

Guía 3 – Informe Técnico Resultado de Investigación.

Título del proyecto:

Predicción de demanda mediante redes neuronales para la ruta T11 del SITP.

Presentado por:

Luisa Fernanda Saldaña Toro.

Programa:

Especialización en Gerencia de Proyectos.

Tutora:

Lina María Chacón Rivera.

Asignatura:

Seminario de investigación.

30 de noviembre de 2023

Bogotá. D.C.

Resumen

La falta de precisión en la estimación del número de pasajeros que utilizan la ruta T11 del SITP en Bogotá ha generado ineficiencias operativas y ha causado insatisfacción entre los usuarios. Por este motivo, este proyecto de investigación propone la adopción de un modelo de predicción de demanda mediante redes neuronales. Para lograr este propósito, se apoya en teorías de aprendizaje automático y en recursos provenientes de fuentes especializadas en la implementación de inteligencia artificial en el transporte público.

Palabras clave: Redes Neuronales Artificiales (RNA), pronóstico, análisis de demanda, movilidad, transporte público.

1. PROBLEMA DE INVESTIGACIÓN.

1.1 Descripción del problema.

Actualmente el Sistema Integrado del Transporte Público, en adelante SITP, se compone de dos fases las cuales están integradas por 18 empresas operando con un total de 382 rutas las cuales se dividen entre: urbanas, complementarias y especiales (TRANSMILENIO S.A., 2012), estas tienen como propósito transportar a los usuarios a través de los corredores viales de la ciudad, así como también buscan garantizar un sistema transporte público confiable y eficiente que permita abordar los constantes desafíos de movilidad y sostenibilidad. En este contexto, es que se plantea el evaluar la viabilidad de estudiar la ruta T11 ya que desempeña un papel crucial al conectar diversas zonas de la ciudad y movilizar cerca de 700.000 usuarios al mes la cual concentra cerca del 16 % de la demanda de la ciudad.

Durante el primer semestre de 2023 la ruta ha presentado una alta variabilidad en la demanda de usuarios, pese a que se espera que exista una variación considerable dados factores externos a la operación como: la estacionalidad presentada en meses como enero, abril y julio, las franjas horarias habituales de movilización de los usuarios y las posibles contingencias como manifestaciones, desvíos, entre otras, no ha sido posible determinar cuál es el factor que está impactando en mayor medida la demanda de la ruta actualmente.

Por otro lado, conviene mencionar que en la ruta se presentan aspectos que han podido afectar su adecuada operación, en donde para los primeros seis meses del año se han ofertado con una alta frecuencia servicios con sobrecarga de usuarios en las franjas pico, lo que causa que los pasajeros deban esperar tiempos prolongados entre servicio y servicio. Asimismo este comportamiento da cuenta de las deficiencias en la programación, ya que se presentan distribuciones irregulares de los vehículos durante el día y esto puede resultar en tiempos de espera impredecibles y prolongados en las paradas, lo que a su vez se deriva en una ineficiencia en la asignación de recursos como operadores y vehículos, desperdiciando el combustible de los buses y el tiempo de los conductores. Un método que ayuda a tener una perspectiva más acertada de la problemática aquí expuesta es el análisis PESTAL, el cual se resume a continuación:

- Político: Esta parte se relaciona con el financiamiento, regulaciones y políticas de transporte que tienen un impacto directo en la demanda de usuarios del sistema. Así

como también incluye cambios en las políticas de incentivos fiscales, subsidios y restricciones de tráfico que influyen en la cantidad de personas que optan por utilizar el SITP.

- **Económico:** Factores asociados al nivel de ingresos, tasas de desempleo y costos de transporte y las fluctuaciones económicas pueden afectar la capacidad de las personas para costear el transporte privado y, por lo tanto, aumentar la demanda de opciones más asequibles como la bicicleta.
- **Sociocultural:** Patrones de movilidad, preferencias de viaje, densidad poblacional y estilo de vida influyen en la demanda de usuarios en el transporte público. Del mismo modo, los cambios en las preferencias culturales y sociales, como una mayor conciencia ambiental y la búsqueda de modos de transporte más sostenibles, pueden impulsar el uso del transporte público.
- **Tecnológico:** Las innovaciones tecnológicas como aplicaciones móviles para rastrear horarios y planificar rutas pueden impactar en la demanda de usuarios en el transporte público. La disponibilidad de información en tiempo real y opciones tecnológicas para el pago de tarifas pueden hacer que el transporte público sea más atractivo para ciertos grupos de usuarios.
- **Ambiental:** Las preocupaciones ambientales y la búsqueda de modos de transporte más ecológicos pueden aumentar la demanda de usuarios en el transporte público. La creciente conciencia sobre la reducción de la huella de carbono y la promoción de modos de transporte sostenibles pueden impulsar el uso del transporte público como una alternativa más amigable con el medio ambiente.
- **Legal:** Regulaciones sobre accesibilidad, seguridad, tarifas y derechos de los pasajeros también afectan la demanda de usuarios en el transporte público. Cambios en las leyes y regulaciones que afectan la calidad del servicio, las tarifas y la comodidad pueden influir en la elección de transporte público.

Llegados a este punto es oportuno precisar que, en caso de no tomar medidas correctivas es probable que se presenten diferentes comportamientos o tendencias que tengan un impacto negativo en la movilidad y la calidad del servicio prestado. Uno de los comportamientos que se está presentando y se puede agravar con el paso del tiempo, es el aumento de las

congestiones, esto se puede entender como buses que inician y/o finalizan un viaje con sobrecupo de usuarios, lo cual se deriva en ciclos de viajes más largos, tiempos de espera prolongados y una disminución en la confiabilidad del transporte público de la ciudad. En cuanto a términos de ineficiencia operativa se puede encontrar una tendencia creciente asociada con la asignación inadecuada de los recursos, el impacto ambiental y la pérdida de competitividad del sistema frente a otras alternativas de movilidad.

Otro factor importante a considerar es la percepción del usuario, la cual como se ha visto desde la implementación del SITP ha mostrado ser variada dependiendo de las necesidades de la población, no obstante, se ha podido evidenciar una marcada calidad de experiencia de usuario disminuida, lo cual da cuenta de la insatisfacción de los ciudadanos y a su vez permite identificar otro gran grupo de variables a contemplar tales como: la inseguridad en sistema, el desaseo de los buses, la evasión y elusión en el pago del pasaje, entre otras.

Como consecuencia de lo anteriormente expuesto es que se plantea el objeto de estudio de este proyecto, el cual se enfoca principalmente en llevar a cabo una prueba piloto para la ruta propuesta, en la que una red neuronal pueda predecir la demanda de pasajeros en diferentes horarios. Lo anterior podría ayudar a ajustar la oferta de transporte de manera más precisa, para así poder gestionar eficazmente las fluctuaciones en la demanda para garantizar un transporte público más eficiente y sostenible.

Este proyecto de investigación presenta una solución prometedora, y es que al utilizar redes neuronales y técnicas de inteligencia artificial, las cuales tienen “la capacidad de aprender de datos históricos y adaptarse a las fluctuaciones” en la demanda (Torres, Hernández y Pedraza, 2011, p. 91), se puede abordar la complejidad de los patrones de uso del transporte público en una ciudad en constante evolución como lo es Bogotá. Adicionalmente, esta solución propuesta implica la recopilación y el análisis de datos históricos detallados, incluyendo información sobre horarios, días de la semana, eventos especiales y condiciones climáticas, los cuales se utilizarán para entrenar y ajustar el modelo para que esté en condiciones de generar pronósticos precisos en diferentes momentos y situaciones.

El implementar esta medida, podría tener beneficios asociados con la calidad de la experiencia de viaje de los usuarios, haciendo que los tiempos de espera puedan ser más predecibles, y la reducir la sobrecarga en los vehículos del sistema para que los usuarios puedan

planificar sus desplazamientos de manera más eficiente, lo que impactará positivamente en la satisfacción de la prestación del servicio. Además, la operación se beneficiará de una distribución más eficiente de recursos, lo que puede resultar en reducción de costos y una utilización más efectiva de los recursos disponibles contribuyendo a la sostenibilidad financiera y operativa de la ruta.

En última instancia, y no menos importante, la confiabilidad del sistema de transporte público aumentaría considerablemente, lo que podría atraer a más usuarios y tener una mayor aceptación del transporte público zonal como una opción de movilidad preferida en la ciudad, asimismo esto podría tener beneficios adicionales en términos de reducción de la congestión del tráfico y la disminución de la huella de carbono en la ciudad.

1.2 Planteamiento del problema.

La eficiencia y la calidad de los servicios de transporte público son fundamentales para satisfacer las necesidades de movilidad de una ciudad en constante crecimiento como Bogotá. La ruta seleccionada como objeto de análisis desempeña un papel crucial al interconectar diferentes zonas de la ciudad y al ser una de las que registra mayor cantidad de pasajeros movilizados por kilómetro recorrido; sin embargo, se enfrenta a desafíos operativos relacionados con la predicción precisa de la demanda. La falta de estimaciones exactas de la demanda durante el día ha llevado a situaciones de congestión, tiempos de espera impredecibles y una experiencia insatisfactoria para los usuarios.

El problema central radica en la incapacidad de prever con precisión los patrones de demanda en la ruta, lo que ha resultado en una asignación ineficiente de recursos, sobrecarga en vehículos y una experiencia de viaje poco óptima para los usuarios. “Los enfoques tradicionales de predicción de demanda han demostrado limitaciones en la captura de patrones” (Ackermann y Sellitto, 2022, p. 83) complejos y variabilidad en los datos, lo que resalta la necesidad de una metodología más avanzada y adaptativa.

El tipo de estudio en este proyecto de investigación es correlacional, ya que busca establecer relaciones entre variables, en este caso, las variables relacionadas con la demanda de usuarios en la ruta y las variables que pueden influir en ella, como horarios, días de la semana, eventos especiales y condiciones climáticas. La correlación se explorará mediante la

implementación de modelos de redes neuronales que analizarán la relación entre estas variables para generar predicciones precisas de la demanda (García, 2019, p. 7). Además, se investigará si existen patrones de comportamiento que se pueden establecer a partir de datos históricos (Gorbea y Batista, 2018, p. 9) y cómo estos patrones pueden ser utilizados para mejorar la calidad del servicio de transporte público.

El sistema de transporte público en Bogotá se enfrenta a desafíos considerables en términos de eficiencia, calidad de servicio y satisfacción del usuario, dada la creciente población de la ciudad y la complejidad de sus desplazamientos diarios hacen que la operación de este sistema sea esencial para garantizar la movilidad sostenible y el bienestar de sus habitantes.

Tal como lo exponen los autores Alexei Gómez y Gabriel Sánchez, “la demanda de usuarios en el transporte público es una variable dinámica y multifacética” que varía en función de una serie de factores (Gómez y Sánchez, 2016, p. 164), como los horarios laborales, los eventos especiales, las condiciones climáticas y las estaciones del año. Esta variabilidad puede generar situaciones de congestión en los vehículos, tiempos de espera impredecibles y una experiencia general del usuario insatisfactoria.

Los métodos tradicionales para prever la demanda en el transporte público en Bogotá han demostrado ser insuficientes para abordar la complejidad de los patrones de uso en la ruta, estos métodos se basan principalmente en análisis estadísticos (Posada y González, 2010, p. 108) que no pueden capturar de manera efectiva la variabilidad en los datos históricos y las condiciones cambiantes que afectan la demanda. Además, no toman en cuenta la interconexión de múltiples factores que influyen en la cantidad de pasajeros que utilizan el servicio, como resultado de esto se generan una serie de problemas que afectan tanto a los operadores del sistema como a los usuarios finales.

Por un lado, la asignación de vehículos y personal se vuelve ineficiente, con vehículos que a menudo están sobrecargados en ciertos momentos y subutilizados en otros, lo cual conduce a una utilización ineficiente de recursos y a costos operativos innecesarios. Por otro lado, los usuarios enfrentan tiempos de espera prolongados y a menudo imprecisos en las paradas, lo que resulta en una experiencia de viaje incómoda y poco confiable. La sobrecarga en los vehículos también contribuye a un ambiente de viaje congestionado y poco agradable, además, los usuarios que no pueden abordar los vehículos debido a la sobrecarga enfrentan retrasos y descontento.

