



UNIVERSIDAD NACIONAL DE CHIMBORAZO
VICERRECTORADO DE INVESTIGACIÓN, VINCULACIÓN Y
POSGRADO
DIRECCIÓN DE POSGRADO

Análisis estadístico de datos de tráfico para optimizar el diseño de carreteras
y sistemas de transporte

Trabajo de titulación para optar al título de Magister en:
Matemática aplicada con mención en Matemática computacional

Autor:

Sangacha Chacha, José Sebastián

Tutor:

Ing. Jenny Patricia Paredes Fierro. PhD.

Riobamba, Ecuador. 2025

Declaración de Autoría y Cesión de Derechos

Yo, **José Sebastián Sangacha Chacha**, con número único de identificación **060574214-7**, declaro y acepto ser responsable de las ideas, doctrinas, resultados y lineamientos alternativos realizados en el presente trabajo de titulación denominado: **“Análisis estadístico de datos de tráfico para optimizar el diseño de carreteras y sistemas de transporte”** previo a la obtención del grado de Magíster en Matemática Aplicada con mención en Matemática Computacional.

- Declaro que mi trabajo investigativo pertenece al patrimonio de la Universidad Nacional de Chimborazo de conformidad con lo establecido en el artículo 20 literal j) de la Ley Orgánica de Educación Superior LOES.
- Autorizo a la Universidad Nacional de Chimborazo que pueda hacer uso del referido trabajo de titulación y a difundirlo como estime conveniente por cualquier medio conocido, y para que sea integrado en formato digital al Sistema de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor, dando cumplimiento de esta manera a lo estipulado en el artículo 144 de la Ley Orgánica de Educación Superior LOES.

Riobamba, 13 de octubre de 2025

Ing. José Sebastián Sangacha Chacha

N.U.I. 0605742147



ACTA DE CULMINACIÓN DE TRABAJO DE TITULACIÓN

En la ciudad de Riobamba, a los 02 días del mes de octubre del año 2025, los miembros del Tribunal designado por la Comisión de Posgrado de la Universidad Nacional de Chimborazo, reunidos con el propósito de analizar y evaluar el Trabajo de Titulación bajo la modalidad Proyecto de titulación con componente investigación aplicada y/o desarrollo, CERTIFICAMOS lo siguiente:

Que, una vez revisado el trabajo titulado: **“ANÁLISIS ESTADÍSTICO DE DATOS DE TRÁFICO PARA OPTIMIZAR EL DISEÑO DE CARRETERAS Y SISTEMAS DE TRANSPORTE”**, perteneciente a la línea de investigación: **Ingeniería Informática** presentado por el maestrante **José Sebastián Sangacha Chacha**, portador de la cédula de ciudadanía No. **0605742147**, estudiante del programa de Maestría en **Matemática Aplicada con mención en Matemática Computacional**, se ha verificado que dicho trabajo cumple al 100% con los parámetros establecidos por la Dirección de Posgrado de la Universidad Nacional de Chimborazo.

Es todo cuanto podemos certificar, en honor a la verdad y para los fines pertinentes.

Atentamente,



Firmado electrónicamente por:
**JENNY PATRICIA
PAREDES FIERRO**
Validar únicamente con FirmaRC

Ing. Jenny Paredes, PhD
TUTOR



Firmado electrónicamente por:
**MANUEL ANTONIO
MENESES FREIRE**
Validar únicamente con FirmaRC

Dr. Antonio Meneses, PhD
MIEMBRO DEL TRIBUNAL 1



Firmado electrónicamente por:
**LIDIA DEL ROCIO
CASTRO CEPEDA**
Validar únicamente con FirmaRC

Ing. Lidia del Rocío Castro, Mgs
MIEMBRO DEL TRIBUNAL 2



Riobamba, 07 de octubre de 2025

CERTIFICADO

De mi consideración:

Yo **Jenny Patricia Paredes Fierro**, certifico que **José Sebastián Sangacha Chacha** con cédula de identidad No. 0605742147 estudiante del programa de Maestría en **Matemática Aplicada con mención en Matemática Computacional**, cohorte Segunda presentó su trabajo de titulación bajo la modalidad de Proyecto de titulación con componente de investigación aplicada/desarrollo denominado: **“ANÁLISIS ESTADÍSTICO DE DATOS DE TRÁFICO PARA OPTIMIZAR EL DISEÑO DE CARRETERAS Y SISTEMAS DE TRANSPORTE”**, el mismo que fue sometido al sistema de verificación de similitud de contenido COMPILATION identificando el porcentaje de similitud < 1% en el texto y el porcentaje de similitud 3% en inteligencia artificial.

Es todo en cuanto puedo certificar en honor a la verdad.

Atentamente,



Ing. Jenny Patricia Paredes Fierro, PhD

Ci: 0202021853

Adj.-

- Resultado del análisis de similitud(Compilation)

Dedicatoria

El presente trabajo y mi título le dedico a mi mama Victoria, por su apoyo incondicional, que a pesar de cualquier dificultad supo sacarme adelante y a quien debo todos mis sueños cumplidos.

Además, dedico a mi familia por su apoyo.

Agradecimiento

Primeramente agradezco a Dios por la vida y la sabiduría para poder cumplir con mis sueños, metas y logros que me ha concedido la vida.

A la Universidad Nacional de Chimborazo, a la Dirección de Posgrados y a mi tutora por permitirme realizar los estudios de posgrados y contribuir una vez más en mi formación profesional.

A mi madre Victoria por su apoyo incondicional, desde pequeño quien me oriento y me ha acompañado paso a paso en el cumplimiento de mis objetivos y en este desafío de mi formación de cuarto nivel.

A mi abuelita Herminia, que es un angelito y sé que me acompaña, me cuida y me protege; así también a mi familia y amigos; quienes me han acompañado.

Índice General

Declaración de Autoría y Cesión de Derechos	ii
Acta de culminación de trabajo de titulación	iii
Certificado de contenido de similitud	iv
Dedicatoria	v
Agradecimiento	vi
Índice General.....	vii
Índice de Tablas	x
Índice de Figuras	xi
Resumen	xii
Abstract	xiii
Introducción	1
Capítulo 1 Generalidades.....	4
1.1 Planteamiento del problema	4
1.2 Justificación de la Investigación.....	5
1.3 Objetivos.....	6
1.3.1 Objetivo General.....	6
1.3.2 Objetivos Específicos	6
Capítulo 2 Estado del Arte y la Práctica	7
2.1 Antecedentes Investigativos	7
2.2 Fundamentación Teórica	12
2.2.1 Estadística descriptiva	12
2.2.2 Series temporales	12
2.2.3 Procesos estocásticos	13
2.2.4 Autocorrelación	14
2.2.5 Procesos Estacionarios	15

2.2.6	Procesos no estacionarios	16
2.2.7	Pruebas de hipótesis	18
2.2.8	Redes Neuronales	18
2.2.9	Evaluación del Modelo	23
2.2.10	Teoría de Colas	24
Capítulo 3 Diseño Metodológico.....		27
3.1	Enfoque de la Investigación	27
3.2	Diseño de la Investigación.....	27
3.3	Tipo de investigación	28
3.4	Nivel de Investigación	28
3.5	Técnicas e Instrumentos de Recolección de Datos.....	28
3.6	Técnicas para el Procesamiento e Interpretación de Datos	29
3.6.1	Estadística descriptiva	29
3.6.2	Estadística Inferencial.....	30
3.6.3	Modelos predictivos	30
3.6.4	Teoría de colas.....	31
3.6.5	Software R Studio.....	31
3.7	Población y Muestra	32
Capítulo 4 Resultados y Discusión		34
4.1	Análisis Descriptivo de los Resultados	34
4.1.1	Análisis vehicular de la ciudad de Riobamba.....	34
4.2	Análisis de la serie temporal de los datos de tráfico vehicular para los modelos predictivos.....	37
4.2.1	Descomposición multiplicativa	37
4.2.2	Estacionariedad.....	38
4.2.3	Autocorrelación	39

4.3	Modelos predictivos de la serie temporal	40
4.3.1	Modelo ARIMA	40
4.3.2	Modelo SARIMA	41
4.3.3	Modelos de redes neuronales.....	43
4.3.4	Comparación entre los modelos de predicción.....	47
4.4	Modelo de Teoría de Colas.....	49
4.5	Discusión de los Resultados	51
	Conclusiones.....	54
	Recomendaciones.....	56
	Referencias Bibliográficas	57
	Apéndice	63
	Apéndice A. Código de los modelos ARIMA y SARIMA	63
	Apéndice B. Código del modelo de redes neuronales multicapa.....	66
	Apéndice C. Código del modelo de redes neuronales recurrentes.....	67
	Apéndice D. Código del modelo de teoría de colas.	68

Índice de Tablas

Tabla 1. <i>Vehículos matriculados en los cantones de la provincia de Chimborazo en el año 2023.</i>	35
Tabla 2. <i>Resultados de la prueba de Dickey Fuller Aumentada</i>	38
Tabla 3. <i>Resultados del modelo ARIMA (1,1,1)</i>	40
Tabla 4. <i>Métricas de evaluación del modelo ARIMA (1,1,1)</i>	40
Tabla 5. <i>Resultados del modelo SARIMA(0,1,2) (0,1,0)₁₂</i>	41
Tabla 6. <i>Métricas de evaluación del modelo (0,1,2) (0,1,0)₁₂</i>	42
Tabla 7. <i>Evaluación de los modelos de predicción mediante el RMSE</i>	48
Tabla 8. <i>Valores de predicción para el año 2023</i>	48
Tabla 9. <i>Resultados de los parámetros operacionales de la intersección del modelo de teoría de colas con múltiples canales M/M/2</i>	50

Índice de Figuras

Figura 1. Modelo de redes neuronales para una serie de tiempo.....	20
Figura 2. Redes neuronales recurrentes LSTM	23
Figura 3. Composición del sistema de colas.	24
Figura 4. Sistema de colas de múltiples servidores	25
Figura 5. Ubicación de la intersección de los datos del tráfico vehicular	32
Figura 6. Serie histórica de vehículos matriculados en Chimborazo	34
Figura 7. Flujo vehicular de la Av. Pedro Vicente Maldonado hacia la Av. Monseñor Leónidas Proaño (Panamericana Norte E35) y hacia la Av. Panamericana Sur E35	36
Figura 8. Serie de Tiempo Número de vehículos 2000-2022.....	36
Figura 9. Descomposición multiplicativa de la serie temporal	37
Figura 10. Función de autocorrelación (ACF) y función de autocorrelación parcial (PACF)	39
Figura 11. Análisis de los residuos del modelo ARIMA(0, 1, 2)(0, 1, 0) ₁₂	42
Figura 12. Datos reales y predicciones del modelo SARIMA(1,0,2)(1,0,1) ₁₂	43
Figura 13. Modelo de la red neuronal multicapa para la serie temporal	44
Figura 14. Datos reales, predicción y pronóstico del Modelo MLP (12, [12,6], 1) reps=200.	44
Figura 15. Análisis de datos para entrenamiento y validación del modelo RNN LSTM	46
Figura 16. Datos reales y predicciones del modelo RNN LSTM (2000-2023).....	47
Figura 17. Predicción de valores de número de vehículos para el año 2023.....	48
Figura 18. Simulación del Proceso de Poisson de vehículos de llegada	51

Resumen

A lo largo de los años en el Ecuador, el crecimiento del parque automotor ha provocado el aumento de los problemas de congestión vehicular. A esto se le añade, la existencia de sistemas de tránsito inadecuados que generan la necesidad de estudios óptimos de sistemas de transporte. Para obtener diseños óptimos y una mejor planificación vial es clave contar con una serie temporal histórica del volumen del tráfico que permita realizar predicciones de una manera correcta. Por tal motivo, la presente investigación tiene el propósito de aplicar técnicas de análisis estadístico de datos de tráfico vehicular para identificar patrones, tendencias y problemas en la infraestructura vial y sistemas de transporte con el fin de obtener recomendación para su optimización. El análisis de datos se los realizó mediante una serie temporal mensual del volumen de tráfico de una intersección de Riobamba entre los años 2000 y 2022. Los modelos estadísticos aplicados para la predicción fueron ARIMA, SARIMA, redes neuronales multicapa (MLP) y redes neuronales recurrentes RNN LSTM, analizados de acuerdo a la naturaleza de los datos y evaluados mediante la raíz del error cuadrático medio RMSE. Adicionalmente, se utilizó el modelo de teoría de colas de dos carriles para analizar el comportamiento vehicular. El modelo predictivo que mejor se ajustó es el modelo SARIMA con un RMSE de 3353,10; y mediante la teoría de colas se identificó que existe un 45% de probabilidad de congestión durante las horas pico y un 28% que exista la posibilidad de espera de un vehículo. De esta manera, se ha evidenciado que los modelos se ajustan de acuerdo a la naturaleza de los datos y permiten tomar decisiones al momento del diseño y planificación vial.

Palabras claves: Serie temporal, ARIMA, SARIMA, redes neuronales, teoría de colas.

ABSTRACT

Over the years, the growth of the vehicle fleet in Ecuador has led to an increase in traffic congestion problems. Added to this is the existence of inadequate traffic systems, which generate the need for optimal transportation system studies. For optimal designs and improved road planning, it is essential to have a historical time series of traffic volumes that allows for accurate predictions. Therefore, this research aims to apply statistical analysis techniques to vehicle traffic data to identify patterns, trends, and problems in road infrastructure and transportation systems, to obtain recommendations for their optimization. Data analysis was performed using a monthly time series of traffic volumes at an intersection in Riobamba between 2000 and 2022. The statistical models applied for prediction were ARIMA, SARIMA, multilayer perceptrons (MLPs), and recurrent neural networks (RNNs), including long short-term memory (LSTMs), analyzed according to the nature of the data and evaluated using the root mean square error (RMSE). Additionally, a two-lane queuing theory model was used to analyze traffic behavior. The best-fitting predictive model was the SARIMA model, with an RMSE of 3353.10. Queuing theory identified a 45% probability of congestion during peak hours and a 28% probability of waiting for a vehicle to arrive. This demonstrates that the models adjust according to the nature of the data, enabling informed decision-making during road design and planning.

Keywords: Time series, ARIMA, SARIMA, neural networks, queuing theory.



Reviewed by:

Mgs. Sofía Freire Carrillo

ENGLISH PROFESSOR

C.C. 0604257881

Introducción

A nivel mundial y en Ecuador, la congestión vehicular representa un problema para la población; esto debido al crecimiento demográfico, la insuficiencia de obras viales, los sistemas de tránsito inadecuados, la falta de mantenimiento vial y el aumento de número de automóviles. Esto genera factores negativos en la sociedad como la reducción de velocidades, incremento en el tiempo de los viajes y aumento de accidentes de tránsito (Ashhad et al., 2020).

El problema de congestión vehicular es más representativo en las principales ciudades del Ecuador como son Quito y Guayaquil, con un tiempo de pérdidas en el tráfico es de 70 y 54 horas respectivamente; ubicando a la capital en el puesto 40 a nivel mundial según el Informe INRIX 2023 (Arias et al., 2023).

Alsubai et al. (2024a) describe una de las formas para solucionar esta problemática, implementando sistemas de tráfico inteligentes basado en la predicción de tráfico eficiente y precisión para mejorar la optimización del tráfico, reducir la accidentabilidad, aliviar la congestión y reducir impactos ambientales. Para lograr esta solución es importante el análisis estadístico de datos del tráfico, mediante modelos estadísticos y matemáticos que proporcionan información importante para identificar patrones y comportamiento del flujo vehicular (Ruiz & Tineo, 2019).

Varias son las investigaciones que se presentan en el campo estadístico y el análisis de datos de tráfico como aplicación de estadística descriptiva y de regresión analizando la problemática del flujo vehicular, tiempos de viajes, distribución de semáforos entre otras variables (Ashhad et al., 2020).

Torres (2012) y Zhang et al. (2016) analizan el tráfico vehicular mediante modelos estocásticos, por otro lado la aplicación de modelo de colas de Poisson Teoría de colas (Daniel et al., 2021) y Redes Preti (Burbano & Ceron, 2007). Así también, Kosun &

Ozdemir (2016) utilizando la superestadística en el análisis avanzado de datos de tráfico con distribuciones q-Gaussianas y Chi Cuadrado para la comprensión de las interacciones entre vehículos y el flujo vehicular.

De la misma manera, la utilización de análisis de series de tiempo por medio de modelos dinámicos lineales permite la identificación de la probabilidad de saturación para la toma de decisiones en ingeniería de tránsito (Pilamunga & Ríos Insua, 2014). Mientras tanto, Kumar & Vanajakshi (2015), Cascante & Alarcon (2018), Rivera & Meneses (2022), Nassiri et al. (2023), Balawi & Tenekeci (2024), estudian las predicciones del tráfico mediante modelos autorregresivos de promedios móviles ARIMA, SARIMA y variantes con la aplicación de redes neuronales.

El aprendizaje automático y las redes neuronales juegan un papel importante en este tema, considerando modelos no lineal así y patrones de la naturaleza compleja del tránsito como Do et al. (2019), Mrudula et al. (2024), adicionalmente Agyemang et al. (2023), Sattarzadeh et al. (2023), Fahs et al. (2023), Alsubai et al. (2024) realizan modelos híbridos para el mejor entendimiento de las variables y patrones de predicción.

De acuerdo a lo antes mencionado, el objetivo de la presente investigación es aplicar técnicas de análisis estadísticos de datos de tráfico mediante el desarrollo de modelos que permitan identificar patrones, tendencias y problemas en la infraestructura vial y sistemas de transporte con el fin de generar recomendaciones para su optimización.

De esta manera, para el presente estudio los datos utilizados son proporcionados por el grupo de investigación de la ESPOCH, inicialmente se analiza por medio de estadística descriptiva, para posterior realizar un análisis estadístico avanzado mediante unos modelos autorregresivos de promedios móviles, de redes neuronales y modelos de Poisson.

Para llevar a cabo el propósito de la presente investigación, se ha organizado todo el estudio en capítulos, donde, en el capítulo 1 se indican las generalidades, el problema de la investigación su justificación y objetivos. En el capítulo 2 se muestra la investigación de artículos similares y la fundamentación teórica abordando los temas de análisis estadístico descriptivo, modelos estadísticos de series de tiempo, redes neuronales y modelos de Poisson. En el capítulo 3 se plasma la metodología y se explica cómo se realiza el análisis de datos. En el capítulo 4 se muestran los resultados obtenidos y su análisis y en el Capítulo 5 se detallan los modelos y su aplicación para la solución de problemas generados por el tránsito; y para finalizar se resumen las principales conclusiones y las recomendaciones.

Capítulo 1

Generalidades

1.1 Planteamiento del problema

El Ecuador es uno de los países con mayor índice de crecimiento poblacional. Esto genera en las grandes e intermedias ciudades el aumento del sector automotor y por ende el incremento de la congestión vehicular, ocurrencia de accidentes de tránsito y contaminación ambiental (Moyolema Chaglla et al., 2022) (Ministerio de Transporte y Obras Públicas, 2023).

De acuerdo con los reportes del Anuario de Estadística del Transporte, 2023 del INEC (2024) existen 3.07 millones de vehículos matriculados y existieron 20994 siniestros hasta el año 2023. Lo que representa un incremento en un 6.42% en relación al 2022.

Para el diseño óptimo de vías y carreteras se deben considerar elementos como el tipo de terreno, velocidad, capacidad, seguridad y el tránsito que mediante el conteo o su volumen se estima el tránsito futuro a través de una proyección para la vida útil que se diseñe (NEVI-12-MTOP, 2013).

