

**MODELO ESPACIO-TEMPORAL PARA LA PREDICCIÓN DE DELITOS
BASADO EN APRENDIZAJE AUTOMÁTICO**

JUAN PABLO PEREZ LEAL

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2024

**MODELO ESPACIO-TEMPORAL PARA LA PREDICCIÓN DE DELITOS
BASADO EN APRENDIZAJE AUTOMÁTICO**

JUAN PABLO PEREZ LEAL

**Una tesis presentada en cumplimiento de los requisitos para el grado de:
Ingeniero de Sistemas e Informática**

Director:

Andrés Sebastián Ríos Gutierrez, Ph.D (C)

Codirector:

David Romo Bucheli, Ph.D

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2024

AGRADECIMIENTOS

Quisiera expresar mi mas sinceros agradecimientos a mi madre y a Natalia, por su constante apoyo emocional durante todos los momentos mas desafiantes. A mi director Andrés, el cual me ha guiado en los momentos mas complicados del desarrollo de este trabajo. A mi codirector David, quien con sus valiosos aportes ha mejorado mi visión crítica del proyecto y a todas las personas que han hecho parte del proceso para la realización de este trabajo de grado.

ÍNDICE GENERAL

	Pág
INTRODUCCIÓN	12
1. FUNDAMENTOS Y TRABAJOS PREVIOS	14
1.1. CONCEPTOS PRELIMINARES	14
1.1.1. Un robo	14
1.1.2. Predicción de crímenes	14
1.1.3. Predicción en series de tiempo	16
1.1.4. Proceso estocástico	16
1.1.5. Ruido blanco	17
1.2. MODELOS DE SERIES DE TIEMPO	18
1.2.1. ARIMA	18
1.2.2. Redes Neuronales	20
1.3. MODELOS CON REDES NEURONALES	22
1.3.1. Redes Neuronales Recurrentes	23
1.3.2. Redes con Gran Memoria a largo corto plazo	24
1.4. ALGORITMOS ESPACIO-TEMPORALES	27
1.4.1. STNN-R	29
1.4.2. STNN-D	29
1.5. SOLUCIONES PLANTEADAS PARA PREDICCIÓN DE CRÍMENES	30
2. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA	32
3. OBJETIVOS	34
4. DESCRIPCIÓN DEL CASO DE ESTUDIO	35

4.1. ZONA DE ESTUDIO	35
4.1.1. Ubicación geográfica	35
4.1.2. Contexto Geoespacial	35
4.1.3. Datos Demográficos	36
4.1.4. División Administrativa	36
4.2. DESCRIPCIÓN DE FUENTES DE DATOS	36
4.2.1. Datos temporales	36
4.2.2. Datos Geoespaciales	38
4.3. EXPLORACIÓN DE LOS DATOS	38
4.3.1. Frecuencia de los robos por año y mes	38
4.4. DISTRIBUCIÓN GEOGRÁFICA DE LOS ROBOS POR COMUNA	39
5. METODOLOGÍA	43
5.1. PROCESAMIENTO DE DATOS	43
5.1.1. Limpieza de datos	43
5.1.2. Construcción del conjunto de datos	43
5.1.3. Representación espacial	44
5.1.4. Ventaneo temporal	44
5.1.5. División de datos en conjuntos de entrenamiento y prueba	45
6. EXPERIMENTOS Y RESULTADOS	48
6.1. CONFIGURACIÓN EXPERIMENTAL	48
6.1.1. Métricas de evaluación	48
6.1.2. Configuración de los vectores de características	49
6.2. EXPERIMENTOS	50
6.3. RESULTADOS	51
6.3.1. Optimización de Hiperparámetros	51
6.3.2. Resultados de Experimentos	52

6.4. VISUALIZACIÓN DE RESULTADOS	56
7. CONCLUSIONES Y PERSPECTIVAS	67
BIBLIOGRAFÍA	69

ÍNDICE DE FIGURAS

	Pág
Figura 1. Distribución geográfica de las comunas de Bucaramanga. Creación propia . . .	37
Figura 2. Número total de robos mensuales para el período de estudio. Creación propia .	39
Figura 3. Distribución geográfica de los robos. Creación propia	40
Figura 4. Número total de robos mensuales por comuna durante el periodo de estudio. Creación propia	42
Figura 5. Diagrama de caja de la distribución de robos por comuna al agrupar semanal- mente. Creación propia	44
Figura 6. Representación gráfica mediante un grafo de las relaciones geográficas entre las comunas. Creación propia	45
Figura 7. Representación de la división del conjunto de datos para validación cruzada. Creación propia	47
Figura 8. Gráfico de caja con los puntajes RMSE obtenidos para cada modelo. Creación propia	55
Figura 9. Gráfico de caja con los puntajes SMAPE obtenidos para cada modelo. Creación propia	57
Figura 10. Visualización de las relaciones encontradas por el modelo STNN-R. Creación propia	59
Figura 11. Visualización de las predicciones realizadas por el modelo ARIMA. Creación propia	61
Figura 12. Visualización de las predicciones realizadas por el modelo STNN. Creación propia	62
Figura 13. Visualización de las predicciones realizadas por el modelo STNN-D. Creación propia	63

Figura 14. Visualización de las predicciones realizadas por el modelo STNN-R. Creación propia 64

Figura 15. Visualización de las predicciones realizadas por el modelo RNN. Creación propia 65

Figura 16. Visualización de las predicciones realizadas por el modelo LSTM. Creación propia 66

ÍNDICE DE CUADROS

	Pág
Tabla 1. Parámetros con mejor ajuste de modelo ARIMA para cada comuna	52
Tabla 2. Parámetros de entrenamiento usados para LSTM y RNN	53
Tabla 3. Resultados usando la métrica de RMSE para cada modelo	54
Tabla 4. Resultados usando la métrica de SMAPE para cada modelo	56

RESUMEN

TÍTULO: MODELO ESPACIO-TEMPORAL PARA LA PREDICCIÓN DE DELITOS BASADO EN APRENDIZAJE AUTOMÁTICO. *

AUTOR: JUAN PABLO PEREZ LEAL **

PALABRAS CLAVE: Predicción de crímenes, Aprendizaje de máquina, Variable espacial, Variable temporal, Redes Neuronales.

DESCRIPCIÓN: Este trabajo se centra en predecir patrones delictivos de robos a peatones en Bucaramanga, Colombia, usando datos de la Alcaldía disponibles en la plataforma de Datos Abiertos ¹. Para el análisis, se desarrollaron modelos ARIMA y redes neuronales recurrentes (RNN) y de gran memoria a corto plazo (LSTM)², junto con modelos espacio-temporales (STNN) que incorporan información geográfica de las comunas. Los resultados muestran que los modelos RNN y LSTM superan a ARIMA en la predicción de robos, mientras que los modelos STNN capturan tendencias generales, aunque con limitaciones para identificar cambios abruptos en las tendencias.

La precisión de los modelos se evaluó con el Error Cuadrático Medio Raíz (RMSE), y se utilizaron técnicas de ventaneo temporal y normalización para mejorar su rendimiento. Estos hallazgos pueden ayudar a optimizar recursos y estrategias de seguridad pública en Bucaramanga, además de ofrecer un marco adaptable a otras ciudades. La investigación aporta bases para futuros estudios sobre predicción delictiva y modelos con variables espacio-temporales.

* Trabajo de investigación

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Andrés Sebastián Ríos Gutierrez

¹ Consulta directa en <https://www.datos.gov.co/Seguridad-y-Defensa/Delitos-ocurridos-en-el-Municipio-de-Bucaramanga/75fz-q98y>

² LSTM significa "Long Short-Term Memory." en inglés

ABSTRACT

TITLE: SPACETIME CRIME PREDICTION MODEL BASED ON MACHINE LEARNING. *

AUTHOR: JUAN PABLO PEREZ LEAL **

KEYWORDS: Crime prediction, Machine learning, Spatial and temporal variables, Neural networks

DESCRIPTION: This work focuses on predicting pedestrian robbery patterns in Bucaramanga, Colombia, using data from the City Hall's Open Data platform¹. The analysis developed ARIMA models and neural network approaches, including Recurrent Neural Networks (RNN) and Long Short-Term Memory networks (LSTM), alongside spatio-temporal models (STNN) that incorporate geographic information from the city's communes. Results show that RNN and LSTM models outperform ARIMA in predicting robberies, while STNN models capture general trends but face limitations in identifying abrupt trend changes.

Model accuracy was evaluated using the Root Mean Square Error (RMSE), with temporal windowing and data normalization techniques applied to improve performance. These findings may help optimize public safety resources and strategies in Bucaramanga, while also providing a framework adaptable to other cities. The research lays the groundwork for future studies on crime prediction and models incorporating spatio-temporal variables.

* Research work

** Faculty of Physical-Mechanical Engineering. School of Systems and Computer Engineering. Advisor: Andrés Sebastián Ríos Gutierrez

¹ Direct access at <https://www.datos.gov.co/Seguridad-y-Defensa/Delitos-ocurridos-en-el-Municipio-de-Bucaramanga-75fz-q98y>

INTRODUCCIÓN

La predicción de patrones delictivos en entornos urbanos se ha convertido en una herramienta crucial para la planificación de la seguridad pública y la prevención de delitos². En particular, la ciudad de Bucaramanga(Colombia), ha experimentado un aumento significativo en los robos a peatones en los últimos años, lo que ha generado una mayor percepción de inseguridad entre sus habitantes³. Esta situación ha llevado a la necesidad de desarrollar modelos predictivos que puedan predecir estadísticamente la ocurrencia de estos delitos y, en consecuencia, permitir una respuesta más eficaz por parte de las autoridades.

El presente trabajo de grado se centra en la evaluación de diferentes enfoques predictivos para la identificación de los robos a mano armada que ocurren en las calles de Bucaramanga cuyas victimas son los peatones. Entre los modelos considerados se incluyen enfoques tradicionales como el modelo ARIMA, las redes neuronales recurrentes (RNN) y las redes de gran memoria a corto plazo (LSTM), también se implementarán modelos recientes espacio-temporales, como las redes neuronales espacio-temporales (STNN).

Para mejorar la eficacia de estos modelos, se aplicarán técnicas de preprocesamiento de datos como la normalización, eliminación de valores atípicos, y ventaneo temporal, que segmenta las series de tiempo en intervalos manejables para facilitar la captura de dinámicas complejas, también se hace una normalización de los datos, que asegura que los modelos trabajen con un conjunto de datos bajo la misma escala, de manera que la variabilidad de los datos no cambie

² Anu Sayal et al. “Crime Detection Using Data Mining Techniques”. En: *2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT)*. 2024, págs. 200-204. DOI: 10.1109/CCICT62777.2024.00041.

³ *Bucaramanga como vamos*. 25 de jun. de 2024. URL: <https://www.bucaramangacomovamos.org/post/informe-de-calidad-de-vida-2022> (visitado 27-09-2022).

significativamente las predicciones.⁴

Al desarrollar y comparar distintos modelos predictivos para la identificación de patrones delictivos en Bucaramanga, este trabajo proporcionará herramientas avanzadas que mejorarán la capacidad de las autoridades para anticipar y responder a los robos de peatones. Esta investigación permite ampliar el conocimiento existente sobre la dinámica delictiva en entornos urbanos, ofreciendo un marco adaptable a otras ciudades. Estos aportes no solo fortalecerán la prevención y el control del delito, sino que también servirán como base para futuras investigaciones en el campo.

⁴ Thomas G. Dietterich. “Machine Learning for Sequential Data: A Review”. En: *Structural, Syntactic, and Statistical Pattern Recognition*. Ed. por Terry Caelli et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, págs. 15-30.

1. FUNDAMENTOS Y TRABAJOS PREVIOS

Este proyecto se centra en el desarrollo y evaluación de un modelo espacio-temporal para predecir robos en Bucaramanga (Colombia). Para entender la base teórica y los métodos utilizados en este estudio, es fundamental revisar los conceptos clave y los trabajos previos relevantes en el campo de la predicción de delitos. En esta sección, se presentan las definiciones básicas para comprender el caso de estudio junto con una recopilación de investigaciones anteriores que han utilizado diferentes enfoques y técnicas para abordar la predicción de robos en contextos urbanos.

1.1. CONCEPTOS PRELIMINARES

1.1.1. Un robo es un delito que implica la apropiación indebida de bienes o valores ajenos mediante el uso de la fuerza, la violencia o la amenaza. Este acto delictivo se caracteriza por la intención del perpetrador de obtener un beneficio económico a costa del perjuicio de la víctima, quien puede sufrir tanto pérdidas materiales como de traumas psicológicos. A diferencia del hurto, el cual no involucra violencia, un robo se distingue por la confrontación directa y la coacción ejercida sobre la víctima.⁵

1.1.2. Predicción de crímenes La predicción de crímenes es el proceso de utilizar datos históricos y técnicas analíticas para anticipar la ocurrencia de delitos en el futuro. Esta disciplina combina los métodos de estadística, el aprendizaje automático y el análisis espacial para identificar patrones y tendencias en los datos delictivos. El objetivo es proporcionar a las organizaciones públicas como las alcaldías o la policía una información valiosa que les permita tomar decisiones con el fin de prevenir y mitigar los delitos.

⁵ Luis Eduardo Mesa Velásquez. “Los delitos de hurto y robo, con referencia al Código Penal Colombiano”. En: *Estudios de Derecho* 18.56 (1959), 129–158. DOI: 10.17533/udea.esde.333074.

Uno de los enfoques más comunes en la predicción de crímenes es el uso de modelos de series temporales⁶, que analizan cómo varía la incidencia delictiva con respecto al tiempo. Estos modelos pueden identificar tendencias y ciclos en los datos, ayudando a predecir periodos de alta actividad delictiva. Además, la incorporación de análisis espaciales permite entender cómo es la ubicación geográfica que influye en los crímenes.⁷

La predicción de crímenes también puede beneficiarse de técnicas de aprendizaje de máquina como las redes neuronales las cuales son capaces de capturar relaciones complejas entre múltiples variables.⁸

En el caso de este trabajo, el problema será analizado como un problema de regresión. La regresión es una técnica estadística utilizada para modelar y analizar la relación entre una variable dependiente cuantitativa y una o más variables independientes cuantitativas.⁹ En su forma más simple, la regresión lineal, se busca ajustar una línea recta a un conjunto de datos para predecir el valor de la variable dependiente en función de las variables independientes. Esta técnica es fundamental en la predicción de valores continuos y se utiliza ampliamente en

-
- ⁶ Sunil Yadav et al. “Crime pattern detection, analysis & prediction”. En: *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*. Vol. 1. 2017, págs. 225-230. DOI: 10.1109/ICECA.2017.8203676.
- ⁷ Anthony A. Braga, Andrew V. Papachristos y David M. Hureau. “The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis”. En: *Justice Quarterly* 31.4 (2014), págs. 633-663. DOI: 10.1080/07418825.2012.673632. eprint: <https://doi.org/10.1080/07418825.2012.673632>.
- ⁸ Eunseo Kwon, Sungwon Jung y Jaewook Lee. “Artificial Neural Network Model Development to Predict Theft Types in Consideration of Environmental Factors”. En: *ISPRS International Journal of Geo-Information* 10.2 (2021). DOI: 10.3390/ijgi10020099.
- ⁹ Donald F Specht et al. “A general regression neural network”. En: *IEEE transactions on neural networks* 2.6 (1991), págs. 568-576.

diversas disciplinas, incluyendo economía, biología y ciencias sociales.^{10,11}

1.1.3. Predicción en series de tiempo La predicción en series de tiempo se empleará como metodología principal para analizar la evolución temporal de los incidentes delictivos y anticipar eventos futuros en Bucaramanga. Esta técnica es crucial para modelar cómo los incidentes se desarrollan a lo largo del tiempo, proporcionando una herramienta potente para entender y prever las dinámicas delictivas basándose en patrones históricos.¹²

Los modelos de series de tiempo se especializan en reconocer y explotar la dependencia temporal entre observaciones consecutivas, lo cual es crucial en contextos donde los eventos históricos son indicativos de futuras ocurrencias.¹³

1.1.4. Proceso estocástico Un proceso estocástico es una colección de variables aleatorias que representan la evolución de un sistema o fenómeno en el tiempo o en el espacio. A diferencia de los procesos deterministas, en los que el estado futuro del sistema está completamente definido por su estado actual, en un proceso estocástico, el estado futuro depende no solo del estado actual, sino también de una cierta incertidumbre o aleatoriedad inherente al sistema.¹⁴

¹⁰ Brian S. Cade, David R. Edmunds y Douglas S. Ouren. “Quantile regression estimates of animal population trends”. En: *The Journal of Wildlife Management* 86.5 (2022), e22228. DOI: <https://doi.org/10.1002/jwmg.22228>. eprint: <https://wildlife.onlinelibrary.wiley.com/doi/pdf/10.1002/jwmg.22228>.

¹¹ Azamat Kibekbaev y Ekrem Duman. “Benchmarking regression algorithms for income prediction modeling”. En: *Information Systems* 61 (2016), págs. 40-52. DOI: <https://doi.org/10.1016/j.is.2016.05.001>.

¹² Romika Yadav y Savita Kumari Sheoran. “Crime Prediction Using Auto Regression Techniques for Time Series Data”. En: *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*. 2018, págs. 1-5. DOI: 10.1109/ICRAIE.2018.8710407.

¹³ Bruno Cavadas, Paula Branco y Sérgio Pereira. “Crime Prediction Using Regression and Resources Optimization”. En: *Progress in Artificial Intelligence*. Ed. por Francisco Pereira et al. Cham: Springer International Publishing, 2015, págs. 513-524.

¹⁴ Athanasios Papoulis y Unnikrishna Pillai. *Probability, Random Variables, and Stochastic Processes, Fourth Edition*. McGraw-Hill, 2002.

