



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

FACULTAD DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA
MAESTRÍA EN OPTIMIZACIÓN Y CÓMPUTO APLICADO

**Predicción de tiempo de llegada entre paradas de autobuses de
transporte público de Cuernavaca mediante redes neuronales
artificiales**

T E S I S

QUE PARA OBTENER EL GRADO DE
MAESTRÍA EN OPTIMIZACIÓN Y CÓMPUTO APLICADO

PRESENTA

EDELMIRA TAPIA ARROYO

DIRECTOR DE TESIS

DR. PEDRO MORENO BERNAL

CO-DIRECTOR

DRA. JESÚS DEL CARMEN PERALTA ABARCA

REVISORES:

DR. PEDRO MORENO BERNAL

DRA. JESUS DEL CARMEN PERALTA ABARCA

DR. SERGIO NESMACHNOW

DR. JOSÉ ALBERTO HERNÁNDEZ AGUILAR

DR. OUTMANE OUBRAM



Facultad de Contaduría,
Administración e Informática

CUERNAVACA, MORELOS.

ABRIL, 2026



Predicción de tiempo de llegada entre paradas de autobuses de transporte público de Cuernavaca mediante redes neuronales artificiales por Edelmira Tapia Arroyo tiene licencia [CC Atribución 4.0](https://creativecommons.org/licenses/by/4.0/).

Dedicatoria

A Dios, por ser mi fortaleza y guía en cada etapa de mi vida.

A la memoria de mi madre Carmen, quien, aunque no pudo acompañarme en esta etapa, sembró en mí los valores, la fortaleza y la perseverancia que hoy me permiten alcanzar esta meta.

A mi esposo Marco Antonio y a mi hijo Marco Alejandro, por su apoyo incondicional y comprensión, y por acompañarme en cada paso hacia la culminación de este logro.

A mis hermanos, por su respaldo constante y por permitirme contar con ellos en todo momento. A toda mi familia y amigos, por su apoyo moral y por estar siempre presentes.

Agradecimientos

Agradezco a la Universidad Autónoma del Estado de Morelos y a la Facultad de Contaduría, Administración e Informática por el respaldo institucional brindado para el desarrollo de esta investigación.

Asimismo, agradezco a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación por el otorgamiento de la beca CONAHCYT, cuyo respaldo fue determinante para el desarrollo de mis estudios de maestría.

De manera particular, agradezco al Dr. Pedro Moreno Bernal, director de esta tesis, por su orientación, acompañamiento y apoyo académico durante el desarrollo del presente trabajo.

A la Dra. Jesús del Carmen Peralta Abarca, codirectora de esta investigación, por su guía y motivación constante. Al Dr. José Alberto Hernández Aguilar, por su disposición como profesor y su orientación académica a lo largo de esta maestría.

Al Dr. Federico Alonso Pecina, Dr. José Crispín Zavala Díaz por su apoyo y orientación durante mi formación académica.

Agradezco al Dr. Sergio Nesmachnow el apoyo tecnológico brindado y por ser parte de mi comité tutorial. Asimismo, extiendo mi agradecimiento a los profesores que contribuyeron a mi formación durante la maestría por su guía, conocimientos y aportaciones para la mejora de este trabajo.

Resumen

El transporte público es un componente fundamental de la movilidad urbana en las ciudades. Sin embargo, uno de los principales desafíos que enfrentan los usuarios es la falta de información precisa sobre los tiempos de llegada en cada parada.

Este trabajo presenta un enfoque de aprendizaje supervisado que utiliza redes neuronales artificiales para estimar el tiempo de viaje entre dos paradas de autobús. El conjunto de datos comprende información de 2,073 viajes, incorporando diversos factores, como horarios de salida y llegada, velocidad, tiempo total de recorrido, número de topes y semáforos, distancia recorrida y otros eventos imprevistos. El estudio de caso se centró en el transporte público urbano en Cuernavaca, Morelos, México, abarcando el periodo del 1 de octubre de 2024 al 20 de febrero de 2025.

Los resultados mostraron que el modelo propuesto de red neuronal artificial alcanzó un alto coeficiente de determinación (R^2), que osciló entre 0.9995 y 0.9998, en comparación con el método de referencia para predecir los tiempos de llegada. El enfoque propuesto resalta el potencial del uso de redes neuronales artificiales como herramienta de apoyo a la toma de decisiones para optimizar los tiempos de llegada del transporte en cada parada, facilitando así una mejor organización y planificación de los horarios de viaje.

Palabras clave: Sistemas de transporte inteligente, estimación de tiempos de llegada, transporte público, redes neuronales artificiales.

Abstract

Public transportation is a crucial component of urban mobility in cities. However, one of the main challenges that users face is the lack of accurate information about arrival times at each stop. This thesis presents a supervised learning approach using artificial neural networks for estimating travel time between any two bus stops. The dataset comprises information from 2 073 trips, incorporating various factors, e.g., starting and arrival times, speed, total travel time, the number of speed bumps and traffic lights, distance traveled, and other unexpected events. The case study focused on urban public transport in Cuernavaca, Morelos, Mexico, covering the period from October 1, 2024 to February 20, 2025. Results showed that the proposed artificial neural network model achieved a high coefficient of determination (R^2), ranging from 0.9995 to 0.9998, compared to the baseline method for predicting arrival times. The proposed approach highlights the potential of using artificial neural networks as a decision-making tool to optimize transportation arrival times at each stop, thus facilitating better organization and planning of travel schedules.

Palabras clave: arrival time estimation, public transportation, artificial neural networks.

Índice general

1. Introducción	1
1.1. Marco teórico	4
1.1.1. Sistema de Transporte Inteligente	4
1.1.2. Transporte público	5
1.1.3. Aprendizaje automático	6
1.1.4. Redes neuronales artificiales	7
1.1.5. Predicción del tiempo de llegada	8
1.2. Pregunta de investigación	10
1.3. Hipótesis	10
1.4. Objetivos	10
1.4.1. Objetivo General	10
1.4.2. Objetivos particulares	11
1.5. Alcances	11
1.6. Contribuciones de la tesis	12
1.7. Estructura de la tesis	13
2. Problema a tratar	14
2.1. Transporte público en México	14
2.2. Transporte público en Cuernavaca	16
2.3. Justificación	18
3. Revisión de antecedentes	21
3.1. Trabajo Relacionado	21
4. Metodología propuesta	54
4.1. RNA para problemas de regresión	54
4.2. RNA para estimar el tiempo de llegada	56
4.2.1. Fuente de datos	56
4.2.2. Preparación de los datos	58
4.2.3. Definición de variables del modelo	63
4.2.4. RNA para la estimación del tiempo de llegada	65

5. Resultados experimentales y discusión	68
5.1. Plataforma de desarrollo y ejecución	68
5.2. Área de estudio	69
5.3. Ajuste de hiperparámetros	69
5.4. Métricas evaluadas	72
5.5. Resultados y discusión	73
6. Conclusiones y trabajo futuro	77
Referencias	79

Índice de figuras

1.1. Arquitectura típica de una Red Neuronal (Fuente: Incio-Flores et al. (2023), Revista Electrónica Educare. CC BY-NC-ND 3.0) . . .	8
2.1. Representación del traslado del autobús entre la parada 1 y la parada 2 (Fuente: Elaboración propia a partir de elementos de dominio público)	17
2.2. Representación del traslado del autobús entre paradas	17
4.1. Matriz de correlación de los datos de entrada	62
4.2. Diagrama de flujo del algoritmo de regresión supervisada	67
5.1. Trayectoria del autobús 13 desde la UAEM hasta Walmart	70
5.2. Histograma de valores de tiempo total	71
5.3. Función de pérdida de MSE	74
5.4. Función R^2	74
5.5. Predicción vs. objetivo (RNA)	75
5.6. Predicción vs. objetivo (MLR)	75

Índice de cuadros

4.1. Estadísticas resumidas de las variables de entrada de la RNA . .	61
---	----

Capítulo 1

Introducción

Hoy en día, el transporte urbano es crucial para el desarrollo sostenible de las ciudades. Ofrece a los ciudadanos opciones de movilidad accesibles, asequibles y eficientes, lo que repercute significativamente en su vida diaria, incluyendo el trabajo, los negocios, la recreación y la atención médica (Al Suleiman, Cortez, Monzón, y Lara, 2023).

Una movilidad urbana eficiente debe proporcionar servicios inteligentes que anticipen los cambios y faciliten la toma de decisiones informada. Sin embargo, gestionar la movilidad urbana en ciudades de alta densidad es complejo debido a la congestión vehicular, la diversidad de modos de transporte y los numerosos orígenes y destinos (Rodrigue, 2020). En este contexto, es necesario optimizar el transporte público urbano para mejorar la movilidad, reducir la congestión vehicular, las emisiones de gases de efecto invernadero y fomentar la equidad social. De esta manera, el transporte público contribuiría a promover la movilidad sostenible, disminuyendo la dependencia del vehículo privado, mejorando la calidad del aire y optimizando la movilidad urbana en general (Nesmachnow

y Hipogrosso, 2022).

El transporte público urbano se enfrenta a diversos desafíos, como la imprevisibilidad de los horarios de llegada de los autobuses y las variaciones en los tiempos de viaje entre paradas. Los problemas del transporte público se deben a factores como la congestión vehicular, la inconsistencia en los horarios de embarque y los retrasos inesperados, como accidentes u obras viales. Además, la falta de un sistema de transporte público bien integrado genera estimaciones de llegada inexactas, lo que resulta en tiempos de espera prolongados, conexiones perdidas y una insatisfacción general de los pasajeros (Kuo, Leung, y Yan, 2023).

El análisis de datos urbanos es fundamental para la evaluación del transporte público y las iniciativas de movilidad sostenible (Nesmachnow y Hipogrosso, 2024). La ausencia de integración de datos en tiempo real y de sistemas de programación adaptativos dificulta que las agencias de transporte público respondan eficazmente a las condiciones cambiantes, como el tráfico, las horas punta y la alta demanda. Para abordar los problemas relacionados con el transporte público urbano, es esencial implementar sistemas inteligentes, como el rastreo GPS, los modelos de aprendizaje automático y el análisis de tráfico en tiempo real, para mejorar la fiabilidad y optimizar la planificación de rutas para los ciudadanos, convergiendo así hacia los Sistemas Inteligentes de Transporte (ITS, por sus siglas en inglés) (European Commission, 2009).

En México, el servicio de transporte público constituye un elemento fundamental para la movilidad urbana, dado que, diariamente millones de personas

requieren trasladarse a diversos destinos con fines laborales, comerciales, recreativos, médicos, entre otros. No obstante, uno de los principales desafíos que enfrentan los usuarios es la falta de información precisa sobre los horarios de llegada del transporte a cada una de las paradas, lo cual dificulta la adecuada organización y previsión del tiempo destinado al traslado.

De acuerdo a datos del Instituto Nacional de Estadística, Geografía e Informática, durante el mes de marzo, se registró un total de 250.6 millones de viajes realizados en transporte público con distintos propósitos, tales como asistir al trabajo, realizar compras, acudir a citas médicas o participar en actividades de esparcimiento (INEGI, 2025). El transporte público desempeña un papel central, ya que es el medio más utilizado para desplazarse dentro de la ciudad. Sin embargo, su operación suele presentar limitaciones en cuanto a control, planificación y organización, factores que inciden directamente en la eficiencia del servicio y en las condiciones generales de movilidad.

León Sosa et al. (León Sosa, Hernández Báez, Morales Morales, y Maldonado Martínez, 2019) describen al transporte público como una forma de movilidad esencial utilizada por prácticamente todos los habitantes, lo que tiene un impacto directo en su vida diaria. León Sosa et al. señalan que, según el PIMUS (Plan Integral de Movilidad Sustentable para la Zona Metropolitana de Cuernavaca), el transporte público realiza más de 300 mil viajes al día, distribuidos en 36 rutas, con un parque vehicular de 2492 vehículos registrados. En cuanto a los traslados escolares, se utilizan diversos medios, como rutas, taxis, combis y otros vehículos públicos. De acuerdo a estas cifras, aproximadamente 36 de

cada 100 personas utilizan el transporte público para este tipo de desplazamientos. El transporte, entendido como un componente básico de la vida cotidiana, posibilita el acceso a actividades laborales, educativas, de salud y de recreación. Asimismo, An et al. (An, Choi, y Choi, 2022) señalan que los ITS se han convertido en un elemento clave para optimizar la movilidad urbana, pues su implementación ofrece a los usuarios herramientas que facilitan la organización de sus desplazamientos, contribuyendo a reducir la saturación vial y a emplear de manera más eficiente la infraestructura existente. Por ello, contar con estimaciones de tiempo de viaje precisas y confiables resulta esencial para una adecuada planificación.

1.1. Marco teórico

En esta sección se presenta el marco teórico de la investigación, con el propósito de exponer los conceptos, antecedentes y fundamentos necesarios para comprender el contexto, la problemática abordada y la relevancia del estudio.

1.1.1. Sistema de Transporte Inteligente

El Instituto Federal de Telecomunicaciones (Instituto Federal de Telecomunicaciones, 2018), define a los ITS como aquellos que utilizan tecnología de la información y comunicación, dispositivos de localización y herramientas de automatización con la finalidad de mejorar la seguridad, la gestión, la eficiencia y la eficacia de la infraestructura y medios de transporte, además, buscan dar respuesta a los problemas de congestión de tráfico cada vez mayores derivados

del constante crecimiento poblacional de las zonas urbanas y los cambios en la densidad demográfica.

De acuerdo con Zuluaga-Isaza y Arango-Monsalve (Zuluaga-Isaza y Arango-Monsalve, 2023), los ITS han cobrado mayor relevancia debido al rápido avance de las técnicas de aprendizaje automático y al acceso a nuevas fuentes de datos. Gracias a estos desarrollos, es posible analizar y prever las condiciones del tráfico con mayor rapidez y precisión.

1.1.2. Transporte público

La Secretaría de Movilidad del Gobierno del Estado de México (Secretaría de Movilidad del Estado de México, 2024) menciona que el transporte público se entiende como aquellos medios de transporte cuyos pasajeros no son propietarios de las unidades, y que son operados por terceros (empresas públicas o privadas), cuya función principal es el traslado de personas de un lugar a otro dentro de una zona determinada, como una ciudad, mediante el pago de una tarifa predeterminada de acuerdo a la distancia o trayecto recorrido.

El transporte público constituye un medio esencial para que las personas puedan desplazarse de un lugar a otro con el propósito de realizar las actividades cotidianas relacionadas con el trabajo, la educación, la recreación, entre otras. En ciudades como Cuernavaca, el servicio de autobuses es fundamental para la movilidad urbana, ya que facilita el traslado de los usuarios hacia distintos destinos. Sin embargo, uno de los principales problemas que enfrentan es la incertidumbre respecto al tiempo de llegada de las unidades a las paradas, lo que

puede generar retrasos en sus actividades y afectar la planificación y disminuir la calidad del servicio.

1.1.3. Aprendizaje automático

El aprendizaje automático es una rama de la informática que se centra en el diseño y la aplicación de algoritmos capaces de aprender a partir de los datos (Raschka y Mirjalili, 2018). Al utilizar estos algoritmos, la información procesada puede convertirse en conocimiento útil. Según estos autores, existen tres modalidades principales de aprendizaje automático.

En el aprendizaje supervisado, el modelo se entrena con datos previamente etiquetados para poder predecir información sobre datos futuros o no observados. Dentro de este enfoque, la regresión permite estimar valores continuos a partir de la relación entre las variables predictoras y la variable dependiente. Por otro lado, el aprendizaje no supervisado trabaja con datos sin etiquetar, lo que facilita explorar la estructura de la información y extraer conocimiento relevante sin necesidad de contar con una variable de resultado conocida. Entre sus técnicas más utilizadas se encuentra el agrupamiento o clustering, que organiza grandes volúmenes de datos en subgrupos cuyos elementos comparten características internas y se diferencian de los demás grupos, de ahí que también se denomine clasificación no supervisada. Otro enfoque importante es la reducción de la dimensionalidad, que se aplica en contextos con numerosas variables y ayuda a mejorar el almacenamiento y el rendimiento de los algoritmos (Raschka y Mirjalili, 2018).

De acuerdo a Raschka y Mirjalili (Raschka y Mirjalili, 2018), en el aprendizaje por refuerzo se busca contar con un sistema capaz de mejorar el rendimiento a partir de las interacciones con el entorno. En este enfoque, el agente recibe información del entorno y una señal de recompensa que evalúa la calidad de las acciones realizadas, sin embargo, a diferencia de este último, la retroalimentación no corresponde con un valor o etiqueta correctos, sino con una valoración de la calidad de la acción ejecutada mediante una función de recompensa. Así, el agente aprende, a través de la interacción continua con el entorno, es decir, mediante una secuencia de acciones que mejoran la recompensa, ya sea mediante un proceso de ensayo y error o a través de una planificación deliberada.

1.1.4. Redes neuronales artificiales

Acevedo et al. (Acevedo, Serna, y Serna, 2017) señalan que una red neuronal puede entenderse como un sistema capaz de establecer relaciones entre entradas y salidas, tomando como inspiración el funcionamiento del sistema nervioso. A diferencia de la computación tradicional, no se basa en algoritmos secuenciales. Las Redes neuronales artificiales (RNA) simulan el comportamiento del cerebro humano, procesando la información de manera paralela y desarrollando la capacidad de aprender y generalizar frente a situaciones que no fueron consideradas durante los procesos de entrenamiento. Se pueden considerar como una herramienta computacional diseñada para abordar problemas complejos, con la capacidad de realizar predicciones en sistemas con relaciones no lineales.

Las RNA funcionan de una manera muy semejante al sistema nervioso

biológico, estas son capaces de simular el comportamiento del cerebro humano, las cuales se consideran un método computacional, con base a las cuales se trata de solucionar problemas complejos. La Figura 1.1 muestra una representación de la arquitectura típica de una Red Neuronal.