Para un análisis del tráfico vehicular y un diseño vial correcto es crucial contar con datos históricos de las vías para realizar estimaciones del número de vehículos y poder dar soluciones a la planificación vial. Es importante indicar que hoy en día no se cuenta con una base de datos histórica del número de vehículos de las vías en la ciudad de Riobamba para un correcto estudio de tráfico; comúnmente para un diseño vial y análisis del tráfico se realiza el conteo del Tráfico Promedio Diario Anual (TPDA). Esta actividad muchas veces genera dificultad al momento de cuantificar los vehículos en las vías y solo se recolecta datos de un periodo de tiempo 7 días.

Por ello la disponibilidad de registros previos sobre el comportamiento del tráfico resulta fundamental para comprender la dinámica y realizar estudios adecuados de tráfico adaptados al sitio de estudio y considerar modelos precisos para diseños correctos y una planificación vial óptima evitando costos en los sistemas de transporte, impactos en la productividad y en la calidad de los habitantes.

1.2 Justificación de la Investigación

El tráfico vehicular es un problema a nivel mundial y afecta directamente con la calidad de vida de las personas, muchas veces no solo se trata de la ampliación de vías y carreteras sino el uso eficiente de todo el sistema de transporte (Rostami-Shahrbabaki et al., 2020), y el análisis de todas las variables que lo componen, como la red vial, densidad poblacional, la educación vial, los accidentes de tránsito, obras viales, entre otras (Rivera-Picado & Meneses-Guzmán, 2022).

El Ecuador no es la excepción en problemas de congestión vehicular, a pesar de que la existencia de planes de movilidad, con el crecimiento demográfico y el aumento del parque automotor las ciudades grandes presentan dificultades en el diseño óptimo de los sistemas de transporte.

Una estrategia para el control del tráfico vehicular son los sistemas de transporte inteligentes cuyo desafío principal es la capacidad de pronosticar el flujo de tráfico y el análisis de sus variables debido a la naturaleza y características dinámicas del tráfico; siendo la predicción confiable del tráfico la clave para proporcionar soluciones para mejorar la congestión y seguridad vial (Sattarzadeh et al., 2023).

Existen modelos tradicionales para la estimación de los volúmenes de tránsito que utilizan regresión lineal, curvilíneas, exponencial, potencial y logarítmica y para el flujo vehicular los modelos microscópicos y macroscópicos (Cal Mayor & Cárdenas, 2018) Por otro lado, los métodos estadísticos avanzados han presentado mejores resultados

destacando los modelos de series de tiempo, así también los modelos híbridos que combinando los modelos de series de tiempo y redes neuronales (Sattarzadeh et al., 2023) presentando mayor precisión y pronóstico del tráfico.

La correcta implementación de un modelo estadístico avanzado en el análisis de los datos de tráfico permite un avance en la ingeniería de tránsito al obtener pronósticos acertados a corto y largo plazo; para proponer soluciones a la congestión vehicular y a la seguridad vial del Ecuador.

1.3 Objetivos

1.3.1 Objetivo General

- Identificar patrones en datos de tráfico vehicular para la optimización en el diseño de carreteras y sistemas de transporte.

1.3.2 Objetivos Específicos

- Identificar patrones en los datos de tráfico vehicular mediante técnicas estadísticas.
- Seleccionar modelos estadísticos adecuados para representar el comportamiento del flujo vehicular.
- Evaluar el desempeño de los modelos aplicados según criterios de ajuste y predicción.
- Proponer recomendaciones para optimizar el diseño de la infraestructura vial.

Capítulo 2

Estado del Arte y la Práctica

2.1 Antecedentes Investigativos

Para el estudio de los problemas de tráfico vehicular una de las soluciones importantes en la aplicación de Sistemas de transporte inteligentes (ITS), dentro del cual la predicción del flujo del tránsito es uno de los conceptos principales aplicados en esta temática; por esta razón en la investigación “Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems” de Boukerche et al. (2020) destacan los principales métodos para la predicción del flujo vehicular como son los modelos basados en estadística como la familia AR/MA (Modelos ARMA, ARIMA, SARIMA, VARIMA) y el suavizado exponencial.

Así también, Bucay et al. (2023) en la investigación “Leveraging Classical Statistical Methods for Sustainable Maintenance in Automotive Assembly Equipment” utilizan los modelos de series temporales como HoltWinters Exponential Smoothing (HWES), Autoregressive Integrated Moving Average (ARIMA) y Seasonal Autoregressive Integrated Moving Average (SARIMA) para datos de enero de 2020 hasta el 2022, identificando que el modelo que mejor se ajusta a los datos es el modelo SARIMA por sus características estacionales.

Por otro lado, los métodos de aprendizaje automático (ML), como subclase los métodos de aprendizaje supervisado como los algoritmos de K-Nearest Neighbours (KNN), regresión lineal, Máquina de vectores de soporte (SVM), regresión lineal, red neuronal recurrente (RNN); así también la subclase de los métodos de aprendizaje no supervisado como los algoritmos de agrupamiento y también modelos basados en estadística y enfoques de machine learning como el filtro del Kalman y la Máquina Oculta de Markov (HMM) (Mrudula et al., 2024).

De acuerdo a lo mencionado anteriormente, a continuación, se explican diferentes estudios que resaltan como antecedentes investigativos para la presente investigación.

Cascante & Alarcón (2018) en su investigación “Estimación del flujo vehicular a través de series de tiempo Caso: Arcabuco, Sáchica y Crucero” realizaron las predicciones del tránsito a corto plazo de una malla vial de Boyacá, a partir de datos obtenidos a través del Instituto Nacional de Vías; para lo cual el estudio consistió en el análisis estadístico descriptivo para determinar que categoría de vehículos es la de mayor flujo vehicular. De esta manera realizaron los pronósticos del tráfico en tres estaciones mediante modelos ARIMA, evidenciando las tendencias, estacionalidad y ciclicidad; y su validación para la selección del mejor modelo con lo cual se determinó que en base a los datos históricos la predicción del modelo es adecuada y presenta resultados satisfactorios.

En el artículo “Predicción flujo de tráfico Ruta 27 en Costa Rica”, Rivera & Meneses (2022) realizan el pronóstico del flujo del tráfico vehicular considerando tres modelos; Arima Estacional (SARIMA), Ingenuo Estacional (SNAIVE) y Autoregresión con Redes Neuronales (NNAR). La metodología utilizada fue primero realizar una limpieza de datos y el análisis mediante estadística descriptiva, análisis de pruebas de normalidad, autocorrelación y estacionariedad; así también la aplicación de los tres métodos mencionados y la comparación mediante métricas de error, para finalizar con la validar los modelos mediante validación cruzada. Concluyendo que el modelo NNAR es el que presenta los resultados más estables y precisos.

Para la mayoría de modelos de predicción del flujo de tráfico se requiere una gran cantidad de datos para el desarrollo pero en el caso de “Short-term traffic flow prediction using seasonal ARIMA model with limited input data” de Kumar & Vanajakshi (2015) utilizan datos de tres días para pronosticar valores del siguiente día utilizando un modelo de ARIMA estacional (SARIMA) que mediante la evaluación por medio del error

porcentual absoluto medio MAPE verificando que el modelo es aceptable para la aplicación de los ITS.

De manera similar Kumar (2017) en “Traffic Flow Prediction using Kalman Filtering Technique” propone un algoritmo basado en la técnica de filtrado de Kalman ya que solo poseen datos históricos de dos días de un tramo de vía de la India, para la evaluación se determinó de igual forma mediante el MAPE que presento estimaciones precisas de acuerdo a los datos; proponiendo un modelo apto cuando se dispone de limitadas bases de datos.

En la investigación “A Comparison of Approaches for the Time Series Forecasting of Motorway Traffic Flow Rate at Hourly and Daily Aggregation Levels” mediante la utilización de modelos de series temporales analizan la predicción de datos agregados horarios y diarios para reducir los problemas de congestión y costos sociales.

Se ha realizado el modelamiento para los datos agregados horarios mediante Ingenio Estacional y el modelo con patrones estacionales complejos mediante suavizado exponencial TBATS (Estacionalidad trigonométrica, transformación Box-Cox, errores ARMA, componentes de tendencia y estacional); y para los datos agregados diarios mediante modelos Ingenio Estacional, SARIMA, SARIMAX, TBATS, redes neuronales recurrentes LSTM. Siendo los modelos TBATS y LSTM que mejor funcionan respectivamente para cada caso descrito mediante una evaluación por el Error absoluto medio MAE (Van Der Bijl et al., 2022).

“An effective spatial-temporal attention based neural network for traffic flow prediction” es un artículo en el que Do et al. (2019) analizaron la predicción del flujo de tráfico a través de un modelo basado en una red neuronal con atenciones espaciales y temporales (STANN); la atención temporal hace énfasis a los segmentos de la carretera y la atención temporal al paso del tiempo.

Fahs et al. (2023) en el proyecto “Traffic Congestion Prediction Based on Multivariate Modelling and Neural Networks Regressions” estudian una propuesta para predecir la congestión del tráfico en Beirut, capital del Líbano. Para el desarrollo de la investigación aplican técnicas de machine learning como redes neuronales de avance (FFNN), redes neuronales de función de base radial (RBFNN), modelo de regresión simple lineal y modelo de regresión lineal polinomial destacando como variables los días de la semana, el tiempo de espera, estado climático si fue lluvioso, si fue día festivo y el tráfico.

La evaluación de los métodos se realizó por varios parámetros como el coeficiente de determinación, error cuadrático medio y error relativo de predicción; en el cual se destaca y presenta mejores resultados el modelo de red neuronal de avance (FFNN).

Para el estudio de los modelos de tránsito y flujo vehicular es esencial la recopilación de datos y en este caso Mrudula et al. (2024) en “ Internet of things and optimized KNN based intelligent transportation system for traffic flow prediction in smart cities” utilizaron dispositivos de Internet de las cosas (IoT) que se implementan en las vías para obtener una base de datos en tiempo real y estadísticas de tráfico históricas; en tanto que para predecir el flujo vehicular realizan la selección óptima de características y optimización del enjambre de partículas; y constituir un modelo de clasificación que utiliza los algoritmos de K-Nearest Neighbours (KNN), perceptrón multicapa (MLP) y red bayesiana.

Los modelos han sido evaluados mediante por el Error cuadrático medio (MSE) y y una tasa de error cuadrático medio (RMSE), presentando mejores resultados y alto nivel de precisión el modelo aplicado mediante KNN.

En el artículo “Hybrid deep learning-based traffic congestion control in IoT environment using enhanced arithmetic optimization technique” describen la aplicación de

un método híbrido que utiliza un algoritmo híbrido de red neuronal convolucional de memoria a largo y corto plazo (HCNN-ALSTM), se deriva una estrategia de ajuste hiperparamétrico basado en IAOA. El modelo es denominado IAOADL-TCC, utilizó los datos de varias carreteras de Gran Bretaña y la validación con un conjunto de datos del repositorio Kaggle, el cual mediante la matriz de confusión se obtuvo buena precisión y mínimo error (Alsubai et al., 2024b).

Otro estudio relacionado es “A performance modeling and analysis of a novel vehicular traffic flow prediction system using a hybrid machine learning-based model”, en el cual Boukerche & Wang (2020) mejoran los modelos de predicción del tráfico implementando un modelo híbrido de aprendizaje profundo mediante el uso de Graph Convolutional Network (GCN) y una estructura de agregación profunda de secuencia a secuencia de Gated Recurrent Unit (GRU). Además proponen una estrategia de predicción en línea basado en un aprendizaje por refinamiento y una nueva métrica de rendimiento llamada ganancia para la evaluación de la relación coste-efectividad del modelo.

Por otro lado, Sattarzadeh et al. (2023) en el artículo titulado “A novel hybrid deep learning model with ARIMA Conv-LSTM networks and shuffle attention layer for short-term traffic flow prediction” en el cual proponen un modelo híbrido para los datos de transporte de California para explorar características de linealidad, no linealidad, espacios temporales y estocasticidad aplicando el modelo de media móvil integral autorregresiva (ARIMA), redes neuronales convolucionales (CNN) y (LSTM) desarrollando un módulo CONV LSTM para la atención aleatoria y componentes LSTM bidireccional (Bi-LSTM) para la periodicidad. Siendo el modelo denominado Shuffle Attention ARIMA Conv-LSTM (SAACL) que posee buena precisión para el pronóstico a corto plazo.

Relacionado al tema de estudio, en el artículo “Análisis estadístico del flujo de tráfico en intersecciones del centro de la ciudad de Riobamba”, Mera Herrera & Colcha

Ortiz (2025) realizaron la aplicación de técnicas estadísticas avanzadas como el modelo ARIMA en 11 intersecciones del centro de la ciudad de Riobamba para identificar componentes de tendencia, estacionalidad y residuos que permite identificar dinámicas del tráfico en horas pico y predecir volúmenes futuros de vehículos. Además, destacan la importancia del modelo ARIMA como una herramienta confiable y robusta para obtener predicciones precisas para el desarrollo de estrategias viales.

Adicionalmente, al estudiar problemas de tráfico es importante analizar por modelos de teorías de colas, así como Yang & Yang (2014) en su investigación “The Application of the Queuing Theory in the Traffic Flow of Intersection”, que estudian la intersección Shenzhen y analizan la configuración del carril de acuerdo al estado actual de la vía y permite analizar el aumento o disminución de los carriles de la vía para evitar ciertos retrasos o tiempos en la cola.

2.2 Fundamentación Teórica

2.2.1 Estadística descriptiva

La estadística es una rama de la matemática que se basa en la recolección, estudio e interpretación de los datos que se obtienen en un estudio. Las principales áreas que aborda esta ciencia son la estadística descriptiva y la estadística inferencia. En este apartado se abordará la estadística descriptiva que se basa en la comprensión de los datos; es decir es un conjunto de técnicas numéricas y gráficas que permiten describir, clasificar, representar y resumir los datos para dar una explicación e interpretación clara de información recolectada (Seoane et al., 2007).

2.2.2 Series temporales

Herrera & Pérez (2019) establecen que una serie temporal es todo aquello que se ha observado en un determinado tiempo, cuyas mediciones se las realiza en un intervalos fijos

que pueden presentar relaciones, siendo así que pueden tener correlación o dependencia para emplearse en predicciones en el futuro.

Una serie de tiempo es una secuencia de observaciones ordenadas cronológicamente. En análisis de tráfico vehicular, las series de tiempo permiten estudiar cómo varían las condiciones de tráfico a lo largo del tiempo para hacer pronósticos y tomar decisiones.

2.2.2.1 Componentes de una serie temporal

- **Tendencia:** Movimiento a largo plazo en la media de la serie.
- **Estacionalidad:** Patrones que se repiten en intervalos regulares (días, meses, años).
- **Ruido:** Variabilidad aleatoria sin patrón definido.

Los modelos se clasifican en:

Modelo Aditivo

$$X_t = T_t + S_t + a_t \quad (1)$$

Modelo Multiplicativo

$$X_t = T_t * S_t * a_t \quad (2)$$

2.2.3 *Procesos estocásticos*

Un proceso estocástico es un concepto matemático que se utiliza para representar aleatoriedad que cambia con el tiempo o para describir una serie de variables aleatorias que evolucionan en función de otra que generalmente es el tiempo.

2.2.3.1 **Proceso estocástico estacionario**

Una serie temporal es estacionaria si sus propiedades estadísticas, como la media y la varianza, permanecen constantes a lo largo del tiempo. Esto es crucial porque muchos modelos estadísticos, incluidos ARIMA y SARIMA, asumen estacionariedad para obtener predicciones confiables (Escudero , Alcoser, & Paredes, 2021). Sea X_t una serie de tiempo entonces:

$$\text{Media } E(X_t) = E(X_{t+k}) = \mu \quad (1)$$

$$\text{Varianza } V(X_t) = V(X_{t+k}) = \sigma^2 \quad (2)$$

$$\text{Covarianza } E[(X_t - \mu)(X_{t+k} - \mu)] = \gamma_k \quad (3)$$

Donde γ_k , la covarianza al rezago k , es la covarianza entre valores de X_t y X_{t+k} que están separados por el tiempo.

2.2.4 Autocorrelación

En una serie temporal, los datos que adquiere una variable a lo largo del tiempo no son independientes unos de otros; por el contrario, cada valor está influenciado por los valores que lo preceden (Yumiceba, 2022).

2.2.4.1 Función de Autocorrelación (ACF)

El coeficiente de autocorrelación en el retardo k , dentro de un proceso estocástico estacionario, representa qué tan fuerte es la relación lineal entre dos valores aleatorios del mismo proceso que están separados por k unidades de tiempo (Yumiceba, 2022).

$$\rho_k = \frac{\text{cov}(X_t, X_{t-k})}{\sqrt{V(X_t)}\sqrt{V(X_{t-k})}} = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\gamma_0} \quad (4)$$

Dado que se trata de un coeficiente de correlación, no está afectado por las unidades.

2.2.4.2 Función de Autocorrelación parcial (PACF)

Evalúa la relación lineal entre dos valores del proceso que están separados por k periodos, sin tener en cuenta la influencia de los valores intermedios que podrían afectar dicha relación (Villavicencio, 2010).

$$\pi_k = \frac{\text{cov}(X_t - \bar{X}_t, \bar{X}_{t-k} - \bar{X}_{t-k})}{\sqrt{V(X_t - \bar{X}_t)}\sqrt{V(\bar{X}_{t-k} - \bar{X}_{t-k})}} \quad (5)$$

2.2.5 Procesos Estacionarios

2.2.5.1 Procesos autorregresivos $AR(p)$

Un modelo AR (autorregresivo) representa un tipo específico de proceso donde los valores actuales pueden estimarse utilizando valores anteriores junto con un término de error aleatorio (Villavicencio, 2010). El caso más simple es el $ARIMA(p,0,0)$ cuya expresión matemática es:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (6)$$

Donde ε_t es la variable aleatoria con $N(0, \sigma^2)$ y $\phi_1, \phi_2, \dots, \phi_p$ son los parámetros a estimar.

2.2.5.2 Procesos de medias móviles $MA(q)$

El modelo de medias móviles (MA) se utiliza para representar series temporales que son estacionarias. En este enfoque, el valor presente se estima principalmente a partir del componente aleatorio actual y, en menor proporción, de los choques aleatorios ocurridos en periodos anteriores (De la Fuente Fernández, 2003).

El modelo $ARIM(0,0,q)$ cuya expresión es:

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (7)$$

Donde ε_t es la variable aleatoria y $\theta_1, \theta_2, \dots, \theta_q$ son los parámetros a estimar.

2.2.5.3 Proceso autorregresivo de medias móviles $ARMA(p, q)$

Una forma de ampliar los modelos $AR(p)$ y $MA(q)$ consiste en combinar ambos enfoques en una sola estructura, que incorpora componentes autorregresivos junto con términos de medias móviles, dando origen a un modelo más completo y se define como $ARIM(p, 0, q)$ (Yumiceba, 2022).

El modelo $ARMA(p, 0, q)$ cuya expresión es:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (8)$$

Donde $\phi_p(L)$ es el polinomio autorregresivo y $\theta_q(L)\varepsilon_t$ es el polinomio de medias móviles.

