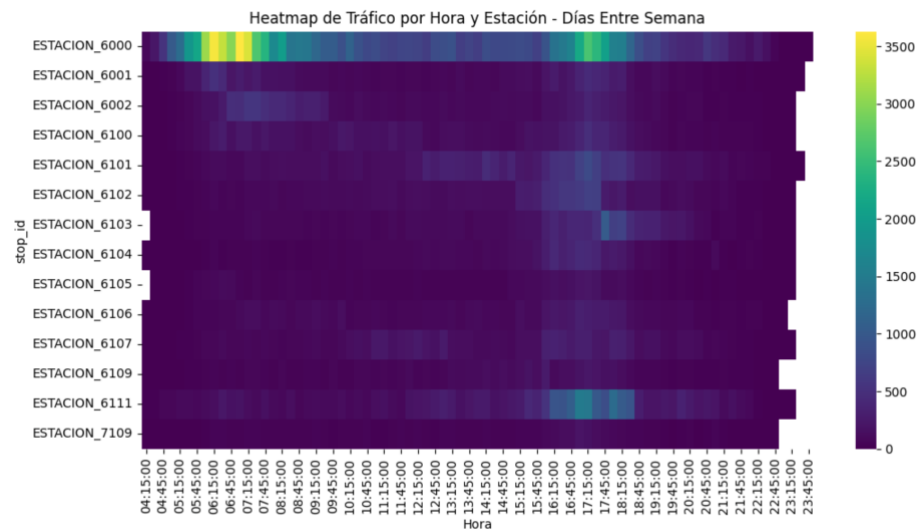


Fig. 15. Mapa de calor de tráfico por hora por estación en días entre de semana

Por otro lado, la figura 15 representa la distribución de los días de fin de semana. Aquí, los patrones de tráfico son notablemente diferentes, con menos consistencia en los picos de tráfico y una disminución general en el volumen de pasajeros. Esto es de esperar, ya que los patrones de desplazamiento los fines de semana suelen ser menos estructurados que los días entre semana debido a la naturaleza de las actividades realizadas en estos días y que la Troncal de la 26 se ve altamente influenciada por los lugares de trabajo que quedan cerca a esta.

Por último, el gráfico 16 muestra una comparación de las salidas por hora entre días entre semana y los fines de semana. Las barras azules representan los días entre semana y las naranjas los fines de semana. Se observa claramente que el volumen de salidas es mayor durante los días entre semana en casi todas las horas del día, lo que refleja el ciclo habitual de viajes de trabajo y escuela.

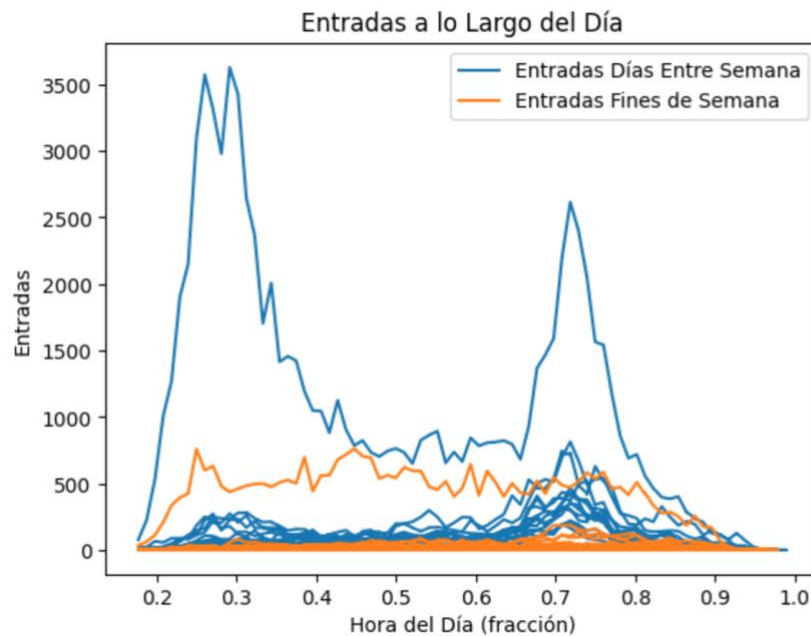


Fig. 16. Entradas a lo largo del día

Estos gráficos reflejan un comportamiento que se alinea estrechamente con los patrones observados en las etapas iniciales del proyecto, validando la capacidad del Digital Twin para replicar con precisión la dinámica del sistema de transporte real.