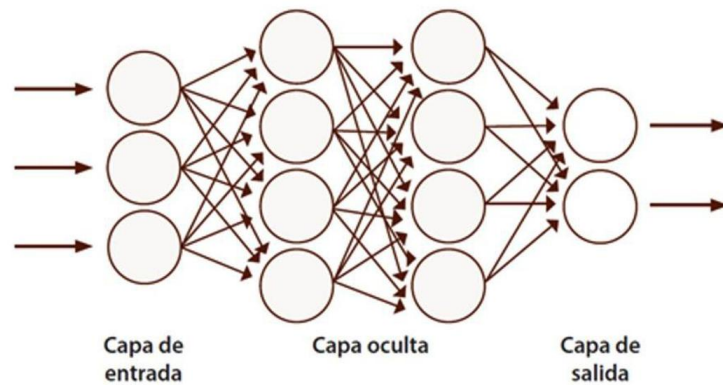


Figura 1.1: Arquitectura típica de una Red Neuronal (Fuente: Incio-Flores et al. (2023), Revista Electrónica Educare. CC BY-NC-ND 3.0)

1.1.5. Predicción del tiempo de llegada

La predicción del tiempo de llegada de los autobuses a las paradas constituye uno de los componentes más relevantes dentro de los Sistemas de Transporte Inteligente, debido a su impacto directo en la experiencia del usuario y en la eficiencia operativa del servicio. Contar con estimaciones confiables del tiempo de arribo permite a los pasajeros planificar mejor sus desplazamientos, reducir los tiempos de espera y aumentar la confianza en el sistema de transporte público.

Tradicionalmente, los sistemas de predicción de tiempos de llegada se han basado en métodos estadísticos simples o en promedios históricos de recorrido.

Sin embargo, estos enfoques presentan limitaciones importantes, ya que no consideran adecuadamente la variabilidad del tráfico, las condiciones climáticas, los eventos imprevistos ni los cambios en la demanda de pasajeros, lo que reduce la precisión de las estimaciones (Cathey y Dailey, 2003). Ante estas limitaciones, en los últimos años se ha incrementado el uso de técnicas de aprendizaje automático para la estimación del tiempo de llegada de los autobuses. Diversos estudios han demostrado que estos métodos permiten modelar relaciones no lineales y complejas entre múltiples variables, logrando predicciones más precisas que los enfoques tradicionales. Por ejemplo, Yu et al. (Yu, Lam, y Tam, 2011) señalan que los modelos basados en aprendizaje automático superan a los métodos clásicos al adaptarse mejor a las condiciones dinámicas del tráfico urbano.

En particular, las redes neuronales artificiales han mostrado un alto desempeño en la predicción de tiempos de viaje y de llegada, debido a su capacidad para aprender patrones a partir de grandes volúmenes de datos y generalizar su comportamiento ante situaciones no observadas durante el entrenamiento. Estas características las hacen especialmente adecuadas para entornos urbanos complejos, donde factores como la congestión vial, el número de paradas, la velocidad del vehículo y las condiciones externas influyen directamente en los tiempos de recorrido (Vlahogianni, Karlaftis, y Golias, 2014).

En este contexto, el uso de redes neuronales artificiales para la predicción del tiempo de llegada entre paradas de autobuses representa una alternativa sólida y eficiente para mejorar la calidad del servicio de transporte público. La

aplicación de estos modelos en ciudades como Cuernavaca permite atender la necesidad de información confiable en tiempo real y contribuye al desarrollo de soluciones tecnológicas orientadas a una movilidad urbana más eficiente e inteligente.

1.2. Pregunta de investigación

¿La implementación de un modelo predictivo basado en redes neuronales artificiales podría estimar el tiempo de llegada entre paradas de autobuses a partir de variables como condiciones climáticas, obstáculos en el trayecto, tiempos de traslado, y tipo de día (laboral o no laboral)?

1.3. Hipótesis

La aplicación de redes neuronales artificiales permite estimar el tiempo de llegada entre paradas de autobuses pertenecientes a la ruta No. 13 en Cuernavaca, considerando distintas variables, presentando una mejora en la planificación del tiempo de los usuarios del transporte público.

1.4. Objetivos

1.4.1. Objetivo General

Desarrollar un modelo de predicción basado en redes neuronales artificiales que permita estimar el tiempo de llegada entre paradas de autobuses pertene-

cientes a la ruta No. 13 en Cuernavaca, Morelos, considerando variables como condiciones climáticas, obstáculos en el trayecto, tiempos de traslado, y tipo de día (laboral o no laboral), con el propósito de mejorar la planificación del tiempo de los usuarios del transporte público.

1.4.2. Objetivos particulares

- Generar una base de datos mediante un estudio de campo que recopile información sobre la trayectoria del autobús No. 13, específicamente en el recorrido entre la Universidad Autónoma del Estado de Morelos y Walmart Jiutepec.
- Identificar y seleccionar las variables relevantes que servirán como datos de entrada para el modelo predictivo.
- Diseñar e implementar un modelo de redes neuronales artificiales que permita estimar el tiempo de llegada entre las paradas del recorrido mencionado.
- Evaluar el modelo propuesto contra un algoritmo de base para garantizar la precisión y confiabilidad del modelo desarrollado.

1.5. Alcances

El presente estudio se circunscribe al análisis y modelado del tiempo de llegada entre paradas del autobús de pasajeros correspondiente a la ruta No. 13 del transporte público en Cuernavaca, Morelos. Específicamente, se enfoca en

el trayecto comprendido entre la Universidad Autónoma del Estado de Morelos (UAEM) y la tienda Walmart ubicada en Jiutepec, así como el recorrido inverso. El modelo desarrollado mediante redes neuronales artificiales se aplica exclusivamente a este tramo de la ruta, considerando variables asociadas al tiempo de desplazamiento, la velocidad de circulación, la duración de las paradas, la distancia recorrida, las condiciones climáticas, la presencia de infraestructura vial como topes y semáforos, el tipo de día (laboral o no laboral), el turno de operación y la ocurrencia de eventos imprevistos durante el trayecto. Los resultados obtenidos y las predicciones generadas por el modelo no son extensibles a otras rutas del sistema de transporte público de Cuernavaca ni a otros contextos urbanos sin una validación adicional. En consecuencia, el alcance de esta investigación se limita a proporcionar una herramienta predictiva aplicable a la ruta No. 13 en el tramo analizado, con el objetivo de reducir la incertidumbre asociada al tiempo de espera de los usuarios.

1.6. Contribuciones de la tesis

Esta tesis presenta un enfoque de aprendizaje supervisado que utiliza RNA para estimar los tiempos de viaje entre dos paradas de autobús. El enfoque propuesto implementa un modelo de RNA multicapa de perceptrón para predecir los tiempos de llegada del transporte público urbano. El conjunto de datos comprende información de 2 073 viajes, incorporando factores como las horas de inicio y llegada, la velocidad promedio, el tiempo total de viaje, el clima, el

número de reductores de velocidad y semáforos encontrados, la distancia recorrida y otros eventos inesperados. El caso de estudio examinó la ruta de transporte público número 13 en Cuernavaca, Morelos, México, durante el período del 1 de octubre de 2024 al 20 de febrero de 2025. Los resultados indicaron que el modelo de RNA propuesto alcanzó un alto coeficiente de determinación (R^2), que oscila entre 0.9995 y 0.9998, superando el método de referencia para predecir los tiempos de llegada. El enfoque propuesto enfatiza el uso de RNA para optimizar los tiempos de llegada del transporte en cada parada, mejorando así la organización y planificación de los horarios de viaje.

1.7. Estructura de la tesis

El contenido de la tesis está organizado de la siguiente manera. El Capítulo 2 describe el problema a tratar, así como la justificación. El Capítulo 3 presenta el trabajo relacionado sobre la predicción de tiempo de llegada entre paradas de autobuses de transporte público mediante técnicas de Machine Learning. El Capítulo 4 describe la red neuronal artificial aplicada para estimar el tiempo de arribo a la parada del autobús, así como la metodología de solución propuesta. El Capítulo 5 presenta los resultados experimentales y discusión. Finalmente, el Capítulo 6 presenta las conclusiones y el trabajo futuro.

Capítulo 2

Problema a tratar

En este capítulo se expone el problema que da origen a la presente investigación, describiendo su contexto dentro del ámbito del transporte público urbano y su relevancia para la movilidad en la ciudad de Cuernavaca, Morelos. Asimismo, se identifican los principales aspectos que requieren atención, lo cual permite delimitar el alcance del estudio y orientar el desarrollo de la propuesta metodológica presentada en los capítulos posteriores.

2.1. Transporte público en México

El servicio de transporte público en México constituye un elemento fundamental para la movilidad urbana, dado que, diariamente millones de personas requieren trasladarse a diversos destinos con fines laborales, comerciales, recreativos, médicos, entre otros.

El transporte público es un medio de movilidad fundamental para millones de personas que se desplazan diariamente a distintos destinos por motivos laborales, educativos, comerciales o de ocio. Según datos recientes, la gran mayoría de la

población depende de este sistema para desplazarse por la ciudad, lo que subraya su importancia para la organización del tiempo y la planificación de los viajes urbanos ([Telediario CdMx, 2025](#)).

No obstante, este sector enfrenta diversos desafíos, entre los que destacan la falta de información precisa y oportuna sobre los horarios y tiempos de llegada de las unidades. Esta situación dificulta la previsión y programación de los desplazamientos lo que afecta tanto a la eficiencia del servicio como a la experiencia de los usuarios. Por ello, disponer de estimaciones fiables de los tiempos de viaje resulta clave para mejorar la planificación de los desplazamientos y fortalecer el papel del transporte público como componente esencial de la movilidad urbana.

La ausencia de estimaciones confiables sobre los tiempos de llegada del autobús genera incertidumbre en los usuarios, incrementa los tiempos de espera y limita la capacidad de organizar adecuadamente las actividades cotidianas. En este contexto, contar con herramientas tecnológicas que permitan estimar de manera precisa los tiempos de viaje resulta clave para fortalecer el papel del transporte público como componente esencial de la movilidad urbana. Según el Instituto Nacional de Estadística y Geografía([INEGI, 2025](#)), en marzo de ese año se registraron 250.6 millones de viajes en transporte público urbano a nivel nacional, lo que representó un incremento del 8.1 % respecto al mismo mes del año anterior. Estas cifras evidencian la alta dependencia de la población hacia este servicio y ponen de manifiesto la necesidad de mejorar su eficiencia y confiabilidad mediante el uso de tecnologías de información y modelos predictivos.

2.2. Transporte público en Cuernavaca

En el caso específico de la ciudad de Cuernavaca, el sistema de transporte público presenta características que incrementan la complejidad de la movilidad urbana. De acuerdo con León Sosa et al. (León Sosa y cols., 2019), el transporte público de la región está conformado por 36 rutas y más de 2 400 unidades. Sin embargo, la frecuencia del servicio y los tiempos de recorrido presentan una alta variabilidad debido a factores como la congestión vial, las condiciones climáticas, el estado de las vialidades, el funcionamiento irregular de las unidades y la falta de información confiable sobre los tiempos de llegada a cada parada. Como consecuencia de estas condiciones, se presentan incrementos en los tiempos de espera y dificultades para planificar los horarios de llegada a los distintos destinos, debido a la necesidad de contar con información precisa sobre los tiempos de arribo de los autobuses entre paradas.

Ante esta situación, la presente investigación se centra en la estimación de los tiempos de llegada de los autobuses correspondientes a la ruta número 13 del transporte público de Cuernavaca. Esta ruta fue seleccionada debido a su relevancia y alta demanda, al conectar zonas de interés como UAEM y el municipio de Jiutepec. En la Figura 2 se tiene una representación del traslado del autobús entre paradas a lo largo del recorrido analizado.

La predicción precisa de los tiempos de llegada de los autobuses representa un factor clave para mejorar la experiencia del usuario y optimizar la operación del sistema de transporte público. Disponer de esta información permite reducir los tiempos de espera, aumentar la confianza en el servicio y apoyar la planificación



Figura 2.1: Representación del traslado del autobús entre la parada 1 y la parada 2 (Fuente: Elaboración propia a partir de elementos de dominio público)

de la movilidad urbana de manera más eficiente y sostenible. La Figura 2.2 muestra una ruta de ejemplo para las paradas de autobús a lo largo de la ruta número 13. En la Figura 2.2, la hora de llegada a cualquier parada de autobús es impredecible debido a la falta de información sobre las horas de salida de los autobuses. Ante esta problemática, se plantea el desarrollo de un modelo basado en RNA, capaz de estimar con alta precisión los tiempos de llegada entre paradas de autobuses, utilizando datos operativos como velocidad promedio, condiciones climáticas y factores de tránsito.

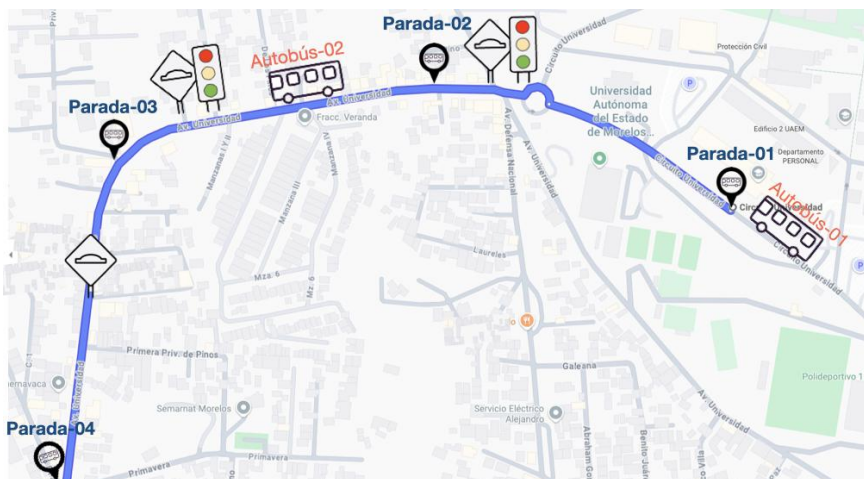


Figura 2.2: Representación del traslado del autobús entre paradas

2.3. Justificación

Ocampo (Ocampo Ortiz, 2021) señala que, a pesar de que el transporte público en la Ciudad de México es ampliamente utilizado, este carece de tecnologías adecuadas para mejorar la experiencia del usuario. Los sistemas de predicción actuales son limitados y no ofrecen acceso a datos provenientes del Sistema de Posicionamiento Global (Global Positioning System, GPS), y no cuentan con formatos estandarizados para su procesamiento e integración. A su vez, el autor destaca el creciente interés en desarrollar proyectos tecnológicos para mejorar las ciudades, respondiendo a nuevas necesidades diarias mediante soluciones tecnológicas, como los ITS. Estos sistemas tienen como objetivo mejorar la seguridad y eficiencia en el transporte facilitando la vida de las personas mediante la provisión de información que optimiza la administración del tiempo y la planificación diaria.

Diversos estudios han demostrado que la implementación de redes neuronales artificiales para estimar los tiempos de llegada de autobuses ha dado resultados positivos en distintos contextos internacionales. Gurmu et al. (Gurmu y Fan, 2014) desarrollaron un modelo basado únicamente en datos GPS y comprobaron que este enfoque es capaz de procesar grandes volúmenes de información y generar predicciones más precisas que los métodos basados en promedios históricos. Del mismo modo, los estudios de Julio et al. (Julio, Giesen, y Lizana, 2016) y Hassanayebi et al. (Hassannayebi, Farjad, Azadnia, Javidi, y Chunduri, 2023) muestran que el uso de técnicas de aprendizaje automático permite mejorar la predicción del comportamiento del transporte público frente a enfoques tra-

dicionales, especialmente en escenarios dinámicos y en tiempo real, como las máquinas de vectores de soporte (SVM) y el modelo autorregresivo integrado de media móvil (AutoRegressive Integrated Moving Average, ARIMA). Estas ventajas se mantienen incluso en escenarios complejos caracterizados por congestión vial o información incompleta, lo que confirma la solidez de este enfoque para la predicción de tiempos de viaje en sistemas de transporte público.

Por otro lado, investigaciones más recientes como las de Alam et al. (Alam, Kush, Emami, y Pouladzadeh, 2021) y Rashvand et al. (Rashvand, Hosseini, Azarbayjani, y Tabkhi, 2024) en sus trabajos de investigación, respaldan la eficacia de las RNA y en particular las variantes recurrentes como LSTM (Long Short-Term Memory por sus siglas en inglés) para capturar relaciones espacio-temporales complejas, adaptarse a condiciones dinámicas del entorno (como el clima o eventos inesperados) y realizar predicciones en tiempo real. Esto es fundamental para entornos urbanos como el de la ciudad de Cuernavaca, donde la variabilidad del tráfico, y diversos factores como accidentes, tipo de avenida, distancias entre una parada y otra y la falta de información en tiempo real impactan directamente en la experiencia del usuario del transporte público. al no contar con la información de tiempo de llegada a cada parada para poder organizar su tiempo y con esto tener una mejor calidad de organización del tiempo y de la vida diaria.

En este contexto, el uso de RNA en el presente proyecto ofrece una respuesta adecuada para predecir con mayor precisión los tiempos de llegada entre paradas del autobús No. 13. Esta tecnología permitirá optimizar la planificación de

los usuarios, reducir la incertidumbre en el servicio y contribuir a una movilidad urbana más eficiente e inteligente. Esta tesis propone desarrollar un Modelo de predicción de tiempo de llegada entre paradas de autobuses de transporte público de Cuernavaca mediante redes neuronales artificiales del tiempo de arribo de autobuses de pasajeros del transporte público a una parada específica, mediante redes neuronales artificiales, utilizando información recabada a partir de los recorridos de los autobuses, considerando variables como tiempos entre paradas, horas de salida y llegada, velocidad, topes, semáforos, datos climáticos, número de ascensos y descensos, embotellamiento, tipo de día, turno matutino o vespertino así como imprevistos ocurridos durante el trayecto.

El siguiente capítulo describe el trabajo relacionado para abordar el problema a tratar.

Capítulo 3

Revisión de antecedentes

Este capítulo presenta el estado del arte de los trabajos que han abordado la predicción de la hora de llegada de autobuses entre paradas, así como los distintos tipos de algoritmos que se utilizan para realizar dicha tarea.