Matemáticamente, un proceso estocástico se define como una familia de variables aleatorias $X_t : t \in T$, donde T es el conjunto de índices que puede representar el tiempo (discreto o continuo) y X_t es el valor que toma la variable en un instante particular t . Cada realización del proceso genera una "trayectoria" que sigue la evolución del sistema a lo largo del tiempo, y debido a la naturaleza aleatoria del proceso, pueden existir múltiples trayectorias posibles.¹⁵

Los procesos estocásticos tienen una amplia gama de aplicaciones, particularmente en campos como la economía, la biología, la ingeniería, y la ciencia de datos.¹⁶ En el contexto de modelos predictivos de patrones delictivos, los procesos estocásticos permiten modelar y analizar fenómenos que presentan incertidumbre o variabilidad, como la ocurrencia de robos en diferentes zonas geográficas a lo largo del tiempo.

1.1.5. Ruido blanco El ruido blanco es una secuencia de variables aleatorias que tienen media cero, varianza constante y son no correlacionadas entre sí. En términos simples, se puede entender como un proceso aleatorio donde no existe ninguna estructura o patrón predecible en los datos. Cada observación es completamente independiente de las demás y no está influenciada por el pasado ni por el futuro.¹⁷

El ruido blanco es un componente fundamental en muchos modelos estadísticos y econométricos, particularmente en la modelación de series de tiempo. En este contexto, se utiliza para representar el componente aleatorio o impredecible de un fenómeno, es decir, las fluctuaciones que no pueden ser explicadas por un modelo.

En algunos casos, cuando las observaciones del ruido blanco además siguen una distribución

¹⁵ Stirzaker David Grimmet Geooffrey. *Probability and Random Processes*. Oxford University Press, 2001.

¹⁶ Paul Embrechts, Rüdiger Frey y Hansjörg Furrer. "Stochastic processes in insurance and finance". En: *Stochastic Processes: Theory and Methods*. Vol. 19. Handbook of Statistics. Elsevier, 2001, págs. 365-412. DOI: [https://doi.org/10.1016/S0169-7161\(01\)19014-0](https://doi.org/10.1016/S0169-7161(01)19014-0).

¹⁷ Peter J Brockwell y Richard A Davis. *Time series: theory and methods*. Springer science & business media, 1991.

normal, se le denomina ruido blanco gaussiano. Este tipo de ruido es particularmente importante en modelos lineales como los modelos ARIMA, donde se asume que los residuos del modelo deberían aproximarse a un proceso de ruido blanco para garantizar la idoneidad del ajuste.¹⁸ En el análisis de patrones delictivos, el ruido blanco puede representar la aleatoriedad inherente en ciertos eventos delictivos que no siguen un patrón claro ni pueden ser anticipados por variables explicativas.

1.2. MODELOS DE SERIES DE TIEMPO

1.2.1. ARIMA Entre las herramientas más comúnmente utilizadas para la previsión de series temporales se encuentra el modelo ARIMA¹⁹. El modelo ARIMA es valioso en situaciones donde los datos exhiben autocorrelación, es decir, hay una correlación entre variables aleatorias pasadas y la variable aleatoria en el tiempo presente. Además, puede manejar series temporales que no son estacionarias, es decir, aquellas que muestran tendencias o variaciones sistemáticas a lo que cambian con respecto al tiempo²⁰. Es importante señalar que el modelo ARIMA es un modelo dinámico de series temporales, lo que significa que las estimaciones futuras se explican por los datos pasados y no por variables independientes.^{17,21} En el modelo ARIMA hay una composición de tres componentes: Autorregresión (AR), Integración (I) y Media Móvil (MA). La Autorregresión (AR) se refiere a la dependencia lineal de una observación actual sobre observaciones pasadas a través del siguiente modelo:

Sea $X_t, t > 0$ un conjunto de variables aleatorias donde X_t representa una variable en un tiempo

¹⁸ George E. P. Box et al. *Time Series Analysis: Forecasting and Control, 5th Edition*. Wiley, 2015.

¹⁹ Autoregressive Integrated Moving Average

²⁰ B Uma Devi, Darshan Sundar y P Alli. "An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50". En: *International Journal of Data Mining & Knowledge Management Process* 3.1 (2013), pág. 65.

²¹ Rob J Hyndman y George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

t el modelo

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t, \quad (1)$$

donde x es la variable dependiente en el tiempo t , c es una constante, $\phi_1, \phi_2, \dots, \phi_p$ son los parámetros de autorregresión, y ϵ_t es un término de error aleatorio en el tiempo t . El cual sigue un proceso de ruido blanco (WN) con varianza σ^2 , para todo $t \geq 0$, se define como el modelo de orden p .

La Media Móvil (MA) modela la relación entre una observación y un término de error residual generado por una media móvil de errores en periodos anteriores. En un modelo MA, la variable dependiente se regresa linealmente sobre los términos de error pasados hasta el orden q . Se define el modelo MA de orden q :

$$x_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t, \epsilon_t \sim WN(0, \sigma^2), \quad (2)$$

donde c es una constante, $\theta_1, \theta_2, \dots, \theta_q$ son los parámetros de media móvil, y $\epsilon_1, \epsilon_2, \dots, \epsilon_{t-q}$ son los términos de error en tiempos anteriores.

La Integración (I) se refiere al proceso de hacer estacionaria una serie temporal, es decir, eliminar tendencias o patrones sistemáticos que puedan afectar la predicción. Esto se logra diferenciando la serie temporal hasta que se vuelva estacionaria. El orden de la diferenciación se denota como d . Esta diferencia se define recurrentemente como:

$$\Delta^d x_t = \Delta^{d-1} x_t - \Delta^{d-1} x_{t-1}, \Delta x_t = \Delta^1 x_t = x_t - x_{t-1} \text{ y } d = 2, 3, \dots \quad (3)$$

El modelo ARIMA combina estos tres componentes en un solo marco y se denota con las siglas ARIMA(p, d, q), donde p es el orden de la autorregresión, d es el orden de la diferenciación para obtener un proceso estacionario ²², y q es el orden de la media móvil.

²² Sidney I Resnick. *Adventures in stochastic processes*. Springer Science & Business Media, 2013.

El *operador de retardo* se define, para todo $u \in \mathbb{Z}^+$, como $L^u x_t := x_{t-u}$. Basado en este operador, el modelo ARIMA(p, d, q) se define como:

Sea $X_t, t \geq 0$ m conjunto de variables aleatorias:

$$\Phi_p(L)(1 - L)^d x_t = c + \Theta_q(L)\epsilon_t \text{ con } \epsilon_t \sim WN(0, \sigma^2), \quad (4)$$

donde $\Phi_p(L), \Theta_q(L)$ son polinomios, respectivamente llamados el polinomio autorregresivo y el polinomio de media móvil. Se define como el modelo ARIMA de orden p, d, q , en consecuencia, d representa el número de las diferencias para que el proceso estocástico $\Phi_p(L)x_t = c + \Theta_q(L)\epsilon_t$ sea estacionario.

1.2.2. Redes Neuronales Las redes neuronales son una clase de modelos de aprendizaje automático inspirados en el funcionamiento del cerebro humano. Estas redes consisten en capas de nodos (o neuronas) conectados entre sí, donde cada conexión tiene un peso que se ajusta durante el proceso de entrenamiento para minimizar el error en las predicciones.²³ Las redes neuronales son especialmente efectivas para capturar patrones complejos y no lineales en los datos, lo que las hace ideales para tareas como el reconocimiento de voz, la visión por computadora y las predicciones en series temporales.

Una red neuronal típica consta de una capa de entrada, una o más capas ocultas, y una capa de salida. La capa de entrada recibe las características del conjunto de datos, las capas ocultas procesan esta información a través de activaciones no lineales mediante funciones de activación como ReLU, sigmoide o tanh, y la capa de salida proporciona la predicción final.

Puede definirse matemáticamente una red neuronal como una función compuesta que transforma las entradas en una salida a través de varias capas de transformaciones. Este proceso se conoce como propagación hacia adelante. Si denotamos las entradas por x y la salida por \hat{y} , la red

²³ Zhihua Zhang. “Artificial Neural Network”. En: *Multivariate Time Series Analysis in Climate and Environmental Research*. Cham: Springer International Publishing, 2018, págs. 1-35. DOI: 10.1007/978-3-319-67340-0_1.

neuronal realiza la siguiente operación:

$$\hat{y} = f(W_n \cdot f(W_{n-1} \cdot \dots \cdot f(W_1 \cdot x + b_1) + b_{n-1}) + b_n) \quad (5)$$

Donde: W_i representa la matriz de pesos de la capa i , b_i es el vector de sesgos de la capa i , f es la función de activación y n es el número de capas de la red.

La operación consiste en aplicar una transformación lineal a las entradas de cada capa (multiplicando por los pesos y sumando el sesgo) y luego aplicar una función de activación no lineal, introduciendo así la capacidad de la red para aprender relaciones complejas y no lineales en los datos. Este proceso se repite capa a capa hasta obtener la predicción final \hat{y} . Una vez que se obtiene la predicción \hat{y} , la calidad de dicha predicción se evalúa mediante una función de pérdida. Esta función mide la discrepancia entre las predicciones de la red y los valores reales del conjunto de datos. Una de las funciones de pérdida más comunes es el error cuadrático medio (MSE), que se define como:

$$L(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (6)$$

Donde: y_i son los valores reales, \hat{y}_i son las predicciones de la red, n es el número total de ejemplos en el conjunto de datos. El objetivo del entrenamiento de la red neuronal es minimizar esta función de pérdida, es decir, hacer que las predicciones \hat{y} estén lo más cerca posible de los valores reales y .

Para reducir el error de predicción, la red ajusta los pesos de sus conexiones mediante un proceso llamado retropropagación del error, que utiliza el descenso de gradiente. El descenso de gradiente ajusta los pesos en la dirección opuesta al gradiente de la función de pérdida respecto a los pesos, con el fin de minimizarla.

Matemáticamente, dado un conjunto de datos (x, y) , donde x representa las características de entrada y y los valores esperados, la función de pérdida L se define como la diferencia entre las predicciones de la red neuronal \hat{y} y los valores reales y . Si W denota el conjunto de pesos de la

red, la actualización de los pesos se realiza mediante la siguiente ecuación:

$$W^{(t+1)} = W^{(t)} - \eta \cdot \nabla_W L(W^{(t)}). \quad (7)$$

Donde $W^{(t)}$ son los pesos en la iteración t , η es la tasa de aprendizaje, un hiperparámetro que controla el tamaño de los pasos en la actualización de los pesos y $\nabla_W L(W^{(t)})$ es el gradiente de la función de pérdida respecto a los pesos en la iteración t .

Este proceso de actualización se realiza para todos los pesos de la red, propagando el error desde la capa de salida hacia las capas ocultas mediante la retropropagación. La retropropagación permite calcular el gradiente de la función de pérdida para cada peso de la red, lo que facilita su ajuste.

El proceso continúa hasta que la red alcanza un valor mínimo aceptable de la función de pérdida, lo que indica que las predicciones de la red son suficientemente precisas. Dando como resultado las predicciones de la red neuronal.

1.3. MODELOS CON REDES NEURONALES

Las RNN y sus variantes avanzadas, como LSTM, se han convertido en herramientas fundamentales para la predicción de series de tiempo. Estas redes son capaces de procesar datos secuenciales y mantener información histórica relevante, lo que les permite capturar patrones temporales complejos y realizar predicciones más precisas. Estas características hacen que RNN y LSTM sean especialmente efectivas en aplicaciones como la predicción de series de tiempo, donde a través de la dinámica temporal se pueden obtener resultados precisos.²⁴

²⁴ Eunseo Kwon, Sungwon Jung y Jaewook Lee. “Artificial Neural Network Model Development to Predict Theft Types in Consideration of Environmental Factors”. En: *ISPRS International Journal of Geo-Information* 10.2 (2021). DOI: 10.3390/ijgi10020099.

1.3.1. Redes Neuronales Recurrentes Las Redes Neuronales Recurrentes (RNN) son una clase de redes neuronales diseñadas para procesar y reconocer patrones en datos secuenciales, como series temporales o texto, lo que las hace especialmente adecuadas para tareas de análisis de series temporales.²⁵ A diferencia de las redes neuronales tradicionales de tipo de propagación hacia adelante mencionadas anteriormente, las RNN tienen conexiones cíclicas que forman ciclos dirigidos, permitiéndoles mantener una "memoria" de las entradas anteriores. Por ejemplo, en la predicción de una serie temporal de temperaturas, una RNN puede recordar que hubo un descenso brusco en días anteriores y ajustar su predicción para los días siguientes en función de ese dato histórico.²⁶ Esto es particularmente útil para tareas donde el contexto histórico es crucial, ya que cada neurona en una RNN no solo recibe información de la capa anterior, sino también de su propia salida en el paso de tiempo previo, lo que influye en las predicciones futuras.²⁷

El componente central de una RNN es la célula recurrente, que es responsable de procesar la secuencia de entrada un paso a la vez, mientras actualiza un estado oculto que refleja la información acumulada de los pasos anteriores. Podemos formular matemáticamente este proceso para una secuencia de entrada x_t , $t = 1, \dots, n$, donde x_t representa la entrada en el paso de tiempo t . En nuestro caso, x_t sería el número de incidentes en la fecha t .²⁸

En cada paso de tiempo t , existe un estado oculto h_t que se calcula utilizando la entrada actual

²⁵ Hansika Hewamalage, Christoph Bergmeir y Kasun Bandara. "Recurrent Neural Networks for Time Series Forecasting: Current status and future directions". En: *International Journal of Forecasting* 37.1 (2021), págs. 388-427. DOI: <https://doi.org/10.1016/j.ijforecast.2020.06.008>.

²⁶ Edward Appau Nketiah et al. "Recurrent neural network modeling of multivariate time series and its application in temperature forecasting". En: *PLOS ONE* 18.5 (mayo de 2023), págs. 1-18. DOI: 10.1371/journal.pone.0285713.

²⁷ Safwan Mahmood Al-Selwi, Mohd Fadzil Hassan, Said Jadid Abdulkadir et al. "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review". En: *Journal of King Saud University-Computer and Information Sciences* (2024), pág. 102068.

²⁸ Zachary C. Lipton, John Berkowitz y Charles Elkan. *A Critical Review of Recurrent Neural Networks for Sequence Learning*. 2015. arXiv: 1506.00019 [cs.LG].

x_t y el estado oculto anterior h_{t-1} :

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \quad (8)$$

donde W_{hx} es la matriz de pesos que conecta la entrada con el estado oculto, W_{hh} es la matriz de pesos que conecta el estado oculto consigo mismo desde el paso de tiempo anterior, b_h es el vector de sesgo, y σ es una función de activación no lineal, típicamente una tangente hiperbólica o una función ReLU porque estas funciones permiten modelar relaciones no lineales y controlar el valor del estado oculto, evitando problemas como la saturación y el desvanecimiento del gradiente ²⁹. La salida \hat{y}_t en el paso de tiempo t se calcula utilizando el estado oculto actual h_t :

$$\hat{y}_t = g(W_{hy}h_t + b_y) \quad (9)$$

donde W_{hy} es la matriz de pesos que conecta el estado oculto con la salida, b_y es el vector de sesgo, y g es típicamente una función lineal o softmax, dependiendo de la tarea. En nuestro caso, g es una función lineal, ya que estamos tratando con una tarea de regresión.

El modelo RNN aprende a ajustar los pesos W_{hx} , W_{hh} y W_{hy} , así como los sesgos b_h y b_y , minimizando una función de pérdida sobre los datos de entrenamiento. Para la predicción de series temporales, una elección común de esta función es el error cuadrático medio (MSE).

En el contexto de este estudio, se emplean RNNs para predecir patrones delictivos basados en un conjunto de datos de series temporales.

1.3.2. Redes con Gran Memoria a largo corto plazo Las redes de Gran Memoria a Corto Plazo (LSTM, por sus siglas en inglés) son una variante avanzada de las Redes Neuronales Recurrentes (RNN), diseñadas para resolver el problema del desvanecimiento y la explosión del gradiente, que ocurre cuando los gradientes, durante el entrenamiento, se vuelven demasiado

²⁹ Jia Li et al. "Hyperbolic linear units for deep convolutional neural networks". En: *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2016, págs. 353-359.

pequeños o grandes, lo que dificulta la actualización de los pesos y provoca que la red aprenda de manera ineficaz o inestable. Las LSTM superan estas limitaciones al incorporar una estructura de celda más compleja, con células de memoria que pueden almacenar información durante largos periodos.³⁰

Dentro de cada célula de una LSTM, hay dos componentes principales: el estado de la celda (c_t), que actúa como la memoria a largo plazo, y el estado oculto (h_t), que contiene la información procesada en el paso de tiempo actual. El control del flujo de información hacia y desde estos estados se gestiona mediante tres puertas fundamentales: la puerta de olvido, la puerta de entrada y la puerta de salida, que regulan qué información debe conservarse, actualizarse o usarse en la salida.

En este contexto, el número de incidentes en la fecha t para una secuencia de entrada $x = (x_1, x_2, \dots, x_t)$, es procesado de manera estructurada por la red LSTM.

La puerta de olvido decide qué parte del estado anterior de la celda (c_{t-1}) debe descartarse, basándose en la entrada actual x_t y el estado oculto anterior h_{t-1} . Esta puerta se define por la siguiente ecuación:

$$f_t = \sigma (W_f x_t + U_f h_{t-1} + b_f) \quad (10)$$

Aquí, W_f es la matriz de pesos que ajusta la entrada actual, mientras que U_f es la matriz de sesgos que transforma el estado oculto anterior. Estos valores combinados determinan cuánto de la memoria anterior debe olvidarse o conservarse, con σ (función sigmoide) limitando los valores entre 0 y 1; b_f define el vector de sesgo. Las matrices de pesos U_f aplican una transformación lineal al estado oculto anterior, permitiendo que la red ajuste su respuesta en función de la información almacenada previamente, antes de que la puerta de olvido decida qué información conservar o descartar.

³⁰ F.A. Gers, J. Schmidhuber y F. Cummins. “Learning to forget: continual prediction with LSTM”. En: *1999 Ninth International Conference on Artificial Neural Networks ICANN99*. Vol. 2. 1999, 850-855 vol.2. DOI: 10.1049/cp:19991218.