2.2.6 Procesos no estacionarios

2.2.6.1 Proceso autorregresivo integrado y de media móvil $ARIMA(p, d, q)$

Los modelos $ARIMA$ (Autorregresivos Integrados de Media Móvil) se emplean para representar procesos que no son estacionarios. En esta clase de modelos, el parámetro p corresponde al orden del componente autorregresivo, d señala cuántas veces se debe diferenciar la serie para hacerla estacionaria, y q representa el orden del componente de media móvil. Cuando la serie ya es estacionaria desde el inicio, entonces ($d = 0$), y el modelo $ARIMA$ se transforma en un modelo $ARMA$. Por lo tanto, la notación $ARIMA(p, d, q)$ puede aplicarse tanto a series estacionarias como a aquellas que requieren diferenciación para lograr la estacionariedad.

$$X_t^d = c + \phi_1 X_{t-1}^d + \dots + \phi_p X_{t-p}^d + \theta_1 \varepsilon_{t-1}^d + \theta_2 \varepsilon_{t-2}^d + \dots + \theta_q \varepsilon_{t-q}^d + \varepsilon_t^d \quad (9)$$

Donde X_t^d son diferencias de orden d , ε_t^d es ruido blanco y $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ son los parámetros para estimar.

2.2.6.1.1 Características

Los modelos $ARIMA$ son herramientas estadísticas que permiten representar y predecir el comportamiento de una serie temporal, ya sea estacionaria o no. A diferencia de otros modelos, no necesitan variables externas, ya que se basan únicamente en los propios datos pasados de la serie. Estos modelos aprovechan los patrones de autocorrelación para realizar pronósticos precisos (Villavicencio, 2010).

La construcción de un modelo $ARIMA$ se basa en la metodología de Box-Jenkins, que incluye cuatro pasos principales:

- **Identificación:** Se analizan los datos para definir los valores adecuados de los parámetros (p, d, q) .

- **Estimación:** Se calculan los parámetros del modelo seleccionado.
- **Validación:** Se verifica si el modelo se ajusta correctamente a los datos mediante pruebas estadísticas.
- **Predicción:** Una vez validado, el modelo se utiliza para generar pronósticos.

2.2.6.2 Proceso autorregresivo integrado y de media móvil estacional

$$ARIMA(p, d, q)(P, D, Q)_s$$

Cuando una serie temporal se registra en intervalos menores a un año, es habitual que exhiba patrones o fluctuaciones periódicas dentro de ese mismo lapso. Estas variaciones regulares, que pueden ocurrir semestral, mensual o diariamente, reflejan la presencia de estacionalidad en la serie (Yumiceba, 2022).

Las series temporales con estacionalidad pueden clasificarse en dos tipos principales: aditivas y multiplicativas. Cada una de estas categorías puede ser, a su vez, estacionaria o no estacionaria. Generalmente, los modelos multiplicativos son más comunes que los aditivos. Estos modelos combinan componentes ordinarios del proceso ARMA con elementos estacionales, e incluyen tanto diferencias regulares como diferencias estacionales para lograr que la serie sea estacionaria (Yumiceba, 2022).

Una de sus características es que incluye una parte $ARIMA(p, d, q)$ que representa la dependencia regular entre observaciones consecutivas, y una parte $ARIMA(P, D, Q)$ que captura la dependencia estacional, relacionada con observaciones separadas por intervalos específicos (Yumiceba, 2022).

La expresión del modelo $ARIMA(p, d, q)(P, D, Q)_s$ es:

$$= c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_1 X_{t-s} + \dots + \theta_p X_{t-ps} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \dots - \varphi_q \varepsilon_{t-q} - \vartheta_1 \varepsilon_{t-s} - \dots - \vartheta_Q \varepsilon_{t-Qs} \quad (10)$$

2.2.7 Pruebas de hipótesis

2.2.7.1 Prueba de Dickey-Fuller Aumentada (ADF)

Es una prueba estadística para verificar si una serie tiene raíz unitaria (no estacionaria).

H_0 : La serie no es estacionaria.

H_1 : La serie es estacionaria.

Interpretación del p-value:

Si p-value < 0.05, se rechaza H_0 → serie estacionaria.

Si p-value > 0.05, no se rechaza H_0 → serie no estacionaria y se requiere diferenciarla.

2.2.8 Redes Neuronales

Las redes neuronales artificiales (RNA) son modelos computacionales inspirados en el funcionamiento del cerebro humano. Tienen la capacidad de aprender a partir de datos y generalizar patrones complejos, incluso cuando los datos contienen ruido o irregularidades. Una de sus grandes ventajas frente a los modelos estadísticos tradicionales, como ARIMA, es que no requieren asumir que el proceso es lineal, lo que las vuelve especialmente útiles en problemas donde las relaciones entre variables son no lineales o difíciles de modelar explícitamente (Escudero et al., 2021).

En su forma más general, una RNA puede modelar el valor de una serie temporal y_t en función de sus valores pasados:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) \quad (11)$$

Donde f representa una función no lineal aprendida automáticamente por la red. Este tipo de estructura formada por una capa de entrada, una o más capas ocultas y una capa de salida, permite aproximar funciones de alta complejidad, lo que las convierte en herramientas poderosas para tareas de predicción temporal (Fernández Salguero, 2021).

2.2.8.1 Redes Neuronales Multicapa (MLP)

Las redes neuronales multicapa (MLP) son modelos computacionales inspirados en el cerebro humano, compuestas por capas de neuronas artificiales. Cada neurona procesa señales mediante funciones de activación y transmite información a la siguiente capa (Escudero et al., 2021).

El modelo MLP en R permite ajustar redes neuronales para predecir series temporales, usando como entrada retardos (lags) de la serie y diferencias si es necesario. El modelo puede predecir valores futuros (por ejemplo, a 12 meses) y su precisión se evalúa con métricas como el RMSE (Escudero et al., 2021).

Las redes neuronales multicapa comprende un conjunto de unidades de procesamiento es decir neuronas, que están organizados y distribuidos en tres capas; una capa de entrada, al menos una capa oculta y una capa de salida. Una serie de tiempo se puede modelar mediante una MLP con p neuronas de entradas, h neuronas en la capa oculta y una neurona de salida que es útil para el análisis de las series de tiempo (Menacho, 2014). El modelo MLP para una serie temporal se muestra en la siguiente expresión:

$$y_t = \alpha_0 + \sum_{j=1}^h \alpha_j G(\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i}) + e_t \quad (12)$$

Donde:

y_t = valor de la serie tiempo en el instante t , representa el valor de salida del modelo MLP.

y_{t-i} = valores rezagados de la serie de tiempo en el instante t , representa el valor de entrada p .

α_j = pesos de la capa oculta a la capa de salida, mientras que α_0 es el sesgo respectivo.

β_{ij} = Coeficientes que representan los pesos de la capa de entrada a la capa oculta y β_o el sesgo

G = Función de activación de la capa de entrada, la cual determina la salida de la capa oculta; la función de activación generalmente es sigmoïdal para la capa de entrada e identidad para la capa de salida.

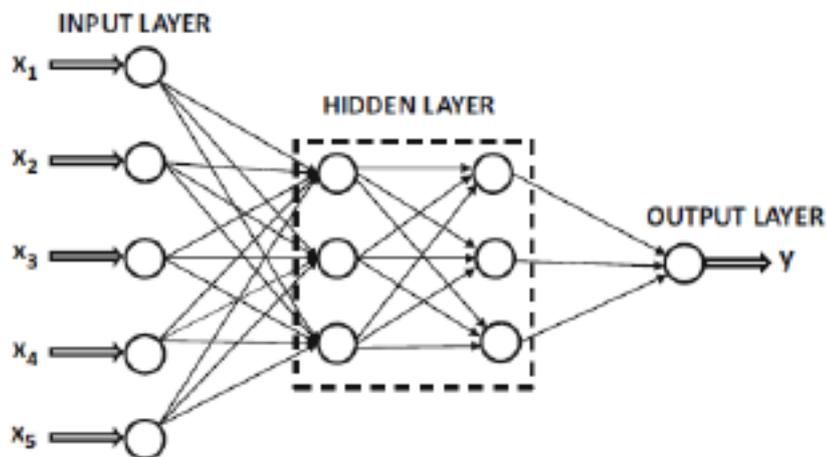
h = Numero de neuronas en la capa oculta

p = Numero de neuronas en la capa de entrada, es el número de rezagos.

e_t = Errores aleatorios del modelo (Menacho, 2014).

De acuerdo con la ecuación 1, la red neuronal con p nodos de entrada, h nodos en las capas ocultas y un nodo de salida, se representa gráficamente en la Figura 1.

Figura 1. Modelo de redes neuronales para una serie de tiempo



Fuente: (Aggarwal, 2018)

Los modelos MLP poseen las siguientes características:

- Tipo de conexión. En este tipo de redes neuronales la conexión es hacia adelante, todas las neuronas se desarrollan hacia adelante a través de las capas de la red.

- Capa de entrada. Es aquella que recibe los datos de la serie temporal, el número de neuronas, nodos o datos de entrada (rezagos), está dado por el número de autocorrelaciones de la serie de tiempo.
- Capas ocultas. Se puede considerar varias capas ocultas con diferente número de neuronas de acuerdo a la complejidad de los datos de la serie de tiempo
- Capa de salida. Está compuesta por una neurona y corresponde al valor y_t de la ecuación 12.
- Función de activación. Como se mencionó anteriormente en la capa de entrada es la sigmoïdal y para la capa de salida es una función identidad.

2.2.8.2 Redes Neuronales Recurrentes RNN

Las redes neuronales recurrentes (RNN) son una extensión más avanzada de las RNA, diseñadas específicamente para trabajar con secuencias temporales. Su principal diferencia con las MLP es que las RNN poseen conexiones recurrentes que les permiten recordar información de estados anteriores, es decir, tienen memoria temporal. Gracias a esta característica, pueden procesar secuencias de longitud variable y modelar relaciones dinámicas entre eventos distribuidos en el tiempo (Fernández Salguero, 2021).

Existen dos tipos principales: redes de impulso finito y redes de impulso infinito (que mantienen bucles de retroalimentación continuos). Esta capacidad de recordar permite que las RNN sean aplicadas con éxito en problemas como reconocimiento de voz, lenguaje natural y series temporales financieras o de tráfico. Además, arquitecturas más modernas como LSTM (Long Short-Term Memory) o GRU (Gated Recurrent Units) han mejorado su desempeño al controlar qué información se debe guardar u olvidar a lo largo del tiempo (Fernández Salguero, 2021).

La red neuronal RNN LSTM se utilizan ampliamente para pronósticos de datos secuenciales debido a su mecanismo de redes de almacenamiento de información a corto y

largo plazo; estos modelos pueden recordar la secuencia de datos y conservarlos durante el tiempo, en tanto almacenar memoria a corto plazo como las RNN básicas como una memoria a largo plazo (Escudero et al., 2021).

Cada bloque de la red LSTM contiene varias células que reemplazan las neuronas y unidades ocultas con un bloque de memoria (Figura 2). Escudero et al. (2021) manifiestan que la entrada y salida se representa mediante x_t y y_t , el vector que representa la memoria a corto plazo se idéntica con h_t y la memoria a largo plazo c_t . Para las predicciones de una serie temporal de x_t , el sistema de RNN LSTM actualiza la celda de memoria c_t y genera un estado oculto h_t para cada paso t , como se indica en las siguientes ecuaciones y Figura 2.

$$i_t = \sigma (W_{xi}^T \cdot x_t + W_{hi}^T \cdot h_{t-1} + b_i) \quad (13)$$

$$f_t = \sigma (W_{xf}^T \cdot x_t + W_{hf}^T \cdot h_{t-1} + b_f) \quad (14)$$

$$O_t = \sigma (W_{xo}^T \cdot x_t + W_{ho}^T \cdot h_{t-1} + b_o) \quad (15)$$

$$g_t = \tanh (W_{xg}^T \cdot x_t + W_{hg}^T \cdot h_{t-1} + b_g) \quad (16)$$

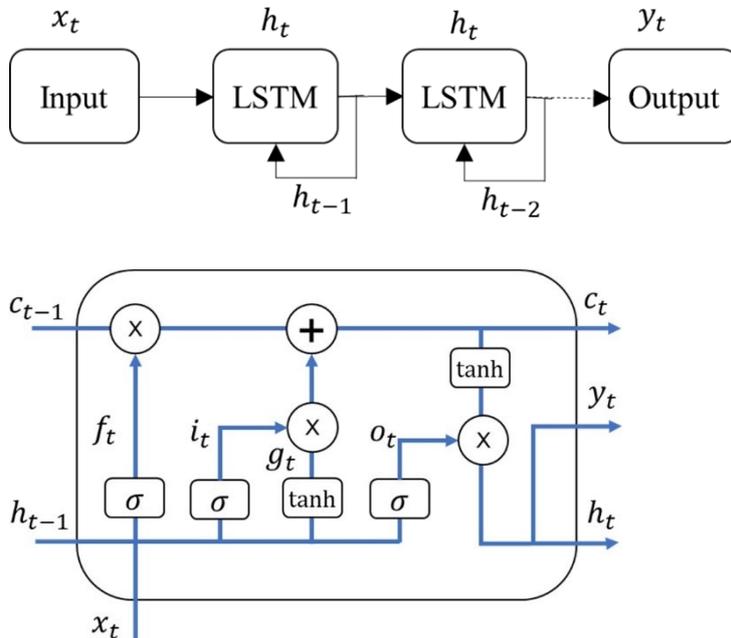
$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t \quad (17)$$

$$y_t = h_t = O_t \otimes \tanh (c_t) \quad (18)$$

La puerta del olvido (forget gate) f_t , la puerta de entrada (output gate) i_t y la puerta de salida (O_t) son alimentadas por los datos de entrada x_t y un estado anterior a corto plazo h_{t-1} que incluye la puerta g_t , donde σ es la función logística sigmoide estándar $\sigma(x) = 1/(1 + e^{-x})$ y \tanh se describe por $\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$. Las matrices de peso $W_{xi}, W_{xf}, W_{xo}, W_{xg}$ y $W_{hi}, W_{hf}, W_{ho}, W_{hg}$ están conectadas al vector de entrada y al estado anterior a corto plazo h_{t-1} ; además b_i, b_f, b_o, b_g son los sesgos para

cada una de las capas en la cual b_f se inicia en 1 segundo para evitar olvidar todo al inicio del entrenamiento (Escudero et al., 2021).

Figura 2. Redes neuronales recurrentes LSTM



2.2.9 Evaluación del Modelo

2.2.9.1 Diagnóstico de residuos

- Se analizan los residuos con la función ACF para verificar la ausencia de autocorrelación.
- Se aplica la prueba de Ljung-Box para confirmar que los residuos son ruido blanco (p-value > 0.05).

2.2.9.2 Criterios de ajuste

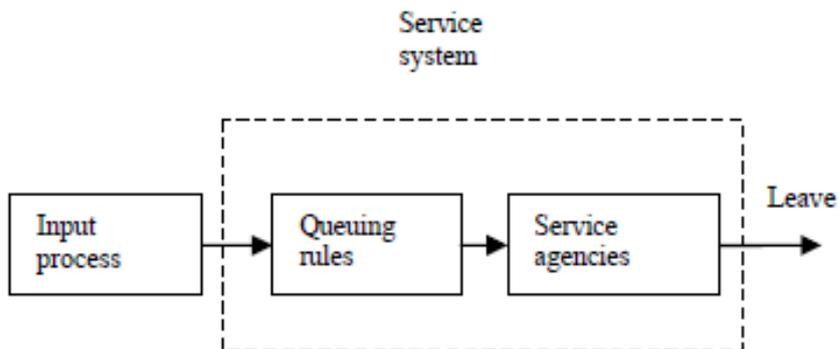
El desempeño de los modelos es importante para escoger los mejores criterios para que cada uno funciones de la mejor manera y comparación entre ellos, para el presente estudio como medida de evaluación se utiliza la raíz del error cuadrático medio (RMSE), en la cual mientras el valor sea más cercano a cero el modelo está prediciendo de mejor manera (Cantuña & Almeida, 2024).

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (19)$$

2.2.10 Teoría de Colas

La teoría de colas es una teoría matemática y un sistema estocástico; en el cual una cola es un sistemas con uno más recursos para una población de usuarios que tienen que esperar para acceder a estos recursos por estar ocupados por otros usuarios. La teoría de colas consta de tres partes: el proceso de entrada, regla de la cola y agencia de servicio. Así también, la teoría de colas se puede dividir en un sistema de colas de un solo canal o en un sistema de colas multicanales (Yang & Yang, 2014).

Figura 3. Composición del sistema de colas.



Fuente: (Yang & Yang, 2014)

2.2.10.1 Sistema de colas de un solo canal

La notación para este modelo es M/M/1, basado en un sistema al que los clientes llegan aleatoriamente siguiendo una distribución de Poisson, en la cual λ es la tasa media de llegada, μ es la tasa media de salida y $\rho = \lambda/\mu$ es la intensidad del tráfico o coeficiente de utilización (Yang & Yang, 2014).

Una condición necesaria para que el sistema sea estable es que $\rho < 1$, es decir que la tasa de llegada sea menor que la tasa de salida y por ende el tráfico en la intersección es fluido. De manera contraria si $\rho \geq 1$, la tasa de llegada es mayor que la tasa de salida la longitud de la cola es infinita, por lo tanto el sistema es inestable (Yang & Yang, 2014).

Yang & Yang (2014) expresan que combinando la fórmula de Little, se obtienen los índices operativos de un sistema de colas de un solo canal o carril, de la siguiente manera:

- Probabilidad de que no haya ningún vehículo en el sistema:

$$P_0 = 1 - \rho \quad (20)$$

- Probabilidad de que haya n vehículos en el sistema:

$$P_n = \rho^n(1 - \rho) \quad (21)$$

- El numero promedio de vehículos en el sistema:

$$L = \frac{\rho}{1-\rho} \quad (22)$$

- El numero promedio de vehículos en el sistema:

$$L_q = L * \rho \quad (23)$$

- El tiempo promedio de permanencia de los vehículos en el sistema:

$$W = \frac{L_q}{\lambda} + \frac{1}{\mu} \quad (24)$$

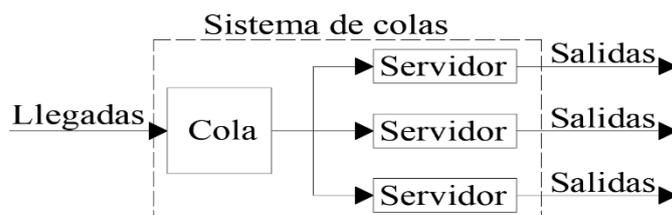
- El tiempo promedio de espera de los vehículos en el sistema:

$$W_q = W - \frac{1}{\mu} \quad (25)$$

2.2.10.2 Sistema de colas de múltiples canales.

El sistema de colas multicanal se denomina M/M/N, a diferencia del sistema de un solo canal, este modelo se basa en dos o más canales de servicio y para ser atendido tendrá más posibilidad que solo esperar en una fila (Paguay & Bejarano, 2020).