5 Conclusiones y trabajo futuro

5.1 Conclusiones

El desarrollo de un Digital Twin (DT) para el sistema de TransMilenio, basado en la robusta metodología GAIA y ejecutado a través de NetLogo, ha demostrado ser una

herramienta invaluable para la simulación y análisis del transporte público urbano. El modelo ha incorporado con éxito múltiples capas de información, incluyendo datos de geolocalización y predicciones de flujo de pasajeros, para crear un entorno virtual dinámico y reactivo que emula fielmente las operaciones del entorno físico. Los agentes del modelo, tanto autobuses como estaciones, reflejan las características y comportamientos de sus contrapartes reales, y están interconectados de tal manera que permiten una simulación detallada y precisa de la actividad diaria. La implementación de este DT ha permitido no solo visualizar y comprender la complejidad de los patrones de tráfico sino también proporcionar un marco para la toma de decisiones informadas y la planificación estratégica. La capacidad del modelo para simular con precisión el tráfico en diferentes escenarios demuestra su potencial para ser una herramienta predictiva y de planificación en el manejo de la infraestructura de transporte urbano. Como tal, el DT se establece como un paso adelante hacia la integración de tecnologías avanzadas en la gestión del transporte público, ofreciendo posibilidades para futuras investigaciones y mejoras operativas.

5.2 Trabajo futuro

Mirando hacia el futuro, el DT puede ser ampliado y mejorado de varias maneras. La integración de datos en tiempo real para actualizar continuamente el modelo podría ofrecer un análisis aún más dinámico y en tiempo real. Además, la expansión del modelo para incluir más rutas y variables ambientales permitirá una comprensión más completa de los sistemas de transporte a gran escala. La colaboración con desarrollos en inteligencia artificial podría ofrecer nuevas formas de optimizar rutas y manejar la congestión. Finalmente, la adaptabilidad del DT para incluir consideraciones de sostenibilidad y la transición hacia opciones de transporte más ecológicas podría ser una valiosa dirección para la investigación futura.

Agradecimientos

A mis padres, Fernando y Nubia que me han acompañado en cada momento.

6 Bibliografía

- [1] A. Valderrama and U. Jørgensen, “Urban Transportation Systems in Bogotá and Copenhagen: An Approach from STS,” *Built Environ*, vol. 34, no. 2, pp. 200–217, 2008, doi: 10.2148/BENV.34.2.200.
- [2] M. Ghazal, R. Hamouda, and S. Ali, “A Smart Mobile System for the Real-Time Tracking and Management of Service Queues,” *International Journal of Computing and Digital Systems*, vol. 5, no. 4, pp. 2210–142, 2016, doi: 10.12785/ijcds/050402.
- [3] A. Kumar Sharma, R. Pandey, S. Tarafdar, and S. Dubey, “Towards Smart Mobility in Cities - Bus Tracking and Booking System,” *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends*