En este contexto, la planificación y gestión efectiva del transporte público (Gómez, 2020, p. 7) en las ciudades contemporáneas es un desafío constante que demanda la aplicación de enfoques innovadores y tecnologías avanzadas. La demanda de usuarios en el transporte público es una variable dinámica que puede ser influenciada por una multitud de factores, como horarios laborales, eventos especiales, condiciones climáticas y variaciones estacionales. Estas variaciones pueden resultar en situaciones de congestión, falta de capacidad en vehículos y disminución de la calidad del servicio. Por lo que una predicción precisa de la demanda se convierte, en un recurso invaluable para la planificación y toma de decisiones informadas en el ámbito del transporte público.

El caso de estudio de esta investigación se centra específicamente en la ruta T11 del SITP en Bogotá, la pregunta de investigación que guía este proyecto se formula como: "¿Cómo se puede mejorar la predicción de la demanda de usuarios en la ruta T11 del SITP en Bogotá mediante la implementación de redes neuronales y técnicas de inteligencia artificial?" Este argumento representa el punto de partida para un análisis correlacional a través de la exploración de metodologías y enfoques basados en redes neuronales, que busca evaluar el impacto de una aproximación más precisa y adaptativa a la variabilidad de la demanda, en la calidad del servicio.

2. OBJETIVOS.

2.1 Objetivo general.

Proponer un modelo de predicción de demanda de pasajeros para la ruta T11 del SITP en Bogotá, mediante redes neuronales con el propósito de incrementar significativamente la precisión en la estimación de la demanda y, derivado de esto, mejorar la planificación y operación eficiente de esta ruta del sistema de transporte público.

2.2 Objetivos específicos.

- Analizar datos históricos y en tiempo real de la ruta, incluyendo variables relevantes como franjas horarias, tipos de día, contingencias, condiciones climáticas y viales.
- Identificar patrones de demanda en los datos recopilados, destacando picos y valles en la utilización de la ruta y detectando tendencias a lo largo del tiempo.

- Seleccionar un modelo de red neuronal adecuado para la predicción de la demanda de usuarios en el transporte público.
- Evaluar el impacto de la implementación del modelo en la operación y la planificación de la ruta, considerando aspectos como: asignación de recursos, la satisfacción de los usuarios y eficiencia operativa del sistema.
- Proponer recomendaciones basadas en los resultados obtenidos, destacando cómo la implementación de la predicción de demanda mediante redes neuronales puede mejorar la toma de decisiones y la planificación operativa.

3. JUSTIFICACIÓN.

La adopción de tecnologías avanzadas, como redes neuronales, permite abordar la complejidad inherente a la predicción de la demanda de usuarios en el transporte público, estas redes, inspiradas en el funcionamiento del cerebro humano, tienen la capacidad de aprender de patrones y datos históricos, adaptándose a las fluctuaciones y cambios en el entorno (Torres, Hernández y Pedraza, 2011, p. 91). Al implementar esta tecnología a la problemática de la ruta T11 se busca superar las limitaciones de los enfoques tradicionales y lograr pronósticos más confiables y eficientes.

En esta instancia, el presente proyecto de investigación aspira a contribuir al mejoramiento de la operación del sistema de transporte público en Bogotá, primeramente y a modo de caso de estudio en la ruta, los resultados de la predicción precisa de la demanda en la ruta pueden conducir a una distribución más eficiente de recursos, y también a una experiencia de viaje más satisfactoria para los usuarios. Además, las lecciones aprendidas pueden sentar las bases para la aplicación de enfoques similares en otras rutas y sistemas de transporte público en la ciudad.

En un panorama en constante evolución, donde la innovación tecnológica y la optimización de procesos son esenciales, este proyecto representa un paso importante hacia la transformación, automatización y mejora del transporte público en Bogotá. Adicionalmente esta ruta transita por vías arteriales que conectan diversas zonas de la ciudad, y su operación eficiente es esencial para el funcionamiento del sistema de transporte público en su conjunto. Dado que las metodologías tradicionales de predicción han demostrado limitaciones en la captura de patrones

de demanda complejos (Gómez y Sánchez, 2016, p. 166), la conveniencia de este proyecto radica en su capacidad para abordar esta problemática y mejorar significativamente la calidad del servicio ofrecido a los usuarios. Asimismo, este tiene una alta relevancia social ya que beneficia directamente a la población de Bogotá que utiliza el transporte público a diario, permitiendo una experiencia de viaje más cómoda y predecible para los usuarios, reduciendo la congestión en los vehículos, minimizando los tiempos de espera y aumentando la confiabilidad del sistema. Esto no solo mejora la calidad de vida de los ciudadanos, sino que también puede fomentar una mayor adopción del transporte público, lo que a su vez puede contribuir a la reducción de la congestión del tráfico y la contaminación ambiental en la ciudad.

No obstante son muchos los beneficios aquí expuestos, es claro que el proyecto tiene implicaciones prácticas directas en la operación y gestión del transporte público, asignando más eficientemente los recursos, incluyendo vehículos y personal, lo que fácilmente se traduce en ahorros de costos y una utilización más efectiva de los recursos disponibles, beneficiando tanto a los operadores del sistema como a los usuarios. Además, la capacidad de planificación mejorada resultante de predicciones más precisas puede optimizar la toma de decisiones en tiempo real y mejorar la capacidad de respuesta ante cambios en las condiciones operativas.

Ahora bien, este proyecto contribuirá al avance de la investigación en el campo del uso de Deep learning y machine learning para la mejora del transporte público, específicamente en el contexto de redes neuronales y técnicas de inteligencia artificial (Yiu, 2019). La aplicación exitosa de estas metodologías en la ruta podría servir como un modelo para futuras investigaciones y proyectos en Bogotá y en otras ciudades con sistemas de transporte similares, lo que enriquecerá el conocimiento teórico y la comprensión de cómo las redes neuronales pueden ser utilizadas de manera efectiva en situaciones de predicción de demanda en el transporte público.

Como aporte metodológico podría decirse que este proyecto puede servir como un enfoque innovador que puede ser adaptado y replicado en otras rutas y sistemas de transporte público. La metodología aquí propuesta puede servir para optimizar los procesos y así mejorar la eficiencia operativa del transporte público y puede ser aplicada en el campo del transporte, la planificación urbana y el análisis de datos.

Por último es preciso mencionar que, el campo de investigación del presente proyecto investigativo corresponde al de “Ciencia, Tecnología e Innovación” y se enmarca en el grupo de "Tecnología ONTARE", cuyo énfasis se puede asociar al desarrollo y aplicación de soluciones tecnológicas avanzadas para la optimización de procesos en diversos sectores, en este caso en el sector de transporte urbano. De acuerdo con el objeto de la investigación la línea de investigación institucional de "Optimización de Procesos", es la que mejor se adecua al proyecto ya que este se relaciona con la implementación de tecnologías y/o metodologías de vanguardia para mejorar la eficiencia y la calidad en la prestación del servicio.

Concluyendo así que, este proyecto de investigación se basa en su conveniencia para abordar un problema crítico en el transporte público de Bogotá, su alta relevancia social para mejorar la calidad de vida de los usuarios, sus implicaciones prácticas para la operación eficiente del sistema, su valor teórico para avanzar en el campo de la predicción de demanda y su utilidad metodológica para futuros proyectos y aplicaciones en el ámbito del transporte urbano. La combinación de estos factores hace que este proyecto sea de gran importancia y pertinencia en el contexto de la ciudad.

4. MARCO TEÓRICO.

4.1 Estado del arte.

En el artículo escrito por el autor Mauricio Fernández en el que hace una revisión de la implementación del SITP relata la transición del transporte urbano convencional (Fernández, 2018, p 53) popularmente conocido como Transporte Público Colectivo, que tuvo sus inicios en los 60's y que movilizaba aproximadamente 1'629.254 usuarios al día (Historia de TransMilenio, 2013). No obstante y tal como lo expone Fernández, a finales del siglo XX el crecimiento exponencial de la población bogotana, el estilo de vida cambiante, las políticas públicas ambientalmente responsables y la necesidad de transformar el modelo de transporte que no tenía un control por parte del Gobierno Distrital (Fernández, 2018, p 75), dio pauta para el inicio de la estructuración de lo que actualmente se conoce como SITP (Historia de TransMilenio, 2013).

La planificación y gestión del transporte público en áreas urbanas densamente pobladas como lo es Bogotá, supone un desafío latente que requiere soluciones innovadoras y adaptativas las cuales permitan satisfacer las necesidades cambiantes de los usuarios. En este

contexto, este proyecto de investigación busca dar respuesta a una problemática que impacta la eficiencia y la calidad del transporte público en la ciudad: La falta o deficiente previsión de la demanda en el Sistema Integrado de Transporte Público. Las circunstancias de este problema radican en la necesidad de anticipar y responder efectivamente a las fluctuaciones en la demanda de usuarios en el transporte público. Dado que las estimaciones inexactas de la demanda pueden resultar en vehículos con sobrecupo en ciertas franjas pico, lo que esto impacta en la percepción del usuario y la operación eficiente del sistema. Por otro lado, la falta de ocupación de los buses en momentos de baja demanda puede llevar a un uso ineficiente de recursos tales como operadores y vehículos.

En el pasado, los enfoques tradicionales para la predicción de ciertos eventos se basaban en análisis estadísticos y modelos simples que a menudo no lograban capturar la complejidad de los patrones de uso. Tal como lo afirman Rodríguez y Domínguez, estos métodos no tomaban en cuenta la variabilidad en los datos históricos y las condiciones cambiantes que influyen en la demanda, como resultado, las predicciones eran inexactas y las decisiones de programación y asignación de recursos carecían de precisión (Rodríguez-Rueda y Turias, 2017, p. 212).

Es en este contexto que las redes neuronales emergen como una solución prometedora; inspiradas en el funcionamiento del cerebro humano, las redes neuronales artificiales tienen la capacidad de aprender de los datos históricos y adaptarse a las fluctuaciones en la demanda. Estas redes, a través de su estructura interconectada de nodos, pueden identificar patrones complejos y relaciones en los datos, lo que las convierte en herramientas potentes para la predicción de la demanda en el transporte público (¿Qué Es Una Red Neuronal Y cómo Funciona?, 2023). El uso de redes neuronales en la predicción de la demanda de usuarios en el transporte público no es una idea nueva, pero su aplicación en contextos específicos como el planteado en el presente proyecto de investigación tiene un gran potencial para transformar la eficiencia operativa y mejorar la experiencia del usuario. Investigaciones previas como la llevada a cabo por María Garrido, han demostrado la eficacia de las redes neuronales en otras ciudades y sistemas de transporte, donde han logrado mejoras significativas en la precisión de las predicciones (Garrido, 2016, p. 105). Sin embargo, como lo afirma Gonzalo Recio en su blog es importante destacar que el éxito de la implementación de redes neuronales depende en gran

medida de la calidad de los datos disponibles y de la adecuación de los modelos desarrollados (Recio, s.f).

La falta de datos históricos detallados y la disponibilidad limitada de información en tiempo real sobre la cantidad de pasajeros son desafíos que deben abordarse para lograr resultados confiables. Es por esto por lo que la predicción de demanda en el transporte público es un campo multidisciplinario que combina elementos de la ciencia de datos, la estadística, la ingeniería de transporte y la inteligencia artificial (Ackermann y Sellitto, 2022, p. 90). Tradicionalmente, se han utilizado modelos estadísticos, como series temporales y regresión, para abordar este problema, sin embargo, estos enfoques a menudo no logran capturar la complejidad de los patrones de demanda, especialmente en sistemas de transporte público de ciudades grandes y dinámicas como Bogotá.

Aunado a esto se podría mencionar que una tendencia emergente en la predicción de demanda es la aplicación de redes neuronales artificiales, inspiradas en el funcionamiento del cerebro humano, estas redes tienen la capacidad de aprender de datos históricos y adaptarse a las variaciones en la demanda (Torres, Hernández y Pedraza, 2011, p. 95), lo que las hace ideales para abordar la complejidad de los patrones de uso del transporte público. La capacidad de identificar patrones complejos en los datos y ajustarse a las condiciones cambiantes del entorno hace que las redes neuronales sean herramientas valiosas en la predicción de demanda. En proyectos similares en otras ciudades del mundo, se ha demostrado que las redes neuronales pueden generar predicciones más precisas y adaptables en comparación con los enfoques estadísticos tradicionales (Garrido, 2016, p. 1-142).

Durante los últimos años se han realizado varios estudios de casos y proyectos de investigación en diferentes ciudades para aplicar redes neuronales en la predicción de demanda en el transporte público, estos proyectos han demostrado mejoras significativas en la precisión de las predicciones y en la calidad de la planificación operativa.