Figura 4. Sistema de colas de múltiples servidores



Fuente: (Paguay & Bejarano, 2020)

La intensidad del tráfico se denota por ρ/N siendo N el numero de canales de servicio, en este caso para que el sistema sea estable $\rho/N < 1$, de lo contrario el sistema sería inestable (Yang & Yang, 2014). Las características operativas del modelo de múltiples canales son las siguientes:

- La probabilidad que no haya vehículos en el sistema:

$$P_0 = \frac{1}{\sum_{k=0}^{N-1} \frac{\rho^k}{k!} + \frac{\rho^N}{N!(1-\rho/N)}} \quad (26)$$

- La probabilidad de k vehículos en el sistema:

$$\begin{cases} \rho_k = \frac{\rho^k}{k!} P_0 & (\text{Cuando, } k < N) \\ \rho_k = \frac{\rho^k}{N! N^{k-N}} P_0 & (\text{Cuando, } k \geq N) \end{cases} \quad (27)$$

- La longitud media de espera de los vehículos en el sistema:

$$L_q = \frac{\rho^{N+1}}{N! N} \cdot \frac{P_0}{(1-\rho/N)^2} \quad (28)$$

- El numero promedio de vehículos en el sistema:

$$L = L_q + \rho \quad (29)$$

- El tiempo promedio de permanencia de los vehículos en el sistema:

$$W = \frac{L_q}{\lambda} + \frac{1}{\mu} \quad (24)$$

- El tiempo promedio de espera de los vehículos en el sistema:

$$W_q = \frac{L_q}{\lambda} \quad (30)$$

Capítulo 3

Diseño Metodológico

3.1 Enfoque de la Investigación

La presente investigación tiene como objetivo analizar los datos del tráfico vehicular mediante modelos estadísticos, considerando el número de vehículos que transitan por una intersección de la ciudad de Riobamba, esta consideración se la realiza debido a que los datos son proporcionados por el grupo de investigación de la ESPOCH ya que no se cuentan con datos históricos del tránsito de la ciudad de otras intersecciones, además la intersección y la avenida analizada es de gran afluencia vehicular ya que conduce a otras ciudades como Guaranda, Guayaquil y Cuenca; y su análisis garantiza que los resultados reflejen un comportamiento del tráfico y la validación de los métodos estadísticos aplicados.

Cabe indicar que es un estudio base y previo que se podría implementar con otros métodos en otros puntos estratégicos de la ciudad.

El enfoque de la investigación es cuantitativo, ya que se fundamenta en el tratamiento de datos numéricos utilizando modelos como ARIMA, SARIMA, redes neuronales artificiales y teoría de colas. Este enfoque permite medir, evaluar y predecir el comportamiento del tráfico vial mediante herramientas matemáticas y computacionales.

3.2 Diseño de la Investigación

De acuerdo con Hernández Sampieri et al. (2014), el diseño de la presente investigación es no experimental, que se analiza una serie de tiempo de los datos del tráfico vehicular (número de vehículos que pasan por una intersección) sin manipulación de las variables; es decir se observa el fenómeno tal como sucede en su contexto natural para analizarlo sin poder influir en las variables.

El estudio se realiza en base en los datos históricos del número de vehículos que transitan por una intersección de Riobamba entre los años 2000 y 2022; lo que permite el análisis del comportamiento vehicular a lo largo del tiempo para identificar patrones y tendencias mediante métodos estadísticos y realizar la predicción para el año 2023.

De esta manera la investigación es no experimental longitudinal de tendencia, considerando que se estudian datos históricos del tráfico y se analiza los cambios al paso del tiempo analizando su tendencia para realizar predicciones. (Hernández Sampieri et al., 2014)

3.3 Tipo de investigación

El tipo de investigación es aplicada, ya que está orientada a resolver un problema real y práctico, aplicando diferentes modelos estadísticos para realizar predicciones y el análisis del tráfico vehicular.

3.4 Nivel de Investigación

El nivel de la investigación es correlacional y explicativos. Hernández Sampieri et al. (2014) destaca que investigación correlacional asocia variables mediante un patrón predecible para un grupo. De esta manera, en el presente estudio se analizan la relación entre las variables del número de vehículos que circulan a lo largo del tiempo por una intersección para realizar predicciones.

Este estudio se complementa con el nivel explicativo para comprender patrones, tendencias del comportamiento vehicular aplicando modelos estadísticos y teoría de colas con el fin de comprender el fenómeno y su proyección adecuada.

3.5 Técnicas e Instrumentos de Recolección de Datos

La recolección de datos se basa en la investigación y revisión documental de fuentes de datos secundarios; (Hernández Sampieri et al., 2014) indica que un método cuantitativo de recolección de datos es mediante datos recolectados por otros

investigadores o secundarios, que implican datos estadísticos de fuentes oficiales, la revisión de documentos, registros públicos y archivos físicos o electrónicos.

En este contexto, los datos han sido investigados en fuentes oficiales estatales como, el Instituto Nacional de Estadísticas y Censos (INEC) para obtener los datos actualizados del parque automotor de Riobamba; y mediante el grupo de investigación de la ESPOCH, por medio de pedido verbal se proporcionó el registro histórico de los datos del tráfico vehicular (número de vehículos) de una intersección de la ciudad de Riobamba.

De acuerdo con el alcance de la investigación el uso de fuentes secundarias permite el acceso a la información objetiva, sistemática y confiable, sin la necesidad de aplicar instrumentos de recolección de datos de forma directa; para poder realizar los análisis cuantitativos sobre la tendencia y patrones de los datos obtenidos.

3.6 Técnicas para el Procesamiento e Interpretación de Datos

En la presente investigación, el procesamiento e interpretación de los datos se llevó a cabo mediante técnicas de estadística descriptiva e inferencial, así como mediante la aplicación de modelos estadísticos avanzados y técnicas de modelación matemática, tales como modelos de series temporales (ARIMA, SARIMA), redes neuronales artificiales y teoría de colas.

Todo el análisis fue desarrollado utilizando el software RStudio, un entorno de programación especializado en el procesamiento de series temporales, el desarrollo de modelos matemáticos y la implementación de redes neuronales artificiales, lo que permitió una evaluación precisa y rigurosa de los datos de tráfico vehicular.

3.6.1 Estadística descriptiva

Los datos obtenidos a través del INEC del número de vehículos registrados a través del tiempo se analizan mediante una gráfica de dispersión, que permite describir, clasificar, representar y resumir los datos.

3.6.2 Estadística Inferencial

La estadística inferencial se aplicó con el objetivo de realizar estimaciones y predicciones sobre el comportamiento futuro del tráfico vehicular, a partir de los datos históricos. Para ello, se emplearon modelos de series temporales como ARIMA y SARIMA, así como redes neuronales artificiales, que permiten inferir patrones ocultos, evaluar la precisión de las predicciones y validar hipótesis sobre el comportamiento de la variable en el tiempo. Esta fase fue fundamental para generar pronósticos confiables que sirvan como base para la toma de decisiones en la planificación vial

3.6.3 Modelos predictivos

Para continuar, se aplicaron modelos de predicción de series temporales, para identificar cual es el mejor modelo y poder realizar la estimación futura de los valores del tráfico vehicular proporcionado por el ANT, los siguientes modelos fueron considerados de acuerdo a la bibliografía.

3.6.3.1 Modelo Autorregresivo Integrado por Media Móvil (ARIMA) y Modelos

Autorregresivo Integrado por media Móvil Estacional (SARIMA).

Inicialmente para los modelos ARIMA y SARIMA, se realizó el análisis de la serie temporal mediante la descomposición multiplicativa, prueba de estacionariedad y autocorrelación, para continuar con el análisis por medio del Modelo ARIMA; cabe indicar que porque existe estacionalidad en los datos el modelo a utilizar fue SARIMA, evaluándolo mediante gráficos de los residuos y verificación de ruido blanco para identificar si el modelo es correcto y está bien ajustado.

3.6.3.2 Red Neuronal Multicapa (MLP)

Para el modelo de redes neuronales MLP, se construyó mediante una capa de entrada, dos capas ocultas y una capa de salida, mediante el cual se entrena los datos de

entrada. Para los datos de entrenamiento se estableció que son del año 2000 hasta el 2012, y para la prueba los restantes del año 2013 hasta el 2022.

3.6.3.3 Red Neuronal Recurrente LSTM

De la misma forma anterior, se utilizó los datos de entrenamiento y prueba, en el cual se manejó un modelo de dos capas ocultas, y una capa densa de salida, con las repeticiones de entrenamiento adecuadas de acuerdo al número de datos.

La evaluación del rendimiento de los modelos se lo realizó mediante la Raíz del Error Cuadrático Medio (RMSE) como parámetro de validación, de esta manera con el modelo que presenta el menor valor del RMSE se realiza las predicciones de la serie temporal.

3.6.4 Teoría de colas

Así también, se aplica un modelo basado en teoría de colas para caracterizar el flujo vehicular en la intersección. En el presente estudio la Av. Pedro Vicente Maldonado posee dos carriles por lo tanto se utiliza el sistema de colas de múltiples canales (dos canales), para determinar los parámetros de la tasa media de llegada λ y la tasa media de servicio o salida μ se realizó la inspección de campo en las horas pico de servicio en la intersección.

Una vez registrado los datos y analizado mediante el sistema de colas de múltiples canales se logró obtener el tiempo promedio en el sistema, tiempo promedio de espera en la cola, número promedio de vehículos en la vía, número promedio de vehículos esperando, la probabilidad de congestión y probabilidad de que el vehículo tenga que esperar.

3.6.5 Software R Studio

Todas las etapas del análisis de los datos se han realizado mediante el Software R Studio, un lenguaje de programación para computación estadística y gráficos que ofrece

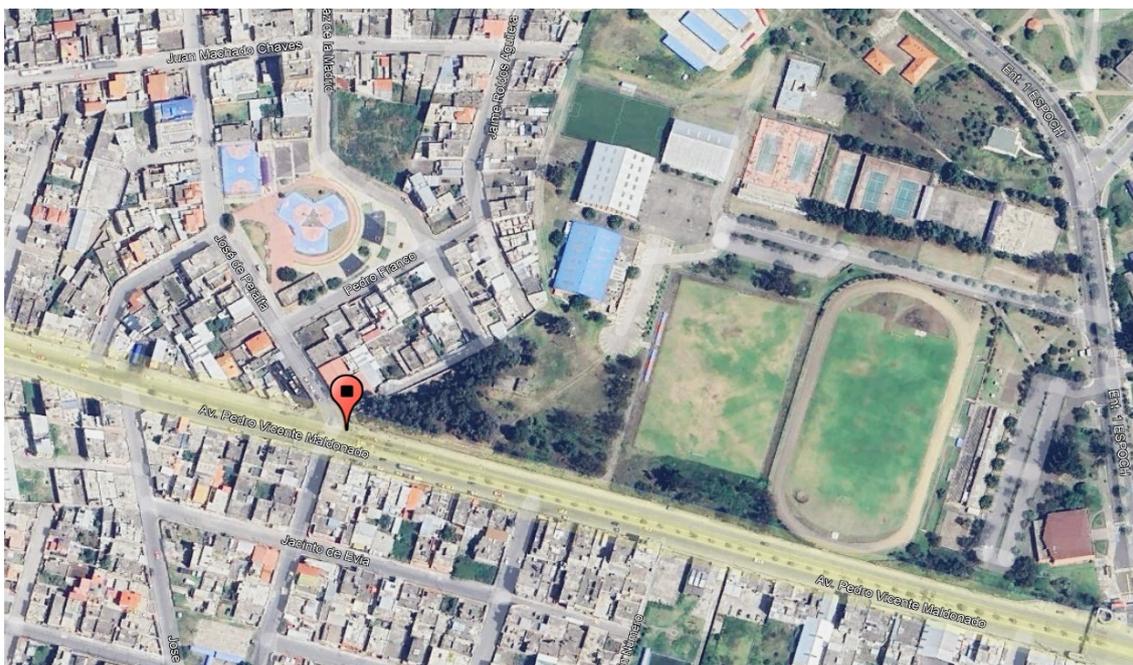
una amplia variedad de técnicas estadísticas como modelos lineales, no lineales, pruebas estadísticas, análisis de series temporales, clasificación, agrupamiento, entre otras (The R Foundation, 2010).

En el software se utilizaron diferentes librerías especializadas para la implementación de los modelos; entre las principales para ARIMA y SARIMA se utilizó *forecast* y *tseries*, para las redes neuronales *keras*, *tensorflow*, *nnfor*, para teoría de colas *queueing*; además las librerías *readxl*, *ggplot2*, *lubricate*, *Metrics* complementarias en los códigos para leer las bases de datos, realizar gráficos y evaluar los modelos.

3.7 Población y Muestra

Para el presente estudio, la población está constituida por los valores del registro histórico de datos proporcionados por el grupo de investigación de la ESPOCH, comprendido de forma mensual entre los años 2000 hasta el 2022 de la intersección entre la Av. Pedro Vicente Maldonado y la Calle José de Peralta (carril de subida) de la ciudad de Riobamba (Figura 5).

Figura 5. Ubicación de la intersección de los datos del tráfico vehicular



Fuente: (Google Earth Pro, 2025)

La información disponible abarca un total de 276 observaciones mensuales, correspondientes a 23 años de datos continuos. En cuanto a la muestra, se consideró el uso de la población completa, es decir, no se aplicó ningún tipo de muestreo. Esta decisión se justifica debido a la naturaleza del estudio, que tiene como objetivo la modelación histórica de la serie temporal para realizar predicciones a partir de todos los valores disponibles. El uso total de los datos permite asegurar una mayor precisión en los modelos estadísticos empleados y en los resultados obtenidos.

Capítulo 4

Resultados y Discusión

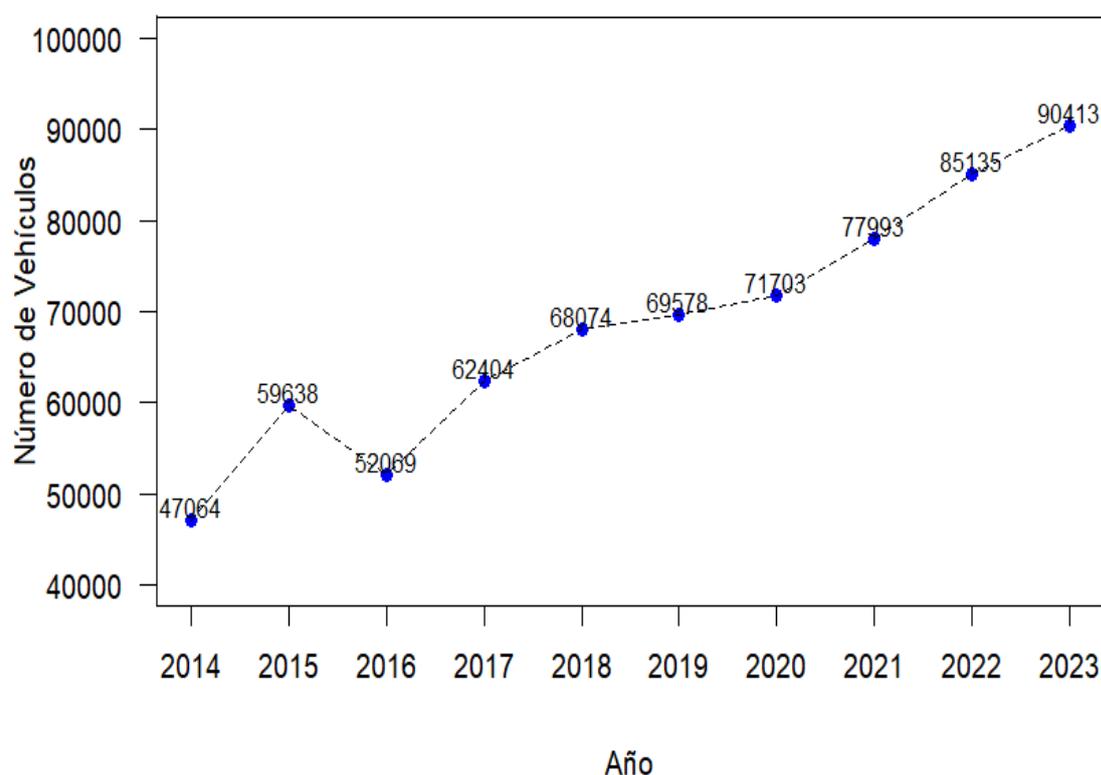
4.1 Análisis Descriptivo de los Resultados

4.1.1 Análisis vehicular de la ciudad de Riobamba

De acuerdo con el INEC (2024), el Ecuador ha experimentado un incremento en vehículos matriculados en un 6.42% para el año 2023 con un total de 3065967 que es evidente a lo largo de los años e indica su registro histórico desde el año 2014 hasta el 2023.

De igual manera, en la provincia de Chimborazo para el año 2023 existen 90413 vehículos matriculados con un crecimiento del 5.84%, es evidente el aumento contante desde el año 2016 hasta el 2023, a excepción que en el año 2016 ha existido un decremento con respecto del 2015 (Figura 6).

Figura 6. Serie histórica de vehículos matriculados en Chimborazo



Fuente: (INEC, 2024)

Tabla 1. Vehículos matriculados en los cantones de la provincia de Chimborazo en el año 2023.

Cantón	Número de Vehículos matriculados	Porcentaje (%)
Riobamba	66028	73,03
Alausí	3076	3,40
Colta	2623	2,90
Chambo	1873	2,07
Chunchi	1264	1,40
Guamote	2941	3,25
Guano	6,806	7,53
Pallatanga	491	0,54
Penipe	2050	2,27
Cumandá	2261	2,50
Total	90413	100

Fuente: (INEC, 2024)

La ciudad de Riobamba es la capital de la provincia de Chimborazo, representa el 73.03% de vehículos matriculados y para el año 2023, con un registro de 66028 unidades (Tabla 1). Es evidente el crecimiento vehicular en la ciudad de Riobamba, como destacan Mera Herrera & Colcha Ortiz (2025) que ha experimentado un aumento significativo en su parque automotor en los últimos años, lo que genera un mayor porcentaje de congestión y tráfico vehicular en las intersecciones.

De esta manera, se realizó el análisis de los datos de tráfico vehicular de la intersección de la Av. Pedro Vicente Maldonado y Calle José de Peralta en el carril de subida. Esta intersección es muy trascurrida especialmente en las horas pico debido a que se encuentra en una Institución de Educación Superior y la avenida alimenta a la Panamericana Norte y Sur E35 (Figura 7).

Los datos de tráfico vehicular están dados por el número de vehículos que pasan por la intersección, organizados cronológicamente en una serie temporal, que está representada

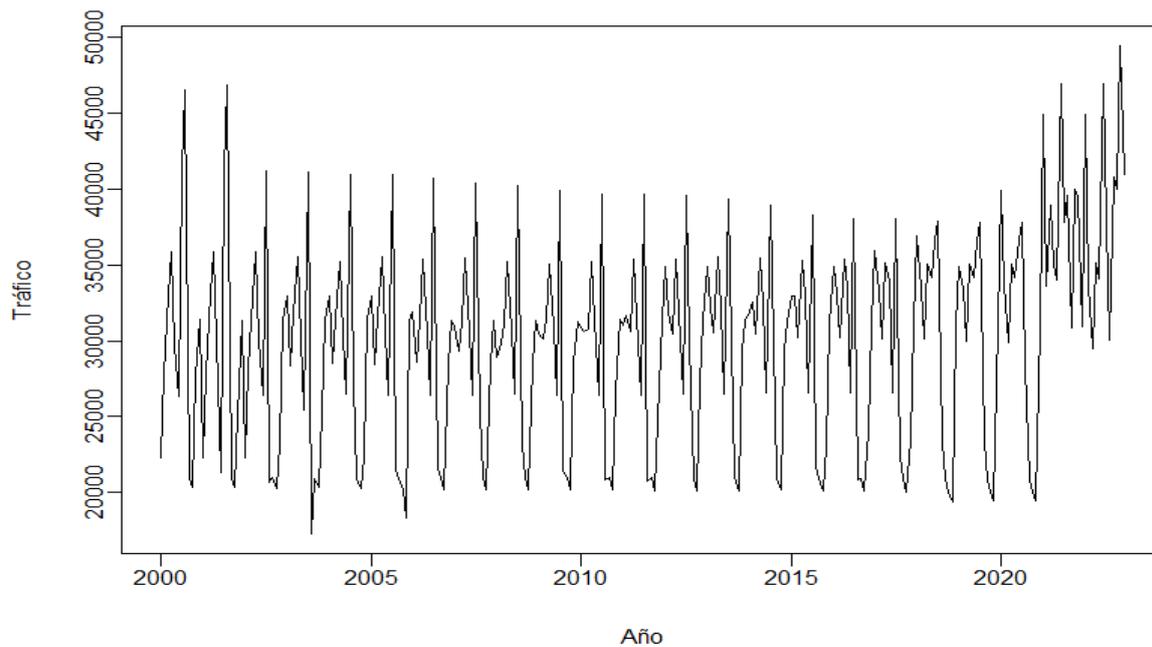
por el número de vehículos contados mensualmente de un registro histórico desde el año 2000 hasta el 2020, así como se plasma en la Figura 8.