La puerta de entrada controla cuánta de la nueva información debe incorporarse al estado de la celda. Se calcula mediante la siguiente ecuación:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (11)$$

Esta puerta utiliza W_i y U_i para transformar la entrada x_t y el estado oculto anterior h_{t-1} . Paralelamente, la red calcula un estado candidato de la celda (\tilde{c}_t), el cual representa la nueva información que podría añadirse al estado de la celda. Este estado candidato se define como:

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (12)$$

Aquí, las matrices W_c y U_c también transforman la entrada actual y el estado oculto anterior, pero esta vez se utiliza la función tangente hiperbólica \tanh , que restringe los valores de \tilde{c}_t a un rango entre -1 y 1, asegurando una actualización más controlada del estado de la celda.

Luego, el estado de la celda se actualiza combinando el estado anterior filtrado por la puerta de olvido con la nueva información regulada por la puerta de entrada. Esta actualización se expresa como:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (13)$$

donde \odot denota la multiplicación elemento por elemento.

Finalmente, la puerta de salida controla qué parte del estado actualizado de la celda se transfiere al estado oculto actual (h_t) y se utiliza como salida de la red. Esto se define mediante las siguientes ecuaciones:

$$\begin{aligned} o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \quad (14)$$

La puerta de salida utiliza W_o y U_o para calcular o_t , que regula la cantidad de información que se extrae del estado de la celda c_t (filtrada a través de la tangente hiperbólica \tanh) y se envía

como salida.

Este mecanismo estructurado permite que las LSTM almacenen, actualicen y recuperen información de manera eficiente, superando las limitaciones de las RNN tradicionales en el manejo de secuencias largas. En este estudio, se empleará LSTM para predecir patrones delictivos utilizando datos de series temporales de cada comuna.

1.4. ALGORITMOS ESPACIO-TEMPORALES

Los modelos espacio-temporales se utilizan para la predicción de crímenes al integrar datos geográficos y temporales, permitiendo identificar patrones complejos en la ocurrencia de delitos. Al considerar tanto la ubicación como el tiempo, estos modelos pueden anticipar la evolución de los crímenes en diferentes áreas, mejorando la precisión de las predicciones.

El modelo de Redes Neuronales Espacio-Temporales³¹ (por sus siglas en inglés STNN) se utiliza para la predicción de crímenes integrando datos geográficos y temporales. Este modelo permite identificar patrones complejos en la ocurrencia de delitos al considerar simultáneamente la ubicación y el tiempo, mejorando así la precisión de las predicciones.

El modelo STNN se formaliza como una red neuronal recurrente que aprende dependencias tanto temporales como espaciales a través de un componente dinámico estructurado en un espacio latente.

Dada una serie de tiempo x_1, x_2, \dots, x_t donde x_t representa los valores observados en el tiempo t , se define la dinámica temporal en un espacio latente para un tiempo t como Z_t . Este valor es una representación latente multidimensional (puede ser un vector o una matriz, dependiendo del número de series de tiempo) que encapsula información tanto temporal como espacial. Su valor se actualiza con el tiempo mediante una función dinámica g . Los espacios latentes Z_t se pueden transformar mediante una función d para representar observaciones reales. Esto se podría definir

³¹ Ali Ziat et al. “Spatio-Temporal Neural Networks for Space-Time Series Forecasting and Relations Discovery”. En: *2017 IEEE International Conference on Data Mining (ICDM)*. 2017, págs. 705-714. DOI: 10.1109/ICDM.2017.80.

matemáticamente como:

$$Z_{t+1} = g(Z_t) \quad (15)$$

$$\hat{X}_t = d(Z_t) \quad (16)$$

Donde g es una función dinámica, que en este caso es una RNN y d es un decodificador lineal el cual transforma los valores latentes a observaciones reales. El valor \hat{X}_t representa una predicción basada en Z_t . De esta manera, la serie temporal completa se modela a través de la evolución de Z_t en el tiempo, capturando tanto las dependencias temporales como espaciales en los datos a través de g y Z_t .

El proceso de aprendizaje se rige por una función de pérdida bi-objetivo $L(g, d, Z)$, que incluye dos componentes principales:

$$L(d, g, Z) = \frac{1}{T} \sum_t \Delta(d(Z_t), X_t) + \lambda \frac{1}{T} \sum_{t=1}^{T-1} \|Z_{t+1} - g(Z_t)\|^2 \quad (17)$$

Donde λ Es un hiperparámetro que se utiliza para equilibrar la función de pérdida bi-objetivo, Δ es una medida de error que denota la diferencia entre $d(Z - t)$ y X_t . Finalmente T es el tamaño de la serie de tiempo. El primer término asegura que el modelo pueda reconstruir los valores observados a partir de los factores latentes. El segundo término captura la dinámica de la serie en el espacio latente, obligando a Z_{t+1} a alinearse con $g(Z_t)$ mediante su distancia eucladiana, dado que los dos valores son vectores de nodos.

Para integrar las dependencias espaciales el modelo utiliza una matriz $W \in \mathbb{R}^{n \times n}$, que representa la influencia mutua entre diferentes series espaciales como pueden ser regiones geográficas. Este contexto espacio-temporal en la función dinámica g se describe como:

$$Z_{t+1} = g(Z_t) = h(Z_t \Theta^{(0)} + W Z_t \Theta^{(1)}) \quad (18)$$

Aquí $\Theta^{(0)}$ es una matriz que captura las dependencias intra-serie, mientras que $\Theta^{(1)}$ es una matriz que captura las dependencias inter-serie mediante la combinación de las representaciones

latentes de las series vecinas definidas por la matriz espacial W . La función h es una función no lineal como \tanh , que añade flexibilidad a la modelización de las dinámicas temporales y espaciales. Esta definición corresponde al modelo STNN.

1.4.1. STNN-R Es posible extender aún más el modelo STNN para aprender los pesos de las relaciones directamente de los datos, permitiendo que el modelo refine relaciones espaciales. Esta variante, denominada STNN-R, utiliza pesos aprendidos Γ para ajustar las relaciones:

$$Z_{t+1} = h(Z_t \Theta^{(0)} + (W \odot \Gamma) Z_t \Theta^{(1)}) \quad (19)$$

Aquí, \odot denota la multiplicación elemento a elemento y Γ es una matriz similar a W , a diferencia de que esta permite aprender o ajustar las relaciones espaciales entre las distintas series de tiempo.

1.4.2. STNN-D En el caso de que no se tenga una matriz de relaciones W el modelo puede operar en un modo de descubrimiento, aprendiendo las relaciones de manera autónoma a través de la matriz Γ , que reemplaza a la matriz W :

$$Z_{t+1} = h(Z_t \Theta^{(0)} + (\Gamma) Z_t \Theta^{(1)}) \quad (20)$$

Aquí, Γ no solo ajusta las relaciones espaciales, sino que las descubre directamente a partir de los datos, representando la influencia aprendida entre series temporales.

Estas distintas configuraciones permiten experimentar con diferentes tipos de relaciones espaciales y temporales, facilitando la comprensión de la dinámica espacio-temporal subyacente en las series de tiempo.

1.5. SOLUCIONES PLANTEADAS PARA PREDICCIÓN DE CRÍMENES

La predicción de robos en las ciudades se ha convertido en un área crítica de investigación para las agencias de seguridad pública y las fuerzas del orden. Al analizar los datos históricos de robos, los investigadores pueden identificar patrones y desarrollar herramientas para anticipar futuras ocurrencias. Varios enfoques de investigación existentes se basan en diferentes técnicas de aprendizaje automático. Por ejemplo, en ³², los autores utilizan el conjunto de datos públicos de robos de Chicago para la predicción de tendencias delictivas basado en un modelo LSTM. También incorpora datos de información de puntos de interés (POI) y emplea un modelo con redes neuronales convolucionales para analizar y comparar la distribución temporal, la distribución espacial y la precisión de la predicción de varios modelos algorítmicos. Los resultados experimentales demuestran una tasa de acierto del 20,1% de los delitos en diferentes experimentos utilizando el algoritmo LSTM.

Otro ejemplo se aplica en Atlanta, una de las ciudades con mayores tasas de criminalidad en los Estados Unidos. En ³³, los autores muestran que los eventos delictivos exhiben agregación espacial y dependencia temporal, indicando que los incidentes criminales son predecibles. En el mencionado artículo, se utilizan datos de robos de Atlanta, que abarcan desde 2009 hasta 2016, para identificar características de distribución espacio-temporal y crear visualizaciones estadísticas de las zonas con mayor criminalidad. En dicho trabajo, la previsión de ocurrencias diarias de robos se obtiene mediante un modelo LSTM. El estudio explora además el impacto de variar las escalas espacio-temporales, es decir, variando el tamaño de las predicciones a semanales y mensuales, como la precisión espacial de coordenadas a barrios. Al usar una serie

³² Nan Jiang et al. “Spatio-temporal prediction of crime based on Data Mining and LSTM network”. En: *2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (IT-NEC)*. Vol. 6. 2023, págs. 672-676. DOI: 10.1109/ITNEC56291.2023.10081985.

³³ Shihuai Wang y Kunxiao-jia Yuan. “Spatiotemporal Analysis and Prediction of Crime Events in Atlanta Using Deep Learning”. En: *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*. 2019, págs. 346-350.

temporal de entrada de 50 días y un tamaño de celda espacial de 0,05 grados, el coeficiente de correlación (valor R) entre los datos de robos predichos y observados supera 0,87.

Finalmente, en ³⁴ se aplica un enfoque espacio-temporal para modelar los robos mensuales en Buenos Aires (Argentina). El objetivo principal de este estudio es implementar modelos predictivos para asociar esquinas específicas y tiempos con la ocurrencia de robos, basándose en los registros mensuales de robos en Buenos Aires desde 2017 hasta 2020. Inicialmente, se emplean modelos de series temporales que utilizan solo los conteos históricos de robos para predecir incidentes criminales en cada esquina. Luego, las predicciones incluyen covariables relacionadas con el entorno circundante y las condiciones climáticas, transformando el problema en un estudio espacio-temporal. El estudio muestra que los robos, al ser una actividad humana, no siguen un comportamiento aleatorio, sino patrones sociales influenciados por el tiempo, el clima y las zonas geográficas. En algunos casos, se observa que algunas características espaciales tienen un impacto en la toma de decisiones de las predicciones en mas de un 13%. ³⁴

³⁴ Rafael Zambrano. “Un enfoque Espacio Temporal para la predicción de delitos en la ciudad de Buenos Aires”. En: *Revista de Investigación de Modelos Matemáticos Aplicados a la Gestión y a la Economía*. Vol. 2. 2022, págs. 38-62.

2. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

En el ámbito de la seguridad ciudadana, la predicción de robos mediante modelos espaciotemporales ha adquirido una gran relevancia a nivel mundial.^{35,36} Estos modelos permiten identificar patrones y tendencias en la ocurrencia de delitos en diferentes áreas geográficas a lo largo del tiempo³⁷, lo que facilita la implementación de estrategias de prevención y combate de la criminalidad. A nivel local, la ciudad de Bucaramanga no ha sido ajena a este desafío, ya que ha experimentado un preocupante incremento en el número de hurtos a personas durante los años 2020 y 2021, alcanzando una tasa de 1.221 robos por cada 100.000 habitantes, según el informe de calidad de vida entregado por el programa Bucaramanga Metropolitana Cómo Vamos³⁸. Este aumento de la actividad delictiva ha generado una mayor percepción de inseguridad entre los habitantes, provocando cambios en su comportamiento y afectando negativamente su calidad de vida.³⁹

Tradicionalmente, se han realizado análisis con modelos paramétricos, utilizando medidas como las correlaciones espaciales para identificar grupos o zonas con mayores índices de delincuencia

³⁵ Xu Zhang et al. “Comparison of Machine Learning Algorithms for Predicting Crime Hotspots”. En: *IEEE Access* 8 (2020), págs. 181302-181310. DOI: 10.1109/ACCESS.2020.3028420.

³⁶ Alcaldía de Bogotá. *Bogotá desarrollará un método de predicción de delitos*. 25 de jun. de 2024. URL: <https://bogota.gov.co/mi-ciudad/seguridad/nuevo-metodo-de-prediccion-de-delitos-sera-desarrollado-en-bogota> (visitado 12-02-2019).

³⁷ Anthony A. Braga, Andrew V. Papachristos y David M. Hureau. “The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis”. En: *Justice Quarterly* 31.4 (2014), págs. 633-663. DOI: 10.1080/07418825.2012.673632. eprint: <https://doi.org/10.1080/07418825.2012.673632>.

³⁸ El informe se puede encontrar en: <https://www.bucaramangacomovamos.org/post/informe-de-calidad-de-vida-2022>

³⁹ Nor Azman Mat Ali Mahadzirah Mohamad Morliyati Mohammad y Zainudin Awang. “The impact of life satisfaction on substance abuse: delinquency as a mediator”. En: *International Journal of Adolescence and Youth* 23.1 (2018), págs. 25-35. DOI: 10.1080/02673843.2016.1267021. eprint: <https://doi.org/10.1080/02673843.2016.1267021>.

⁴⁰ o modelos autoregresivos⁴¹ para identificar tendencias ⁴². En la literatura, se han propuesto varias estrategias con aprendizaje de máquina para predecir robos utilizando análisis de series de tiempo. Entre estas técnicas se encuentran las Máquinas de Soporte Vectorial (SVM), reconocidas por su versatilidad y facilidad de implementación para predecir con un enfoque temporal.⁴³

Recientemente, se han utilizado redes neuronales para predecir series tiempo con el objetivo de mejorar el ajuste en comparación con los modelos anteriormente mencionados, ya que los modelos paramétricos pueden tener problemas a la hora de predecir grandes volúmenes de datos y los modelos de redes neuronales logran mejores ajustes que los modelos de aprendizaje de máquina.⁴⁴

En este contexto, el presente trabajo de investigación plantea como objetivo principal el desarrollo de un modelo de Aprendizaje de Máquina, utilizando redes neuronales recurrentes para series espacio temporales, con el fin de realizar predicciones de robos en Bucaramanga. Se espera que este enfoque permita identificar las distintas tendencias de robo en la ciudad de Bucaramanga comparado con los métodos tradicionales.

⁴⁰ Martin A. Andresen. “Estimating the probability of local crime clusters: The impact of immediate spatial neighbors”. En: *Journal of Criminal Justice* 39.5 (2011), págs. 394-404. DOI: <https://doi.org/10.1016/j.jcrimjus.2011.05.005>.

⁴¹ Un modelo autorregresivo es un tipo de modelo estadístico en el que el valor actual de una variable se predice utilizando una combinación lineal de sus valores pasados.

⁴² Romika Yadav y Savita Kumari Sheoran. “Crime Prediction Using Auto Regression Techniques for Time Series Data”. En: *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*. 2018, págs. 1-5. DOI: 10.1109/ICRAIE.2018.8710407.

⁴³ Ping Wang et al. “Predicting Criminal Recidivism with Support Vector Machine”. En: *2010 International Conference on Management and Service Science*. 2010, págs. 1-9. DOI: 10.1109/ICMSS.2010.5575352.

⁴⁴ Sri Adi Pavan Naidu Kavala et al. “Crime analysis and Prediction using Neural Networks”. En: *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Vol. 1. 2022, págs. 2047-2050. DOI: 10.1109/ICACCS54159.2022.9785081.

3. OBJETIVOS

Objetivo general

Implementar una técnica de aprendizaje automático para la predicción espacio-temporal de la ocurrencia de robos.

Objetivos específicos

- Construir una base de datos refinada a partir de los datos provenientes de la plataforma de datos abiertos Colombia⁴⁵, mediante la aplicación de métodos de depuración de bases de datos.
- Realizar un análisis estadístico descriptivo y exploratorio de los robos cometidos en Bucaramanga, con el fin de identificar patrones, tendencias y características relevantes.
- Implementar un modelo predictivo usando el modelo de ARIMA que funcione como línea base para la comparación del modelo propuesto.
- Implementar un modelo con redes neuronales para series de tiempo ajustada por el conjunto de datos de robos acontecidos en Bucaramanga.
- Implementar una red neuronal recurrente para series espacio-temporales ajustada por el conjunto de datos de robos acontecidos en Bucaramanga.
- Evaluar los modelos implementados empleando la raíz del error cuadrático medio como métrica para el análisis de series espacio-temporales.

⁴⁵ Cuyo enlace corresponde a <https://www.datos.gov.co/>

4. DESCRIPCIÓN DEL CASO DE ESTUDIO

En esta sección se realizará la descripción del área geográfica en la cual se realizó modelo de predicción de delitos. También se expondrán los datos de interés utilizados en el desarrollo del modelo.

4.1. ZONA DE ESTUDIO

La zona de estudio se centra en la ciudad de Bucaramanga, Colombia. Específicamente en su zona urbana, la cual está dividida en 17 comunas.

4.1.1. Ubicación geográfica Bucaramanga es la capital del departamento de Santander, Colombia. Está ubicada en una meseta a 960 metros sobre el nivel del mar, con una población de aproximadamente 600,000 habitantes. Situada en los Andes orientales de Colombia, Bucaramanga se caracteriza por su topografía ondulante y su densa urbanización.

4.1.2. Contexto Geoespacial Bucaramanga, al estar situada en una meseta sobre la cordillera de los Andes y extenderse sobre una serie de colinas y valles, resulta en variaciones en la accesibilidad y visibilidad a lo largo de diferentes áreas causada por los tres grandes cerros se destacan a lo largo del territorio: Morro Rico al este de la ciudad, Alto de San José al noreste de la ciudad y El Cacique al sur de la ciudad. Los cuales aislan ligeramente las comunas ubicadas al norte y el oriente de la ciudad. La temperatura media es de 23°C.⁴⁶

⁴⁶ Área Metropolitana de Bucaramanga. *Bucaramanga*. AMB. 10 de abr. de 2012. URL: <https://www.amb.gov.co/bucaramanga/> (visitado 31-08-2023).

4.1.3. Datos Demográficos De acuerdo con el último censo realizado por el DANE ⁴⁷, la población de la zona de estudio son 528.855 personas efectivamente censadas y la distribución de la población por sexo es la siguiente 47.5 % para hombres y 52.5 % mujeres.