3.1. Trabajo Relacionado

En los últimos años, predecir con precisión los tiempos de llegada de los autobuses se ha convertido en un desafío importante en los ITS, particularmente en áreas urbanas densamente pobladas con paradas de autobús muy cercanas entre sí. Para abordar el problema de las predicciones de llegada de autobuses en tiempo real entre dos paradas cualesquiera, es esencial adoptar tecnologías avanzadas para gestionar y monitorear eficazmente los autobuses a lo largo de sus rutas. A continuación se presenta una breve reseña del trabajo relacionado.

Gurmu et al. (Gurmu y Fan, 2014) abordaron el problema de la predicción precisa del tiempo de viaje en autobuses, una necesidad fundamental para mejorar la información en tiempo real proporcionada a los usuarios del trans-

porte público. Los autores identificaron que disponer de datos confiables sobre los tiempos de llegada y salida permitía a los pasajeros planificar sus trayectos y reducir los tiempos de espera, por lo que se propusieron desarrollar un modelo dinámico de RNA capaz de predecir los tiempos de viaje utilizando exclusivamente datos GPS. Para resolver este problema, emplearon una RNA tipo perceptrón multicapa con tres capas: entrada, oculta y salida. El entrenamiento del modelo se efectuó en MATLAB, mientras que el procesamiento inicial y la gestión de datos se realizaron mediante el lenguaje R. Los autores aplicaron funciones de entrenamiento basadas en Regularización Bayesiana y en el algoritmo de retropropagación Levenberg–Marquardt con el fin de optimizar el aprendizaje de la red y reducir el sobreajuste. Previamente, normalizaron variables clave como la hora del día y el tiempo de viaje actual, con el objetivo de mejorar la estabilidad y el desempeño del modelo durante la fase de ajuste. El estudio se sustentó en un conjunto de datos extenso, compuesto por más de un millón de registros generados entre noviembre de 2008 y mayo de 2009 para los autobuses de la línea LT11 en Macaé, Brasil. El sistema de recolección se basó en tecnología de Localización de vehículos automáticos (Automatic Vehicle Location, AVL), en el cual los receptores GPS instalados en los autobuses se conectaron mediante módems GSM para transmitir posiciones georreferenciadas expresadas en pares latitud–longitud, además de registrar velocidades, fechas y horas. La ruta analizada incluyó 35 paradas y proporcionó la base empírica para evaluar el desempeño del método propuesto. El conjunto de datos se ordenó por número de semana y se dividió en tres subconjuntos: 70 % para entrenamiento, 30 %

para prueba, y de esta última fracción se extrajo un 20 % adicional para validación. La métrica empleada para evaluar la precisión del modelo fue el Error Porcentual Absoluto Medio (MAPE), seleccionada por su capacidad para medir discrepancias entre tiempos predichos y observados en términos porcentuales. Los resultados obtenidos mostraron que la RNA superó de manera consistente al método tradicional basado en promedios históricos, evidenciando menor error y una mayor capacidad para capturar la variabilidad temporal de los tiempos de viaje. En conjunto, la metodología propuesta demostró ser eficaz para predecir tiempos de viaje en sistemas de transporte que disponen únicamente de datos GPS, contribuyendo así al desarrollo de sistemas de información más precisos para los usuarios del transporte público.

Byon y Liang (Byon y Liang, 2014) abordaron el problema de la detección en tiempo real del modo de transporte, con el propósito de evaluar si los teléfonos inteligentes podían ofrecer un desempeño comparable o superior al de los registradores GPS convencionales. Los autores plantearon que la identificación automática del modo de movilidad como caminar, conducir o utilizar transporte público, requería modelos capaces de reconocer patrones dinámicos en los datos sensoriales, por lo que propusieron el uso de RNA como enfoque principal. Para resolver este problema, implementaron diferentes arquitecturas de RNA mediante el software comercial Neurosolutions, el cual permitió entrenar y evaluar los modelos bajo múltiples configuraciones. El enfoque se centró en comparar dos tipos de fuentes de datos: registradores GPS tradicionales y teléfonos inteligentes equipados con GPS asistido, acelerómetro y magnetómetro. La combinación

de sensores en los dispositivos móviles permitió capturar no solo coordenadas y velocidad, sino también información sobre aceleración y orientación, añadiendo una dimensión adicional para la detección de modos de transporte. El conjunto de datos se recopiló durante el verano de 2011 en el Área Metropolitana de Toronto, a lo largo de rutas reales que incluyeron automóvil, autobús, tranvía, bicicleta y caminata. Participaron cinco personas utilizando cinco registradores GPS y cinco teléfonos inteligentes (iPhone 4), con una frecuencia de muestreo de 1 Hz en ambos dispositivos. El estudio reunió aproximadamente 100 horas de datos en días laborables, distribuidos equitativamente: 50 horas provenientes de los GPS convencionales y 50 horas recolectadas por los teléfonos inteligentes. Para el entrenamiento de las RNA se destinó el 60 % de los datos, mientras que el 40 % restante se utilizó para evaluar el desempeño del modelo. La métrica principal empleada para medir la eficacia del método fue la tasa de detección correcta de modo de transporte, calculada para distintos escenarios y configuraciones de red. Los resultados indicaron que las RNA fueron capaces de identificar correctamente los modos de transporte incluso cuando los datos de cada modo se obtuvieron en rutas distintas, lo que mostró la capacidad del modelo para reconocer patrones físicos característicos de cada tipo de movilidad. Asimismo, los autores observaron que el desempeño mejoró de manera notable cuando las redes se ajustaron específicamente para la detección del modo automóvil, lo que sugirió que la especialización del modelo incrementaba su precisión. En conjunto, el estudio demostró que los teléfonos inteligentes, gracias a su combinación de sensores, ofrecieron un rendimiento competitivo y, en algunos casos, superior

al de los dispositivos GPS tradicionales en la tarea de detección de modos de transporte.

Julio et al. (Julio y cols., 2016) abordaron el problema de la predicción precisa de las velocidades de viaje de autobuses en tiempo real dentro del sistema Transantiago, en Santiago de Chile. Los autores señalaron que la congestión vehicular generaba incrementos y variabilidad en los tiempos de viaje, afectando la confiabilidad de los sistemas de control y favoreciendo el agrupamiento de vehículos. Para enfrentar este problema, propusieron un modelo de predicción basado en ondas de choque de tráfico combinado con algoritmos de aprendizaje automático, con el objetivo de estimar velocidades más confiables en escenarios dinámicos. El método se sustentó en la utilización de datos GPS generados por los autobuses del sistema de transporte público, cuyos dispositivos enviaban un pulso cada 30 segundos independientemente de la ubicación del vehículo. Estos pulsos fueron procesados para obtener velocidades asociadas a segmentos específicos de ruta y a intervalos temporales determinados. El enfoque incorporó dos tipos de velocidades: históricas, correspondientes a patrones registrados en fechas previas, y velocidades en tiempo real, derivadas de las condiciones actuales del tráfico. A partir de esta combinación, los autores entrenaron diversos algoritmos de aprendizaje automático, cuyos parámetros fueron ajustados empíricamente ante la ausencia de configuraciones óptimas previamente establecidas. Para evaluar el desempeño de los modelos, se emplearon tres métricas: el error cuadrático medio (MSE), la raíz del error cuadrático medio (RMSE) y MAPE. El análisis comparó los resultados de los modelos basados en aprendizaje

automático con dos enfoques de referencia comúnmente utilizados en la predicción de velocidades de viaje. Los experimentos mostraron que todos los modelos perdieron precisión bajo condiciones de congestión intensa; sin embargo, las RNAs obtuvieron el mejor desempeño relativo. En particular, el modelo basado en RNA logró mejoras de hasta un 23 % en el RMSE respecto al mejor modelo de referencia y de 3.3 % adicional frente al segundo mejor algoritmo evaluado, lo que evidenció su capacidad superior para capturar la variabilidad del tráfico en tiempo real. Los autores concluyeron que la integración de información histórica, datos GPS de alta frecuencia y modelos basados en aprendizaje automático representó una estrategia eficaz para mejorar la predicción de velocidades en sistemas de transporte urbano sujetos a congestión.

Deb et al. (Deb, Khan, Hasan, Khan, y Alam, 2019) abordaron el problema de la predicción del tiempo de viaje en contextos urbanos altamente congestionados, analizando particularmente cómo las condiciones meteorológicas afectaban la movilidad en la ciudad de Mumbai, India. Los autores señalaron que la congestión inducida por factores climáticos generaba incrementos sustanciales en los tiempos promedio de viaje, lo que dificultaba la planificación del transporte y la gestión eficiente del tráfico. Para ello, propusieron un enfoque basado en aprendizaje automático orientado a pronosticar los niveles de congestión mediante el análisis de series temporales enriquecidas con datos climáticos. El estudio empleó datos secundarios provenientes de Uber Movement, los cuales incluían registros anonimizados derivados de más de dos mil millones de viajes alrededor del mundo, y específicamente datos de la ciudad de Mumbai reco-

pilados desde 2016. La aplicación Uber Partner registraba latitud, longitud y marca temporal cada 4 segundos, proporcionando información suficientemente granular para estimar tiempos de recorrido entre nodos urbanos. En paralelo, los autores incorporaron variables meteorológicas obtenidas de Wunderground para el mismo periodo, incluyendo temperatura promedio, humedad, punto de rocío, velocidad del viento, presión atmosférica y precipitación. Los datos fueron organizados en archivos CSV y divididos en conjuntos de entrenamiento y prueba. Como primera etapa analítica, calcularon el coeficiente de correlación producto-momento entre cada parámetro climático y los tiempos promedio de viaje entre nodos, con el fin de identificar la magnitud del impacto de cada variable. Posteriormente, entrenaron varios algoritmos de aprendizaje automático, entre ellos modelos de regresión, para predecir tiempos medios de viaje. Los autores indicaron que el ajuste de parámetros se realizó de manera empírica, dado que no existían configuraciones óptimas establecidas previamente para este tipo de aplicación. Para evaluar el desempeño del modelo, utilizaron métricas como el RMSE, lo que permitió cuantificar la precisión de las predicciones. Los resultados mostraron que el tiempo promedio de viaje aumentó significativamente durante el tercer trimestre del año en comparación con el segundo, un periodo que coincidió con la temporada de lluvias intensas en Mumbai. Asimismo, el modelo de aprendizaje automático alcanzó una precisión de hasta 85%, siendo la regresión el enfoque que presentó el mejor desempeño entre los algoritmos evaluados. Los autores señalaron también que los viajes realizados por la tarde tendieron a presentar mayores tiempos de recorrido que los efectuados en ho-

ras pico tradicionales, y que ciertos días festivos generaron patrones atípicos de tráfico. En cuanto al impacto del clima, la precipitación resultó ser la variable con mayor influencia sobre los tiempos de viaje, mientras que otras variables meteorológicas mostraron efectos menos significativos. En conjunto, el estudio demostró que la integración de grandes volúmenes de datos de movilidad con información climática y modelos de aprendizaje automático permitió mejorar la predicción de tiempos de viaje en entornos urbanos complejos.

Pang et al. (Pang y cols., 2019) abordaron el problema de la predicción precisa del tiempo de llegada de autobuses en contextos urbanos donde múltiples factores como son las condiciones climáticas, la variabilidad del tráfico y la dinámica de operación, introducían incertidumbre en las estimaciones convencionales. Para enfrentarlo, los autores propusieron un modelo basado en redes neuronales recurrentes (RNN), específicamente utilizando una arquitectura de memoria a largo corto plazo (LSTM), con la finalidad de capturar dependencias espacio-temporales de largo alcance presentes en los registros históricos y en tiempo real del tránsito. El enfoque integró datos heterogéneos provenientes del sistema de transporte público de Beijing. El conjunto de datos incluyó dos tipos de información: datos dinámicos sobre la posición y operación de los autobuses, obtenidos mediante GPS con una frecuencia de muestreo promedio de un punto cada 30 segundos y un error de localización inferior a 50 metros; y datos estáticos sobre la red vial, como las ubicaciones de paradas, intersecciones, rutas y direcciones de viaje. Cada registro GPS contenía latitud, longitud, marca temporal, ID del autobús e ID de la ruta. El estudio utilizó información correspondiente a

22 días de operación, con 64 245 viajes para entrenamiento y 3 464 viajes para la fase de prueba. Para resolver el problema, los autores construyeron vectores de características que combinaban mediciones dinámicas y atributos estáticos, los cuales fueron utilizados como entrada del modelo LSTM. Durante el entrenamiento, la red ajustó sus pesos internos para aprender la relación entre patrones observados en las paradas previas y los tiempos de llegada en paradas futuras. El enfoque se formuló como una tarea de predicción de múltiples pasos adelante (MSAP), lo que requirió capturar dependencias temporales de largo alcance, una capacidad intrínseca de las LSTM frente a otros modelos tradicionales. La evaluación del desempeño se realizó mediante métricas estándar en problemas de regresión temporal, incluyendo el error absoluto medio (MAE) y RMSE. Estas métricas permitieron comparar el modelo LSTM con métodos de referencia y con otros enfoques de última generación. Los resultados mostraron que el modelo propuesto obtuvo el mejor rendimiento general, alcanzando mejoras de al menos 10% en MAE y RMSE respecto a los demás métodos. Los autores concluyeron que la integración de datos heterogéneos y la capacidad de las LSTM para modelar dependencias a largo plazo constituían una solución eficaz para predecir tiempos de llegada en sistemas de transporte urbanos complejos.

Chen et al. (Chen, Wang, Yuan, Jia, y Yao, 2020) abordaron el problema de predecir con mayor precisión el tiempo de viaje de autobuses en áreas urbanas, donde factores como la variabilidad del tráfico y las condiciones operativas introducían incertidumbre significativa en las estimaciones tradicionales. Con el fin de mejorar la confiabilidad de estas predicciones, los autores propusie-

ron un modelo basado en redes de creencias profundas (Deep Belief Networks, DBN) combinado con una etapa de retropropagación supervisada para refinar los resultados. El estudio se llevó a cabo en la ciudad de Shenyang, China, utilizando datos obtenidos durante cinco días consecutivos, del 17 al 21 de julio de 2017. El conjunto de datos se construyó mediante la integración de dos fuentes: reconocimiento automatizado de vehículos a partir de video y datos de localización automática de vehículos (AVL) registrados por unidades GPS instaladas en los autobuses. El sistema de reconocimiento visual proporcionó información sobre el número de autobús, el tiempo de permanencia en paradas y el tiempo de recorrido, mientras que los datos GPS permitieron corregir errores del proceso visual. Tras mapear ambas fuentes y eliminar valores atípicos, se obtuvieron 50 876 observaciones válidas correspondientes a la ruta número 232. Para resolver el problema, los autores emplearon una arquitectura DBN compuesta por Máquinas de Boltzmann Restringidas, utilizando una variante Gaussiana–Bernoulli para el tratamiento de datos continuos. Esta estructura permitió realizar un preentrenamiento no supervisado para extraer representaciones jerárquicas de los datos. Posteriormente, aplicaron una red neuronal con retropropagación (BP) para ajustar los parámetros y mejorar la precisión de las predicciones del tiempo de viaje. El enfoque se implementó como un modelo híbrido de profundo aprendizaje, diseñado para capturar tanto relaciones no lineales como dependencias temporales implícitas en los datos. El desempeño del modelo se evaluó comparándolo con métodos ampliamente utilizados, como k-vecinos más cercanos (k-NN), RNAs, máquinas de vectores de soporte (SVM)

y bosques aleatorios (RF). Las métricas empleadas incluyeron MAE, RMSE y MAPE, analizadas en ambas direcciones de la ruta. Los resultados indicaron que el modelo DBN superó significativamente a los demás algoritmos en términos de MAE y MAPE. En la dirección sur, el DBN obtuvo reducciones muy notables en MAPE y RMSE respecto a todos los modelos base. En la dirección norte, aunque el DBN mantuvo un buen desempeño, el modelo SVM alcanzó el mejor valor de RMSE. Aun así, de manera general, la arquitectura DBN evidenció una mejora sustancial en la precisión de la predicción del tiempo de viaje y demostró ser especialmente efectiva en la captura de variaciones complejas del tráfico urbano.

Alam et al. (Alam y cols., 2021) abordaron el problema de predecir irregularidades en los tiempos de llegada de autobuses, un desafío relevante debido a que los retrasos generaban afectaciones en actividades cotidianas como traslados a trabajos o citas programadas. Para mejorar la precisión de dichas predicciones, los autores propusieron un enfoque basado en RNN de tipo LSTM (LSTM-RNN), capaces de capturar dependencias temporales de largo plazo en series de datos espacio-temporales. El objetivo del estudio fue modelar las fluctuaciones en los tiempos de llegada mediante la integración de información histórica de operación, ubicaciones GPS, horarios, configuraciones de las rutas y datos meteorológicos. El conjunto de datos utilizado provino de la Comisión de Tránsito de Toronto. Abarcó registros AVL obtenidos cada 20 segundos a partir de unidades GPS instaladas en los autobuses, así como horarios y ubicaciones de paradas provenientes del estándar GTFS. Adicionalmente, incorporaron da-

tos meteorológicos horarios adquiridos desde una estación cercana al centro de Toronto, los cuales incluyeron variables como temperatura, precipitación, humedad y velocidad del viento. La combinación de estas fuentes permitió construir vectores de características que integraron tanto información operacional como condiciones ambientales para alimentar el modelo LSTM. Para resolver el problema, el modelo LSTM se entrenó ajustando los pesos internos para capturar dependencias de largo alcance entre las secuencias de datos, lo que permitió predecir los tiempos de llegada en múltiples paradas futuras. Con el fin de evaluar el desempeño del enfoque, los autores compararon la arquitectura LSTM con cuatro métodos de referencia ampliamente utilizados: RNAs, regresión de vectores de soporte (SVR), modelos ARIMA y el promedio histórico de tiempos de llegada. Las métricas empleadas incluyeron medidas de error como MAE y RMSE, utilizadas para cuantificar la precisión de los modelos bajo distintos escenarios. Los resultados mostraron que el modelo LSTM obtuvo la mayor precisión entre todos los métodos de comparación, debido a su capacidad para manejar dependencias temporales y variaciones abruptas en los datos. Asimismo, los autores demostraron que la incorporación de variables meteorológicas mejoró la precisión de la predicción en 48 % de los casos analizados, lo que evidenció el rol significativo del clima en la dinámica de los tiempos de llegada. Dado este hallazgo, desarrollaron un segundo modelo RNN enfocado exclusivamente en la característica climática, cuyo resultado fue combinado con el modelo LSTM principal. Este enfoque híbrido produjo una mejora adicional superior al 500 % en la precisión de las predicciones, lo que subrayó la relevancia de integrar mo-

delación climática específica en los sistemas avanzados de predicción de llegada de autobuses.