Figura 7. Flujo vehicular de la Av. Pedro Vicente Maldonado hacia la Av. Monseñor Leónidas Proaño (Panamericana Norte E35) y hacia la Av. Panamericana Sur E35



Fuente: (GoogleEarth, 2024)

Figura 8. Serie de Tiempo Número de vehículos 2000-2022



Fuente: (ANT, 2025)

La figura 8 muestra la evolución del tráfico vehicular mensual en el período comprendido entre los años 2000 y 2022. En el eje horizontal se representan los años y en el eje vertical, el número de vehículos registrados. La línea conecta los datos mes a mes, permitiendo observar el comportamiento general del tráfico a lo largo del tiempo.

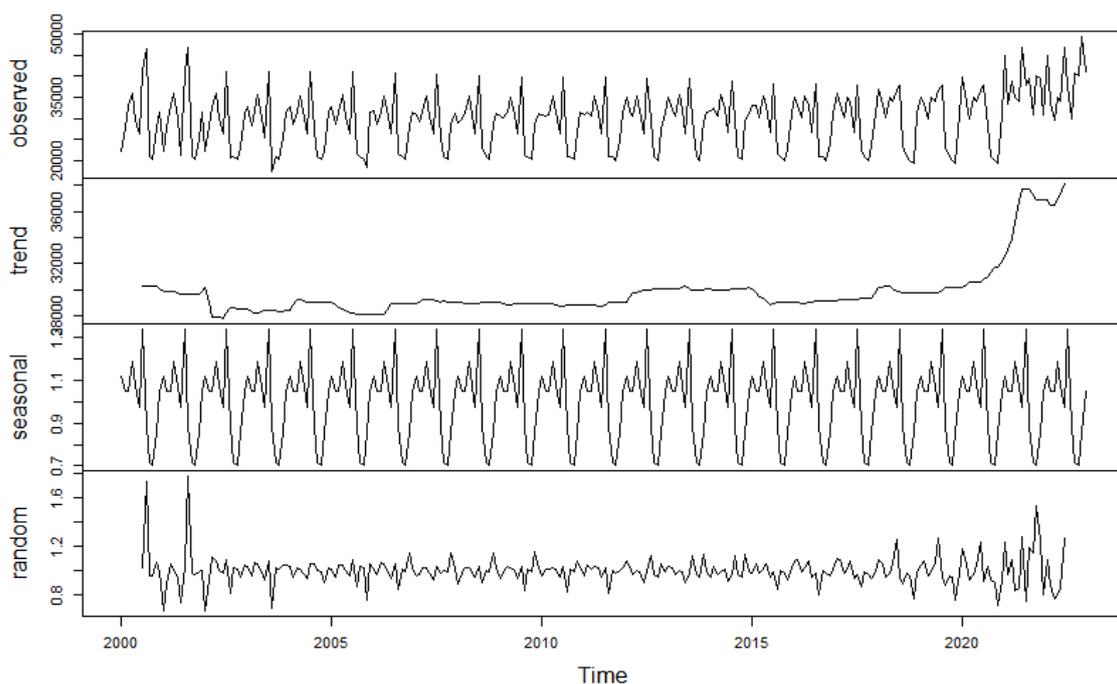
Una característica de la serie temporal es la presencia de estacionalidad, que es evidente en cada año ya que el tráfico presenta picos y caídas regulares; de la misma manera existe una tendencia de la serie que durante los años 2000 hasta el 2015 se mantiene estable; sin embargo, a partir del 2016 y más notorio en el periodo 2020 a 2022 existe un aumento progresivo y notorio en el volumen de tráfico.

4.2 Análisis de la serie temporal de los datos de tráfico vehicular para los modelos predictivos.

4.2.1 Descomposición multiplicativa

Se realizó el análisis de los datos vehiculares por medio de la gráfica de descomposición multiplicativa, en la cual se puede evidenciar las gráficas de los datos observados, tendencias, estacionalidad y residuos.

Figura 9. Descomposición multiplicativa de la serie temporal



En la Figura 9 se identifica lo siguiente:

- **Datos Observados:** indica el grafico de la serie temporal original, en la cual los datos siguen un patrón y una ciclicidad repetitiva durante cada año, así también existe una tendencia de crecimiento a partir del año 2020.
- **Tendencia:** la tendencia es contante a lo largo de todos los años hasta el 2020 que presenta un crecimiento significativo hacia el 2022.
- **Estacionalidad:** la gráfica muestra claramente que existe un patrón que se repite a lo largo del tiempo de manera cíclica coincidiendo los valores entre 1,3 y 0,7 durante todos los años.
- **Residuos:** desde el año 2000 hasta el 2002 y a partir del 2020 existe un aumento de ruido, los demás valores son bajos e indican la tendencia y estacionalidad.

4.2.2 Estacionariedad

Para verificar si la serie temporal es estacionaria, se aplicó la prueba de Dickey Fuller Aumentada (ADF) por medio del software R. Mediante esta prueba se pudo evaluar la presencia de la raíz unitaria de la serie y así determinar la aplicabilidad de los modelos de predicción ARIMA. A continuación, se presentan los resultados obtenidos:

Tabla 2. Resultados de la prueba de Dickey Fuller Aumentada

Augmented Dickey-Fuller Test	
Dickey-Fuller	-7.8475
Lag order	6
p-value	0.01
alternative hypothesis: stationary	

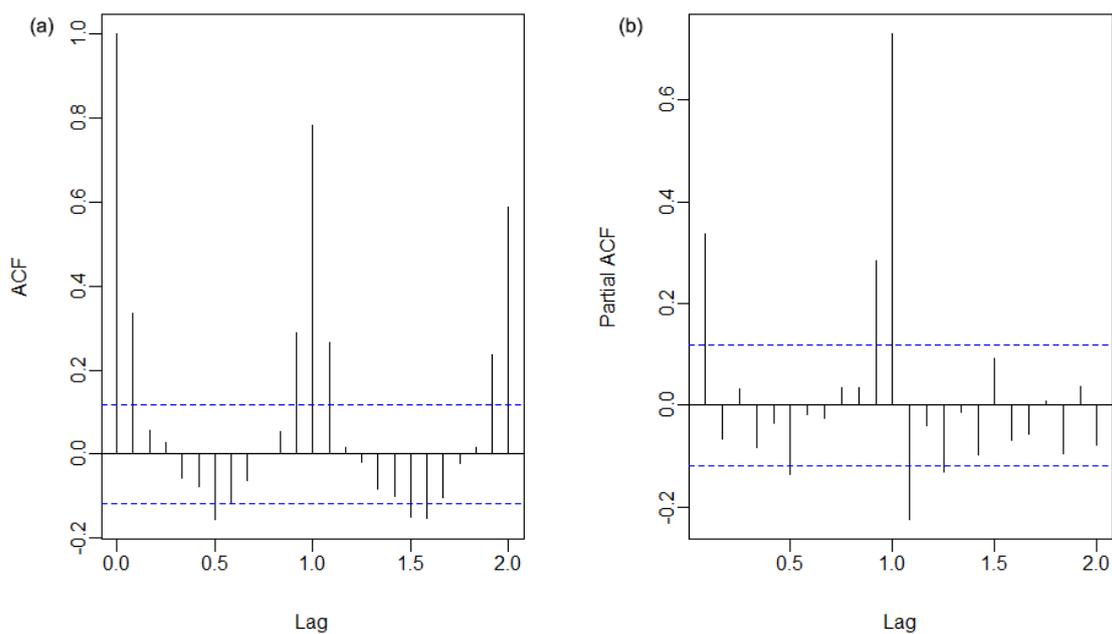
De acuerdo con la Tabla 2, el valor estadístico de prueba es de -7.8475, el número de retardos es 6 y el p valor es de 0.01, como es menor a 0.05 no hay suficiente evidencia

estadística para aceptar la hipótesis nula lo que indica que la serie es estacionaria, por lo tanto, no es necesaria la realización de diferenciaciones para realizar los modelos ARIMA y SARIMA ($d=0$).

4.2.3 Autocorrelación

Mediante la prueba de Dickey Fuller se ha obtenido el valor d para los modelos autorregresivos, ahora es importante determinar los valores de p (autorregresivo) y q (promedio móvil), que se lo realiza mediante la función de autocorrelación (ACF) y autocorrelación parcial (PACF) (Salcedo et al., 2007). Mediante el software R se generaron estas graficas indicadas en la Figura 10.

Figura 10. (a) *Función de autocorrelación (ACF)* y (b) *función de autocorrelación parcial (PACF)*



La imagen presenta dos gráficos fundamentales en el análisis de series temporales: la Función de Autocorrelación (ACF) a la izquierda y la Función de Autocorrelación Parcial (PACF) a la derecha. Ambos gráficos se utilizan para entender cómo los valores pasados de la serie (llamados lags) están relacionados con los valores actuales.

En el gráfico ACF, se observa que existe una fuerte autocorrelación estacional, lo que confirma que el tráfico tiene un comportamiento que se repite aproximadamente cada 12 meses. En el gráfico PACF, también se aprecia un pico importante en el primer rezago (lag 1), lo que sugiere que el valor del tráfico en un mes está influenciado por el valor del mes anterior. Después de ese punto, las barras bajan rápidamente, lo que sugiere que con un número bajo de rezagos es posible capturar la relación más importante en los modelos autorregresivos (AR).

Cabe indicar que para la presente investigación se utilizó la función autoarima para estimar los parámetros de los modelos autorregresivos.

4.3 Modelos predictivos de la serie temporal

4.3.1 Modelo ARIMA

A partir del análisis realizado en el software RStudio, se estimó un modelo ARIMA(1,1,1) para la serie temporal de tráfico vehicular. Los coeficientes obtenidos y sus errores estándar se presentan en la Tabla 3, junto con las métricas de evaluación del modelo:

Tabla 3. Resultados del modelo ARIMA (1,1,1)

Coefficientes:	ar1	ma1
	0.297	-0.9684
s.e.	0.061	0.0172

Que se describe mediante la siguiente ecuación:

$$X_t^d = c + 0.297X_{t-1}^d + \varepsilon_t^d - 0.9684 \varepsilon_{t-1}^d \quad (31)$$

Tabla 4. Métricas de evaluación del modelo ARIMA (1,1,1)

Métricas:	Estimación
RMSE	6263.28
p-value (Ljung-Box)	0.6997

La ecuación 31 representa el modelo predictivo ARIMA (1,1,1). En cuanto a su desempeño, el modelo presenta un Error Cuadrático Medio (RMSE) de 6.263,28, lo que indica una precisión moderada en la predicción.

Además, Mediante la prueba de Ljung Box aplicada a los residuos del modelo se determinó un valor de 0.6997, lo que permite aceptar la hipótesis nula; estableciendo que los residuos no presentan autocorrelación significativa y por ende existe ruido blanco, lo cual valida el ajuste del modelo.

Planteamiento de Hipótesis:

H_0 : *Los residuos no presentan autocorrelación significativa*

H_1 : *Los residuos presentan autocorrelación significativa*

A pesar del buen ajuste reflejado por la prueba de residuos, el valor del RMSE sugiere que el modelo podría no capturar completamente la estructura estacional de la serie. Por ello, en la siguiente sección se desarrolla el modelo SARIMA, que incorpora la componente estacional y podría mejorar el desempeño predictivo debido a la estacionalidad evidente en ellos por medio de la Figura 9.

4.3.2 Modelo SARIMA

De la igual manera, mediante el software R se ha estimado los parámetros para el modelo SARIMA(0,1,2) (0,1,0)₁₂, en el cual sus coeficientes son los siguientes:

Tabla 5. Resultados del modelo SARIMA(0,1,2) (0,1,0)₁₂

Coefficientes:	ma1	ma2
	-0.8177	-0.1348
s.e.	0.0586	0.0582

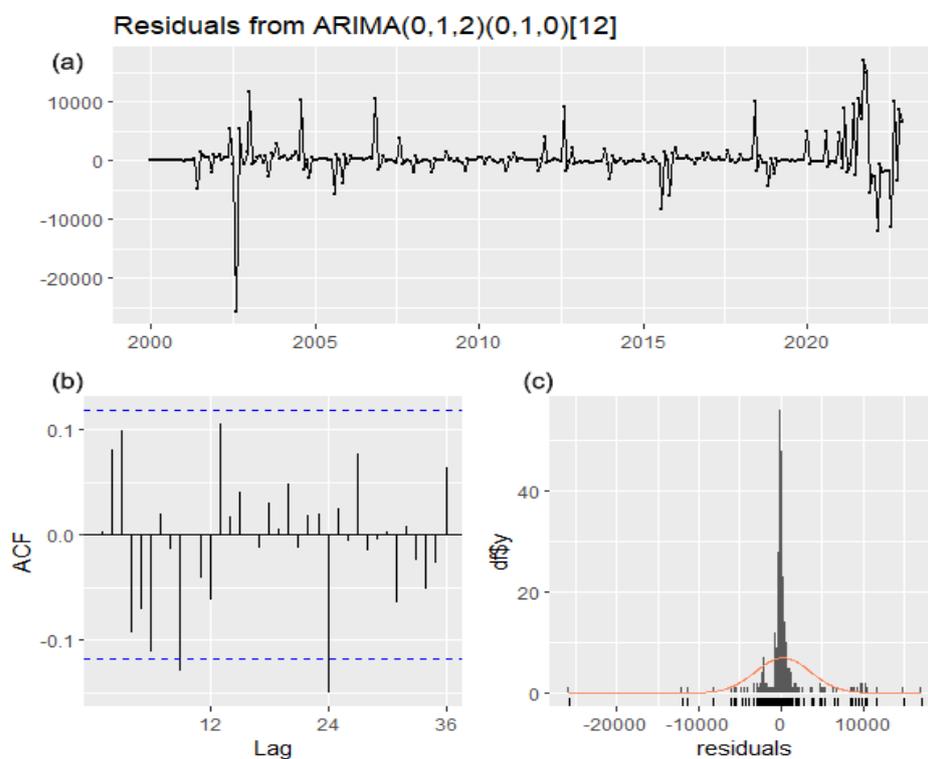
La ecuación que representa el modelo es la siguiente:

$$(X_t - X_{t-1}) - (X_{t-12} - X_{t-13}) = \varepsilon_t - 0.3944 \varepsilon_{t-1} + 0.0378 \varepsilon_{t-2} \quad (32)$$

Tabla 6. Métricas de evaluación del modelo $(0,1,2) (0,1,0)_{12}$

Métricas:	Estimación
RMSE	3.353,10
p-value (Ljung-Box)	0,9606

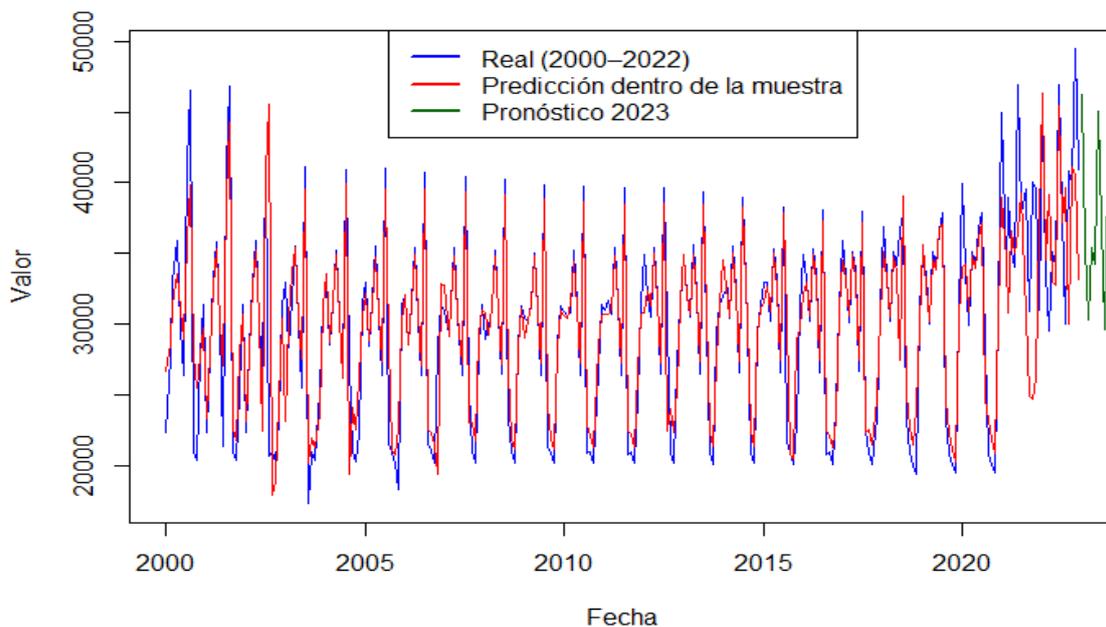
Mediante la prueba de Ljung Box se evaluó el modelo, que con un p valor = 0.9606, que significa que el modelo posee ruido blanco por lo tanto los residuos son aleatorios y el modelo está bien ajustado.

Figura 11. Análisis de los residuos del modelo $ARIMA(0, 1, 2)(0, 1, 0)_{12}$ 

Así también, mediante la Figura 11 (a) se puede identificar en el gráfico de los residuos que existe ciertos valores estables alrededor de cero a pesar de existir unos picos en los primeros y últimos años, en cuanto Figura 11 (b) de la función de autocorrelación indica que la mayoría todos los valores se encuentran dentro de las bandas de significación y Figura 11 (c) sobre el histograma los residuos indica que sigue una distribución normal a pesar de los grandes picos.

Adicionalmente, el modelo fue evaluado mediante el RMSE cuyo valor es de 3353,10. Este resultado indica un mejor desempeño en comparación con el modelo ARIMA, lo cual se evidencia en la Figura 12, que indica la gráfica de los valores reales en contraste con las predicciones y el pronóstico para el año 2023. La visualización refleja un buen comportamiento predictivo, lo que respalda la efectividad del modelo empleado.

Figura 12. Datos reales y predicciones del modelo $SARIMA(1,0,2)(1,0,1)_{12}$



4.3.3 Modelos de redes neuronales

4.3.3.1 Redes neuronales multicapa MLP

Las redes neuronales multicapa o perceptrón multicapa MLP tienen un gran potencial de aprender a representar correlaciones y dependencias entre un conjunto de entrada y salida (Mrudula et al., 2024). En series temporales permiten analizar y obtener pronósticos precisos porque comprenden los patrones y tendencias con respecto al tiempo; es decir comprenden el conjunto de nodos que están organizados por medios de tres capas, una capa de entrada, al menos una capa oculta y una capa de salida (Menacho, 2014).

De acuerdo con ello, Mediante la librería *nnfor* del software R, se estableció el modelo propuesto que tiene 200 repeticiones de entrenamientos, 12 neuronas de entrada

(retardos) ; dos capas ocultas (una con 12 neuronas y la otra con 6 neuronas) cada una interconectadas con la capa anterior y una capa de salida que predice los valores del número de vehículos en función del tiempo. Además, en concordancia con Menacho (2014) el tipo de conexión es hacia adelante y la función de activación es sigmoideal para la capa de entrada y capas ocultas y lineal para la capa de salida.