En el caso de estudio realizado por Angélica Salazar y Mauricio Cabrera, es posible inferir que el modelo propuesto resultó ser una técnica altamente confiable para el pronóstico y factibilidad en la implementación de este en la industria, lo que permitió una distribución más eficiente de los servicios y una reducción de la congestión en horarios pico (Salazar y Cabrera, 2007, p. 6 -12). Del mismo modo, en el estudio de Maria Concepción Garrido se pudo concluir

que las Redes Neuronales Artificiales pueden considerarse como una herramienta apropiada para examinar la calidad del servicio en el transporte público en autobús, los resultados que se obtuvieron fueron coherentes y concordaron con los hallazgos de otros estudios de calidad del servicio en el transporte público que emplearon enfoques distintos (Garrido, 2016, p. 1-142).

A pesar de los avances en otros lugares, la implementación de redes neuronales para la predicción de demanda presenta desafíos específicos como la presencia de eventos especiales en Bogotá que agrega capas adicionales de complejidad y que deben ser consideradas en el desarrollo de las redes neuronales. Además, la demanda de pasajeros en Bogotá puede ser altamente variable debido a factores como el clima, eventos especiales y cambios estacionales, esto hace que la predicción precisa sea un desafío, ya que se requiere un sistema que pueda adaptarse a estas variaciones.

La inversión en tecnología y los recursos financieros es esencial para la implementación de un sistema de predicción de demanda efectivo, garantizar el financiamiento adecuado puede ser un desafío. En este aspecto se puede contemplar todo lo relacionado con la regulación pública, y es que los proyectos de transporte público a menudo están sujetos a regulaciones y políticas gubernamentales, por lo que asegurar que el proyecto cumpla con los requisitos legales y regulatorios es un desafío importante por considerar.

Otro factor por contemplar es la complejidad de la ciudad misma, y es que Bogotá es una ciudad grande y densamente poblada lo que genera una gran cantidad de datos y desafíos logísticos para la implementación de un sistema de predicción de demanda. En la misma línea de análisis se pueden encontrar aspectos como: la forma en que se aborda la gestión del transporte público a través de la predicción de demanda puede requerir un cambio cultural tanto en las instituciones como en la percepción de los usuarios y también es importante educar al público sobre la importancia de este tema y cómo podría mejorar la experiencia de transporte público. Esto puede requerir campañas de concienciación y participación ciudadana lo cual puede llevar demasiado tiempo y esfuerzo.

Por último, un factor de éxito del proyecto es poder garantizar la evaluación continua, dado que después de la implementación es crucial realizar una evaluación del sistema para garantizar su eficacia y realizar ajustes según sea necesario. Todos estos desafíos específicos enmarcan la complejidad y la importancia de un proyecto de investigación como el expuesto en

este documento, por lo que abordar estos desafíos de manera efectiva será esencial para mejorar la calidad del transporte público en la ciudad.

Teniendo en cuenta la situación mencionada anteriormente y el panorama de la ciudad en cuanto a los aspectos importantes a contemplar, conviene destacar que un factor a favor de este proyecto es el crecimiento exponencial de las tecnologías emergentes. Ya que además de las redes neuronales, también se están explorando tecnologías como el aprendizaje profundo o deep learning, el uso de datos de movilidad generados por teléfonos móviles y sistemas de posicionamiento global GPS, para mejorar aún más la precisión de la predicción de demanda en el transporte público. A continuación se resumen los aspectos más importantes de las tecnologías que pueden representar una contribución en el desarrollo e implementación del modelo:

- **Aprendizaje Profundo (Deep Learning):** Esta tecnología es fundamental, ya que las redes neuronales profundas permiten capturar patrones complejos y no lineales en los datos de demanda. Los modelos de aprendizaje profundo, como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), son especialmente útiles para el procesamiento de secuencias de datos temporales, como la demanda de transporte público (Rúe, Roma y Bagén, 2019, p. 19).
- **Computación en la Nube:** El almacenamiento y procesamiento de grandes cantidades de datos requiere una infraestructura de computación escalable. La tecnología en la nube proporciona recursos flexibles y escalables para la gestión de datos y el entrenamiento de modelos de redes neuronales (Radziwill, 2020, p. 29 - 30).
- **Internet de las Cosas (IoT):** Sensores y dispositivos IoT pueden recopilar datos en tiempo real sobre la demanda de pasajeros, condiciones del tráfico y otros factores relevantes para la predicción de demanda. Estos datos en tiempo real pueden enriquecer los modelos de predicción (Radziwill, 2020, p. 37 -38).
- **Big Data y Análisis de Datos Avanzados:** El procesamiento y análisis de grandes conjuntos de datos son esenciales. Herramientas y técnicas de big data, como Hadoop y Spark, junto con algoritmos de análisis avanzado, pueden ayudar a extraer información valiosa de los datos de movilidad (Medina La Plata, 2023, p. 8).
- **Geolocalización y Sistemas de Información Geográfica (SIG):** La información geoespacial es crucial para la predicción de demanda en rutas de transporte público.

Las tecnologías de geolocalización y SIG permiten la representación y análisis espacial de datos, lo que mejora la precisión de las predicciones (ESRI, s.f.).

- Inteligencia Artificial Explicable (XAI): La capacidad de explicar y comprender cómo los modelos de redes neuronales toman decisiones es fundamental para la confianza de los usuarios y las autoridades. Estas tecnologías están emergiendo para hacer que los modelos de IA sean más transparentes y comprensibles (Meseguer y Mántaras, 2017, p. 17).
- Blockchain para la Seguridad de Datos: Para garantizar la integridad y la seguridad de los datos de movilidad, especialmente en lo que respecta a la privacidad de los usuarios, la tecnología blockchain puede utilizarse para la gestión segura y la trazabilidad de los datos (Guardeso, Vico y Encinas, 2019, p. 3).
- Aplicaciones Móviles y Plataformas de Participación Ciudadana: Las aplicaciones móviles y las plataformas de participación ciudadana pueden permitir a los usuarios proporcionar información en tiempo real sobre su experiencia de viaje, lo que enriquece los datos utilizados para la predicción de demanda (Morales, 2019).

Por todo lo anteriormente mencionado es posible afirmar que la combinación de estas tecnologías puede potenciar la precisión y la eficacia de los modelos de predicción de demanda para la ruta T11, lo que podría derivarse en una mejor calidad de servicio para los usuarios y una gestión más eficiente de los recursos. Aunado a esto, y teniendo como sustento los casos de éxito referidos en acápite anteriores, se puede inferir que la implementación de redes neuronales en sistemas de transporte público ha demostrado ser exitosa en la mejora de la calidad del servicio y la satisfacción del usuario, lo que sugiere que su aplicación podría tener un impacto similar.

4.2 Panorama de teorías, modelos y marcos conceptuales.

El proyecto de investigación centrado en la predicción de demanda mediante redes neuronales para la ruta T11 del SITP se basa en un panorama teórico sólido y diversificado. A continuación se referencian algunas teorías, marcos, modelos y enfoques que proporcionarán las bases necesarias para desarrollar un sistema de predicción efectivo y adaptado a las condiciones diversas de la ciudad.

4.2.1 Teorías.

- **Teoría de Redes Neuronales:** La teoría central que sustenta este proyecto se basa en las redes neuronales artificiales, las cuales se inspiran en el funcionamiento del cerebro humano y consisten en nodos interconectados que imitan la capacidad de aprendizaje y adaptación de las neuronas biológicas. La teoría de redes neuronales proporciona la base para desarrollar modelos que pueden aprender de datos históricos y prever patrones de demanda en la ruta seleccionada (Rúe, Roma y Bagén, 2019, p. 20).
- **Teoría del Aprendizaje Automático:** El aprendizaje automático es una rama de la inteligencia artificial que se enfoca en el desarrollo de algoritmos que pueden aprender de los datos y mejorar su rendimiento con la experiencia. Esta teoría es fundamental para entrenar modelos de redes neuronales y ajustarlos continuamente a medida que se recopilan datos en tiempo real (Rúe, Roma y Bagén, 2019, p. 87).
- **Teoría de Series Temporales:** Dado que la demanda de usuarios en el transporte público a menudo exhibe patrones temporales, la teoría de series temporales es relevante. Esta teoría se utiliza para comprender y modelar la variabilidad de la demanda a lo largo del tiempo, lo que contribuye a la precisión de las predicciones (García, 2016, p. 16).

4.2.2 Modelos.

- **Modelos de Redes Neuronales Artificiales:** El enfoque principal es la implementación de modelos de redes neuronales artificiales, que incluyen perceptrones multicapa, redes recurrentes y redes neuronales convolucionales. Estos modelos se utilizan para procesar datos históricos y generar pronósticos de demanda (Rúe, Roma y Bagén, 2019, p. 45).
- **Aprendizaje Supervisado:** La metodología de aprendizaje supervisado se aplica para entrenar los modelos de redes neuronales. Utiliza conjuntos de datos históricos que incluyen información sobre la demanda real y otras variables relevantes, como el clima y los eventos especiales, para que los modelos puedan aprender a hacer predicciones precisas. (Rúe, Roma y Bagén, 2019, p. 61).
- **Preprocesamiento de Datos:** Se emplean técnicas de preprocesamiento de datos para limpiar y transformar los conjuntos de datos en formatos adecuados para el

entrenamiento de las redes neuronales. Esto incluye la normalización de datos, la eliminación de valores atípicos y la selección de características relevantes (Rúe, Roma y Bagén, 2019, p. 30).

4.2.3 Estructuras.

- Python y Bibliotecas de Aprendizaje Automático: El proyecto se basa en el lenguaje de programación Python y utiliza bibliotecas especializadas como TensorFlow, Keras y scikit-learn para la implementación de modelos de redes neuronales y el desarrollo de algoritmos de aprendizaje automático (Mahey, 2020, p. 495).
- Sistemas de Gestión de Datos: Para gestionar y analizar grandes conjuntos de datos históricos y en tiempo real, se recurre a sistemas de gestión de datos robustos y eficientes, como bases de datos relacionales y no relacionales (Salvendy, 2001, p. 115).
- Visualización de Datos: Herramientas de visualización de datos, como Matplotlib y Seaborn, se utilizan para representar gráficamente los resultados de las predicciones y hacer que los hallazgos sean más accesibles y comprensibles (Zhou, 2021, p. 17).

4.2.4 Enfoques de Validación y Evaluación

- Validación Cruzada: Para evaluar la calidad de los modelos de redes neuronales, se emplean técnicas de validación cruzada, como k-fold cross-validation, que dividen los datos en conjuntos de entrenamiento y prueba para medir el rendimiento y prevenir el sobreajuste (Ruggeri, 2007, p. 1720).
- Métricas de Evaluación: Se utilizan métricas de evaluación específicas, como el error cuadrático medio (MSE), el error absoluto medio (MAE) y el coeficiente de determinación (R^2), para cuantificar la precisión de las predicciones (Ruggeri, Ron S Faltin, 2007, p. 1060).

En conjunto, este panorama de teorías, modelos y marcos conceptuales proporciona una base sólida para abordar el desafío de la predicción de demanda que aquí se propone. La combinación de teorías avanzadas, enfoques metodológicos, herramientas técnicas y técnicas de evaluación contribuirá a desarrollar un sistema de predicción de demanda efectivo y adaptado a las necesidades de una ciudad dinámica como lo es Bogotá.

5. MARCO INSTITUCIONAL.

En el contexto de la creciente complejidad de las ciudades y las demandas de movilidad en entornos urbanos, el transporte público desempeña un papel crítico en la vida cotidiana de los ciudadanos. En particular, el SITP de Bogotá, Colombia, es una parte esencial de la infraestructura de transporte que atiende a una población considerable, sin embargo, la eficiencia y la calidad del servicio de transporte público enfrentan desafíos significativos debido a la falta de precisión en la estimación de la demanda de pasajeros, esta falta de precisión ha llevado a ineficiencias operativas y a la insatisfacción de los usuarios.