- and Future Directions), *ICRITO 2021*, 2021, doi: 10.1109/ICRITO51393.2021.9596492.
- [4] S. Geetha and D. Cicilia, "IoT enabled intelligent bus transportation system," *Proceedings of the 2nd International Conference on Communication and Electronics Systems, ICCES 2017*, vol. 2018-January, pp. 7–11, Jul. 2018, doi: 10.1109/CESYS.2017.8321235.
- [5] C. D. Cottrill and S. Derrible, "Leveraging Big Data for the Development of Transport Sustainability Indicators," *Journal of Urban Technology*, vol. 22, no. 1, pp. 45–64, 2015, doi: 10.1080/10630732.2014.942094.
- [6] J. A. Erkoyuncu, I. F. del Amo, D. Ariensyah, D. Bulka, R. Vrabich, and R. Roy, "A design framework for adaptive digital twins," *CIRP Annals*, vol. 69, no. 1, pp. 145–148, 2020, doi: 10.1016/j.cirp.2020.04.086.
- [7] R. Stark, C. Freseman, and K. Lindow, "Development and operation of Digital Twins for technical systems and services," 2019, doi: 10.1016/j.cirp.2019.04.024.
- [8] K. Ding, F. T. S. Chan, X. Zhang, G. Zhou, and F. Zhang, "Defining a Digital Twin-based Cyber-Physical Production System for autonomous manufacturing in smart shop floors," *Int J Prod Res*, vol. 57, no. 20, pp. 6315–6334, Oct. 2019, doi: 10.1080/00207543.2019.1566661.
- [9] C. Bachechi, "Digital Twins for Urban Mobility," *Communications in Computer and Information Science*, vol. 1652 CCIS, pp. 657–665, 2022, doi: 10.1007/978-3-031-15743-1_61/COVER.
- [10] T. Y. Melesse, V. Di Pasquale, and S. Riemma, "Digital Twin Models in Industrial Operations: A Systematic Literature Review," *International Conference on Industry 4.0 and Smart Manufacturing (IS)*, vol. 42, pp. 267–272, 2020, doi: 10.1016/j.promfg.2020.02.084.
- [11] K. Josifovska, E. Yigitbas, and G. Engels, "Reference Framework for Digital Twins within Cyber-Physical Systems," in *2019 IEEE/ACM 5th International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS)*, IEEE, May 2019, pp. 25–31. doi: 10.1109/SEsCPS.2019.00012.
- [12] K. M. Alam and A. El Saddik, "C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems," *IEEE Access*, vol. 5, pp. 2050–2062, 2017, doi: 10.1109/ACCESS.2017.2657006.
- [13] S. A. P. Kumar, R. Madhumathi, P. R. Chelliah, L. Tao, and S. Wang, "A novel digital twin-centric approach for driver intention prediction and traffic congestion avoidance," *J Reliab Intell Environ*, vol. 4, no. 4, pp. 199–209, Dec. 2018, doi: 10.1007/S40860-018-0069-Y/METRICS.
- [14] Z. Fan *et al.*, "Online Trajectory Prediction for Metropolitan Scale Mobility Digital Twin," *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, vol. 18, p. 12, Jun. 2022, doi: 10.1145/3557915.3561040.
- [15] M. A. Khamis and W. Gomaa, "Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework," *Eng Appl Artif Intell*, vol. 29, pp. 134–151, Mar. 2014, doi: 10.1016/J.ENGAPPAI.2014.01.007.