Ocampo Ortiz ([Ocampo Ortiz, 2021](#)) abordó el problema de la limitada disponibilidad de sistemas precisos de predicción de tiempos de arribo en el transporte público de la Ciudad de México, una deficiencia que afectaba la experiencia del usuario y dificultaba la planificación eficiente de los viajes. En respuesta, Ocampo Ortiz planteó el desarrollo de un sistema de predicción basado en el estándar GTFS y en datos capturados con dispositivos GPS, orientado a generar información en tiempo real confiable para los usuarios de autobús en un entorno urbano complejo. En la primera fase del estudio, el autor recopiló rutas, paradas y horarios en formato GTFS, creando una base de datos estructurada que integró información oficial de agencias de transporte. Paralelamente, desarrolló una aplicación móvil para Android que permitió georreferenciar las paradas mediante recorridos reales, utilizando el GPS del dispositivo como hardware principal de adquisición. Posteriormente, obtuvo información histórica de tiempos de llegada mediante viajes completos por dos rutas, una con once paradas y otra con diez, lo que generó un conjunto de datos empírico que combinó tiempos reales de arribo con la estructura programada del GTFS. En la segunda fase, Ocampo Ortiz implementó diversos métodos de predicción, incluyendo K-Nearest Neighbors, K-means, modelos de filtrado de Kalman y un modelo de transferencia histórica basado en cadenas de Markov. Estos algoritmos fueron integrados dentro de un sistema de software diseñado para procesar la información histórica, estructurada y georreferenciada. Cada enfoque requirió

parámetros ajustados empíricamente según las características de los datos y la dinámica de las rutas; por ejemplo, los modelos de Kalman utilizaron matrices de transición y ruido adaptadas al comportamiento de los autobuses, mientras que K-means permitió agrupar patrones de viaje para mejorar la estimación. El sistema fue validado mediante pruebas experimentales en campo, comparando las predicciones generadas por los algoritmos con los tiempos de llegada reales observados durante nuevos recorridos. La evaluación empleó métricas de error como diferencia absoluta y error porcentual, lo que permitió cuantificar el desempeño de los distintos modelos. Los resultados mostraron que el sistema tenía el potencial de mejorar significativamente la planificación del viaje al proporcionar estimaciones más cercanas a los tiempos reales observados. Asimismo, el autor señaló que la integración de sensores adicionales y el refinamiento de los algoritmos podrían incrementar la precisión y confiabilidad del sistema, lo cual abriría nuevas posibilidades para su implementación dentro de infraestructura ITS en ciudades con alta demanda de transporte público.

Hassanayebi et al. ([Hassannayebi y cols., 2023](#)) abordaron el problema de la predicción confiable del tiempo de llegada de autobuses en sistemas urbanos, donde los datos extensivos de AVL presentaban errores, interrupciones y valores faltantes que dificultaban la generación de estimaciones precisas. Los autores identificaron la necesidad de un enfoque eficiente basado en datos que mejorara la calidad de la información proporcionada a los usuarios y apoyara la operación del sistema de autobuses. En respuesta, propusieron un marco híbrido de análisis de datos fundamentado en técnicas de inteligencia artificial, cuyo

objetivo principal fue diseñar un modelo integrado capaz de predecir tiempos de llegada y estimar la fiabilidad de los recorridos. El enfoque metodológico consistió en desarrollar un modelo de predicción híbrido que combinó RNAs y SVM, optimizados mediante un algoritmo memético. Este algoritmo metaheurístico permitió ajustar de manera eficiente los parámetros del modelo para mejorar su capacidad de aprendizaje a partir de datos AVL históricos. Dado que los registros brutos incluían errores y valores faltantes debidos a fallas de transmisión, los autores aplicaron un preprocesamiento exhaustivo para depurar la base de datos y garantizar la consistencia de las series temporales empleadas en el entrenamiento. El conjunto de datos se dividió en 70 % para entrenamiento y 30 % para evaluación, permitiendo comparar el desempeño del modelo propuesto con métodos existentes como SVM y ARIMA. La implementación del sistema predictivo se llevó a cabo en el software IBM SPSS Modeler, y las pruebas se ejecutaron en una computadora equipada con un procesador Intel Core i7 a 2.4 GHz y 8 GB de RAM. Para evaluar la precisión de los modelos, los autores emplearon métricas como MAE y diversos indicadores estadísticos adicionales. Los resultados mostraron que el modelo híbrido RNA-SVM optimizado con el algoritmo memético ofreció una alta precisión en la predicción de los tiempos de llegada. En particular, el componente de RNA alcanzó un MAE aproximado de 44 segundos, un desempeño considerablemente superior al de los métodos de referencia. Tanto SVM como ARIMA registraron errores mayores, lo que evidenció la superioridad del enfoque híbrido, especialmente del modelo neuronal, en términos de confiabilidad y exactitud. En conjunto, los autores demostraron

que un modelo integrado y optimizado mediante técnicas metaheurísticas podía mejorar significativamente la previsibilidad del transporte urbano, incluso bajo condiciones de datos incompletos y ruidosos.

Yu et al. (Yu y cols., 2011) abordaron el problema de la alta variabilidad en los tiempos de viaje de los autobuses, una situación que afectaba la confiabilidad del transporte público debido a fluctuaciones en la velocidad, detenciones operativas y condiciones cambiantes del tráfico. Estas variaciones producían tiempos de llegada impredecibles y mayores esperas para los usuarios. Para enfrentar este problema, los autores desarrollaron un modelo basado en SVM, proponiendo esta técnica como alternativa eficaz frente a métodos tradicionales como el promedio histórico, ARIMA y las redes neuronales artificiales. El enfoque metodológico se centró en emplear SVM con núcleo de Función de Base Radial (RBF), el cual permitió proyectar los datos en espacios de alta dimensión y capturar relaciones no lineales entre la velocidad del autobús y el tiempo de viaje. El modelo utilizó como variable principal la velocidad registrada durante los recorridos, a partir de la cual se estimaron los tiempos futuros entre segmentos. El diseño del sistema implicó ajustar parámetros del modelo SVM para optimizar su capacidad de predicción, lo que incluyó la calibración del parámetro de penalización y del ancho del kernel RBF, elementos fundamentales para mejorar la precisión del algoritmo. Para evaluar el desempeño del modelo, los autores compararon la técnica SVM con métodos ampliamente utilizados en la predicción de tiempos de viaje, como redes neuronales artificiales, modelos ARIMA y enfoques basados en promedios históricos. Las métricas empleadas incluyeron

MAE, MAPE, RMSE y MRE, lo que permitió realizar una valoración integral de la exactitud de las predicciones bajo diferentes escenarios. Los resultados mostraron que el modelo SVM superó a los demás enfoques analizados, obteniendo los menores valores en MAE, MAPE, RMSE y MRE, lo que evidenció su capacidad para capturar patrones complejos asociados al comportamiento del tráfico. Mientras que las redes neuronales generaron resultados menos óptimos y con mayor variabilidad, SVM logró predicciones consistentemente más precisas. Además, los autores observaron que los tiempos de viaje durante horas pico fueron más predecibles, ya que la congestión redujo la variabilidad de las velocidades y estabilizó los patrones de flujo vehicular. En conjunto, el estudio demostró que SVM constituía la alternativa más efectiva para la predicción del tiempo de viaje en autobuses, destacándose por su capacidad para manejar relaciones no lineales y por ofrecer mejores resultados que los métodos tradicionales y otros enfoques de aprendizaje automático utilizados en ese momento.

Rashvand et al. (Rashvand y cols., 2024) abordaron el problema de la discrepancia existente entre los horarios programados y los tiempos reales de salida de los autobuses, una situación que generaba retrasos recurrentes, insatisfacción entre los pasajeros y disminución en la eficiencia operativa de los sistemas de transporte público. Los autores señalaron que, en promedio, las desviaciones respecto al horario oficial alcanzaban casi cuatro minutos, lo que justificó el desarrollo de un sistema de predicción en tiempo real orientado a entornos de transporte inteligente basados en IoT. Para resolver este problema, propusieron un modelo basado en una Red Neuronal Totalmente Conectada (Fully

Connected Neural Network, FCNN) diseñada para capturar relaciones no lineales complejas en datos operativos multivariados. El estudio empleó el conjunto de datos MBTA Bus Departure Times 2023, del cual se extrajo información referente a horarios de salida, secuencias de paradas, duración de los viajes y características operativas. El proceso de preparación incluyó limpieza de datos para eliminar registros incompletos o inconsistentes, así como la construcción de variables derivadas relacionadas con factores temporales, distancia, condiciones climáticas y tráfico. Los autores segmentaron los datos por viaje, abarcando rutas compuestas por entre 2 y 14 paradas, lo que permitió capturar variaciones intra-ruta y analizar patrones específicos de cada secuencia. Para resolver el problema, entrenaron una FCNN optimizada mediante el algoritmo Adam y utilizando MSE como función de pérdida. Este diseño permitió ajustar los parámetros internos de la red para predecir desviaciones en los tiempos de salida con precisión creciente durante el proceso de aprendizaje. El modelo se evaluó mediante métricas como RMSE y MAE, comparándolo con métodos de regresión tradicionales. La implementación se orientó a su operación en entornos IoT con capacidades computacionales limitadas, demostrando la viabilidad de ejecutar el modelo con eficiencia. Los resultados mostraron que el modelo propuesto superó de forma significativa los enfoques tradicionales, logrando predecir los tiempos de salida con una precisión promedio menor a 80 segundos. En total, el sistema fue probado en 151 rutas, obteniendo un RMSE de 77.83 % y un MAE de 55.63 %, valores considerablemente inferiores a los registrados por los métodos comparativos. Estas cifras evidenciaron una mejora sustancial frente a

la desviación promedio de casi cuatro minutos observada en el sistema previo, así como la capacidad del modelo para operar eficazmente en tiempo real. En conjunto, los autores demostraron la pertinencia del enfoque FCNN como herramienta para fortalecer la confiabilidad y eficiencia de los sistemas de transporte público integrados a plataformas IoT.

Yu et al. (H.Yu, Wu, Chen, y Ma, 2017) abordaron el problema de predecir de manera precisa y confiable el intervalo entre autobuses, una variable crítica para la programación eficiente de los servicios de transporte y para la provisión de información útil a los pasajeros. Los autores señalaron que los intervalos presentaban alta variabilidad debido a factores operativos y de demanda, lo que dificultaba las predicciones mediante métodos tradicionales. Para resolver este problema, propusieron un modelo basado en la Regresión de Máquina de Vectores de Relevancia (RVM), una técnica probabilística capaz de generar intervalos de confianza asociados a las predicciones, lo que representó una ventaja significativa respecto a enfoques deterministas. El estudio se sustentó en datos provenientes de tarjetas inteligentes, que aportaron información sobre la demanda de pasajeros, tiempos de viaje previos y patrones operativos asociados a las rutas analizadas. A partir de estos datos, los autores construyeron vectores de características que permitieron identificar los factores más influyentes en la variación de los intervalos entre autobuses. El modelo RVM fue entrenado para generar predicciones probabilísticas, mientras que otros cinco métodos de referencia como son SVM, GA-SVM, Filtro de Kalman (KF), K-NN y RNA, fueron implementados para comparar su desempeño. El procedimiento meto-

dológico incluyó el ajuste de los parámetros del modelo RVM, la construcción de intervalos de confianza y la evaluación del rendimiento mediante métricas de precisión y confiabilidad. Aunque el artículo no detalla el hardware o software específico utilizado, se describe un flujo completo de preprocesamiento, modelado y validación comparable con los sistemas típicos de análisis de grandes volúmenes de datos de transporte. Los resultados numéricos mostraron que el modelo RVM superó consistentemente a los demás métodos evaluados. En particular, el 97.87 % de los intervalos predichos para la Ruta A y el 98.70 % para la Ruta B se ubicaron dentro de los intervalos de confianza estimados, lo que evidenció una capacidad superior para capturar tanto la tendencia central como la incertidumbre asociada a los intervalos entre autobuses. Estos valores reflejaron una mayor confiabilidad respecto a SVM, GA-SVM, KF, K-NN y RNA, cuyos niveles de ajuste quedaron por debajo de los logrados por RVM. Los autores concluyeron que la incorporación de factores relacionados con la demanda de pasajeros y los tiempos de viaje anteriores contribuyó de manera sustancial a mejorar la predicción probabilística y la gestión operativa del servicio. En conjunto, el estudio demostró que un enfoque basado en RVM permitía predecir intervalos entre autobuses con alto nivel de confiabilidad, representando un avance relevante para sistemas de transporte que requieren información robusta para optimizar la frecuencia, reducir tiempos de espera y mejorar la satisfacción del usuario.

Arhin et al. (Arhin, Manandhar, Baba-Adam, y Gatiba, 2020) abordaron el problema de la falta de predicciones precisas de los tiempos de viaje en el sis-

tema de transporte público de Washington, D.C., una limitación que afectaba la toma de decisiones operativas y la satisfacción de los usuarios. Con el fin de mejorar la confiabilidad del servicio, los autores desarrollaron modelos basados en RNAs para estimar los tiempos de viaje de autobuses urbanos y proporcionar herramientas analíticas que apoyaran la planificación del sistema de transporte. El método propuesto consistió en la implementación de una RNA tipo perceptrón multicapa (MLP) entrenada mediante un algoritmo quasi-Newton, con el propósito de determinar la arquitectura óptima, en particular, el número de perceptrones ocultos, capaz de minimizar el error en periodos de alta demanda. Para ello, el estudio utilizó el software NeuralDesigner, una plataforma especializada en ciencia de datos y aprendizaje automático que permitió construir, entrenar y validar el modelo neuronal, además de generar las ecuaciones matemáticas aproximadas de los tiempos de viaje para distintos periodos pico. El conjunto de datos se dividió en un 75 % para entrenamiento y 25 % para prueba, permitiendo un ajuste robusto del modelo previo a su validación. El entrenamiento se realizó mediante un proceso iterativo de propagación hacia adelante y retropropagación del error, hasta cumplir con el criterio de normalización del gradiente o un límite de 1 000 épocas, garantizando así la convergencia del modelo. La arquitectura final incluyó tres capas: entrada, oculta y salida, lo que permitió capturar relaciones no lineales entre las variables operativas del sistema de autobuses. Para evaluar la eficacia del modelo, los autores utilizaron métricas como el error cuadrático normalizado (NSE), MAE y el error porcentual medio (MPE), proporcionando una evaluación integral bajo distintos enfoques

de precisión. Los resultados mostraron que los modelos basados en RNA fueron capaces de predecir eficazmente los tiempos de viaje en las rutas seleccionadas, alcanzando errores porcentuales mínimos, lo que los posicionó como alternativas superiores frente a métodos convencionales. En particular, los autores concluyeron que la arquitectura MLP entrenada con quasi-Newton ofreció estimaciones estables y precisas durante las horas pico, un periodo crítico para la planificación del transporte público debido al aumento de la demanda y la variabilidad en el tráfico. En conjunto, el estudio demostró que las redes neuronales artificiales, implementadas mediante herramientas especializadas como NeuralDesigner, constituían una solución efectiva para mejorar las predicciones de tiempos de viaje en autobuses, proporcionando una base sólida para decisiones operativas orientadas a aumentar la eficiencia y la satisfacción del usuario en sistemas de transporte urbano.

Jabamony y Shanmugavel ([Jabamony y Ganesh, 2020](#)) abordaron el problema de la predicción precisa del tiempo de llegada de autobuses en sistemas de transporte público inteligente, una tarea crítica para mejorar la confiabilidad del servicio en entornos urbanos donde el tráfico, las condiciones climáticas y factores operativos introducían alta variabilidad. Los autores señalaron que la ausencia de predicciones confiables afectaba la planificación de los usuarios y la eficiencia de los sistemas de transporte, por lo que desarrollaron un modelo basado en RNA integrado a un entorno IoT (Internet of Things), orientado a generar estimaciones en tiempo real. El estudio utilizó datos históricos recopilados durante 10 días, los cuales incluían parámetros de tráfico, clima y operación

del sistema de autobuses. El conjunto de datos se obtuvo en la ruta entre Tampuspanoor (Trivandrum) y Kazhakootum, en la India, mediante sensores GPS y sistemas de monitoreo instalados en los autobuses. La arquitectura de la red neuronal incluyó capas de entrada, ocultas y salida, con nodos interconectados que emplearon funciones de activación no lineales. El proceso de aprendizaje se realizó de forma iterativa: el sistema registró un dato por vez, generó una predicción y comparó el valor estimado con el real; posteriormente, los errores se retroalimentaron a la red para actualizar los pesos mediante retropropagación, mejorando progresivamente la precisión del modelo. Con el fin de evaluar el enfoque propuesto, los autores compararon el desempeño de la RNA con el método K-NN, incorporando de manera conjunta los factores ambientales y de tráfico que influyen en los tiempos de llegada. Las métricas utilizadas incluyeron MAE, RMSE y MAPE. Los resultados numéricos evidenciaron que el sistema basado en RNA superó significativamente al método K-NN. El modelo neuronal alcanzó valores de MAE=134.58, RMSE=197.73 y MAPE=28.87, reflejando una mayor capacidad para capturar la variabilidad del tráfico y las condiciones ambientales respecto a la técnica comparativa. Estos valores confirmaron que la RNA ofrecía una predicción más precisa del tiempo de llegada de los autobuses en el corredor analizado, destacando su utilidad para mejorar la puntualidad, la planificación del viaje por parte de los usuarios y la eficiencia operativa en sistemas de transporte público inteligentes basados en IoT.