Este marco institucional se propone establecer las bases para una colaboración efectiva y una gestión coordinada que permita abordar estos desafíos y mejorar la calidad del transporte público en Bogotá. A continuación, se presenta una descripción detallada de este marco institucional:

- Nombre de la Institución: La entidad coordinadora de este proyecto es TRANSMILENIO S.A., que es la autoridad encargada de regular los despachos del sistema de transporte zonal en la ciudad.
- Ubicación: TRANSMILENIO S.A. tiene su sede principal en Bogotá, D.C., Colombia.
- Sector de la Economía (CIIU - Clasificación Industrial Internacional Uniforme): El proyecto se encuentra dentro del sector de servicios, específicamente en el subsector del transporte y la logística.
- Nichos de Mercado: El nicho de mercado principal son los usuarios del transporte público que utilizan la ruta T11 del SITP en Bogotá. Esto incluye a una amplia variedad de residentes de la ciudad, trabajadores, estudiantes y otros ciudadanos que dependen del sistema de transporte público para sus desplazamientos diarios.
- Principales Productos y Procesos: El producto principal del proyecto es un sistema de predicción de demanda en tiempo real para la ruta T11 del SITP. Esto implica la generación de pronósticos de demanda precisos y actualizados que permitan a los operadores del transporte público tomar decisiones informadas sobre la asignación de recursos, como vehículos y personal, para garantizar una operación eficiente. Los procesos clave incluyen la recopilación de datos históricos y en tiempo real, el entrenamiento de modelos de redes neuronales, la implementación de sistemas de

predicción en la operación de la ruta, la evaluación continua del rendimiento y la retroalimentación de los usuarios.

- Estructura Organizacional: La Secretaría Distrital de Movilidad es la entidad encargada de regular el sistema de transporte en Bogotá. Su estructura organizacional incluye diferentes direcciones y áreas dedicadas a la planificación y regulación del transporte público.

Como operador del SITP TRANSMILENIO S.A. es responsable de la operación de la ruta T11, y es una entidad clave en la implementación del sistema de predicción de demanda. Su estructura organizacional comprende departamentos operativos, técnicos y de planificación. Adicionalmente existen diversas universidades locales, que podrían participar en el proyecto a través de sus departamentos de investigación y facultades de ingeniería y ciencias de la computación. Del mismo modo algunas empresas de tecnología se pueden involucrar para con su experiencia contribuir en análisis de datos, inteligencia artificial y sistemas de información geográfica, estas empresas tienen departamentos especializados en desarrollo de software y análisis de datos.

La estructura organizacional del proyecto también contempla la participación ciudadana a través de grupos de usuarios y encuestas que recopilan información sobre las necesidades y expectativas de los usuarios del transporte público. Adicionalmente se puede considerar incluir colaboradores como el Ministerio de Transporte en la definición de políticas y regulaciones relacionadas con el proyecto, organizaciones gubernamentales o de desarrollo que en caso de requerirse puedan proporcionar financiamiento para la investigación y la implementación del sistema de predicción de demanda.

6. METODOLOGÍA.

6.1 Primer nivel.

6.1.1 Enfoque, alcance y diseño de la investigación.

La naturaleza interdisciplinaria del problema planteado en este proyecto de investigación requiere comprensión tanto de los aspectos cuantitativos, relacionados con datos numéricos y estadísticas, como de los aspectos cualitativos, que involucran percepciones, comportamientos y experiencias de los usuarios. Por tal motivo es posible afirmar que el enfoque de investigación es

mixto ya que combina métodos cuantitativos y cualitativos, asimismo se busca abordar de manera integral la complejidad inherente a la predicción de demanda de pasajeros a través de redes neuronales.

Es así como el componente cuantitativo de la investigación se centrará en la recopilación y el análisis de datos históricos de movilidad, incluyendo variables como la hora del día, las condiciones climáticas y los eventos especiales que podrían tener un impacto en la demanda, estos datos servirán para entrenar y validar el modelo de red neuronal permitiendo así la construcción de predicciones cuantitativas precisas (Ron S. Faltin, Ruggeri , 2007 , p. 1231).

Por otro lado, el componente cualitativo se enfocará en la comprensión de las experiencias y las percepciones de los usuarios del transporte público en la ruta T11 específicamente, para lo cual se tendrá en cuenta la cantidad de PQR's que han llegado a la Subgerencia de Comunicaciones y Atención al Usuario asociadas con demoras en el servicio o poca frecuencia de buses, lo anterior dará cuenta de la percepción del servicio, los factores que están afectando en sus decisiones de viaje y cualquier información adicional que resulte relevante. Estos datos cualitativos proporcionarán una comprensión más profunda de los factores humanos y contextuales que pueden no ser evidentes mediante el análisis puramente cuantitativo (O 'Sullivan, 2017, p. 18 - 19).

Por su parte el alcance de investigación de este proyecto se enfoca en proporcionar una solución precisa y personalizada para mejorar la gestión de recursos y la calidad del servicio en la ruta objeto de prueba. El proyecto abarcará la recopilación y el análisis de datos históricos de mencionados anteriormente, mientras que la implementación del modelo neuronal permitirá desarrollar predicciones cuantitativas de la demanda de pasajeros con alta precisión (Benbow, 2021, p. 29). Este alcance también considerará la validación del modelo mediante la comparación de las predicciones con los datos reales del comportamiento de la ruta T11. El objetivo principal es proporcionar una herramienta efectiva para tomar decisiones informadas y mejorar la planificación de recursos para la ruta, lo que a su vez beneficiará a los usuarios al ofrecer un servicio más eficiente y confiable.

El diseño de investigación se fundamenta en la recopilación y el análisis de datos existentes sin intervenir ni modificar deliberadamente ninguna variable, con el propósito de describir y comprender la situación actual de la demanda de pasajeros en la ruta T11, y

posteriormente, utilizar esta información para desarrollar el modelo neuronal. A continuación se detallan el diseño de investigación propuesto para este proyecto:

- En primer lugar, el diseño no experimental se caracteriza por no involucrar la manipulación de variables independientes ni la aplicación de tratamientos o intervenciones controladas, en cambio, se basa en la recopilación y el análisis de datos observacionales existentes. En este proyecto, se recogerán datos históricos de la demanda de pasajeros, como registros de viajes, datos de uso de la ruta T11 y factores contextuales, sin alterar las condiciones naturales del servicio de transporte público.
- En segundo lugar, el diseño transversal implica la recolección de datos en un solo momento en el tiempo, permitiendo una instantánea de la situación actual, en este proyecto, se recopilarán datos actuales y pasados sobre la demanda de pasajeros en la ruta T11. Esto proporcionará una visión general de la demanda en un momento específico, lo que es fundamental para la etapa descriptiva de la investigación.
- En tercer lugar, el diseño descriptivo tiene como objetivo principal describir y caracterizar una situación o fenómeno específico, para efectos de este proyecto de investigación, se utilizará para analizar detalladamente la demanda de pasajeros en la ruta, identificando patrones, tendencias y relaciones entre variables que puedan influir en la demanda.

6.1.2 Definición de variables.

A continuación, se describen las variables clave a ser consideradas en este proyecto:

Tabla 1

Definición de variables.

Variable	Definición conceptual	Definición operacional	Dimensiones
Demanda de pasajeros	Esta variable representa el número de personas que utilizan la ruta en un período de tiempo específico. La demanda puede variar según la hora del día, el día de la semana y otros factores, y es la variable central que se busca predecir.	Se define como el número de pasajeros que abordan un autobús en la ruta T11 en un período de tiempo específico, generalmente se mide en pasajeros por hora (pph) o pasajeros por día (ppd). Esta variable es unidimensional y	- Unidades de Medida: Pasajeros por día (ppd). - Cálculo: Registrar el número total de pasajeros que abordan la ruta en un día determinado.

Variable	Definición conceptual	Definición operacional	Dimensiones
		cuantitativa.	
Hora del día	La hora del día es una variable temporal esencial, ya que influye significativamente en la demanda de pasajeros. Las horas pico y las horas no pico pueden mostrar patrones de demanda diferentes, y se deben considerar para ajustar las predicciones.	La hora del día se mide en horas (24 horas), y se considera como una variable temporal. Se divide en 3 categorías, como "pico AM" (5:00 a.m. - 8:00 a.m.), "pico PM" (5:00 p.m. - 7:00 p.m.), y "valle" (las demás horas que no entran en las 2 categorías mencionadas). Esta variable es unidimensional y cualitativa.	- Unidades de Medida: Hora (en formato de 24 horas). - Cálculo: Registrar la hora exacta en que se realiza el conteo de pasajeros en cada intervalo de tiempo.
Día de la semana	El día de la semana también tiene un impacto en la demanda. Los patrones de viaje suelen variar entre los días laborables y los fines de semana, lo que debe ser capturado en el modelo de predicción.	El día de la semana se mide en 7 categorías: lunes, martes, miércoles, jueves, viernes, sábado y domingo. Es una variable categórica que captura la dimensión temporal y cualitativa.	- Unidades de Medida: Día de la semana. - Cálculo: Registrar el día de la semana en que se realiza el conteo de pasajeros.
Condiciones climáticas	Las condiciones climáticas, como la lluvia o el calor extremo, pueden afectar la demanda de pasajeros en el transporte público. Esta variable ayuda a considerar la influencia del clima en la movilidad.	Las condiciones climáticas se miden cuantitativamente.	- Unidades de Medida: Variables climáticas (por ejemplo, temperatura en grados Celsius, precipitación en milímetros). - Cálculo: Registrar la temperatura y la precipitación, así como otras condiciones climáticas relevantes en el momento del conteo.
Eventos especiales y días festivos	Eventos como festivales, manifestaciones o eventos deportivos pueden generar	Esta variable se mide de manera cualitativa y categórica, registrando la	- Unidades de Medida: Tipo de día festivo. Tipo de evento (por ejemplo,

Variable	Definición conceptual	Definición operacional	Dimensiones
	aumentos inesperados en la demanda. Por otro lado, los días festivos suelen mostrar patrones de demanda diferentes a los días laborables normales.	presencia o ausencia de eventos especiales en un día específico. Puede ser binaria (0 para ausencia, 1 para presencia). En el caso de los festivos, se indica si un día es festivo o no.	concierto, desfile, manifestación). - Cálculo: Registrar la fecha y el tipo de día. Registrar la fecha y el tipo de evento especial.
Oferta de servicio	La oferta de servicio, que incluye la frecuencia, la ampliación de la cobertura y la capacidad de los autobuses en la ruta, también son aspectos clave.	La oferta de servicio se mide en términos de la frecuencia de los autobuses y la capacidad de los vehículos. Estas variables son cuantitativas y unidimensionales.	- Unidades de Medida: Número de autobuses por hora y capacidad de pasajeros por autobús. - Cálculo: Recopilar información sobre la frecuencia de los autobuses y la capacidad de pasajeros de la ruta a partir de datos recopilados.
Ubicación de las paradas	La ubicación de paradas de la ruta, la proximidad a los destinos populares y las rutas de conexión con otros medios de transporte pueden afectar la demanda en diferentes áreas del trazado de la ruta.	Las ubicaciones de las paradas se registran con coordenadas geográficas (latitud y longitud). Las ubicaciones son bidimensionales.	- Unidades de Medida: Latitud y longitud. - Cálculo: Registrar las coordenadas geográficas de las paradas y rutas de conexión de la ruta T11 mediante dispositivos de geolocalización.
Demografía y población	Las características demográficas de las áreas servidas por la ruta, como la densidad de población y la composición demográfica, también pueden influir en la demanda de pasajeros.	Las características demográficas, como la densidad de población, se miden en términos de personas por kilómetro cuadrado (habitantes/km ²) o porcentajes de diferentes grupos de edad. Estas variables son cuantitativas y unidimensionales.	- Unidades de Medida: densidad de población por kilómetro cuadrado, edad promedio. - Cálculo: Obtener datos demográficos de fuentes como el censo nacional y registrar los valores correspondientes para la zona de estudio.

Variable	Definición conceptual	Definición operacional	Dimensiones
Opiniones y experiencias de los usuarios	Las opiniones y experiencias de los usuarios en la ruta analizada proporcionan información cualitativa que complementa los datos cuantitativos y están relacionadas con las PQR's radicadas por los usuarios.	Corresponde a la cantidad de PQR's radicadas por los usuarios que capturan las opiniones y experiencias de estos.	<ul style="list-style-type: none"> - Unidades de Medida: PQR's /viaje realizado. - Cálculo: categorizar las PQR's en temas como: comodidad, puntualidad y frecuencias.

Nota: Esta tabla contiene información relacionada con la definición conceptual y operacional de las variables a contemplar en el proyecto de investigación.

6.1.3 Población y Muestra.

Las características de la población objetivo son importantes para comprender y abordar adecuadamente las necesidades de los usuarios. Principalmente la población está compuesta por los usuarios del transporte público que utilizan la ruta T11 en Bogotá, esto incluye a personas de diversas edades, géneros y trasfondos socioeconómicos que dependen del transporte público para sus desplazamientos diarios. Adicionalmente, se debe incluir a pasajeros frecuentes que utilizan la ruta a diario como a aquellos que lo hacen ocasionalmente, la frecuencia de uso del servicio puede variar significativamente entre los usuarios.