4.1.4. División Administrativa Bucaramanga se dividen en 17 comunas. Dentro de cada comuna hacen parte los barrios, asentamientos, urbanizaciones y otros (sectores con población flotante). Las comunas que componen a Bucaramanga son las siguientes: Comuna 1 Norte, Comuna 2 Nororiental, Comuna 3 San Francisco, Comuna 4 Occidental, Comuna 5 García Rovira, Comuna 6 La Concordia, Comuna 7 La Ciudadela, Comuna 8 Sur Occidente, Comuna 9 La Pedregosa, Comuna 10 Provenza, Comuna 11 Sur, Comuna 12 Cabecera del Llano, Comuna 13 Oriental, Comuna 14 Morrórico, Comuna 15 Centro, Comuna 16 Lagos del Cacique y 17 Mutis.⁴⁸ En la figura 1 se puede ver la distribución espacial de cada una de las comunas.

4.2. DESCRIPCIÓN DE FUENTES DE DATOS

En esta sección se detallarán los procedimientos, datos relevantes, métodos de procesamiento, y otras consideraciones previstas para el tratamiento y manipulación de los datos.

4.2.1. Datos temporales Los datos utilizados para el estudio se encuentran disponibles en el portal de datos abiertos. Estos datos son suministrados por la Alcaldía de Bucaramanga cuya publicación es realizada por la oficina asesora de TIC bajo el proyecto "Datos abiertos observatorio digital municipal de Bucaramanga".

Los datos obtenidos incluyen registros de los delitos cometidos en Bucaramanga, con cada fila representando un delito específico. Los delitos son tomados diariamente y agregados a la base

⁴⁷ Departamento Administrativo Nacional de Estadística (DANE). *Censo Nacional de Población y Vivienda 2018*, DANE. 10 de mayo de 2019.

⁴⁸ Alcaldía de Bucaramanga. *División político urbana y rural*. Alcaldía de Bucaramanga. 30 de ago. de 2024. URL: <https://www.bucaramanga.gov.co/division-politico-urbana/> (visitado 30-08-2024).

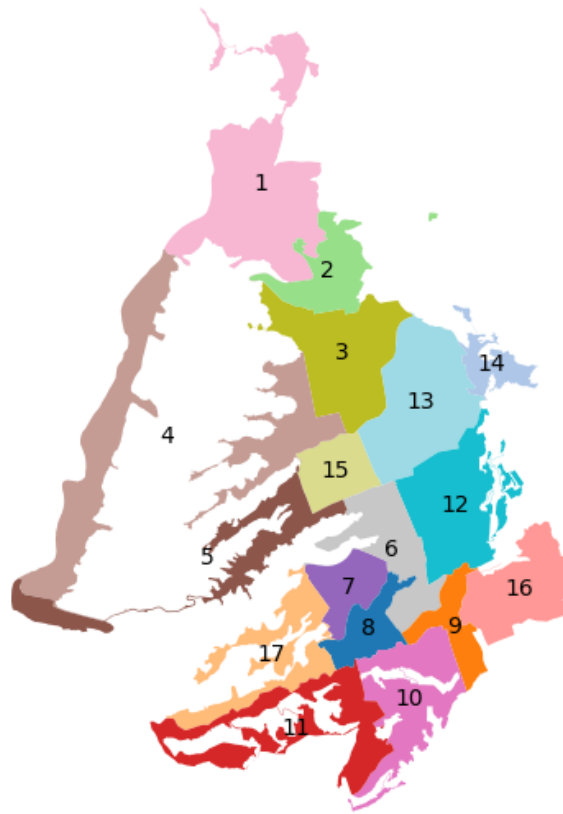


Figura 1. Distribución geográfica de las comunas de Bucaramanga. Creación propia

de datos. Para que un delito sea registrado, debe haberse presentado una denuncia, por lo que los delitos no denunciados no se incluyen. El periodo de registro abarca desde enero de 2016 hasta junio de 2023. Los datos poseen las siguientes características: descripción del delito, armas/medios utilizados, barrio donde ocurrió el delito, fecha del delito, edad de la víctima, sexo de la víctima, rango de edad de la víctima, clasificación del delito, nombre de la comuna, número de la comuna y cantidad de afectados.

El enfoque de este trabajo es modelar las tendencias de los robos en las diferentes comunas de manera espacio-temporal. Por esta razón, se filtrarán únicamente los delitos clasificados como robos a peatones, junto con sus fechas y la comuna correspondiente.

4.2.2. Datos Geoespaciales Los datos geográficos del área de Bucaramanga fueron suministrados por la Oficina de Instrumentos Públicos. Estos datos representan geográficamente a través de polígonos cada una de las comunas de la ciudad de Bucaramanga y sus fronteras.

4.3. EXPLORACIÓN DE LOS DATOS

En esta sección se analizan los datos de robos en Bucaramanga mediante diversas representaciones gráficas. Estas visualizaciones permiten contextualizar patrones, tendencias y distribuciones espaciales y temporales de los robos, proporcionando una comprensión más profunda del fenómeno delictivo en la ciudad. Este análisis es fundamental para informar el desarrollo y la validación de los modelos predictivos utilizados en el estudio.

4.3.1. Frecuencia de los robos por año y mes A través de la observación del número de robos mensuales en la figura 2 durante el período de enero de 2016 hasta junio de 2023, se puede identificar una tendencia al alza en la incidencia de robos en la ciudad de Bucaramanga. Un aspecto notable es la marcada disminución en el número de robos durante el año 2020, lo cual coincide con el período de la pandemia de COVID-19, cuando se implementaron medidas de confinamiento y restricciones de movilidad.⁴⁹ Sin embargo, tras el levantamiento de las restricciones, se observa una tendencia a repunte significativo en el número de robos, superando incluso los niveles previos a la pandemia. Este aumento podría ser una consecuencia directa de la crisis económica y social generada por la pandemia, así como también a cambios en las dinámicas urbanas y comportamientos delictivos.⁵⁰

⁴⁹ Jorge Luis Vargas Valencia et al. “Criminalidad en contexto COVID año 2020 y aproximación de una propuesta en neurocriminología”. En: *Revista Criminalidad* 63.3 (2022), 9–31. DOI: 10.47741/17943108.310.

⁵⁰ *Se agudiza la criminalidad en Colombia: cada día más de mil personas son víctimas de hurtos y/o extorsión*. 25 de jun. de 2024. URL: <https://cej.org.co/destacados-home-page/se-agudiza-la-criminalidad-en-colombia-cada-dia-mas-de-mil-personas-son-victimas-de-hurtos-y-o-extorsion/> (visitado 21-09-2023).

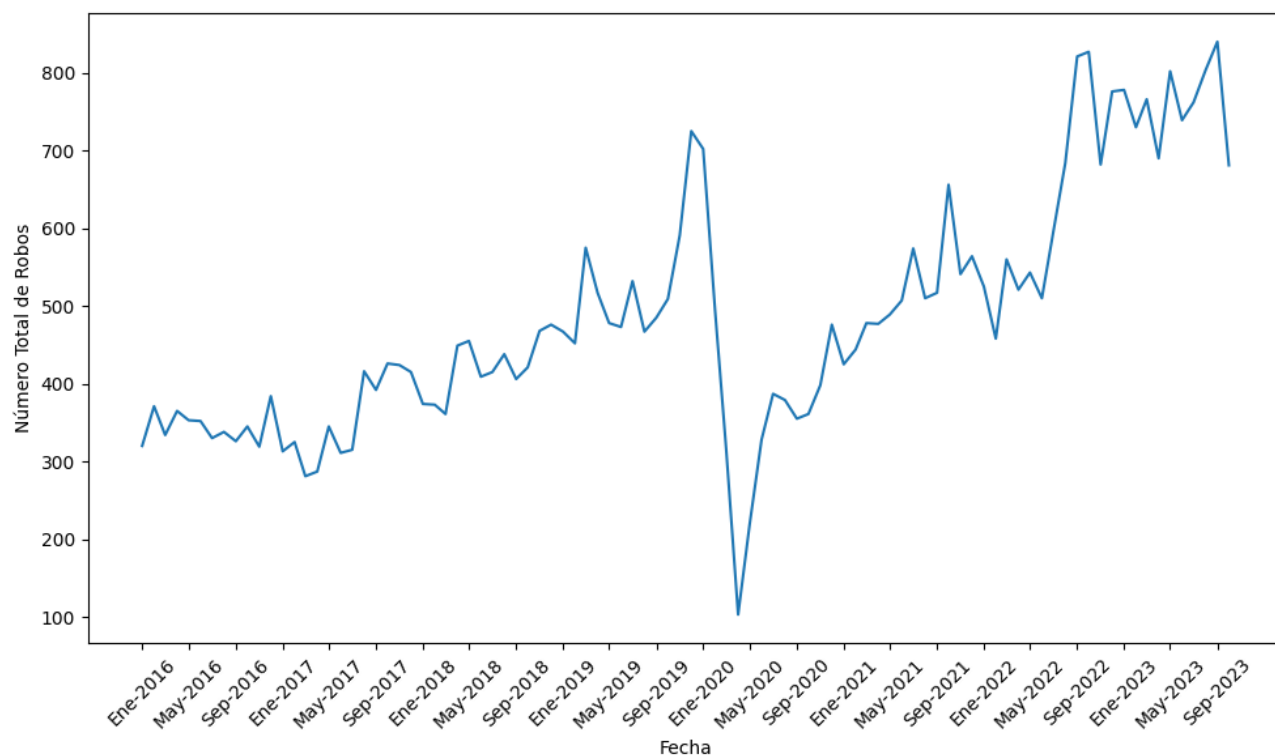


Figura 2. Número total de robos mensuales para el período de estudio. Creación propia

4.4. DISTRIBUCIÓN GEOGRÁFICA DE LOS ROBOS POR COMUNA

En la figura 3 se puede evidenciar la distribución geográfica de los robos para cada una de las comunas durante todo el periodo del estudio. La comuna 13 (Oriental) destaca con un color rojo intenso, lo que indica la mayor frecuencia de robos, seguida por las comunas 12 (Cabecera del Llano) y 15 (Centro). A pesar de que la comuna 13 tiene menor actividad comercial, en ella se encuentran barrios con alta estratificación e ingresos, lo que podría influir en su elevada tasa de robos. Por otro lado, las comunas 12 y 15 corresponden a zonas mayormente comerciales, donde el constante flujo de peatones y la afluencia de personas durante el día pueden explicar las altas tasas de incidencia delictiva en estas áreas.

Por otro lado, las comunas 7 (La Ciudadela), 8 (Sur Occidente), 10 (Provenza) y 17 (Mutis) presentan tonos azulados, lo que indica una menor incidencia de robos. Esto puede atribuirse a

su carácter predominantemente residencial y a un menor flujo de personas en comparación con otras áreas de la ciudad. Además, estas comunas están geográficamente aisladas de los barrios circundantes debido a la presencia de cerros en sus alrededores, lo que podría contribuir a la reducción de la actividad delictiva al limitar el acceso y el tránsito en estas zonas.

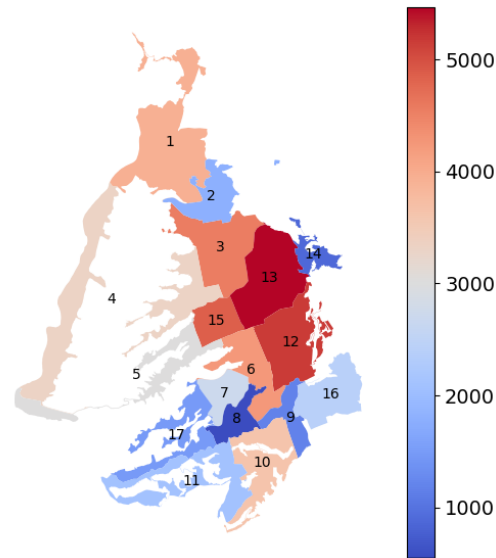


Figura 3. Distribución geográfica de los robos. Creación propia

De igual manera, las series de tiempo presentadas en la figura 4 permiten identificar patrones y tendencias particulares en cada comuna. Por ejemplo, algunas comunas, como la 3, 12, 13 y 15, muestran un aumento constante en la frecuencia de robos a lo largo del tiempo, especialmente a partir de 2019-2020, coincidiendo con su naturaleza comercial o de alto tránsito peatonal. Por otro lado, comunas como la 6 y 8 presentan un aumento moderado, mientras que otras, como la 1, 4, 5 y 14, muestran una menor variabilidad, sin una tendencia claramente creciente.

Además, algunas comunas como la 2, 7, 10 y 11 revelan patrones cíclicos con picos recurrentes en ciertas épocas del año, lo que sugiere la existencia de factores estacionales que influyen en la incidencia de robos.

También se observan picos y valles en el número de robos durante el período de la pandemia de COVID-19 y en el periodo post-pandemia después del 2021. La caída notable durante el 2020 está asociada con las restricciones de movilidad y confinamiento, mientras que el aumento

posterior refleja la relajación de estas restricciones y los cambios en las dinámicas sociales y económicas. En muchas comunas, como la 12 y 13, este incremento post-pandemia es particularmente pronunciado.

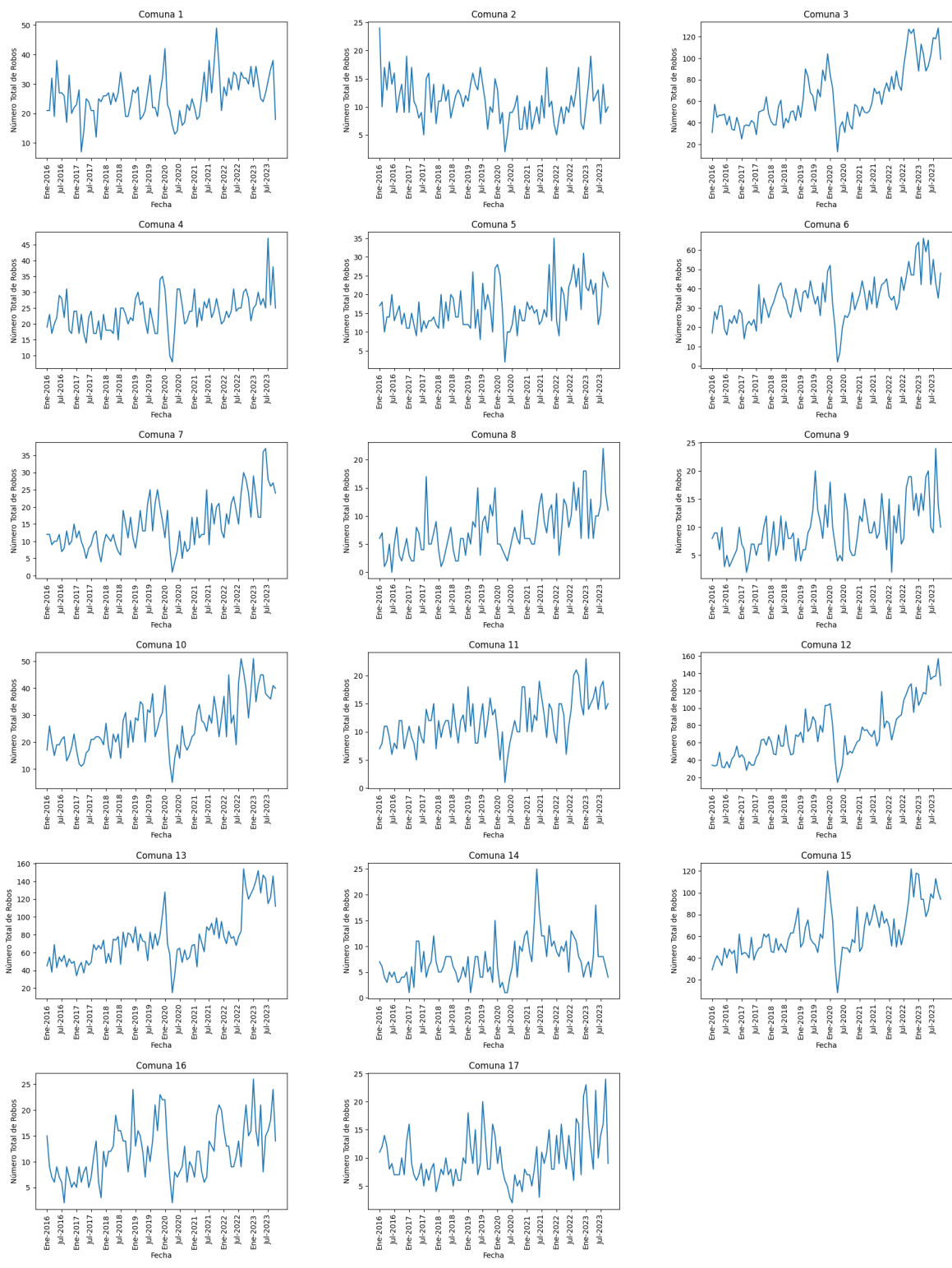


Figura 4. Número total de robos mensuales por comuna durante el periodo de estudio. Creación propia

5. METODOLOGÍA

5.1. PROCESAMIENTO DE DATOS

El procesamiento de datos es una etapa crucial que incluye la limpieza, transformación y estandarización de los datos, asegurando su calidad y coherencia para el análisis posterior. Esta etapa corresponde a las subetapas:

5.1.1. Limpieza de datos En la etapa inicial del tratamiento de datos, se filtró el conjunto de datos para crear una serie de tiempo específica para cada comuna. Para ello, se calculó el total de delitos diarios y se agrupó por comuna. El resultado fue una tabla en la que cada columna representa el número de delitos diarios ocurridos en cada comuna, junto con una columna con la del registro. En el caso de que no haya registros para una fecha específica, se asigna un valor de 0 robos.

5.1.2. Construcción del conjunto de datos Debido a la alta frecuencia de días con valores de 0 robos y con el objetivo de modelar de manera más efectiva las tendencias en la dinámica de robos en la ciudad, se decidió agrupar los delitos por semanas. Esta agrupación semanal, en lugar de un enfoque diario fue elegida no considerar las fluctuaciones diarias. Esta metodología también permite capturar patrones estacionales semanales y facilita la planificación y evaluación de estrategias de seguridad pública, que a menudo se desarrollan en ciclos semanales como en el trabajo realizado en ³³.