Al-Naim y Lytkin (Al-Naim y Lytkin, 2021) abordaron el problema de la baja precisión en las predicciones de tiempos de viaje para autobuses urba-

nos, una limitación relevante para los sistemas avanzados de transporte público (APTS) que buscan proporcionar información en tiempo real a los usuarios. Los autores señalaron que los métodos basados en promedios históricos no podían capturar la variabilidad del tráfico ni los cambios repentinos en las condiciones operativas, por lo que propusieron el desarrollo de un modelo dinámico de RNA entrenado con datos GPS, capaz de actualizar las predicciones para paradas posteriores con mayor exactitud. El conjunto de datos utilizado fue proporcionado por la empresa StarLine Ltd. en la ciudad de San Petersburgo, Rusia, e incluyó 298 089 trayectorias registradas durante un día laboral mediante dispositivos GPS instalados en autobuses. Cada trayectoria contenía información del vehículo como ubicación geográfica, velocidad, marca temporal e identificador único, con longitudes de hasta dos kilómetros. La duración real de cada trayecto se calculó a partir de las marcas temporales del primer y último punto GPS. Estos datos se emplearon para construir un conjunto de características adecuado para los modelos predictivos. La metodología consistió en desarrollar y comparar múltiples enfoques de aprendizaje automático, entre ellos RNA, SVR, Random Forest Regression (RFR), Gradient Boosting Regression (GBR), así como el modelo profundo DeepTTE y un motor de enrutamiento comercial (Valhalla). El entrenamiento de todos los modelos se realizó de manera off-line, mientras que su aplicación se orientó a estimaciones en tiempo real para paradas posteriores. Los autores destacaron la importancia de seleccionar cuidadosamente las variables de entrada y salida para mejorar la precisión del modelo, especialmente dada la variabilidad de los datos GPS. Para evaluar el desempeño del

sistema, se utilizaron métricas como MAE, MAPE y la desviación estándar, lo que permitió valorar tanto la precisión como la consistencia de las predicciones. Los resultados mostraron diferencias claras entre los modelos comparados. El método SVR obtuvo el mejor desempeño global, logrando un MAE de 91.65 y un MAPE inferior al 30 %, mientras que RFR registró la menor desviación estándar, indicando mayor estabilidad en sus estimaciones. Los modelos GBR y RNA alcanzaron resultados satisfactorios y similares entre sí, aunque inferiores a los de SVR. El modelo profundo DeepTTE presentó el rendimiento más bajo, lo que los autores atribuyeron a las limitaciones del conjunto de datos disponible. Por su parte, el motor de enrutamiento Valhalla tendió a subestimar sistemáticamente los tiempos de viaje. En conjunto, el estudio demostró que los enfoques basados en aprendizaje automático, particularmente SVR, superaron a los métodos tradicionales basados en promedios históricos y al motor de enrutamiento de referencia, proporcionando predicciones más precisas y robustas para sistemas APTS y mejorando la calidad de la información en tiempo real ofrecida a los usuarios.

Ford (Ford, 2020) abordó el problema de la falta de precisión en los tiempos de llegada informados a los usuarios del transporte público en la ciudad de Reading, donde los horarios programados no reflejaban las condiciones reales de operación. Esta discrepancia generaba incertidumbre entre los pasajeros y afectaba la eficiencia del sistema. Con el objetivo de mejorar la calidad de la información proporcionada a los usuarios, el autor desarrolló un sistema de predicción basado en ciencia de datos y técnicas de aprendizaje automático,

capaz de estimar los tiempos de llegada utilizando datos históricos reales. El estudio empleó datos abiertos proporcionados por Reading Open Data Service (RODS), obtenidos mediante una API, que incluían información detallada sobre rutas, paradas, horarios planificados, tiempos reales de llegada y salida, así como variables espaciales y temporales asociadas a cada recorrido. Estos datos permitieron construir un conjunto robusto de características representativas del comportamiento del sistema de transporte. El proceso metodológico comprendió varias etapas: recopilación de datos mediante la API, limpieza para eliminar valores atípicos e inconsistencias, preprocesamiento y visualización exploratoria para identificar patrones relevantes antes del modelado predictivo. El diseño del sistema de predicción se basó en técnicas de aprendizaje supervisado, para lo cual el conjunto de datos se dividió en subconjuntos de entrenamiento y prueba. Ford entrenó varios modelos de aprendizaje automático, comparando su desempeño con enfoques tradicionales basados en promedios históricos. Aunque el estudio no especificó hardware particular, su implementación siguió un flujo estándar de análisis de datos reproducible en plataformas comunes de ciencia de datos. Para evaluar el rendimiento de los modelos, el autor utilizó principalmente MSE, métrica que permitió cuantificar la diferencia entre los tiempos de llegada reales y los estimados. Esta medida proporcionó una base objetiva para comparar la precisión de los modelos predictivos frente a los métodos tradicionales. Los resultados mostraron que los modelos de aprendizaje automático ofrecieron estimaciones significativamente más precisas y consistentes que los enfoques basados en promedios históricos. En particular, Ford evidenció que la

capacidad de los modelos para captar patrones temporales, variaciones operativas y tendencias no lineales permitió mejorar la predicción de los tiempos de llegada. Asimismo, el uso de datos abiertos y la integración de modelos avanzados contribuyeron a la construcción de un sistema más confiable para los usuarios y más útil para la gestión del transporte urbano. En conjunto, el estudio demostró que la aplicación de ciencia de datos y aprendizaje automático constituía una alternativa eficaz para optimizar la predicción de los tiempos de llegada en sistemas de transporte público, promoviendo una movilidad urbana más eficiente, orientada a las necesidades del usuario y acorde con los principios de las ciudades inteligentes.

Johansson ([Johansson, 2019](#)), en su tesis de maestría, abordó el problema de la baja precisión en los sistemas tradicionales de predicción de tiempos de llegada de autobuses, los cuales se basaban principalmente en promedios históricos y modelos estadísticos incapaces de capturar la alta variabilidad del tráfico urbano. Esta limitación afectaba particularmente al sistema de transporte público de Östgötatrafiken (ÖT), en el condado de Östergötland, Suecia, donde las estimaciones existentes no satisfacían las necesidades de los usuarios. Con el fin de mejorar la confiabilidad de la información proporcionada a los pasajeros, el autor propuso el desarrollo de modelos basados en RNN y LSTM, debido a su capacidad para modelar dependencias temporales complejas en series de datos secuenciales. Para resolver el problema, Johansson empleó datos reales provenientes de los sistemas de AVL de ÖT, generados por transmisores GPS con una frecuencia de 1 Hz. Estos datos fueron estructurados en formatos GTFS y

GTFS-Realtime, incorporando tanto información estática (horarios programados, estructura de rutas, identificadores de paradas) como dinámica (posiciones geográficas, tiempos reales de llegada y salida, variables temporales como tiempo transcurrido desde medianoche y dirección del recorrido). La metodología se fundamentó en aprendizaje supervisado con series temporales, a partir de las cuales se generaron secuencias de pasos temporales que combinaban variables dinámicas y atributos estáticos relevantes para la predicción del tiempo restante de viaje. Antes del entrenamiento, los datos fueron normalizados para mejorar la estabilidad del aprendizaje. Los modelos se implementaron utilizando arquitecturas LSTM, entrenadas mediante el optimizador Adam, con mini-lotes de tamaño 32 y aplicación de dropout del 30% para evitar sobreajuste. Se evaluaron también distintas funciones de pérdida basadas en el error cuadrático medio para analizar su efecto en la precisión y utilidad práctica para distintos perfiles de usuarios. El desempeño de los modelos se midió utilizando métricas ampliamente aceptadas en la literatura, tales como MSE, RMSE, MAE y MAPE, así como métricas modificadas que incorporaron un margen de tolerancia de ± 1 minuto para identificar predicciones correctas, subestimadas o sobreestimadas en función de las necesidades reales de los pasajeros. Los resultados mostraron que los modelos basados en RNN con LSTM superaron consistentemente al sistema de predicción existente, obteniendo valores inferiores en todas las métricas de error utilizadas y logrando una proporción significativamente mayor de predicciones correctas en múltiples rutas analizadas. El estudio evidenció además que la selección adecuada de la función de pérdida influyó directamente

en el equilibrio entre precisión numérica y confiabilidad percibida por diferentes usuarios. En conjunto, la investigación demostró que las LSTM constituían una solución eficiente y robusta para mejorar la predicción de los tiempos de llegada en sistemas de transporte urbano, aportando información más precisa y útil para los pasajeros, además de sentar bases sólidas para el desarrollo de sistemas avanzados de movilidad inteligente.

Wai y Zhou (Wai y Zhou, 2020) abordaron el problema de generar predicciones precisas y en tiempo real de los tiempos de salida de autobuses, una tarea clave para mejorar la confiabilidad de los sistemas de transporte público, donde los horarios programados suelen diferir de las condiciones reales de operación. Con el fin de resolver esta limitación, los autores propusieron un enfoque basado en modelos de aprendizaje automático altamente granulares, entrenados por segmento de ruta y por parada, incorporando variables temporales, meteorológicas y operativas. El propósito central consistió en estimar sucesivamente los tiempos de permanencia en parada (*dwell time*) y los tiempos de recorrido (*run time*) para generar predicciones precisas de salida para cada autobús. El método se estructuró en dos tipos de modelos: uno dedicado a predecir el *dwell time* y otro al *run time*, cada uno con dos etapas. En la primera etapa, se estimó la desviación entre el tiempo programado y el tiempo real; en la segunda, dicha desviación se integró como una nueva característica para calcular la predicción final. Con el fin de asegurar la robustez estadística del entrenamiento, los autores establecieron un umbral mínimo de 500 instancias de datos por segmento o por parada. El conjunto de datos combinó información AVL,

bases de datos históricas de machine learning, información operativa de rutas y paradas, y datos meteorológicos en vivo, integrados mediante un flujo de datos de alta velocidad diseñado para aplicaciones en tiempo real. Para integrar las predicciones individuales, los autores desarrollaron un algoritmo agregador que combinó las estimaciones de run time y dwell time para producir los tiempos de salida predichos (PBDTs). Aunque el sistema fue diseñado para ser independiente del modelo subyacente, Wai y Zhou seleccionaron XGBoost debido a su velocidad de entrenamiento y robustez. Sus hiperparámetros fueron ajustados mediante Grid Search, mientras que la arquitectura general se implementó con un patrón de diseño basado en colas de trabajo y uso intensivo de caché, lo que permitió alcanzar altos niveles de rendimiento computacional y resiliencia operativa. A nivel de infraestructura, los autores describieron un sistema capaz de procesar más de 20 000 predicciones por minuto, atendiendo simultáneamente múltiples autobuses y paradas en tiempo real. El sistema incluyó mecanismos de autoescalado para responder a variaciones en la demanda y detectó deriva en los modelos, activando procesos automáticos de reentrenamiento con datos recientes. Las métricas empleadas para evaluar el desempeño incluyeron los errores absolutos y relativos utilizados comúnmente en predicción de tiempos de viaje, aunque el estudio destacó principalmente la estabilidad y precisión de las predicciones integradas. Los resultados indicaron que el enfoque logró predicciones altamente precisas y consistentes, superando ampliamente las limitaciones de métodos tradicionales basados en promedios históricos o modelos globales. Los autores demostraron que su sistema permitía generar tiempos de salida en

tiempo real adecuados a la complejidad real del transporte público urbano, articulando modelos granulares basados en XGBoost con una arquitectura escalable y confiable. En conjunto, el estudio proporcionó un marco sólido para el desarrollo de sistemas avanzados de predicción en transporte inteligente, capaces de responder a escenarios de operación dinámicos y heterogéneos.

Armita et al. (Amitya, Jain, y Garg, 2016) abordaron el problema de la falta de precisión en la estimación del tiempo de viaje de los autobuses, un elemento fundamental para garantizar un servicio eficiente y confiable en sistemas de transporte urbano. Los autores señalaron que los modelos tradicionales no incorporaban adecuadamente la variabilidad derivada del tráfico, las detenciones operativas ni el comportamiento de los pasajeros, lo que afectaba la calidad de las predicciones. Para resolver este problema, propusieron un modelo basado en RNA apoyado en datos provenientes de GPS y observaciones directas del sistema de autobuses de Delhi. El estudio se desarrolló en la ruta 832 de la Delhi Transport Corporation, con una longitud de 17.4 km y 43 paradas. La recolección de datos se realizó durante seis días de octubre de 2013, abarcando horas punta y valle. Se utilizaron dispositivos GPS portátiles para registrar tiempos de llegada, salida y detención en cada parada, mientras que los ascensos y descensos de pasajeros se contabilizaron manualmente, generando información sobre la carga del autobús, el tiempo promedio sin movimiento y los retrasos por detenciones. En total, se obtuvieron 40 viajes, 20 en hora pico y 20 en hora valle. Un análisis ANOVA confirmó que no existían diferencias significativas entre ambos periodos, permitiendo unificar el conjunto de datos. Para resolver el problema

de predicción, los autores implementaron tres enfoques, *i*) una RNA utilizando datos GPS y de sistemas AVL; *ii*) un modelo de regresión para estimar tiempos entre paradas; y *iii*) un filtro de Kalman para predecir tiempos de llegada y salida. La arquitectura RNA incluyó capas de entrada, ocultas y salida, ajustadas para capturar relaciones no lineales entre velocidad, flujo de pasajeros, detenciones y distancia. El proceso de entrenamiento empleó retropropagación del error y técnicas de ajuste supervisado, apoyándose en herramientas de análisis estadístico para validar el rendimiento. El desempeño de los modelos se evaluó mediante métricas ampliamente utilizadas como MAE, RMSE y MAPE, lo que permitió comparar la precisión entre las técnicas. Los resultados mostraron que el modelo basado en RNA presentó menor error que los enfoques de regresión y filtro de Kalman, evidenciando una mayor capacidad para representar la variabilidad operacional del sistema de autobuses en Delhi. En particular, la RNA logró capturar con mayor detalle los efectos de los ascensos y descensos de pasajeros y de las detenciones en paradas, mejorando la estimación del tiempo de viaje entre segmentos. En conjunto, el estudio demostró que las redes neuronales artificiales constituían una herramienta eficaz para predecir el tiempo de viaje en rutas urbanas complejas, superando a métodos estadísticos tradicionales y ofreciendo una base sólida para el desarrollo de sistemas de transporte inteligente basados en datos reales de operación.

La literatura revisada muestra que la predicción del tiempo de llegada de autobuses es un problema relevante en el transporte público urbano, debido a la variabilidad de las condiciones de operación y a las limitaciones de los métodos

tradicionales basados en promedios históricos. Estas limitaciones han impulsado el uso de técnicas de aprendizaje automático como alternativa para mejorar la precisión de las estimaciones. En particular, las redes neuronales artificiales han demostrado una alta capacidad para modelar relaciones no lineales y dependencias temporales, logrando mejores resultados en la predicción del tiempo de llegada entre paradas cuando se utilizan datos históricos reales. En consecuencia, se justifica la aplicación de estos modelos al transporte público de Cuernavaca, con el fin de adaptar las metodologías existentes a un contexto local y mejorar la confiabilidad de la información para los usuarios. Este capítulo presenta y analiza metodologías propuestas en la literatura, asimismo, se identifican sus principales características y resultados, y se establece un marco de referencia para el desarrollo del presente estudio.

A continuación, el siguiente capítulo describe la metodología de solución.

Capítulo 4

Metodología propuesta

El presente capítulo describe la metodología de solución propuesta para la predicción del tiempo de llegada entre paradas de autobuses de transporte público de Cuernavaca mediante redes neuronales artificiales, así como la recolección de datos del recorrido de la ruta número 13.

4.1. RNA para problemas de regresión

El aprendizaje automático (ML) es un campo de la inteligencia artificial que se centra en el desarrollo de algoritmos capaces de aprender directamente de los datos (Barbierato y Gatti, 2024). Además, el ML emplea técnicas estadísticas para identificar patrones en los datos, lo que permite su aplicación en casos reales. Una RNA es un modelo computacional no lineal inspirado en la estructura y función de las redes neuronales biológicas del cerebro humano (McCulloch y Pitts, 1943). En este contexto, una RNA es una técnica de ML que consta de capas de nodos interconectados, conocidos como neuronas. Las neuronas de una RNA procesan los datos aprendiendo patrones y relaciones, lo que da como

resultado un sistema dinámico, adaptable y cada vez más inteligente (Bishop, 2006). El mecanismo de aprendizaje de las RNA proporciona un método eficaz para entrenar redes de propagación hacia adelante multicapa mediante el algoritmo de aprendizaje supervisado conocido como retropropagación (Haykin, 1999). De esta manera, las técnicas ANN son prácticas para resolver problemas de regresión, donde el objetivo es predecir una salida continua en situaciones complejas del mundo real que exhiben relaciones no lineales entre entradas y salidas (Hornik, Stinchcombe, y White, 1989).

De acuerdo con Goodfellow et al. (Goodfellow, Bengio, y Courville, 2016), las RNAs son modelos computacionales compuestos por múltiples capas de unidades interconectadas, conocidas como neuronas artificiales. Las capas operan de manera jerárquica, permitiendo que el modelo aprenda representaciones progresivamente más complejas de los datos a través de un proceso de entrenamiento. Durante el proceso de entrenamiento, la RNA ajusta sus parámetros internos con el objetivo de minimizar el error entre las salidas estimadas y los valores reales, lo que le otorga la capacidad de identificar patrones de alta complejidad. Gracias a estas características, las redes neuronales artificiales han sido ampliamente utilizadas en diversas áreas, tales como la visión por computadora, el procesamiento del lenguaje natural y la predicción de series temporales.