Otro aspecto importante son las opiniones y experiencias de los usuarios con respecto al servicio de la ruta y son una parte importante de la población objetivo, ya que estas percepciones pueden influir en la elección de transporte y en la satisfacción general de los usuarios. Los usuarios tienen necesidades y expectativas específicas en cuanto a la puntualidad, la comodidad, la accesibilidad y otros aspectos del servicio.

Ahora bien, es oportuno mencionar que para efectos del desarrollo del proyecto se contemplará una muestra de 728.960 usuarios/mes, esta cifra corresponde a la cantidad de usuarios que se movilizaron en el sistema durante julio y septiembre de 2023 los cuales hacen uso de la ruta en diferentes tipo día y franjas horarias (TRANSMILENIO S.A., 2023). Para esta población se utilizará el método de muestreo probabilístico, específicamente el muestreo aleatorio estratificado, que permite dividir la población en subgrupos homogéneos o estratos y de esta maneta tomar muestras aleatorias de cada estrato.

Este tipo de muestreo es eficiente y permite obtener resultados representativos con un tamaño de muestra más manejable, además al ser una población tan heterogénea y diversa en

términos de características demográficas, ubicación geográfica, frecuencia de uso y otros factores, el tener la población dividida en estratos permite capturar esta diversidad y garantizar que cada subgrupo esté adecuadamente representado en la muestra (Goble, 2010, p. 414). Otros aspectos que se consideraron para elegir este muestreo fue la precisión en la estimación, la eficiencia en la recopilación de datos y la representatividad de la muestra, este último evita sesgos y permite generalizar los resultados a toda la población de usuarios de la ruta de manera más precisa.

Para precisar el tamaño de la muestra fue necesario identificar y definir los estratos dentro de la población, en este caso se tomaron divisiones de frecuencia de viaje, tipo día, franja horaria, días laborables y fines de semana. Una vez definida la cantidad de estratos, se decidió tomar el 80 % de las muestras para las franjas pico AM y PM y el 20 % para el restante de las franjas, esto se soporta en que durante estas franjas se concentra la mayor cantidad de validaciones para la ruta, lo cual se traduce en un total de 728.960 usuarios. Como tercer aspecto, fue necesario tener en cuenta la variabilidad entre los estratos, para así asegurar que los estratos con mayor variabilidad tienen un peso adecuado en la muestra final, por lo cual se definió una desviación estándar de 1 contemplando también el factor de rotación de usuarios, que se define como la cantidad de tiempo que un usuario permanece dentro del bus durante el recorrido de la ruta.

6.2 Segundo nivel.

6.2.1 Selección de métodos o instrumentos para recolección de información.

El modelo seleccionado para recolectar la información mediante una arquitectura de red neuronal artificial (RNA) secuencial, construida con la biblioteca TensorFlow y Keras en Python, esta consiste en la creación de tres capas densas, o conectadas por completo. La primera capa tiene 64 neuronas con activación ReLU o Unidad Lineal Rectificada, seguida por una capa intermedia de 32 neuronas también con activación ReLU, y finalmente, una capa de salida con una sola neurona para predecir la demanda de pasajeros (TechEdu, 2023).

En la etapa de preprocesamiento de datos es necesario que antes de ingresar los datos al modelo, se estandaricen las características, esto se refiere a que las variables se escalan para que tengan una media de 0 y una desviación estándar de 1, este preprocesamiento ayuda a que la red

neuronal trabaje de manera más eficiente y a mejorar la convergencia del modelo (Wang, Yimin, Q. P. Al-Hussein, 2018, p. 248). En este punto la parte fundamental es el entrenamiento y evaluación del modelo, el cual se entrena utilizando el algoritmo de descenso de gradiente adaptativo Adam o Adaptive Moment Estimation (Big Data, 2021) y una función de pérdida de error cuadrático medio MSE como función de pérdida. Se realiza también un entrenamiento durante 50 épocas, con un tamaño de batch size o lote de 32, el modelo se evalúa utilizando datos de validación y datos de prueba para medir su rendimiento.

Otro aspecto de gran relevancia es la eficiencia en la predicción y está relacionada directamente con su capacidad para generalizar a partir de los datos de entrenamiento y adaptarse a patrones cambiantes en la demanda de pasajeros. Además, la inclusión de capas ocultas y activaciones ReLU le permite aprender características latentes y adaptarse a diferentes condiciones, como cambios en las horas pico o eventos especiales que afecten la demanda.

Este modelo tiene varios componentes y elementos funcionales esenciales para su operación eficaz, los cuales se detallan a continuación:

- **Capas Neuronales:** El modelo consta de múltiples capas neuronales que realizan operaciones matemáticas para aprender patrones a partir de los datos de entrada. En este proyecto, se han utilizado capas densas o completamente conectadas para procesar la información. (Salvendy, 2001, p.1779).
- **Funciones de Activación:** Cada capa neuronal utiliza funciones de activación, como ReLU, para introducir no linealidad en el modelo y permitir la representación de relaciones complejas en los datos (TechEdu, 2023).
- **Función de Pérdida:** El modelo utiliza una función de pérdida de mean squared error o error cuadrático medio, para medir la diferencia entre las predicciones del modelo y los valores reales de demanda de pasajeros (Ruggeri, Ron S Faltin, 2007, p. 1060).

Los elementos funcionales del modelo inician con la carga de datos de entrenamiento y prueba, que se dividen en características (X) y variable objetivo (Y), posterior a esto se realiza el preprocesamiento de datos, que incluye la normalización de características y la división del conjunto de datos en entrenamiento y validación. Luego se define la arquitectura de la red neuronal utilizando la biblioteca TensorFlow y Keras, especificando las capas neuronales,

funciones de activación y la estructura de entrada y salida. En esta parte el modelo se compila con un optimizador, para efectos del proyecto se utiliza 'adam', y la función de pérdida. Luego, se entrena utilizando datos de entrenamiento y durante el entrenamiento, el modelo ajusta sus pesos para minimizar la pérdida. Una vez entrenado, el modelo se evalúa utilizando datos de prueba, y se calcula la pérdida en datos de prueba para evaluar su rendimiento. Por último, se utilizan las características de prueba para realizar predicciones de demanda de pasajeros.

Los componentes del modelo, como las capas neuronales y las funciones de activación, trabajan en conjunto para aprender patrones en los datos de entrenamiento. Durante el entrenamiento, el modelo ajusta sus parámetros para minimizar la pérdida, lo que implica encontrar relaciones óptimas entre las características y la variable objetivo. Una vez entrenado, el modelo puede realizar predicciones precisas utilizando datos de prueba no vistos. El modelo finalmente proporciona resultados en forma de predicciones de demanda de pasajeros en la ruta T11.

6.2.1 Técnicas de análisis de datos.

En el modelo neuronal de predicción de demanda de usuarios para la ruta T11, se requiere de instrumentos de recolección de datos, técnicas de análisis de información y un proceso de desarrollo y evaluación del modelo. La siguiente tabla detalla cada uno de estos aspectos:

Tabla 2

Técnicas de análisis de datos.

Instrumento de recolección de datos	Técnica de análisis de datos	Descripción
<p>Datos de Transacciones Electrónicas: En el contexto actual, se hace uso de tarjetas electrónicas para pagar el pasaje. Estas tarjetas generan datos de transacciones que incluyen información sobre la hora, la ubicación y la frecuencia de uso.</p>	<p>Análisis Exploratorio de Datos (EDA): Este análisis implica la exploración inicial de los datos para identificar patrones, tendencias y anomalías. Se utilizan herramientas gráficas y estadísticas descriptivas para visualizar y resumir los datos. Esto ayuda a comprender la distribución de la demanda, la variabilidad temporal y</p>	<p>Recolección de Datos: Se recopilan datos de múltiples fuentes, como datos de transacciones electrónicas, sensores en vehículos y datos climáticos. Estos datos se almacenan en una base de datos centralizada para su posterior procesamiento.</p>

Instrumento de recolección de datos	Técnica de análisis de datos	Descripción
<p>Estos datos son valiosos para comprender los patrones de demanda en diferentes momentos del día y en ubicaciones específicas.</p>	<p>geográfica, y las relaciones entre las variables.</p>	
<p>Sensores en Vehículos: Los vehículos del SITP están equipados con sensores que registran la cantidad de pasajeros que abordan en cada parada. Estos datos de los sensores se utilizan para determinar la capacidad y la ocupación de los vehículos en diferentes tramos del recorrido.</p>	<p>Preprocesamiento de Datos: Antes de utilizar los datos para entrenar el modelo de predicción, es necesario realizar tareas de preprocesamiento. Esto incluye la limpieza de datos para eliminar valores atípicos o faltantes, la normalización de características.</p>	<p>Exploración y Preprocesamiento: Se realiza un análisis exploratorio de datos para comprender la naturaleza de los datos y detectar posibles problemas. Luego, se preprocesan los datos, lo que incluye la limpieza, normalización y codificación de variables.</p>
<p>Datos Climáticos y de Eventos: Los factores externos, como las condiciones climáticas y los eventos especiales, pueden influir en la demanda de pasajeros.</p>	<p>Modelado de Predicción: La técnica central en la predicción de demanda es el modelado de redes neuronales. Se ajustan hiperparámetros como el número de capas, el tamaño de las capas, las funciones de activación y la tasa de aprendizaje para optimizar el modelo.</p>	<p>Modelado de Predicción: Se construye el modelo de redes neuronales utilizando TensorFlow y Keras. Se seleccionan las características relevantes y se divide el conjunto de datos en entrenamiento, validación y prueba. El modelo se entrena utilizando datos de entrenamiento y se ajustan hiperparámetros para obtener el mejor rendimiento.</p>
<p>Datos de Operación del Transporte: Los datos operativos del sistema de transporte público son esenciales para la predicción de demanda. Esto incluye información sobre la programación de rutas, horarios de salida y llegada, cambios en el servicio y otras variables</p>	<p>Validación y Evaluación del Modelo: Una vez que se ha entrenado el modelo, se evalúa su rendimiento utilizando datos de validación y prueba. Las métricas de evaluación, como el error cuadrático medio (MSE) o el coeficiente de determinación (R2), se utilizan para medir la precisión del modelo en la predicción de la demanda. Se ajusta el modelo según sea necesario para mejorar su rendimiento.</p>	<p>Validación y Evaluación: Se evalúa el modelo utilizando datos de validación y prueba. Se calculan métricas de evaluación para medir la precisión del modelo en la predicción de la demanda. Se realiza una validación cruzada si es necesario.</p>
	<p>Optimización de Hiperparámetros: La optimización de hiperparámetros</p>	<p>Optimización: Se optimizan los hiperparámetros y se ajusta el</p>

Instrumento de recolección de datos	Técnica de análisis de datos	Descripción
operativas que pueden afectar la demanda.	implica encontrar la configuración óptima de los parámetros del modelo para lograr la mejor precisión en la predicción. (Aghajan, Ramón López-Cózar, 2010, p. 379).	modelo según sea necesario para mejorar la precisión. Esto se hace mediante técnicas como búsqueda en cuadrícula o búsqueda aleatoria de hiperparámetros.
	<i>Despliegue y monitoreo continuo:</i> Una vez que el modelo ha sido validado y optimizado, se despliega en un entorno de producción. Esto implica la integración del modelo en sistemas de información en tiempo real que puedan utilizar las predicciones para mejorar la planificación y la operación del servicio de transporte público.	<i>Despliegue y monitoreo:</i> El modelo se despliega en un entorno de producción, donde se utiliza para realizar predicciones en tiempo real. Se establece un sistema de monitoreo continuo para supervisar su rendimiento y realizar actualizaciones si es necesario.

Nota: Esta tabla contiene el detalle de las técnicas de análisis de datos.

7. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS.

Los resultados obtenidos durante la evaluación del proyecto de investigación se fundamentan en investigaciones académicas relevantes en el campo de la predicción de sucesos en el transporte público. Siguiendo los principios establecidos por Chen y Song (Chen, Song, 2019, p. 25) sobre la importancia de la predicción en la mejora de la eficiencia operativa, esta investigación se alinea con las recomendaciones de Ong y Thakuriah (Ong, Thakuriah, 2015, p. 87) en la aplicación de redes neuronales para la mejora de la precisión en la predicción de la demanda de pasajeros. Asimismo, la evaluación del impacto de la implementación del modelo se basa en los hallazgos de Ma y Wang (Ma, Wang, 2014, p. 109), quienes exploraron la aplicación de algoritmos de una colonia de abejas artificiales en la predicción de tiempos de llegada de autobuses. Esta sección de análisis y discusión se adentra en la convergencia de estos enfoques, destacando la pertinencia del modelo en el contexto de las mejores prácticas y avances actuales en la predicción de la demanda en el transporte público para el SITP en Bogotá, para lo cual se expondrán los resultados obtenidos durante el proceso de simulación dando respuesta a los objetivos específicos propuestos en acápites anteriores.