Para implementar esta metodología, se dividió el año en semanas y se calculó el número total de robos para cada semana. Se estableció un límite superior de 23 robos por semana en una comuna como umbral para identificar valores atípicos, correspondiente a la diferencia entre el cuartil 3 y 1,5 veces el rango intercuartil. En la figura 5 se puede observar la distribución de robos por comuna mediante un diagrama de cajas. Las fechas fueron etiquetadas en el formato

año-semana para facilitar su cálculo y análisis.

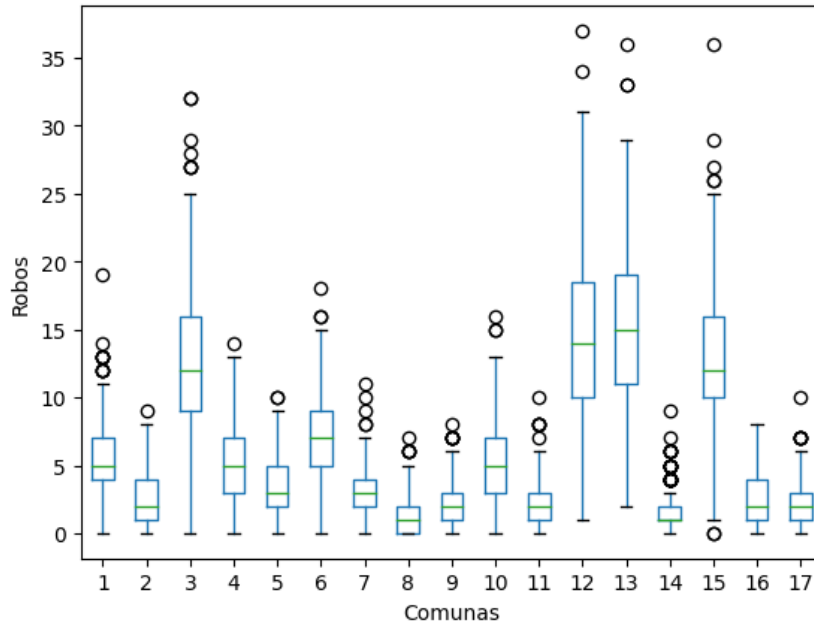


Figura 5. Diagrama de caja de la distribución de robos por comuna al agrupar semanalmente. Creación propia

5.1.3. Representación espacial Para representar espacialmente las comunas de Bucaramanga, se utilizó una matriz de adyacencia, que es una estructura matemática utilizada para describir la relación de vecindad entre diferentes elementos. En esta matriz, se asignó un valor de 1 cuando existía una frontera compartida entre dos comunas, y un valor de 0 cuando no había tal frontera. Los límites entre las comunas se definieron utilizando el mapa de Bucaramanga de la figura 1 como referencia.

En la figura 6 se observa una representación por medio de un grafo de las relaciones entre cada una de las comunas. Se puede observar el aislamiento geográfico de las comunas 1 y 2 del resto de la ciudad.

5.1.4. Ventaneo temporal Se utilizó la técnica de ventaneo temporal para estructurar el conjunto de datos de manera que los modelos predictivos logren capturar las dinámicas

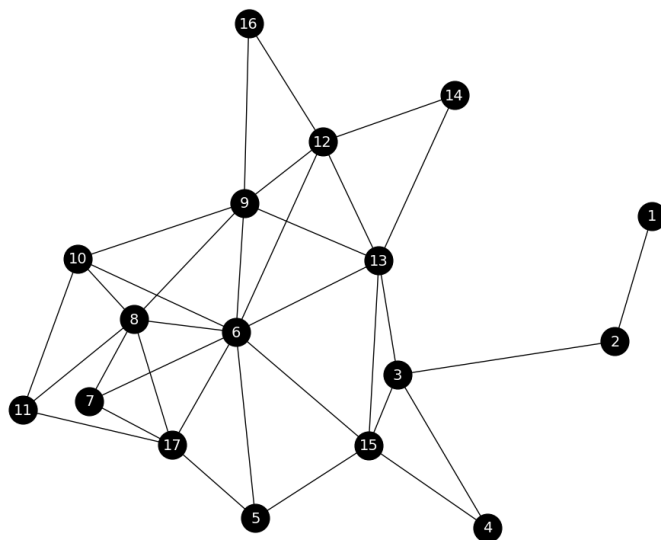


Figura 6. Representación gráfica mediante un grafo de las relaciones geográficas entre las comunas. Creación propia

temporales inherentes a la serie de tiempo⁵¹. El ventaneo temporal implica dividir la serie en ventanas de longitud fija, donde cada ventana representa un subconjunto secuencial de datos que el modelo utiliza para hacer predicciones. En este caso, se empleó un ventaneo de 4 semanas, debido a que es el periodo de estacionalidad observado en el análisis de datos.

El uso de un ventaneo de 4 semanas permite capturar de manera efectiva los ciclos mensuales y estacionales asegurando que el modelo aprenda patrones tanto a corto como a largo plazo. Al exponer el modelo a diversas condiciones temporales.

5.1.5. Division de datos en conjuntos de entrenamiento y prueba Para el desarrollo de los modelos, se dividió el conjunto de datos en un conjunto de entrenamiento y un conjunto

⁵¹ Rafael Carmona-Cabezas et al. “A sliding window-based algorithm for faster transformation of time series into complex networks”. En: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.10 (oct. de 2019), pág. 103121. DOI: 10.1063/1.5112782. eprint: https://pubs.aip.org/aip/cha/article-pdf/doi/10.1063/1.5112782/16714740/103121_1_online.pdf.

de prueba. El conjunto de entrenamiento incluye datos desde enero de 2016 hasta el 10 de octubre de 2021, cubriendo un período de casi seis años. El conjunto de prueba, abarca desde octubre de 2021 hasta finales de noviembre de 2022.

Adicionalmente, con el fin de evaluar la capacidad del modelo para predecir bajo diferentes dinámicas temporales, se implementó una validación cruzada temporal con 5 particiones. Cada partición representa el 20 % del total de datos, resultando en cuatro particiones de 59 semanas y una primera partición de 63 semanas, que cubren un total de 359 semanas. El proceso consistió en entrenar el modelo con la primera partición y evaluarlo con la segunda. Luego, entrenar con las dos primeras particiones y se evaluar con la tercera, y así sucesivamente, hasta que se haya utilizado todo el conjunto de datos. El RMSE final se calculó promediando los errores obtenidos en cada etapa de validación. La figura 7 representa un ejemplo de cómo fue la división en el conjunto de datos sobre la comuna 12. En color verde se puede observar el conjunto de datos entrenado y en rojo el conjunto de datos evaluado.

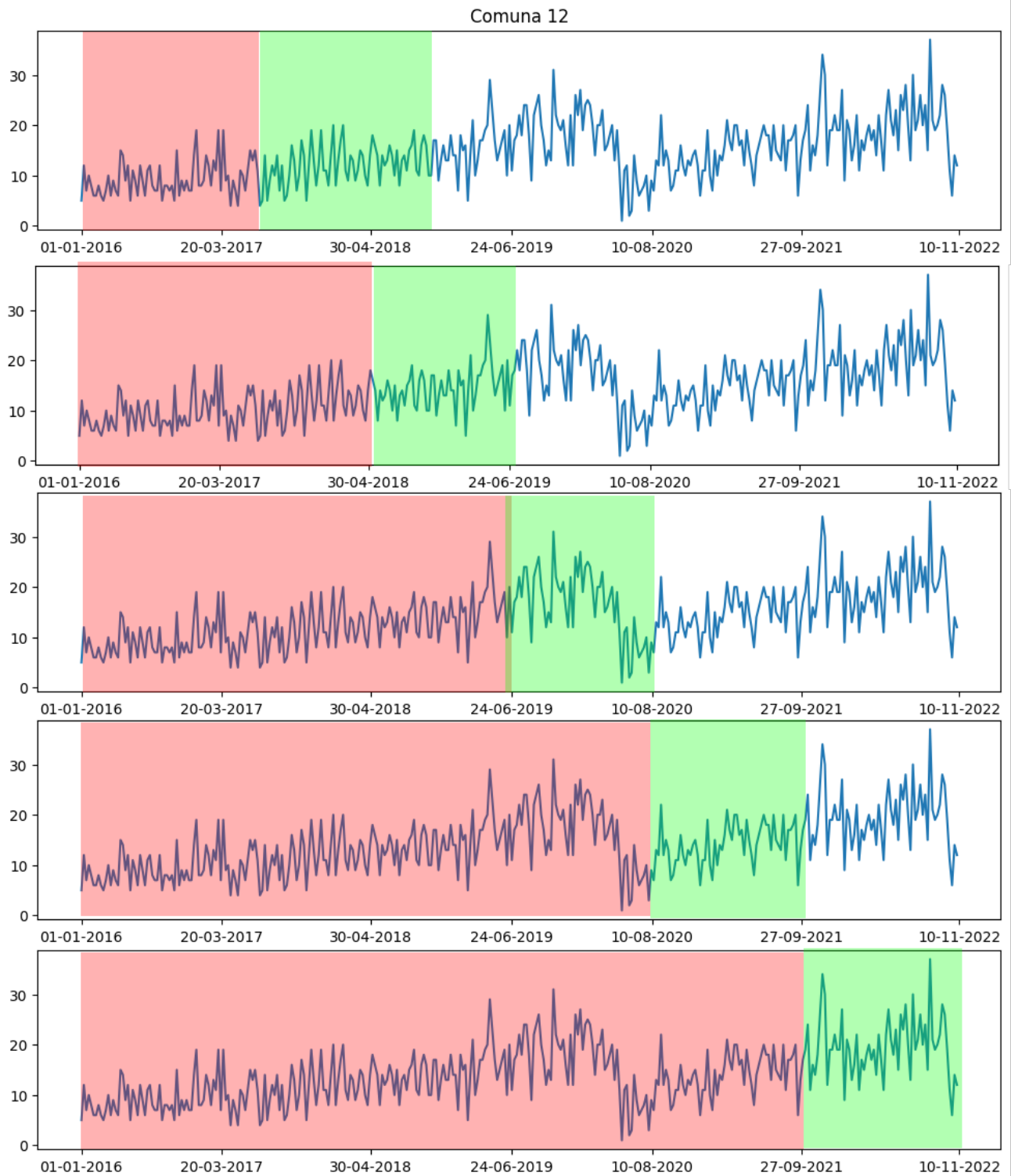


Figura 7. Representación de la división del conjunto de datos para validación cruzada. Creación propia

6. EXPERIMENTOS Y RESULTADOS

En este capítulo se definen las métricas utilizadas para evaluar los modelos. Así mismo, se definirán también las distintas configuraciones experimentales usadas para los modelos y los resultados de las predicciones de cada uno.

6.1. CONFIGURACIÓN EXPERIMENTAL

6.1.1. Métricas de evaluación

RMSE El rendimiento de los modelos propuestos para predecir patrones de crimen se evalúa utilizando la Raíz del Error Cuadrático Medio (RMSE). RMSE es una métrica comúnmente utilizada para tareas de regresión y se define de la siguiente manera:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2} \quad (21)$$

donde n es el número de observaciones, \hat{x}_i es el valor predicho para la i -ésima observación y x_i es el valor real para la i -ésima observación.

Valores más bajos de RMSE indican un mejor ajuste del modelo a los datos, ya que implica que las diferencias entre los valores predichos y los valores reales son menores.⁵²

SMAPE El SMAPE⁵³ es una métrica utilizada para evaluar el rendimiento de los modelos de predicción, especialmente en series temporales. A diferencia de otras métricas, el SMAPE tiene la ventaja de ser simétrico, lo que significa que penaliza tanto los errores de sobreestimación

⁵² “Mean Squared Error”. En: *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, págs. 337-339. DOI: 10.1007/978-0-387-32833-1_251.

⁵³ Error Porcentual Absoluto Medio Simétrico, por sus siglas en inglés

como los de subestimación de manera equitativa. Esta métrica se expresa como un porcentaje, lo que facilita su interpretación en diferentes contextos.

La ecuación para calcular el SMAPE es la siguiente:

$$SMAPE = \frac{100\%}{n} \sum_{t=1}^n \frac{|x_t - \hat{x}_t|}{\frac{(|x_t| + |\hat{x}_t|)}{2}} \quad (22)$$

Donde x_t es el valor real en el instante de tiempo t . \hat{x}_t es el valor predicho por el modelo en el instante t y n es el número total de observaciones.

El SMAPE varía entre 0 % y 200 %, donde 0 % indica una predicción perfecta y valores cercanos a 200 % reflejan un modelo con predicciones inexactas. Cuanto más bajo sea el SMAPE, mejor será el ajuste del modelo a los datos observados. Es importante destacar que, debido a su estructura, el SMAPE es más adecuado para datos que no contengan valores cercanos a cero, ya que en estos casos la métrica podría producir resultados sesgados.

6.1.2. Configuración de los vectores de características Una vez que los datos fueron preprocesados, se procedió a estandarizarlos antes del entrenamiento de cada uno de los modelos. En este caso, se utilizó el método de escalado min-max debido a las diferencias en las magnitudes de los datos.⁵⁴ El escalado min-max consiste en ajustar los valores de los datos para que se ubiquen dentro de un intervalo específico, en este caso $[0, 1]$. Este proceso puede definirse matemáticamente como:

$$z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (23)$$

Donde z_i es el valor normalizado, x_i es el valor original X_{min} es el valor mínimo de los datos y X_{max} es el valor máximo de los datos.

⁵⁴ Samit Bhanja y Abhishek Das. *Impact of Data Normalization on Deep Neural Network for Time Series Forecasting*. 2019. arXiv: 1812.05519 [cs.LG].

6.2. EXPERIMENTOS

En esta sección, se describen en detalle los experimentos realizados, abordando la configuración experimental y la metodología específica empleada para evaluar el rendimiento de los modelos. Se explican las decisiones tomadas en cada experimento y cómo se diseñaron para abordar preguntas específicas de investigación.

Para evaluar y comparar el rendimiento de los diferentes enfoques predictivos, se implementaron los siguientes modelos y experimentos:

1. ARIMA: Este modelo, conocido por su fácil implementación y capacidad para generar predicciones aceptables. Se empleará como línea base para medir el desempeño de otros modelos.
2. Red Neuronal recurrente: Con el fin de contrastar la capacidad de las predicciones espacio-temporales, se utiliza una RNN que solo va a aprender de la información temporal de cada una de las comunas, aprendiendo de cada comuna de manera independiente.
3. Red LSTM: Se implementa una red LSTM para capturar dependencias a largo plazo en las series de tiempo. Este modelo se utiliza para comparar el rendimiento de la RNN con un enfoque más avanzado en el manejo de patrones temporales complejos.

Para el modelo espacio temporal (STNN), se definen tres posibles experimentos adicionales utilizando las distintas configuraciones que ofrece el algoritmo utilizado:

4. STNN: Este modelo recibirá la información de las relaciones como una matriz de adyacencia.
5. STNN-R (STNN-Refine): Se aplica el mismo diseño de STNN, pero esta vez, el modelo también tiene la oportunidad de aprender y mejorar las relaciones que se reciben en su matriz de adyacencia.

6. STNN-D (STNN-Discover): En este experimento no se utiliza ninguna matriz de adyacencia como valor inicial, sino que en este caso, el modelo aprenderá y definirá su propia matriz de adyacencia a partir de la interacción de los datos durante el entrenamiento.

Para los modelos ARIMA, RNN y LSTM, se optimizarán los hiperparámetros de manera independiente para cada comuna, con el objetivo de obtener los resultados óptimos respecto a la métrica de evaluación RMSE. En cambio, para los experimentos con STNN, se emplearán los mismos hiperparámetros en todas las comunas, ya que el enfoque principal se centra en la búsqueda de dependencias espaciales. Las variaciones en los parámetros se aplicarán únicamente para las diferentes configuraciones experimentales específicas de cada modelo STNN.

En cada uno de los experimentos, los hiperparámetros se optimizarán de manera independiente con el objetivo de obtener las mejores predicciones posibles en función de la métrica RMSE. Este proceso de optimización se llevará a cabo para cada modelo utilizado, mediante múltiples iteraciones en las que se ajustarán los parámetros de manera gradual. El propósito de estas iteraciones es encontrar los valores óptimos de los hiperparámetros que maximicen la precisión del modelo y minimicen el error.

El código para el desarrollo de cada uno de los experimentos se implementó utilizando el lenguaje de programación Python, y la librería "TensorFlow Keras". Los experimentos se ejecutaron sobre la nube de Google Collab.

6.3. RESULTADOS

Para validar la validación de los experimentos, se evaluó el RMSE de los modelos para cada una de las comunas.

6.3.1. Optimización de Hiperparámetros Los distintos parámetros p, d, q del modelo ARIMA se pueden ver en la tabla 1, los parámetros óptimos se calcularon mediante la librería `pdarima` y su función `auto_arima`.

COMUNA	p	d	q
1	5	1	0
2	5	1	0
3	0	1	1
4	5	1	0
5	5	1	0
6	0	1	1
7	0	1	1
8	5	1	0
9	5	1	0
10	0	1	1
11	5	1	0
12	0	1	1
13	2	1	3
14	5	1	0
15	2	1	1
16	0	1	1
17	5	0	0

Tabla 1. Parámetros con mejor ajuste de modelo ARIMA para cada comuna

Para los modelos RNN y LSTM, se observó que la elección del optimizador y la tasa de aprendizaje afectaban en mayor medida los resultados, por este motivo se utilizaron diferentes para cada modelo. Se utilizaron los parámetros fijos del numero de épocas = 90, tamaño de los lotes = 8, Ventaneo = 4, representando la variabilidad mensual y dropout de 0,2. La tabla 2 resume los hiperparámetros utilizados en cada comuna para RNN y LSTM, destacando las combinaciones de optimizador y tasa de aprendizaje que ofrecieron los mejores resultados.