En el contexto de los ITS, las RNAs se han consolidado como una herramienta eficaz para abordar problemas asociados a la movilidad urbana, en los que intervienen múltiples factores variables como el tiempo, la distancia, las condiciones del tráfico y, en algunos casos, el clima. Un ejemplo representativo

es el estudio realizado por Rashvand et al. (Rashvand y cols., 2024), quienes desarrollaron un modelo neuronal totalmente conectado para predecir en tiempo real los horarios de salida de autobuses. Los resultados obtenidos mostraron un desempeño superior en comparación con métodos convencionales, evidenciando el potencial de este tipo de modelos para mejorar la eficiencia operativa, la fiabilidad del servicio y la satisfacción de los usuarios del transporte público.

4.2. RNA para estimar el tiempo de llegada

Esta sección describe de manera detallada el proceso metodológico que se sigue para el desarrollo del modelo de RNA propuesto. El proceso abarca desde la recolección de los datos obtenidos mediante trabajo de campo en la ruta de estudio, hasta su depuración y preparación para el análisis, con el objetivo de garantizar la calidad y consistencia de la información utilizada. Asimismo, la sección explica el diseño de la arquitectura de la RNA implementada.

4.2.1. Fuente de datos

La presente investigación se sustentó en el análisis sistemático de datos relacionados con la operación del transporte público urbano en la ruta 13 de Cuernavaca, Morelos, México, con el propósito de emplearlos posteriormente en el diseño, entrenamiento y validación de la RNA. La información fue recopilada de manera presencial mediante un procedimiento estructurado de observación y registro, cuyo objetivo consistió en identificar las variables que influyen de manera directa o indirecta en los tiempos de desplazamiento entre paradas consecutivas.

Este proceso consideró tanto las características intrínsecas del recorrido como factores operativos y ambientales observables en campo.

Para la recolección de datos se diseñó un formato estructurado en forma de tabla, impreso y empleado como bitácora de campo durante los recorridos. Este instrumento permitió registrar de manera ordenada variables como fecha, día de la semana, parada de origen y destino, hora de salida y llegada, velocidad promedio (km/h), número de topes y semáforos, tipo de día, turno, distancia recorrida, coordenadas geográficas (latitud y longitud), número de pasajeros que abordaron y descendieron, condiciones climáticas e imprevistos operativos ocurridos durante el trayecto. A partir de estos registros se conformó una base de datos depurada y estructurada que constituyó el insumo principal para el desarrollo del modelo predictivo.

En total, se efectuaron 26 recorridos entre el 1 de octubre de 2024 y el 20 de febrero de 2025, obteniéndose 2 073 registros derivados del trabajo de campo. La ruta analizada inicia en la Universidad Autónoma del Estado de Morelos (UAEM) y concluye en Walmart Jiutepec, considerando 37 paradas en el sentido de ida y 41 paradas en el sentido de regreso. Los recorridos se realizaron en distintos días de la semana, de lunes a domingo, y en dos turnos operativos, matutino (6:00 a 12:00 h) y vespertino (12:00 a 20:00 h), con el fin de capturar variaciones horarias representativas en el comportamiento del servicio.

Los datos fueron inicialmente registrados en papel mediante la bitácora de campo y posteriormente digitalizados en una hoja de cálculo para su organización, verificación y análisis preliminar. Las paradas de origen y destino se utili-

zaron como identificadores nominales, mientras que las coordenadas geográficas oficiales de cada parada se obtuvieron mediante la herramienta Google Maps. Las horas de salida y llegada correspondieron al momento exacto en que el autobús partió de una parada y arribó a la siguiente. La velocidad promedio se registró empleando la aplicación Waze, tomando al menos tres mediciones por tramo. El número de topes y semáforos se consideró constante para cada segmento del recorrido. Finalmente, las condiciones climáticas, clasificadas como soleado, nublado o lluvia, y los imprevistos, como accidentes, reparaciones viales o bloqueos, fueron registrados directamente en campo.

Este conjunto de datos permitió construir un marco de referencia sólido para el modelado del tiempo de viaje mediante redes neuronales artificiales, al integrar tanto variables operativas del sistema como factores contextuales que afectan la dinámica del transporte público urbano.

4.2.2. Preparación de los datos

Los datos recopilados durante el trabajo de campo fueron sometidos a un proceso exhaustivo de preprocesamiento con el propósito de garantizar su consistencia, integridad y adecuación para el entrenamiento del modelo de red neuronal artificial. Los registros originales fueron almacenados en un archivo con formato CSV (Comma-Separated Values), lo que facilitó su manipulación mediante herramientas computacionales, así como su posterior análisis, transformación y validación.

En una primera etapa, se llevó a cabo la limpieza del conjunto de datos,

identificando y corrigiendo errores de captura, inconsistencias sintácticas, valores duplicados y registros que no cumplían con criterios mínimos de calidad. Asimismo, se revisó la presencia de valores faltantes, nulos o negativos, los cuales fueron eliminados cuando comprometían la coherencia del análisis o la estabilidad del modelo. Este proceso permitió depurar la base de datos y asegurar que únicamente se conservaran observaciones fiables para la etapa de modelado. Posteriormente, se realizaron transformaciones sobre las variables temporales, convirtiendo las horas de salida y llegada a unidades de tiempo expresadas en segundos absolutos, mediante la descomposición estándar de horas, minutos y segundos. Esta homogeneización permitió trabajar con una escala numérica continua y fue fundamental para el cálculo preciso de los tiempos de recorrido entre paradas consecutivas.

La distancia entre paradas se obtuvo mediante mediciones realizadas con la herramienta Google Maps, a partir de las cuales se calculó la velocidad real del autobús (km/h) utilizando la relación entre distancia y tiempo registrado en campo. Asimismo, el tiempo total entre paradas se definió como la suma del tiempo de recorrido más el tiempo de detención correspondiente a la permanencia del autobús en cada parada.

Una vez estructurada la información, se aplicó un proceso de normalización mediante el método Min-Max, que escaló cada variable numérica dentro del intervalo $[0,1]$ utilizando la expresión:

$$\text{variable normalizada} = \frac{\text{valor original} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}}$$

Esta técnica permitió preservar las relaciones proporcionales entre los datos y evitar que variables con magnitudes superiores influyeran de manera desproporcionada en el proceso de aprendizaje del modelo. Para garantizar la coherencia de la escala, los valores mínimos y máximos de referencia fueron establecidos a partir de las observaciones capturadas en campo y de las condiciones operativas del transporte urbano. Por ejemplo, la variable día se acotó en un rango de 1 a 7; la velocidad real se delimitó entre 0 y 80 km/h; el número de topes y semáforos entre 0 y 5; los tiempos de recorrido y tiempo total entre 0 y 500 segundos; y la distancia entre paradas dentro del intervalo de 0 a 1 500 metros. Estos rangos evitaban distorsiones generadas por valores extremos y permitieron una normalización adecuada. Las variables binarias, como tipo de día, turno, clima e imprevistos, se conservaron en su representación numérica original (0 y 1), al no requerir un proceso adicional de reescalamiento.

Posteriormente, se realizó un análisis exploratorio de datos (EDA) mediante visualizaciones gráficas individuales, con el objetivo de identificar valores atípicos, verificar la coherencia de las distribuciones y examinar patrones generales en cada variable.

Una vez realizado el análisis EDA, los datos disponibles se preprocesaron para identificar y eliminar registros erróneos. Los datos sin procesar se capturaron y almacenaron en un formato de archivo de valores separados por comas. El preprocesamiento de datos implicó la identificación de entradas duplicadas, la corrección de errores de captura y de sintaxis, la corrección de campos vacíos y de valores negativos o nulos. El preprocesamiento garantizó la consistencia de

los datos mediante la normalización, el escalado de los datos dentro de un rango específico o la aproximación uniforme de los valores dentro de dicho rango. La Tabla 4.1 proporciona un resumen de las estadísticas de cada variable numérica en el conjunto de datos.

Tabla 4.1: Estadísticas resumidas de las variables de entrada de la RNA

#	Variable	Media	Mediana	Máx	Mín
1	Hora de salida (seg)	44 480.73	45 559.00	70 286.00	22 622.00
2	Hora de llegada (seg)	44 552.72	45 629.00	70 330.00	22 705.00
3	Distancia recorrida (m)	394.92	388.39	1 150.00	74.27
4	Tiempo de viaje (seg)	71,98	54,00	557,00	7,00
5	Tiempo de parada (s)	13,89	5,00	278,00	0,00
6	Velocidad real (km/h)	7,00	6,38	19,82	0,31
7	Reductores de velocidad	0,23	0,00	2,00	0,00
8	Semáforos	0,67	0,00	3,00	0,00
9	Tipo de día (laborable/fin de semana)	0,70	1,00	1,00	0,00
10	Turno (mañana/tarde)	0,53	1,00	1,00	0,00
11	Clima (frío/cálido)	0.04	0.00	1.00	0.00
12	Eventos inesperados	0.016	0.00	1.00	0.00

El procesamiento de las variables de datos dentro del conjunto de datos es el siguiente. En la Tabla 4.1, los tiempos de salida (t_{salida}) y llegada ($t_{llegada}$) entre dos paradas de autobús se convirtieron a segundos (s) mediante las siguientes ecuaciones. Para el tiempo total de salida, la ecuación fue $t_{salida} = hora_{salida} * 3600 + minutos_{salida} * 60 + segundos_{salida}$, y para el tiempo total de llegada, la ecuación utilizada fue $t_{llegada} = hora_{llegada} * 3600 + minutos_{llegada} * 60 + segundos_{llegada}$. La distancia (d) entre dos paradas de autobús cualesquiera se midió utilizando Google Maps. Los tiempos de viaje (t_{viaje}) y parada (t_{parada}) del autobús se midieron en persona. Esto implicó recopilar las marcas de tiempo de cuándo el autobús viajó entre paradas, así como el momento en que se detuvo

y comenzó a moverse de nuevo en cada parada. La velocidad real (v) se calcula en km/h utilizando la fórmula $v = d/t_{viaje}$. Los parachoques, semáforos y eventos inesperados cuentan el número de reductores de velocidad, semáforos y cualquier incidente en la trayectoria del camino. El tipo de día clasifica el día según si los datos recopilados corresponden a un día laborable o a un fin de semana. El turno indica si la información recopilada corresponde a la mañana o a la tarde. Por último, el tiempo se refiere a la temperatura de ese día, distinguiendo entre condiciones frías y cálidas.

Se analizaron las variables del conjunto de datos para identificar los cambios más significativos en las diferentes variables. La figura 4.1 muestra la matriz de correlación de las variables de entrada utilizadas en el modelo de RNA.

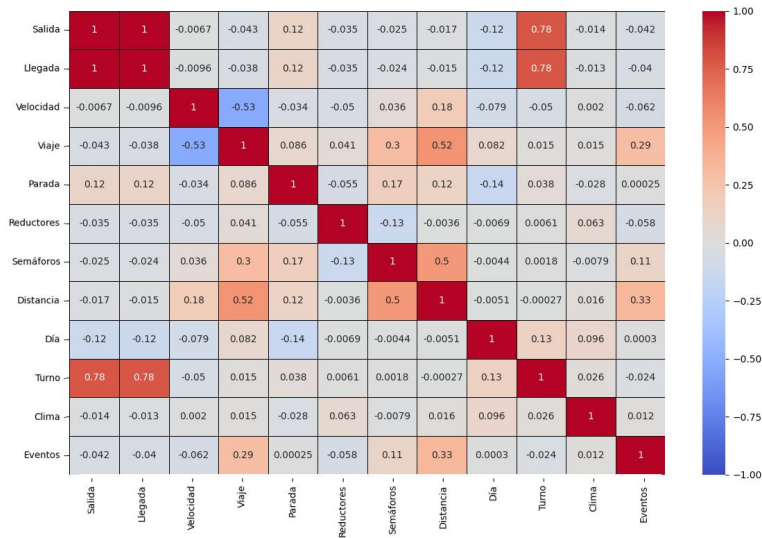


Figura 4.1: Matriz de correlación de los datos de entrada

La matriz de correlación de la Fig. 4.1 reveló una correlación de Pearson positiva perfecta (r) entre la hora de salida y la hora de llegada ($r=1,0$), lo

que indica que los viajes que comienzan más tarde también tienden a terminar más tarde. La variable “Turno” mostró una fuerte correlación positiva tanto con la hora de salida como con la hora de llegada ($r=0,78$), lo que sugiere que los diferentes “Turnos” se asocian con ventanas de tiempo específicas. Además, el tiempo de viaje y la velocidad real mostraron una correlación negativa ($r=-0,53$), lo que indica que las velocidades más altas se asocian con tiempos de viaje más cortos. Asimismo, la distancia se correlaciona positivamente tanto con el tiempo de viaje ($r=0,52$) como con el número de semáforos ($r=0,50$), lo que indica que las rutas más largas generalmente toman más tiempo y pasan por más intersecciones. Finalmente, el tiempo total de viaje (t_{total}) entre dos paradas de autobús cualesquiera se calcula en segundos mediante la ecuación $t_{total} = t_{viaje} + t_{parada}$. El tiempo total de viaje corresponde al vector de salida utilizado para entrenar el modelo ANN con el conjunto de datos etiquetado.

4.2.3. Definición de variables del modelo

Una vez concluido el proceso de preparación, depuración y normalización del conjunto de datos, se definieron las variables que integran el modelo de RNA propuesto. El conjunto definitivo quedó conformado por doce variables de entrada (variables independientes) y una variable de salida (variable dependiente), seleccionadas por su relevancia para explicar la variabilidad en los tiempos de desplazamiento entre paradas consecutivas en la ruta analizada.

Variables de entrada (x). Las variables de entrada representan las condiciones temporales, operativas y contextuales del recorrido entre dos paradas sucesivas. El conjunto se compone de los siguientes atributos:

- Hora de salida (seg)
- Hora de llegada (seg)
- Distancia recorrida (m)
- Tiempo de viaje (seg)
- Tiempo de parada (s)
- Velocidad real (km/h)
- Reductores de velocidad
- Semáforos
- Tipo de día (laborable/fin de semana)
- Turno (mañana/tarde)
- Clima (frío/cálido)
- Eventos inesperados

Las variables de naturaleza numérica fueron previamente normalizadas con el método Min–Max, con el fin de mantenerlas dentro de un rango común [0,1] y evitar que diferencias en magnitud afectaran desproporcionadamente el proceso de entrenamiento del modelo. Por su parte, las variables categóricas binarias

(tipo de día, turno, clima e imprevistos) se conservaron en su codificación original (0–1), ya que este formato resulta compatible con el proceso de aprendizaje y no requiere transformaciones adicionales.

Variable de salida (y) La variable dependiente corresponde al tiempo total requerido por el autobús para desplazarse entre dos paradas consecutivas. Este valor incluye tanto el tiempo de movimiento como el tiempo de detención, y constituye el parámetro que el modelo busca estimar con base en las características operativas y contextuales registradas en cada tramo. Al igual que las variables numéricas independientes, el tiempo total fue normalizado para mantener consistencia en la escala de los datos.

En síntesis, el modelo de red neuronal artificial fue estructurado para aprender la relación no lineal existente entre las doce variables de entrada y el tiempo total de viaje entre paradas, con el fin de generar predicciones robustas y precisas para el sistema de transporte público de la ruta 13 de Cuernavaca. La definición rigurosa de las variables garantizó que el modelo contara con información representativa y adecuadamente preprocesada para su entrenamiento, mejorando así su capacidad de generalización y desempeño predictivo.

4.2.4. RNA para la estimación del tiempo de llegada

El algoritmo de regresión de aprendizaje supervisado propuesto basada en RNA se entrena y valida utilizando datos etiquetados que indican el tiempo total de viaje entre dos paradas de autobús. El algoritmo propuesto consta de

cuatro pasos:

1. Determinar el tiempo total de viaje entre todas las paradas de autobús a lo largo de la ruta 13.
2. Crear un conjunto de datos etiquetados basado en la información recopilada.
3. Construir un vector con 12 variables reportadas en la Tabla 4.1. El vector se organiza en una matriz, donde cada fila representa las condiciones de un viaje realizado, para entrenar el algoritmo de regresión de aprendizaje supervisado.
4. Proponer una RNA como algoritmo de regresión. La capa de entrada contiene 12 neuronas, correspondientes a las 12 características utilizadas para simular las condiciones del tiempo de viaje, lo que permite estimar los tiempos de llegada a cualquier parada de autobús. La RNA incluye de tres a cuatro capas ocultas, cada una con un número variable de neuronas, que utilizan la función de activación de la Unidad Lineal Rectificada (ReLU). La capa de salida consta de una sola neurona con una función de activación ReLu para estimar el tiempo de llegada a una parada de autobús.

La figura 4.2 muestra un diagrama de flujo del algoritmo de regresión supervisada propuesto.

La figura 4.2 muestra los cuatro pasos empleados por el algoritmo de regresión de aprendizaje supervisado. El modelo de RNA toma 12 características

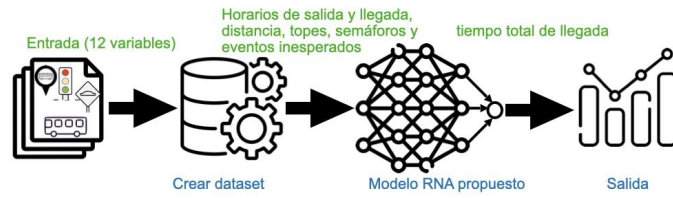


Figura 4.2: Diagrama de flujo del algoritmo de regresión supervisada

como entradas, que se describen en la tabla 4.1. Las doce características de entrada se estandarizaron durante el entrenamiento y las pruebas, además de aplicar el método de normalización para mejorar la convergencia del modelo y evitar que variables con rangos numéricos mayores opaquen los resultados.

La siguiente sección presenta el análisis experimental y analiza los resultados.

Capítulo 5

Resultados experimentales y discusión

En este Capítulo, se presenta y discute el análisis experimental del enfoque propuesto para la estimación del tiempo de llegada.

5.1. Plataforma de desarrollo y ejecución

El algoritmo de regresión para aprendizaje supervisado se implementó en Python, utilizando bibliotecas científicas como Pandas, Matplotlib y NumPy. El método de RNA se desarrolló, entrenó y evaluó utilizando los frameworks TensorFlow y Scikit-learn. La evaluación experimental se realizó en un procesador Intel Xeon Gold 6138 de 40 núcleos y 128 GB de RAM, utilizando recursos del Centro Nacional de Supercomputación de Uruguay (ClusterUY) (Nesmachnow y Iturriaga, 2019).