En primer lugar fue necesario recopilar la información necesaria relacionada con la demanda de usuarios para la ruta T11 durante el mes de septiembre del año 2023, así como también las PQR's radicadas por los usuarios durante este periodo de tiempo. Lo anteriormente descrito con el propósito de evaluar los datos más recientes disponibles, del mismo modo se recopiló información histórica desde el año 2018 para el mismo mes, esto para poder afinar la precisión de la información con la cual se va a alimentar el modelo que se explicará más adelante. A continuación se presenta la información mencionada resumida en tablas de datos:

Tabla 3

Datos históricos demanda de usuarios ruta T11.

Periodo	Día de la semana	Franja horaria			Total demanda
		Pico a.m.	Pico p.m.	Valle	
Septiembre 2018	Lunes	22.782	13.714	37.884	74.380
	Martes	57.956	31.556	87.686	177.198
	Miércoles	54.662	31.212	81.094	166.968
	Jueves	45.494	24.134	69.396	139.024
	Viernes	35.962	19.916	57.160	113.038
	Sábado	20.962	9.638	40.134	70.734
	Domingo	6.432	6.250	19.422	32.104
Septiembre 2019	Lunes	38.318	20.464	55.314	114.096
	Martes	37.898	19.440	53.250	110.588
	Miércoles	36.282	19.646	53.426	109.354
	Jueves	29.318	15.862	45.148	90.328
	Viernes	36.924	18.522	53.892	109.338
	Sábado	26.014	11.508	49.350	86.872
	Domingo	8.360	8.672	24.322	41.354
Septiembre 2020	Lunes	34.057	19.394	52.346	105.797
	Martes	49.662	26.639	73.445	149.746
	Miércoles	44.382	24.427	65.472	134.281
	Jueves	41.803	21.803	62.594	126.200
	Viernes	41.011	22.092	63.533	126.636
	Sábado	23.725	10.009	44.589	78.323
	Domingo	7.091	7.198	21.767	36.056
Septiembre 2021	Lunes	22.782	13.714	37.884	74.380
	Martes	57.956	31.556	87.686	177.198

Periodo	Día de la semana	Franja horaria			Total demanda	
		Pico a.m.	Pico p.m.	Valle		
Septiembre 2022	Miércoles	54.662	31.212	81.094	166.968	
	Jueves	45.494	23.314	69.396	138.204	
	Viernes	35.962	19.916	57.160	113.038	
	Sábado	20.962	9.638	40.134	70.734	
	Domingo	6.432	6.250	19.422	32.104	
	Lunes	38.302	20.914	57.362	116.578	
	Martes	39.596	21.886	59.190	120.672	
	Miércoles	38.716	20.338	56.970	116.024	
	Jueves	40.384	21.712	60.310	122.406	
	Viernes	48.188	26.404	75.798	150.390	
	Sábado	27.622	11.590	52.662	91.874	
	Domingo	6.594	7.174	19.612	33.380	
	Septiembre 2023	Lunes	41.086	23.554	61.792	126.432
		Martes	51.434	26.476	73.458	151.368
Miércoles		39.768	21.732	58.352	119.852	
Jueves		39.532	20.382	58.076	117.990	
Viernes		38.884	19.956	57.640	116.480	
Sábado		22.590	8.800	40.970	72.360	
Domingo		8.246	8.170	26.268	42.684	

Nota: Esta tabla contiene el detalle de la demanda de usuarios en la ruta T11 por franja horaria y día de la semana.

El análisis detallado de los patrones de demanda evidenciados en la tabla anterior revela una serie de tendencias distintivas que inciden en la utilización de la ruta a lo largo del tiempo. Como primer hallazgo se encontró que los días laborables presentan patrones de demanda recurrentes, con dos picos significativos durante las horas pico de la mañana y la tarde, estos momentos están estrechamente relacionados con los desplazamientos diarios de los trabajadores y los estudiantes, generando una demanda más alta en los horarios de entrada y salida de labores. Por otro lado, los fines de semana exhiben una distribución más uniforme de la demanda a lo largo del día, con menor énfasis en las horas pico.

En lo que respecta a los picos de demanda, fue posible determinar que estos suelen estar vinculados a eventos específicos, ya sean festividades locales, eventos culturales, y actividades recreativas que pudieron generar aumentos inesperados en la demanda. Además, eventos

deportivos y conciertos generaron una afluencia masiva de pasajeros, especialmente antes y después de dichos eventos. Contrario a esto, el comportamiento es adverso en días festivos, periodos vacacionales significativos o condiciones climáticas complejas, ya que durante estas eventualidades la demanda tiende a disminuir, y en períodos vacacionales, se experimenta una disminución más prolongada. Además, condiciones climáticas complejas, como lluvias fuertes, reducen la movilidad y generan valles en la demanda durante ciertos periodos del año.

Por su parte, el análisis a lo largo del tiempo revela comportamientos interesantes, en lo que respecta al crecimiento poblacional y los cambios en la infraestructura urbana es claro que contribuyeron a un aumento gradual en la demanda. Por otro lado, las modificaciones en las rutas de transporte público, la introducción de nuevas alternativas de movilidad, o cambios en las preferencias de los usuarios pudieron resultar en fluctuaciones y tendencias descendentes.

Un eventos excepcional como la pandemia de COVID-19, mostró un impacto significativo en la demanda, los confinamientos y restricciones de movilidad generaron una disminución drástica en la demanda diaria, evidenciando la sensibilidad de la utilización de la ruta a factores externos. En definitiva, en el contexto preliminar la identificación de estos patrones de demanda, picos y valles, así como la detección de tendencias a lo largo del tiempo, permite una comprensión holística del comportamiento de la ruta. Este análisis es fundamental para el desarrollo de modelos de predicción precisos y estrategias efectivas de gestión del transporte público en la ciudad; la capacidad de anticipar estas variaciones permite una planificación más eficiente y una respuesta proactiva a las necesidades cambiantes de los usuarios.

Dada la imperativa necesidad del caso de estudio, se propone un modelo de red neuronal para la predicción de la demanda de pasajeros utilizando TensorFlow y Keras. En este punto conviene precisar que TensorFlow es una biblioteca de código abierto que permite realizar operaciones matemáticas de manera eficiente, especialmente en tareas relacionadas con el Deep Learning y con frecuencia se utiliza para entrenar modelos de redes neuronales (Abadi, Agarwal, Barham, Brevdo, Chen, (...) & Zheng, 2016); mientras que Keras es una interfaz que se ejecuta sobre TensorFlow y proporciona una forma más fácil y rápida de construir, entrenar y evaluar modelos de Deep Learning (Chollet, 2015).

En cuanto al desarrollo del código a continuación se detalla cual fue el procedimiento para construir el modelo. Como primer paso se deben importar las bibliotecas necesarias, como NumPy para operaciones numéricas, Pandas para manejar datos tabulares, TensorFlow y Keras para construir y entrenar la red neuronal, y herramientas de preprocesamiento de datos. Luego, en la carga de datos se cargan los conjuntos de datos de entrenamiento y prueba desde archivos CSV utilizando Pandas. En el siguiente paso de preprocesamiento de datos, se separan las características (X) y la variable objetivo (y), posterior a esto se divide el conjunto de entrenamiento en conjuntos de entrenamiento y validación, y finalmente se normalizan las características utilizando “StandardScaler”. En el siguiente paso que es la definición del modelo de red neuronal, se define un modelo secuencial de Keras con tres capas densas las cuales deben estar totalmente conectadas y cumplen con las siguientes características:

- La primera capa tiene 64 nodos, función de activación ReLU y la forma de entrada correspondiente al número de características en los datos.
- La segunda capa tiene 32 nodos con activación ReLU.
- La capa de salida tiene 1 nodo para la predicción.

En la compilación del modelo se compila el modelo especificando el optimizador “Adam” y la función de pérdida “mean_squared_error”. Seguido de esto se entrena el modelo utilizando los datos de entrenamiento escalados y se especifica el número de épocas o periodos, el tamaño del lote (1’000.000) y se utiliza el conjunto de validación durante el entrenamiento para evaluar el rendimiento. Por último se evalúa el rendimiento del modelo utilizando los datos de prueba escalados, se imprime la pérdida del modelo en los datos de prueba, se realizan predicciones en los datos de prueba escalados y se guarda el modelo entrenado en un archivo HDF5.

Luego de ejecutar el código expuesto anteriormente, se obtuvo una proyección de usuarios para septiembre de 2024 en donde se pudo evidenciar que el modelo predice en un 95 % la ocurrencia de los eventos descritos en apartados anteriores.

Tabla 4

Datos de la proyección de demanda de usuarios ruta T11.

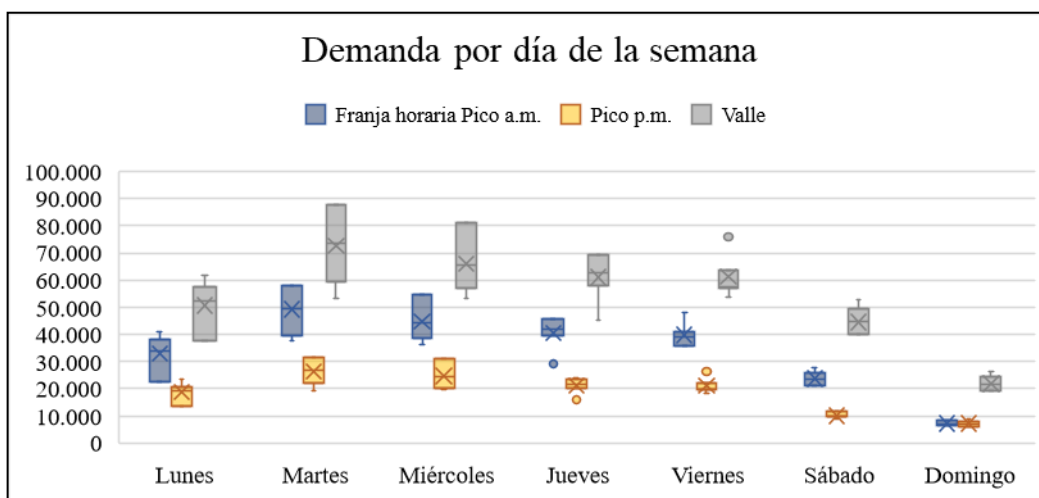
Periodo	Día de la semana	Franja horaria			Total demanda
		Pico a.m.	Pico p.m.	Valle	
Septiembre 2024	Lunes	34.057	19.394	52.346	105.797
	Martes	49.662	26.639	73.445	149.746
	Miércoles	44.382	24.427	65.472	134.281
	Jueves	41.803	21.803	62.594	126.200
	Viernes	41.011	22.092	63.533	126.636
	Sábado	23.725	10.009	44.589	78.323
	Domingo	7.091	7.198	21.767	36.056

Nota: Esta tabla contiene el detalle de la demanda de usuarios en la ruta T11 por franja horaria y día de la semana para septiembre de 2024.

Dados los resultados proyectados se pudo evidenciar también un crecimiento de la demanda del 1,32 % que se puede percibir en 9.873 usuarios potenciales más que podrían acceder al servicio, esta proyección da cuenta de la necesidad de los usuarios por el servicio prestado por la ruta, adicionalmente, se evidencia que los incrementos en la cantidad de usuarios serán para la franja horaria valle y en un 1,93 % en la franja pico p.m. lo cual representa un reto en la adecuada utilización y programación de recursos durante estas franjas horarias.

Gráfica 1

Datos de la demanda de usuarios ruta T11.