6.3.2. Resultados de Experimentos Los resultados del RMSE para los modelos ARIMA, RNN, LSTM y las variantes de STNN (STNN, STNN-D, STNN-D) en la tabla 3 revelan un desempeño variado en la predicción de patrones delictivos en las diferentes comunas de Bucaramanga. Adicionalmente, la figura 8 representa gráficamente estos resultados, donde se observa que los modelos de RNN y LSTM obtuvieron los mejores desempeños.

Observando los resultados en la tabla, se observa que el modelo LSTM demostró un rendimiento

Comuna	Optimizador	Tasa de aprendizaje
1	Adam	0,000001
2	RMSprop	0,0001
3	Adam	0,000001
4	Adam	0,000001
5	Adam	0,000001
6	Adam	0,0001
7	Adam	0,00001
8	RMSprop	0,00001
9	Adam	0,0001
10	Adadelta	0,00001
11	Adam	0,0001
12	Adadelta	0,01
13	Adam	0,000001
14	Adam	0,0001
15	Adadelta	0,01
16	RMSprop	0,00001
17	Adam	0,000001

Tabla 2. Parámetros de entrenamiento usados para LSTM y RNN

superior en la mayoría de las comunas, con un RMSE promedio más bajo en comparación con los demás modelos. Esto indica que las LSTM, son particularmente eficaces en este contexto. Además, esta solución muestra la menor desviación estándar, lo que sugiere un rendimiento más fiable a través de las distintas comunas.

El modelo RNN también mostró un buen rendimiento, con un RMSE promedio de 2,47, siendo superado por LSTM en la mayoría de las comunas. Por otro lado, el modelo ARIMA, a pesar de ser un modelo más simple, logró resultados similares en su métrica de RMSE en algunas comunas, como la 2, 4, 8, 9, 16, y 17, donde sus valores fueron mejores que los de RNN y comparables a los de LSTM. Esto sugiere que, en ciertas comunas, los patrones delictivos pueden ser capturados adecuadamente por modelos ARIMA estacionales dada su poca variabilidad en el tiempo.

Las variantes de STNN no lograron superar a LSTM o RNN en la mayoría de las comunas. En comunas con mayor variabilidad en los datos, como la 12 y 13, los modelos STNN presentaron los valores de RMSE más altos, lo que podría indicar que la integración de información espacial

Comuna	ARIMA	STNN	STNN-R	STNN-D	RNN	LSTM
1	3,53	3,24	4,38	4,87	2,70	2,57
2	1,32	1,72	1,22	1,35	1,43	1,43
3	4,54	7,63	9,94	11,05	4,54	4,02
4	2,79	2,35	3,37	3,74	2,28	2,23
5	2,39	1,98	2,69	2,99	2,11	2,00
6	3,26	4,38	4,41	4,89	3,14	2,66
7	2,30	1,58	2,74	3,04	1,60	1,57
8	1,42	2,04	1,56	1,73	1,33	1,31
9	1,64	1,83	1,71	1,90	1,53	1,45
10	3,65	2,79	4,74	5,27	2,60	2,82
11	1,95	1,63	1,99	2,22	1,68	1,53
12	4,95	9,17	10,81	12,01	4,13	4,16
13	4,26	11,05	10,31	11,46	4,71	4,16
14	1,74	2,06	1,77	1,96	1,12	1,15
15	4,03	9,21	7,90	8,78	4,06	3,77
16	1,80	1,89	1,90	2,11	1,46	1,49
17	1,73	1,84	1,75	1,94	1,50	1,43
Promedio	2,78	3,91	4,30	4,78	2,47	2,34
Desviación estándar	1,19	3,20	3,33	3,70	1,22	1,09

Tabla 3. Resultados usando la métrica de RMSE para cada modelo

no fue beneficiosa en este caso, esto podría relacionarse al tamaño de las áreas geográficas disponibles para el estudio.

Particularmente, STNN-D y STNN-R, que intentan descubrir relaciones espaciales durante el entrenamiento, no mostraron mejoras notables sobre el STNN estándar, lo que sugiere que la capacidad del modelo para aprender relaciones espaciales de manera autónoma no se tradujo en un mejor rendimiento predictivo.

En la tabla 4 se puede observar la comparación entre el rendimiento de acuerdo a la métrica SMAPE de los distintos modelos utilizados. El modelo LSTM tiene el mejor rendimiento general, con los valores de SMAPE más bajos en la mayoría de las comunas. Por ejemplo, en las comunas 1, 3, 5, y 11 lo que indica una mayor precisión en la predicción. Por otro lado, Aunque es un modelo tradicional, ARIMA tiene un rendimiento relativamente bueno, con valores bajos de SMAPE en varias comunas, como la comuna 2 y la comuna 9. El modelo de RNN también

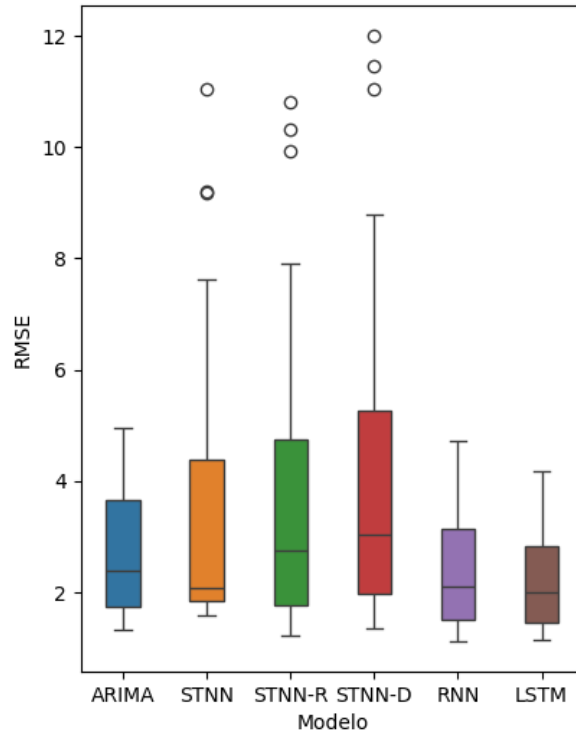


Figura 8. Gráfico de caja con los puntajes RMSE obtenidos para cada modelo. Creación propia

muestra resultados consistentes, aunque generalmente es superado por LSTM.

Finalmente, los modelos de STNN presentan valores de SMAPE significativamente más altos en comparación con otros modelos en la mayoría de las comunas. El modelo STNN-R, diseñado para refinar las relaciones espaciales, tiene las peores predicciones en general, con valores extremadamente altos de SMAPE. Esto de nuevo, sugiere que las relaciones espaciales, al menos tal como están modeladas aquí, no aportan una mejora significativa en la precisión de las predicciones.

En la figura 9 se puede observar una comparación visual del desempeño de los distintos modelos de acuerdo a la métrica SMAPE, se puede observar también el rendimiento superior de los modelos RNN y LSTM.

Con respecto a los tiempos de ejecución, el modelo de RNN obtuvo un tiempo de ejecución de promedio de 140 segundos por comuna, seguido por LSTM con 150 segundos. Los modelos de

Comuna	ARIMA	STNN	STNN-D	STNN-R	RNN	LSTM
1	61,10	67,63	65,89	102,43	41,72	38,61
2	78,07	65,09	59,89	126,87	60,77	60,64
3	25,66	127,54	78,92	91,06	33,54	39,00
4	75,83	57,24	62,93	55,34	47,93	47,54
5	50,09	51,15	65,69	162,22	51,09	51,46
6	33,17	84,65	61,75	146,23	53,33	40,67
7	46,40	47,84	70,62	169,31	55,85	54,76
8	63,51	60,83	78,00	181,91	76,09	75,40
9	67,45	67,51	82,29	173,28	78,04	83,16
10	38,78	75,47	71,86	155,61	47,63	39,98
11	57,29	56,31	66,84	176,11	65,93	60,25
12	29,76	136,19	86,35	97,58	51,02	33,31
13	63,66	133,28	82,80	93,99	25,87	35,02
14	73,71	65,84	71,93	59,60	73,43	70,30
15	23,37	123,34	65,54	77,23	31,57	26,85
16	85,36	50,12	67,31	161,04	55,44	56,76
17	68,49	66,94	78,38	183,80	75,83	68,87
Promedio	55,39	78,65	71,59	130,21	54,42	54,80
Desviación estándar	19,52	30,88	8,11	44,72	15,95	16,24

Tabla 4. Resultados usando la métrica de SMAPE para cada modelo

STNN, tomaron 50 segundos en promedio para entrenar todas las comunas, mientras el modelo de ARIMA tomó 45 segundos por comuna.

6.4. VISUALIZACIÓN DE RESULTADOS

Los modelos RNN y LSTM en las figuras 15 y 16 muestran una capacidad notable para intentar predecir picos y valles en la serie de tiempo. Esto es particularmente importante en el contexto delictivo, donde los picos representan períodos de alta incidencia de robos, y la capacidad de predecir estos eventos puede ser crucial para la planificación en seguridad pública. Sin embargo, aunque estos modelos capturan mejor la dinámica temporal, su éxito en predecir picos es parcial dada la naturaleza del evento a predecir, pues son eventos esporádicos afectados por diferentes variables externas que no se pueden predecir, como distintas políticas de seguridad o eventos

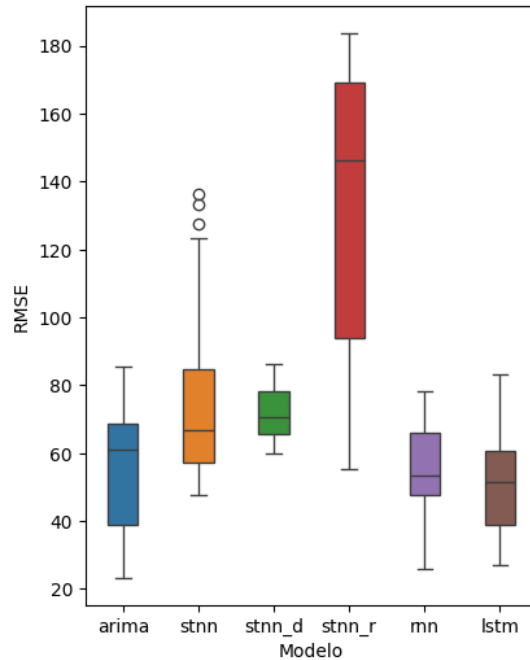


Figura 9. Gráfico de caja con los puntajes SMAPE obtenidos para cada modelo. Creación propia

extracurriculares ⁵⁵. Aún así, su capacidad para reflejar las tendencias de los datos los convierte en herramientas valiosas para la predicción a corto plazo y para la identificación de tendencias fluctuantes con el fin de mejorar la planificación de las distintas organizaciones de control en la ciudad.

En contraste, en el modelo ARIMA en la figura 11 se observa que no logra capturar la variabilidad y los picos presentes en los datos; se puede ver que las predicciones se limitan a una línea recta, lo que indica que la predicción depende de los valores pasados más recientes. Este comportamiento hace que este modelo no sea adecuado para escenarios donde las fluctuaciones y los picos son comunes, ya que su simplicidad en la predicción puede no proporcionar la información necesaria para intervenciones efectivas. Algunos ejemplos de comunas en las que se ve

⁵⁵ Nurul Hazwani Mohd Shamsuddin, Nor Azizah Ali y Razana Alwee. “An overview on crime prediction methods”. En: *2017 6th ICT International Student Project Conference (ICT-ISPC)*. 2017, págs. 1-5. DOI: 10.1109/ICT-ISPC.2017.8075335.

esta situación son las comunas 1, 3, 5, 10, 12, 13 y 15.

Las variantes de STNN (incluyendo STNN, STNN-d, y STNN-r) en las figuras 12, 13, 14 muestran un desempeño similar. Estos modelos no logran predecir los picos de manera efectiva, lo que sugiere que la integración de la información espacial no es necesaria para mejorar la detección picos. Sin embargo, una ventaja que presentan estos modelos es su capacidad para identificar algunas líneas de tendencias, específicamente en el modelo STNN. Aunque no son capaces de capturar el comportamiento de altas y bajas con precisión, su capacidad para estimar la media general de la serie temporal puede ser útil para predicciones a mediano y largo plazo.

El modelo STNN-d, fué utilizado para descubrir relaciones espaciales de manera autónoma durante el proceso de entrenamiento. Sin embargo, al finalizar el entrenamiento, el modelo no encontró ninguna nueva relación espacial entre las comunas, sugiriendo que no encontró evidencia suficiente para establecer nuevas conexiones significativas espacialmente. Este resultado puede indicar que, según los datos disponibles y la estructura del modelo, las interacciones espaciales no son variables explicativas para el número de robos en Bucaramanga. Esto se sugiere en necesidad de explorar otros enfoques o ajustes como es el caso de los modelos utilizados de RNN y LSTM. Por otro lado, el modelo STNN-R mostró resultados de altas variaciones a la hora de modelar los datos, producto de la poca precisión del modelo a la hora de predecir muchos puntos en una serie.

Adicionalmente, en la figura 10 se pueden observar las nuevas relaciones espaciales descubiertas por el modelo STNN-D a partir de los resultados obtenidos en sus matrices de adyacencia. Estas relaciones permiten identificar patrones subyacentes que no fueron definidos explícitamente en la matriz inicial, lo que sugiere que los modelos han sido capaces de detectar conexiones y similitudes en la ocurrencia de delitos entre diferentes comunas. A partir de estas nuevas relaciones, se puede definir una serie de subconjuntos dentro del área de Bucaramanga que comparten ciertas características y fenómenos delictivos en común. Estos grupos reflejan cómo ciertas comunas experimentan patrones delictivos similares debido a factores compartidos como la proximidad a zonas comerciales, vías de acceso, o características socioeconómicas similares. Estos grupos

son: El subgrupo 1 compuesto por las comunas 1 y 2 las cuales se encuentran aisladas del resto de la ciudad por el final de la meseta de Bucaramanga; subgrupo 2 compuesto las comunas 3 y 4, ubicadas en el centro de la ciudad; subgrupo 3 compuesto las comunas 5,6,7,8,15 y 17, ubicadas en la región occidental; subgrupo 4 compuesto por las comunas 10 y 11, ubicadas fuera de la meseta de Bucaramanga y separadas por la Puerta del Sol; subgrupo 5 compuesto por las comunas 9 , 16 y el sur de la comuna 12 separadas por la autopista a Floridablanca y la puerta del Sol; y el subgrupo 6 compuesto por el norte de la comuna 12 con las comunas 13 y 14, donde suelen habitar las clases altas de la ciudad.

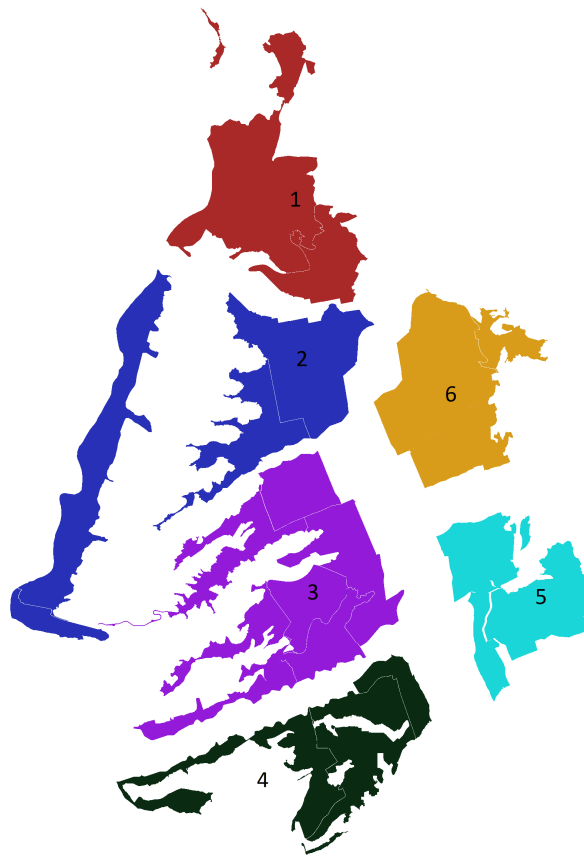


Figura 10. Visualización de las relaciones encontradas por el modelo STNN-R. Creación propia

Este análisis sugiere que los modelos STNN, a pesar de no predecir picos de robos con precisión, son capaces de identificar estructuras espaciales complejas que pueden ser clave para el control y

la prevención del crimen a nivel urbano. Estas nuevas agrupaciones pueden servir como base para futuros estudios y para la implementación de políticas públicas más dirigidas, que consideren las interacciones entre comunas que aborden los problemas de seguridad de manera más holística y coordinada.