En la Figura 13, se indica el modelo de la red neuronal multicapa para los datos de la serie temporal que se está trabajando con 12 retardos de acuerdo a la estacionalidad de los datos mensuales alrededor de un año (12 meses), de esta manera el modelo MLP utiliza 12 neuronas de entrada (input).

Figura 13. Modelo de la red neuronal multicapa para la serie temporal

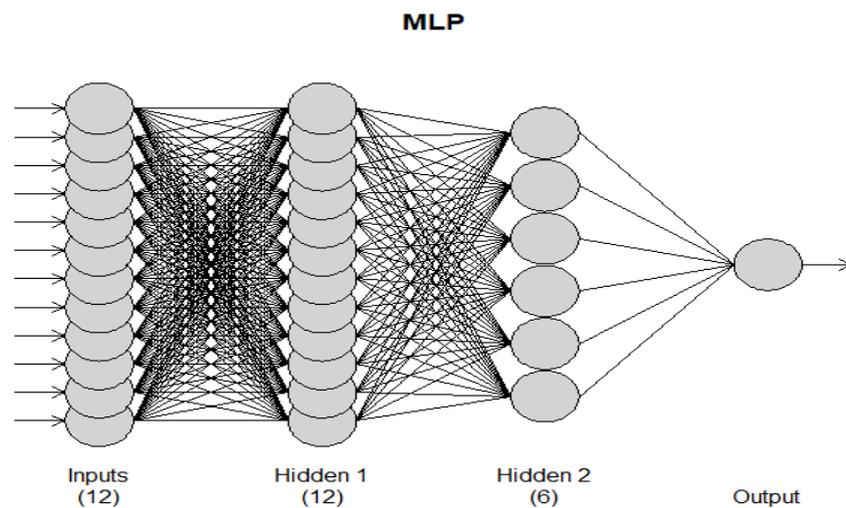
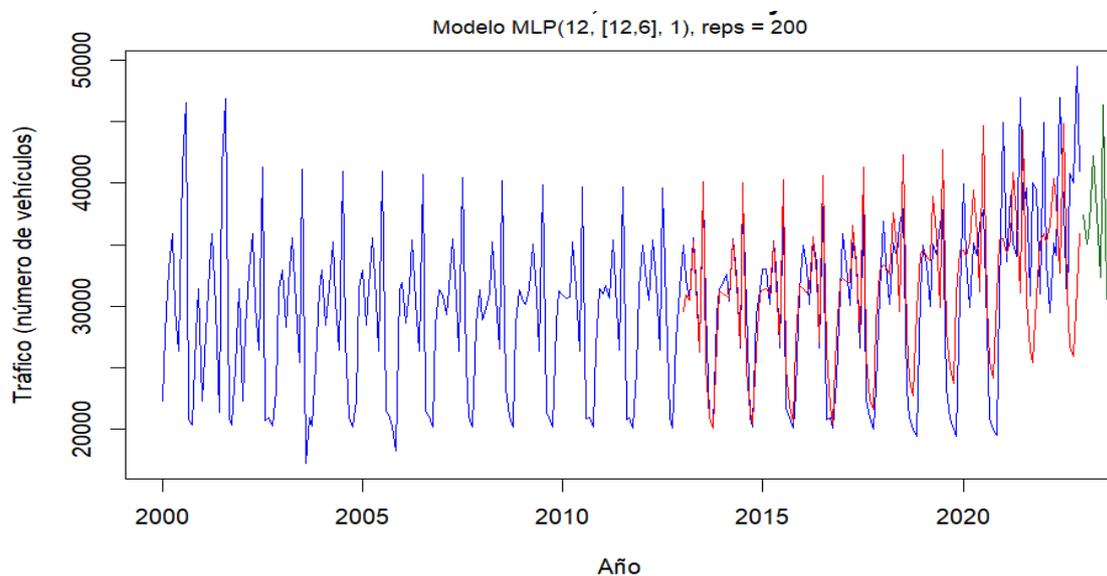


Figura 14. Datos reales, predicción y pronóstico del Modelo MLP (12, [12,6], 1) reps=200.



El mediante el modelo propuesto de MLP se estableció los valores de predicción que se observan en la figura 14, en la cual se grafica los datos reales, las predicciones de los datos de prueba y la predicción para el año 2023; se puede evidenciar que el modelo predictivo genera valores con un patrón estacional que posee un crecimiento constante, que de acuerdo a la métrica de evaluación del modelo el valor de RMSE es de 5150,16.

4.3.3.2 Redes neuronales recurrentes RNN LSTM

En las series temporales con valores tomados de la vida real presentan un problema en el aprendizaje y en las predicciones siendo un análisis autorregresivo; en este sentido las redes neuronales recurrentes son estructuras especializadas para aprender modelos con gran complejidad en los datos (Cantuña & Almeida, 2024).

El modelo implementado se basa en neuronales recurrentes LSTM por su capacidad para procesar información de datos secuenciales y tomar en cuenta las dependencias temporales; en el presenta caso se tiene series temporales mensuales.

Mediante la librería *keras* del software R se implementó la aplicación de las redes neuronales recurrentes LSTM, por la naturaleza de los datos; han sido tomados de forma temporal continua para conservar la dependencia temporal.

4.3.3.2.1 Hiperparámetros de las RNN LSTM

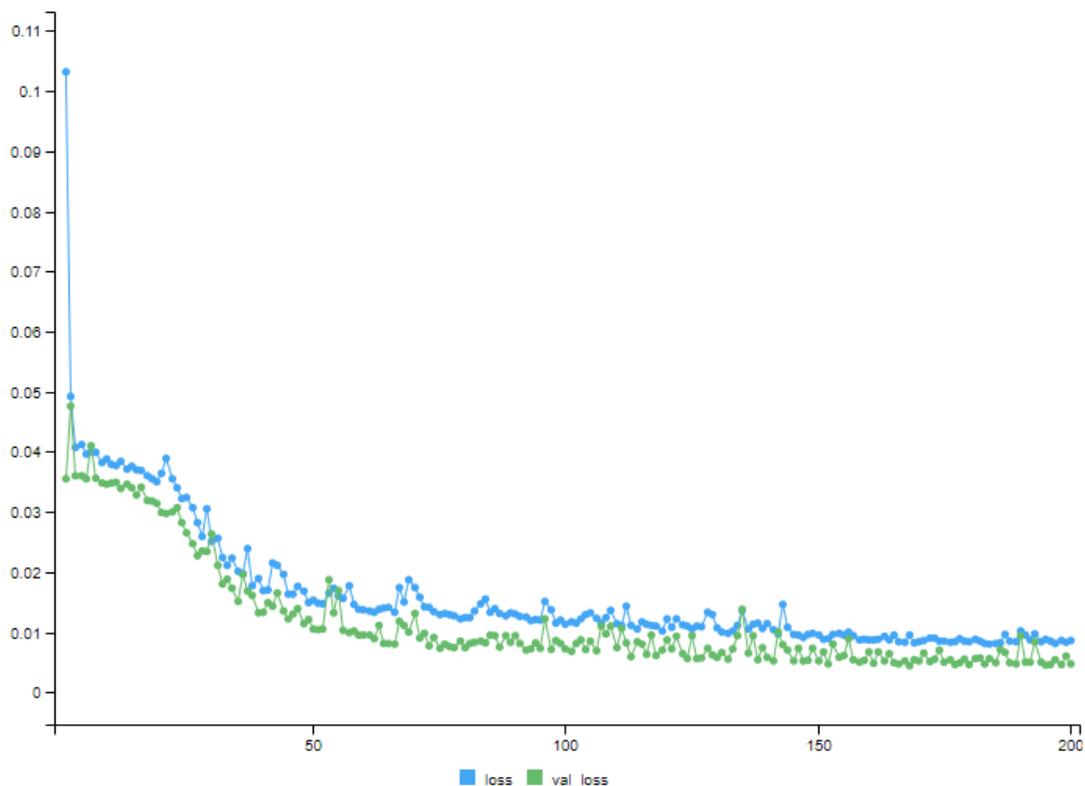
Para la construcción del modelo de RNN LSTM se ha considerado tres capas, dos ocultas y una capa densa de salida. La primera capa posee 50 neuronas, la segunda tiene 25 neuronas y la capa densa de salida posee una sola neurona que permite la predicción final. La longitud de la secuencia de entrada es de 12 debido a que toma ese número de observaciones mensuales pasadas para la predicción del valor del mes siguiente, esto es debido al patrón estacional y de tendencia que existe anualmente de acuerdo a los datos.

Como función objetivo se ha establecido el Error Cuadrático Medio MSE que evalúa la diferencia entre el valor calculado y el valor R, así también se ha considerado

como optimizador al algoritmo Adam que permite una convergencia rápida, estable y halla mejores mínimos por su comportamiento refinado con gradientes dispersos (Zohra, 2024).

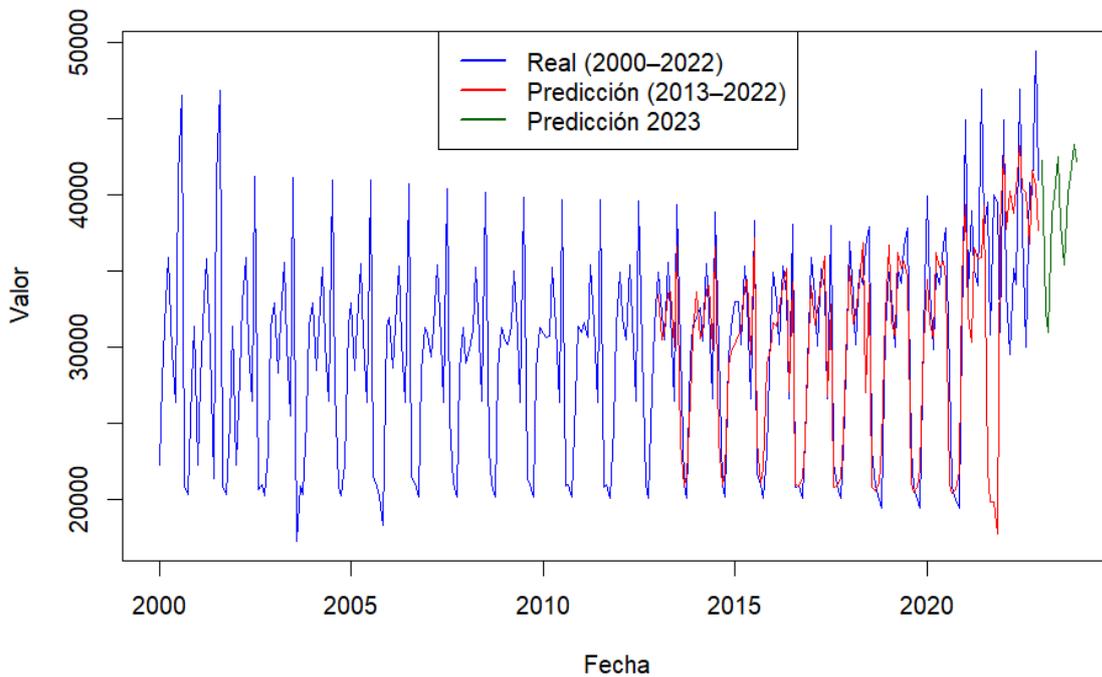
Para el entrenamiento se lo realizó mediante 200 iteraciones (número de épocas), con un tamaño de lote de 16 y establece que el 20 % de los datos de entrenamiento son utilizados para la validación, cabe indicar que no se utiliza la técnica de dropout para parar el entrenamiento debido a que no se evidencian sobreajustes en el número de iteraciones establecidas.

Figura 15. *Análisis de datos para entrenamiento y validación del modelo RNN LSTM*



La Figura 15 describe el comportamiento de la pérdida o error del modelo a lo largo de las épocas de entrenamiento, se evidencia que mientras aumentan las iteraciones (épocas) la curva tiende a llegar a cero y verificar que el valor de 200 iteraciones es correcto ya que al final se estabilizar el modelo y no se presenta sobreajustes.

Figura 16. Datos reales y predicciones del modelo RNN LSTM (2000-2023)



De esta manera, se realizó las predicciones que se puede observar en la Figura 16, que muestra la gráfica de los valores reales de los datos de tráfico, la predicción para los años de prueba (2013-2022) y la predicción para el año 2023. Cabe indica que el modelo RNN LSTM evaluado mediante la métrica del RMSE tiene un valor de 4694.05.

4.3.4 Comparación entre los modelos de predicción

Los modelos analizados anteriormente presentan un comportamiento predictivo estacional de acuerdo a la naturaleza de los datos obtenido para la comparación de resultados, se analiza los modelos SARIMA, redes neuronales MLP y redes neuronales recurrentes RNN LSTM.

De acuerdo con la Figura 12, 14 y Figura 16; el modelo SARIMA presenta un mejor comportamiento comparando los datos reales con las predicciones (Figura 12), del mismo modo el comportamiento grafico respalda los resultados de la métrica de evaluación RMSE que alcanza un valor de 3353.10, que indica un mejor desempeño de predicción (Tabla 7).

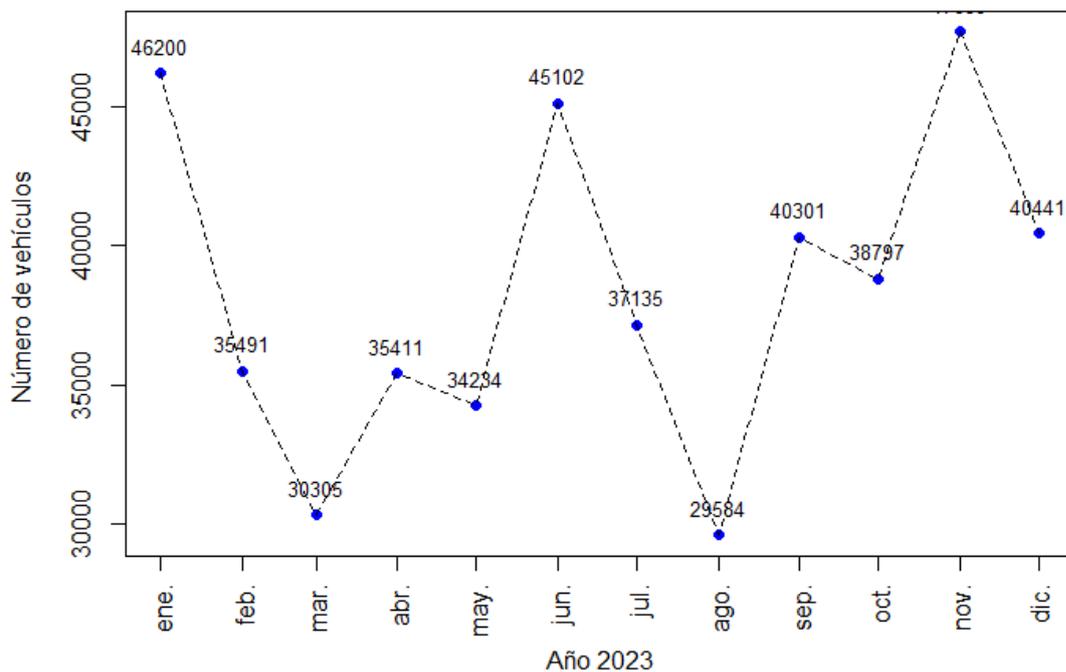
Tabla 7. Evaluación de los modelos de predicción mediante el RMSE

Modelo de predicción	RMSE
SARIMA	3353,10
Redes Neuronales MLP	5150,16
Redes Neuronales RNN LSTM	4694.05

De esta manera, el modelo utilizado para las predicciones es SARIMA (1,0,2)(1,0,1)12; en la Tabla 8 y en la Figura 17 se puede identificar los valores de predicción para el año 2023 y puede aplicarse para el año que requiera ser estudiado.

Tabla 8. Valores de predicción para el año 2023

Predicciones para el año 2023	
Mes	Numero de vehículo
Enero	46200
Febrero	35491
Marzo	30305
Abril	35411
Mayo	34234
Junio	45102
Julio	37135
Agosto	29584
Septiembre	40301
Octubre	38797
Noviembre	47689
Diciembre	40441

Figura 17. Predicción de valores de número de vehículos para el año 2023

La gráfica muestra las predicciones mensuales del número de vehículos para el año 2023, generadas mediante el modelo SARIMA (0,1,2)(0,1,0)[12], que considera tanto componentes estacionales como no estacionales. Se observa un comportamiento fluctuante a lo largo del año, con variaciones significativas entre los distintos meses. El valor más alto estimado corresponde al mes de noviembre, con 47.689 vehículos, seguido de enero (46.200) y junio (45.102), lo que sugiere una mayor circulación en estos periodos. Por el contrario, los meses con menor número de vehículos proyectados son agosto (29.584) y marzo (30.305), lo que podría deberse a una disminución de actividades o menor demanda de movilidad.

Estos patrones reflejan una marcada estacionalidad en el tráfico vehicular, lo cual es coherente con las características del modelo SARIMA empleado. Esta información resulta útil para la planificación operativa y la toma de decisiones estratégicas en función de los picos y descensos esperados en la circulación vehicular.

4.4 Modelo de Teoría de Colas

Para complementar el estudio del tráfico vehicular, se utilizó el modelo de teoría de colas con múltiples canales que en el presente caso son dos canales M/M/2 al existir dos carriles en la avenida de estudio, adicionalmente como datos λ y μ se establecieron de acuerdo al análisis del número de vehículos de la intersección en horas pico y son los siguientes:

$\lambda = 45$ vehículos que llegan a la intersección por minuto

$\mu = 50$ vehículos que son atendidos por minuto

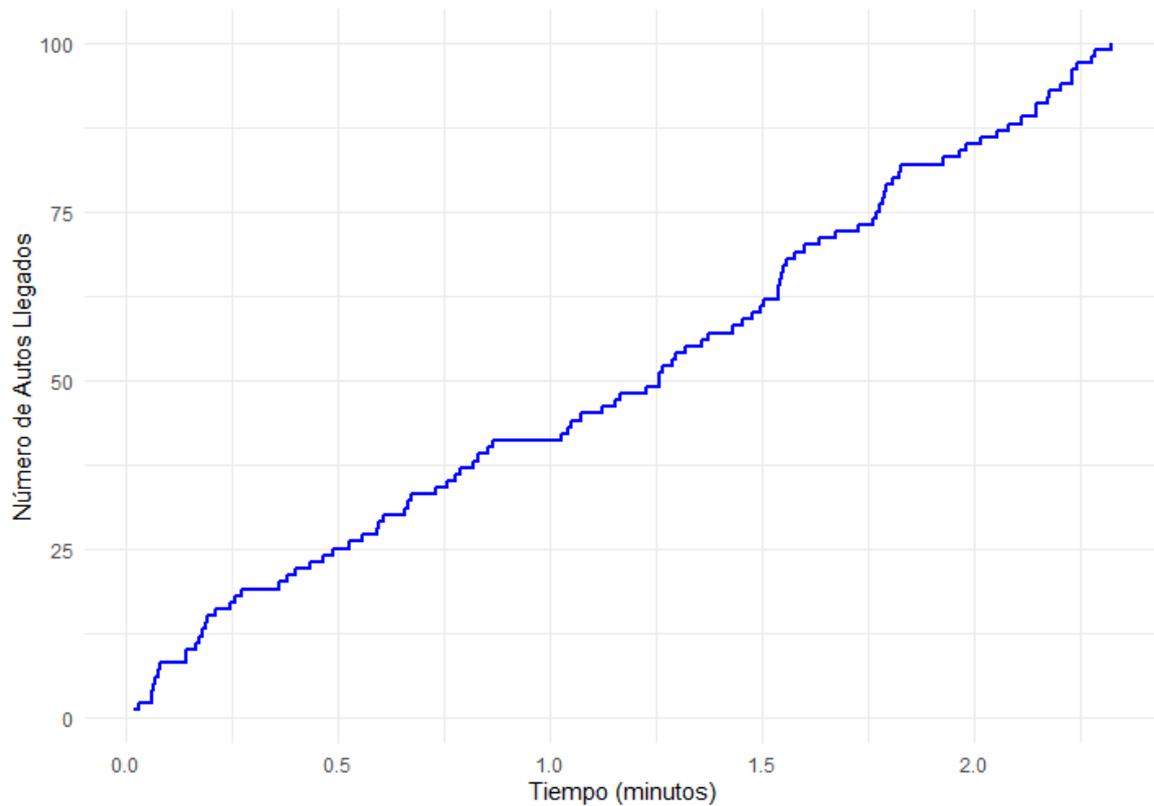
Tabla 9. Resultados de los parámetros operacionales de la intersección del modelo de teoría de colas con múltiples canales M/M/2

Parámetro	Valor	Unidad
Tiempo promedio en el sistema [W]	0.025	Min
Tiempo promedio de esperar en la cola [Wq]	0.005	Min
Número promedio de vehículos en la carretera [L]	1.106	Vehículos
Número promedio de vehículos esperando [Lq]	2.06	Vehículos
Probabilidad de congestión	0.45	-
Probabilidad de que un vehículo tenga que esperar [Pw]	0.279	-

En la Tabla 9 se indican los resultados de la aplicación del modelo de teoría de colas con 2 canales de servicio, en los cuales el tiempo promedio en el sistema es de 0.025 minutos equivalente a 1.5 segundos, el tiempo promedio de esperar en la cola es de 0.005 minutos es decir 0.3 segundos, el número promedio de vehículos en la vía es aproximadamente 3 vehículos, de igual forma el número promedio de vehículos esperando es de 3 vehículos; así también la probabilidad de congestión del 45% y la probabilidad que un vehículo tenga que esperar es de 28%.