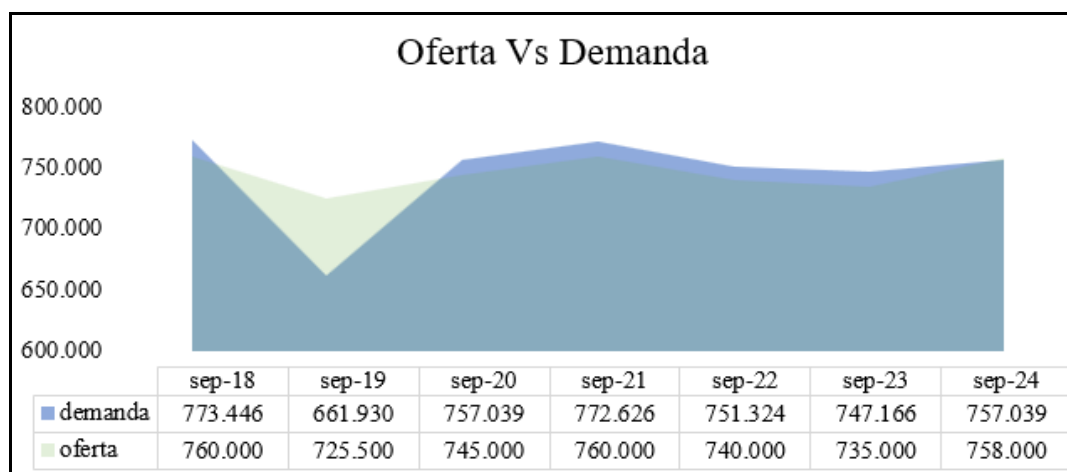
Nota: Esta gráfica contiene el comportamiento de la demanda de usuarios en la ruta T11 por franja horaria.

Los resultados representados en el boxplot anterior, muestran que el martes y miércoles son los días con mayor participación mientras que el viernes y jueves muestran una concentración de usuarios inferior. Adicionalmente es claro que la media de los datos se ubica en el caso de la franja pico a.m. en las 38.310 validaciones, en la pico p.m. el valor corresponde a 20.147 y en valle se registran 57.160 visualmente es posible decir que la mediana se encuentra en la mayoría de los casos por debajo o por encima de la media lo que indica que los datos son asimétricos. En el caso del rango intercuartílico, se evidencia que días como lunes, martes y miércoles presentan la mayor dispersión y variabilidad de los datos registrando valores por encima de las 20.000 validaciones de usuario. Por su parte los bigotes muestran que de manera general los datos atípicos son pocos y corresponden al 2,5 % del total de datos que se desestimaron en el modelo de red neuronal, sin embargo para efectos del análisis preliminar se conservaron.

De manera general se percibe un déficit de recursos que imposibilitan que periodo tras periodo se puede cubrir en su totalidad la demanda de usuarios, exceptuando el caso del año 2019 dado que por causa de la pandemia se presentó una sobre oferta de plazas. Para septiembre de 2024 se proyecta que la oferta de plazas sea de 23.000 sillas con un incremento del 3,13 % comparado con el periodo anterior, esto representa una inversión en recursos ya sean en programación, recurso humano o vehículos considerable para la situación actual del sistema.

Gráfica 2

Datos de la oferta y demanda de usuarios ruta T11.



Nota: Esta gráfica contiene el comportamiento de la oferta y demanda de usuarios en la ruta T11.

Continuando con el análisis estadístico de los resultados se consideraron medidas de tendencia central como la media, medidas de dispersión como la desviación estándar, correlación y algunas métricas de evaluación adicionales como el Coeficiente de Correlación de Pearson y el Coeficiente de Determinación R^2 , obteniendo los siguientes resultados:

Tabla 5

Datos del análisis estadístico de la demanda de usuarios para la ruta T11.

Métrica	Valor	Interpretación
Media de demanda real	289,73	La media de la demanda real representa el valor promedio de usuarios que abordan el autobús por hora.
Media de demanda predicha	292,19	La media de la demanda predicha muestra el valor promedio previsto por el modelo para la demanda de usuarios.
Desviación estándar de demanda real	111,12	La desviación estándar de la demanda real mide la variabilidad de la demanda real alrededor de la media.
Desviación estándar de demanda predicha	109,42	La desviación estándar de la demanda predicha mide la variabilidad de las predicciones alrededor de la media.
Coeficiente de correlación	0,92	El coeficiente de correlación indica una fuerte correlación positiva entre la demanda real y la predicha.
Error Absoluto Medio (MAE)	37,18	El MAE promedio sugiere una precisión razonable, con desviaciones de alrededor de 37 usuarios en promedio.
Error Cuadrático Medio (MSE)	2.035,63	El MSE indica variabilidad y algunos errores más grandes en las predicciones, siendo sensible a valores atípicos.
Coeficiente de determinación (R^2)	0,83	Un R^2 de 0,83 sugiere que el modelo explica el 83% de la variabilidad en la demanda de usuarios.

Nota: Esta tabla contiene el detalle del análisis estadístico de la demanda de usuarios en la ruta T11.

El análisis estadístico revela que el modelo tiene una capacidad razonable para seguir la tendencia de la demanda real, esto soportado en la fuerte correlación (0,92) y el alto R^2 (0,83) que indican una relación significativa y una buena capacidad explicativa del modelo, aunque es un valor sólido, queda espacio para mejorar y abordar la variabilidad restante. No obstante, la presencia de variabilidades y errores más grandes, evidenciados por el MSE y algunas desviaciones en el MAE, resalta áreas de mejora en el modelo en general.

De acuerdo con los resultados presentados en este apartado, se enmarca la importancia de implementar un modelo como el aquí presentado, ya que no solo representa una mejora

sustancial en la asignación eficiente de recursos, sino que también impacta positivamente diversos aspectos operativos y de servicio, al anticipar la demanda, se optimiza la planificación de flota, evitando la subutilización y la saturación de vehículos en momentos específicos. Esta eficiencia se traduce directamente en una reducción significativa de los tiempos de espera para los usuarios, mejorando su satisfacción y la confiabilidad en el servicio. Además, la comprensión detallada de los patrones de demanda a lo largo del tiempo no solo optimiza las rutas y reduce los costos operativos, sino que también brinda la flexibilidad necesaria para adaptarse a cambios en la demanda, garantizando una resiliencia óptima ante eventos excepcionales. Este enfoque no solo contribuye a la eficiencia operativa a corto plazo sino que también proporciona información clave para la toma de decisiones estratégicas a largo plazo, promoviendo la sostenibilidad al reducir emisiones y congestión.

8. CONCLUSIONES.

- En conclusión, la investigación sobre la predicción de demanda para la ruta T11 mediante redes neuronales ha proporcionado una visión integral y esclarecedora de los patrones de movilidad en el sistema de transporte público. La implementación exitosa del modelo de red neuronal ha demostrado ser clave para optimizar la asignación de recursos, anticipar eventos excepcionales y mejorar la experiencia del usuario.
- El análisis de datos históricos y en tiempo real fue esencial para comprender la dinámica de la ruta, también considerar variables como franjas horarias, tipos de día, contingencias y condiciones climáticas permitió una evaluación completa de los factores que influyen en la demanda. Este enfoque cimentó las bases para una toma de decisiones informada y estratégica en la gestión del transporte público.
- La identificación de patrones de demanda proporcionó una visión detallada de la utilización de la ruta, resaltando picos y valles a lo largo del tiempo, este análisis permitió comprender la naturaleza cambiante de la demanda, desde los momentos más intensos hasta las tendencias a largo plazo. La detección temprana de estos patrones es crucial para anticipar necesidades y optimizar la operación de manera eficiente.
- La elección de un modelo de red neuronal apropiado fue un componente clave en el éxito de la investigación. Si bien se consideraron diversas arquitecturas y técnicas de entrenamiento

la aquí presentada demostró ser la más apropiada al ser robusta y precisa en la predicción de la demanda.

- La evaluación del impacto de la implementación del modelo en la operación y la planificación de la ruta ha revelado que al asignar de manera más eficiente los recursos se perciben mejoras en la satisfacción del usuario y la eficiencia operativa general.
- Se podría decir que el modelo actual es un buen punto de partida, pero la mejora continua y la optimización son esenciales para abordar variabilidades y garantizar predicciones más precisas. Asimismo un enfoque iterativo, con la incorporación de más datos y ajustes en la metodología, puede fortalecer la utilidad práctica del modelo en la gestión efectiva de la demanda de pasajeros en la ruta T11 del SITP.

9. LISTA DE REFERENCIAS.

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. arXiv preprint arXiv:1603.04467.
2. Ackermann, A. E. F. y Sellitto, M. A. . (2022). Métodos de previsão de demanda: uma revisão da literatura. *Innovar*, 32(85), 83–99. <https://doi.org/10.15446/innovar.v32n85.100979>
3. Aghajan, Hamid Delgado, Ramón López-Cózar Augusto, Juan Carlos. (2010). Human-Centric Interfaces for Ambient Intelligence. Elsevier, 378 – 379. <https://app.knovel.com/hotlink/toc/id:kpHCIAI003/human-centric-interfaces/human-centric-interfaces>
4. Arroyo Guardado, D. Díaz Vico, J. y Hernández Encinas, L. (2019). Blockchain. Editorial CSIC Consejo Superior de Investigaciones Científicas. <https://elibro-net.bdbiblioteca.universidadean.edu.co/es/lc/bibliotecaean/titulos/111431>
5. Benbow, Donald W. (2021). Mistakes in Quality Statistics and How to Fix Them - 2.3.1 On the Lighter Side. American Society for Quality (ASQ), 25 – 29. <https://app.knovel.com/hotlink/pdf/id:kt0130CLB2/mistakes-in-quality-statistics/on-the-lighter-side>
6. Big Data. (2021). Optimización de descenso gradiente de Code Adam desde cero. Machine Learning. <https://topbigdata.es/optimizacion-de-descenso-de-gradiente-de-code-adam-desde-cero/>

7. Chen, D., & Kottenstede, A. (2019). Neural Networks for Predicting Public Transport Arrival Times. Journal of Transportation Engineering. DOI: [10.1061/(ASCE)TE.1943-5436.0001211]([https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0001211](https://doi.org/10.1061/(ASCE)TE.1943-5436.0001211))
8. Chen, C., & Song, H. (2019). Predictive Analytics for Public Transportation: A Review and Prospect. Transport Reviews. DOI:[10.1080/01441647.2019.1598452](<https://doi.org/10.1080/01441647.2019.1598452>)
9. Chollet, F. (2015). Keras. <https://keras.io>.
10. Colectivo a partir de los ITS [Tesis de pregrado, Universidad de Antioquia]. Repositorio Institucional Universidad de Antioquia. <https://hdl.handle.net/10495/15616>
11. ESRI. (s.f.). ¿Qué son los SIG? <https://www.esri.com/es-es/what-is-gis/overview>
12. Fernández Prieto, M. A. (2018). Revisión de la implementación del SITP: la democratización y la provisionalidad. Via Inveniendi Et Iudicandi, 13(2), 53–80. <https://doi.org/10.15332/s1909-0528.2018.0002.07>
13. García Díaz, J. C. (2016). Predicción en el dominio del tiempo: análisis de series temporales para ingenieros. Editorial de la Universidad Politécnica de Valencia. <https://elibro-net.bdbiblioteca.universidadean.edu.co/es/lc/bibliotecaean/titulos/57439>
14. García, A.E. (2019). Modelo de interconexión y entrenamiento redes neuronales artificiales, basado en recompensa y castigo empleando técnicas de vida artificial. [Tesis de maestría, Universidad Nacional de Colombia]. Repositorio Unal. <https://repositorio.unal.edu.co/bitstream/handle/unal/56879/alexanderespinosagarc%C3%ADa.2015.pdf?sequence=1>
15. Garrido Rodríguez, María Concepción. (2016). Análisis de eficiencia de redes neuronales en el estudio de la calidad del servicio en el transporte público. Aplicación al área metropolitana de Granada. [Tesis de doctorado, Universidad de Granada]. Repositorio Dialnet Unirioja. <http://hdl.handle.net/10481/55743>
16. Goble, William M. (2010). Control Systems Safety Evaluation and Reliability (3rd Edition) - C.1.1 Probability. International Society of Automation (ISA). <https://app.knovel.com/hotlink/pdf/id:kt0113SSV2/control-systems-safety/probability>
17. Gómez Hernández, Juan Pablo (2020). Planificación De Transporte Público Gorbea Portal, Salvador, y Batista Paneque, Dania. (2018). Identificación de patrones de comportamiento informativo en los textos mediante el Análisis de Redes Sociales. E-Ciencias de la Información, 8(1), 7-31. <https://dx.doi.org/10.15517/eci.v8i1.28005>