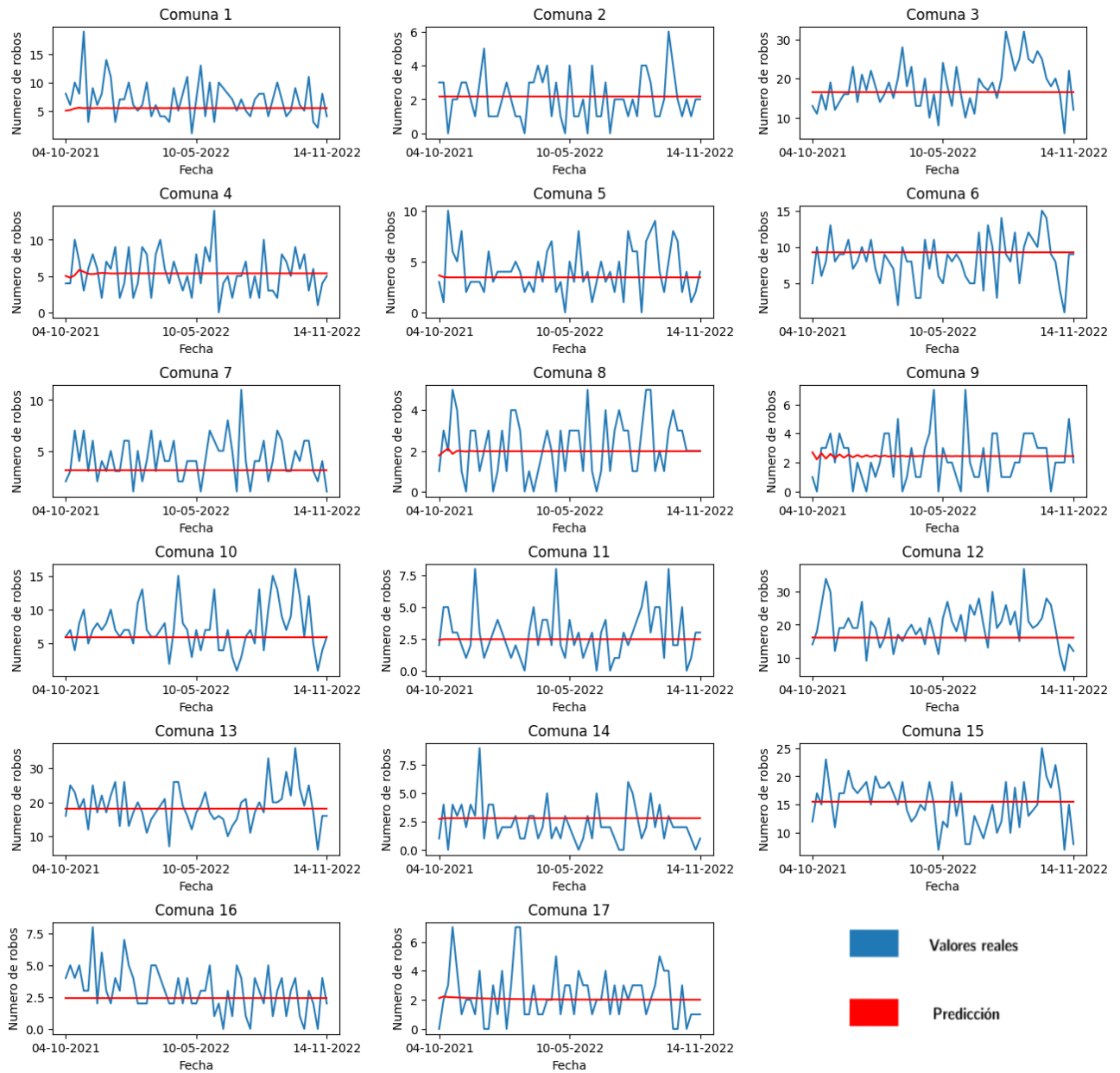


Figura 11. Visualización de las predicciones realizadas por el modelo ARIMA. Creación propia

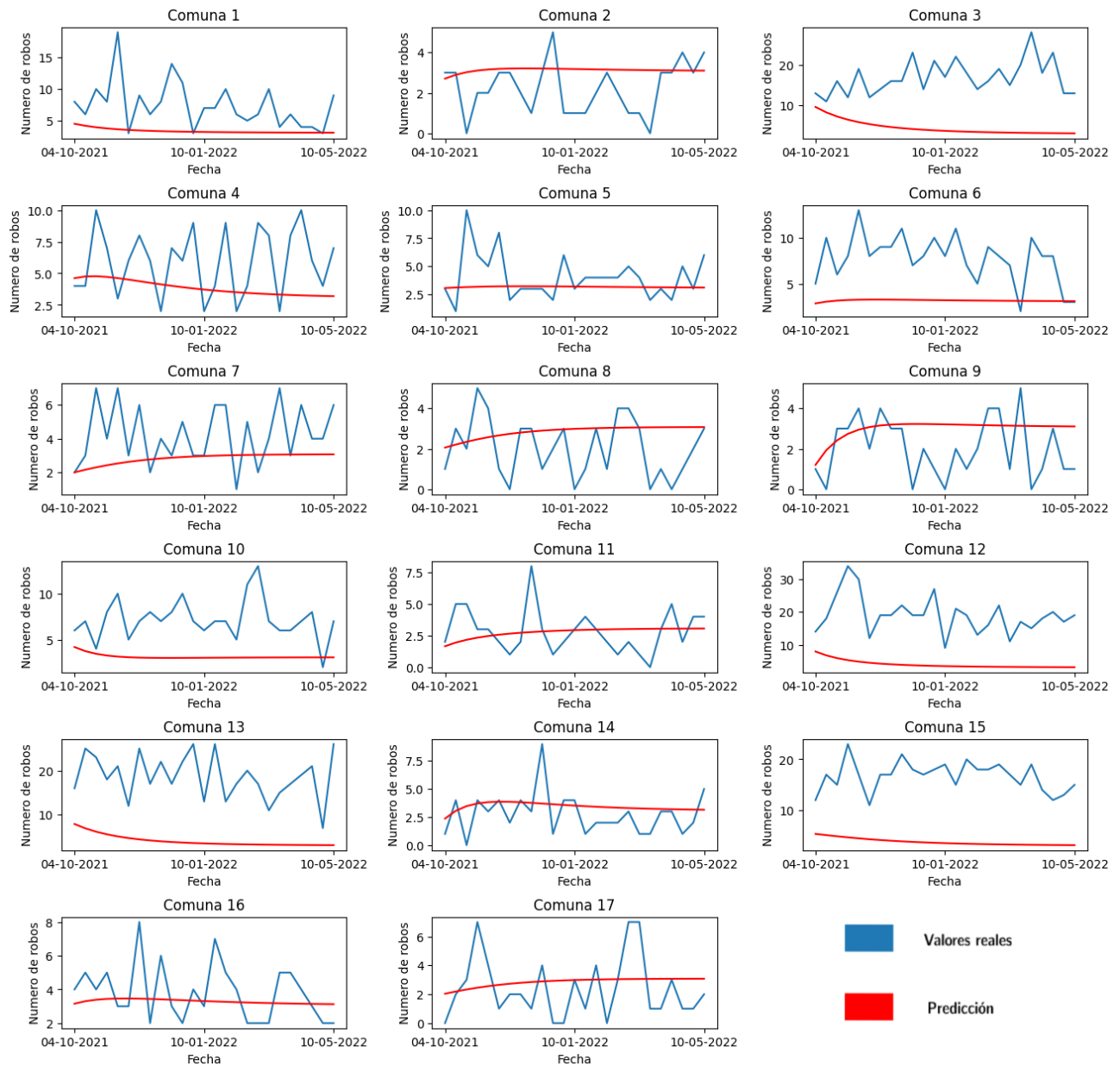


Figura 12. Visualización de las predicciones realizadas por el modelo STNN. Creación propia

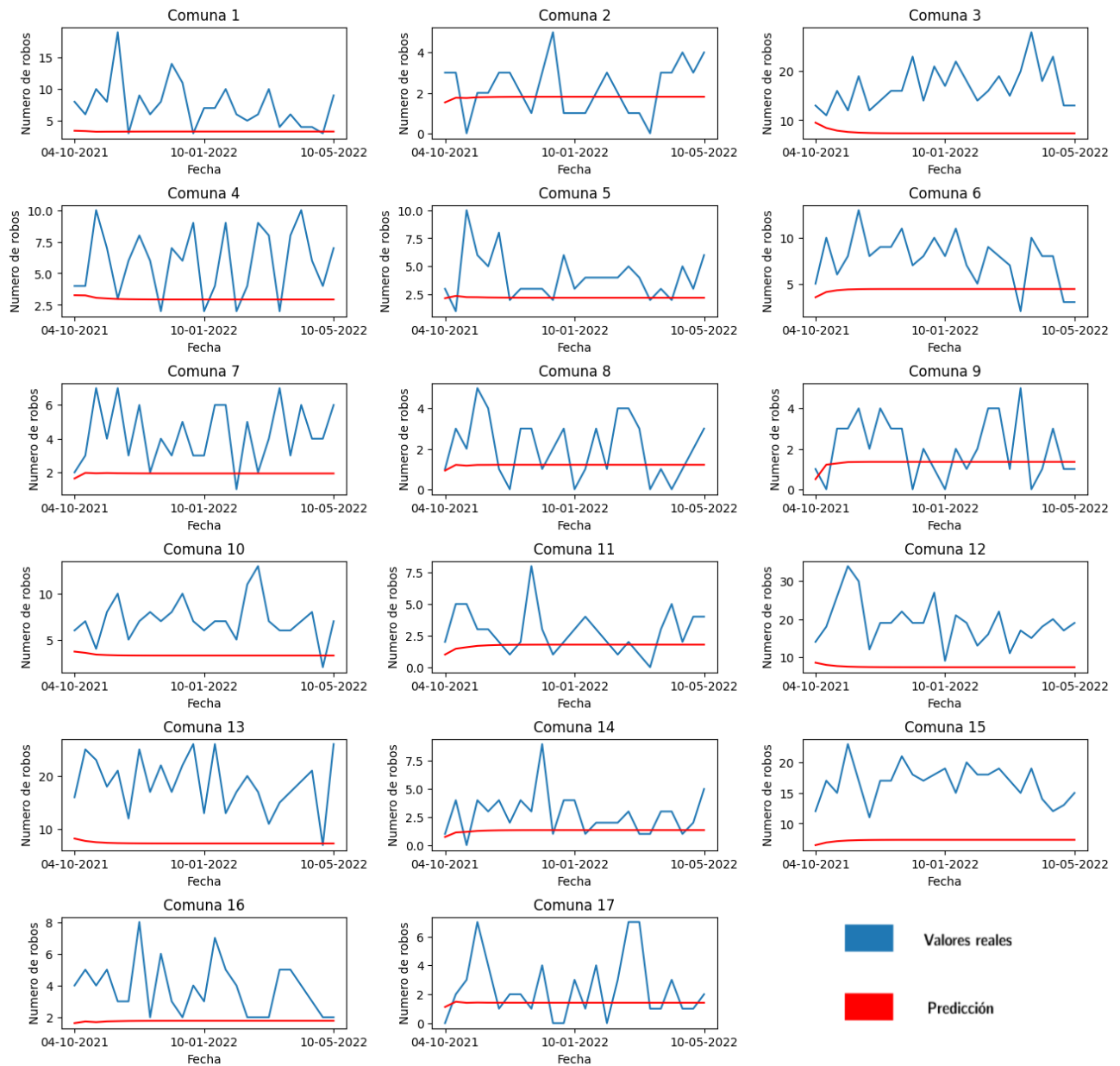


Figura 13. Visualización de las predicciones realizadas por el modelo STNN-D. Creación propia

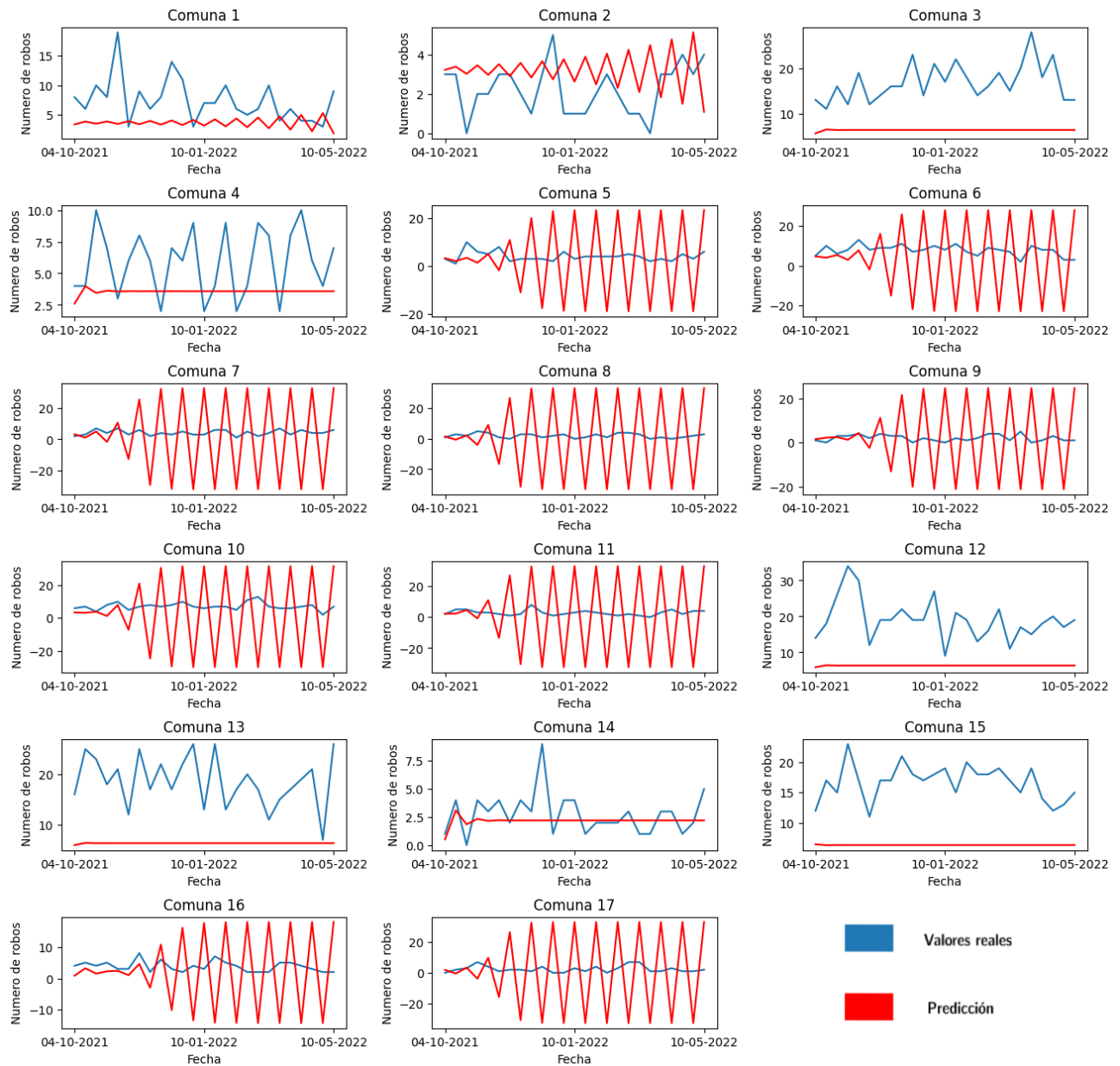


Figura 14. Visualización de las predicciones realizadas por el modelo STNN-R. Creación propia

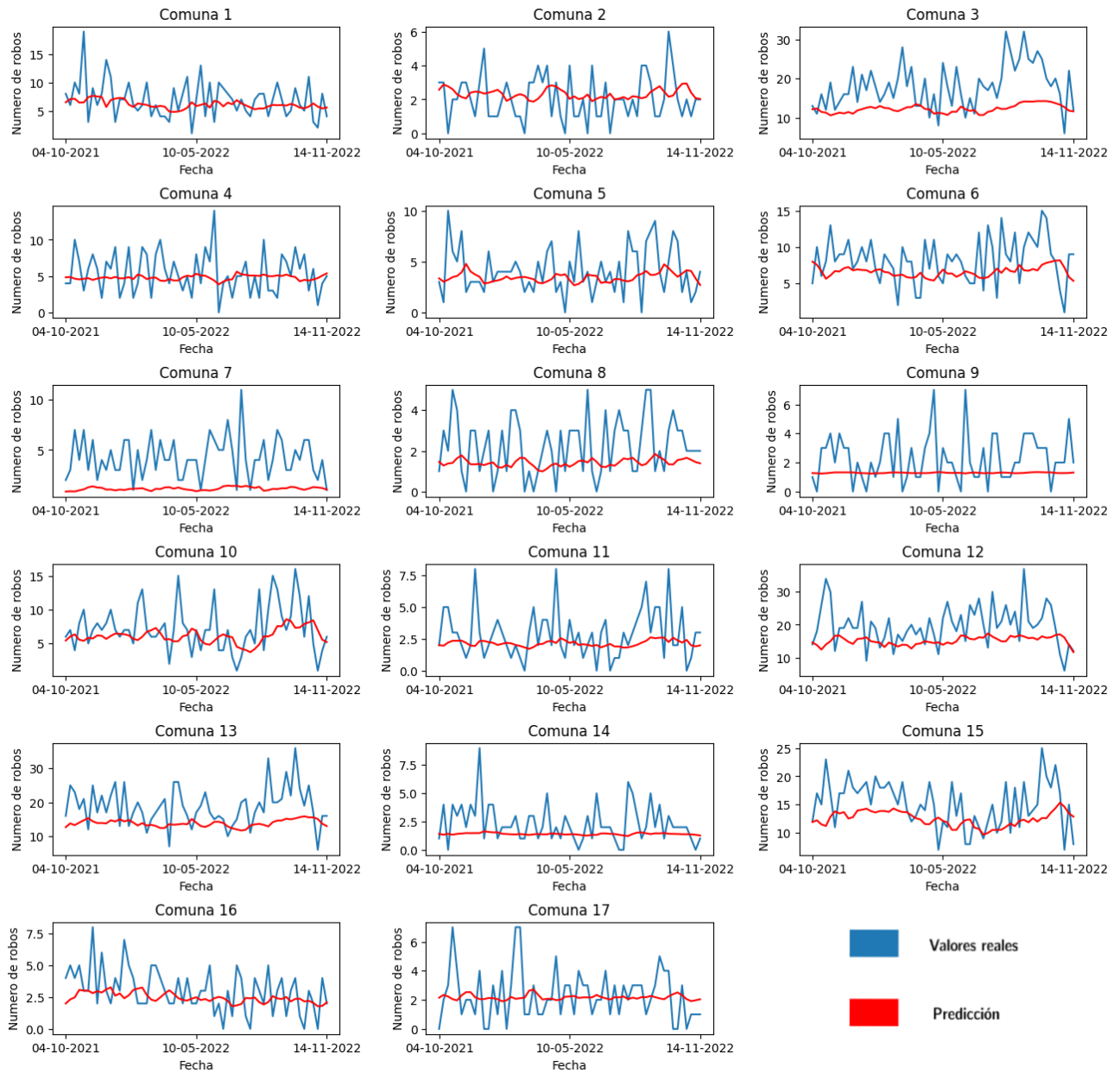


Figura 15. Visualización de las predicciones realizadas por el modelo RNN. Creación propia