La simulación del proceso de Poisson de número de vehículos que llegan durante el tiempo se identifica en la Figura 18, representando la simulación estocástica y aleatoria de la llegada de vehículos a la intersección, en concordancia con el valor de $\lambda = 45$ vehículos que llegan a la intersección por minuto, por de aproximadamente a los 2 minutos llegarían 100 vehículos, además se describe como en los tramos con pendiente fuertes los vehículos llegan rápidamente en contraste de los tramos con pendientes nulas que no llegan vehículos.

Figura 18. Simulación del Proceso de Poisson de vehículos de llegada



4.5 Discusión de los Resultados

El crecimiento vehicular a nivel nacional y local es evidente de acuerdo a las cifras proporcionadas por el INEC (2024) con un incremento vehicular del 6.42% del año 2022 al 2023 en el país y un 5.84% en la provincia de Chimborazo. Este crecimiento genera cada vez más problemas en Ingeniería de tránsito como congestión vehicular y déficit de planificación vial en las diferentes zonas del país y en la ciudad de Riobamba.

La clave principal del estudio se basó en la obtención de datos históricos del volumen de tráfico vehicular de una intersección de la ciudad de Riobamba para poder analizarlos; en este caso los datos fueron proporcionados por la ANT de la intersección de la Av. Pedro Vicente Maldonado y calle José de Peralta, una avenida concurrida a lo largo del día especialmente en horas pico.

Varios estudios sustentan y son base de la investigación, en los cuales se realizó predicciones del volumen del tráfico como Boukerche et al. (2020) que utiliza modelos estadísticos basados en la familia AR/MA, Cascante & Alarcón (2018) y Mera Herrera & Colcha Ortiz (2025) desarrollan modelos ARIMA, Rivera & Meneses (2022) destacan los modelos SARIMA, SNAIVE y NNAR, Kumar & Vanajakshi (2015) con un modelo SARIMA, (Van Der Bijl et al., 2022) adiciona redes neuronales recurrentes RNN LSTM, Mrudula et al. (2024) analiza algoritmos K-Nearest Neighbours (KNN), perceptrón multicapa (MLP) y red bayesiana. Además destacan la evaluación de los modelos mediante el error cuadrático medio (MSE), la raíz del error cuadrático medio (RMSE), error porcentual absoluto medio (MAPE) y error absoluto medio (MAE).

De esta manera, el análisis de los datos se ejecutó mediante los modelos estadísticos de predicción ARIMA, SARIMA, redes neuronales MLP y RNN LSTM; siendo el mejor resultado de predicción que se obtuvo mediante el modelo SARIMA(0,1,2)(0,1,0)₁₂, cuyo valor del RSME fue el más bajo de 3353,10 en contraste con los otros modelos; con lo que se evidencia claramente la estacionalidad.

Mediante el análisis de la serie temporal se pudo identificar que los datos poseen un patrones de ciclicidad hasta aproximadamente el año 2015, sin embargo a partir de este año y más evidente en el 2020 poseen una tendencia en crecimiento hasta el 2022; existiendo en los datos una estacionalidad y estacionariedad de acuerdo la figura 8, Figura 9, así como también a la prueba de Dickey Fuller Aumentada ADF.

Los resultados obtenidos de la predicción de los datos se deben a la naturaleza del fenómeno del flujo vehicular y al caso de la intersección estudiada; con similitud al estudio de Kumar & Vanajakshi (2015) que utilizan un modelo de ARIMA estacional (SARIMA) y obtiene resultados aceptables para la aplicación de sistemas de transporte inteligentes. Pero con diferencia a Cascante & Alarcón (2018) y Mera Herrera & Colcha Ortiz (2025)

resuelven por medio de modelos ARIMA, evidenciando tendencias, estacionalidad y ciclicidad con resultados adecuados y satisfactorios.

Por otro lado, para tener un análisis adicional del comportamiento vehicular de la intersección se realizó el estudio de teoría de colas con múltiples canales por tener en la vía dos carriles, se determinó que la probabilidad de congestión es del 45 % y la probabilidad de que un vehículo tenga que esperar es de 28%, que son valores que coinciden con la realidad del sitio, especialmente en horas pico.

Conclusiones

Los datos de la serie temporal fueron recopilados por medio del grupo de investigación de la ESPOCH, de la intersección de la Av. Pedro Vicente Maldonado y Calle José de Peralta. Los valores representan el número histórico de vehículos desde el año 2000 hasta el 2022 de forma mensual y se analizó mediante los modelos estadísticos ARIMA, SARIMA, redes neuronales MLP y RNN LSTM para identificar patrones y tendencia de los datos de tráfico vehicular.

Se evidencia que la serie temporal posee estacionalidad a lo largo de los datos con un patrón repetitivo y comportamiento cíclico, siguiendo una tendencia en crecimiento a partir del año 2015 hasta el 2022. De esta manera, el modelo que mejor se ajustó a estas características fue el SARIMA; que mediante la gráfica y el valor del RMSE de 3353,10 representa el más bajo en comparación con los otros modelos.

Se obtuvo los datos hasta el 2022 y se realizó una predicción únicamente para el 2023, dado que a más años de proyección los errores se incrementan y los modelos estadísticos como ARIMA y SARIMA ofrecen mejor precisión en plazos cortos o medianos. El modelo SARIMA, que presentó menor error mostró que en el 2023 gran variación mensual con grandes picos en diciembre, enero y junio; y disminuciones en los meses de agosto y marzo. Que plasmado con la realidad del sector es debido a la disminución de actividades por lo que existe menor demanda de movilidad; además, esto refleja la estacionalidad que el tráfico posee y la coherencia del modelo utilizado para las predicciones.

A partir del estudio de teoría de Colas se pudo comprobar el comportamiento vehicular en la intersección; en la cual el tiempo promedio en el sistema es de 0.025 minutos equivalente a 1.5 segundos, el tiempo promedio de esperar en la cola es de 0.005 minutos es decir 0.3 segundos, el número promedio de vehículos en la vía es

aproximadamente 3 vehículos, de igual forma el número promedio de vehículos esperando es de 3 vehículos; así también la probabilidad de congestión del 45% y la probabilidad que un vehículo tenga que esperar es de 28%.

La variación entre un modelo y otro que presenta mejores resultados de acuerdo a diferentes autores es debido a la naturaleza de los datos, en el presente estudio la estacionalidad es un factor clave al momento de evaluación del modelo; por esta razón el modelo que presenta mejores resultados es el SARIMA que refleja un buen comportamiento para las predicciones.

Con estos resultados, es importante destacar que los modelos estadísticos permiten realizar predicción de datos de volúmenes de tráfico vehicular y simular su comportamiento, lo que permite la toma de decisiones al momento de un diseño vial, aplicación de semaforización, optimización de tiempos de viajes y mejorar la planificación vial.

Recomendaciones

Para el presente estudio la información de los datos fue proporcionada por la ANT, se recomienda que para próximas investigaciones se pueda ampliar a diferentes intersecciones de las cuales se cuente con una base de datos histórica similar a la presentada y pueda ayudar de mejor manera al análisis del tráfico.

El modelo SARIMA(1,0,2)(1,0,1)₁₂ es el modelo que mejor comportamiento presenta al analizar los datos históricos de la intersección estudiada, cabe indicar que en los sectores aledaños pueda existir el mismo comportamiento; pero sería adecuado realizar el estudio similar o con otros modelos para una correcta predicción.

Adicionalmente, los datos corresponden a un análisis de serie temporal de varios años, es decir largo plazo. Si se realiza otras investigaciones con datos de corto plazo, también podría variar de acuerdo con los modelos analizados.

Los modelos estadísticos son una forma correcta de estimar y realizar predicciones de datos del tráfico vehicular que pueden utilizarse para el diseño vial de pavimentos, así también la planificación de una intersección para evitar el deterioro constante de las vías y congestión vehicular en zonas de alto tráfico.

De acuerdo con el análisis de tráfico de la intersección analizada en las horas pico existe gran afluencia de tráfico y congestionamiento por lo cual es recomendable la utilización de un semáforo programable para los diferentes horarios que abastezca la demanda y logre que exista fluidez en el paso de los vehículos.

Por otro lado, no se evidencia espacios destinados a ciclovía, que hoy en día es una forma de movilización alternativa y podría reducir la dependencia de los vehículos particulares en tramos cortos.

Referencias Bibliográficas

- Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*.
<https://doi.org/https://doi.org/10.1007/978-3-319-94463-0>
- Agyemang, E. F., Mensah, J. A., Ocran, E., Opoku, E., & Northey, E. N. N. (2023). Time series based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints. *Heliyon*, 9(12).
<https://doi.org/10.1016/j.heliyon.2023.e22544>
- Alsubai, S., Dutta, A. K., & Sait, A. R. W. (2024a). Hybrid deep learning-based traffic congestion control in IoT environment using enhanced arithmetic optimization technique. *Alexandria Engineering Journal*, 105, 331–340.
<https://doi.org/10.1016/j.aej.2024.06.072>
- Alsubai, S., Dutta, A. K., & Sait, A. R. W. (2024b). Hybrid deep learning-based traffic congestion control in IoT environment using enhanced arithmetic optimization technique. *Alexandria Engineering Journal*, 105, 331–340.
<https://doi.org/10.1016/j.aej.2024.06.072>
- Arias, D. C., Salvador, M. A., Hidalgo, R., León, J. M., Parrado, C., Meléndez, G., Simbaña, L., Villavicencio, J., Arteaga, C., & Yopez, D. (2023). *Informe de Calidad de Vida 2023*.
- Ashhad, T., Cabrera, F., & Roa, O. (2020). Análisis del congestionamiento vehicular para el mejoramiento de vía principal en Guayaquil-Ecuador. *Revista Gaceta Técnica*, 21, 4–23. <https://doi.org/https://doi.org/10.13140/RG.2.2.21905.04960>
- Balawi, M., & Tenekeci, G. (2024). Time series traffic collision analysis of London hotspots: Patterns, predictions and prevention strategies. *Heliyon*, 10(4).
<https://doi.org/10.1016/j.heliyon.2024.e25710>

- Boukerche, A., Tao, Y., & Sun, P. (2020). Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. In *Computer Networks* (Vol. 182). Elsevier B.V. <https://doi.org/10.1016/j.comnet.2020.107484>
- Boukerche, A., & Wang, J. (2020). A performance modeling and analysis of a novel vehicular traffic flow prediction system using a hybrid machine learning-based model. *Ad Hoc Networks*, 106. <https://doi.org/10.1016/j.adhoc.2020.102224>
- Bucay, J., Escudero, P., Paredes, J., & Ayala, M. (2023). Leveraging Classical Statistical Methods for Sustainable Maintenance in Automotive Assembly Equipment. *Sustainability (Switzerland)*, 15(21). <https://doi.org/10.3390/su152115604>
- Burbano, P., & Cerón, E. (2007). *Análisis de Tráfico vehicular en una zona crítica de la ciudad de Quito*. Escuela Politécnica Nacional.
- Cal Mayor, R. R., & Cárdenas, J. G. (2018). *Ingeniería de tránsito. Fundamentos y aplicaciones* (Alfaomega Grupo Editor, Ed.; 9na ed.). <http://www.alfaomega.com.co>
- Cantuña, J., & Almeida, C. (2024). *Marco de evaluación para modelos predictivos de series de tiempo heterogéneas usando inteligencia artificial*.
- Cascante, S., & Alarcón, R. (2018). *Estimación del flujo vehicular a través de series de tiempo*.
- Daniel, A., Karthik, S., & Rajkumar, N. (2021). Optimizing spectral efficiency based on poisson queuing model for procuring cognitive intelligence in vehicular communication. *Microprocessors and Microsystems*, 82. <https://doi.org/10.1016/j.micpro.2020.103819>

- Do, L. N. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, *108*, 12–28. <https://doi.org/10.1016/j.trc.2019.09.008>
- Escudero, P., Alcocer, W., & Paredes, J. (2021). Recurrent neural networks and ARIMA models for euro/dollar exchange rate forecasting. *Applied Sciences (Switzerland)*, *11*(12). <https://doi.org/10.3390/app11125658>
- Fahs, W., Chbib, F., Rammal, A., Khatoun, R., Attar, A. El, Zaytoun, I., & Hachem, J. (2023). Traffic Congestion Prediction Based on Multivariate Modelling and Neural Networks Regressions. *Procedia Computer Science*, *220*, 202–209. <https://doi.org/10.1016/j.procs.2023.03.028>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. del P. (2014). *Metodología de la Investigación*.
- Herrera, D., & Pérez, F. (2019). *Predicción de demanda eléctrica mediante la aplicación de modelos Arima y Sarima en lenguaje de programación R - Caso de estudio en la Empresa Eléctrica Quito*. Escuela Politécnica Nacional.
- INEC. (2024). *Anuario de Estadísticas de Transporte, 2023*.
- Kosun, C., & Ozdemir, S. (2016). A superstatistical model of vehicular traffic flow. *Physica A: Statistical Mechanics and Its Applications*, *444*, 466–475. <https://doi.org/10.1016/j.physa.2015.10.042>
- Kumar, S. V. (2017). Traffic Flow Prediction using Kalman Filtering Technique. *Procedia Engineering*, *187*, 582–587. <https://doi.org/10.1016/j.proeng.2017.04.417>

- Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*, 7(3). <https://doi.org/10.1007/s12544-015-0170-8>
- Menacho, C. H. (2014). Comparación de los métodos de series de tiempo y redes neuronales. *Anales Científicos*, 75(2), 245. <https://doi.org/10.21704/ac.v75i2.960>
- Mera Herrera, W. E., & Colcha Ortiz, A. R. (2025). Análisis estadístico del flujo de tráfico en intersecciones del centro de la ciudad de Riobamba. *Revista Científica Multidisciplinaria InvestiGo*, 6(14), 692–708. <https://doi.org/10.56519/szj4zh91>
- Ministerio de Transporte y Obras Públicas. (2023). *Política Nacional de Movilidad Urbana Sostenible del Ecuador*.
- Moyolema Chaglla, Á. G., Toscano Guerrero, E., Toalombo Rojas, B. M., & Killkana, R. (2022). Simulador de tráfico vehicular mediante el modelo matemático macroscópico LWR. *Killkana Técnica*, 5(1), 19–32. <https://doi.org/10.26871/killkanatecnica.v5i1.959>
- Mrudula, S. T., Meenakshi, Ritonga, M., Sivakumar, S., Jawarneh, M., F, S., Keerthika, T., Rane, K. P., & Roy, B. (2024). Internet of things and optimized knn based intelligent transportation system for traffic flow prediction in smart cities. *Measurement: Sensors*, 35, 101297. <https://doi.org/10.1016/j.measen.2024.101297>
- Nassiri, H., Mohammadpour, S. I., & Dahaghin, M. (2023). Forecasting time trend of road traffic crashes in Iran using the macro-scale traffic flow characteristics. *Heliyon*, 9(3). <https://doi.org/10.1016/j.heliyon.2023.e14481>
- NEVI-12-MTOP. (2013). *Norma para estudios y diseños viales*.

- Paguay, J., & Bejarano, C. (2020). *Estudio de teoría de colas en el área de información de la Dirección de Movilidad Tránsito y Transporte del GADM Riobamba.*
- Pilamunga, M., & Ríos Insua, D. (2014). Un modelo de predicción de tráfico en la ciudad de Ambato. *MASKANA, I+D+ingeniería 2014.*
- Rivera, C., & Meneses, M. (2022). Predicción flujo de tráfico vehicular Ruta 27 en Costa Rica. *Revista Tecnología En Marcha.* <https://doi.org/10.18845/tm.v35i4.5892>
- Rostami-Shahrbabaki, M., Safavi, A. A., Papageorgiou, M., Setoodeh, P., & Papamichail, I. (2020). State estimation in urban traffic networks: A two-layer approach. *Transportation Research Part C: Emerging Technologies, 115.* <https://doi.org/10.1016/j.trc.2020.102616>
- Ruiz, L., & Tineo, A. (2019). Una revisión de modelos de tráfico automotor usando autómatas celulares. *Novasinería Revista Digital de Ciencia, Ingeniería y Tecnología, 2(2), 7–16.* <https://doi.org/10.37135/unach.ns.001.04.01>
- Salcedo, O., Augusto Hernández Suárez, C., & Díaz, A. E. (2007). Diseño de un modelo de tráfico a través de series de tiempo para pronosticar tráfico Wimax Design of a traffic model through time series to forecast Wimax traffic. *Ingeniería, 12(1), 4–13.*
- Sattarzadeh, A. R., Kutadinata, R. J., Pathirana, P. N., & Huynh, V. T. (2023). A novel hybrid deep learning model with ARIMA Conv-LSTM networks and shuffle attention layer for short-term traffic flow prediction. *Transportmetrica A: Transport Science.* <https://doi.org/10.1080/23249935.2023.2236724>
- Seoane, T., Martín, J., Sánchez, M., Segovia, L., & Moreno, A. (2007). Introduction course to clinical research. Chapter 7: Statistics: Descriptive and inferential statistics. *Semergen, 33(9), 466–471.* [https://doi.org/10.1016/S1138-3593\(07\)73945-X](https://doi.org/10.1016/S1138-3593(07)73945-X)

- The R Foundation. (2010, August 6). *¿Qué es R?*
- Torres, P. (2012). Simulación del tráfico en una vía expresa y análisis estadístico de los resultados. *Ingeniería Industrial*, 30(1025–9929), 45–79.
- Van Der Bijl, B., Gijsbertsen, B., Van Loon, S., Reurich, Y., De Valk, T., Koch, T., & Dugundji, E. (2022). A Comparison of Approaches for the Time Series Forecasting of Motorway Traffic Flow Rate at Hourly and Daily Aggregation Levels. *Procedia Computer Science*, 201(C), 213–222. <https://doi.org/10.1016/j.procs.2022.03.030>
- Yang, S., & Yang, X. (2014). The Application of the Queuing Theory in the Traffic Flow of Intersection. *World Academy of Science, Engineering and Technology International Journal of Mathematical and Computational Sciences*, 8 No. 6, 986–989.
- Zhang, T., De Grande, R. E., & Boukerche, A. (2016). Design and analysis of stochastic traffic flow models for vehicular clouds. *Ad Hoc Networks*, 52, 39–49. <https://doi.org/10.1016/j.adhoc.2016.07.009>
- Zohra. (2024). *Predicción de series temporales: LSTM, BI-LSTM, GRU*.