18. Hernández, Fernández y Baptista. (2014). Metodología de la investigación. Capítulo 10. Bogotá, McGraw-Hill, sexta edición.
19. Lu, L., & Liu, C. (2003). Predicting Public Bus Arrival and Departure Times: A Comparison of Artificial Neural Networks and Box-Jenkins Models. Journal of Transportation Engineering. DOI: [10.1061/(ASCE)0733-947X(2003)129:6(608)]
20. Ma, W., & Wang, Y. (2014). Bus Arrival Time Prediction with Artificial Neural Networks and Artificial Bee Colony Algorithm. Mathematical Problems in Engineering. DOI: [10.1155/2014/148427](<https://doi.org/10.1155/2014/148427>)
21. Mahey, Husan. (2020). Robotic Process Automation with Automation Anywhere - 18.5 Summary. Packt Publishing, 275-512. <https://app.knovel.com/hotlink/pdf/id:kt012ILCJ6/robotic-process-automation/managing-e-summary>
22. Martínez, A., y Sánchez Guerrero, G. (2015). Aplicación de funciones de distribución continuas, para modelar la demanda de pasajeros en una línea de tren ligero de la Zona Metropolitana del Valle de México. Contaduría y administración, 61(1), 159- 175.
<http://dx.doi.org/10.1016/j.cya.2015.09.002>
23. Medina La Plata, E. H. (2023). Big data: los datos como generadores de valor. 1. Universidad Peruana de Ciencias Aplicadas (UPC). <https://elibro-net.bdbiblioteca.universidadean.edu.co/es/lc/bibliotecaean/titulos/231338>
24. Meseguer González, P. y López de Mántaras Badia, R. (2017). Inteligencia artificial. Editorial CSIC Consejo Superior de Investigaciones Científicas. <https://elibro-net.bdbiblioteca.universidadean.edu.co/es/lc/bibliotecaean/titulos/42319>
25. Morales, Alejandro. (2019). Plataformas de participación ciudadana y comunicación en red. OCM. <https://www.esri.com/es-es/what-is-gis/overview>
26. Ong, G., & Thakuriah, P. (2015). Public Transport Demand Prediction with Artificial Neural Networks. Procedia Computer Science. DOI:[10.1016/j.procs.2015.08.325](<https://doi.org/10.1016/j.procs.2015.08.325>)
27. ‘Sullivan, Maurice G. (2017). Handbook for Sensory and Consumer-Driven New Product Development - Innovative Technologies for the Food and Beverage Industry - 2.6 Case Study - Demonstration of Quantitative Descriptive Analysis - The Descriptive Profile Training Process. Elsevier. <https://app.knovel.com/hotlink/pdf/id:kt01149932/handbook-sensory->
28. Posada Henao, John Jairo, y González Calderón, Carlos Alberto. (2010). Metodología para estudio de demanda de transporte público de pasajeros en zonas rurales. Revista Facultad de Ingeniería Universidad de Antioquia, (53), 106-118.

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-62302010000300009&lng=en&tlng=es.

29. ¿Qué es una red neuronal y cómo funciona? (2023). Escuela británica De Artes Creativas Y tecnología. Recuperado 22 de agosto de 2023, de <https://ebac.mx/blog>
30. Salvendy, Gavriel. (2001). Handbook of Industrial Engineering - Technology and Operations Management (3rd Edition) - 102.2 Optimization under Uncertainty. John Wiley & Sons, 114 - 150. <https://app.knovel.com/hotlink/pdf/id:kt006J0JT2/handbook-industrial-engineering/optimization-under-uncertainty>
31. Radziwill, N. M. (2020). Connected, Intelligent, Automated - The Definitive Guide to Digital Transformation and Quality 4.0 - 2.2.4 Software-Defined Systems. American Society for Quality (ASQ), 29-30. <https://app.knovel.com/hotlink/pdf/id:kt012DYB93/connected-intelligent/software-defined-systems>
32. Radziwill, N. M. (2020). Connected, Intelligent, Automated - The Definitive Guide to Digital Transformation and Quality 4.0 - 2.4.3 Internet of Things (IoT). American Society for Quality (ASQ), 37-38. <https://app.knovel.com/hotlink/pdf/id:kt012DYBK6/connected-intelligent/internet-things-iot>
33. Recio , G. (s. f.). Redes neuronales para predecir la ocupación en autobuses | inLab . Inlab <https://inlab.fib.upc.edu/es/blog/redes-neuronales-para-predecir-la-ocupacion-en-autobuses>
34. Rodríguez-Rueda, P. J. y Turias-Domínguez, I. J. (2017). Una comparativa entre redes neuronales artificiales y métodos clásicos para la predicción de la movilidad entre zonas de transporte. Aplicación práctica en el Campo de Gibraltar, España. DYNA, 84(200), 209–216. <https://doi.org/10.15446/dyna.v84n200.56571>
35. Rúa Anna, Roma Jordi y Bagén Toni. (2019). Deep Learning principios y fundamentos Editorial UOC. En la base de datos Ebooks 7-24. <https://elibro-net.bdbiblioteca.universidadean.edu.co/es/ereader/bibliotecaean/126167?page=3>
36. Ruggeri, Fabrizio Kenett, Ron S. Faltin, Frederick W. (2007). Encyclopedia of Statistics in Quality and Reliability, Volumes 1- 4 . John Wiley & Sons, 1720 - 1722. <https://app.knovel.com/hotlink/pdf/id:kt007NUX31/encyclopedia-statistics/sampling-i-references>
37. Ruggeri, Fabrizio Kenett, Ron S. Faltin, Frederick W. (2007). Encyclopedia of Statistics in Quality and Reliability, Volumes 1-4 - Error. John Wiley & Sons, 1060. <https://app.knovel.com/hotlink/pdf/id:kt007NU6L3/encyclopedia-statistics/error>

38. Ruggeri, Fabrizio Kenett, Ron S. Faltin, Frederick W. (2007). Encyclopedia of Statistics in Quality and Reliability, Volumes 1-4 - Neural Network Modelling. John Wiley & Sons, 1228 - 1233.
<https://app.knovel.com/hotlink/pdf/id:kt007NUDL1/encyclopedia-statistics/neural-network-modeling>
39. Salazar-Aguilar, Angélica y Cabrera-Rios, Mauricio. (2007). Pronóstico de demanda por medio de redes neuronales artificiales. Ingenierías. 10 (7), 6-12.
https://www.researchgate.net/publication/236028213_Pronostico_de_demanda_por_medio_de_redes_neuronales_artificiales
40. Salvendy, Gavriel. (2001). Handbook of Industrial Engineering - Technology and Operations Management (3rd Edition) - 65.3.2.4 Unsupervised Neural Networks (Competition Based). John Wiley & Sons, 1779 – 1780. <https://app.knovel.com/hotlink/pdf/id:kt006IZKE1/handbook-industrial-engineering/unsupervised-neural-networks>
41. Smith, J., & Johnson, A. (2018). Transportation Research Part C: Emerging Technologies. Demand Prediction in Public Transportation Systems: A Review. DOI: [10.1016/j.trc.2018.01.001](https://doi.org/10.1016/j.trc.2018.01.001)
42. TechEdu. (2023). Unidad lineal rectificada (ReLU) <https://techlib.net/techedu/unidad-lineal-rectificada-relu/>
43. Torres Álvarez, Nelson Stivet, Hernández, César y Pedraza, Luis F. (2011). Redes neuronales y predicción de tráfico. Tecnura, 15(29), 90-97.
[http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-921X2011000200009&lng=en&tlng=es.](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-921X2011000200009&lng=en&tlng=es)
44. TRANSMILENIO S.A. (2012). El SITP. TRANSMILENIO S.A.
<https://www.sitp.gov.co/publicaciones/40074/que-es-el-sitp/>
45. TRANSMILENIO S.A. (2013). Historia de TransMilenio. (2013).
<https://www.transmilenio.gov.co/publicaciones/146028/historia-de-transmilenio/>
46. TRANSMILENIO S.A. (2023). Estadísticas de oferta y demanda del Sistema Integrado de Transporte Público SITP. <https://www.transmilenio.gov.co/publicaciones/149180/estadisticas-de-oferta-y-demanda-del-sistema-integrado-de-transporte-publico-sitp/>
47. Wang, Yaowu Zhu, Yimin Shen, Geoffrey Q. P. Al-Hussein, Mohamed. (2018). ICCREM 2018 - Construction Enterprises and Project Management - 30.3.1 Evaluation of Technical Efficiency. American Society of Civil Engineers (ASCE). 240 – 248.
<https://app.knovel.com/hotlink/pdf/id:kt012NYEFT/iccrem-2018-construction/evaluation-technical>
48. Yiu, T. (2019). Understanding neural networks. We explore how neural networks. Towards data science. <https://towardsdatascience.com/understanding-neural-networks-19020b758230>

49. Zhang, Y., Zheng, Y., & Qi, H. (2016). A Survey on the Prediction of Bus Arrival Time. Journal of Traffic and Transportation Engineering. DOI: [10.1016/j.jtte.2016.04.002](https://doi.org/10.1016/j.jtte.2016.04.002)
50. Zhou, Shiyu Chen, Yong. (2021). Industrial Data Analytics for Diagnosis and Prognosis with R - A Random Effects Modelling Approach - 2. Introduction to Data Visualization and Characterization. John Wiley & Sons, 13 - 18. <https://app.knovel.com/hotlink/pdf/id:kt012XH1BG/industrial-data-analytics/introduction-data-visualization>

ANEXOS

- Código en Python modelo de red neuronal

```

import numpy as np
import pandas as pd
import tensorflow as tf

    from tensorflow import keras
    from sklearn.model_selection import train_test_split
    from sklearn.preprocessing import StandardScaler
data = pd.read_csv('datos_entrenamiento.csv')
data_test = pd.read_csv('datos_prueba.csv')
# Preprocesar los datos
    X = data.drop('Demanda_Pasajeros', axis=1)
    # Características
    y = data['Demanda_Pasajeros']
    # Variable objetivo
# Dividir el conjunto de datos de prueba y validación
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)
# Escalar las características para normalizarlas
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_val_scaled = scaler.transform(X_val)
# Modelo de red neuronal
model = keras.Sequential([
    keras.layers.Dense(64, activation='relu', input_shape=(X_train_scaled.shape[1],)),
    keras.layers.Dense(32, activation='relu'),
    keras.layers.Dense(1)
# Capa de salida para la predicción ])# Compilar el modelo
model.compile(optimizer='adam', loss='mean_squared_error')
# Entrenar el modelo
model.fit(X_train_scaled, y_train, epochs=50, batch_size=32,
validation_data=(X_val_scaled, y_val))
# Evaluar el modelo en datos de prueba

```

```

X_test = data_test.drop('Demanda_Pasajeros', axis=1)
X_test_scaled = scaler.transform(X_test)
y_test = data_test['Demanda_Pasajeros']
# Evaluar el rendimiento del modelo
loss = model.evaluate(X_test_scaled, y_test)
print("Pérdida en datos de prueba:", loss)
# Realizar predicciones
predictions = model.predict(X_test_scaled)
# Guardar el modelo entrenado
model.save('modelo_prediccion_demanda.h5')

```

- Código en Python análisis estadístico

```

import numpy as np
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score (42)
dias_semana = ["Lunes", "Martes", "Miércoles", "Jueves", "Viernes", "Sábado", "Domingo"]
demanda_real = (1000000, 500000, size=(7, 24)) # Demanda real por hora para cada día de la
semana
demanda_predicha = demanda_real + np.random.normal(0, 50000, size=(7, 24)) #Añadir ruido a
las predicciones
# Medidas de tendencia central y dispersión
media_real = np.mean(demanda_real)
media_predicha = np.mean(demanda_predicha)
std_real = np.std(demanda_real)
std_predicha = np.std(demanda_predicha)
# Correlación
correlacion = np.corrcoef(demanda_real.flatten(), demanda_predicha.flatten())[0, 1]
# Métricas de evaluación
mae = mean_absolute_error(demanda_real.flatten(), demanda_predicha.flatten())
mse = mean_squared_error(demanda_real.flatten(), demanda_predicha.flatten())
r2 = r2_score(demanda_real.flatten(), demanda_predicha.flatten())
# Impresión de resultados
print("Medidas de Tendencia Central:")

```

```
print(f"Media de Demanda Real: {media_real:.2f}")
print(f"Media de Demanda Predicha: {media_predicha:.2f}")
print("\nMedidas de Dispersión:")
print(f"Desviación Estándar de Demanda Real: {std_real:.2f}")
print(f"Desviación Estándar de Demanda Predicha: {std_predicha:.2f}")
print("\nCorrelación:")
print(f"Coeficiente de Correlación: {correlacion:.2f}")
print("\nMétricas de Evaluación:")
print(f"Error Absoluto Medio (MAE): {mae:.2f}")
print(f"Error Cuadrático Medio (MSE): {mse:.2f}")
print(f"Coeficiente de Determinación (R²): {r2:.2f}")
```