5.2. Área de estudio

El estudio de caso examinó la ruta de transporte público 13 en Cuernavaca, México. Esta ruta cuenta con el servicio de autobús número 13. Esta ruta es importante para conectar a los residentes de Jiutepec con el centro de Cuernavaca, especialmente a los estudiantes de la UAEM. La ruta estudiada comienza en la central de autobuses de la UAEM (coordenadas UTM: 18.9774512, -99.2355954) y termina en la parada de Walmart en Jiutepec, Morelos (coordenadas UTM: 18.9028787, -99.1749039), con una longitud aproximada de 15.25 km. Se seleccionó la parada de Walmart porque, al pasar por Walmart en Jiutepec, la ruta se bifurca en diferentes rutas, por ejemplo, San Gaspar y Las Fuentes, ambas ubicadas en el municipio de Jiutepec. La ruta de la UAEM a Walmart cuenta con 37 paradas intermedias, mientras que la ruta de regreso de Walmart a la UAEM cuenta con 41 paradas intermedias. La Fig. 5.1 muestra la ubicación geográfica de la Ruta 13 de la UAEM a Walmart. La imagen fue tomada de Google Maps y se utiliza únicamente con fines académicos de acuerdo con los derechos de autor de “uso legítimo”.

5.3. Ajuste de hiperparámetros

El ajuste de parámetros se realizó mediante el método GridSearch (GS). GS es un enfoque de fuerza bruta que identifica la configuración óptima de hiperparámetros dentro del espacio de búsqueda de la arquitectura para modelos de aprendizaje automático. GS evalúa diversas combinaciones de hiperparámetros



Figura 5.1: Trayectoria del autobús 13 desde la UAEM hasta Walmart

del algoritmo de entrenamiento para determinar el conjunto óptimo que produce el mejor rendimiento del modelo. La variable objetivo (t_{total}) en el conjunto de datos tiene una media de 86.02, una mediana de 67.00, valores del primer y tercer cuartil de 45.00 y 107.00, respectivamente, y valores mínimo y máximo de 11 y 579, respectivamente. En este contexto, las diferentes escalas de salida influyen en el proceso de aprendizaje de la RNA.

La figura 5.2 muestra la distribución y dispersión de la variable objetivo (valores de la duración total del viaje en el eje x) en relación con los tiempos de

llegada entre dos paradas de autobús (frecuencia, en el eje y).

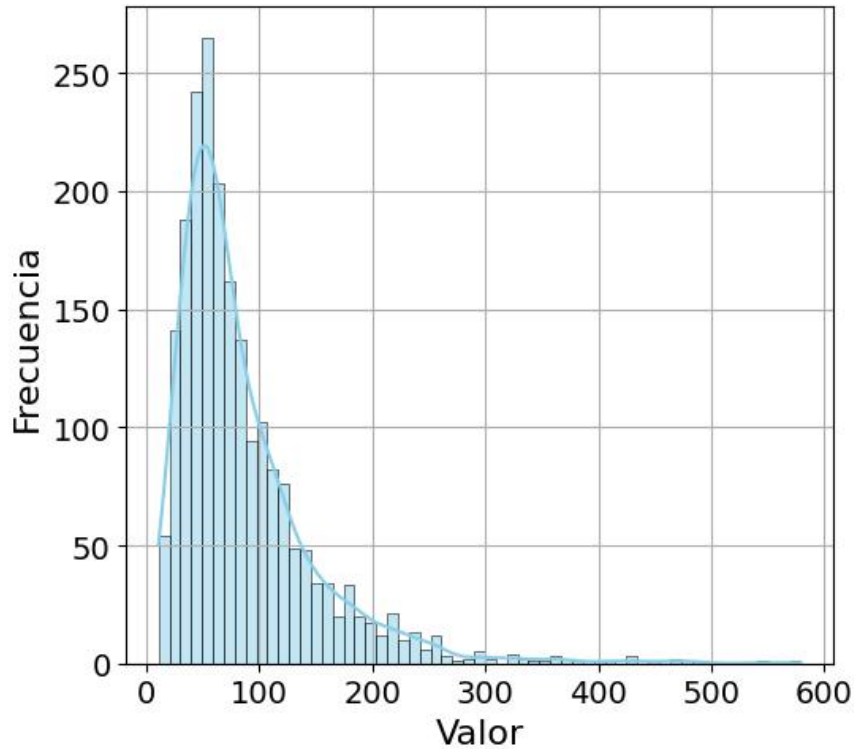


Figura 5.2: Histograma de valores de tiempo total

La Fig. 5.2 muestra que la mayoría de los valores se concentran entre 0 y 100 segundos, con un pico significativo entre los 40 y 60 segundos. Además, una cola larga se extiende a duraciones mayores, llegando a casi 600 segundos. La distribución presenta un sesgo positivo, lo que indica que la mayoría de los viajes son relativamente cortos, mientras que algunos son significativamente más largos.

Los valores de hiperparámetros estudiados en el GS para el conjunto de datos incluyeron un tamaño de lote= $\{16, 32, 64, 128, 256\}$, épocas= $\{50, 100, 300,$

500, 700, 1000}, tasa de aprendizaje={0,00001, 0,0001, 0,001, 0,01}, neuronas en la capa 1={64, 32, 24, 16, 12, 8}, neuronas en la capa 2={32, 24, 16, 12, 8, 6}, neuronas en la capa 3={24, 16, 12, 8, 4, 2}, neuronas en la capa 4={12, 8, 6, 4, 2, 1} y optimizadores={Adam, RMSprop}. Los resultados de GS revelaron la configuración óptima de hiperparámetros de la RNA para el objetivo t_{total} . La configuración óptima para tres capas ocultas consistió en un tamaño de lote (batch size) de 32, 1000 épocas, una tasa de aprendizaje de 0.01, una arquitectura de capas de 32-16-4 y el optimizador Adam, logrando una puntuación media de 0.9996. La configuración óptima para una arquitectura de cuatro capas consistió en un tamaño de lote de 16 (batch size), 300 épocas, una tasa de aprendizaje de 0.001 y el optimizador Adam, con una estructura de capas de 64-6-16-6, lo que resultó en una puntuación media de 0.9995.

5.4. Métricas evaluadas

Las métricas de evaluación para problemas de regresión con RNA son cruciales para evaluar el rendimiento del modelo. Las métricas de evaluación estándar para la regresión con RNA incluyen las siguientes:

- El Error Absoluto Medio (MAE) mide la magnitud promedio de los errores en un conjunto de predicciones, lo que proporciona información sobre la magnitud de los errores sin considerar su dirección. La ecuación de MAE se expresa como $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$
- El Error Cuadrático Medio (RMSE) calcula la raíz cuadrada del promedio

de los errores al cuadrado, otorgando mayor peso a los errores mayores.

La ecuación RMSE se expresa como $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$

- R-cuadrado (R^2), también conocido como coeficiente de determinación, indica la proporción de varianza en la variable dependiente que las variables independientes pueden explicar. La ecuación R^2 se expresa como

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Las métricas son esenciales para comprender el rendimiento de un modelo de RNA y orientar las mejoras.

5.5. Resultados y discusión

Se realizó una evaluación experimental utilizando una división de datos 80–20 para entrenamiento y prueba. Además, el 80 % de los datos de entrenamiento se dividió en conjuntos de entrenamiento y validación, también utilizando una división 80–20. El rendimiento del modelo de RNA se comparó con un método de referencia de regresión lineal multivariante (MLR). Los resultados indicaron que el modelo de RNA tuvo un excelente rendimiento en la estimación del tiempo total en los conjuntos de datos de entrenamiento y prueba. El valor de MAE obtenido fue de 0.5951 y el MSE de 0.6292. El valor de RMSE fue de 0.7932 y el coeficiente de determinación (R^2) de 0.9998. De esta manera, la RNA produjo un resultado promedio un 14.99 % mejor que el método de referencia. Las Figuras 5.3 y 5.4 muestra el comportamiento de la función de pérdida para los datos de entrenamiento y validación, utilizando las métricas MSE (Fig. 5.3)

y R^2 (Fig. 5.4). Además, las Figuras 5.5 y 5.6 compara los valores estimados de la RNA (Fig. 5.5) con los valores objetivo de la MLR (Fig. 5.6).

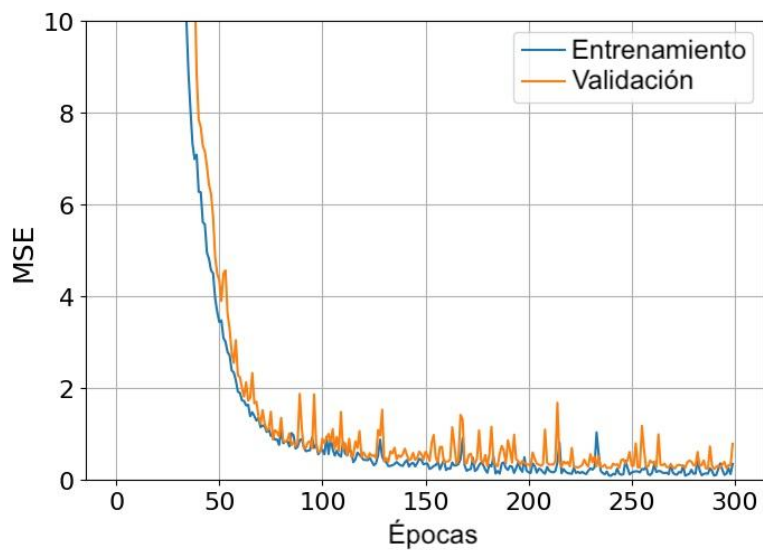


Figura 5.3: Función de pérdida de MSE

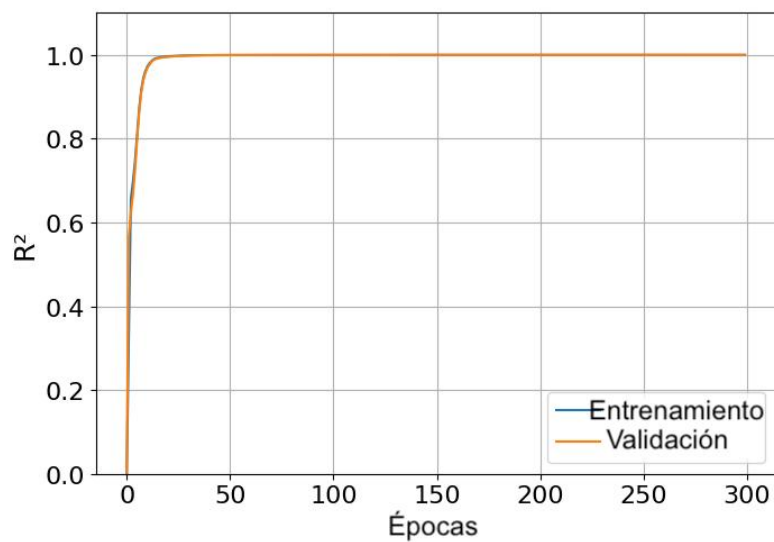


Figura 5.4: Función R^2

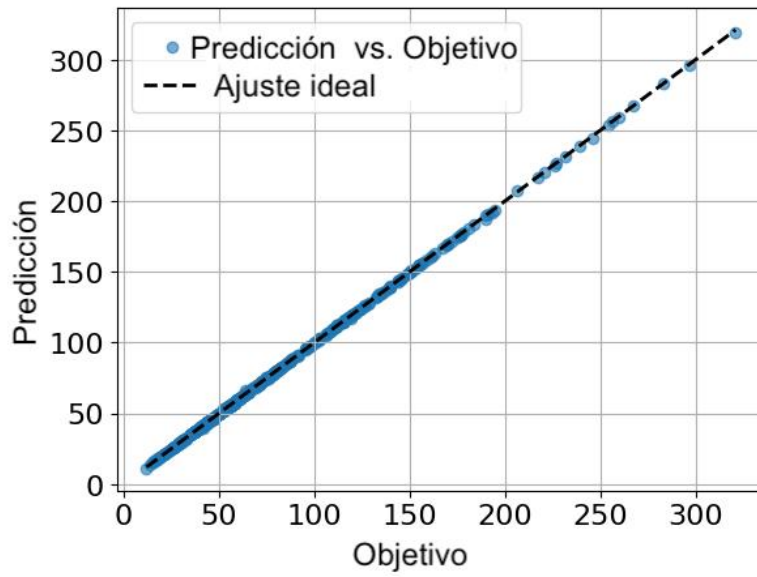


Figura 5.5: Predicción vs. objetivo (RNA)

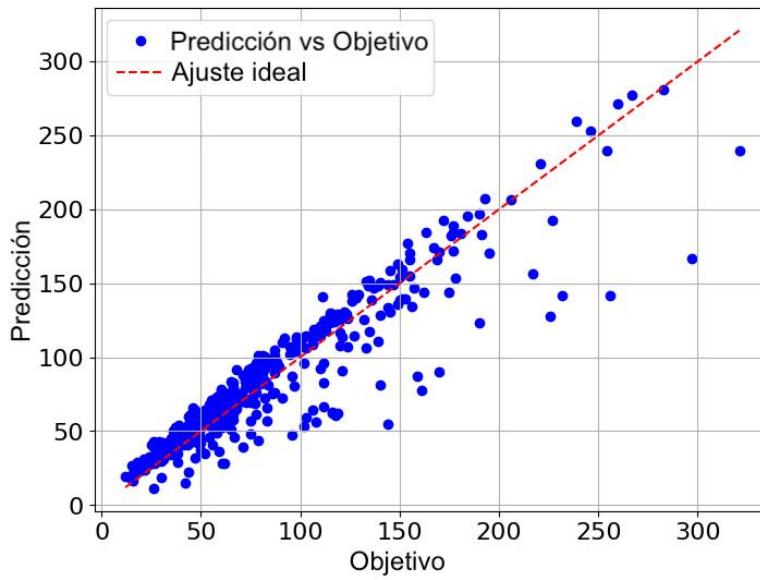


Figura 5.6: Predicción vs. objetivo (MLR)

En la Fig. 5.5, los puntos de datos están estrechamente alineados con la línea diagonal ideal, lo que indica una fuerte correlación entre los valores predichos y los reales. De esta manera, el modelo de RNA se generaliza correctamente, demostrando un sobreajuste mínimo y una alta precisión predictiva. Además, la diferencia porcentual promedio entre los valores reales y las predicciones realizadas por la RNA es del 0.80 % para el conjunto de entrenamiento y del 0.63 % para el conjunto de prueba. Esto indica que, en promedio, las predicciones de la RNA se desvían de los valores reales de entrenamiento en menos del 1 %, lo que demuestra un alto nivel de precisión para los datos de entrenamiento.

Capítulo 6

Conclusiones y trabajo futuro

Esta tesis presenta un enfoque de aprendizaje supervisado que utiliza una RNA para estimar los tiempos de viaje entre dos paradas de autobús. El método de RNA se basó en una arquitectura perceptrón multicapa. El caso de estudio analizó la ruta de transporte público número 13 en Cuernavaca, Morelos, México, durante el período comprendido entre el 1 de octubre de 2024 y el 20 de febrero de 2025. El conjunto de datos comprende información de 2 073 viajes, que incorpora factores como las horas de inicio y de llegada, la velocidad, el tiempo total de viaje, las condiciones meteorológicas, el número de reductores de velocidad y semáforos encontrados, la distancia recorrida y otros eventos inesperados.

Los resultados mostraron que la RNA propuesta superó consistentemente el rendimiento del modelo base, alcanzando valores de R^2 más altos, de 0.9995 a 0.9998, y superando al método base en la predicción de los tiempos de llegada. De esta manera, la RNA produjo un resultado promedio un 14.99% mejor que

el método base. El modelo de RNA propuesto se generaliza eficazmente y supera al modelo base MLR en la estimación del tiempo total. Los modelos propuestos sirven como herramientas prácticas para los sistemas de transporte inteligente.

Las principales líneas de trabajo futuras consisten en mejorar los modelos de RNA para que se adapten a las ubicaciones GPS, así como al número de pasajeros que suben y bajan del autobús, y para que respondan a eventos inesperados. Además, es necesario generar nuevos conjuntos de datos utilizando datos sintéticos creados con redes generativas antagónicas para optimizar el rendimiento de diversos algoritmos de regresión de aprendizaje automático.