Apéndice

Apéndice A. Código de los modelos ARIMA y SARIMA

```

library(readxl)
library(ggplot2)
library(tseries)
library(forecast)
library(lubridate)

Datos <- read_excel("C:/Users/José Sangacha/OneDrive/MAESTRIA MATEMATICA/TITULACION/TESIS/DATOS/Datos.xlsx")
serie<- ts(Datos$VALOR, start = c(2000,1), frequency = 12)

#Graficar la serie
library(ggplot2)
plot(serie, main = "serie de Tráfico 2000-2022", ylab = "Tráfico", xlab = "Año")
options(scipen = 999)

# Crear la serie temporal
serie <- ts(Datos$VALOR, start = c(2000, 1), frequency = 12)

# Ajustar márgenes y fuente similar a Times
par(mar = c(4.5, 5, 3, 2), mfp = c(3, 0.5, 0))

# Graficar
plot(serie,
     main = "Serie de tiempo Número de vehículos 2000-2022",
     ylab = "Tráfico (Número de vehículos)",
     xlab = "Tiempo (Años)",
     col = "black",
     lwd = 1,
     cex.main = 1.5,
     cex.lab = 1.2,
     cex.axis = 1,
     las = 1 # Hace el texto del eje Y horizontal
)

##Análisis de la serie
summary(serie)
sd(serie)

#Evaluar si la serie es estacionaria
#Una serie estacionaria tiene media y varianza constantes
library(tseries)
#Hipótesis del test:
#H0 (hipótesis nula): La serie NO es estacionaria (tiene raíz unitaria).
#H1 (hipótesis alternativa): La serie SÍ es estacionaria.
adf.test(serie, alternative = "stationary") #prueba de dickey fuller aumentada adf test)
#p-value > 0.05: La serie no es estacionaria, se rechaza H0, hay que diferenciar.
#p-value < 0.05: La serie sí es estacionaria no se rechaza H0, aplicar ARIMA directamente.
#sí es estacionaria

library(forecast)
library(tseries)
# Gráficas de autocorrelación
par(mfrow=c(1,2)) # Poner gráficos en dos columnas
acf(serie)
pacf(serie)
dev.off()
#ACF → Te ayuda a definir q (el número de lags para el modelo de media móvil).
#PACF → Te ayuda a definir p (el número de lags para el modelo autorregresivo).
#los datos están dentro del límite, hay una relación significativa entre los valores de la serie

```

```

#Ajustar modelos
#####
#ARIMA
#Modelo ARIMA (p, d, q):
modeloarima <- Arima(serie, order = c(1, 0, 0))
summary(modeloarima)

#En el grafico: Un gráfico ACF de residuos: si las barras están dentro de las bandas azules,
#todo bien.
#p-value > 0.05, los residuos no tienen correlación y el modelo está bien ajustado.
#p-value < 0.05, los residuos todavía tienen un patrón, ajustar el modelo.
checkresiduals(modeloarima)

Box.test(residuals(modeloarima), type = "Ljung-Box")
#p-value > 0.05 → El modelo está bien ajustado.
# los residuos de ambos modelos no tienen correlación significativa.
#Es decir, los residuos son aleatorios o ruido blanco.
#el modelo está bien ajustados.

#Cálculo del RMSE
residuos <- residuals(modeloarima)
rmse <- sqrt(mean(residuos^2, na.rm = TRUE));rmse

#Modelo con auto.arima()

modeloarima2 <- auto.arima(serie, seasonal = FALSE) # Busca estacionalidad automáticamente
summary(modeloarima2)
checkresiduals(modeloarima2) #p-value > 0.05, los residuos no tienen correlación y el modelo está bien ajustado.

Box.test(residuals(modeloarima2), type = "Ljung-Box") #p-value > 0.05 → El modelo está bien ajustado.

#Cálculo del RMSE
residuos2 <- residuals(modeloarima2)
rmse2 <- sqrt(mean(residuos2^2, na.rm = TRUE));rmse2

#Predicción dentro de la muestra
fittedvals <- fitted(modeloarima2)
forecastvals <- forecast(modeloarima2, h = 12)
#mejor modelo con auto.arima

# Extraer los parámetros del modelo ARIMA
p <- modeloarima2$arima[1]
d <- modeloarima2$arima[6]
q <- modeloarima2$arima[2]

# Crear fechas para los datos y pronósticos
h<-12
fechas <- seq(as.Date("2000-01-01"), by = "month", length.out = length(serie))
fechas_2023 <- seq(as.Date("2023-01-01"), by = "month", length.out = h)

# Gráfico final similar al solicitado
plot(fechas, serie, type = "l", col = "blue", lwd = 1,
      ylab = "Valor", xlab = "Fecha",
      main = paste("Datos Reales y Predicciones ARIMA (", p, ", ", d, ", ", q, ")", sep = ""))

# Línea de predicción dentro de la muestra (fitted values)
lines(fechas, fittedvals, col = "red", lwd = 1)

# Línea de pronóstico (2023)
lines(fechas_2023, forecastvals$mean, col = "darkgreen", lwd = 1)

# Agregar leyenda
legend("top",
      legend = c("Real (2000-2022)", "Predicción (2000-2022)", "Pronóstico 2023"),
      col = c("blue", "red", "darkgreen"), lty = 1, lwd = 2)

```

```
#####
#MODELO SARIMA
# Ajustar el modelo SARIMA
modelo_sarima <- Arima(serie, order = c(1, 0, 2), seasonal = c(1, 0, 1))
summary(modelo_sarima)

#modelo_sarima2 <- auto.arima(serie, seasonal = TRUE) # Busca estacionalidad automáticamente
#summary(modelo_sarima2)

# Verificar residuos del modelo SARIMA
checkresiduals(modelo_sarima)
#p-value es mayor que 0.05
#Entonces, los residuos son aleatorios, lo que significa:
#El modelo SARIMA está bien ajustado

Box.test(residuals(modelo_sarima), type = "Ljung-Box")
#el p-value es mayor que 0.05, significa que los residuos son aleatorios y
#el modelo está bien ajustado.

#modelo 2
library(forecast)
modelo_auto <- auto.arima(serie, seasonal = TRUE) # Busca estacionalidad automáticamente
summary(modelo_auto)

# Verificar residuos del modelo SARIMA
checkresiduals(modelo_auto)
#Tu p-value = 0.4124 → eso es mayor que 0.05
#Entonces, los residuos son aleatorios, lo que significa:
#El modelo SARIMA automatico está bien ajustado

Box.test(residuals(modelo_auto), type = "Ljung-Box")
#el p-value es mayor que 0.05, significa que los residuos son aleatorios y
#el modelo está bien ajustado.

#Cálculo del RMSE
residuos3 <- residuals(modelo_auto)
rmse3 <- sqrt(mean(residuos3^2, na.rm = TRUE));rmse3

#Predicción dentro de la muestra
fitted_vals <- fitted(modelo_sarima)

# Pronóstico
forecast_vals <- forecast(modelo_sarima, h = 12)
forecast_vals
round(forecast_vals$mean)
# Crear fechas para los datos y pronósticos
fechas2 <- seq(as.Date("2000-01-01"), by = "month", length.out = length(serie))
fechas2_2023 <- seq(as.Date("2023-01-01"), by = "month", length.out = h)
fechas2_2023

# Gráfico final con los parámetros del modelo en el título
plot(fechas2, serie, type = "l", col = "blue", lwd = 1,
     ylab = "valor", xlab = "Fecha",
     main = paste("Datos Reales y Predicciones SARIMA (",
                  0, ",", 1, ",", 2, ")(",
                  0, ",", 1, ",", 0, ") [12]", sep = ""))

# Línea de predicción dentro de la muestra (fitted values)
lines(fechas2, fitted_vals, col = "red", lwd = 1)

# Línea de pronóstico (2023)
lines(fechas2_2023, forecast_vals$mean, col = "darkgreen", lwd = 1)

# Agregar leyenda
legend("top",
      legend = c("Real (2000-2022)", "Predicción dentro de la muestra", "Pronóstico 2023"),
      col = c("blue", "red", "darkgreen"), lty = 1, lwd = 2)

# Gráfico final con los parámetros del modelo en el título
plot(fechas2_2023, forecast_vals$mean, type = "n", col = "blue", lwd = 2,
     ylab = "Número de vehículos ", xlab = "Año 2023", xaxt = "n",
     main = paste("Predicciones Modelo SARIMA (",
                  0, ",", 1, ",", 2, ")(",
                  0, ",", 1, ",", 0, ") [12]", sep = ""))

points(fechas2_2023, forecast_vals$mean, pch = 19, col = "blue")
lines(fechas2_2023, forecast_vals$mean, col = "black", lwd = 1, lty = 2)

# Agregar etiquetas mensuales al eje x
axis(1, at = fechas2_2023, labels = format(fechas2_2023, "%b"), las = 2)

# Etiquetas sobre los puntos
text(fechas2_2023, forecast_vals$mean + 1000,
     labels = round(forecast_vals$mean),
     cex = 0.8)
```

Apéndice B. Código del modelo de redes neuronales multicapa

```
# RED NEURONAL MULTICAPA CON ENTRENAMIENTO Y PRUEBA

library(readxl)
library(nnfor)
library(Metrics)

# Cargar los datos mediante readxl desde la ubicación del archivo
Datos <- read_excel("C:/Users/José Sangacha/OneDrive/MAESTRÍA MATEMÁTICA/TITULACION/TESIS/DATOS/Datos.xlsx")
serie <- ts(Datos$VALOR, start = c(2000, 1), frequency = 12)

# Serie temporal
y <- serie
y
# Dividir datos: entrenamiento y prueba
y_train <- window(y, start = c(2000, 1), end = c(2012, 12)) # 2000-2012
y_test <- window(y, start = c(2013, 1), end = c(2022, 12)) # 2013-2022

# Modelo MLP
fit <- mlp(y_train, reps = 200, hd = c(12, 6), sel.lag = FALSE, lags=1:12, allow.det.season = FALSE )

plot(fit)

# Predicción total: 2013-2022 (10 años) + 2023 (1 año) = 132 meses
p_total <- forecast(fit, h = 132)

# Separar predicción en dos partes
serie_pred_test <- ts(p_total$mean[1:120], start = c(2013, 1), frequency = 12) # 2013-2022
serie_pronostico <- ts(p_total$mean[121:132], start = c(2023, 1), frequency = 12) # 2023

# Calcular RMSE solo sobre el periodo de prueba real
rmse_test <- rmse(y_test, serie_pred_test)
cat("RMSE para el periodo 2013-2022:", rmse_test, "\n")

# Graficar resultados
rango_y <- range(c(y, serie_pred_test, serie_pronostico), na.rm = TRUE)

plot(y, type = "l", col = "blue", lwd = 1.5,
      main = "Tráfico: Datos Reales, Predicción y Pronóstico",
      xlab = "Año", ylab = "Tráfico (número de vehículos)",
      ylim = rango_y)

mtext("Modelo MLP(12, [12,6], 1), reps = 200", side = 3, line = 0.5, cex = 0.9)

lines(serie_pred_test, col = "red", lwd = 1.5) # Predicción 2013-2022
lines(serie_pronostico, col = "darkgreen", lwd = 1.5) # Pronóstico 2023

legend("top",
       legend = c("Datos Reales (2000-2022)", "Predicción (2013-2022)", "Pronóstico 2023"),
       col = c("blue", "red", "darkgreen"), lty = 1, lwd = 2)

rm(list = ls())
```

Apéndice C. Código del modelo de redes neuronales recurrentes

```

#REDES NEURONALES RECURRENTE
# install.packages("keras")
# install.packages("tensorflow")
# install.packages("readxl")

# Cargar librerías
library(keras)
library(tensorflow)
library(readxl)

# Leer archivo Excel
Datos <- read_excel("C:/Users/José Sangacha/OneDrive/MAESTRÍA MATEMÁTICA/TITULACION/TESIS/DATOS/Datos.xlsx")
serie <- ts(Datos$VALOR, start = c(2000,1), frequency = 12)

# Normalizar La serie temporal
min_val <- min(serie)
max_val <- max(serie)
serie_norm <- (serie - min_val) / (max_val - min_val)

# Convertir a vector
serie_norm <- as.numeric(serie_norm)

serie_norm
# Parámetros
lag <- 12
n <- length(serie_norm)

# Crear estructura de datos para LSTM
X <- array(0, dim = c(n - lag, lag, 1))
Y <- array(0, dim = c(n - lag))

for (i in 1:(n - lag)) {
  X[i,,1] <- serie_norm[i:(i + lag - 1)]
  Y[i] <- serie_norm[i + lag]
}

# División: 13 años (156 datos) - 144 muestras de entrenamiento
train_limit <- 156

train_samples <- train_limit - lag
X_train <- X[1:train_samples,,]
Y_train <- Y[1:train_samples]

X_test <- X[(train_samples + 1):(n - lag),,, drop = FALSE]
Y_test <- Y[(train_samples + 1):(n - lag)]

# Construcción del modelo LSTM
model <- keras_model_sequential() %>%
  layer_lstm(units = 50, input_shape = c(lag, 1), return_sequences = TRUE) %>%
  layer_lstm(units = 25, return_sequences = FALSE) %>%
  layer_dense(units = 1)

model %>% compile(loss = 'mean_squared_error', optimizer = 'adam')

# Entrenamiento
model %>% fit(
  X_train, Y_train,
  epochs = 200,
  batch_size = 16,
  validation_split = 0.2,
  verbose = 1#,
  #callbacks = list(callback_early_stopping)
)

print(model)

# Predicción sobre conjunto de prueba
pred_test <- model %>% predict(X_test)
pred_test_rescaled <- pred_test * (max_val - min_val) + min_val
Y_test_rescaled <- Y_test * (max_val - min_val) + min_val

```

```

# Calcular MSE y RMSE
Y_test_rescaled <- as.numeric(Y_test_rescaled)
pred_test_rescaled <- as.numeric(pred_test_rescaled)
mse <- mean((Y_test_rescaled - pred_test_rescaled)^2)
rmse <- sqrt(mse)
cat("MSE:", round(mse, 4), "\n")
cat("RMSE:", round(rmse, 4), "\n")

# Crear fechas reales desde enero de 2000
fechas <- seq(as.Date("2000-01-01"), by = "month", length.out = length(serie))
fechas_test <- fechas[(train_limit + 1):length(serie)]

# PREDICCIÓN AÑO 2023
input_seq <- tail(serie_norm, lag)
predictions_2023 <- numeric(12)

for (i in 1:12) { input_array <- array(input_seq[(length(input_seq) - lag + 1):length(input_seq)],
  dim = c(1, lag, 1))
  pred_norm <- model %>% predict(input_array)
  predictions_2023[i] <- pred_norm
  input_seq <- c(input_seq, pred_norm)
}

# Desnormalizar
pred_2023 <- predictions_2023 * (max_val - min_val) + min_val

fechas_2023 <- seq(as.Date("2023-01-01"), by = "month", length.out = 12)

print(pred_2023)
# === GRÁFICA FINAL (DATOS REALES + PREDICCIONES) ===
fechas_completas <- seq(as.Date("2000-01-01"), by = "month", length.out = length(serie) + 12)

# Expandir serie real con valores predichos
# Crear fechas completas (2000-2023)
fechas_completas <- seq(as.Date("2000-01-01"), by = "month", length.out = length(serie) + 12)

# Expandir serie real con valores NA para las predicciones futuras
serie_completa <- c(serie, rep(NA, 12))

# Gráfico
plot(fechas_completas, serie_completa, type = "l", col = "blue", lwd = 1,
  ylab = "valor", xlab = "Fecha", main = "Datos Reales y Predicciones RNN LSTM (2000-2023)")

# Línea de predicciones del modelo 2013-2022
lines(fechas_test, pred_test_rescaled, col = "red", lwd = 1)

# Línea de predicciones 2023
lines(fechas_2023, pred_2023, col = "darkgreen", lwd = 1)

# Marcadores para los valores de predicción 2023
lines(fechas_2023, pred_2023, col = "darkgreen", pch = 1)

# Agregar leyenda
legend("top", legend = c("Real (2000-2022)", "Predicción (2013-2022)", "Predicción 2023"),
  col = c("blue", "red", "darkgreen"), lty = 1, lwd = 2)

```

Apéndice D. Código del modelo de teoría de colas.

```

# Teoría de colas
if (!require(queueing)) install.packages("queueing", dependencies = TRUE)
library(queueing)

# Una sola carretera
lambda <- 45 # Tasa de llegada de vehículos por minuto
mu <- 50     # Tasa de servicio (vehículos atendidos por minuto)

```

```

c <- 2 # N?mero de carriles disponibles
lambda <- 45 # Tasa de llegada de veh?culos por minuto
mu <- 50
# crear el modelo M/M/c

rho <- lambda / (c * mu)
fact <- function(n) ifelse(n == 0, 1, n * fact(n - 1))

P0 <- 1 / (sum(sapply(0:(c-1), function(k) (lambda/mu)^k / fact(k))) + ((lambda/mu)^c / (fact(c) * (1 - rho))))
Pw <- ((lambda/mu)^c / (fact(c) * (1 - rho))) * P0
Lq <- (Pw * rho) / (1 - rho) * (lambda/mu)
L <- Lq + (lambda/mu)
W <- L / lambda
wq <- Lq / lambda

cat("\n??? Resultados del Modelo M/M/", c, " (" , c, " carriles):\n")
cat("??? Tiempo promedio en el sistema (w):", round(w, 3), "min\n")
cat("????? Tiempo promedio de espera en la cola (wq):", round(wq, 3), "min\n")
cat("???? N?mero promedio de veh?culos en la carretera (L):", round(L, 3), "\n")
cat("???? N?mero promedio de veh?culos esperando (Lq):", round(Lq, 3), "\n")
cat("??? Probabilidad de congesti?n (??):", round(rho, 3), "\n")
cat("???? Probabilidad de que un veh?culo tenga que esperar (Pw):", round(Pw, 3), "\n")

# Modelo de Poisson

# Instalar y cargar paquetes necesarios
if (!require(ggplot2)) install.packages("ggplot2", dependencies = TRUE)
library(ggplot2)

# Par?metro lambda: tasa de ocurrencia de eventos (por ejemplo, autos por minuto)
lambda <- 45 # 5 autos llegan por minuto en promedio

# Simulaci?n de un proceso de Poisson
set.seed(123) # Para reproducibilidad
tiempos_llegada <- cumsum(rexp(100, rate = lambda)) # 100 eventos

# Crear un DataFrame con los tiempos de llegada
data_poisson <- data.frame(evento = 1:100, tiempo = tiempos_llegada)

# Graficar las llegadas de eventos en el tiempo
ggplot(data_poisson, aes(x = tiempo, y = evento)) +
  geom_step(color = "blue", size = 1) +
  labs(title = "Proceso de Poisson: Llegadas de Autos",
       x = "Tiempo (minutos)",
       y = "N?mero de Autos Llegados") +
  theme_minimal()

```