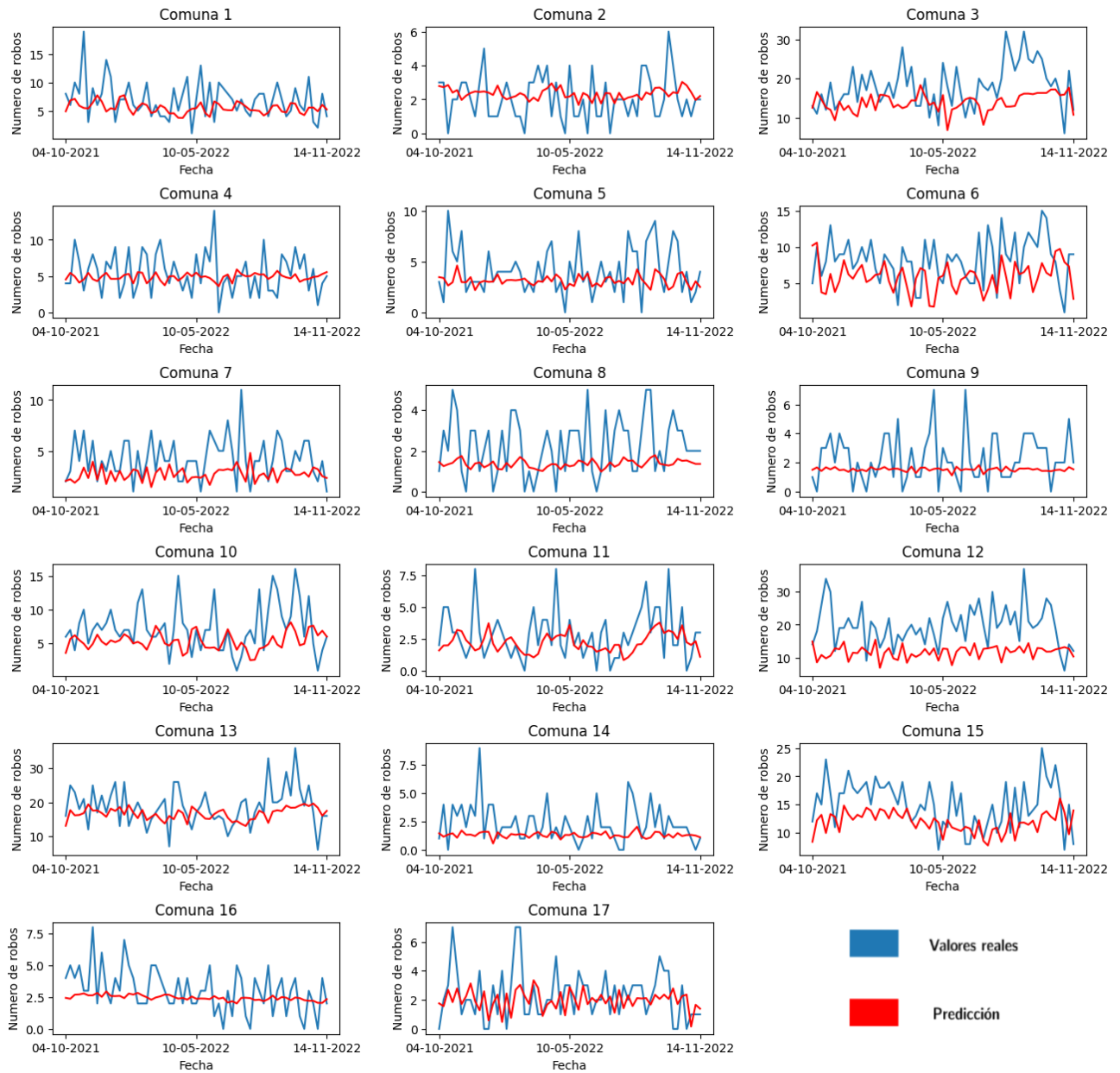


Figura 16. Visualización de las predicciones realizadas por el modelo LSTM. Creación propia

7. CONCLUSIONES Y PERSPECTIVAS

Este estudio evaluó la efectividad de distintos enfoques predictivos, incluyendo modelos tradicionales como ARIMA, redes neuronales recurrentes (RNN), redes de gran memoria a corto plazo (LSTM) y modelos espacio-temporales (STNN), en la predicción de patrones de robos en Bucaramanga. Los resultados demostraron que los modelos RNN y LSTM son más eficaces para capturar los picos y valles y predecir picos de actividad delictiva, superando a ARIMA en términos de precisión, con LSTM destacándose por su capacidad para manejar dependencias a largo plazo.

El uso de técnicas como el ventaneo temporal y la normalización de datos fue crucial para obtener buenos resultados. El ventaneo permitió capturar mejor las dinámicas temporales al dividir la serie de tiempo en segmentos manejables, mientras que la normalización mediante escalado del mínimo y el máximo garantizó que los modelos trabajaran con datos en una misma escala, mejorando la precisión de las predicciones. Además, se realizaron múltiples configuraciones experimentales, ajustando hiperparámetros y probando diversas combinaciones de modelos, para maximizar la extracción de información relevante.

Aunque las variantes de STNN fueron diseñadas para integrar información espacial y temporal, no lograron mejorar significativamente las predicciones. Un hallazgo clave fue que el modelo STNN-d, que comenzó con todas las relaciones en 0, mantuvo estas relaciones sin cambios después del entrenamiento, lo que sugiere que la representación espacial basada en comunas puede ser demasiado amplia para capturar interacciones delictivas relevantes.

Estas observaciones resaltan la importancia de utilizar una escala espacial más precisa en el modelado de relaciones entre comunas. Dado el amplio tamaño del área cubierta por cada comuna, se recomienda colaborar con la fuente de datos, como la alcaldía, para obtener información más detallada. Utilizar coordenadas o direcciones específicas permitiría modelar relaciones espaciales en un área más pequeña, lo que podría revelar patrones más precisos y mejorar significativamente la capacidad predictiva de los modelos STNN. Integrar esta georreferenciación con datos

adicionales, como información socioeconómica y de movilidad, podría enriquecer los modelos espacio-temporales y ofrecer una visión más completa de los factores que influyen en la criminalidad.

En conclusión, el ajuste de los modelos seleccionados es aceptable para el conjunto de datos de patrones delictivos en Bucaramanga, considerando la complejidad del problema y la amplitud del área de estudio. Los resultados obtenidos validan la posibilidad de relacionar variables temporales y geográficas en contextos específicos como la predicción de delitos, reforzando la idea de que con un enfoque adecuado y un análisis detallado, es posible integrar dimensiones espaciales y temporales para mejorar la comprensión y anticipación de fenómenos complejos como la criminalidad urbana.

BIBLIOGRAFÍA

- Al-Selwi, Safwan Mahmood, Mohd Fadzil Hassan, Said Jadid Abdulkadir et al. “RNN-LSTM: From applications to modeling techniques and beyond—Systematic review”. En: *Journal of King Saud University-Computer and Information Sciences* (2024), pág. 102068 (vid. pág. 23).
- Alcaldía de Bucaramanga. *División político urbana y rural*. Alcaldía de Bucaramanga. 30 de ago. de 2024. URL: <https://www.bucaramanga.gov.co/division-politico-urbana/> (visitado 30-08-2024) (vid. pág. 36).
- Alcaldía de Bogotá. *Bogotá desarrollará un método de predicción de delitos*. 25 de jun. de 2024. URL: <https://bogota.gov.co/mi-ciudad/seguridad/nuevo-metodo-de-prediccion-de-delitos-sera-desarrollado-en-bogota> (visitado 12-02-2019) (vid. pág. 32).
- Andresen, Martin A. “Estimating the probability of local crime clusters: The impact of immediate spatial neighbors”. En: *Journal of Criminal Justice* 39.5 (2011), págs. 394-404. DOI: <https://doi.org/10.1016/j.jcrimjus.2011.05.005> (vid. pág. 33).
- Bhanja, Samit y Abhishek Das. *Impact of Data Normalization on Deep Neural Network for Time Series Forecasting*. 2019. arXiv: 1812.05519 [cs.LG] (vid. pág. 49).
- Box, George E. P. et al. *Time Series Analysis: Forecasting and Control, 5th Edition*. Wiley, 2015 (vid. pág. 18).
- Braga, Anthony A., Andrew V. Papachristos y David M. Hureau. “The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis”. En: *Justice Quar-*

terly 31.4 (2014), págs. 633-663. DOI: 10.1080/07418825.2012.673632. eprint: <https://doi.org/10.1080/07418825.2012.673632> (vid. págs. 15, 32).

Brockwell, Peter J y Richard A Davis. *Time series: theory and methods*. Springer science & business media, 1991 (vid. págs. 17, 18).

Bucaramanga como vamos. 25 de jun. de 2024. URL: <https://www.bucaramangacomovamos.org/post/informe-de-calidad-de-vida-2022> (visitado 27-09-2022) (vid. pág. 12).

Cade, Brian S., David R. Edmunds y Douglas S. Ouren. “Quantile regression estimates of animal population trends”. En: *The Journal of Wildlife Management* 86.5 (2022), e22228. DOI: <https://doi.org/10.1002/jwmg.22228>. eprint: <https://wildlife.onlinelibrary.wiley.com/doi/pdf/10.1002/jwmg.22228> (vid. pág. 16).

Carmona-Cabezas, Rafael et al. “A sliding window-based algorithm for faster transformation of time series into complex networks”. En: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.10 (oct. de 2019), pág. 103121. DOI: 10.1063/1.5112782. eprint: https://pubs.aip.org/aip/cha/article-pdf/doi/10.1063/1.5112782/16714740/103121\1_online.pdf (vid. pág. 45).

Cavadas, Bruno, Paula Branco y Sérgio Pereira. “Crime Prediction Using Regression and Resources Optimization”. En: *Progress in Artificial Intelligence*. Ed. por Francisco Pereira et al. Cham: Springer International Publishing, 2015, págs. 513-524 (vid. pág. 16).

Departamento Administrativo Nacional de Estadística (DANE). *Censo Nacional de Población y Vivienda 2018, DANE*. 10 de mayo de 2019 (vid. pág. 36).

Devi, B Uma, Darshan Sundar y P Alli. “An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50”. En: *International Journal of Data Mining & Knowledge Management Process* 3.1 (2013), pág. 65 (vid. pág. 18).

- Dietterich, Thomas G. “Machine Learning for Sequential Data: A Review”. En: *Structural, Syntactic, and Statistical Pattern Recognition*. Ed. por Terry Caelli et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, págs. 15-30 (vid. pág. 13).
- Embrechts, Paul, Rüdiger Frey y Hansjörg Furrer. “Stochastic processes in insurance and finance”. En: *Stochastic Processes: Theory and Methods*. Vol. 19. Handbook of Statistics. Elsevier, 2001, págs. 365-412. DOI: [https://doi.org/10.1016/S0169-7161\(01\)19014-0](https://doi.org/10.1016/S0169-7161(01)19014-0) (vid. pág. 17).
- Gers, F.A., J. Schmidhuber y F. Cummins. “Learning to forget: continual prediction with LSTM”. En: *1999 Ninth International Conference on Artificial Neural Networks ICANN99*. Vol. 2. 1999, 850-855 vol.2. DOI: [10.1049/cp:19991218](https://doi.org/10.1049/cp:19991218) (vid. pág. 25).
- Grimmet Geoffrey, Stirzaker David. *Probability and Random Processes*. Oxford University Press, 2001 (vid. pág. 17).
- Hewamalage, Hansika, Christoph Bergmeir y Kasun Bandara. “Recurrent Neural Networks for Time Series Forecasting: Current status and future directions”. En: *International Journal of Forecasting* 37.1 (2021), págs. 388-427. DOI: <https://doi.org/10.1016/j.ijforecast.2020.06.008> (vid. pág. 23).
- Hyndman, Rob J y George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018 (vid. pág. 18).
- Jiang, Nan et al. “Spatio-temporal prediction of crime based on Data Mining and LSTM network”. En: *2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. Vol. 6. 2023, págs. 672-676. DOI: [10.1109/ITNEC56291.2023.10081985](https://doi.org/10.1109/ITNEC56291.2023.10081985) (vid. pág. 30).

- Kibekbaev, Azamat y Ekrem Duman. “Benchmarking regression algorithms for income prediction modeling”. En: *Information Systems* 61 (2016), págs. 40-52. DOI: <https://doi.org/10.1016/j.is.2016.05.001> (vid. pág. 16).
- Kwon, Eunseo, Sungwon Jung y Jaewook Lee. “Artificial Neural Network Model Development to Predict Theft Types in Consideration of Environmental Factors”. En: *ISPRS International Journal of Geo-Information* 10.2 (2021). DOI: 10.3390/ijgi10020099 (vid. págs. 15, 22).
- Li, Jia et al. “Hyperbolic linear units for deep convolutional neural networks”. En: *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2016, págs. 353-359 (vid. pág. 24).
- Lipton, Zachary C., John Berkowitz y Charles Elkan. *A Critical Review of Recurrent Neural Networks for Sequence Learning*. 2015. arXiv: 1506.00019 [cs.LG] (vid. pág. 23).
- Mahadzirah Mohamad Morliyati Mohammad, Nor Azman Mat Ali y Zainudin Awang. “The impact of life satisfaction on substance abuse: delinquency as a mediator”. En: *International Journal of Adolescence and Youth* 23.1 (2018), págs. 25-35. DOI: 10.1080/02673843.2016.1267021. eprint: <https://doi.org/10.1080/02673843.2016.1267021> (vid. pág. 32).
- “Mean Squared Error”. En: *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, págs. 337-339. DOI: 10.1007/978-0-387-32833-1_251 (vid. pág. 48).
- Mesa Velásquez, Luis Eduardo. “Los delitos de hurto y robo, con referencia al Código Penal Colombiano”. En: *Estudios de Derecho* 18.56 (1959), 129–158. DOI: 10.17533/udea.esde.333074 (vid. pág. 14).
- Nketiah, Edward Appau et al. “Recurrent neural network modeling of multivariate time series and its application in temperature forecasting”. En: *PLOS ONE* 18.5 (mayo de 2023), págs. 1-18. DOI: 10.1371/journal.pone.0285713 (vid. pág. 23).

Papoulis, Athanasios y Unnikrishna Pillai. *Probability, Random Variables, and Stochastic Processes, Fourth Edition*. McGraw-Hill, 2002 (vid. pág. 16).

Pavan Naidu Kavala, Sri Adi et al. “Crime analysis and Prediction using Neural Networks”. En: *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Vol. 1. 2022, págs. 2047-2050. DOI: 10.1109/ICACCS54159.2022.9785081 (vid. pág. 33).

Resnick, Sidney I. *Adventures in stochastic processes*. Springer Science & Business Media, 2013 (vid. pág. 19).

Sayal, Anu et al. “Crime Detection Using Data Mining Techniques”. En: *2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT)*. 2024, págs. 200-204. DOI: 10.1109/CCICT62777.2024.00041 (vid. pág. 12).

Se agudiza la criminalidad en Colombia: cada día más de mil personas son víctimas de hurtos y/o extorsión. 25 de jun. de 2024. URL: <https://cej.org.co/destacados-home-page/se-agudiza-la-criminalidad-en-colombia-cada-dia-mas-de-mil-personas-son-victimas-de-hurtos-y-o-extorsion/> (visitado 21-09-2023) (vid. pág. 38).

Shamsuddin, Nurul Hazwani Mohd, Nor Azizah Ali y Razana Alwee. “An overview on crime prediction methods”. En: *2017 6th ICT International Student Project Conference (ICT-ISPC)*. 2017, págs. 1-5. DOI: 10.1109/ICT-ISPC.2017.8075335 (vid. pág. 57).

Specht, Donald F et al. “A general regression neural network”. En: *IEEE transactions on neural networks* 2.6 (1991), págs. 568-576 (vid. pág. 15).

Vargas Valencia, Jorge Luis et al. “Criminalidad en contexto COVID año 2020 y aproximación de una propuesta en neurocriminología”. En: *Revista Criminalidad* 63.3 (2022), 9-31. DOI: 10.47741/17943108.310 (vid. pág. 38).

- Wang, Ping et al. “Predicting Criminal Recidivism with Support Vector Machine”. En: *2010 International Conference on Management and Service Science*. 2010, págs. 1-9. DOI: 10.1109/ICMSS.2010.5575352 (vid. pág. 33).
- Wang, Shihuai y Kunxiaoja Yuan. “Spatiotemporal Analysis and Prediction of Crime Events in Atlanta Using Deep Learning”. En: *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*. 2019, págs. 346-350 (vid. págs. 30, 43).
- Yadav, Romika y Savita Kumari Sheoran. “Crime Prediction Using Auto Regression Techniques for Time Series Data”. En: *2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*. 2018, págs. 1-5. DOI: 10.1109/ICRAIE.2018.8710407 (vid. págs. 16, 33).
- Yadav, Sunil et al. “Crime pattern detection, analysis & prediction”. En: *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*. Vol. 1. 2017, págs. 225-230. DOI: 10.1109/ICECA.2017.8203676 (vid. pág. 15).
- Zambrano, Rafael. “Un enfoque Espacio Temporal para la predicción de delitos en la ciudad de Buenos Aires”. En: *Revista de Investigación de Modelos Matemáticos Aplicados a la Gestión y a la Economía*. Vol. 2. 2022, págs. 38-62 (vid. pág. 31).
- Zhang, Xu et al. “Comparison of Machine Learning Algorithms for Predicting Crime Hotspots”. En: *IEEE Access* 8 (2020), págs. 181302-181310. DOI: 10.1109/ACCESS.2020.3028420 (vid. pág. 32).
- Zhang, Zhihua. “Artificial Neural Network”. En: *Multivariate Time Series Analysis in Climate and Environmental Research*. Cham: Springer International Publishing, 2018, págs. 1-35. DOI: 10.1007/978-3-319-67340-0_1 (vid. pág. 20).

Ziat, Ali et al. “Spatio-Temporal Neural Networks for Space-Time Series Forecasting and Relations Discovery”. En: *2017 IEEE International Conference on Data Mining (ICDM)*. 2017, págs. 705-714. DOI: 10.1109/ICDM.2017.80 (vid. pág. 27).

Área Metropolitana de Bucaramanga. *Bucaramanga*. AMB. 10 de abr. de 2012. URL: <https://www.amb.gov.co/bucaramanga/> (visitado 31-08-2023) (vid. pág. 35).