Referencias

- Acevedo, E., Serna, A., y Serna, E. M. (2017). *Principios y características de las redes neuronales artificiales* (2.^a ed.). Instituto Antioqueño de Investigación.
- Al Suleiman, S., Cortez, A., Monzón, A., y Lara, A. (2023). How to improve public transport usage in a medium-sized city: key factors for a successful bus system. *Eur. Transp. Res. Rev.*, 15(47). doi: 10.1186/s12544-023-00616-y
- Alam, O., Kush, A., Emami, A., y Pouladzadeh, P. (2021). Predicting irregularities in arrival times for transit buses with recurrent neural networks using gps coordinates and weather data. *Journal of Ambient Intelligence and Humanized Computing*, 12(7), 7813–7826. doi: 10.1007/s12652-020-02507-9
- Al-Naim, R., y Lytkin, Y. (2021). Review and comparison of prediction algorithms for the estimated time of arrival using geospatial transportation data. En *Procedia computer science* (Vol. 193, pp. 13–21). doi: 10.1016/j.procs.2021.11.003
- Amita, J., Jain, S., y Garg, P. (2016). Prediction of bus travel time using ann: A case study in delhi. En *Transportation research procedia* (Vol. 17, pp. 263–272). doi: 10.1016/j.trpro.2016.11.091
- An, H. Y., Choi, W. S., y Choi, S. G. (2022). Data preprocessing method for ann-based travel time estimation using insufficient gps data. En *International conference on information and communication technology convergence* (pp. 1917–1922). doi: 10.1109/ICTC55196.2022.9952906
- Arhin, S., Manandhar, B., Baba-Adam, H., y Gatiba, A. (2020). Predicting travel times of bus transit in washington, d.c. using artificial neural networks. *Civil Engineering Journal*, 6(11), 2245–2261. doi: 10.28991/cej-2020-03091615
- Barbierato, E., y Gatti, A. (2024). The challenges of machine learning: A critical review. *Electronics*, 13(2). doi: 10.3390/electronics13020416
- Bishop, C. (2006). *Pattern recognition and machine learning*. Springer.
- Byon, Y., y Liang, S. (2014). Real-time transportation mode detection using smartphones and artificial neural networks. *Journal of Intelligent Transportation Systems*, 18(3), 264–272. doi: 10.1080/15472450.2013.824762
- Cathey, F., y Dailey, D. (2003). A prescription for transit arrival/departure prediction using automatic vehicle location data. *Transportation Research*

- Part C*, 11(3–4), 241–264. doi: 10.1016/S0968-090X(03)00023-8
- Chen, C., Wang, H., Yuan, F., Jia, H., y Yao, B. (2020). Bus travel time prediction based on deep belief network with back propagation. *Neural Computing and Applications*, 32, 10435–10449. doi: 10.1007/s00521-019-04579-x
- Deb, S., Khan, R., Hasan, K., Khan, A., y Alam, M. (2019). Travel time prediction using machine learning and weather impact on traffic conditions. En *2019 IEEE 5th international conference for convergence in technology* (pp. 1–8). doi: 10.1109/I2CT45611.2019.9033922
- European Commission. (2009). *European commission mandate M/453*.
- Ford, J. (2020). *Reading bus time prediction: A data science approach*. Undergraduate dissertation. Descargado de https://ojhavk.github.io/data/Student_Projects/UG/2019_20_Jade_Ford.pdf
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. MIT Press. Descargado de <https://www.deeplearningbook.org/>
- Gurmu, Z., y Fan, W. (2014). Artificial neural network travel time prediction model for buses using only gps data. *Journal of Public Transportation*, 17, 45–65. doi: 10.5038/2375-0901.17.2.3
- Hassannayebi, E., Farjad, A., Azadnia, A., Javidi, M., y Chunduri, R. (2023). A data analytics framework for reliable bus arrival time prediction using artificial neural networks. *International Journal of Data Science and Analytics*, 20, 337–356. doi: 10.1007/s41060-023-00391-y
- Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice-Hall.
- Hornik, K., Stinchcombe, M., y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–406. doi: [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- H.Yu, Wu, Z., Chen, D., y Ma, X. (2017). Probabilistic prediction of bus headway using relevance vector machine regression. *IEEE Transactions on Intelligent Transportation Systems*, 18(7), 1772–1781. doi: 10.1109/TITS.2016.2620483
- INEGI. (2025). *Estadística de transporte urbano de pasajeros (etup): marzo 2025*. Descargado de https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2025/transporteurbano/trans_urb2025_05.pdf
- Instituto Federal de Telecomunicaciones. (2018). *Estudio sobre sistemas de transporte inteligente*. Descargado de https://www.ift.org.mx/sites/default/files/estudio_sistemas_de_transporte_inteligente.pdf
- Jabamony, J., y Ganesh, R. (2020, 02). Iot based bus arrival time prediction using artificial neural network (ann) for smart public transport system (spts). *International Journal of Intelligent Engineering and Systems*, 13, 312–323. doi: 10.22266/ijies2020.0229.29
- Johansson, C. (2019). *Arrival time predictions for buses using recurrent neural networks* (Tesis de Master, Linköping University). Descargado de <https://liu.diva-portal.org/smash/get/diva2:1424184/FULLTEXT01.pdf>

- Julio, N., Giesen, R., y Lizana, P. (2016). Real-time prediction of bus travel speeds using traffic shockwaves and machine learning algorithms. *Research in Transportation Economics*, 59, 250–257. doi: 10.1016/j.retrec.2016.07.019
- Kuo, Y.-H., Leung, J., y Yan, Y. (2023). Public transport for smart cities: Recent innovations and future challenges. *European Journal of Operational Research*, 306(3), 1001–1026. doi: 10.1016/j.ejor.2022.06.057
- León Sosa, S., Hernández Báez, I., Morales Morales, C., y Maldonado Martínez, H. (2019). El traslado de pasajeros en transporte público en el área metropolitana de cuernavaca morelos. *ANFEI Digital*, 6(11), 1–11.
- McCulloch, W., y Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133. doi: <https://doi.org/10.1007/BF02478259>
- Nesmachnow, S., y Hipogrosso, S. (2022). Transit oriented development analysis of parque rodó neighborhood, montevideo, uruguay. *World Development Sustainability*, 1, 100017. doi: 10.1016/j.wds.2022.100017
- Nesmachnow, S., y Hipogrosso, S. (2024). Assessment of sustainable mobility initiatives developed in montevideo, uruguay. *Urban Science*, 8(2), 52. doi: 10.3390/urbansci8020052
- Nesmachnow, S., y Iturriaga, S. (2019). Cluster-UY: Collaborative Scientific High Performance Computing in Uruguay. En *Supercomputing* (Vol. 1151, pp. 188–202). Springer. doi: 10.1007/978-3-030-38043-4_16
- Ocampo Ortiz, A. (2021). *Sistema de predicción de tiempos de arribo utilizando el sistema gtfs* (Tesis de Master, Tecnológico Nacional de México). Descargado de <https://rinacional.tecnm.mx/handle/TecNM/4429>
- Pang, J., Huang, J., Du, Y., Yu, H., Huang, Q., y Yin, B. (2019). Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network. *IEEE Transactions on Intelligent Transportation Systems*, 20(9), 3283–3293. doi: 10.1109/TITS.2018.2873747
- Raschka, S., y Mirjalili, V. (2018). *Aprendizaje automático y aprendizaje profundo con python, scikit-learn y tensorflow* (2.ª ed.). Marcombo.
- Rashvand, N., Hosseini, S. S., Azarbayjani, M., y Tabkhi, H. (2024). Real-time bus departure prediction using neural networks for smart iot public bus transit. *IoT*, 5(4), 650–665. doi: 10.3390/iot5040029
- Rodrigue, J. (2020). *The geography of transport systems*. Routledge. doi: 10.4324/9780429346323
- Secretaría de Movilidad del Estado de México. (2024). *Transporte público*. Descargado de https://smovilidad.edomex.gob.mx/transporte_publico
- Telediario CdMx. (2025). *Cuál es el transporte público más utilizado por los mexicanos*. Descargado de <https://www.telediario.mx/comunidad/transporte-publico-mexico-cual-es-el-mas-utilizado-inegi-2025>
- Vlahogianni, E., Karlaftis, M., y Golias, J. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C*, 43, 3–19. doi: 10.1016/j.trc.2014.01.005

- Wai, B., y Zhou, W. (2020). Designing and implementing real-time bus time predictions using artificial intelligence. *Transportation Research Record*, 2674(11), 636–648. doi: 10.1177/0361198120947715
- Yu, B., Lam, W., y Tam, M. (2011). Bus arrival time prediction at bus stop using support vector machines. *Transportation Research Part C*, 19(4), 615–628. doi: 10.1016/j.trc.2010.06.009
- Zuluaga-Isaza, J., y Arango-Monsalve, A. (2023). Una revisión a los sistemas de transporte inteligente. *Revista de Ingenierías Interfaces*, 6(1), 1–18.



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Facultad de Contaduría,
Administración e Informática

FACULTAD DE CONTADURÍA, ADMINISTRACIÓN e INFORMÁTICA

SECRETARÍA DE INVESTIGACIÓN

Cuernavaca, Morelos; Marzo 4 de 2026

DR. FELIPE DE JESÚS BONILLA SÁNCHEZ
DIRECTOR DE LA FACULTAD DE CONTADURÍA,
ADMINISTRACIÓN E INFORMÁTICA.
PRESENTE

En mi carácter de director de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, de la estudiante **Edelmira Tapia Arroyo**, con matrícula **10072372**, con el título **Predicción de tiempo de llegada entre paradas de autobuses de transporte público de Cuernavaca mediante redes neuronales artificiales**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

FIRMA DIGITAL

Dr. Pedro Moreno Bernal
Profesor-Investigador de la
Facultad de Contaduría, Administración e Informática





UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANO INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

PEDRO MORENO BERNAL | Fecha:2026-03-04 19:42:32 | FIRMANTE

m9zp59k3ZT+FMpRtFVBqAMXzaSu1k/VKgoz7ruYLI//+TrnXNukl86ezANIKbE2psftRzMC7+4ySr3j/+82vnjWeAdzWbvSd7SeoN3bN3ZOB5ZZt7kmRUU+MriICxQzdTAs7iB+kVmBrZ4JmZBAmAECeWwcmAAangkoymasABfweBEUB8/bK5dn4WR0V7K0apcfW92fkji9JNhqOv3OG5SbzRzwutEpj7epK/OuGtJBhVVyUI7+8umcfwTRP93dte7ciGWxeRsAb1whw4qbHEZTEQi9Jj8HmguQ36FSQV5qYz69RyK9V1HBSj3Fie0Ghspka2FUKgvLeIMJ/7kFg==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



[rJPt9mjoE](#)

<https://efirma.uaem.mx/noRepudio/U186EZF07klhQTKxQCPLN8Kqwlmhf19U>



UAEM
RECTORÍA
2023-2029



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Facultad de Contaduría,
Administración e Informática

FACULTAD DE CONTADURÍA, ADMINISTRACIÓN e INFORMÁTICA

SECRETARÍA DE INVESTIGACIÓN

Cuernavaca, Morelos a 15 de abril de 2026

MTRO. ELMER IVÁN SÁNCHEZ RABADÁN
ENCARGADO DE DESPACHO DE LA DIRECCIÓN DE LA
FACULTAD DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA
PRESENTE

En mi carácter de revisor de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante **Edelmira Tapia Arroyo**, con matrícula **10072372**, con el título **Predicción de tiempo de llegada entre paradas de autobuses de transporte público de Cuernavaca mediante redes neuronales artificiales**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

FIRMA DIGITAL

Dra. Jesús del Carmen Peralta Abarca
Profesora Investigadora de Tiempo Completo
Facultad de Contaduría, Administración e Informática





UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANO INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

JESUS DEL CARMEN PERALTA ABARCA | Fecha:2026-04-17 15:28:47 | FIRMANTE

J3gk1AgpvQRIFmbzjj530IEDPKBVpU5Zpn3i1jLMGmLSeq+qLVj/Z/rzME9BpauGv16Kc8N0Ahho4F05PNDIkVHTTPRG2bJqRy4r6ucdxCdBXvy+ra15ejxMsufM6iVWlj5UdW4RdC
btNmJ7qcjExkRzbcDkXcXIKxKDCwELhORdoghJ7rQqeVp/1Pkc1HaP6KB/hQv+Gdf98gD+ETqQA6/+UjBmZSVn5wL6G2ZAVSuw0JS9C+KvvmCFv2Tpk4HPelwveWRCbHSkgf
gMK1L7QEK6+Uu94rHq0PSof737OB8OK2Du5JRI3EEwfleCKGyB9te8J0n/mQhgH0taTE0ROg==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o
escaneando el código QR ingresando la siguiente clave:



TCrucPXV0

<https://efirma.uaem.mx/noRepudio/tT3krbBddBpK2ziPnOLqB4JOMSKQvPUU>



UAEM
RECTORÍA
2023-2029



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Facultad de Contaduría,
Administración e Informática

FACULTAD DE CONTADURÍA, ADMINISTRACIÓN e INFORMÁTICA

SECRETARÍA DE INVESTIGACIÓN

Cuernavaca, Morelos a 16 de abril de 2026

MTRO. ELMER IVÁN SÁNCHEZ RABADÁN
ENCARGADO DE DESPACHO DE LA DIRECCIÓN DE LA
FACULTAD DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA
PRESENTE

En mi carácter de revisor de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante **Edelmira Tapia Arroyo**, con matrícula **10072372**, con el título **Predicción de tiempo de llegada entre paradas de autobuses de transporte público de Cuernavaca mediante redes neuronales artificiales**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

FIRMA DIGITAL

Dr. Sergio Nesmachnow
Universidad de la República, Uruguay



Av. Universidad 1001 Col. Chamilpa, Cuernavaca Morelos, México, 62209, edificio 2B, Tel. (777) 329 70 00, Ext. 7917
<https://www.uaem.mx/fcaei> correo: posgrado.fcaei@uaem.mx

UAEM
RECTORÍA
2023-2029



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANO INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

SERGIO NESMACHNOW | Fecha:2026-04-17 11:02:23 | FIRMANTE

UMdmX1xEZi+x2LFeRZC/qV99c1jaKbSL2uqJdKBxnGKX7ZugZh0Yst+An1uuzBXWELoEGFFPPC+FbqUfMGuFDepDN+GK0tXjKlfwr7EkjamLd3zdwf/YPIPARewaCAWvATP4s
mLm6tswSPp1iRts+8UyY5//u2QOFSdk5gWSIKjDGy3BQbtjUIO4Pkj7C8P4CJG+TrxxMnF+nkrGLBtc6AGwSebU8HgW/hdr2+IGSQjFaiYxqekDnok93tl5pzLI2HOzcodfwFYnuJgf+
wlqHuleuCXrT0V8lzXVM/RR0fPHHiER5x3VnOKRinhuDWc03CG9/csKVhCaacdH3iMCFa==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o
escaneando el código QR ingresando la siguiente clave:



[gadjYHDti](#)

<https://efirma.uaem.mx/noRepudio/mO2ARsTP8m8RTFQw5DcVPaKRLcZy5Rho>



UAEM
RECTORÍA
2023-2029

Cuernavaca, Morelos a 13 de abril del 2026

MTRO. ELMER IVÁN SÁNCHEZ RABADÁN
ENCARGADO DE DESPACHO DE LA DIRECCIÓN DE LA FACULTAD
DE CONTADURÍA, ADMINISTRACIÓN E INFORMÁTICA
PRESENTE

En mi carácter de revisor de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante **Edelmira Tapia Arroyo**, con matrícula **10072372**, con el título **Predicción de tiempo de llegada entre paradas de autobuses de transporte público de Cuernavaca mediante redes neuronales artificiales**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

FIRMA DIGITAL

Dr. José Alberto Hernández Aguilar
Profesor Investigador de la
Facultad de Contaduría, Administración e Informática



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANO INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

JOSE ALBERTO HERNANDEZ AGUILAR | Fecha:2026-04-17 13:45:13 | FIRMANTE

2TOx8enOKNi8FmienY1A8zgozCITagK86qj51lcm1eCVAjOi0Ao5/CcXUX3fRrHTbo1Ek6sd5QmAYkUvw5iQ5B9f6HYRK0ulla3Z95Y+/i/CueV+5wfCTuPAODh07yWGDrf4Af+/eFbTdmKZQPzh4325F8VkJZuiMBCi87ehHKTtY4pWtayatBluruvWXM0nDFaKryBESIOlQVHG7Fk0XVzc367wctrlt7mM4i1T2xiE19eRy89NjND/8V1aJCWO4Sk8cKsvgridgYJJ04VQQWLaZswL3sj1/kDuWNZTXB5T+qLdyeCwloM5donbzqbtSNOUFT+bS0jpwiiqq1WC2w==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



4upRTKqDP

<https://efirma.uaem.mx/noRepudio/q0Yophwfl56LDCgEwFviimPhE1r3YEP3>



UAEM
RECTORÍA
2023-2029

Cuernavaca, Morelos a 14 de abril del 2026.

MTRO. ELMER IVAN SANCHEZ RABADAN
ENCARGADO DE DESPACHO DE LA DIRECCION DE LA FACULTAD
DE CONTADURIA, ADMINISTRACION E INFORMATICA DE LA UAEM.
PRESENTE.

En mi carácter de revisor de tesis, hago de su conocimiento que he leído con interés la tesis para obtener el grado de la Maestría en Optimización y Cómputo Aplicado, del estudiante **Edelmira Tapia Arroyo**, con matrícula **10072372**, con el título **Predicción de tiempo de llegada entre paradas de autobuses de transporte público de Cuernavaca mediante redes neuronales artificiales**, por lo cual, me permito informarle que después de una revisión cuidadosa de dicha tesis, me permito expresar mi **VOTO APROBATORIO** por lo que de mi parte no existe inconveniente para que el estudiante continúe con los trámites que la Universidad Autónoma del Estado de Morelos tenga establecidos para obtener el grado mencionado.

Atentamente
Por una humanidad culta

Dr. Outmane Oubram
Profesor Investigador de Tiempo Completo
Facultad de Ciencias Químicas e Ingeniería



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento con firma electrónica UAEM, soportada por el certificado vigente a la fecha de su elaboración y con efectos plenos de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS PUBLICADOS en el ÓRGANO INFORMATIVO UNIVERSITARIO "ADOLFO MENÉNDEZ SAMARÁ" número 117 de fecha 20 de abril de 2021.

Sello electrónico

OUTMANE OUBRAM | Fecha:2026-04-17 12:04:56 | FIRMANTE

MCzVImW5eV7YCjn0/2cmpsNG4PXyzUr0WOORtp602zoHHdLKWLRJzeAPvAaRRCFRK45Thz24SL7D7DuyxKOqrOfWk4NAPtK0wOV0Fbebdo1KF5lrakwsfrVH4focxY9Wk11cP+wW5zL7/58kgNzlg1SZ0yZStp99UBT3ia+Vj/U+63Ga1P0xQ5HHi+ohLv65prXCa3sP/grWWKwt/zgmYOvaeYfTfj8I5oDOUPCLMtFPdAvRYxeWfG9wulKO5/l5I8Ni70N10siYJO7ZwRItZCPTNDywdFC+/Hfd7ZGuF/XBrOE+FolgcNvuysfdbD0JlyUzHoKaMBcAl3HwnxMA==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



[TFidbAgsf](#)

<https://efirma.uaem.mx/noRepudio/SOsvPNpJnJFb9zFJ4SKWTme7JNxyrYh>



UAEM
RECTORÍA
2023-2029