- [16] S. Wang, F. Zhang, and T. Qin, "Urban traffic optimal solution based on digital twin system technology simulation," *International Conference on Computer Modeling and Simulation*, pp. 83–87, Jun. 2022, doi: 10.1145/3547578.3547591.
- [17] A. Pratelli *et al.*, "A Framework to Develop Urban Aerial Networks by Using a Digital Twin Approach," *Drones*, vol. 6, no. 12, Dec. 2022, doi: 10.3390/DRONES6120387.
- [18] T. Clemen *et al.*, "Multi-Agent Systems and Digital Twins for Smarter Cities," *SIGSIM Principles of Advanced Discrete Simulation*, pp. 45–55, May 2021, doi: 10.1145/3437959.3459254.
- [19] Q. Long and W. Zhang, "An integrated framework for agent based inventory-production-transportation modeling and distributed simulation of supply chains," *Inf Sci (N Y)*, vol. 277, pp. 567–581, 2014, doi: 10.1016/j.ins.2014.02.147.
- [20] M. Mahunnah and K. Taveter, "A scalable multi-agent architecture in environments with limited connectivity: Case study on individualised care for healthy pregnancy," in *2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, IEEE, Jul. 2013, pp. 84–89. doi: 10.1109/DEST.2013.6611334.
- [21] B. Chen, H. H. Cheng, and J. Palen, "Integrating mobile agent technology with multi-agent systems for distributed traffic detection and management systems," *Trans Res Part C Emerg Technol*, vol. 17, no. 1, pp. 1–10, 2009, doi: 10.1016/j.trc.2008.04.003.
- [22] H. P.-M. systems: a modern approach to and undefined 1999, "Industrial and practical applications of DAI," *books.google.comHVD ParunakMultiagent systems: a modern approach to distributed artificial, 1999•books.google.com*, Accessed: Nov. 15, 2023. [Online]. Available: https://books.google.com/books?hl=es&lr=&id=JYcznFCN3xcC&oi=fnd&pg=PA377&dq=+Industrial+and+practical+applications+of+DAI&ots=IK_QpFNlYy&sig=dMKYul9UkL5_30OzwYUOGH32L6s
- [23] Z. Lv, D. Chen, H. Feng, A. K. Singh, W. Wei, and H. Lv, "Computational Intelligence in Security of Digital Twins Big Graphic Data in Cyber-physical Systems of Smart Cities," *ACM Trans Manag Inf Syst*, vol. 13, no. 4, p. 39, Aug. 2022, doi: 10.1145/3522760.
- [24] T. Clemen *et al.*, "Multi-Agent Systems and Digital Twins for Smarter Cities," *SIGSIM Principles of Advanced Discrete Simulation*, pp. 45–55, May 2021, doi: 10.1145/3437959.3459254.
- [25] H. Nguyen, A. Hussein, M. A. Garratt, and H. A. Abbass, "Swarm Metaverse for Multi-Level Autonomy Using Digital Twins," *Sensors (Basel)*, vol. 23, no. 10, May 2023, doi: 10.3390/S23104892.
- [26] J. C. García-Ojeda, J. De, J. Pérez-Alcázar, and A. E. Arenas, "Extending the Gaia Methodology with Agent-UML," 2004. Accessed: May 23, 2021. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.7051&rep=rep1&type=pdf>

- [27] W. Huang, E. El-Darzi, and L. Jin, "Extending the Gaia Methodology for the Design and Development of Agent-based Software Systems," in *31st Annual International Computer Software and Applications Conference - Vol. 2 - (COMPSAC 2007)*, IEEE, Jul. 2007, pp. 159–168. doi: 10.1109/COMPSAC.2007.114.
- [28] "Misión, Visión, funciones y deberes de TRANSMILENIO." Accessed: Nov. 16, 2023. [Online]. Available: <https://www.transmilenio.gov.co/publicaciones/146031/mision-y-vision-de-transmilenio/>
- [29] "Plan Estratégico." Accessed: Nov. 16, 2023. [Online]. Available: <https://www.transmilenio.gov.co/publicaciones/146045/plan-estrategico/>
- [30] "Proyectos de Inversión." Accessed: Nov. 16, 2023. [Online]. Available: <https://www.transmilenio.gov.co/publicaciones/146047/proyectos-de-inversion/>
- [31] "Estadísticas de oferta y demanda del Sistema Integrado de Transporte Público SITP." Accessed: Nov. 16, 2023. [Online]. Available: <https://www.transmilenio.gov.co/publicaciones/149180/estadisticas-de-oferta-y-demanda-del-sistema-integrado-de-transporte-publico-sitp/>
- [32] TransMilenio S.A, "GTFS Estáticos | Portal de Datos Abiertos de TransMilenio." Accessed: Nov. 16, 2023. [Online]. Available: <https://datosabiertos-transmilenio.hub.arcgis.com/search?collection=Document&tags=gtfs>
- [33] Y. Ye, L. Chen, and F. Xue, "Passenger flow prediction in bus transportation system using ARIMA models with big data," *Proceedings - 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2019*, pp. 436–443, Oct. 2019, doi: 10.1109/CYBERC.2019.00081.
- [34] Y. Su and Y. Ye, "Daily Passenger Volume Prediction in the Bus Transportation System using ARIMAX Model with Big Data," *Proceedings - 2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2020*, pp. 291–300, Oct. 2020, doi: 10.1109/CYBERC49757.2020.00055.
- [35] J.-F. Jiménez, "Dynamic and hybrid architecture for the optimal reconfiguration of control systems: application to manufacturing control," 2017, doi: 10.1515/eko-2017-0018.